

Programmable RNA-binding protein composed of repeats of a single modular unit

 Katarzyna P. Adamala^{a,1}, Daniel A. Martin-Alarcon^{b,1}, and Edward S. Boyden^{a,b,c,2}
^aMedia Lab, Massachusetts Institute of Technology, Cambridge, MA 02139; ^bDepartment of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139; and ^cMcGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139

Edited by Robert H. Singer, Albert Einstein College of Medicine of Yeshiva Uni, Bronx, NY, and approved March 30, 2016 (received for review September 29, 2015)

The ability to monitor and perturb RNAs in living cells would benefit greatly from a modular protein architecture that targets unmodified RNA sequences in a programmable way. We report that the RNA-binding protein PumHD (Pumilio homology domain), which has been widely used in native and modified form for targeting RNA, can be engineered to yield a set of four canonical protein modules, each of which targets one RNA base. These modules (which we call Pumby, for Pumilio-based assembly) can be concatenated in chains of varying composition and length, to bind desired target RNAs. The specificity of such Pumby-RNA interactions was high, with undetectable binding of a Pumby chain to RNA sequences that bear three or more mismatches from the target sequence. We validate that the Pumby architecture can perform RNA-directed protein assembly and enhancement of translation of RNAs. We further demonstrate a new use of such RNA-binding proteins, measurement of RNA translation in living cells. Pumby may prove useful for many applications in the measurement, manipulation, and biotechnological utilization of unmodified RNAs in intact cells and systems.

RNA-binding protein | Pumilio | gene expression monitoring | protein engineering | translation initiation

Many scientific questions and bioengineering goals relate to the monitoring and control of RNA functions in living cells. A powerful strategy is to modify a target RNA by inserting an exogenous sequence such as MS2 or PP7, so that the corresponding RNA-binding protein can deliver a reporter or RNA modification enzyme to an RNA of interest (1–3). Ideally one could target unmodified RNA, both for simplicity and to preserve as much native RNA structure and function as possible (4, 5). It has been proposed that proteins such as the *Caenorhabditis elegans* Puf (6), the human PumHD (Pumilio homology domain) (7), or members of the pentatricopeptide family (8) could serve such a purpose. Each of these proteins is made of many similar units, each of which targets one RNA base. The most extensively studied protein architecture, in the context of single-stranded RNA targeting in mammalian cells, is the human PumHD (9–12). PumHD is a protein with 10 units, of which 8 units bind to the bases of an eight-nucleobase target RNA sequence (Fig. 1*A*), called the Nanos response element (NRE), in the reverse orientation 3'-AUAUAUGU-5' (Fig. 1*B*) (13–19). X-ray structures of the PumHD–NRE complex indicate that three key amino acids interact with each RNA nucleobase (14, 20).

A number of pioneering studies have shown that modifications of the wild-type PumHD can indeed bind to many sequences other than the NRE (summarized in *SI Appendix, Table S18*), strongly pointing toward the modularity of PumHD (we here use the shorthand “Pum” to denote any protein homologous to, or derived from, PumHD). We set out to determine whether, given the rich set of previous findings related to Pum proteins, we could devise a set of four canonical protein modules, each of which targets one RNA base with high specificity and could be concatenated in chains of varying composition and length so as to bind desired target RNAs. A similar protein architecture, the transcription activator-like (TAL) effector, has been rendered in this single-module

form and has proven to be useful for targeting DNA with various proteins, because of its modularity (21, 22). There are four canonical TALE protein modules, each of which targets one DNA base with high specificity. If analogous Pum modules could be developed, they could be easily designed and used: Simply concatenate a chain of modules according to the sequence of a natural target RNA, and then the protein (perhaps equipped with various reporters and effectors) could be targeted to a desired RNA.

Results

A number of studies have mutated different units of PumHD to target different bases, testing various mutations in various cell-free or cellular contexts. Eleven of these studies used mammalian cells to explore 19 out of the 24 possible mutant units (i.e., three different bases at eight different sites; *SI Appendix, Table S18*). Because no single study tested PumHD variants binding to all four nucleotides at each unit's position, in the same condition, we first assessed whether all 24 PumHD single-unit mutants could target their respective 8-nt sequences. We used an assay commonly used in Pumilio evaluation and also useful in cell biology—RNA-based GFP complementation in mammalian cells (Fig. 2*A* and *B*). This assay is sometimes seen as qualitative, because it does not indicate actual binding affinities, but it has proven useful in the study of RNA-binding proteins such as Pumilio because it allows for such interactions to be measured at a functional level in living cells (23–26). In particular, split fluorescent protein reconstitution was used to test on-target binding of three different Pum variants to NRE variants, and also previously used to visualize binding of PumHD variants to the mRNAs for human β -actin and NADH dehydrogenase subunit 6 (23–26). Based upon earlier literature,

Significance

The ability to monitor and perturb RNAs in living cells would benefit greatly from a protein architecture that targets RNA sequences in a programmable way. We report four protein building blocks, which we call Pumby modules, each of which targets one RNA base and can be concatenated in chains of varying composition and length. The Pumby building blocks will open up many frontiers in the measurement, manipulation, and biotechnological utilization of unmodified RNAs in intact cells and systems.

Author contributions: K.P.A., D.A.M.-A., and E.S.B. designed research; K.P.A. and D.A.M.-A. performed research; K.P.A. and D.A.M.-A. contributed new reagents/analytic tools; K.P.A., D.A.M.-A., and E.S.B. analyzed data; and K.P.A., D.A.M.-A., and E.S.B. wrote the paper.

Conflict of interest statement: The authors applied for a patent that is owned by Massachusetts Institute of Technology.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. KU900022–KU900031).

¹K.P.A. and D.A.M.-A. contributed equally to this work.

²To whom correspondence should be addressed. Email: esb@media.mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1519368113/-DCSupplemental.

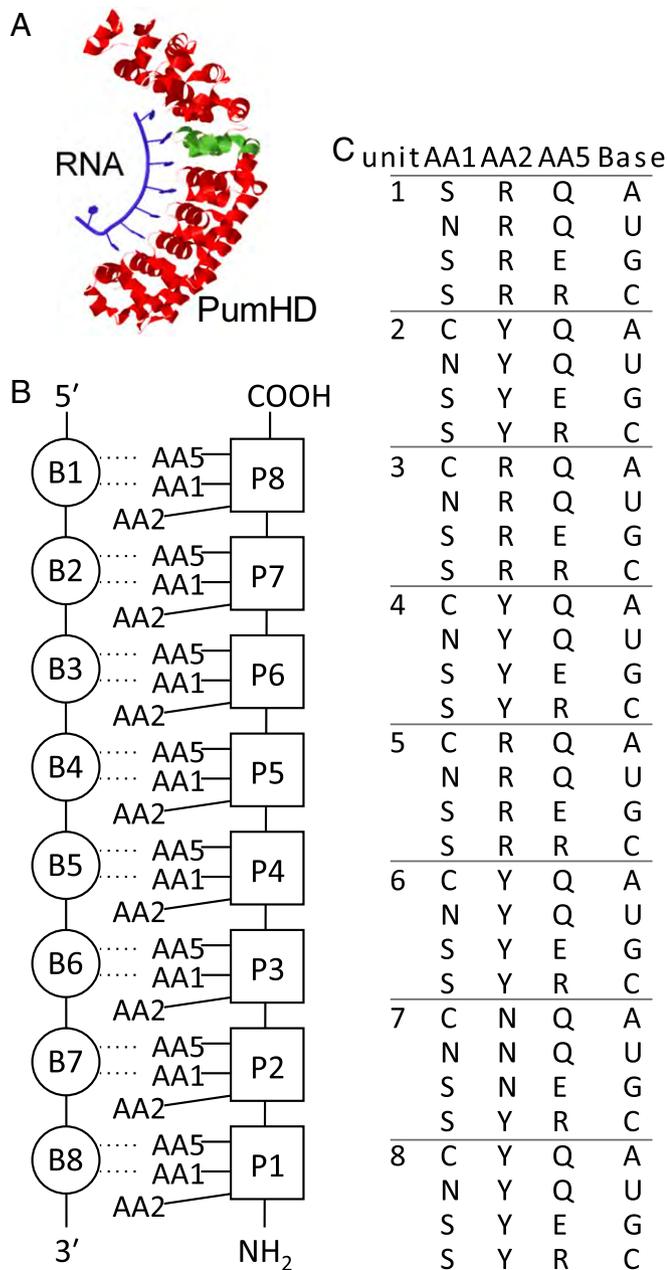


Fig. 1. A proposed amino acid code for universal PumHD binding. (A) Crystal structure of the wild-type human PumHD (red) with its cognate RNA (blue). One protein unit is highlighted in green; data are from PDB ID code 1M8X (14). (B) Schematic representation of RNA bases (labeled B1 to B8) and their respective PumHD protein units (labeled P8 to P1). Note the binding direction: The carboxyl terminus of the Pum protein binds to the 5' end of the target RNA. Three amino acids (labeled AA1, AA2, and AA5) are key for recognizing the target nucleobases. (C) A proposed consensus sequence of key amino acids that allow PumHD to bind RNA sequences that have any base at any position in the eight-base target sequence.

we hypothesized a consensus sequence for how to modify each unit of PumHD so that its base preference could be tuned to any of the four RNA bases (Fig. 1C). We adapted—from the TAL effector field—a Golden Gate assembly method to rapidly create PumHD variants (*SI Appendix, Fig. S1 and Supplementary Results*).

We used a reference PumHD variant [used in a prior fluorescent protein reconstitution study (23), and that binds 3'-AUA-GAUGU-5'] to assess the efficacy of our hypothesized consensus sequence (Fig. 2F). Throughout our experiments we used two

PumHD proteins (a variable Pum, denoted Pum1, and a wild-type, denoted Pum2), each fused to one part of a split GFP, which bind two adjacent sequences before the stop codon of a transcript that codes for mRuby (Fig. 2B). Pum1 binding was assessed with on- as well as off-target RNA sequences (for off-target cases, purines were swapped with pyrimidines at all eight positions; see *SI Appendix, Table S2* for all sequences used). We found that on-target RNA sequences supported effective Pum1 binding and green fluorescence, whereas off-target RNA sequences did not (individual examples, Fig. 2C–E; population data, Fig. 2F; $P < 0.0001$ for factor of on- vs. off-target; two-way ANOVA with factors of on- vs. off-target and target sequence; $n = 3$ biological replicates each; for full statistics for Fig. 2, refer to *SI Appendix, Table S1*). All 24 PumHD variants had binding preferences for on-target vs. off-target sequences that were indistinguishable from the wild type ($P > 0.05$, Dunnett's post hoc test comparing target sequence vs. wild-type, for the ANOVA above). Thus, as expected given the prior literature, PumHD variants can indeed support any unit targeting any base.

To explore the robustness of PumHD binding in varying contexts, we tested binding of the wild-type PumHD with varying bases upstream and downstream of the NRE (Fig. 2G) and found successful binding, albeit with statistically significant differences in GFP reconstitution from one set of bases to another (statistics in *SI Appendix, Table S1*). Given that any protein–RNA interaction will be susceptible to environmental changes or RNA secondary structure arising from the specific sequences involved, this result suggests that PumHD variants should be vetted on a per-case basis. However, PumHD variants were generally capable of binding their target regardless of the bases immediately upstream and downstream of the core eight bases.

To assess the specificity of PumHD mutants for target sequence even more quantitatively, we assessed binding of two different PumHD variants to RNA targets that were on-target vs. those off-target at one, two, or three specific bases (see *SI Appendix, Table S2* for all of the sequences used), using the GFP reconstitution method described above. Although some Pum-mediated GFP reconstitution was observed for RNA targets off by two bases, RNA targets off by three bases did not support GFP reconstitution any more than did completely different (i.e., off by eight bases) RNA sequences (Fig. 2H; $P = 0.9999$ for comparison of three vs. eight mismatches; Dunnett's post hoc test for the factor of mismatch number, after a two-way ANOVA with factors of mismatch number and Pum identity; $n = 3$ biological replicates; see *SI Appendix, Table S1* for full statistics).

We then set out to make a set of four canonical protein modules, each of which targets one RNA base with high specificity (Fig. 3A). As in Fig. 2, we tested binding for both on-target and off-target Pum pairs in live mammalian cells, using GFP reconstitution. For simplicity we kept AA2 (the “stacking” amino acid) the same for all four modules in our design. Because most of the PumHD units of Fig. 1C had either Y or R for AA2, we decided not to use unit 7, which mostly used N. Then, we examined which units had been most thoroughly mutated by the most groups (*SI Appendix, Table S18*) and thus had been the most vetted in a variety of contexts and chose units 3 and 6 of PumHD as candidates for a Pumby module starting material. We screened variants of units 3 and 6 using the GFP reconstitution assay (see *SI Appendix, Table S21* for all of the Pumby candidates that we tested). Using unit 3 and stacking amino acid R, assemblies that we tested seemed to hamper cell survival (*SI Appendix, Fig. S6A*). Using unit 3 and stacking amino acid Y, the tested assemblies did not hamper cell survival, but no Pum-mediated GFP reconstitution was observed (*SI Appendix, Fig. S6B*). Using unit 6 and stacking amino acid R, we found that the tested assemblies expressed well, but very weak Pum-mediated GFP reconstitution was observed for all tested sequences (*SI Appendix, Fig. S6C*). Finally, we tested unit 6 with stacking amino acid Y and found normal cell health and also

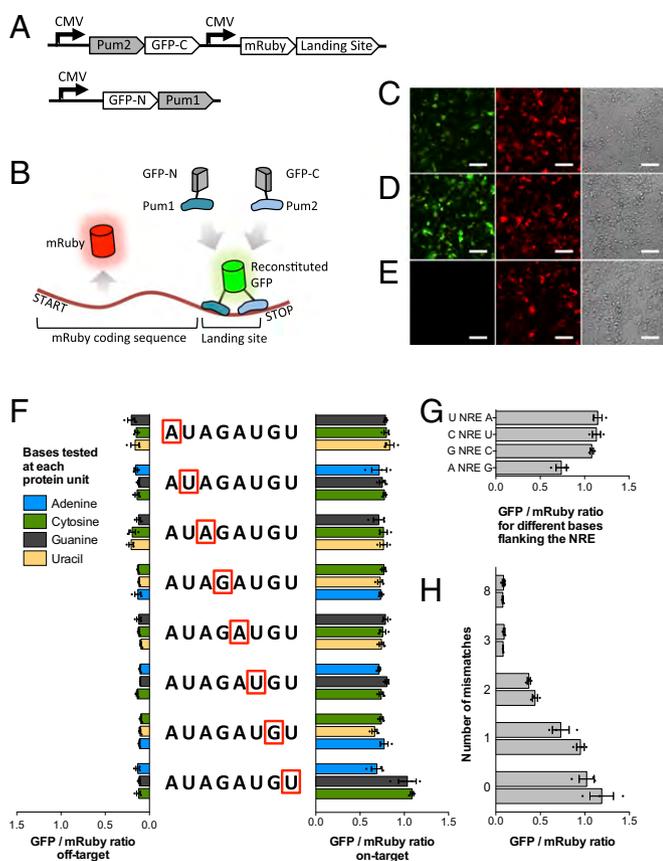


Fig. 2. Evaluation of PumHD mutant binding, with every unit mutated to target each of the four RNA bases. (A) Schematic of the plasmids used in the binding assay for validating the PumHD consensus sequence of Fig. 1C. (B) Schematic of the binding event that results from using the plasmids in A. A PumHD variant (denoted Pum1) and the wild-type PumHD (denoted Pum2) are each fused to one part of split GFP (for a full list of the sequences used in this figure, see *SI Appendix, Table S2*; for full statistics and n values of replicates, see *SI Appendix, Table S1*). Pum1 and Pum2 each target one 8-mer sequence within the landing site inserted before the stop codon of mRuby. The mRuby landing site transcript serves as a scaffold for GFP reconstitution upon PumHD binding, and the mRuby protein provides a control for overall cell density and transfection efficiency. (C–E) Representative fluorescent microscopy images of HEK293FT cells expressing the system of A, showing the green (GFP), red (mRuby), and bright-field channels for the same cells. (C) The transfected construct is PumHD with module 7 mutated to bind U (abbreviated 7-U), with on-target RNA present. (D) The transfected construct is PumHD 4-C, with on-target RNA. (E) The transfected construct is PumHD 4-C, with off-target RNA present. (All scale bars, 100 μm .) (F) Binding of on-target vs. off-target PumHD variants. We varied the target sequence of Pum1, changing each unit in turn to target each of the four bases of RNA, according to the key amino acid consensus sequence in Fig. 1C. The starting target sequence for Pum1 was 3'-AUAGAUGU-5', which we mutated unit by unit to test the targeting of three other bases, at each position. Each cluster of three horizontal bars in this panel corresponds to the test results for the unit framed in red; the colors of the bars (blue, green, black, and yellow) indicate the specific base targeted according to the color key at left. The readout for this assay is fluorescence from reconstituted split GFP, normalized to mRuby expression. Bars to the right show the GFP/mRuby ratio for on-target Pum1 (i.e., in which the protein sequence exactly matches the RNA target in the landing site), and bars on the left show the ratio for off-target Pum1 (i.e., in which there are eight out of eight possible mismatches between Pum1 and its RNA target). (G) GFP/mRuby ratios for wild-type PumHD tested against the wild-type target sequence (called the NRE) flanked by different adjacent nucleotides. The bar at bottom, A NRE G, is for the pair of flanking bases used in the rest of Fig. 2. (H) Tolerance of PumHD to protein–RNA mismatches. We tested two Pum1 sequences against RNA targets with zero, one, two, three, and eight mismatches. Values throughout this figure are mean \pm SEM.

GFP reconstitution (Fig. 3C; compare with Fig. 3D, which shows another example of the failed unit 3/stacking amino acid Y candidate highlighted in *SI Appendix, Fig. S6B*), which resulted in the hypothesized Pumby (Pumilio-based assembly) module set of Fig. 3B.

We validated the hypothesized Pumby module set of Fig. 3 using GFP reconstitution as in Fig. 2 (for a full list of the sequences used in Fig. 4, see *SI Appendix, Table S4*). We found that, for Pumby-based chains that were eight units long (abbreviated Pumby8 below), on-target Pum pairs resulted in significantly higher GFP reconstitution compared with off-target pairs (Fig. 4A; $P < 0.0001$ for factor of on- vs. off-target; two-way ANOVA with factors of on- vs. off-target and specific target sequence; $n = 3$ biological replicates each; see *SI Appendix, Table S3* for full statistics for Fig. 4), as it had for the PumHD variants (Fig. 2F). We also explored the effect of varying flanking bases around the Pumby target sequence (as for PumHD variants in Fig. 2G) and again found successful binding, albeit with, as expected, quantitative differences in GFP reconstitution magnitude (Fig. 4B). We used purified PumHD variants as well as Pumby8 chains to measure K_d 's for on- vs. off-target pairs, obtaining K_d 's in the nanomolar range for both Pumby8 and PumHD variants (*SI Appendix, Fig. S8, Table S16, and Supplementary Results*); off-target pairs had no detectable binding. We performed off-by-one, -two, and -three mismatch assessment for Pumby8, as we did for PumHD earlier, and found that some split GFP reconstitution was observed for one or two mismatched units, implying some degree of Pumby8 mismatch tolerance, but three mismatches did not support GFP reconstitution any more than did completely different (i.e., off by eight bases) RNA sequences (Fig. 4C; $P = 0.9999$ for comparison of three vs. eight mismatches; Dunnett's post hoc test across values of mismatch number, after the previous ANOVA; $n = 3$ biological replicates). We investigated the stability of Pumby8 proteins compared with PumHD proteins that bind the same RNA target sequence. We used a thermal assay, the measuring of fluorescence of SYPRO Orange as it is bound by unfolding protein. The resulting melting curves show that all Pum variants have a melting temperature (T_m) between 50–60 $^{\circ}\text{C}$, Pumby8 and PumHD alike (*SI Appendix, Fig. S7*; for a full list of the sequences used in *SI Appendix, Fig. S7, see SI Appendix, Table S15*).

Having demonstrated the performance of Pumby chains eight units long (Pumby8 for short), we next explored Pumby chains that could bind to shorter or longer RNA sequences—ranging in length from 6 to 18 units long (denoted Pumby6 to Pumby18). We found that, for Pumby-based chains of variable length, on-target pairs resulted in significantly higher GFP reconstitution compared with off-target pairs (Fig. 4D; $P < 0.0001$ for factor of on- vs. off-target; two-way ANOVA with factors of on- vs. off-target and specific target sequence; $n = 3$ biological replicates; full statistics in *SI Appendix, Table S3*). The Pumby chains ranging from length 6 to length 18 were similar to Pumby8 in terms of their GFP reconstitution effects (Fig. 4D; statistics in *SI Appendix, Table S3*). Thus, Pumby modules can indeed support the generation of RNA-binding proteins that are specific and that are longer in length than wild-type PumHD, which have efficacy comparable to the 8-mer Pumby (Fig. 4A). We also explored sequences shorter than Pumby8, synthesizing and testing Pumby chains that were six units long, and found on-target pairs to yield significantly higher GFP reconstitution than off-target pairs (Fig. 4E; $P < 0.0001$ for factor of on- vs. off-target; two-way ANOVA with factors of on- vs. off-target and specific target sequence; $n = 3$ biological replicates), with no difference between any of the Pumby6's tested and the 4-U variant—that is, the equivalent of the truncated wild type, which was assessed in Fig. 4D ($P > 0.05$, Dunnett's post hoc test across specific target sequence for the ANOVA above).

We next developed a novel use of programmable RNA-binding proteins: the monitoring of translation in live cells. Our initial experiments showed how Pum proteins can recruit split GFP to produce green fluorescence in the presence of a target RNA (as in

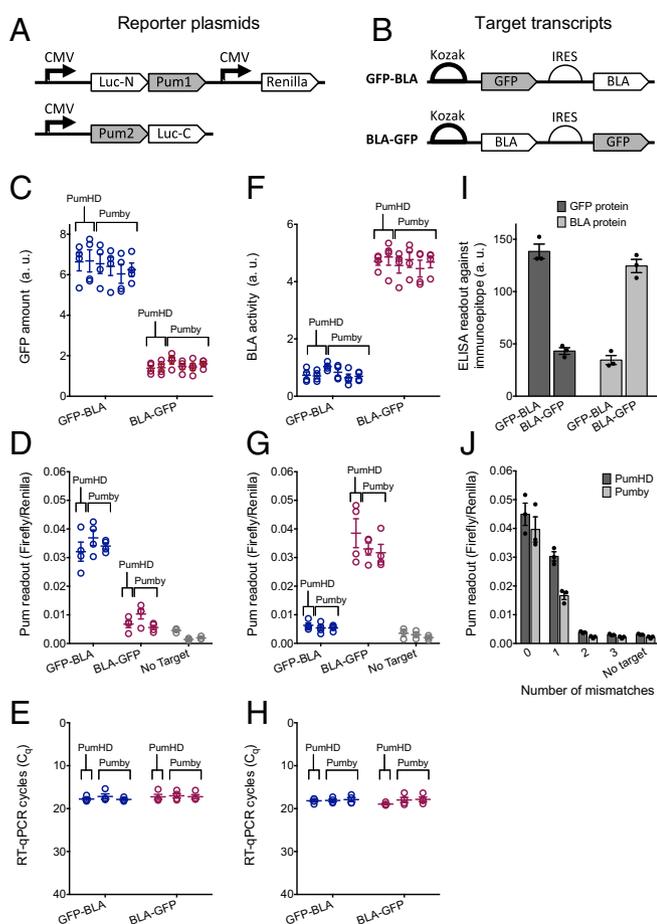


Fig. 5. Pumby-mediated monitoring of RNA translation in live cells. (A) Schematic of reporter plasmids used to measure translation. The plasmids encode for two Pum proteins (designed to bind to various sequences within the target RNAs shown in B), each fused to half of split firefly luciferase. One plasmid also encodes for a control gene, Renilla luciferase, which helps quantify transfection efficiency and cell density. (B) Schematics of two different target mRNAs used to systematically test the Pum vectors shown in A. Only one of the two target mRNAs is used in each experiment. The mRNAs contain sequences encoding for GFP and BLA behind strong (Kozak sequence) or weak (IRES) translation start positions. They are labeled GFP-BLA and BLA-GFP for the (GFP strong, BLA weak) and (BLA strong, GFP weak) conditions, respectively. Three Pums were targeted to each of the two ORFs, aiming for stretches of RNA with low secondary structure (see *SI Appendix, Table S6* for a full list of the sequences used in this figure, *SI Appendix, Supplementary Results* for more on how these sequences were chosen, and *SI Appendix, Table S5* for full statistics). (C) GFP levels (arbitrary units) measured for cells transfected with either GFP-BLA or BLA-GFP (with the choice of target transcript marked on the x axis), as well as both reporter plasmids from A; $n = 4$ biological replicates. (D) Firefly luciferase reconstitution (normalized to Renilla luciferase levels) mediated by Pum reassembly on RNA scaffolds, for three Pum binding sites in the GFP sequence, for cells transfected with either GFP-BLA or BLA-GFP (or no target) as well as both reporter plasmids from A; $n = 4$ biological replicates for the GFP-BLA and BLA-GFP cases; $n = 3$ biological replicates for the case of no target. (E) RT-qPCR measurement of the GFP transcript for the experiments of D, where C_q is the quantification cycle (50) ($n = 4$ biological replicates). (F) As in C, but for BLA activity (from the same set of biological replicates). (G) As in D, but for Pum binding sites in the BLA sequence. (H) As in E, but for the experiments of G. (I) Amount of GFP or BLA protein for cells transfected with one of the two target transcripts from B, as measured by ELISA against a small immunopeptide (6xHis) fused to either BLA or GFP. (J) Sensitivity of translation measurement to mismatches between Pum proteins and their target RNA. We tested two proteins in the role of Pum1 (a Pumby8 against a target in GFP and a PumHD against a target in BLA), each paired with the target transcript that would create high expression for their target gene (GFP-BLA and BLA-GFP, respectively), and varied the target RNA to have zero, one,

compared with when it was immediately downstream of the IRES; this was observed for both GFP (Fig. 5C; $P < 0.0001$ for factor of GFP location; two-way ANOVA with factors of GFP location and Pum type; see *SI Appendix, Table S5* for the full statistics for Fig. 5, as well as n values for replicates) and for BLA (Fig. 5F; $P < 0.0001$ for factor of BLA location). The amount of translation did not depend on whether a Pumby8 or a PumHD was targeted to the mRNA sequence (Fig. 5C, $P = 0.6517$ for factor of Pum type; two-way ANOVA with factors of Pum type and GFP location; Fig. 5F, $P = 0.7198$ for factor of Pum type in the analogous BLA case).

We sought to independently verify the results of these translation measurements in a way that did not depend on the reporter nature of the two proteins (GFP and BLA) that we used in our demonstration but that could potentially apply to any protein. Thus, we fused GFP and BLA to 6xHis, an immunopeptide, and measured expression levels with ELISA, a standard way of gauging protein levels (Fig. 5I). As in the reporter-based readout (Fig. 5C and F), we saw that the gene behind the Kozak sequence consistently yielded higher levels of protein production than the one behind the IRES (Fig. 5I; $P = 0.00024$ for variations in GFP protein level caused by position in the target transcript; $P = 0.0003$ for BLA protein level; multiple t tests using the Holm-Sidak method; see *SI Appendix, Table S5* for full statistics). Thus, we were able to validate through both direct reporter detection and ELISA immunopeptide quantitation the modulation of translation by gene position in our constructs.

Having validated our assay, we next assessed the hypothesis that Pum-mediated luciferase reconstitution could also measure protein translation. We observed greater Pum-mediated luciferase reconstitution when Pums targeted the coding sequence behind the Kozak sequence than behind the IRES (for landing sites within GFP, Fig. 5D; $P < 0.0001$ for factor of GFP location; two-way ANOVA with factors of Pum type and GFP location; for landing sites within BLA, Fig. 5G; $P < 0.0001$ for factor of BLA location). Pumby8 and PumHD showed indistinguishable behavior in this experiment (Fig. 5D; $P = 0.5261$ for factor of Pum type; two-way ANOVA with factors of Pum type and GFP location; Fig. 5G; $P = 0.0854$ for factor of Pum type in the analogous BLA case). We verified that in these experiments mRNA levels were unaffected by the order of the ORFs within, using reverse-transcription quantitative PCR (RT-qPCR) to quantitate the amount of target transcript mRNA (Fig. 5E, $P = 0.2589$ for factor of GFP location; two-way ANOVA with factors of Pum type and GFP location; Fig. 5H, $P = 0.5634$ for factor of BLA location). The RT-qPCR mRNA counts for GFP were indistinguishable when Pumby8 vs. PumHD were used (Fig. 5E; $P = 0.6236$ for factor of Pum type; two-way ANOVA with factors of Pum type and gene order; Fig. 5H; $P = 0.1092$ for factor of Pum type). Thus, the Pum-based reconstitution assays, and the more conventional protein measurement assays above, represent mRNA translation and not mRNA transcript copy number change.

Next, we tested the tolerance of our translation monitoring assay, assessing mismatches between the Pum protein and its target RNA sequence. We mutated two particular Pum sequences (one PumHD and one Pumby8) to contain zero, one, two, or three mismatches (Fig. 5J). We observed some luciferase reconstitution above baseline when one unit was mismatched, but in this case even two mismatches were sufficient to effectively eliminate luciferase reconstitution (Fig. 5J; $P > 0.99$ for comparison of two or three mismatches vs. No Target; Dunnett's post hoc test across values of mismatch number, after two-way ANOVA with factors of Pum protein and mismatch number).

two, or three mismatches; we also included a case in which the target transcript was absent. Circles (C–H) and dots (I and J) represent individual data points; error bars show mean \pm SEM.

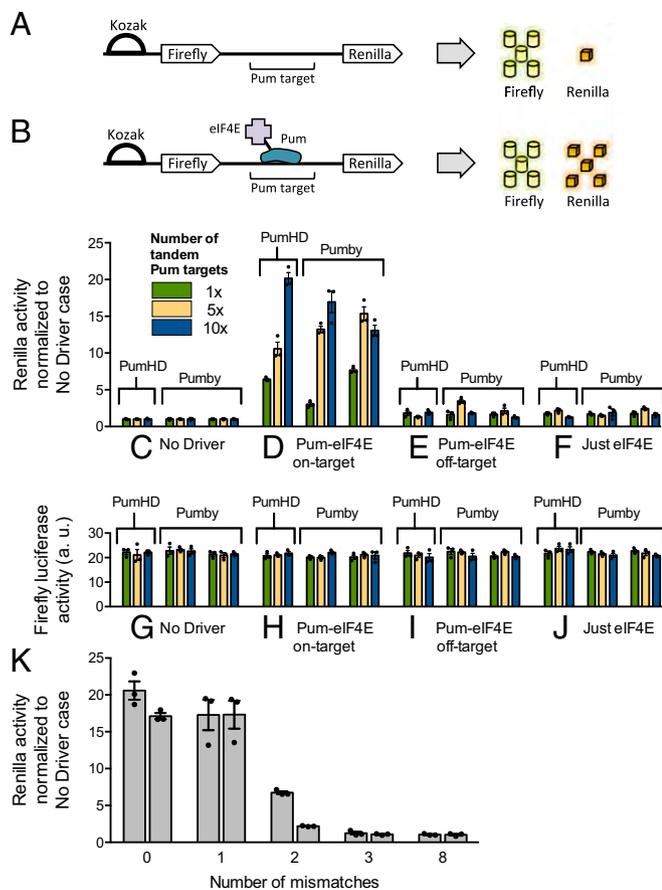


Fig. 6. Gene translation targeted to specific sequences by modular RNA-binding proteins. (A) Schematic of a reporter transcript containing genes for firefly and Renilla luciferases, with a Kozak sequence immediately upstream of firefly luciferase but not of Renilla luciferase; under these conditions, the Renilla ORF yields much lower levels of translation (30). (B) Schematic of how expressing translation initiation factor eIF4E fused to a Pum protein (from a separate driver plasmid) could in principle be used to drive translation of a downstream ORF, causing in this case the production of more Renilla luciferase. (C–F) Renilla luciferase activity as a measure of Pum-eIF4E-mediated translation initiation facilitation, using reporter transcripts bearing three different Pum target sites, in tandem repeats of 1, 5, or 10 copies in a row, in conjunction with various different driver plasmids. The data in C–F were normalized to their respective means in C (for a full list of the target binding sequences used in this figure, see *SI Appendix, Table S8*; for full statistics, see *SI Appendix, Table S7*). Specifically: C, Renilla levels when only the reporter plasmid of A is used, with no driver plasmid. (D) Renilla levels when the reporter plasmid of A is used with an on-target driver plasmid, as in B. (E) Renilla levels when the reporter plasmid is used with an off-target driver plasmid. (F) Renilla levels when the reporter plasmid is used with a driver plasmid where eIF4E is present but not fused to Pum. (G–J) Firefly luciferase activity, from the first ORF of the bicistronic luciferase vectors. (K) Sensitivity of translation initiation to mismatches between Pum-eIF4E and the RNA target. We tested two Pum proteins against targets with zero, one, two, and three mismatches. $n = 3$ biological replicates; values throughout are mean \pm SEM.

Another useful mRNA operation is translation initiation, previously demonstrated by fusing wild-type PumHD (or two of its mutants) to the translation activation factor eIF4E (30, 31). We assessed the performance of Pumby in this context by simultaneously measuring the expression of two ORFs from a single transcript (Fig. 6A). We created a transcript containing a Kozak sequence, a firefly luciferase ORF, and a Renilla luciferase ORF, in that order. The Kozak sequence allows translation of the more proximal firefly ORF, with only a weak spillover effect on the Renilla ORF. Between the ORFs were one of three

mRNA target sequences (for PumHD or Pumby binding), present in 1, 5, or 10 copies. We combined this target transcript with various Pum-eIF4E fusion proteins to drive translation (Fig. 6B; one Pum was a PumHD variant and two were Pumby8 chains; see *SI Appendix, Table S8* for a full list of the sequences used in Fig. 6). We found that, compared with baseline Renilla expression with any of the nine target vectors on its own, expression with the correct on-target Pum-eIF4E driver increased Renilla luciferase translation by about an order of magnitude (Fig. 6C and D; $P < 0.0001$ for post hoc comparison of these two conditions; Tukey's post hoc test after three-way ANOVA with factors of copy number, driver plasmid, and Pum type used throughout this paragraph; see *SI Appendix, Table S7* for full statistics related to Fig. 6, as well as n values of replicates). More tandem repeats led to higher boosts in expression; for example, the 10 \times array produced several times higher expression than the 1 \times (Fig. 6D; $P = 0.0006$ for post hoc comparison of these two conditions). In contrast, expression was indistinguishable from baseline for off-target Pum proteins fused to eIF4E (Fig. 6E; $P = 0.9899$ for post hoc comparison of these two conditions), or for eIF4E administered alone (Fig. 6F; $P = 1$ for post hoc comparison of these two conditions). As a control, firefly luciferase activity did not vary with target copy number or Pum type (Fig. 6G–J; $P = 0.7826$ and $P = 0.4676$ for each factor, respectively). Thus, Pum proteins make it possible to up-regulate translation of proteins without any need for modified translation initiation sites. We found that Pumby8 and PumHD had the same effect as each other throughout this experiment (Fig. 6C–F, $P = 0.4656$ for factor of Pum type; Fig. 6G–J, $P = 0.4676$ for factor of Pum type).

We tested the tolerance of our translation initiation assay to mismatches between the Pum protein and its target RNA sequence (Fig. 6K). We mutated two particular Pum sequences to contain zero, one, two, or three mismatches. We observed some translation above baseline for one or two mismatched units, but three mismatches were sufficient to effectively eliminate the Pum-eIF4E translation boost (Fig. 6K; $P = 0.9998$ for comparison of three vs. eight mismatches; Dunnett's post hoc test across values of mismatch number, after a two-way ANOVA with factors of Pum protein and mismatch number; $n = 3$ biological replicates each). Thus, Pums can mediate target-specific translation initiation, as well as PumHD. We also tested our Pum proteins in an assay for gene silencing (see *SI Appendix, Fig. S4 and Supplementary Results*), as well as further tests of Pum orthogonality (see *SI Appendix, Fig. S3 and Supplementary Results*) and found in all cases equivalent performance between Pumby8 and PumHD. Thus, through all these experiments we showed that PumHD and Pumby modules can enable a wide variety of protein-mediated mRNA measurements and perturbations, to be easily performed on unmodified mRNA sequences. We also discovered a new use of such RNA-binding proteins, the monitoring of translation level in living cells.

Discussion

We have discovered a modular protein architecture comprising four protein building blocks derived from the Pumilio protein that enable universal RNA targeting and engineered it for concatenation in chains ranging from 6 to 18 modules in length. Previous works had demonstrated, using proteins that bind to specific RNA sequences, the measurement of mRNA expression level (23, 24), imaging of mRNA dynamics (23–26, 32), and enhancement and suppression of mRNA translation (6, 30, 31, 33) with variants of natural RNA-binding proteins. We demonstrated that our Pumby architecture, which uses a single repeated module to support protein generation (analogous to the TALE design), enables performance equivalent to the original Pumilio protein. We also demonstrate a novel application of modular mRNA-binding proteins—the measurement of translation in live cells. This simple and modular technology may support, as the

ability to systematically map the static distribution of RNAs in situ becomes available (34, 35), the dynamic mapping and control of RNAs to assess their causal role in cellular processes such as those explored here. Pumby was able to support specific binding, with sequences differing by as few as two or three bases resulting in less, or even functionally zero, binding.

A significant part of this functionality in Pumby results from its modular architecture, which makes it possible to target sequences of varying length, not just eight bases long like with the wild-type Pumilio. Longer target sequences are less likely to be found at random in the transcriptome, which helps avoid off-target effects. Furthermore, some investigations require the recognition of a long target: Differentially spliced or highly repetitive transcripts, in particular, can only be uniquely identified through sequences longer than their constitutive parts. Pumby may allow for the creation of varying-length footprints for protection against nucleases or other RNA-binding proteins and may provide a malleable tool for tuning the energy balance of RNA secondary structure in living cells. Many engineering applications are also possible, such as assembling complex scaffolded protein-based reaction pathways in mammalian cells in an RNA-programmable fashion, as has been done before in bacteria (36).

RNA takes on complex secondary structures in live cells and is frequently bound by endogenous RNA-binding proteins; this behavior affects all technologies that rely on in vivo interactions with RNA. Pum proteins are no exception to this rule, and our use of several arbitrary target sequences should not be interpreted as evidence that any arbitrary Pum sequence will bind successfully, or that a Pum protein that worked in one cellular environment will work in all others. In our experiments, roughly three-fifths of the protein sequences we tested in a new RNA context behaved as expected (see *SI Appendix, Supplementary Results* for details on how this occurred in the context of our translation measurement experiments). Our several mismatch experiments, furthermore, showed that RNA sequences differing by one or two bases from a Pum's target sequence (but not three or more) can result in measurable binding effects using our assays. With these benchmarks in mind, researchers applying PumHD and Pumby to a new experiment should always validate new sequences in their final biological context.

Previous studies had probed whether PumHD variants could bind a wide diversity of NRE mutants. Here, in a single study, we tested PumHD variants binding to all four possible nucleotides at all positions under the same set of conditions. For many applications, especially if the number of bases targeted is not a key issue, or if a modular design is not required, this dataset may help with application of PumHD variants themselves to the mapping and control of RNA functions. Along these lines, other members of the Puf family have also been used to engineer selective binding between functional effector proteins and RNA targets. One of the most extensively studied is the *C. elegans* Fem-3 mRNA-binding factor 2 (FBF-2), which is an analog of PumHD (6, 37–40). Cooke et al. (41) linked wild-type FBF-2 to the translation activator GLD2 to trigger poly(A) signal addition and up-regulate translation in *Xenopus* oocytes. Conversely, they linked the FBF-2 domain to the translational repressor CAF1 to trigger poly(A) removal and subsequent translation down-regulation. Campbell et al. (6) also activated translation in human U2OS cells by fusing the yeast poly(A) binding protein to an FBF-2 protein mutant that targets a specific mRNA segment of the human cyclin B1. Such architectures, if tested with every unit mutated to bind every base, or if they yield single-module building blocks, may present the kinds of utility shown here for the Pumilio protein.

The seemingly simple modular binding nature of PumHD masks a great wealth of complexity in the way that the diverse units of the protein contribute to overall protein binding. For example, it has been observed that stacking residues affect the specificity of base-binding differently at different units, that changes to the three key

amino acids binding one base affect binding to neighboring bases as well as at the mutant site, and that C-terminal repeats are in general more specific than N-terminal repeats (6). PumHD variants from yeast and nematodes have been shown to bind nine-nucleotide RNA sequences even though they have only eight protein units (18). Human PumHD may bind the fifth RNA base in its target sequence using different in vivo binding modes depending on the base at that position (42). Pumby presents an array in which all units can be selected from the same set of four modules. Thus, Pumby may present a simplified context in which to insert Pumilio modules to study how specific amino acids contribute to the emergent properties of modular RNA binding, independent of position-specific effects. Such future insights into the architecture of Pumilio may not only provide basic science insights into this interesting class of proteins but also help with the design of next-generation RNA-binding tools.

Materials and Methods

Golden Gate Compatible Mammalian and Bacterial Expression Vectors. We prepared Golden Gate compatible mammalian expression vectors by eliminating BsaI sites from previously used vectors as follows. The human CMV major immediate-early gene enhancer/promoter expression vector, called pCI-CMV-GG, was made from the commercially available pCI vector (Promega) by removing BsaI sites from the CMV region (specifically from the β -globin/IgG chimeric intron located downstream of the enhancer/promoter) and from the ampicillin resistance gene. The BsaI site in the chimeric intron, and thus the introduced mutation, was outside of the two known intron splice sites (43). For lower expression levels we created a vector called pCI-GG-UB, in which we replaced the CMV promoter with the human polyubiquitin C (UBC) promoter and introduced a single point mutation to remove the BsaI site from the UBC promoter. The efficiency of the two newly mutated promoters was confirmed by comparing the expression of the firefly luciferase under the original promoters with that under the Golden Gate compatible mutated versions (data not shown). In both cases, the expression levels of luciferase from the original and mutated versions of the promoter were nearly identical. All key sequences are deposited at GenBank (KU900022–KU900031), and all key plasmids will be available from the nonprofit distributor Addgene.

Golden Gate Cloning of PumHD Variants. Our PumHD units were assembled by adapting the Golden Gate protocol from a prior TAL effector study (44). See *SI Appendix, Fig. S1* for a general scheme of our cloning procedure. We first purchased—as synthetic oligonucleotides (IDT)—four base-specific variants of each of the eight RNA-binding units in PumHD, as well as nonsequence-specific units 0 and 9. The units were designed with BsmBI and BsaI restriction sites at the ends (see *SI Appendix, Table S19*).

To assemble the 10 units (eight RNA-binding units plus units 0 and 9) required for the PumHD architecture as used in Fig. 1, two intermediate pentamer assemblies were first prepared. The Golden Gate reaction (digestion with BsmBI at 37 °C and ligation with T7 ligase at 16 °C, repeated 25 times) created circular pentamers; for each PumHD assembly, one pentamer contained units 0, 1, 2, 3, and 4 and the second pentamer contained units 5, 6, 7, 8, and 9.

Any incorrect, noncircularized assemblies were digested with an ATP-dependent DNase that acts only on linear DNA (Plasmid-Safe ATP-Dependent DNase; Epicentre). The DNase digestion reaction mixture was then used as a PCR template to amplify the linear pentamers. The PCR, performed using Herculase polymerase (Herculase II Fusion DNA Polymerase; Agilent) yielded several unspecific products (“smudged bands”), as was previously described in the case of TAL assembly; this phenomenon has been attributed to polymerases “slipping” on repetitive templates, an occurrence that can be almost entirely avoided by preheating the PCR plus silicone oil to 98 °C and adding Herculase plus dNTPs to the hot mixture through the silicone oil. Pentamer products of the correct size were separated on a 2% (wt/vol) agarose gel and extracted from the gel. Two linear pentamers were assembled into the final construct by the second Golden Gate reaction, using BsaI (digestion with BsaI at 37 °C and ligation with T7 ligase at 16 °C, repeated 25 times) followed by a final digestion with Plasmid-Safe ATP-Dependent DNase. The digestion mixture was used to transform Z-Competent Stbl3 *Escherichia coli* (Zymo). Bacteria were always incubated at 30 °C, because slower growth is reported to prevent scrambling of the repetitive array plasmids. The plasmids were purified using standard Miniprep kits (Zymo). See *SI Appendix, Supplementary Results and Fig. S1* for details on the design of the cloning procedure.

Golden Gate Cloning of Pumby. Proteins based on the Pumby module were assembled using the general Golden Gate scheme described above, with unit 6 of PumHD used on all positions in the assembly and Tyrosine as AA2 (the stacking amino acid). The full list of sequences used to prepare hexamers for Pumby construction is given in *SI Appendix, Table S20*. One major difference with PumHD is that the total length of Pumby chains may vary; consequently, the four base-specific variants of each Pumby unit were prepared with cloning overhangs to circularize into n-mer cloning intermediates of whatever length was needed. We used cloning intermediates with between three and six units to assemble final Pumby chains of up to 24 units. To create a 10-mer Pumby, for example, we prepared one hexamer and one tetramer to reach the total of 10 units in the final assembly. All bacterial amplification was done at 30 °C, as above. Because of difficulty in sequencing highly repetitive arrays, for each assembly three correct clones were selected, purified, and mixed (to minimize the chance of having undetected mutations because of lack of comprehensive sequencing coverage of the highly repetitive area). See *SI Appendix, Supplementary Results and Fig. S1* for details on the design of the cloning procedure.

Transfections and Cell Culture. HEK293FT and HeLa cells were purchased from ATCC. All cells purchased from ATCC are tested for *Mycoplasma* contamination before shipping. All transfections of HEK293FT (used in all figures except *SI Appendix, Fig. S4*) and HeLa (used in *SI Appendix, Fig. S4*) cells were performed using Mirus X2 transfection reagent, according to the manufacturer's directions. Cells were grown in D10 medium (DMEM, supplemented with 10% vol/vol heat-inactivated FBS, 100 IU penicillin, 100 µg/mL streptomycin, and 1 mM sodium pyruvate). For imaging, cells were grown in Matrigel (Corning)-coated glass 24-well plates. For qPCR, luciferase, and BLA assays cells were grown in polystyrene six-well plates (Greiner Bio-One). In all experiments, cells used were no older than passage 18, typically passage 7–15. All batches of cells were assigned randomly to receive one set of transfected genes or pharmacological conditions vs. another. No blinding was used.

For transfection of cells in 24-well plates, we transfected 250 ng of plasmid with 250 ng of diluent DNA (pUC19 plasmid) to keep the total amount of DNA introduced at 500 ng per well of the 24-well plate. If multiple plasmids were cotransfected, they were always in equal proportion and the total amount of plasmid DNA was always 250 ng per well of the 24-well plate (plus 250 ng of pUC19, for 500 ng of total DNA). At 24 h posttransfection, we always exchanged the cell growth media with fresh D10 to remove any remaining transfection reagent.

PumHD and Pumby Binding in Live Mammalian Cells Measured via Pum-Mediated GFP Reconstitution Normalized to mRuby Red Fluorescence (the "Green Red Screen"). All images (Figs. 2 and 3 and *SI Appendix, Figs. S6 and S11*) were captured using cultured HEK293FT cells after a 60-h incubation posttransfection [48 h at 37 °C followed by 12 h at 30 °C, as has been done in previous split GFP experiments (23, 24)]. All images for samples presented in a given figure were taken with the same light source, filter cubes, and objective settings.

RNA Quantification for Translation Measurement Assays. RNA was quantified by RT-qPCR with a LightCycler480 (Roche), using a CellsDirect One-Step qRT-PCR Kit (Life Technologies). Hydrolysis probes were designed against the sequences of EGFP, BLA, and the N-terminal fragment of split luciferase using the Custom TaqMan Assay Design Tool (Life Technologies). Life Technologies did not disclose the sequence of the probes used in this work. HEK293FT cells were grown in 24-well plates, transfected at ~70% confluence, and harvested after 24 h. For harvesting, cells were washed with DMEM (Corning), digested with 100 µL 0.05% trypsin-EDTA (Corning) for 5 min, diluted with 800 µL PBS, and transferred to 1.5-mL microtubes. Cells were centrifuged at 200 relative centrifugal force (rcf) for 5 min, resuspended in 1 mL PBS, and counted with a Scepter 2.0 Handheld Cell Counter (Millipore). A given cell number for each condition depending on availability (4,000 cells per condition for half of the biological replicates, 2,000 cells for the other half) was extracted, centrifuged at 200 rcf for 5 min, and resuspended in PBS. The cells were then treated according to the CellsDirect protocol. Briefly, cells from each condition were mixed with lysis buffer and frozen at –80 °C until further use, then lysed, digested with DNase I, and divided into RT-qPCR wells. The 20-µL reactions were carried out in 96-well plates (Roche). Each reaction included steps for reverse transcription (15 min at 50 °C) and 40 cycles of qPCR (30 s at 60 °C). Quantification cycle (C_q) calculations were carried out in the LightCycler480 software by the Fit Points Method (Roche). Statistical analysis of the C_q values was carried out in Microsoft Excel 2011, GraphPad Prism 6, and JMP Pro-11.

For experiments in Fig. 5, the data for GFP, BLA, and Pum-readout luciferase, as well as corresponding RT-qPCR data for each sample, were collected

from the same biological replicates (cells grown and transfected at the same time, in adjacent wells of a microwell plate). HEK293FT cells for those experiments were harvested 72 h posttransfection.

For the gene silencing experiments of *SI Appendix, Fig. S4*, the Renilla luciferase, firefly luciferase, and RT-qPCR data for each sample were collected from the same biological replicates (HeLa cells grown and transfected at the same time, in adjacent wells of a microwell plate). Cells for those experiments were harvested 48 h posttransfection.

Orthogonality Tests. For the orthogonality tests of *SI Appendix, Fig. S3*, luciferase and APEX2 assays were performed on all technical replicates on the same day, with the same batch of reagents. APEX2 activity served as a transfection control; that is, we screened all our biological samples for peroxidase activity and used its presence as an indicator that the well had been successfully transfected with a target vector. We chose APEX2 for this purpose because it is a modified peroxidase that shows strong activity in the mammalian cytosol and to provide a verifiably translated gene in which to place the landing site. The landing site needed to be within the ORF of a translated gene, in order for a large amount of split firefly luciferase to be reconstituted (as described before for Fig. 2). We intended to exclude any samples that displayed zero peroxidase activity but in the end excluded none of our samples from the study for this reason. APEX2 activity was assayed with an Amplex Red Hydrogen Peroxide/Peroxidase Assay Kit (Invitrogen). Each biological replicate consisted of the HEK cells from one 24-well plate well, transfected with three plasmids encoding the following: Pum fused to N-terminal split firefly luciferase, Pum fused to C-terminal split firefly luciferase, and APEX2 fused to the landing site. All replicates were transfected with the same Pum fused to C-terminal split firefly luciferase, so reconstitution was determined solely by the correspondence between the Pum fused to N-terminal split firefly luciferase and its binding site. Each tile in *SI Appendix, Fig. S3* presents the average of three biological replicates.

Firefly and Renilla Luciferase Activity Assay. The activity of Renilla luciferase and firefly luciferase was measured using the Dual-Glo Luciferase Assay System (Promega) according to the manufacturer's instructions. It is important to note that the measured luciferase activity, especially for the reconstituted split luciferase, differs significantly between experiments if the reconstituted luciferin reagent is allowed to go through more than one freeze-thaw cycle. This has been previously noted by others using a luciferase detection kit based on the same chemistry (29). For results described in this paper, each "batch" of experiments (samples directly compared with each other, that is, all biological replicates in a single figure panel) was analyzed using the same, freshly prepared batch of reagents.

For the translation quantification experiments of Fig. 5, the data for GFP, BLA, and Pum readout luciferase, as well as corresponding RT-qPCR data for each sample, were collected from the same biological replicates (cells grown and transfected at the same time, in adjacent wells of a microwell plate). The cell harvesting protocol for those experiments is described in *Materials and Methods, RNA Quantification for Translation Measurement Assays*.

For gene silencing experiments of *SI Appendix, Fig. S4*, the Renilla luciferase, firefly luciferase, and RT-qPCR data for each sample were collected from the same biological replicates (HeLa cells grown and transfected at the same time, in adjacent wells of a microwell plate). The cell harvesting protocol for those experiments is described in *Materials and Methods, RNA Quantification for Translation Measurement Assays*.

For the translation initiation experiments of Fig. 6, cells were harvested 36 h posttransfection by digestion with Glo Lysis Buffer (Promega), according to manufacturer's instructions.

BLA Activity Assay. The BLA activity assays were performed using GeneBLazer In Vitro Detection Kit (Invitrogen) according to the manufacturer's instructions. For the translation measurement experiments of Fig. 5, the data for GFP, BLA, and Pum readout luciferase, as well as corresponding RT-qPCR data for each sample, were collected from the same biological replicates (cells grown and transfected at the same time, in adjacent wells of a microwell plate). The cell harvesting protocol for those experiments is described in *Materials and Methods, RNA Quantification for Translation Measurement Assays*.

Quantitative GFP Assay. The GFP activity was quantitated using GFP Quantitation Kit (BioVision) according to the manufacturer's instructions. For translation measurement experiments of Fig. 5, the data for GFP, BLA, and Pum readout luciferase, as well as corresponding RT-qPCR data for each sample, were collected from the same biological replicates (cells grown and transfected at the same time, in adjacent wells of a microwell plate). Thus,

the cell harvesting protocol for those experiments is described in *Materials and Methods, RNA Quantification for Translation Measurement Assays*.

His-Tag ELISA Expression Assay. A 6x poly-histidine tag (6xHis) was cloned at the N terminus of the GFP and BLA constructs used in the translation measurement experiments of Fig. 5. We measured expression of these proteins with a 6xHis-tag ELISA Kit (Abcam) according to the manufacturer's instructions.

Measurement of Native ATF4 Translation via Pum-Mediated Fluorophore Reconstitution. For the experiments described in *SI Appendix, Fig. S2 B and C*, HEK293FT cells were seeded and transfected with a pair of Pum GFP vectors and imaged as described above for the "green red screen." At 24 h posttransfection, 0.5 μ M thapsigargin was added. Cells were imaged again after 12 h, as described above. Each experiment was performed in three biological replicates (cells grown and transfected at the same time, in adjacent wells of a microwell plate). ATF4 protein expression was quantified using an ELISA Kit for Activating Transcription Factor 4 (Cloud-Clone Corp.). The cells were harvested at indicated time points and the ELISAs performed according to the manufacturer's instructions. Each experiment was performed in three biological replicates (cells grown and transfected at the same time, in adjacent wells of a microwell plate).

Protein Expression and Purification. A custom Golden Gate compatible bacterial expression vector was prepared, based on the pBadHisB (6xHis tag) vector backbone, removing BsaI site from the BLA coding sequence. Pum arrays were cloned into this vector as described above. His-tagged Pum variants were expressed in *E. coli* strain DH5 α , grown in 100 mL RM media induced with 0.005% arabinose, at 18 $^{\circ}$ C, 200 rpm, for 18–24 h (until the colony reached OD₆₀₀ of 0.7). Bacterial pellets were lysed with BugBuster Protein Extraction Reagent (5 mL per 1 g of wet bacteria paste; EMD Millipore) with lysozyme (0.50 mg/mL final concentration; Thermo Scientific). The proteins were purified using Talon Spin Columns (Clontech). The purified proteins were stored in aliquots in 25% (vol/vol) glycerol at -80° C.

Binding of Pum Variants to RNA Measured by Fluorescence Anisotropy. We used fluorescence anisotropy to measure the kinetics of binding of purified Pum proteins to their cognate and noncognate RNA. Fluorescence anisotropy is widely used to investigate steady-state, dynamic equilibrium binding between protein and RNA (45–47).

The cognate and noncognate RNA targets for the purified Pum variant proteins were synthesized with 5'-labeled FAM, 6-carboxyfluorescein (IDT). The activity of the purified Pum variants was estimated with a saturation assay for each protein and its cognate RNA as described before (7). Fifty nanomolar cognate RNA was mixed with increasing concentration of the protein (measured by NanoDrop; Thermo Scientific) in the binding buffer (25 mM Tris-HCl, pH 7.5, 0.5 mM EDTA, 50 mM KCl, and 0.1 mg/mL BSA). The 100- μ L samples were assayed, in duplicates, for fluorescence anisotropy using a Cary Eclipse fluorimeter (Varian) with Manual Polarizer Accessory (Varian). The cognate RNA is always the sequence exactly matching the whole Pum protein binding sequence, flanked as CCAGAAU*Pum_{sequence}*UUCG (for full list of sequences, see *SI Appendix, Table S16*) with flanking bases selected according to previously published studies (7, 23). Fluorescence anisotropy was calculated as a unitless ratio defined as $R = (I_{\parallel} - I_{\perp}) / (I_{\parallel} + 2I_{\perp})$, where I is the emission intensity parallel (I_{\parallel}) or perpendicular (I_{\perp}) to the direction of polarization of the excitation source. The stoichiometric point of each saturation plot was used to estimate the active protein fraction (See *SI Appendix, Fig. S8* for example plots). The K_d of each protein to its cognate and noncognate RNA was subsequently measured, using the

protein concentration corrected to the active protein fraction, with constant concentration of RNA. The K_d was calculated from a nonlinear fit in IgorPro 6.22 of the anisotropy vs. protein concentration plot to the equation (48)

$$F([\text{protein}]) = \left(\frac{((([\text{protein}] * K_a + [\text{RNA}] * K_a + 1) - ((([\text{protein}] * K_a + [\text{RNA}] * K_a + 1)^2 - 4 * K_a^2 * [\text{RNA}] * [\text{protein}])^{.5}) / (2 * K_a)) * (F_b - F_0) / ([\text{protein}]) + F_0}{1} \right)$$

where [protein] is the concentration of the active fraction of the protein and [RNA] is the RNA concentration. Example anisotropy measurement plots are shown in *SI Appendix, Fig. S8* and the K_d values for binding of PumHD variants and Pumby to cognate and noncognate RNA are shown in *SI Appendix, Table S16*.

Stability of Pum Variants Measured by a Thermal Shift Assay. The T_m of purified PumHD and Pumby variants was measured using a thermal shift assay with SYPRO Orange (Invitrogen) dye according to the previously described protocol (51). Briefly, the 2.5 μ M peptide samples were prepared in 100 mM Hepes (pH 7.4), 150 mM NaCl and 5 \times SYPRO Orange dye. Fluorescence vs. temperature was measured with a LightCycler480 (Roche) with a ramp rate of 1.2 $^{\circ}$ C/min. The melting temperature was obtained as a midpoint of the thermal unfolding curve by fitting the slope of the curve to the sigmoid equation in Igor Pro-6.37:

$$F = \text{base} + (\text{max} / (1 + \exp((T_m - x) / (\text{rate}))))$$

The reported T_m is an arithmetic average of four replicates; T_m obtained from all independent replicates was within 1 $^{\circ}$ C from the reported average value. See *SI Appendix, Fig. S7* for melting plots and T_m results.

Statistics. The reasoning behind the sample sizes was not based upon a power analysis, because this study is primarily about exploring a new technology. As noted in ref. 49 and recommended by the NIH, "In experiments based on the success or failure of a desired goal, the number of [experiments] required is difficult to estimate..." We want to evaluate how a new technology works, and outcomes are not anticipatable, because the technology has not existed before, to our knowledge. As noted in ref. 49, "The number of experiments required is usually estimated by experience instead of by any formal statistical calculation, although the procedures will be terminated when the goal is achieved." In our case, we attempted to validate the tool by trying many different biological validations, in different contexts, as we have done in the past, to understand the biological impact of the tool in the context of different questions. Each experiment was repeated on a minimum of nine technical replicates; see n values given with each experiment.

ACKNOWLEDGMENTS. We thank Katriona Guthrie-Honea and Paul Reginato for assistance with cloning and Kiryl Piatkevich, Jacob Becraft, Daniel Schmidt, and Stuart Levine for advice. This work was supported by NIH Grant 1R01NS075421, National Science Foundation Chemical, Bioengineering, Environmental, and Transport Systems Grant 1344219, Jeremy and Joyce Wertheimer, NIH Brain Research through Advancing Innovative Neurotechnologies Initiative Grant 1U01MH106011, the New York Stem Cell Foundation Robertson Award, NIH Director's Transformative Award 1R01MH103910, and NIH Director's Pioneer Award 1DP1NS087724 (to E.S.B.), the Janet and Sheldon Razin (1959) Fellowship (to D.A.M.-A.), and the Massachusetts Institute of Technology Media Lab.

- Buxbaum AR, Haimovich G, Singer RH (2015) In the right place at the right time: Visualizing and understanding mRNA localization. *Nat Rev Mol Cell Biol* 16(2):95–109.
- Lionnet T, et al. (2011) A transgenic mouse for in vivo detection of endogenous labeled mRNA. *Nat Methods* 8(2):165–170.
- Tyagi S (2009) Imaging intracellular RNA distribution and dynamics in living cells. *Nat Methods* 6(5):331–338.
- Re A, Joshi T, Kulberkyte E, Morris Q, Workman CT (2014) RNA-protein interactions: An overview. *Methods Mol Biol* 1097:491–521.
- Chen Y, Varani G (2013) Engineering RNA-binding proteins for biology. *FEBS J* 280(16):3734–3754.
- Campbell ZT, Valley CT, Wickens M (2014) A protein-RNA specificity code enables targeted activation of an endogenous human transcript. *Nat Struct Mol Biol* 21(8):732–738.
- Abil Z, Denard CA, Zhao H (2014) Modular assembly of designer PUF proteins for specific post-transcriptional regulation of endogenous RNA. *J Biol Eng* 8(1):7.
- Cocquille S, et al. (2014) An artificial PPR scaffold for programmable RNA recognition. *Nat Commun* 5:5729.
- Filipovska A, Rackham O (2012) Modular recognition of nucleic acids by PUF, TALE and PPR proteins. *Mol Biosyst* 8(3):699–708.
- Moore FL, et al. (2003) Human Pumilio-2 is expressed in embryonic stem cells and germ cells and interacts with DAZ (Deleted in AZoospermia) and DAZ-like proteins. *Proc Natl Acad Sci USA* 100(2):538–543.
- Lunde BM, Moore C, Varani G (2007) RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol* 8(6):479–490.
- Wickens M, Bernstein DS, Kimble J, Parker R (2002) A PUF family portrait: 3'UTR regulation as a way of life. *Trends Genet* 18(3):150–157.
- Spassov DS, Jurcic R (2002) Cloning and comparative sequence analysis of PUM1 and PUM2 genes, human members of the Pumilio family of RNA-binding proteins. *Gene* 299(1-2):195–204.
- Wang X, Zamore PD, Hall TM (2001) Crystal structure of a Pumilio homology domain. *Mol Cell* 7(4):855–865.
- Wang X, McLachlan J, Zamore PD, Hall TMT (2002) Modular recognition of RNA by a human pumilio-homology domain. *Cell* 110(4):501–512.
- Cheong C-G, Hall TMT (2006) Engineering RNA sequence specificity of Pumilio repeats. *Proc Natl Acad Sci USA* 103(37):13635–13639.
- Zamore PD, Williamson JR, Lehmann R (1997) The Pumilio protein binds RNA through a conserved domain that defines a new class of RNA-binding proteins. *RNA* 3(12):1421–1433.

18. Miller MT, Higgin JJ, Hall TM (2008) Basis of altered RNA-binding specificity by PUF proteins revealed by crystal structures of yeast Puf4p. *Nat Struct Mol Biol* 15(4):397–402.
19. Qiu C, et al. (2012) Divergence of Pumilio/fem-3 mRNA binding factor (PUF) protein specificity through variations in an RNA-binding pocket. *J Biol Chem* 287(9):6949–6957.
20. Chen Y, Varani G (2011) Finding the missing code of RNA recognition by PUF proteins. *Chem Biol* 18(7):821–823.
21. Miller JC, et al. (2011) A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* 29(2):143–148.
22. Sander JD, et al. (2011) Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat Biotechnol* 29(8):697–698.
23. Ozawa T, Natori Y, Sato M, Umezawa Y (2007) Imaging dynamics of endogenous mitochondrial RNA in single living cells. *Nat Methods* 4(5):413–419.
24. Yamada T, Yoshimura H, Inaguma A, Ozawa T (2011) Visualization of nonengineered single mRNAs in living cells using genetically encoded fluorescent probes. *Anal Chem* 83(14):5708–5714.
25. Yoshimura H, Inaguma A, Yamada T, Ozawa T (2012) Fluorescent probes for imaging endogenous β -actin mRNA in living cells using fluorescent protein-tagged pumilio. *ACS Chem Biol* 7(6):999–1005.
26. Tilsner J, et al. (2009) Live-cell imaging of viral RNA genomes using a Pumilio-based reporter. *Plant J* 57(4):758–770.
27. Schwartz EC, Saez L, Young MW, Muir TW (2007) Post-translational enzyme activation in an animal via optimized conditional protein splicing. *Nat Chem Biol* 3(1):50–54.
28. Chong S, et al. (1996) Protein splicing involving the *Saccharomyces cerevisiae* VMA intein. The steps in the splicing pathway, side reactions leading to protein cleavage, and establishment of an in vitro splicing system. *J Biol Chem* 271(36):22159–22168.
29. Selgrade DF, Lohmueller JJ, Lienert F, Silver PA (2013) Protein scaffold-activated protein trans-splicing in mammalian cells. *J Am Chem Soc* 135(20):7713–7719.
30. Cao J, et al. (2013) Light-inducible activation of target mRNA translation in mammalian cells. *Chem Commun (Camb)* 49(75):8338–8340.
31. Cao J, Arha M, Sudrik C, Schaffer DV, Kane RS (2014) Bidirectional regulation of mRNA translation in mammalian cells by using PUF domains. *Angew Chem Int Ed Engl* 53(19):4900–4904.
32. Tilsner J (2015) Pumilio-based RNA in vivo imaging. *Methods Mol Biol* 1217:295–328.
33. Choudhury R, Tsai YS, Dominguez D, Wang Y, Wang Z (2012) Engineering RNA endonucleases with customized sequence specificities. *Nat Commun* 3:1147.
34. Chen F, Tillberg PW, Boyden ES (2015) Expansion microscopy. *Science* 347(6221):543–548.
35. Lee JH, et al. (2014) Highly multiplexed subcellular RNA sequencing in situ. *Science* 343(6177):1360–1363.
36. Delebecque CJ, Lindner AB, Silver PA, Aldaye FA (2011) Organization of intracellular reactions with rationally designed RNA assemblies. *Science* 333(6041):470–474.
37. Campbell ZT, et al. (2012) Cooperativity in RNA-protein interactions: Global analysis of RNA binding specificity. *Cell Reports* 1(5):570–581.
38. Wang Y, Opperman L, Wickens M, Hall TMT (2009) Structural basis for specific recognition of multiple mRNA targets by a PUF regulatory protein. *Proc Natl Acad Sci USA* 106(48):20186–20191.
39. Opperman L, Hook B, DeFino M, Bernstein DS, Wickens M (2005) A single spacer nucleotide determines the specificities of two mRNA regulatory proteins. *Nat Struct Mol Biol* 12(11):945–951.
40. Bernstein D, Hook B, Hajarnavis A, Opperman L, Wickens M (2005) Binding specificity and mRNA targets of a *C. elegans* PUF protein, FBF-1. *RNA* 11(4):447–458.
41. Cooke A, Prigge A, Opperman L, Wickens M (2011) Targeted translational regulation using the PUF protein family scaffold. *Proc Natl Acad Sci USA* 108(38):15870–15875.
42. Lu G, Hall TMT (2011) Alternate modes of cognate RNA recognition by human PUMILIO proteins. *Structure* 19(3):361–367.
43. Matsumoto K, Wassarman KM, Wolffe AP (1998) Nuclear history of a pre-mRNA determines the translational activity of cytoplasmic mRNA. *EMBO J* 17(7):2107–2121.
44. Sanjana NE, et al. (2012) A transcription activator-like effector toolbox for genome engineering. *Nat Protoc* 7(1):171–192.
45. Shi X, Herschlag D (2009) Fluorescence polarization anisotropy to measure RNA dynamics. *Methods Enzymol* 469:287–302.
46. Heyduk T, Ma Y, Tang H, Ebright RH (1996) Fluorescence anisotropy: Rapid, quantitative assay for protein-DNA and protein-protein interaction. *Methods Enzymol* 274:492–503.
47. Dinman JD, ed (2013) *Biophysical Approaches to Translational Control of Gene Expression* (Springer, New York).
48. Qu X, Chaires JB (2000) Analysis of drug-DNA binding data. *Methods Enzymol* 321:353–369.
49. Dell RB, Holleran S, Ramakrishnan R (2002) Sample size determination. *ILAR J* 43(4):207–213.
50. Bustin SA, et al. (2009) The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* 55(4):611–622.
51. Biggar KK, Dawson NJ, Storey KB (2012) Real-time protein unfolding: A method for determining the kinetics of native protein denaturation using a quantitative real-time thermocycler. *Biotechniques* 53(4):231–238.

Supplementary Materials

for the manuscript

"Programmable RNA binding protein composed of repeats of a single modular unit"

Katarzyna Adamala, Daniel A. Martin-Alarcon, Edward S. Boyden

Supplementary Results

General assembly of custom Pum repeats

The cloning of proteins like Pumilio, with highly repetitive structures, is challenging. Recent studies have presented assembly methods for Pumilio proteins based on the wild-type architecture, based on single-step Golden Gate cloning procedures (1). In this work we have modified a two-step Golden Gate cloning protocol previously developed for TAL effectors (2). Our protocol allows the efficient construction of assemblies with variable length and sequence.

At the beginning of our cloning procedure, we prepare a library of “monomers” with Golden Gate cloning overhangs, where each monomer is the sequence for a Pumilio unit (**Figs. S1B - S1C**). For PumHD architecture assembly, that library is comprised of 8 different units in 4 versions each: every unit of PumHD architecture in a variant that binds to each of the 4 canonical RNA bases. For Pumby, the library contains the four versions of one screened and optimized binding unit. These libraries can be used to construct PumHD or Pumby chains of any sequence. Each of the monomers is prepared in one of 5 (for the PumHD architecture) or in one of 6 (e.g., for the 6-mer, 12-mer, and 18-mer Pumby) variants, with GoldenGate cloning overhangs placing it in the correct position of circular cloning intermediate composed of 5 (for PumHD architecture) or 6 (for 6-mer, 12-mer, and 18-mer Pumby) Pum units (**Fig. S1D**). To prepare custom assemblies, in the first Golden Gate reaction we prepare circular cloning intermediate pentamers (for PumHD architecture) or hexamers (for 6-mer, 12-mer, or 18-mer Pumby chains) of Pum units. For Pumby chains that are not a multiple of 6, one can of course combine different sets of building blocks, e.g. a 6-mer and a 4-mer can be combined to make a 10-mer. Those circular n-mers are subsequently linearized to produce linear cloning intermediates (**Fig. S1E**). The linear intermediates are then assembled into the destination vector in the second Golden Gate reaction (**Fig. S1F**). For each PumHD chain, two pentamers were assembled into the final vector (total of 10 units: 8 RNA-binding units plus non-binding units 0 and 9). For this work, we have created several destination vectors compatible with Golden Gate reactions. These vectors contain point mutations to remove BsaI

enzyme sites from the CMV and UBC promoters, from the pCI vector backbone, and from the β La antibiotic resistance gene (see **Material and Methods** for details).

Preliminary investigation of transcription / translation modulation by Pum

We chose as a Pum target the Activating Transcription Factor 4 (ATF4) mRNA, whose transcription and translation is induced by cell exposure to thapsigargin (3, 4). Using Pum-anchored split GFP (**Fig. S2A**) targeted to different parts of the ATF4 gene, we longitudinally estimated protein production levels in living cells. We observed significant differences of Pum-mediated GFP reconstitution in response to thapsigargin (**Fig. S2B**; $P < 0.0001$ for post-hoc comparison of the 12 hour thapsigargin positive case and the other two cases; Tukey's range test after ANOVA with factors of specific target and experimental condition; see **Table S9** for the full statistics related to **Fig. S2**, including n's for replicates; see **Table S10** for all the sequences used in this figure). The increases in Pum-mediated GFP reconstitution were associated with qualitative increases in the ATF4 protein, as measured by ELISA (**Fig. S2C**; $P < 0.0001$ for post-hoc comparison of the 12 hour thapsigargin positive case and the other two cases; Tukey's range test after ANOVA with factors of specific target and experimental condition). We observed no significant difference in performance between Pumby8 and PumHD in this assay (**Fig. S2B**; $P = 0.3248$ for factor of Pum type; two-way ANOVA with factors of experimental condition and Pum type).

As with split GFP reconstitution before (**Figs. 2H** and **4C**), we tested the tolerance of our endogenous gene translation monitoring assay to mismatches between the Pum protein and its target RNA sequence. We mutated two particular Pum sequences (one PumHD and one Pumby8) to contain 0, 1, 2, or 3 mismatches. In this assay, we observed that 1 to 2 mismatches decrease binding affinity (**Fig. S2D**). During our tests, both the particular protein sequence we were testing and the number of mismatches had an effect on overall binding (**Fig. S2D**; $P = 0.0043$ for factor of Pum protein and $P < 0.0001$ for factor of mismatch number; two-way ANOVA with factors of Pum protein and mismatch number; $n = 3$ biological replicates each). We observed some luciferase reconstitution above baseline for 1 mismatched unit, but in this case even 2 mismatches were sufficient to effectively eliminate luciferase reconstitution (**Fig. S2D**; $P = 0.9577$ for comparison of 2 vs. 8 mismatches and $P > 0.9999$ for 3 vs. 8 mismatches; Dunnett's *post-hoc* test across values of mismatch number, after the previous ANOVA).

Sequence of Pum targeting gene of interest for quantification of translational activity

mRNA in live cells has complex folding that is often not well understood (5, 6). As with all technologies targeting RNA in live cells, it is advised to utilize multiple Pum targeting sequences to validate a lack of

nonspecific binding or a lack of secondary structure that prevents binding to the targeted region of interest. In the case of imaging translation, in addition to the sequences reported in **Table S6**, we tested three additional pairs of sequences targeting the GFP gene and two targeting the BLA gene. We observed either no measureable Pum-mediated split luciferase reconstitution (suggesting that Pum binding to the target mRNA does not happen, presumably due to the secondary structure of the mRNA region), or split luciferase reconstitution not corresponding to the translation activity of the gene (suggesting the Pum binding accidentally targets native, constitutively expressed genes). See **Table S17** for the list of those sequences.

Promiscuity of Pum unit 4

It has been previously suggested that unit 4 of PumHD does not distinguish between U, A, or C nucleotides. We investigated this by measuring the K_d of binding to the target, with the nucleotide binding Pum unit 4 mutated to each of the 4 possible bases (A, U, C and G). Indeed, the K_d of unit 4 binding to A, U and C is similar, whereas introducing G on this position in the RNA template causes a significant decrease in binding affinity. See **Fig. S9** for binding curves.

Pum-mediated mRNA silencing, and orthogonality tests

A general endonuclease PIN domain has been previously fused with wild type PumHD and 5 different Pum mutants, creating a sequence-specific nuclease that works well in cultured cells (7). In this experiment we demonstrate that Pumby can be fused to the PIN domain to direct nuclease activity towards transcripts in cultured cells. We used both the PumHD architecture and Pumby chains to create series of Pum-PIN constructs targeting different areas of the Firefly luciferase gene. We tested several PumHD architecture and Pumby variants, and showed silencing of the luciferase in response to the Pum-mediated nuclease activity (**Fig. S4**). We prepared a bicistronic reporter vector containing Firefly luciferase (the gene targeted for silencing) and Renilla luciferase (used as a control for cell density, transfection efficiency and non-specific nuclease activity). We co-expressed this double luciferase vector with the vector containing Pum-PIN constructs (where Pum is either PumHD architecture-based or Pumby module-based, binding different RNA recognition sequences within the Firefly luciferase gene; see **Table S14** for all sequences used). We also prepared a control (“No Pum-PIN”) where the Pum-PIN was left out and only the reporter plasmid was present, with PumHD protein not targeting any sequence of the luciferase vector (**Fig. S4A**). It has been previously shown that PIN domain alone, without an RNA

binding protein fused to it, does not exhibit gene silencing activity. Therefore, the RNA binding protein domain is necessary to localize the PIN nuclease domain to a target (7).

In each experiment, we co-transfected HeLa cells with one of the Pum-PIN vectors (where Pum was either PumHD or Pumby) with the double luciferase vector. We observed decreased copy number of the Firefly luciferase mRNA relative to the Renilla luciferase mRNA, as measured by RT-qPCR experiments (**Fig. S4B**; $P = 0.0003$ for factor ‘Pum Target Site’; one-way ANOVA; see **Table S13** for full statistics for **Fig. S4**, including post-hoc tests indicating which specific Pum targets differed from the control condition), as well as decreased Firefly luminescence relative to Renilla (**Fig. S4C**; $P < 0.0001$ for factor ‘Pum Target Site’; one-way ANOVA). For some of the Pum target sites, the difference in RT-qPCR cycles equated to a reduction of Firefly vs. Renilla by 2-4 cycles, with the corresponding protein reduction of around 70%. It is worth noting that, as it is the case with siRNA and all other techniques relying on binding of a tool to a gene (RNA or DNA) in live cells, there is the potential for non-specific interactions caused by binding of the tool to sequences similar to the target sequence. Also, secondary structure formation on the mRNA of the targeted gene can prevent efficient binding to that region. Therefore, it is necessary to test several candidate sequences targeting different areas of the gene of interest, as with all RNA-binding tools.

As a final test of Pumby and PumHD, we checked the intrinsic orthogonality between Pum proteins designed for various target sequences (**Fig. S3**). Specifically, we tested 7 of the Pums used in **Figs. 5** and **S2** for their ability to work without cross-talk, using the luciferase reconstitution assay of **Fig. 5**. Instead of using the full genes for GFP and BLA, we created a new set of target transcripts in which the required target sequences were placed in a landing site at the end of the coding sequence for APEX2 peroxidase (8), which serves as a transfection control (see **Table S12** for the landing site sequences). A match between the Pum and the landing site sequence was key for Firefly luciferase reconstitution (**Fig. S3**; $P < 0.0001$ for factor of target match; three-way ANOVA with factors of Pum protein, mRNA target, and target match; $n = 3$ biological replicates; see **Table S11** for the full statistics for **Fig. S3**). Throughout this assay, Pumby8 was indistinguishable from the PumHD equivalents (**Fig. S3**; $P = 0.0709$ for factor of Pum type).

Supplementary Figures

Fig. S1

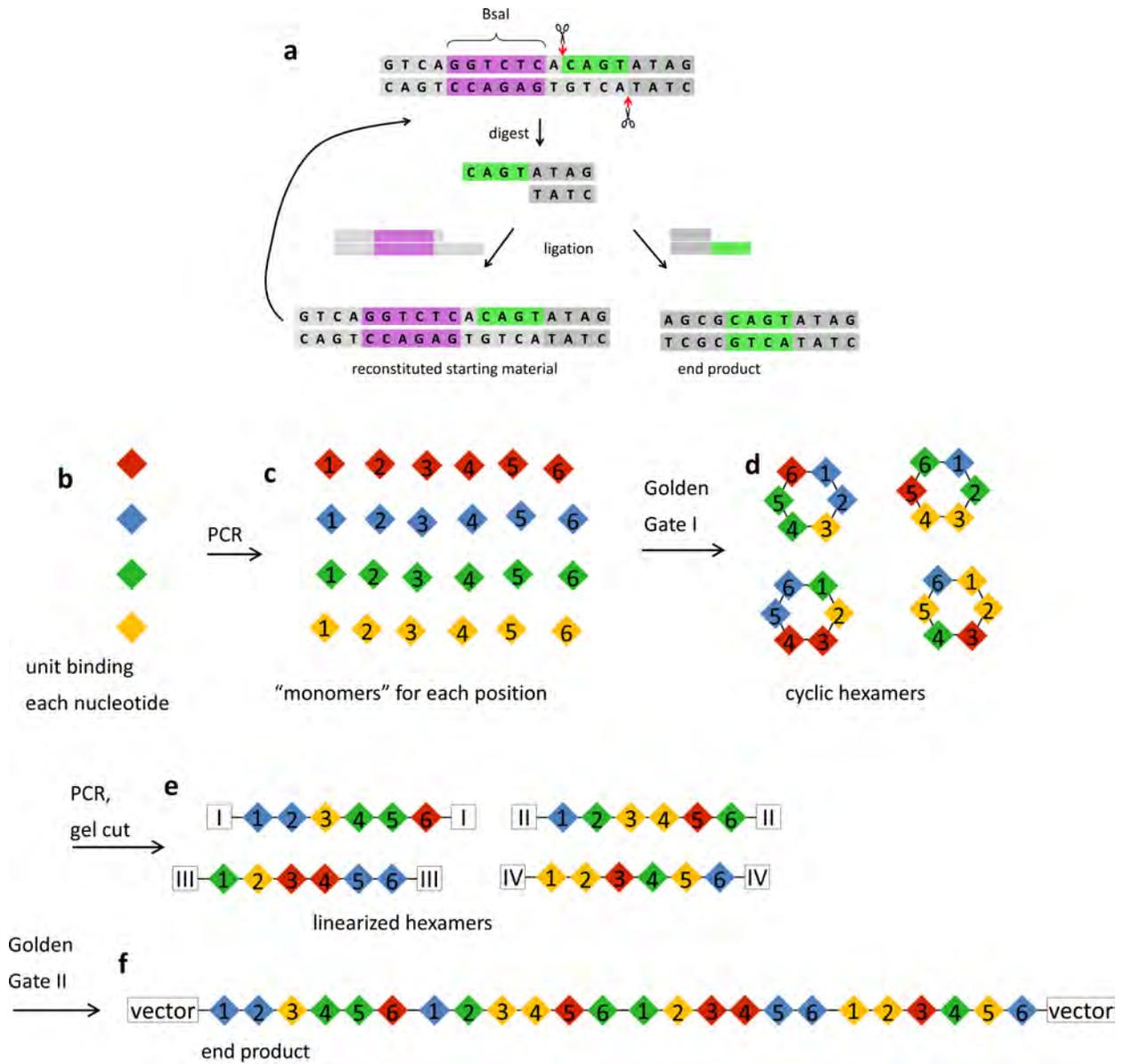
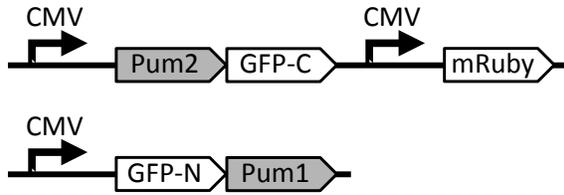


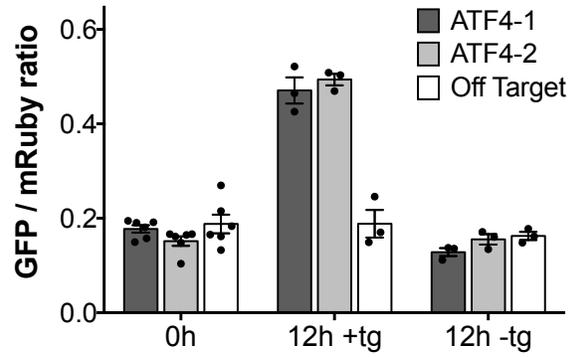
Fig. S1. Golden Gate reaction for PumHD and Pumby creation. **A**, The rationale of the Golden Gate reaction strategy. It is a two-step process of digestion with type II restriction enzymes and subsequent ligation with T7 ligase. The enzyme recognition sequence (for BsaI, 5'-GGTCTC-3') is one nucleotide removed from the cut site. The “sticky end” created after the enzyme digestion on the strand that does not contain the cut site can be ligated back with its original partner, reconstituting the original site (which can then react again), or it can be ligated into complementary “sticky end” created with the use of reversed BsaI site. The enzyme recognition site is no longer present in the latter case, resulting in a stable end product. The reaction is repeated 15-25 times (with digestion at 37°C and ligation at 16-20°C), driving the reaction toward product formation. **B**, The first step in making custom Pumby and PumHD architecture assemblies is to prepare a library of “monomers”, where each monomer encodes for one of the four canonical Pumby modules (**Fig. 3B**) or the appropriate PumHD unit (**Fig. 1C**), as needed to bind the corresponding RNA base. **C**, We use PCR to add Golden Gate sites to the monomers; the overhangs determine the position that the monomers will acquire in a circular cloning intermediate. We have labeled these intermediates “cyclic hexamers” because they may contain up to 6 PumHD or Pumby monomers. The number of monomers that build the cloning intermediate is always 5 for PumHD (since the overall PumHD chain always contains 10 units), but varies for Pumby because it depends on the final length of the chain. Pumby6 can be built with a single cloning intermediate; Pumby8 can be made with one 5-mer intermediate and one 3-mer, or with two 4-mers; Pumby10 would take two 5-mer intermediates. Shown are 6 monomers, for the Pumby case. **D**, An initial Golden Gate reaction assembles the monomers into a circular pentamer (for PumHD) or other n-mer (for Pumby). Shown are hexamers. **E**, We use PCR and agarose electrophoresis purification to amplify the circular hexamers into linear hexamers that contain the cloning overhangs (white squares) for a second Golden Gate reaction. **F**, A second Golden Gate reaction assembles the hexamers into a destination vector. We have prepared a mammalian expression destination vector with point mutations in the chimeric intron of the CMV promoter and in the bLa antibiotic resistance gene to remove BsaI sites, another mammalian expression destination vector based on the pCI backbone but with the human UBC promoter, and the bacterial expression vector pBad with BsaI sites removed; see **Materials and Methods** for details.

Fig. S2

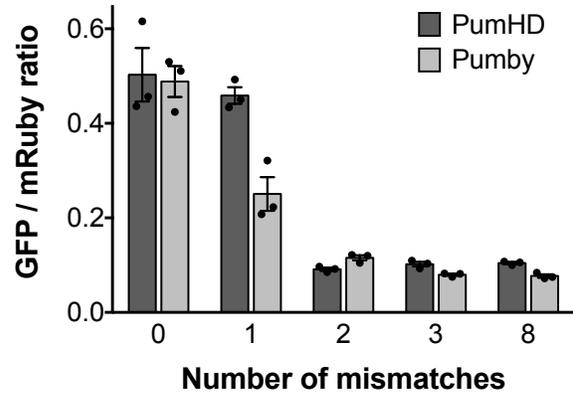
A. ATF4 reporter plasmids



B.



D.



C.

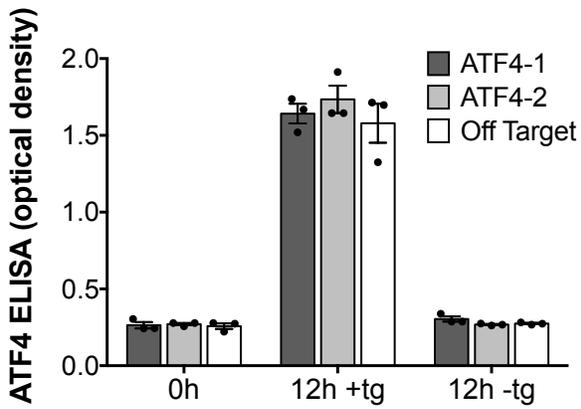


Fig. S2. Measurement of endogenous gene expression. **A**, Schematic of reporter plasmid as in **Fig. 5A**, but with Pum1 and Pum2 aimed at target sites in the mRNA for ATF4 (for a full list of the target binding sequences used in this figure, see **Table S10**; for the full statistics, see **Table S9**). **B**, Pum-guided GFP reconstitution using the reporters of **A**. “Time 0” represents the beginning of the experiment (6 biological replicates). Half of the samples (3 biological replicates) were exposed to thapsigargin (+tg), half were not (-tg), and both were imaged 12 hours later. **C**, ATF4 protein, quantitated via ELISA, in samples prepared as in **B**. **D**, Sensitivity of translation measurement to mismatches between the Pum proteins and endogenous RNA targets. For the two RNA target sites used in **B**, we created protein variants with 1, 2, 3, and 8 mismatches and tested them under the same conditions of thapsigargin-mediated ATF4 expression. Error bars are s. e. m. throughout this figure.

Fig. S3

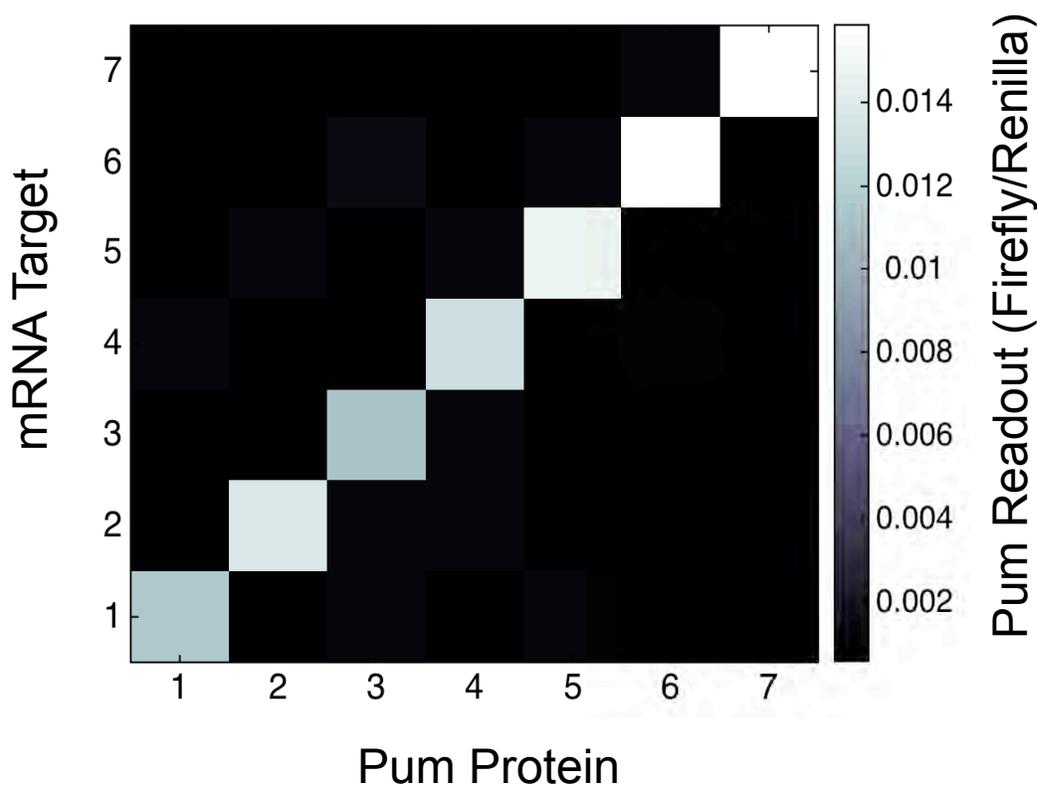
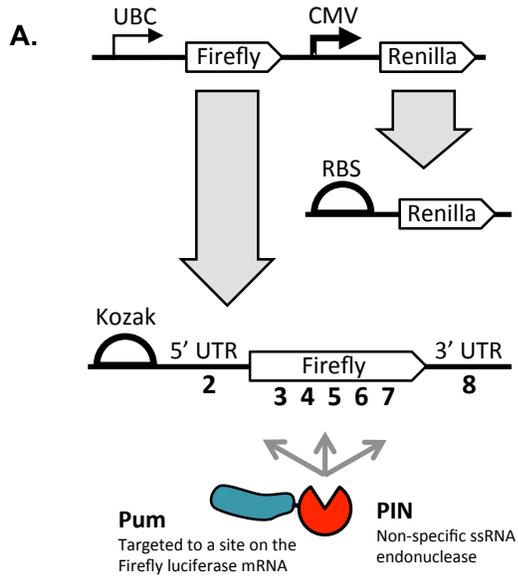
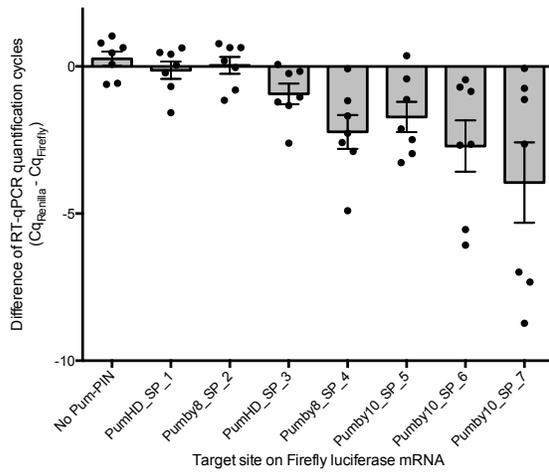


Fig. S3. Orthogonality of binding for modular RNA-binding proteins. We tested seven of the Pums (all targeting 8-mer sequences) used in **Figs. 5** and **S2** for crosstalk between each other, as measured by Firefly luciferase reconstitution normalized to Renilla luciferase expression. We created a series of seven target plasmids, each containing an APEX2 peroxidase (8) (as a transfection control; see **Materials and Methods**) coding sequence with a 24-bp landing site inserted immediately before the stop codon. This landing site, as for those used in **Fig. 2B**, contains two Pum binding targets (for a full list of the target binding sequences used in this figure, see **Table S12**; for the full statistics, see **Table S11**). One of the RNA targets was designed, across all 7 landing sites, to bind Pumby8_TM_GFP_4B carrying C-terminal split Firefly luciferase (sequence CAGCGUGU), and the other RNA target was designed to bind only one of seven Pums carrying N-terminal split Firefly luciferase. The plasmids carrying the Pums are as depicted in **Fig. 5A**; the plasmid encoding for N-terminal Firefly luciferase also encodes for Renilla luciferase, which we use to normalize for cell count and transfection efficiency. Thus, the values reported in this graph have units of Firefly/Renilla luminescence (arbitrary units) and are 3 biological replicates (see **Materials and Methods**).

Fig. S4



B.



C.

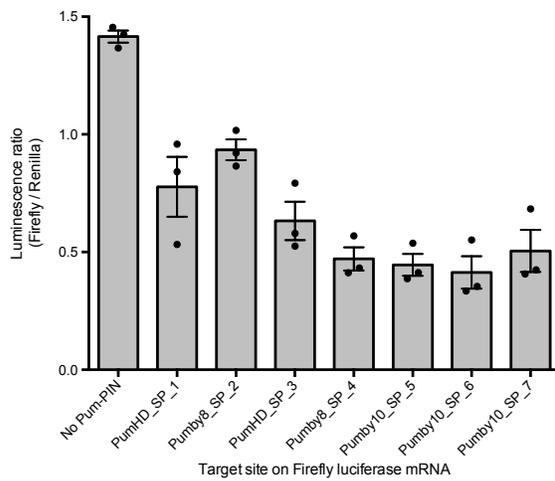


Fig. S4. Targeted transcript silencing via Pum-endonuclease fusion protein. **A**, The RNA silencing assay uses a bicistronic target vector in which Firefly luciferase and Renilla luciferase are expressed from independent promoters. This results in two separate transcripts, of which only the one carrying Firefly luciferase is targeted. The Firefly mRNA is targeted by a protein fusion of Pum with the non-specific ssRNA endonuclease PIN. We targeted 7 different sites on the Firefly mRNA located before, within, and after the coding sequence (for complete list of target sequences, see **Table S14**; Pums were uniquely identified for easy reference as PumHD_SP or Pumby[number]_SP, where [number] represents the size of that particular Pumby and SP stands for “silencing, PIN”). **B**, RT-qPCR measurement of Renilla vs. Firefly relative transcript levels, expressed as differences in quantification cycle (C_q difference), measured from HeLa cells transfected with Pum-PIN vectors targeted to various sites on the Firefly luciferase mRNA, as indicated by numbers in panel **A**. Error bars are s. e. m. for 7 biological replicates and dots represent individual data points. **C**, Ratio of Firefly luciferase luminescence to Renilla luciferase luminescence for HeLa cells transfected with Pum-PIN vectors targeted to the sites numbered in panel **A** on the Firefly luciferase mRNA. Error bars are s.e.m. for 3 biological replicates and dots represent individual data points. See **Fig. S5** for mFold predictions of the RNA structures of those target sites.

Fig. S5

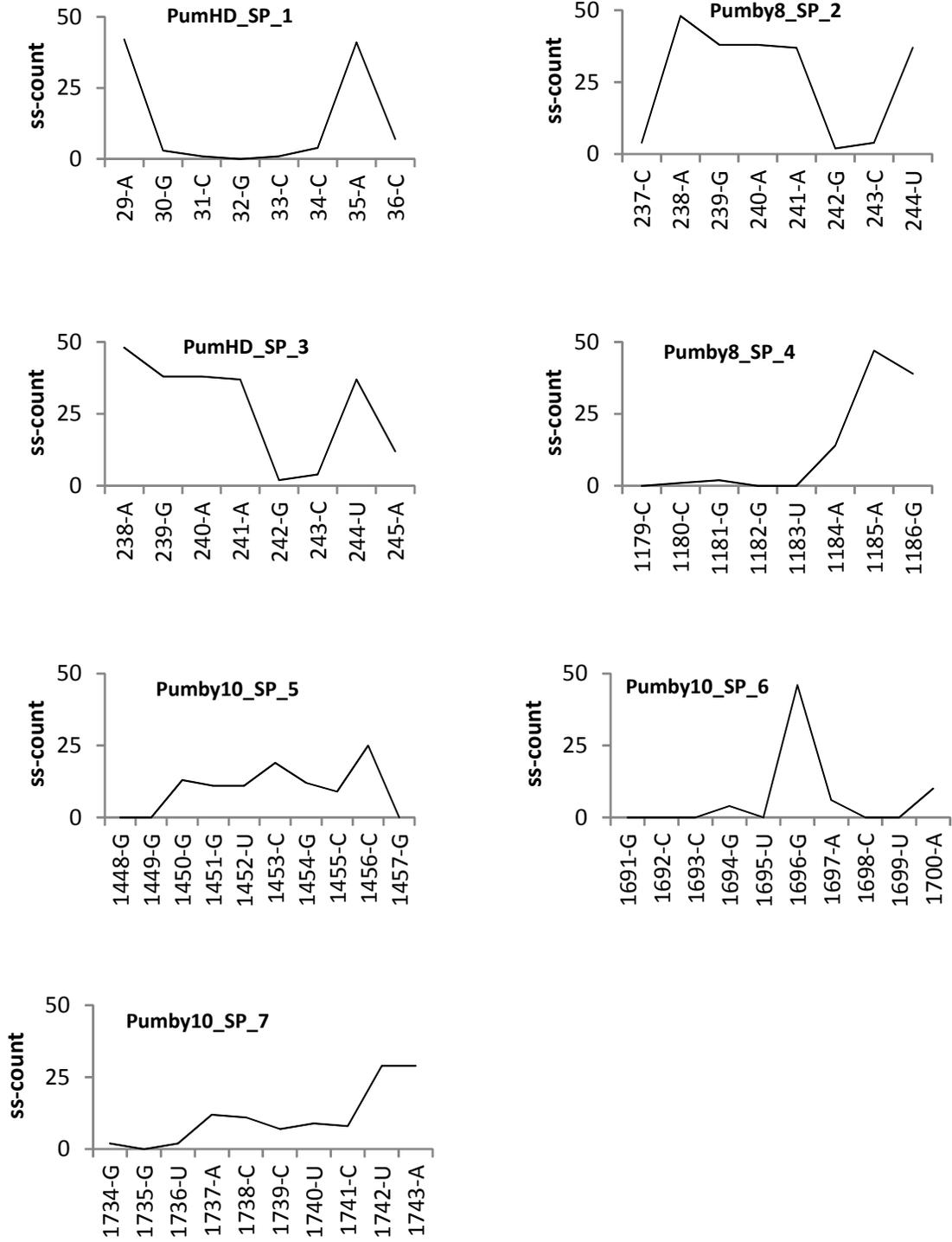
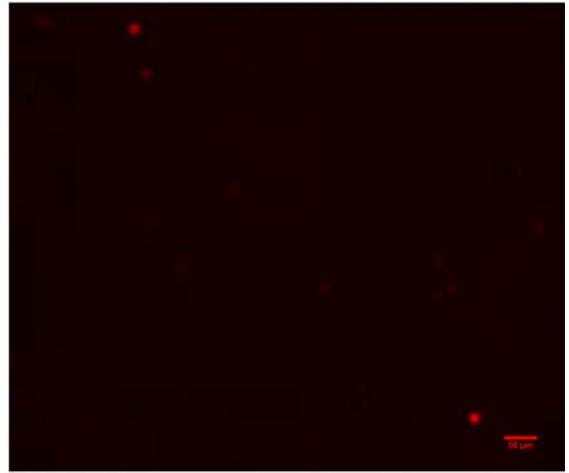
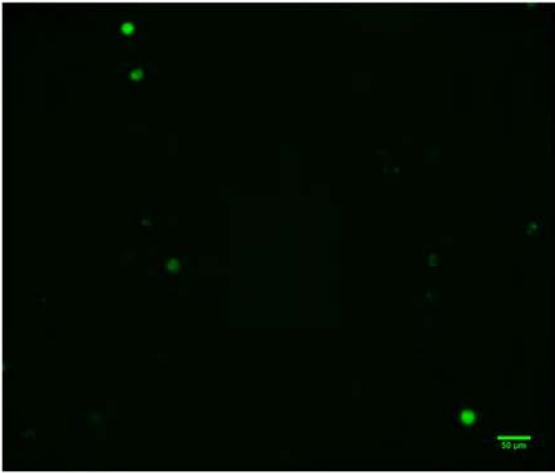


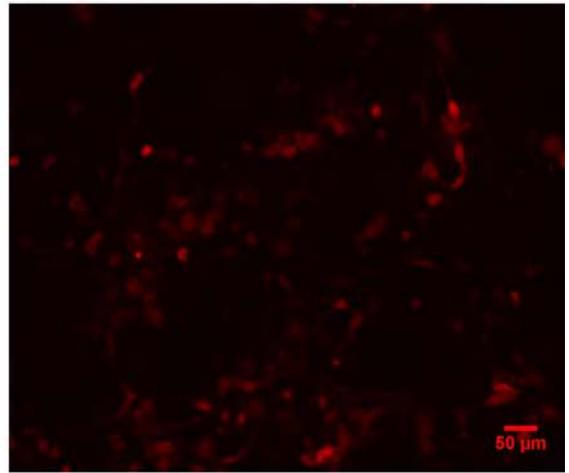
Fig. S5. Prediction of single-stranded RNA prevalence in Pum target sites for Pum-PIN. Single-stranded RNA prevalence (ss-count), as defined by the mFold web server (9), is presented for the Pum target sequences in the luciferase plasmid in **Fig. S4**. One hypothesis that emerges is that the Pum target sequences with the best silencing results (as observed by changes in mRNA count and luciferase activity, **Fig. S4**) have a high probability of single stranded sequence near the 5' end of the RNA.

Fig. S6

A.



B.



C.

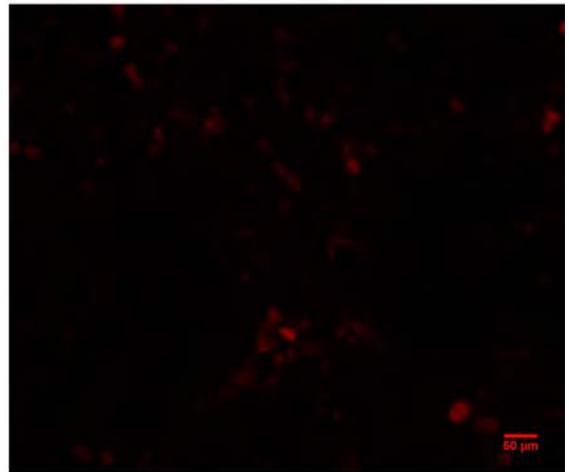


Fig. S6. Examples of failed Pumby candidates. Results of the Pum-mediated GFP reconstitution assay (as in **Figs. 2C-2E, 3C-3D**), with Pumby code candidates (**Fig. 3B**). The Pumby variants were prepared using different single units of PumHD as a repeated 8-mer, with different stacking amino acids on position AA2. All images were taken 72 hours post-transfection, as was also done for **Figs. 2** and **3**. All pictures show constructs with on-target Pum landing site, with the sequence of Pum1, AUAGAUGU and the sequence of Pum-2, GCGAGCAC. **A**, Pumby candidate made of unit 3 of PumHD (see **Fig. 1C** for the PumHD units) with stacking AA2 R. **B**, Pumby candidate made of unit 3 of PumHD with stacking AA2 Y. **C**, Pumby candidate made of unit 6 of PumHD with stacking AA2 R. See **Table S21** for a list of all the Pumby candidates that we tested.

Fig. S7

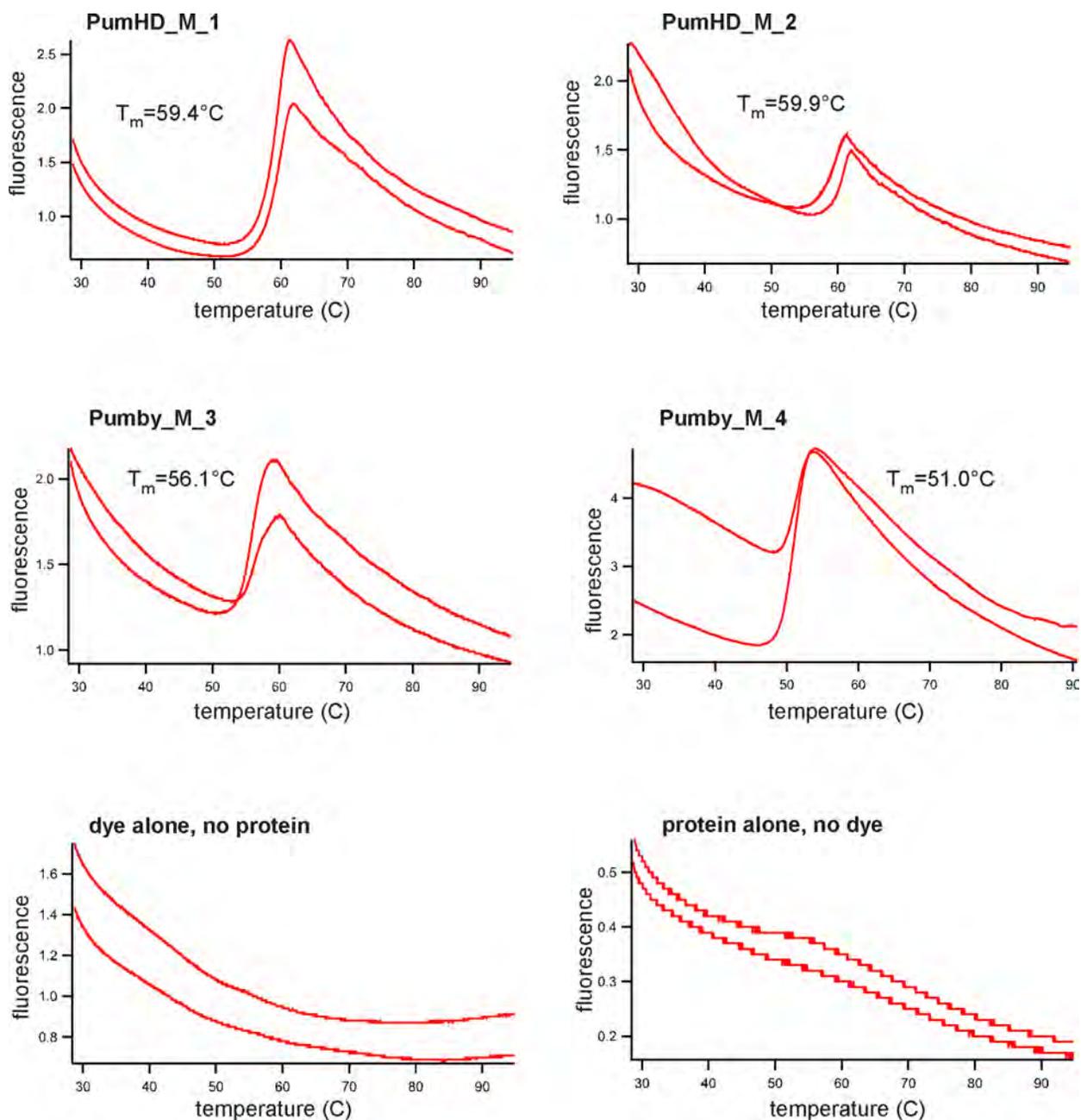


Fig. S7. Stability of Pum variants measured via a thermal shift assay. Each plot shows 2 representative melt graphs for each protein. See **Materials and Methods** for experimental details. For the list of sequences used in this figure, see **Table S15**.

Fig. S8

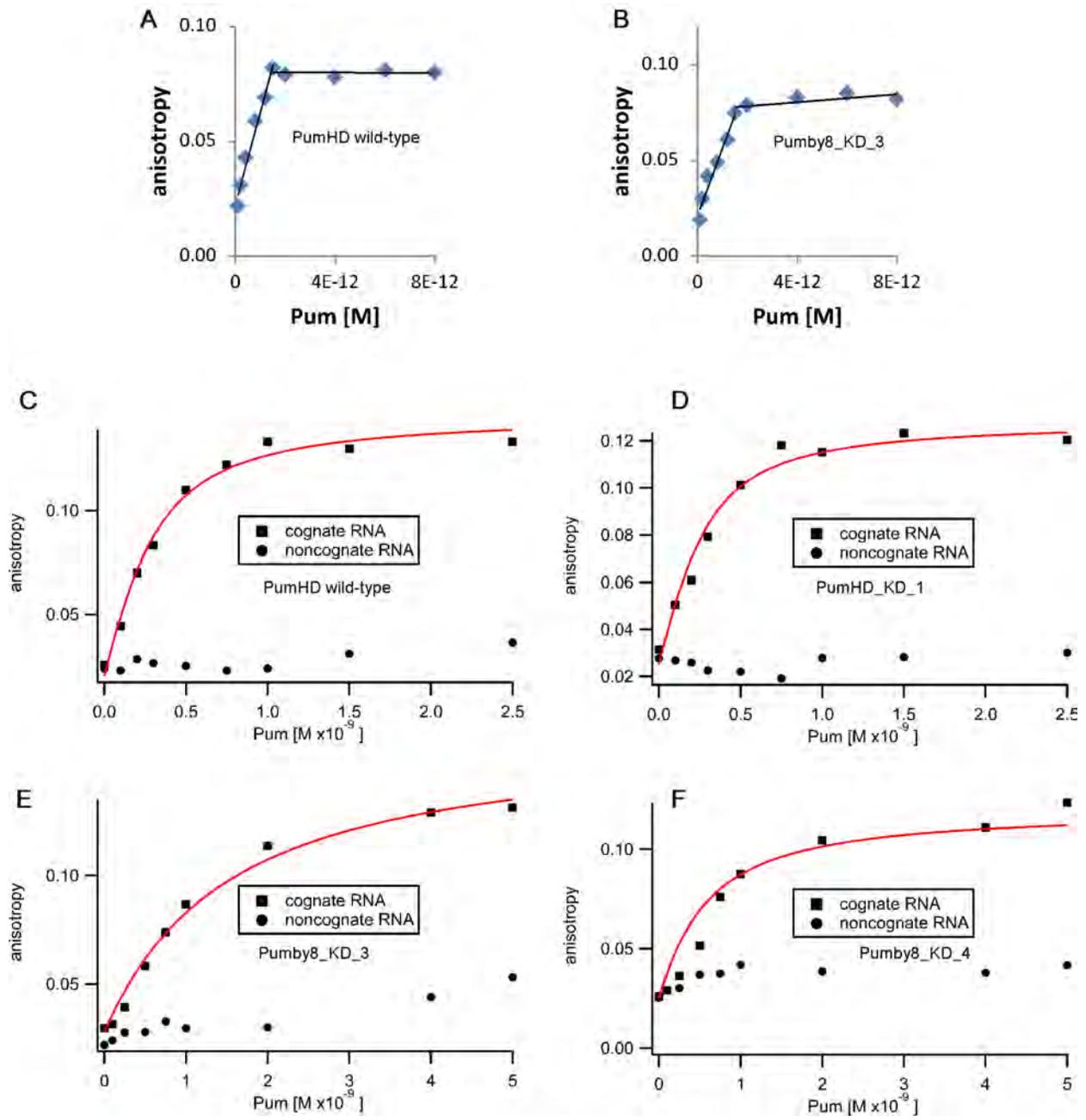


Fig. S8. Cell-free measurement of binding affinity of modular RNA binding proteins. Throughout this figure, the cognate RNA is always the sequence exactly matching the whole Pum protein binding sequence, flanked as CCAGAAU*Pum_sequence*UUCG. The sequence of the bases flanking the RNA target sequence was selected from previously published studies (1, 10). **A**, Saturation experiment (Job plot) to estimate the active fraction of purified protein, for PumHD wild-type sequence. For a full list of the target binding sequences used in this figure, see **Table S16**. **B**, Saturation experiment (Job plot) to estimate the active fraction of purified protein, for Pumby8_KD_3; Pums were uniquely identified for easy reference as PumHD_KD or Pumby8_KD, where KD refers to binding affinity. **C-F**, K_d measurement for cognate RNA (with nonlinear fit) and non-cognate RNA for various Pum variants, listed below. The K_d values for non-cognate RNA targets were not estimated (the attempted fits did not converge). For estimated fractions of the active protein, the calculated K_d values, and standard deviations of the fits, see **Table S16**. **C**, PumHD wild type sequence. **D**, PumHD_KD_1. **E**, Pumby8_KD_3. **F**, Pumby8_KD_4.

Fig. S9

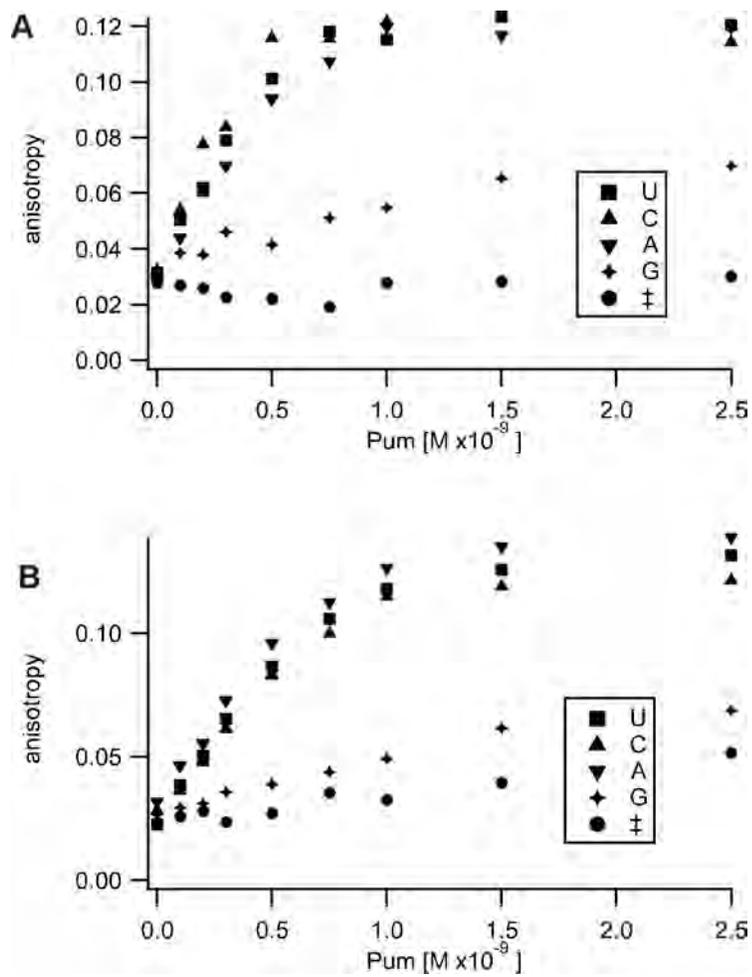


Fig. S9. The promiscuity of Pum unit 4. The proteins and cognate RNAs are as used in the experiments shown in **Fig. S8** (**A**, PumHD AUAUAUGU; **B**, PumHD CGUAUGAC), except that the base binding to Pum unit 4 was replaced with the base indicated for each trace. See **Supplementary Results** for details.

Fig. S10

Split GFP reconstitution after swapping Pum1 and Pum2

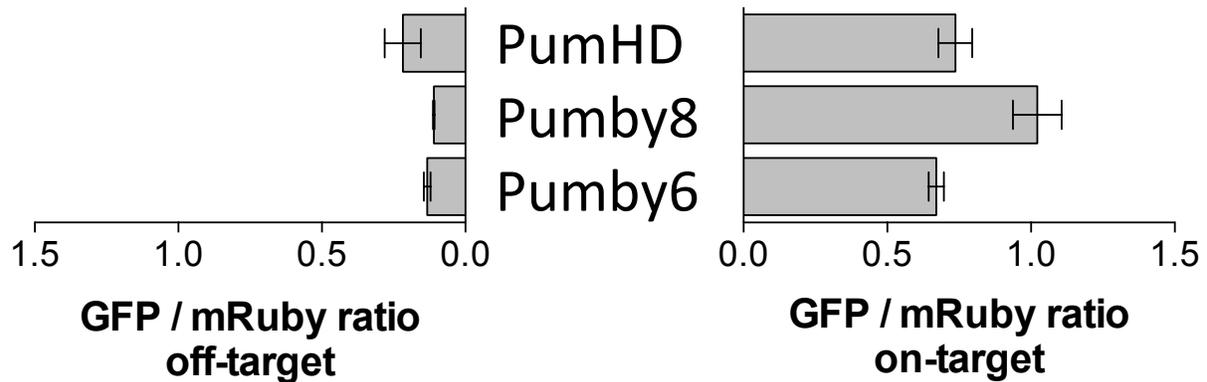


Fig. S10. Split GFP reconstitution values as in **Fig. 2F**, but for three cases where the sequence most commonly used for Pum2 (see **Fig. 2B**) was swapped with the sequence for Pum2. This was in order to ensure that our results through **Figs. 2** and **4** did not depend on the arbitrary choice of binding sequences for Pum1 and Pum2. SWAP data, as we call it, was part of the statistical analysis for **Figs. 2** and **4**. See **Tables S2** and **S4** for the full list of sequences used in this figure.

Fig. S11

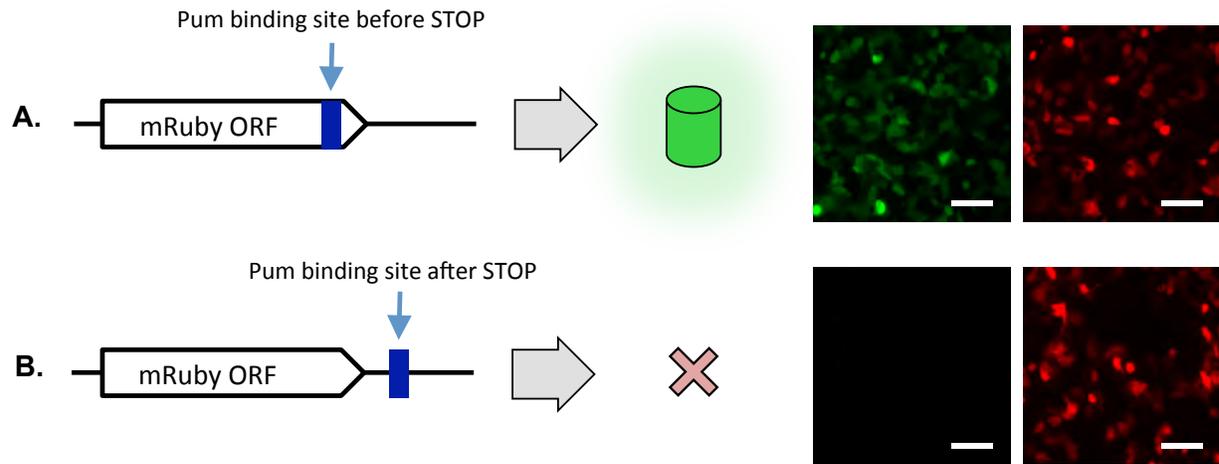


Fig. S11. Pum-mediated GFP reconstitution is much more effective when the binding site is located within an actively translated open reading frame. **A**, A Pum binding site within the open reading frame of mRuby leads to GFP reconstitution. **B**, The same binding site, placed after a STOP codon, produces no GFP reconstitution. In both cases, the Pum is PumHD with protein sequence AUAGAUGU (SWAP from **Table S2**). The red-green assay is the same as used in Figures 2 and 4.

Supplementary Tables

Table S1

Statistics for **Fig. 2**.

No samples were excluded from the statistical analysis. The threshold for significance throughout this figure is $\alpha = 0.05$; $n = 3$ biological replicates throughout this figure.

Abbreviations

Diff	Difference
SE	Standard Error
DoF	Degrees of Freedom
Sig?	Significant?
ns	Not significant

Fig. 2F

Two-way ANOVA with factors ‘ON or OFF Target’, which compares on-target against off-target data (Both sides of **Fig. 2F**); and ‘Target Sequence’, which compares among the bars. On-target pairs presented a GFP/mRuby ratio on average 6.1-fold greater (standard deviation of 1.37 fold) than the corresponding ratio for off-target pairs. This statistical analysis included all the bars in **Fig. 2F**, and also one more data point called SWAP. SWAP represents one additional test case in which we swapped the sequences usually tested with Pum1 and Pum2, in order to make sure that our results weren’t the result of positional effects (see **Table S2**). The values for SWAP were included in all the following global analyses, and these values were also plotted in **Fig. S10**.

Source of Variation	% of total variation	P value	Sig?	F (DFn, DFd)
Interaction	2.103	< 0.0001	Yes	F (24, 100) = 4.416
Target Sequence	2.132	< 0.0001	Yes	F (24, 100) = 4.476
ON or OFF Target	93.78	< 0.0001	Yes	F (1, 100) = 4726

Dunnett’s multiple comparisons test across the values of ‘Target Sequence’, against the value for wild-type PumHD (i.e., 4-U), after the two-way ANOVA with factors ‘ON or OFF Target’ and ‘Target Sequence’. Note that, for 22 of the 24 variants, on-vs-off target behavior was indistinguishable from that of the wild-type PumHD (i.e., 4-U), and for two variants on-target binding was significantly enhanced.

Dunnett’s multiple comparisons test	Mean Diff.	95% CI of diff.	Sig?	Summary	Adjusted P Value
-------------------------------------	------------	-----------------	------	---------	------------------

SWAP vs. 4-U	0.01249	-0.08717 to 0.1121		ns	0.9995
8-C vs. 4-U	0.1821	0.08242 to 0.2817	Yes	****	< 0.0001
8-G vs. 4-U	0.1518	0.05218 to 0.2515	Yes	***	0.0003
8-A vs. 4-U	-0.0121	-0.1118 to 0.08756		ns	0.9995
7-A vs. 4-U	0.02028	-0.07937 to 0.1199		ns	0.9991
7-U vs. 4-U	-0.03792	-0.1376 to 0.06174		ns	0.9748
7-C vs. 4-U	-0.003842	-0.1035 to 0.09582		ns	0.9999
6-C vs. 4-U	0.01648	-0.08318 to 0.1161		ns	0.9993
6-G vs. 4-U	0.03144	-0.06822 to 0.1311		ns	0.9939
6-A vs. 4-U	-0.01318	-0.1128 to 0.08648		ns	0.9995
5-U vs. 4-U	-0.003792	-0.1035 to 0.09587		ns	0.9999
5-C vs. 4-U	0.01451	-0.08515 to 0.1142		ns	0.9994
5-G vs. 4-U	0.03057	-0.06909 to 0.1302		ns	0.9942
4-A vs. 4-U	0.009042	-0.09062 to 0.1087		ns	0.9997
4-C vs. 4-U	0.02631	-0.07335 to 0.1260		ns	0.9957
3-U vs. 4-U	0.06305	-0.03661 to 0.1627		ns	0.5056
3-C vs. 4-U	0.05724	-0.04242 to 0.1569		ns	0.6416
3-G vs. 4-U	-0.005045	-0.1047 to 0.09461		ns	0.9998
2-C vs. 4-U	0.03478	-0.06488 to 0.1344		ns	0.9866
2-G vs. 4-U	0.01436	-0.08530 to 0.1140		ns	0.9994
2-A vs. 4-U	0.01229	-0.08737 to 0.1120		ns	0.9995
1-U vs. 4-U	0.07584	-0.02382 to 0.1755		ns	0.2613
1-C vs. 4-U	0.05136	-0.04830 to 0.1510		ns	0.7778
1-G vs. 4-U	0.07415	-0.02551 to 0.1738		ns	0.2876

Each bar in **Fig. 2G** vs. the combined on-target replicates of **Fig. 2F** (excluding SWAP).

Unpaired two-sided t-tests, one for each bar of **Fig. 2G** (n = 3 biological replicates) against the aggregate on-target replicates of **Fig. 2F** (excluding SWAP). We use the Bonferroni method to correct for multiple comparisons and do not assume consistent standard deviation between bars.

Bar	Sig?	P value	Mean on-target	Mean off-target	SE of difference	t ratio	DoF
U NRE A	Yes	2.15102E-07	1.14845	0.777674	0.0647659	5.72483	73
C NRE U	Yes	5.41003E-07	1.13356	0.777674	0.06475	5.49624	73
G NRE C	Yes	0.000011557	1.08073	0.777674	0.0643563	4.7091	73
A NRE G		0.533228	0.736956	0.777674	0.0650387	0.626061	73

Fig. 2H

Two-way ANOVA with factors of “Pum Tested” (comparing between two particular proteins into which we introduced mismatch mutations) and “Mismatch number”.

Source of Variation	P value	Summary	Sig?	F (DFn, DFd)	DoF
Interaction	0.2399	ns	No	F (4, 20) = 1.500	P = 0.2399
Mismatch number	< 0.0001	****	Yes	F (4, 20) = 114.3	P < 0.0001
Pum protein	0.0327	*	Yes	F (1, 20) = 5.264	P = 0.0327

Dunnett’s multiple comparison test across the values of “Mismatch number”, and against the off-target case, after the ANOVA with factors of “Pum Tested” and “Mismatch number”.

Test details	Mean Diff.	SE of diff.	Adjusted P Value	Summary	Sig?
0 vs. 8	1.021	0.06038	< 0.0001	****	Yes
1 vs. 8	0.7565	0.06038	< 0.0001	****	Yes
2 vs. 8	0.3222	0.06038	0.0001	***	Yes
3 vs. 8	0.004359	0.06038	0.9999	ns	No

All raw datapoints are plotted in the figure so that the entire dataset is available to the reader, since each individual condition has less than 5 datapoints, and thus formal evaluations of normality are not appropriate.

Standard deviation of each on-target row in **Fig. 2F**.

Bar	Mean	SD	N
SWAP	0.7369556	0.1013903	3
8-C	1.086346	0.03166456	3
8-G	1.034407	0.172783	3
8-A	0.689422	0.1035738	3
7-A	0.7724454	0.07534922	3
7-U	0.6638123	0.05124879	3
7-C	0.741838	0.03347131	3

6-C	0.7393286	0.04098918	3
6-G	0.803802	0.0312377	3
6-A	0.7136006	0.02238935	3
5-U	0.7412313	0.04215925	3
5-C	0.754919	0.05954938	3
5-G	0.7867377	0.0463917	3
4-A	0.7330766	0.02671649	3
4-U	0.727691	0.04971226	3
4-C	0.7694757	0.03153829	3
3-U	0.765128	0.0708895	3
3-C	0.764703	0.07300427	3
3-G	0.7116267	0.103213	3
2-C	0.7756546	0.02432792	3
2-G	0.7521863	0.0564747	3
2-A	0.7156966	0.1453089	3
1-U	0.834829	0.08177497	3
1-C	0.7976527	0.04601417	3
1-G	0.7885616	0.02018688	3

Standard deviation of each off-target row in **Fig. 2F**.

Bar	Mean	SD	N
SWAP	0.1329597	0.02086228	3
8-C	0.1227627	0.03905559	3
8-G	0.1142143	0.01488234	3
8-A	0.1313207	0.03604711	3
7-A	0.113069	0.009095185	3
7-U	0.1052877	0.01537993	3
7-C	0.09542333	0.01387945	3

6-C	0.138577	0.01609951	3
6-G	0.1040313	0.002148154	3
6-A	0.104991	0.01091833	3
5-U	0.09612999	0.01423749	3
5-C	0.119041	0.01377349	3
5-G	0.1193397	0.03951928	3
4-A	0.1299513	0.0594724	3
4-U	0.1172537	0.01049802	3
4-C	0.128089	0.01065953	3
3-U	0.2059217	0.04441654	3
3-C	0.1947207	0.06232889	3
3-G	0.1232273	0.0390928	3
2-C	0.1388563	0.03903133	3
2-G	0.1214687	0.01914786	3
2-A	0.1538327	0.03096628	3
1-U	0.161786	0.08166267	3
1-C	0.1500143	0.03140079	3
1-G	0.204686	0.06953154	3

Standard deviation of each row in **Fig. 2G**.

Bar	Mean	SD	N
U NRE A	1.148447	0.08097368	3
C NRE U	1.133555	0.07963145	3
G NRE C	1.080734	0.03158507	3
A NRE G	0.7369556	0.1013903	3

Ratio of each on-target bar in **Fig. 2F** to the corresponding off-target.

Bar	Ratio
-----	-------

1-G	3.9
1-C	5.3
1-U	5.2
2-A	4.7
2-G	6.2
2-C	5.6
3-G	5.8
3-C	3.9
3-U	3.7
4-C	6.0
4-U	6.2
4-A	5.6
5-G	6.6
5-C	6.3
5-U	7.7
6-A	6.8
6-G	7.7
6-C	5.3
7-C	7.8
7-U	6.3
7-A	6.8
8-A	5.2
8-G	9.1
8-C	8.8
SWAP	5.5

Table S2

RNA target sequences of the Pums used in experiments of **Fig. 2**. Each landing site at the end of the mRuby mRNA contains two 8-base binding sites, one for each of the two Pum proteins needed to reconstitute a split GFP: Pum1 is fused with the N-terminal portion of split GFP (GFP-N); Pum2 is fused with the C-terminal portion of split GFP (GFP-C). The sequence of Pum2 almost always corresponds to the wild-type, but we replaced it with a different sequence in order to test the wild-type sequence in the position of Pum1 (as in mutant 4-U below) and in order to test the reference Pum sequence in the position of Pum1 (a case which we called SWAP and included in our data analyses.)

Fig. 2F

Mutant (Unit – base it binds)	Pum1 binding site on-target RNA sequence	Pum1 binding site off-target RNA sequence	Pum2 binding site RNA sequence
SWAP	AUAGAUGU	UAUCUACA	GCGAGCAC
1-G	GUAGAUGU	CAUCUACA	AUAUAUGU
1-C	CUAGAUGU	GAUCUACA	AUAUAUGU
1-U	UUAGAUGU	AAUCUACA	AUAUAUGU
2-A	AAAGAUGU	UUUCUACA	AUAUAUGU
2-G	AGAGAUGU	UCUCUACA	AUAUAUGU
2-C	ACAGAUGU	UGUCUACA	AUAUAUGU
3-G	AUGGAUGU	UACCUACA	AUAUAUGU
3-C	AUCGAUGU	UAGCUACA	AUAUAUGU
3-U	AUUGAUGU	UAACUACA	AUAUAUGU
4-C	AUACAUGU	UAUGUACA	AUAUAUGU
4-U	AUAUAUGU	UAUAUACA	GCGAGCAC
4-A	AUAAAUGU	UAUUUACA	AUAUAUGU
5-G	AUAGGUGU	UAUCCACA	AUAUAUGU
5-C	AUAGCUGU	UAUCGACA	AUAUAUGU

Pum1 = WT PumHD

5-U	AUAGUUGU	UAUCAACA	AUAUAUGU
6-A	AUAGAAGU	UAUCUUCA	AUAUAUGU
6-G	AUAGAGGU	UAUCUCCA	AUAUAUGU
6-C	AUAGACGU	UAUCUGCA	AUAUAUGU
7-C	AUAGAUCU	UAUCUAGA	AUAUAUGU
7-U	AUAGAUUU	UAUCUAAA	AUAUAUGU
7-A	AUAGAUAU	UAUCUAUA	AUAUAUGU
8-A	AUAGAUGA	UAUCUACU	AUAUAUGU
8-G	AUAGAUGG	UAUCUACC	AUAUAUGU
8-C	AUAGAUGC	UAUCUACG	AUAUAUGU

Fig. 2G

Label	Pum1 binding site RNA sequence	Pum2 binding site RNA sequence
A NRE G	AUAUAUGU	GCGAGCAC
G NRE C	AUAUAUGU	GCGAGCAC
C NRE U	AUAUAUGU	GCGAGCAC
U NRE A	AUAUAUGU	GCGAGCAC

Fig. 2H

Name	Pum Type	Mismatch number	RNA target with mismatches	Pum protein	Fusion
2-C_Pum1_PumHD	PumHD	0	UGUAGACA	ACAGAUGU	N-GFP
2-C_Pum1_PumHD	PumHD	1	UGUAGAGA	ACAGAUGU	N-GFP
2-C_Pum1_PumHD	PumHD	2	UGUAGUGA	ACAGAUGU	N-GFP
2-C_Pum1_PumHD	PumHD	3	UGUAGUGU	ACAGAUGU	N-GFP
8-A_Pum1_PumHD	PumHD	0	AGUAGAU A	AUAGAUGA	N-GFP

8-A_Pum1_PumHD	PumHD	1	AGUAGAUU	AUAGAUGA	N-GFP
8-A_Pum1_PumHD	PumHD	2	AGUAGAAU	AUAGAUGA	N-GFP
8-A_Pum1_PumHD	PumHD	3	AGUAGGAU	AUAGAUGA	N-GFP

Table S3

Statistics for Fig. 4.

No samples were excluded from the statistical analysis. The threshold for significance throughout this figure is $\alpha = 0.05$; $n = 3$ biological replicates throughout this figure. The statistical analysis for Figs. 4A and 4E includes all the data points displayed in those figures, and also one more data point in each data set called SWAP. SWAP represents one additional test case in which we swapped the sequences usually tested with Pum1 and Pum2, in order to make sure that our results weren't the result of positional effects (see Table S4). The values for SWAP were included in all the following global analyses, and these values were also plotted in Fig. S10.

Abbreviations

Diff	Difference
SE	Standard Error
DoF	Degrees of Freedom
Sig?	Significant?

On-target vs. off-target data in Fig. 4A.

Two-way ANOVA with factors 'ON or OFF Target', which compares on-target against off-target data; and 'Target Sequence', which compares among the bars. Average fold-increase in signal between on-target and off-target samples is 7.9, with standard deviation of 2.2.

Source of Variation	% of total variation	P value	Sig?	F (DFn, DFd)
Interaction	2.31	0.0043	Yes	F (24, 100) = 2.162
Target Sequence	2.164	0.0082	Yes	F (24, 100) = 2.026
ON or OFF Target	91.07	< 0.0001	Yes	F (1, 100) = 2046

On-target vs. off-target data in Fig. 4D.

Two-way ANOVA with factors 'ON or OFF Target', which compares On-target against off-target data; and 'Target Sequence', which compares among the bars. Average fold-increase in signal between on-target and off-target samples is 4.2, with standard deviation of 0.91.

Source of Variation	% of total variation	P value	Sig?	F (DFn, DFd)
Interaction	2.717	< 0.0001	Yes	F (12, 52) = 5.075
Target Sequence	2.119	0.0002	Yes	F (12, 52) = 3.959
ON or OFF Target	92.84	< 0.0001	Yes	F (1, 52) = 2081

On-target vs. Off-target data in **Fig. 4E**.

Two-way ANOVA with factors ‘ON or OFF Target’, which compares On-target against Off-target data; and ‘Target Sequence’, which compares among the bars. Average fold-increase in signal between on-target and off-target samples is 2.6, with standard deviation of 0.58.

Source of Variation	% of total variation	P value	Sig?	F (DFn, DFd)
Interaction	3.031	0.0094	Yes	F (18, 76) = 2.197
Target Sequence	3.334	0.0041	Yes	F (18, 76) = 2.417
ON or OFF Target	87.81	< 0.0001	Yes	F (1, 76) = 1146

Fig. 4C

Two-way ANOVA with factors of “Pum Tested” (comparing between two particular proteins into which we introduced mismatch mutations) and “Mismatch number”.

Source of Variation	P value	Summary	Sig?	F (DFn, DFd)	DoF
Interaction	0.1051	ns		F (4, 20) = 2.206	P = 0.1051
Mismatch number	< 0.0001	****	Yes	F (4, 20) = 265.8	P < 0.0001
Pum protein	0.0115	*	Yes	F (1, 20) = 7.739	P = 0.0115

Fig. 4C

Dunnnett’s multiple comparison test across the values of “Mismatch number”, and against the off-target case, after the ANOVA with factors of “Pum Tested” and “Mismatch number”.

Test details	Mean Diff.	SE of diff.	Adjusted P Value	Summary	Significant?
0 vs. 8	0.9016	0.0342	< 0.0001	****	Yes
1 vs. 8	0.5842	0.0342	< 0.0001	****	Yes
2 vs. 8	0.2146	0.0342	< 0.0001	****	Yes
3 vs. 8	0.000168	0.0342	> 0.9999	ns	

Dunnnett’s multiple comparisons test across the values of ‘Target Sequence’ in **Fig. 4D**, against the value for Pumby8 (i.e., 8mer), after the two-way ANOVA with factors ‘ON or OFF Target’ and ‘Target Sequence’.

Dunnnett’s multiple comparisons test	Mean Diff.	95% CI of diff.	Sig?	Summary	Adjusted P Value
6mer vs. 8mer	0.07745	-0.005967 to 0.1609		ns	0.0816
7mer vs. 8mer	0.005622	-0.07780 to 0.08904		ns	0.9997
9mer vs. 8mer	0.03216	-0.05126 to 0.1156		ns	0.9047
10mer vs. 8mer	0.03372	-0.04970 to 0.1171		ns	0.8781
11mer vs. 8mer	-0.003537	-0.08696 to 0.07988		ns	0.9999
12mer vs. 8mer	0.0243	-0.05912 to 0.1077		ns	0.9848
13-mer vs. 8mer	0.0384	-0.04502 to 0.1218		ns	0.7797
14-mer vs. 8mer	0.1182	0.03479 to 0.2016	Yes	**	0.0017
15-mer vs. 8mer	-0.007827	-0.09125 to 0.07559		ns	0.9996
16-mer vs. 8mer	0.04001	-0.04341 to 0.1234		ns	0.7409

17-mer vs. 8mer	0.05058	-0.03285 to 0.1340		ns	0.4758
18-mer vs. 8mer	-0.0434	-0.1268 to 0.04002		ns	0.6557

Dunnett's multiple comparisons test across the values of 'Target Sequence' in **Fig. 4E**, against the value for the truncated wild-type variant (i.e., 4-U), after the two-way ANOVA with factors 'ON or OFF Target' and 'Target Sequence'.

Dunnett's multiple comparisons test	Mean Diff.	95% CI of diff.	Sig?	Summary	Adjusted P Value
SWAP vs. 4-U	-0.05117	-0.1616 to 0.05925		ns	0.8395
6-C vs. 4-U	-0.02003	-0.1305 to 0.09039		ns	0.9993
6-G vs. 4-U	0.04137	-0.06905 to 0.1518		ns	0.9596
6-A vs. 4-U	0.02523	-0.08519 to 0.1357		ns	0.999
5-U vs. 4-U	-0.04687	-0.1573 to 0.06356		ns	0.9036
5-C vs. 4-U	-0.01808	-0.1285 to 0.09234		ns	0.9993
5-G vs. 4-U	-0.06209	-0.1725 to 0.04833		ns	0.6279
4-A vs. 4-U	-0.02163	-0.1321 to 0.08879		ns	0.9991
4-C vs. 4-U	-0.01376	-0.1242 to 0.09667		ns	0.9995
3-U vs. 4-U	-0.05312	-0.1635 to 0.05730		ns	0.8056
3-C vs. 4-U	0.05937	-0.05106 to 0.1698		ns	0.684
3-G vs. 4-U	-0.0386	-0.1490 to 0.07182		ns	0.9774
2-C vs. 4-U	0.07052	-0.03991 to 0.1809		ns	0.4597
2-G vs. 4-U	-0.01926	-0.1297 to 0.09116		ns	0.9993
2-A vs. 4-U	-0.008006	-0.1184 to 0.1024		ns	0.9997
1-U vs. 4-U	-0.03497	-0.1454 to 0.07545		ns	0.9883
1-C vs. 4-U	0.01194	-0.09848 to 0.1224		ns	0.9996
1-G vs. 4-U	-0.08388	-0.1943 to 0.02655		ns	0.2481

Each bar in **Fig. 4B** vs. the combined on-target replicates in **Fig. 4A** (excluding SWAP).

Unpaired two-sided t-tests, one for each bar of **Fig. 4B** (n = 3 biological replicates) against the combined replicates of **Fig. 4A** (on-target, excluding SWAP). We use the Bonferroni method to correct for multiple comparisons and do not assume consistent standard deviation between bars.

Bar	Sig?	P value	Mean 1	Mean 2	SE of difference	t ratio	DoF
U NRE A		0.0169841	0.756696	0.925343	0.0690323	2.44302	73
C NRE U	Yes	5.2077E-05	0.631211	0.925343	0.0684026	4.30002	73
G NRE C		0.0249934	0.768555	0.925343	0.0685057	2.28868	73
A NRE G		0.166916	1.02206	0.925343	0.0692799	1.3961	73

All raw datapoints are plotted in the figure so that the entire dataset is available to the reader, since each individual condition has less than 5 datapoints, and thus formal evaluations of normality are not appropriate.

Standard deviation for **Fig. 4A** (8-mer)

Bar	On Target			Off Target		
	Mean	SD	N	Mean	SD	N
SWAP	1.022065	0.1460573	3	0.1099796	0.005015381	3
8-C	0.9120812	0.04729243	3	0.06979913	0.02135131	3
8-G	0.8185704	0.1476841	3	0.1088798	0.02035209	3
8-A	0.8687229	0.1148349	3	0.1194246	0.06077832	3
7-A	0.866118	0.02463895	3	0.1012546	0.03986008	3
7-U	0.8997113	0.009695103	3	0.1373508	0.0807678	3
7-C	0.8557666	0.03262806	3	0.07865977	0.02199873	3
6-C	0.8961539	0.05833427	3	0.148298	0.04288156	3
6-G	0.90022	0.07743163	3	0.1570926	0.09693857	3
6-A	0.8611347	0.02775115	3	0.1005414	0.009915407	3
5-U	0.8539314	0.01321622	3	0.09478492	0.01215006	3
5-C	0.8817373	0.03742461	3	0.09711954	0.01211225	3
5-G	0.8755319	0.03334989	3	0.1146943	0.0106131	3
4-A	0.8608447	0.002063088	3	0.1146783	0.002956916	3
4-U	0.8797872	0.05415945	3	0.1205901	0.01516778	3
4-C	0.7986001	0.04934525	3	0.1828711	0.07905062	3
3-U	0.9233379	0.02562596	3	0.099484	0.007701178	3
3-C	0.9060388	0.05820748	3	0.1577154	0.07083459	3
3-G	0.8833146	0.04568162	3	0.4577023	0.639156	3
2-C	0.8851156	0.0417971	3	0.09441459	0.009389973	3
2-G	1.157364	0.1069792	3	0.1097169	0.006463079	3
2-A	1.097334	0.06262239	3	0.1179531	0.02179759	3
1-U	1.117275	0.03061979	3	0.1213214	0.04721319	3
1-C	1.066407	0.1378023	3	0.1619717	0.06218273	3
1-G	1.14314	0.06345791	3	0.1474879	0.05961747	3

Standard deviation for **Fig. 4D** (n-mer)

Bar	On Target			Off Target		
	Mean	SD	N	Mean	SD	N
6mer	0.7560282	0.1359919	3	0.1948956	0.03511899	3
7mer	0.6321488	0.04692709	3	0.1751124	0.02012509	3
8mer	0.6187699	0.05195605	3	0.1772477	0.008615075	3
9mer	0.6804538	0.02238468	3	0.1798818	0.004294916	3
10mer	0.695353	0.01607241	3	0.1680999	0.02478947	3
11mer	0.6137316	0.06969854	3	0.1752113	0.01051353	3
12mer	0.6910031	0.04404068	3	0.1536241	0.03558193	3
13-mer	0.7036877	0.03102987	3	0.1691392	0.006199401	3
14-mer	0.8816705	0.1055437	3	0.1507607	0.01278519	3
15-mer	0.6725989	0.02139873	3	0.1077651	0.004304052	3
16-mer	0.7214001	0.09481295	3	0.1546466	0.0471785	3
17-mer	0.7300651	0.06687612	3	0.1671027	0.04583143	3
18-mer	0.5237896	0.01679627	3	0.1854292	0.01940445	3

Standard deviation for **Fig. 4E** (6-mer)

Bar	On Target			Off Target		
	Mean	SD	N	Mean	SD	N
SWAP	0.6706445	0.04527973	3	0.21798	0.1092039	3
6-C	0.6482364	0.02666046	3	0.3026592	0.02838244	3
6-G	0.7857187	0.09828691	3	0.2879847	0.07042043	3
6-A	0.6844471	0.04801351	3	0.356985	0.05486156	3
5-U	0.6351328	0.0421611	3	0.2621006	0.02571849	3
5-C	0.6438665	0.008325058	3	0.3109422	0.08461593	3
5-G	0.5984381	0.04323539	3	0.2683385	0.06455794	3
4-A	0.6723599	0.04897044	3	0.2753476	0.01745891	3

4-U	0.6821637	0.04208474	3	0.3088014	0.02377533	3
4-C	0.6748123	0.0268278	3	0.2886397	0.05315387	3
3-U	0.6773398	0.04725474	3	0.2073803	0.02189145	3
3-C	0.8433485	0.1732628	3	0.2663467	0.07541873	3
3-G	0.6435395	0.03345279	3	0.2702217	0.03175108	3
2-C	0.8006291	0.1203763	3	0.3313685	0.08130261	3
2-G	0.6493838	0.01130306	3	0.3030574	0.0629117	3
2-A	0.6661043	0.02757233	3	0.3088488	0.02160324	3
1-U	0.6321438	0.0257313	3	0.2888803	0.03777402	3
1-C	0.7759755	0.1478537	3	0.2388762	0.05531069	3
1-G	0.6679561	0.07293265	3	0.1552551	0.01866407	3

Standard deviation for **Fig. 4B** (bases flanking 8-mer)

Bar	Mean	SD	N
U NRE A	0.7566957	0.1331674	3
C NRE U	0.6312112	0.09292551	3
G NRE C	0.7685555	0.1005967	3
A NRE G	1.022065	0.1460573	3

Ratio of on-target vs. off-target values for **Fig. 4A** (8-mer Pumby)

Bar	Ratio
SWAP	9.3
8-C	13.1
8-G	7.5
8-A	7.3
7-A	8.6
7-U	6.6
7-C	10.9

6-C	6.0
6-G	5.7
6-A	8.6
5-U	9.0
5-C	9.1
5-G	7.6
4-A	7.5
4-U	7.3
4-C	4.4
3-U	9.3
3-C	5.7
3-G	1.9
2-C	9.4
2-G	10.5
2-A	9.3
1-U	9.2
1-C	6.6
1-G	7.8

Ratio of on-target vs. off-target values for **Fig. 4D** (n-mer Pumby)

Bar	Ratio
6mer	3.9
7mer	3.6
8mer	3.5
9mer	3.8
10mer	4.1
11mer	3.5
12mer	4.5
13-mer	4.2
14-mer	5.8
15-mer	6.2
16-mer	4.7
17-mer	4.4
18-mer	2.8

Ratio of on-target vs. off-target values for **Fig. 4E** (6-mer Pumby)

Bar	Ratio
SWAP	3.1
6-C	2.1
6-G	2.7
6-A	1.9
5-U	2.4
5-C	2.1
5-G	2.2
4-A	2.4
4-U	2.2
4-C	2.3
3-U	3.3
3-C	3.2
3-G	2.4
2-C	2.4
2-G	2.1
2-A	2.2
1-U	2.2
1-C	3.2
1-G	4.3

Table S4

RNA target sequences of the Pums used in experiments of **Fig. 4**. Each landing site at the end of the mRuby mRNA contains two binding sites (of various lengths), one for each of the two Pum proteins needed to reconstitute a split GFP: Pum1 is fused with the N-terminal portion of split GFP (GFP-N); Pum2 is fused with the C-terminal portion of split GFP (GFP-C).

Fig. 4A

Mutant (Unit – base it binds)	Pum1 binding site on-target RNA sequence	Pum2 binding site off-target RNA sequence	Pum2 binding site RNA sequence
SWAP	AUAGAUGU	UAUCUACA	GCGAGCAC
1-G	GUAGAUGU	CAUCUACA	AUAUAUGU
1-C	CUAGAUGU	GAUCUACA	AUAUAUGU
1-U	UUAGAUGU	AAUCUACA	AUAUAUGU
2-A	AAAGAUGU	UUUCUACA	AUAUAUGU
2-G	AGAGAUGU	UCUCUACA	AUAUAUGU
2-C	ACAGAUGU	UGUCUACA	AUAUAUGU
3-G	AUGGAUGU	UACCUACA	AUAUAUGU
3-C	AUCGAUGU	UAGCUACA	AUAUAUGU
3-U	AUUGAUGU	UAACUACA	AUAUAUGU
4-C	AUACAUGU	UAUGUACA	AUAUAUGU
4-U	AUAUAUGU	UAUAUACA	GCGAGCAC
4-A	AUAAAUGU	UAUUUACA	AUAUAUGU
5-G	AUAGGUGU	UAUCCACA	AUAUAUGU
5-C	AUAGCUGU	UAUCGACA	AUAUAUGU
5-U	AUAGUUGU	UAUCAACA	AUAUAUGU
6-A	AUAGAAGU	UAUCUUCA	AUAUAUGU

Wild-type target sequence

6-G	AUAGAGGU	UAUCUCCA	AUAUAUGU
6-C	AUAGACGU	UAUCUGCA	AUAUAUGU
7-C	AUAGAUCU	UAUCUAGA	AUAUAUGU
7-U	AUAGAUUU	UAUCUAAA	AUAUAUGU
7-A	AUAGAUAU	UAUCUAUA	AUAUAUGU
8-A	AUAGAUGA	UAUCUACU	AUAUAUGU
8-G	AUAGAUGG	UAUCUACC	AUAUAUGU
8-C	AUAGAUGC	UAUCUACG	AUAUAUGU

Mutant (Unit – base it binds)	Pum1 binding site on-target RNA sequence	Pum2 binding site RNA sequence
A NRE G	AUAUAUGU	GCGAGCAC
G NRE C	AUAUAUGU	GCGAGCAC
C NRE U	AUAUAUGU	GCGAGCAC
U NRE A	AUAUAUGU	GCGAGCAC

Fig. 4B

Mutant (Unit – base it binds)	Pum1 binding site RNA sequence	Pum2 binding site RNA sequence
A NRE G	AUAUAUGU	GCGAGCAC
G NRE C	AUAUAUGU	GCGAGCAC
C NRE U	AUAUAUGU	GCGAGCAC
U NRE A	AUAUAUGU	GCGAGCAC

Fig. 4C

Name	Pum Type	Mismatch number	RNA target with mismatches	Pum protein	Fusion
-------------	-----------------	----------------------------	---------------------------------------	--------------------	---------------

2-C_Pum1_Pumby8	Pumby	0	UGUAGACA	ACAGAUGU	N-GFP
2-C_Pum1_Pumby8	Pumby	1	UGUAGAGA	ACAGAUGU	N-GFP
2-C_Pum1_Pumby8	Pumby	2	UGUAGUGA	ACAGAUGU	N-GFP
2-C_Pum1_Pumby8	Pumby	3	UGUAGUGU	ACAGAUGU	N-GFP
8-A_Pum1_Pumby8	Pumby	0	AGUAGAU	AUAGAUGA	N-GFP
8-A_Pum1_Pumby8	Pumby	1	AGUAGAUU	AUAGAUGA	N-GFP
8-A_Pum1_Pumby8	Pumby	2	AGUAGAAU	AUAGAUGA	N-GFP
8-A_Pum1_Pumby8	Pumby	3	AGUAGGAU	AUAGAUGA	N-GFP

Fig. 4D

Label	Pum1 binding site on-target RNA sequence	Pum1 binding site off-target RNA sequence	Pum2 binding site RNA sequence
6mer	AUAU	UAUA	AUAU
7mer	AUAUAUG	UAUAUAC	AUAUAUG
8mer	AUAUAUGU	UAUAUACA	AUAUAUGU
9mer	AUAUAUGUA	UAUAUACAU	AUAUAUGU
10mer	AUAUAUGUAA	UAUAUACAUI	AUAUAUGU
11mer	AUAUAUGUAAG	UAUAUACAUIU	AUAUAUGU
12mer	AUAUAUGUAAGG	UAUAUACAUIUC	AUAUAUGU
13-mer	AUAUAUGUAAGGC	UAUAUACAUIUCG	AUAUAUGU
14-mer	AUAUAUGUAAGGCG	UAUAUACAUIUCGC	AUAUAUGU
15-mer	AUAUAUGUAAGGCGG	UAUAUACAUIUCGCC	AUAUAUGU
16-mer	AUAUAUGUAAGGCGGC	UAUAUACAUIUCGCCG	AUAUAUGU
17-mer	AUAUAUGUAAGGCGGCU	UAUAUACAUIUCGCCGA	AUAUAUGU
18-mer	AUAUAUGUAAGGCGGCUU	UAUAUACAUIUCGCCGAA	AUAUAUGU

Fig. 4E

Label	Pum1 binding site on-target RNA sequence	Pum1 binding site off-target RNA sequence	Pum2 binding site RNA sequence
SWAP	AUAGAU	UAUCUA	GCGAGCAC
1-G	GUAGAU	CAUCUA	AUAUAUGU
1-C	CUAGAU	GAUCUA	AUAUAUGU
1-U	UUAGAU	AAUCUA	AUAUAUGU
2-A	AAAGAU	UUUCUA	AUAUAUGU
2-G	AGAGAU	UCUCUA	AUAUAUGU
2-C	ACAGAU	UGUCUA	AUAUAUGU
3-G	AUGGAU	UACCUA	AUAUAUGU
3-C	AUCGAU	UAGCUA	AUAUAUGU
3-U	AUUGAU	UAACUA	AUAUAUGU
4-C	AUACAU	UAUGUA	AUAUAUGU
4-U	AUAU AU	UAUAUA	AUAUAUGU
4-A	AUAAAU	UAUUUA	AUAUAUGU
5-G	AUAGGU	UAUCCA	AUAUAUGU
5-C	AUAGCU	UAUCGA	AUAUAUGU
5-U	AUAGUU	UAUCAA	AUAUAUGU
6-A	AUAGAA	UAUCUU	AUAUAUGU
6-G	AUAGAG	UAUCUC	AUAUAUGU
6-C	AUAGAC	UAUCUG	AUAUAUGU

Truncated wild-type sequence

Table S5Statistics for **Fig. 5**.

No samples were excluded from the statistical analysis. The threshold for significance throughout this figure is $\alpha = 0.05$.

Abbreviations

SD	Standard Deviation
SE	Standard Error
DoF	Degrees of Freedom
Sig?	Significant?
Diff.	Difference

Number of Biological Replicates

Figure	Variable	Replicates
5C	All	4
5F	All	4
5D	No target	3
5D	All others	4
5E	All	4
5G	No target	3
5G	All others	4
5H	All	4
5I	All	3
5J	All	3

Two-way ANOVAs for **Figs. 5C, 5D, 5E, 5F, 5G, and 5H**, factors of ‘Pumby8 or PumHD’ and either ‘GFP-BLA or BLA-GFP’ or ‘GFP-BLA or BLA-GFP or No Target’, depending on the panel.

Group	Source of effects	Number of Parameters	DoF	Sum of Squares	F Ratio	Prob > F	Sig?
5C	GFP-BLA or BLA-GFP	1	1	289.89566	714.6966	<.0001	Yes
5D	GFP-BLA or BLA-GFP or No Target	2	2	0.00649816	234.077	<.0001	Yes
5E	GFP-BLA or BLA-GFP	1	1	1.119456	1.3466	0.2589	
5F	GFP-BLA or BLA-GFP	1	1	182.18818	1369.5557	<.0001	Yes
5G	GFP-BLA or BLA-GFP or No Target	2	2	0.00687906	182.7482	<.0001	Yes
5H	GFP-BLA or BLA-GFP	1	1	0.2320667	0.3448	0.5634	
5C	Pumby8 or PumHD	1	1	0.08376	0.2065	0.6517	
5D	Pumby8 or PumHD	1	1	0.00000572	0.4119	0.5261	
5E	Pumby8 or PumHD	1	1	0.2062815	0.2481	0.6236	
5F	Pumby8 or PumHD	1	1	0.01734	0.1303	0.7198	

5G	Pumby8 or PumHD	1	1	0.0000597	3.1722	0.0854	
5H	Pumby8 or PumHD	1	1	1.8841687	2.7991	0.1092	

Tukey's tests for **Figs. 5D** and **5G**, after two-way ANOVA with factors of ‘Pumby8 or PumHD’ and ‘GFP-BLA or BLA-GFP or No Target’

Panel	Factor 1	Factor 2	Difference	SE	t Ratio	Prob > t	Sig?
5D	Pumby8	PumHD	0.0008829	0.0013758	0.64	0.5261	
5D	BLA-GFP	GFP-BLA	-0.026781	0.001521	-17.61	<.0001	Yes
5D	BLA-GFP	No Target	0.004921	0.0016429	3	0.0149	Yes
5D	GFP-BLA	No Target	0.031702	0.0016429	19.3	<.0001	Yes
5G	Pumby8	PumHD	-0.002853	0.001602	-1.78	0.0854	
5G	BLA-GFP	GFP-BLA	0.0287012	0.0017711	16.21	<.0001	Yes
5G	BLA-GFP	No Target	0.031551	0.001913	16.49	<.0001	Yes
5G	GFP-BLA	No Target	0.0028497	0.001913	1.49	0.3105	

Fig. 5J

Two-way ANOVA with factors of “Pum Tested” (comparing between two particular proteins into which we introduced mismatch mutations) and “Mismatch number”.

Figure	Source of Variation	P value	Summary	Sig?	F (DFn, DFd)	DoF
5J	Interaction	0.0196	*	Yes	F (4, 20) = 3.751	P = 0.0196
5J	Mismatch number	< 0.0001	****	Yes	F (4, 20) = 161.7	P < 0.0001
5J	Pum tested	0.0018	**	Yes	F (1, 20) = 12.96	P = 0.0018

Dunnnett’s multiple comparison test across the values of “Mismatch number”, and against the no-target case, after the ANOVA with factors of “Pum Tested” and “Mismatch number”.

Figure	Test details	Mean Diff.	SE of diff.	Adjusted P Value	Summary	Sig?
5J	0 vs. No target	0.03964	0.001977	< 0.0001	****	Yes
5J	1 vs. No target	0.02081	0.001977	< 0.0001	****	Yes
5J	2 vs. No target	0.0004234	0.001977	0.9983	ns	
5J	3 vs. No target	-0.00004058	0.001977	> 0.9999	ns	

Fig. 5I

Two-way ANOVA with factors of “GFP-BLA or BLA-GFP” (which of the two target transcripts was used) and “protein measured” (whether the immunoepitope we were measuring was attached to GFP or β -lactamase).

Source of Variation	P value	P value summary	Sig?	F (DFn, DFd)	DoF
Interaction	< 0.0001	****	Yes	F (1, 8) = 291.7	1

GFP-BLA or BLA-GFP	0.6322	ns		F (1, 8) = 0.2476	1
Protein measured	0.0706	ns		F (1, 8) = 4.345	1
Residual					8

Tukey's multiple comparison test after the previous ANOVA.

Test details	Mean Diff.	SE of diff.	Summary	Sig?	DoF
GFP prot:BLA-GFP vs. GFP prot:GFP-BLA	-95.27	7.666	Yes	****	8
BLA prot:GFP-BLA vs. GFP prot:GFP-BLA	-103.9	7.666	Yes	****	8
BLA prot:BLA-GFP vs. GFP prot:GFP-BLA	-14	7.666		ns	8
BLA prot:GFP-BLA vs. GFP prot:BLA-GFP	-8.602	7.666		ns	8
BLA prot:BLA-GFP vs. GFP prot:BLA-GFP	81.27	7.666	Yes	****	8
BLA prot:BLA-GFP vs. BLA prot:GFP-BLA	89.88	7.666	Yes	****	8

Multiple t-tests on **Fig. 5I** using the Holm-Sidak method with no assumption of consistent SD.

	Sig?	P value	Mean for GFP-BLA	Mean for BLA-GFP	Diff	SE of Diff	t ratio	DoF
GFP protein	*	0.000239 531	138.423	43.1535	95.2698	7.65372	12.4475	4
BLA protein	*	0.000304 581	34.5515	124.427	-89.8758	7.67787	11.7058	4

All raw datapoints are plotted in the figure so that the entire dataset is available to the reader, since each individual condition has fewer than 5 datapoints, and thus formal evaluations of normality are not appropriate.

Standard deviation for **Fig. 5C**

Site	GFP-BLA			BLA-GFP		
	Mean	SD	N	Mean	SD	N
Pumby8_TM_GFP_4A	6.414861	0.8985428	4	1.468694	0.2909553	4
PumHD_TM_GFP_1A	6.636467	0.8914505	4	1.386541	0.2600522	4
Pumby8_TM_GFP_3A	6.541493	0.9614081	4	1.763351	0.3417693	4
Pumby8_TM_BLA_6A	6.248656	0.6633087	4	1.601341	0.1936187	4

Pumby8_TM_BLA_5A	6.042879	0.933061	4	1.446533	0.3549457	4
PumHD_TM_BLA_2A	6.683953	1.093118	4	1.411393	0.3269526	4

Standard deviation for **Fig. 5F**

Site	GFP-BLA			BLA-GFP		
	Mean	SD	N	Mean	SD	N
Pumby8_TM_GFP_4A	0.8361412	0.2452421	4	4.768704	0.5417642	4
PumHD_TM_GFP_1A	0.7317386	0.2285844	4	4.701451	0.2578983	4
Pumby8_TM_GFP_3A	1.034556	0.1371141	4	4.561729	0.5299556	4
Pumby8_TM_BLA_6A	0.6951778	0.1560119	4	4.682626	0.401232	4
Pumby8_TM_BLA_5A	0.6442034	0.2536526	4	4.454893	0.585783	4
PumHD_TM_BLA_2A	0.7079814	0.18155	4	4.859115	0.5737917	4

Standard deviation for **Fig. 5D**

Site	GFP-BLA			BLA-GFP			No Target		
	Mean	SD	N	Mean	SD	N	Mean	SD	N
Pumby8_T M_GFP_4A	0.034 03527	0.002017148	4	0.005589337	0.00147515	4	0.001932808	0.000547496	3
PumHD_T M_GFP_1A	0.032 07211	0.006654412	4	0.006769207	0.002454604	4	0.004535947	0.000736295	3
Pumby8_T M_GFP_3A	0.036 89607	0.005445201	4	0.01030112	0.003455467	4	0.001428475	0.000377058	3

Standard deviation for **Fig. 5E**

Site	GFP-BLA			BLA-GFP		
	Mean	SD	N	Mean	SD	N
Pumby8_TM_GFP_4A	17.86833	0.4249315	4	17.22833	0.9934322	4
PumHD_TM_GFP_1A	17.7725	0.6059726	4	17.26167	1.177897	4
Pumby8_TM_GFP_3A	17.165	1.208373	4	17.02	1.035292	4

Standard deviation for **Fig. 5G**

Site	GFP-BLA			BLA-GFP			No Target		
	Mean	SD	N	Mean	SD	N	Mean	SD	N

Site	Mean	SD	N	Mean	SD	N	Mean	SD	N
Pumby8_TM_BLA_6A	0.005399815	0.000990153	4	0.03167689	0.005847114	4	0.001999483	0.000750605	3
Pumby8_TM_BLA_5A	0.005374884	0.001800345	4	0.03301405	0.004251305	4	0.003026238	0.001434976	3
PumHD_TM_BLA_2A	0.006318842	0.001627106	4	0.03850631	0.01008892	4	0.003518587	0.001844866	3

Standard deviation for **Fig. 5H**

Site	GFP-BLA			BLA-GFP		
	Mean	SD	N	Mean	SD	N
Pumby8_TM_BLA_6A	17.9	0.9496784	4	17.84667	1.067004	4
Pumby8_TM_BLA_5A	18.1025	0.6163591	4	17.98833	1.213143	4
PumHD_TM_BLA_2A	18.175	0.5881581	4	18.9325	0.3446888	4

Table S6

List of the Pum target sequences for the experiments of **Fig. 5**.

Each mRNA target site contains two 8-base binding sites, one for each of the two Pum proteins needed to reconstitute a split reporter protein: Pum1 (fused to N-terminal portion of split luciferase, N-Luc) binds to the target site whose name ends in “A”; Pum2 (fused to C-terminal portion of split luciferase, C-Luc) binds to the binding site whose name ends in “B”.

Proteins pairs (Pum1 and Pum2; see next table) corresponding to the data points in **Figs. 5C – 5H**.

Panel	Position of each data point within a given condition, left-to-right	Protein pairs (Pum1 and Pum2)
5C	1	PumHD_TM_GFP_1
5C	2	PumHD_TM_BLA_2
5C	3	Pumby8_TM_GFP_3
5C	4	Pumby8_TM_GFP_4
5C	5	Pumby8_TM_BLA_5
5C	6	Pumby8_TM_BLA_6
5D	1	PumHD_TM_GFP_1
5D	2	Pumby8_TM_GFP_3
5D	3	Pumby8_TM_GFP_4
5E	1	PumHD_TM_GFP_1
5E	2	Pumby8_TM_GFP_3
5E	3	Pumby8_TM_GFP_4
5F	1	PumHD_TM_GFP_1
5F	2	PumHD_TM_BLA_2
5F	3	Pumby8_TM_GFP_3
5F	4	Pumby8_TM_GFP_4
5F	5	Pumby8_TM_BLA_5
5F	6	Pumby8_TM_BLA_6
5G	1	PumHD_TM_BLA_2
5G	2	Pumby8_TM_BLA_5
5G	3	Pumby8_TM_BLA_6
5H	1	PumHD_TM_BLA_2

5H	2	Pumby8_TM_BLA_5
5H	3	Pumby8_TM_BLA_6

Protein and RNA target sequences for all the proteins in **Figs. 5C – 5H**. The “_A” and “_B” suffixes refer to Pum1 and Pum2, respectively.

Name	Pum Type	RNA sequence Pum binds to (5' to 3')	Protein sequence	Fusion
PumHD_TM_GFP_1A	PumHD	GAAGGCUA	AUCGGAAG	N-Luc
PumHD_TM_GFP_1B	PumHD	AGGAGCGC	CGCGAGGA	C-Luc
PumHD_TM_BLA_2A	PumHD	GACAACAG	GACAACAG	N-Luc
PumHD_TM_BLA_2B	PumHD	CGAUUGGA	AGGUUAGC	C-Luc
Pumby8_TM_GFP_3A	Pumby	GCCCGACA	ACAGCCCG	N-Luc
Pumby8_TM_GFP_3B	Pumby	UACCUGAG	GAGUCCAU	C-Luc
Pumby8_TM_GFP_4A	Pumby	ACGGCCAC	CACCGGCA	N-Luc
Pumby8_TM_GFP_4B	Pumby	CAGCGUGU	UGUGCGAC	C-Luc
Pumby8_TM_BLA_5A	Pumby	GAGCGACA	ACAGCGAG	N-Luc
Pumby8_TM_BLA_5B	Pumby	GCGGCUAA	AAUCGGCG	C-Luc
Pumby8_TM_BLA_6A	Pumby	CUGCUGUG	GUGUCGUC	N-Luc
Pumby8_TM_BLA_6B	Pumby	CAGUGUUG	GUUGUGAC	C-Luc

Protein and RNA target sequences for all the proteins in **Fig. 5J**

Name	Pum Type	Mismatch number	RNA target with mismatches	Pum protein	Fusion
Pumby8_TM_GFP_4A	Pumby	0	ACGGCCAC	CACCGGCA	N-Luc
Pumby8_TM_GFP_4A-M1	Pumby	1	ACGGCCAG	GACCGGCA	N-Luc
Pumby8_TM_GFP_4A-M2	Pumby	2	ACGGCUAG	GAUCGGCA	N-Luc
Pumby8_TM_GFP_4A-M3	Pumby	3	ACGGCUGG	GGUCGGCA	N-Luc
PumHD_TM_BLA_2A	PumHD	0	GACAACAG	GACAACAG	N-Luc
PumHD_TM_BLA_2A-M1	PumHD	1	GACAACAA	AACAACAG	N-Luc

PumHD_TM_BLA_2A-M2	PumHD	2	GACAACUA	AUCAACAG	N-Luc
PumHD_TM_BLA_2A-M3	PumHD	3	GACAAGUA	AUGAACAG	N-Luc

Table S7

Statistics for **Fig. 6**.

No samples were excluded from the statistical analysis. The threshold for significance throughout this figure is $\alpha = 0.05$. A total of three Pum proteins were tested throughout **Fig. 6**, with various combinations of target RNA. Each cluster of 3 bars in **Figs. 6C – 6J** contains the results of a single protein; the three bars represent different numbers of tandem RNA target repeats (distinguished by color). The three proteins are PumHD_TI_1, Pumby8_TI_2, and Pumby8_TI_3, and their corresponding clusters always appear in that order (see **Table S8** for the full sequences of these proteins and their RNA targets).

In order to create the relative values of Renilla activity in **Figs. 6C – 6F**, the values under each test condition (on-target, off-target, etc.) were normalized to the expression (mean of the 3 biological replicates) for that same Pum protein and number of tandem repeats in the No Driver condition. Thus, for example, the raw expression value of the 3 biological replicates for Pumby8_TI_2 (second cluster) with a 5x tandem target (yellow bar) in the condition of “Pum-eIF4E off-target” (**Fig. 6E**) were divided by the mean of the three biological replicates for that same protein and that same repeat number under the No Driver condition (**Fig. 6C**). The same normalization was applied to the biological replicates in the No Driver condition itself (**Fig. 6C**), which is why their mean value is 1 but their standard deviation is greater than zero.

All raw values for Firefly luciferase luminescence (**Figs. 6G - 6J**) were divided by 10,000 prior to analysis and graphing, given that the units are arbitrary.

Abbreviations

SD	Standard Deviation
SE	Standard Error
DoF	Degrees of Freedom
Sig?	Significant?
Diff.	Difference
CI	Confidence Interval
N Parm	Number of paramters

Biological Replicates

Panel	Replicates
All panels	3

Three-way ANOVAs for **Fig. 6** with factors of ‘Copy Number’, ‘Driver Plasmid’, and ‘Pum Type’.

Group	Source of	Number of	DoF	Sum of	F Ratio	Prob > F	Sig?
--------------	------------------	------------------	------------	---------------	----------------	--------------------	-------------

	effects	Parameters		Squares			
CDEF	Copy Number	2	2	143.2743	8.0124	0.0006	Yes
CDEF	Driver Plasmid	3	3	1862.353	69.4325	<.0001	Yes
CDEF	Pum Type	1	1	0.3179	0.0356	0.8508	
LMNO	Copy Number	2	2	1.28378574	0.2458	0.7826	
LMNO	Driver Plasmid	3	3	23.257516	2.9681	0.0355	Yes
LMNO	Pum Type	1	1	1.388363	0.5315	0.4676	

Tukey's honest significance post-hoc tests on the above ANOVAs.

Dataset	Factor 1	Factor 2	Difference	Std Error	t Ratio	Prob > t	Significant
CDEF	just eIF4E	none	0.04072	0.8154879	0.05	1	
CDEF	just eIF4E	Pum-eIF4E OFF	-0.20663	0.8218949	-0.25	0.9944	
CDEF	just eIF4E	Pum-eIF4E ON	-9.64233	0.8218949	-11.73	<.0001	Yes
CDEF	none	Pum-eIF4E OFF	-0.24735	0.8067719	-0.31	0.9899	
CDEF	none	Pum-eIF4E ON	-9.68306	0.8067719	-12	<.0001	Yes
CDEF	Pum-eIF4E OFF	Pum-eIF4E ON	-9.43571	0.8138082	-11.59	<.0001	Yes
CDEF	10x	1x	2.70356	0.7051425	3.83	0.0006	Yes
CDEF	10x	5x	0.64778	0.7047786	0.92	0.6295	
CDEF	1x	5x	-2.05578	0.7051425	-2.92	0.0121	Yes
CDEF	Pumby	PumHD	-0.115168	0.6107764	-0.19	0.8508	
GHIJ	just eIF4E	none	0.458997	0.4407706	1.04	0.7256	
GHIJ	just eIF4E	Pum-eIF4E OFF	0.982064	0.4442336	2.21	0.1273	
GHIJ	just eIF4E	Pum-eIF4E ON	1.204718	0.4442336	2.71	0.0387	Yes
GHIJ	none	Pum-eIF4E OFF	0.523067	0.4360596	1.2	0.6286	
GHIJ	none	Pum-eIF4E ON	0.745721	0.4360596	1.71	0.3238	
GHIJ	Pum-eIF4E OFF	Pum-eIF4E ON	0.222654	0.4398627	0.51	0.9574	
GHIJ	10x	1x	-0.206839	0.381129	-0.54	0.8504	
GHIJ	10x	5x	-0.249753	0.3809323	-0.66	0.7896	
GHIJ	1x	5x	-0.042913	0.381129	-0.11	0.993	
GHIJ	Pumby	PumHD	-0.240683	0.3301242	-0.73	0.4676	

Fig. 6K

Two-way ANOVA with factors of “Pum Tested” (comparing between two particular proteins into which we introduced mismatch mutations) and “Mismatch number”.

Figure	Source of Variation	P value	Summary	Sig?	F (DFn, DFd)	DoF
6K	Interaction	0.0729	ns	No	F (4, 20) = 2.526	P = 0.0729

6K	Mismatch number	< 0.0001	****	Yes	F (4, 20) = 160.9	P < 0.0001
6K	Pum tested	0.0161	*	Yes	F (1, 20) = 6.907	P = 0.0161

Dunnett's multiple comparison test across the values of "Mismatch number", and against the off-target case, after the ANOVA with factors of "Pum Tested" and "Mismatch number".

Figure	Test details	Mean Diff.	SE of diff.	Adjusted P Value	Summary	Sig?
6K	0 vs. 8	17.81	0.9819	< 0.0001	****	Yes
6K	1 vs. 8	16.25	0.9819	< 0.0001	****	Yes
6K	2 vs. 8	3.423	0.9819	0.0082	**	Yes
6K	3 vs. 8	0.1174	0.9819	0.9998	ns	No

All raw data points are plotted in the figure so that the entire dataset is available to the reader, since each individual condition has less than 5 data points, and thus formal evaluations of normality are not appropriate.

Standard deviation for **Fig. 6C**

Panel	Pum	Tandem repeats	Mean	SD	N
6C	PumHD_TI_1	1x	1	0.05679457	3
6C	PumHD_TI_1	5x	1	0.03343772	3
6C	PumHD_TI_1	10x	1	0.09560168	3
6C	Pumby8_TI_2	1x	1	0.1253747	3
6C	Pumby8_TI_2	5x	1	0.04206137	3
6C	Pumby8_TI_2	10x	1	0.132707	3
6C	Pumby8_TI_3	1x	1	0.06171444	3
6C	Pumby8_TI_3	5x	1	0.0724223	3
6C	Pumby8_TI_3	10x	1	0.06606679	3

Standard deviation for **Fig. 6D**

Panel	Pum	Tandem repeats	Mean	SD	N
6D	PumHD_TI_1	1x	6.432531	0.2441904	3
6D	PumHD_TI_1	5x	10.58404	1.540903	3
6D	PumHD_TI_1	10x	20.15385	1.358123	3
6D	Pumby8_TI_2	1x	3.018663	0.4372596	3
6D	Pumby8_TI_2	5x	13.21637	0.7242065	3
6D	Pumby8_TI_2	10x	16.95128	2.518701	3
6D	Pumby8_TI_3	1x	7.682245	0.4350015	3
6D	Pumby8_TI_3	5x	15.35425	1.621024	3
6D	Pumby8_TI_3	10x	13.07604	1.236269	3

Standard deviation for **Fig. 6E**

Panel	Pum	Tandem repeats	Mean	SD	N
6E	PumHD_TI_1	1x	1.82257	0.4971374	3
6E	PumHD_TI_1	5x	1.311775	0.1026023	3
6E	PumHD_TI_1	10x	1.911702	0.324498	3
6E	Pumby8_TI_2	1x	1.620898	0.5932796	3
6E	Pumby8_TI_2	5x	3.45328	0.431692	3
6E	Pumby8_TI_2	10x	1.820353	0.1111846	3
6E	Pumby8_TI_3	1x	1.655751	0.2836693	3
6E	Pumby8_TI_3	5x	2.133241	0.6505566	3
6E	Pumby8_TI_3	10x	1.262076	0.1543146	3

Standard deviation for **Fig. 6F**

Panel	Pum	Tandem repeats	Mean	SD	N
6F	PumHD_TI_1	1x	1.768859	0.1599723	3
6F	PumHD_TI_1	5x	2.182862	0.3021834	3

6F	PumHD_TI_1	10x	1.268697	0.0608768	3
6F	Pumby8_TI_2	1x	1.710847	0.1939064	3
6F	Pumby8_TI_2	5x	1.491666	0.1267124	3
6F	Pumby8_TI_2	10x	1.90308	0.8379387	3
6F	Pumby8_TI_3	1x	1.765795	0.3371155	3
6F	Pumby8_TI_3	5x	2.433637	0.192457	3
6F	Pumby8_TI_3	10x	1.5874	0.2521338	3

Standard deviation for **Fig. 6G**

Panel	Pum	Tandem repeats	Mean	SD	N
6G	PumHD_TI_1	1x	22.23065	1.7792	3
6G	PumHD_TI_1	5x	21.10016	3.832503	3
6G	PumHD_TI_1	10x	22.14464	0.9083386	3
6G	Pumby8_TI_2	1x	22.81259	2.624784	3
6G	Pumby8_TI_2	5x	23.23346	1.024698	3
6G	Pumby8_TI_2	10x	22.58942	1.810628	3
6G	Pumby8_TI_3	1x	21.41398	1.472258	3
6G	Pumby8_TI_3	5x	20.96817	1.68418	3
6G	Pumby8_TI_3	10x	21.42419	0.8597766	3

Standard deviation for **Fig. 6H**

Panel	Pum	Tandem repeats	Mean	SD	N
6H	PumHD_TI_1	1x	20.73316	1.441055	3
6H	PumHD_TI_1	5x	21.05666	0.7110565	3
6H	PumHD_TI_1	10x	21.78599	1.142843	3
6H	Pumby8_TI_2	1x	20.15019	0.7608042	3
6H	Pumby8_TI_2	5x	20.00925	0.9271255	3

6H	Pumby8_TI_2	10x	22.10006	0.8095853	3
6H	Pumby8_TI_3	1x	20.14858	1.522769	3
6H	Pumby8_TI_3	5x	21.17119	1.085817	3
6H	Pumby8_TI_3	10x	20.87999	2.39997	3

Standard deviation for **Fig. 6I**

Panel	Pum	Tandem repeats	Mean	SD	N
6I	PumHD_TI_1	1x	21.80757	1.928916	3
6I	PumHD_TI_1	5x	20.8567	1.431555	3
6I	PumHD_TI_1	10x	20.09514	2.652624	3
6I	Pumby8_TI_2	1x	22.48278	2.152899	3
6I	Pumby8_TI_2	5x	22.22386	0.7464249	3
6I	Pumby8_TI_2	10x	20.47038	2.220453	3
6I	Pumby8_TI_3	1x	20.61409	1.122162	3
6I	Pumby8_TI_3	5x	22.3222	0.9081312	3
6I	Pumby8_TI_3	10x	20.35442	0.8144501	3

Standard deviation for **Fig. 6J**

Panel	Pum	Tandem repeats	Mean	SD	N
6J	PumHD_TI_1	1x	21.69381	1.651343	3
6J	PumHD_TI_1	5x	23.65166	1.69335	3
6J	PumHD_TI_1	10x	23.28157	2.055748	3
6J	Pumby8_TI_2	1x	22.36742	0.8868849	3
6J	Pumby8_TI_2	5x	21.45267	0.9391241	3
6J	Pumby8_TI_2	10x	20.91501	1.462175	3
6J	Pumby8_TI_3	1x	22.71314	1.119192	3
6J	Pumby8_TI_3	5x	21.78995	1.635835	3
6J	Pumby8_TI_3	10x	20.79808	0.5037156	3

Table S8

List of the RNA and protein sequences used in the experiments of **Fig. 6**. Each cluster of 3 bars in **Figs. 6C – 6J** contains the results of a single protein; the three bars represent different numbers of tandem RNA target repeats (distinguished by color). The three proteins are PumHD_TI_1, Pumby8_TI_2, and Pumby8_TI_3, and their corresponding clusters always appear in that order. The RNA targets were included 1, 5, or 10 times, along with 9, 5, or 0 “dummy” sequences (AUAUAU) used to pad the length and keep the overall size of the mRNA constant.

Figs. 6C – 6J

Name	Pum Type	Protein sequence	On-target RNA sequence (5' to 3')	Off-target RNA sequence	Fusion
PumHD_TI_1	PumHD	GUCAGCUC	CUCGACUG	GAGUUGGA (Pumby8_TI_2)	eIF4E
Pumby8_TI_2	Pumby	AGGUUGAG	GAGUUGGA	UAGACUGG (Pumby8_TI_3)	eIF4E
Pumby8_TI_3	Pumby	GGUCAGAU	UAGACUGG	CUCGACUG (PumHD_TI_1)	eIF4E

Fig. 6K.

Name	Pum Type	Mismatch number	RNA target with mismatches	Pum protein	Fusion
Pumby8_TI_2	Pumby	0	GAGUUGGA	AGGUUGAG	eIF4E
Pumby8_TI_2	Pumby	1	GACUUGGA	AGGUUGAG	eIF4E
Pumby8_TI_2	Pumby	2	GACAUGGA	AGGUUGAG	eIF4E
Pumby8_TI_2	Pumby	3	GUCAUGGA	AGGUUGAG	eIF4E
Pumby8_TI_2	Pumby	off-target	UAGACUGG	AGGUUGAG	eIF4E
Pumby8_TI_3	Pumby	0	UAGACUGG	GGUCAGAU	eIF4E
Pumby8_TI_3	Pumby	1	UAGACUAG	GGUCAGAU	eIF4E
Pumby8_TI_3	Pumby	2	UAGACGAG	GGUCAGAU	eIF4E
Pumby8_TI_3	Pumby	3	UAGAUGAG	GGUCAGAU	eIF4E
Pumby8_TI_3	Pumby	off-target	CUCGACUG	GGUCAGAU	eIF4E

Table S9

Statistics for **Fig. S2**.

No samples were excluded from the statistical analysis. The threshold for significance throughout this figure is $\alpha = 0.05$.

Abbreviations

SD	Standard Deviation
SE	Standard Error
DoF	Degrees of Freedom
Sig?	Significant?
Diff.	Difference

Number of Biological Replicates

Figure	Variable	Replicates
S2B	Time = 0h	6
S2B	All others	3
S2C	All	3
S2D	All	3

Two-way ANOVAs for **Figs. S2B** and **S2C**.

ANOVAs with factors of ‘Target Site’ and ‘Treatment Point’

Group	Source of effects	Number of Parameters	DoF	Sum of Squares	F Ratio	Prob > F	Sig?
S2B	Target Site	2	2	0.02561342	2.17	0.1312	
S2B	Treatment Point	2	2	0.3295258	27.9181	<.0001	Yes
S2C	Target Site	2	2	0.013292	0.7249	0.4956	
S2C	Treatment Point	2	2	11.401527	621.796	<.0001	Yes

ANOVAs with factors of ‘Treatment Point’ and ‘Pumby8 or PumHD’

Group	Source of effects	Number of Parameters	DoF	Sum of Squares	F Ratio	Prob > F	Sig?
S2B	Treatment Point	2	2	0.3295258	26.0695	<.0001	Yes
S2B	Pumby8 or PumHD	1	1	0.00632042	1	0.3248	
S2C	Treatment Point	2	2	11.401527	634.8298	<.0001	Yes
S2C	Pumby8 or PumHD	1	1	0.008453	0.9413	0.342	

ANOVA for **Fig. S2D**

Two-way ANOVA with factors of “Pum Tested” (comparing between two particular proteins into which we introduced mismatch mutations) and “Mismatch number”.

Figure	Source of Variation	P value	Summary	Sig?	F (DFn, DFd)	DoF
--------	---------------------	---------	---------	------	--------------	-----

S2D	Interaction	0.0011	**	Yes	F (4, 20) = 6.968	P = 0.0011
S2D	Mismatch number	< 0.0001	****	Yes	F (4, 20) = 118.4	P < 0.0001
S2D	Pum tested	0.0043	**	Yes	F (1, 20) = 10.35	P = 0.0043

Tukey's tests for **Figs. S2B** and **S2C**, after two-way ANOVA with factors of 'Target site' and 'Treatment Point'.

Panel	Factor 1	Factor 2	Difference	SE	t Ratio	Prob > t	Sig?
S2B	ATF4-1	ATF4-2	0.0002448	0.0313625	0.01	1	
S2B	ATF4-1	Off-Target	0.0567054	0.0313625	1.81	0.1836	
S2B	ATF4-2	Off-Target	0.0564606	0.0313625	1.8	0.1862	
S2B	0h	12h -tg	0.02361	0.0313625	0.75	0.7342	
S2B	0h	12h +tg	-0.211955	0.0313625	-6.76	<.0001	Yes
S2B	12h -tg	12h +tg	-0.235565	0.0362143	-6.5	<.0001	Yes
S2C	ATF4-1	ATF4-2	-0.021139	0.0451374	-0.47	0.8867	
S2C	ATF4-1	Off-Target	0.032792	0.0451374	0.73	0.7506	
S2C	ATF4-2	Off-Target	0.053931	0.0451374	1.19	0.4686	
S2C	0h	12h -tg	-0.01793	0.0451374	-0.4	0.917	
S2C	0h	12h +tg	-1.38738	0.0451374	-30.74	<.0001	Yes
S2C	12h -tg	12h +tg	-1.36944	0.0451374	-30.34	<.0001	Yes

Fig. S2D

Dunnnett's multiple comparison test across the values of "Mismatch number", and against the off-target case, after the ANOVA with factors of "Pum Tested" and "Mismatch number".

Figure	Test details	Mean Diff.	SE of diff.	Adjusted P Value	Summary	Sig?
S2D	0 vs. 8	0.4049	0.02438	< 0.0001	****	Yes
S2D	1 vs. 8	0.264	0.02438	< 0.0001	****	Yes
S2D	2 vs. 8	0.01269	0.02438	0.9577	ns	
S2D	3 vs. 8	0.0001464	0.02438	> 0.9999	ns	No

All raw datapoints are plotted in the figure so that the entire dataset is available to the reader, since each individual condition has fewer than 5 datapoints, and thus formal evaluations of normality are not appropriate.

Standard deviation for Fig. S2B

	ATF4-1	ATF4-2	Off Target

Bar	Mean	SD	N	Mean	SD	N	Mean	SD	N
0h	0.1773049	0.01924965	6	0.1517214	0.0244773	6	0.1880018	0.04844195	6
12h +tg	0.4707347	0.04815675	3	0.4938084	0.02126303	3	0.1883514	0.05082738	3
12h -tg	0.1283059	0.01440251	3	0.1554198	0.01895663	3	0.1624738	0.01484854	3

Standard deviation for **Fig. S2C**

	ATF4-1			ATF4-2			Off Target		
Bar	Mean	SD	N	Mean	SD	N	Mean	SD	N
0h	0.263875	0.03540149	3	0.2709583	0.01219652	3	0.2580417	0.03105699	3
12h +tg	1.641667	0.1108708	3	1.734292	0.1546578	3	1.579042	0.2200438	3
12h -tg	0.3042917	0.02908645	3	0.268	0.006557445	3	0.274375	0.007512483	3

Table S10

List of the Pum target sequences for the experiments of **Fig. S2**.

Each mRNA target site contains two 8-base binding sites, one for each of the two Pum proteins needed to reconstitute a split reporter protein: Pum1 (fused to N-terminal portion of split GFP, N-GFP) binds to the target site whose name ends in “A”; Pum2 (fused to C-terminal portion of split GFP, C-GFP) binds to the binding site whose name ends in “B”.

Figs. S2B – S2C

Name	Pum Type	RNA sequence Pum binds to (5' to 3')	Protein sequence	Fusion
PumHD_TM_7A	PumHD	UGAGCUUC	CUUCGAGU	N-GFP
PumHD_TM_7B	PumHD	CAGCGAGG	GGAGCGAC	C-GFP
Pumby8_TM_8A	Pumby	GACAGAUU	UUAGACAG	N-GFP
Pumby8_TM_8B	Pumby	UUGGAGAA	AAGAGGUU	C-GFP
PumHD_TM_9A	PumHD	AUAGGUGU	UGUGGAUA	N-GFP
PumHD_TM_9B	PumHD	GCGAGCAC	CACGAGCG	C-GFP

Fig. S2D

Name	Pum Type	Mismatch number	RNA target with mismatches	Pum protein	Fusion
PumHD_TM_7A	PumHD	0	UGAGCUUC	CUUCGAGU	N-GFP
PumHD_TM_7A-M1	PumHD	1	UGAGCUUG	GUUCGAGU	N-GFP
PumHD_TM_7A-M2	PumHD	2	UGAGCUAG	GAUCGAGU	N-GFP
PumHD_TM_7A-M3	PumHD	3	UGAGCAAG	GAACGAGU	N-GFP
Pumby8_TM_8A	Pumby	0	GACAGAUU	UUAGACAG	N-GFP
Pumby8_TM_8A-M1	Pumby	1	GACAGAUA	AUAGACAG	N-GFP
Pumby8_TM_8A-M2	Pumby	2	GAGAGAUA	AUAGAGAG	N-GFP
Pumby8_TM_8A-M3	Pumby	3	UAGAGAUA	AUAGAGAU	N-GFP

Table S11

Statistics for **Fig. S3**.

No samples were excluded from the statistical analysis. The threshold for significance throughout this figure is $\alpha = 0.05$.

Abbreviations

SD	Standard Deviation
SS	Sum of Squares
DoF	Degrees of Freedom
Sig?	Significant?
Diff.	Difference
NParm	Number of parameters

Biological Replicates

Tile	Replicates
All tiles	3

Three-way ANOVA for **Fig. S3** with factors of 'Pum Protein', 'mRNA Target', and 'Target Match'.

Source	Number of Parameters	DoF	Sum of Squares	F ratio	Prob > F	Sig?
Pum Protein	6	6	0.00000957	2.7372	0.0154	Yes
mRNA target	6	6	0.00000859	2.4578	0.0276	Yes
Target Match	1	1	0.00291382	5002.7474	<.0001	Yes

Two-way ANOVA for **Fig. S3** with factors of 'Pum Protein' and 'mRNA Target', for the data where 'Target Match' has a value of 'No'.

Source	Number of Parameters	DoF	Sum of Squares	F ratio	Prob > F	Sig?
Pum Protein	6	6	2.94E-07	0.7325	0.6244	
mRNA target	6	6	3.70E-07	0.9221	0.4819	

Tukey's test for values of the factor 'Target Match', after three-way ANOVA with factors of 'Pum Protein', 'mRNA Target', and 'Target Match'.

Target Match 1	Target Match 2	Diff.	SE of Diff.	t Ratio	Prob > t 	Sig?
No	Yes	-0.012723	0.0001799	-70.73	<.0001	Yes

Tukey's test for values of the factor 'Pum Protein', after three-way ANOVA with factors of 'Pum Protein', 'mRNA Target', and 'Target Match'.

Pum Protein 1	Pum Protein 2	Diff.	SE of Diff.	t Ratio	Prob > t 	Sig?
Pumby8_TM_1A	Pumby8_TM_3A	-0.00016	0.0002355	-0.07	1	
Pumby8_TM_1A	Pumby8_TM_4A	-0.000294	0.0002355	-1.25	0.8735	
Pumby8_TM_1A	Pumby8_TM_5A	-0.000554	0.0002355	-2.35	0.2276	
Pumby8_TM_1A	Pumby8_TM_8A	-0.000605	0.0002355	-2.57	0.1447	
Pumby8_TM_1A	PumHD_TM_2A	-0.000431	0.0002355	-1.83	0.5316	
Pumby8_TM_1A	PumHD_TM_6A	-0.000679	0.0002355	-2.88	0.0676	
Pumby8_TM_3A	Pumby8_TM_4A	-0.000278	0.0002355	-1.18	0.9007	
Pumby8_TM_3A	Pumby8_TM_5A	-0.000538	0.0002355	-2.28	0.2601	
Pumby8_TM_3A	Pumby8_TM_8A	-0.000588	0.0002355	-2.5	0.1684	
Pumby8_TM_3A	PumHD_TM_2A	-0.000414	0.0002355	-1.76	0.5779	
Pumby8_TM_3A	PumHD_TM_6A	-0.000662	0.0002355	-2.81	0.0807	
Pumby8_TM_4A	Pumby8_TM_5A	-0.00026	0.0002355	-1.1	0.9261	
Pumby8_TM_4A	Pumby8_TM_8A	-0.000311	0.0002355	-1.32	0.8422	
Pumby8_TM_4A	PumHD_TM_2A	-0.000136	0.0002355	-0.58	0.9973	
Pumby8_TM_4A	PumHD_TM_6A	-0.000384	0.0002355	-1.63	0.6618	
Pumby8_TM_5A	Pumby8_TM_8A	-0.000051	0.0002355	-0.22	1	
Pumby8_TM_5A	PumHD_TM_2A	0.000123	0.0002355	0.52	0.9985	
Pumby8_TM_5A	PumHD_TM_6A	-0.000125	0.0002355	-0.53	0.9984	
Pumby8_TM_8A	PumHD_TM_2A	0.000174	0.0002355	0.74	0.9898	
Pumby8_TM_8A	PumHD_TM_6A	-0.000074	0.0002355	-0.31	0.9999	
PumHD_TM_2A	PumHD_TM_6A	-0.000248	0.0002355	-1.05	0.9404	

Tukey's test for values of the factor 'mRNA Target', after three-way ANOVA with factors of 'Pum Protein', 'mRNA Target', and 'Target Match'.

mRNA Target 1	mRNA Target 2	Diff.	SE of Diff.	t Ratio	Prob > t 	Sig?
Pumby8_TM_1A	Pumby8_TM_3A	-0.000043	0.0002355	-0.18	1	
Pumby8_TM_1A	Pumby8_TM_4A	-0.000258	0.0002355	-1.1	0.928	
Pumby8_TM_1A	Pumby8_TM_5A	-0.000533	0.0002355	-2.26	0.2696	
Pumby8_TM_1A	Pumby8_TM_8A	-0.000576	0.0002355	-2.45	0.1875	
Pumby8_TM_1A	PumHD_TM_2A	-0.000294	0.0002355	-1.25	0.8741	
Pumby8_TM_1A	PumHD_TM_6A	-0.000661	0.0002355	-2.81	0.0816	
Pumby8_TM_3A	Pumby8_TM_4A	-0.000215	0.0002355	-0.91	0.9697	
Pumby8_TM_3A	Pumby8_TM_5A	-0.00049	0.0002355	-2.08	0.3699	
Pumby8_TM_3A	Pumby8_TM_8A	-0.000534	0.0002355	-2.27	0.2688	
Pumby8_TM_3A	PumHD_TM_2A	-0.000251	0.0002355	-1.06	0.9372	
Pumby8_TM_3A	PumHD_TM_6A	-0.000618	0.0002355	-2.63	0.1268	
Pumby8_TM_4A	Pumby8_TM_5A	-0.000275	0.0002355	-1.17	0.9053	
Pumby8_TM_4A	Pumby8_TM_8A	-0.000318	0.0002355	-1.35	0.8265	
Pumby8_TM_4A	PumHD_TM_2A	-0.000035	0.0002355	-0.15	1	
Pumby8_TM_4A	PumHD_TM_6A	-0.000403	0.0002355	-1.71	0.6103	
Pumby8_TM_5A	Pumby8_TM_8A	-0.000043	0.0002355	-0.18	1	
Pumby8_TM_5A	PumHD_TM_2A	0.000239	0.0002355	1.02	0.9495	
Pumby8_TM_5A	PumHD_TM_6A	-0.000128	0.0002355	-0.54	0.9981	
Pumby8_TM_8A	PumHD_TM_2A	0.000283	0.0002355	1.2	0.8929	
Pumby8_TM_8A	PumHD_TM_6A	-0.000085	0.0002355	-0.36	0.9998	

PumHD_TM_2A	PumHD_TM_6A	-0.000367	0.0002355	-1.56	0.7077	
-------------	-------------	-----------	-----------	-------	--------	--

Three-way ANOVA for **Fig. S3** with factors of 'Pum Type', 'mRNA Target', and 'Target Match'.

Source	Number of Parameters	DoF	Sum of Squares	F ratio	Prob > F	Sig?
mRNA target	6	6	0.00000859	2.3244	0.0361	Yes
Target Match	1	1	0.00291382	4731.1751	<.0001	Yes
Pum Type	1	1	0.00000204	3.312	0.0709	

Two-way ANOVA for **Fig. S3** with factors of 'Pum Type' and 'mRNA Target', for the data where 'Target Match' has a value of 'No'.

Source	Number of Parameters	DoF	Sum of Squares	F ratio	Prob > F	Sig?
mRNA target	6	6	3.13E-07	0.7892	0.5801	
Pum Type	1	1	5.14E-08	0.7782	0.3795	

Table S12

List of Pum proteins for the experiments in **Fig. S3**.

Each experiment was the combination of one protein carrying N-luc, the protein Pumby8_TM_GFP_4B carrying C-Luc, and one landing site.

Column number	Pum protein	Pum target	Fusion
All columns	Pumby8_TM_GFP_4B	CAGCGUGU	C-Luc
1	Pumby8_TM_GFP_4A	ACGGCCAC	N-Luc
2	PumHD_TM_GFP_1A	GAAGGCUA	N-Luc
3	Pumby8_TM_GFP_3A	GCCCGACA	N-Luc
4	Pumby8_TM_BLA_6A	CUGCUGUG	N-Luc
5	Pumby8_TM_BLA_5A	GAGCGACA	N-Luc
6	PumHD_TM_BLA_2A	GACAACAG	N-Luc
7	Pumby8_TM_8A	GACAGAUU	N-Luc

Full sequence of all landing sites used in **Fig. S3**:

Row number	Pums binding to the left site (green). Pumby8_TM_GFP_4B always binds to the right site (magenta).	Full landing site sequence Spacer1 Pum with N-terminal luciferase Spacer 2 Pum with C-terminal luciferase Spacer 3
1	Pumby8_TM_GFP_4A	ACACGGCCACCGUCCAGCGUGUC
2	Pumby8_TM_GFP_4B	ACGAAGGCUACGUCCAGCGUGUC
3	PumHD_TM_GFP_1A	ACGCCGACACGUCCAGCGUGUC
4	Pumby8_TM_GFP_3A	ACCUGCUGUGCGUCCAGCGUGUC
5	Pumby8_TM_BLA_6A	ACGAGCGACACGUCCAGCGUGUC
6	Pumby8_TM_BLA_5A	ACGACAACAGCGUCCAGCGUGUC
7	Pumby8_TM_8A	ACGACAGATTGCGUCCAGCGUGUC

Table S13Statistics for **Fig. S4**.

No samples were excluded from the statistical analysis. The threshold for significance throughout this figure is $\alpha = 0.05$.

Abbreviations

SD	Standard Deviation
SE	Standard Error
DoF	Degrees of Freedom
Sig?	Significant?
Diff.	Difference
CI	Confidence Interval

Biological Replicates

Panel	Replicates
S2B	7
S2C	3

One-way ANOVAs for **Figs. S4B** and **S4C**, with factor of ‘Pum Target Site’.

Figure	P value	P value summary	Sig?	F (DFn, DFd)	DoF
S2B	0.0003	***	Yes	F (7, 48) = 4.997	55
S2C	< 0.0001	****	Yes	F (7, 16) = 21.75	23

Post-hoc tests for **Fig. S4B**

Tukey’s honest significance post-hoc test after one-way ANOVA with factor of ‘Pum Target Site’.

Tukey's multiple comparisons test	Mean Diff.	95% CI of diff.	Sig?	Summary	Adjusted P value
No Pum-PIN vs. PumHD_SP_1	0.389	-2.605 to 3.383		ns	0.9999
No Pum-PIN vs. Pumby8_SP_2	0.2233	-2.771 to 3.217		ns	> 0.9999
No Pum-PIN vs. PumHD_SP_3	1.192	-1.802 to 4.186		ns	0.9082
No Pum-PIN vs. Pumby8_SP_4	2.485	-0.5092 to 5.479		ns	0.1708
No Pum-PIN vs. Pumby10_SP_5	1.978	-1.016 to 4.972		ns	0.434
No Pum-PIN vs. Pumby10_SP_6	2.965	-0.02875 to 5.959		ns	0.0539
No Pum-PIN vs. Pumby10_SP_7	4.206	1.212 to 7.200	Yes	**	0.0012
PumHD_SP_1 vs. Pumby8_SP_2	-0.1657	-3.160 to 2.828		ns	> 0.9999
PumHD_SP_1 vs. PumHD_SP_3	0.8033	-2.191 to 3.797		ns	0.989
PumHD_SP_1 vs. Pumby8_SP_4	2.096	-0.8983 to 5.090		ns	0.3598
PumHD_SP_1 vs. Pumby10_SP_5	1.589	-1.405 to 4.583		ns	0.699
PumHD_SP_1 vs. Pumby10_SP_6	2.576	-0.4178 to 5.570		ns	0.1396
PumHD_SP_1 vs. Pumby10_SP_7	3.817	0.8231 to 6.811	Yes	**	0.0044
Pumby8_SP_2 vs. PumHD_SP_3	0.969	-2.025 to 3.963		ns	0.9682

Pumby8 SP 2 vs. Pumby8 SP 4	2.261	-0.7326 to 5.255		ns	0.2681
Pumby8 SP 2 vs. Pumby10 SP 5	1.754	-1.240 to 4.748		ns	0.5862
Pumby8 SP 2 vs. Pumby10 SP 6	2.742	-0.2521 to 5.736		ns	0.0948
Pumby8 SP 2 vs. Pumby10 SP 7	3.983	0.9889 to 6.977	Yes	**	0.0026
PumHD SP 3 vs. Pumby8 SP 4	1.292	-1.702 to 4.286		ns	0.8671
PumHD SP 3 vs. Pumby10 SP 5	0.7852	-2.209 to 3.779		ns	0.9904
PumHD SP 3 vs. Pumby10 SP 6	1.773	-1.221 to 4.767		ns	0.5733
PumHD SP 3 vs. Pumby10 SP 7	3.014	0.01982 to 6.008	Yes	*	0.0474
Pumby8 SP 4 vs. Pumby10 SP 5	-0.5071	-3.501 to 2.487		ns	0.9994
Pumby8 SP 4 vs. Pumby10 SP 6	0.4805	-2.514 to 3.474		ns	0.9996
Pumby8 SP 4 vs. Pumby10 SP 7	1.721	-1.273 to 4.715		ns	0.609
Pumby10 SP 5 vs. Pumby10 SP 6	0.9876	-2.006 to 3.982		ns	0.9647
Pumby10 SP 5 vs. Pumby10 SP 7	2.229	-0.7654 to 5.223		ns	0.285
Pumby10 SP 6 vs. Pumby10 SP 7	1.241	-1.753 to 4.235		ns	0.8894

Post-hoc tests for **Fig. S4C**

Tukey's honest significance post-hoc test after one-way ANOVA with factor of 'Pum Target Site'.

Tukey's multiple comparisons test	Mean Diff.	95% CI of diff.	Sig?	Summary	Adjusted P Value
No Pum-PIN vs. PumHD SP 1	0.6382	0.2800 to 0.9965	Yes	***	0.0003
No Pum-PIN vs. Pumby8 SP 2	0.4809	0.1227 to 0.8392	Yes	**	0.0051
No Pum-PIN vs. PumHD SP 3	0.7831	0.4249 to 1.141	Yes	****	< 0.0001
No Pum-PIN vs. Pumby8 SP 4	0.9444	0.5861 to 1.303	Yes	****	< 0.0001
No Pum-PIN vs. Pumby10 SP 5	0.9699	0.6117 to 1.328	Yes	****	< 0.0001
No Pum-PIN vs. Pumby10 SP 6	1.002	0.6438 to 1.360	Yes	****	< 0.0001
No Pum-PIN vs. Pumby10 SP 7	0.9107	0.5525 to 1.269	Yes	****	< 0.0001
PumHD SP 1 vs. Pumby8 SP 2	-0.1573	-0.5155 to 0.2009		ns	0.7866
PumHD SP 1 vs. PumHD SP 3	0.1448	-0.2134 to 0.5031		ns	0.8447
PumHD SP 1 vs. Pumby8 SP 4	0.3061	-0.05210 to 0.6643		ns	0.1243
PumHD SP 1 vs. Pumby10 SP 5	0.3317	-0.02654 to 0.6899		ns	0.0802
PumHD SP 1 vs. Pumby10 SP 6	0.3638	0.005595 to 0.7220	Yes	*	0.0452
PumHD SP 1 vs. Pumby10 SP 7	0.2725	-0.08574 to 0.6307		ns	0.2134
Pumby8 SP 2 vs. PumHD SP 3	0.3021	-0.05608 to 0.6604		ns	0.1329

Pumby8_SP_2 vs. Pumby8_SP_4	0.4634	0.1052 to 0.8216	Yes	**	0.007
Pumby8_SP_2 vs. Pumby10_SP_5	0.489	0.1308 to 0.8472	Yes	**	0.0043
Pumby8_SP_2 vs. Pumby10_SP_6	0.5211	0.1629 to 0.8793	Yes	**	0.0024
Pumby8_SP_2 vs. Pumby10_SP_7	0.4298	0.07156 to 0.7880	Yes	*	0.0133
PumHD_SP_3 vs. Pumby8_SP_4	0.1613	-0.1969 to 0.5195		ns	0.7665
PumHD_SP_3 vs. Pumby10_SP_5	0.1868	-0.1714 to 0.5450		ns	0.6254
PumHD_SP_3 vs. Pumby10_SP_6	0.219	-0.1392 to 0.5772		ns	0.445
PumHD_SP_3 vs. Pumby10_SP_7	0.1276	-0.2306 to 0.4858		ns	0.9098
Pumby8_SP_4 vs. Pumby10_SP_5	0.02556	-0.3327 to 0.3838		ns	> 0.9999
Pumby8_SP_4 vs. Pumby10_SP_6	0.05769	-0.3005 to 0.4159		ns	0.999
Pumby8_SP_4 vs. Pumby10_SP_7	-0.03364	-0.3919 to 0.3246		ns	> 0.9999
Pumby10_SP_5 vs. Pumby10_SP_6	0.03213	-0.3261 to 0.3903		ns	> 0.9999
Pumby10_SP_5 vs. Pumby10_SP_7	-0.0592	-0.4174 to 0.2990		ns	0.9988
Pumby10_SP_6 vs. Pumby10_SP_7	-0.09133	-0.4495 to 0.2669		ns	0.9836

One-way ANOVAs for **Fig. S4B**, with factor of 'Pum Type'.

Source	Number of parameters	DoF	Sum of Squares	F Ratio	Prob > F
Pum Type	3	3	71.71247	6.6245	0.0007

Post-hoc tests for **Fig. S4B**

Tukey's honest significance post-hoc test after one-way ANOVA with factor of 'Pum Type'.

Factor 1	Factor 2	Diff.	SE of Diff.	t Ratio	Prob > t
None	Pumby10	3.04968	0.8290518	3.68	0.003
None	Pumby8	1.35405	0.8793422	1.54	0.4216
None	PumHD	0.79071	0.8793422	0.9	0.8053
Pumby10	Pumby8	-1.69563	0.655423	-2.59	0.0586
Pumby10	PumHD	-2.25897	0.655423	-3.45	0.006
Pumby8	PumHD	-0.56333	0.7179799	-0.78	0.8611

All raw datapoints are plotted in the figure so that the entire dataset is available to the reader, since each individual condition has less than 5 datapoints, and thus formal evaluations of normality are not appropriate.

Standard deviation for **Fig. S4B**

Panel	Mean	SD	N
No Pum-PIN	0.2595238	0.6526511	7
PumHD_SP_1	-0.1295238	0.7800437	7
Pumby8_SP_2	0.03619047	0.7542802	7
PumHD_SP_3	-0.9328571	0.9240359	7
Pumby8_SP_4	-2.225238	1.515329	7
Pumby10_SP_5	-1.718095	1.358509	7
Pumby10_SP_6	-2.705715	2.307108	7
Pumby10_SP_7	-3.946667	3.616957	7

Standard deviation for **Fig. S4C**

Panel	Mean	SD	N
No Pum-PIN	1.415493	0.04474581	3
PumHD_SP_1	0.7772467	0.2202205	3
Pumby8_SP_2	0.9345433	0.07671259	3
PumHD_SP_3	0.6324067	0.1415307	3
Pumby8_SP_4	0.4711367	0.08493729	3
Pumby10_SP_5	0.4455767	0.08058548	3
Pumby10_SP_6	0.4134367	0.1197959	3
Pumby10_SP_7	0.50477	0.1548788	3

Table S14

The list of sequences for experiments of **Figs. S4** and **S5**.

Name	Pum Type	Location of the target sequence in mRNA molecule	RNA sequence Pum binds to (5' to 3')	Protein sequence	Fusion
PumHD_SP_1	PumHD	5'UTR	AGCGCCAC	CACCGCGA	PIN nuclease
Pumby8_SP_2	Pumby	ORF1	CAGAAGCU	UCGAAGAC	PIN nuclease
PumHD_SP_3	PumHD	ORF2	CUCAGCGU	UGCGACUC	PIN nuclease
Pumby8_SP_4	Pumby	ORF3	CCGGUAAG	GAAUGGCC	PIN nuclease
Pumby10_SP_5	Pumby (10 units)	ORF4	GGGUGC GCCG	GCCGCU GGGG	PIN nuclease
Pumby10_SP_6	Pumby (10 units)	ORF5	GCCGUGACUA	AUCAGUGCCG	PIN nuclease
Pumby10_SP_7	Pumby (10 units)	3'UTR	GGUACCUCUA	AUCUCCAUGG	PIN nuclease

Table S15Sequences of Pum proteins used on **Fig. S7**.

Name	Pum Type	RNA sequence Pum binds to (5' to 3')	Protein sequence	Fusion
PumHD_M_1	PumHD	UGUAUAUA	AUAUAUGU	None
PumHD_M_2	PumHD	CAGUGUGC	CGUGUGAC	None
Pumby_M_3	Pumby	UGUAUAUA	AUAUAUGU	None
Pumby_M_4	Pumby	CAGUGUGC	CGUGUGAC	None

Table S16

Binding of PumHD variants and Pumby variants to cognate and noncognate RNA as measured via fluorescence anisotropy of the FAM-labeled RNA target (**Fig. S8**).

Protein	Active fraction	Pum protein sequence	Cognate RNA	K_a	STDev K_a	K_d nM	Noncognate RNA
PumHD wild-type	0.35	AUAUAUGU	UGUAUAUA	1.13E+10	$\pm 1.71e+009$	0.088	ACAUAUAU
PumHD_KD_1	0.34	CGCUCGUG	GUGCUCGC	8.50E+09	$\pm 9.46e+008$	0.118	CACGAGCG
PumHD_KD_2	0.22	GACUGUAC	CAUGUCAG	2.77E+09	$\pm 2.97e+008$	0.362	GUACAGUC
Pumby8_KD_3	0.27	AUAGAUGU	UGUAGUAU	7.44E+08	$\pm 2.71e+008$	1.343	ACAUCUAU
Pumby8_KD_4	0.26	GACUGUAC	CAUGUCAG	2.28E+09	$\pm 1.71e+009$	0.439	GUACAGUC

Table S17

Non-specific and incorrect binding of Pum sequences to the GFP and BLA genes used in **Fig. 5**. “Luc” stands for split firefly luciferase.

Name	Pum Type	RNA sequence Pum binds to (5' to 3')	Protein sequence	Fusion
Pumby8_TM_10A	Pumby	GAAACACU	UCACAAAG	N-Luc
Pumby8_TM_10B	Pumby	AGGUGAAG	GAAGUGGA	C-Luc
Pumby8_TM_11A	Pumby	GGAACCGG	GGCCAAGG	N-Luc
Pumby8_TM_11B	Pumby	AGCCGAAA	AAAGCCGA	C-Luc
Pumby8_TM_12A	Pumby	GCUGACCC	CCCAGUCG	N-Luc
Pumby8_TM_12B	Pumby	UUCAUCUG	GUCUACUU	C-Luc
PumHD_TM_13A	PumHD	AGGGCAUC	CUACGGGA	N-Luc
PumHD_TM_13B	PumHD	CAAGGAGG	GGAGGAAC	C-Luc
Pumby8_TM_14A	Pumby	GGAUCACU	UCACUAGG	N-Luc
Pumby8_TM_14B	Pumby	CAUGGACG	GCAGGUAC	C-Luc

Table S18

Previously reported mutants of PumHD binding different 8-mer target RNA sequences, tested in cells and/or cell-free systems.

PumHD mutant	Usage in the literature
1G	Binding to cognate and non-cognate targets via electrophoretic mobility shift; mutants tested along with testing WT PumHD. The non-cognate sequence was the NRE. (11)
1C	New C binding code, tested using a yeast three-hybrid system, as well as by an electrophoretic mobility shift assay. (12)
2C	New C binding code, tested using a yeast three-hybrid system, as well as by an electrophoretic mobility shift assay. (12)
	New C binding code, tested in yeast. (13)
3C	New C binding code, tested using a yeast three-hybrid system, as well as by an electrophoretic mobility shift assay. (12)
	New C binding code, tested in yeast. (13)
3G	Binding to cognate and non-cognate targets via electrophoretic mobility shift; mutants tested along with testing WT PumHD. The non-cognate sequence was the NRE. (11)
	New C binding code, tested using a yeast three-hybrid system, as well as by an electrophoretic mobility shift assay. (12)
	Use split GFP complementation to visualize binding of PumHD in mammalian cells. (14)
	Silence mRNA transcripts, both in <i>E. coli</i> and in the mitochondria of cultured human HEK293 cells, by fusing the non-specific nuclease PIN domain protein to the Pum protein mutants. (7)
	Alter the splicing of human gene Bcl-X. (15)
	Enhance translation via eIF4E fusion. (16)
3U	New C binding code, tested using a yeast three-hybrid system, as well as by an electrophoretic mobility shift assay. (12)
	Binding to cognate and non-cognate targets via electrophoretic mobility shift; mutants tested along with testing WT PumHD. The non-cognate sequence was

	the NRE. (11)
4C	New C binding code, tested using a yeast three-hybrid system, as well as by an electrophoretic mobility shift assay. (12)
4G	Image viral RNA in plant cells. (17)
5C	New C binding code, tested using a yeast three-hybrid system, as well as by an electrophoretic mobility shift assay. (12)
	New C binding code, tested in yeast. (13)
6C	New C binding code, tested using a yeast three-hybrid system, as well as by an electrophoretic mobility shift assay. (12)
	New C binding code, tested in yeast. (13)
6G	New C binding code, tested in yeast. (13)
	Visualize the human mitochondrial gene NT-ND6 through Pum-mediated split fluorescent protein reconstitution, and assessed cognate and non-cognate binding via electrophoretic mobility shift.(10)
6A	New C binding code, tested in yeast. (13)
7C	New C binding code, tested in yeast. (13)
	New C binding code, tested using a yeast three-hybrid system, as well as by an electrophoretic mobility shift assay. (12)
7U	Suppress translation using the post-transcriptional regulator TTP in human cells. (1)
7A	Suppress translation using the post-transcriptional regulator TTP in human cells. (1)
8C	New C binding code, tested using a yeast three-hybrid system, as well as by an electrophoretic mobility shift assay. (12)
1G/3G	Visualize the localization of β -actin mRNA. (18)
2C/6C	New C binding code, tested in yeast. (13)
3U/7U	Binding to cognate and non-cognate targets via electrophoretic mobility shift; mutants tested along with testing WT PumHD. The non-cognate sequence was the NRE. (11)
6G/7U	Alter the splicing of human gene Bcl-X. (15)
	Binding to cognate and non-cognate targets via electrophoretic mobility shift;

	mutants tested along with testing WT PumHD. The non-cognate sequence was the NRE. (11)
	Silence mRNA transcripts, both in E. coli and in the mitochondria of cultured human HEK293 cells, by fusing the non-specific nuclease PIN domain protein to the Pum protein mutants. (7)
	Enhance translation via eIF4E fusion. (16)
1U/8A	Test binding of PumHD and mutants to cognate and non-cognate targets by fluorescent anisotropy assay. (1)
4G/6C/7A	Suppress translation using the post-transcriptional regulator TTP in human cells. (1)
2G/3G/5G	
1U/2A/8A	
2A/5U/7C	
1G/3G/5G	Alter the splicing of human gene Bcl-X. (15)
	Silence mRNA transcripts, both in E. coli and in the mitochondria of cultured human HEK293 cells, by fusing the non-specific nuclease PIN domain protein to the Pum protein mutants. (7)
1G/6G/7U	Silence mRNA transcripts, both in E. coli and in the mitochondria of cultured human HEK293 cells, by fusing the non-specific nuclease PIN domain protein to the Pum protein mutants. (7)
1U/2G/3U/5G	Use split GFP complementation to visualize binding of PumHD in mammalian cells and visualize the localization of β -actin mRNA. (14, 18)
1U/2A/3U/4A	Suppress translation using the post-transcriptional regulator TTP in human cells. (1)
5U/6A/7C/8A	
3U/4A/5U/6A	Test binding of PumHD and mutants to cognate and non-cognate targets by fluorescent anisotropy assay. (1)
3G/4G/5U/6A	Visualize the human mitochondrial gene NT-ND6 through Pum-mediated split fluorescent protein reconstitution, and assessed cognate and non-cognate binding via electrophoretic mobility shift.(10)
1G/3C/4G/5U/6C	New C binding code, tested in yeast. (13)
1G/2A/6G/7U/8G	Silence mRNA transcripts, both in E. coli and in the mitochondria of cultured human HEK293 cells, by fusing the non-specific nuclease PIN domain protein to the Pum protein mutants. (7)

1U/3G/4A/5U/6A	Image viral RNA in plant cells. (17)
1U/3U/4A/5U /6A/8A	Test binding of PumHD and mutants to cognate and non-cognate targets by fluorescent anisotropy assay. (1)
1C/2G/3U/4C /5G/7C/8G	New C binding code, tested in yeast. (13)
1U/2G/3U/4A/5U /6A/7U/8A	Test binding of PumHD and mutants to cognate and non-cognate targets by fluorescent anisotropy assay. (1)

Table S19

The list of PumHD units with Golden Gate cloning overhangs (the "units" used in the PumHD assembly, see **Materials and Methods**)

Unit	Sequence
0	GTCATGCGTCTCCGGATCGATAGTAGCGGTCTCCCGAGGCATGGGCCGAGCCGCCTTTTGGAAGATTT CGAAACAACCGGTACCCCAATTTACAACCTGCGGGAGATTGCCGGAGGAGACGGAGTGT
1A	GTCATGCGTCTCCCGGACATATAATGGAATTTCCCAAGACCAGCATGGGTCCAGATTCATTCAGCTGAA ACTGGAGCGTGCCACACCAGCTGAGCGCCAGCTTGCTTCAATGAAATCCTCCAGGGAGACGGAGTGT
1C	GTCATGCGTCTCCCGGACATATAATGGAATTTCCCAAGACCAGCATGGGTCCAGATTCATTCGCTGAA ACTGGAGCGTGCCACACCAGCTGAGCGCCAGCTTGCTTCAATGAAATCCTCCAGGGAGACGGAGTGT
1G	GTCATGCGTCTCCCGGACATATAATGGAATTTCCCAAGACCAGCATGGGTCCAGATTCATTGAGCTGAA ACTGGAGCGTGCCACACCAGCTGAGCGCCAGCTTGCTTCAATGAAATCCTCCAGGGAGACGGAGTGT
1U	GTCATGCGTCTCCCGGACATATAATGGAATTTCCCAAGACCAGCATGGGAACAGATTCATTCAGCTGA AACTGGAGCGTGCCACACCAGCTGAGCGCCAGCTTGCTTCAATGAAATCCTCCAGGGAGACGGAGTGT
2A	GTCATGCGTCTCCCCAGGCTGCCTACCAACTCATGGTGGATGTGTTTGGTTGTTACGTCATTCAGAAGTT CTTTGAATTTGGCAGTCTTGAACAGAAGCTGGCTTTGGCAGAACGGATTTCGAGGTGGAGACGGAGTGT
2C	GTCATGCGTCTCCCCAGGCTGCCTACCAACTCATGGTGGATGTGTTTGGTAGTTACGTCATTCGCAAGTT CTTTGAATTTGGCAGTCTTGAACAGAAGCTGGCTTTGGCAGAACGGATTTCGAGGTGGAGACGGAGTGT
2G	GTCATGCGTCTCCCCAGGCTGCCTACCAACTCATGGTGGATGTGTTTGGTAGTTACGTCATTCAGAAGTT CTTTGAATTTGGCAGTCTTGAACAGAAGCTGGCTTTGGCAGAACGGATTTCGAGGTGGAGACGGAGTGT
2U	GTCATGCGTCTCCCCAGGCTGCCTACCAACTCATGGTGGATGTGTTTGGTAATTACGTCATTCAGAAGTT CTTTGAATTTGGCAGTCTTGAACAGAAGCTGGCTTTGGCAGAACGGATTTCGAGGTGGAGACGGAGTGT
3A	GTCATGCGTCTCCAGGTCACGTCTGTTCATTGGCACTACAGATGTATGGCTGCCGTGTTATCCAGAAAGC TCTTGAGTTTATTCCTTCAGACCAGCAGAATGAGATGGTTCGGGAACTAGATGGCGGAGACGGAGTGT
3C	GTCATGCGTCTCCAGGTCACGTCTGTTCATTGGCACTACAGATGTATGGCTCCCGTGTATCCGCAAAGC TCTTGAGTTTATTCCTTCAGACCAGCAGAATGAGATGGTTCGGGAACTAGATGGCGGAGACGGAGTGT
3G	GTCATGCGTCTCCAGGTCACGTCTGTTCATTGGCACTACAGATGTATGGCTCCCGTGTATCCGAGAAAGC TCTTGAGTTTATTCCTTCAGACCAGCAGAATGAGATGGTTCGGGAACTAGATGGCGGAGACGGAGTGT
3U	GTCATGCGTCTCCAGGTCACGTCTGTTCATTGGCACTACAGATGTATGGCAACCGTGTATCCAGAAAGC TCTTGAGTTTATTCCTTCAGACCAGCAGAATGAGATGGTTCGGGAACTAGATGGCGGAGACGGAGTGT
4A	GTCATGCGTCTCCTGGCCATGTCTTGAAGTGTGTGAAAGATCAGAATGGCTGTTACGTGGTTCAGAAATG CATTGAATGTGTACAGCCCCAGTCTTTGCAATTTATCATCGATGCGTTTAAGGGACAGGAGAGACCGGA TGGCAGAGGATGGAGACGGAGTGT
4C	GTCATGCGTCTCCTGGCCATGTCTTGAAGTGTGTGAAAGATCAGAATGGCAGTTACGTGGTTCGAAATG CATTGAATGTGTACAGCCCCAGTCTTTGCAATTTATCATCGATGCGTTTAAGGGACAGGAGAGACCGGA TGGCAGAGGATGGAGACGGAGTGT

4G	GTCATGCGTCTCCTGGCCATGTCTTGAAGTGTGTGAAAGATCAGAATGGCAGTTACGTGGTTGAGAAAT GCATTGAATGTGTACAGCCCCAGTCTTTGCAATTTATCATCGATGCGTTTAAGGGACAGGAGAGACCGG ATGGCAGAGGATGGAGACGGAGTGT
4U	GTCATGCGTCTCCTGGCCATGTCTTGAAGTGTGTGAAAGATCAGAATGGCAATTACGTGGTTCAGAAAT GCATTGAATGTGTACAGCCCCAGTCTTTGCAATTTATCATCGATGCGTTTAAGGGACAGGAGAGACCGG ATGGCAGAGGATGGAGACGGAGTGT
5A	GTCATGCGTCTCCGGATCGATAGTAGCGGTCTCCAGGTATTTGCCTTATCCACACATCCTTATGGCTGC CGAGTGATTAGAGAATCCTGGAGCACTGTCTCCCTGACCAGACACTCCCTATTTTAGAGGAGCTTCACC AGCACAGGAGACGGAGTGT
5C	GTCATGCGTCTCCGGATCGATAGTAGCGGTCTCCAGGTATTTGCCTTATCCACACATCCTTATGGCTCC CGAGTGATTAGAGAATCCTGGAGCACTGTCTCCCTGACCAGACACTCCCTATTTTAGAGGAGCTTCACC AGCACAGGAGACGGAGTGT
5G	GTCATGCGTCTCCGGATCGATAGTAGCGGTCTCCAGGTATTTGCCTTATCCACACATCCTTATGGCTCC CGAGTGATTAGAGAATCCTGGAGCACTGTCTCCCTGACCAGACACTCCCTATTTTAGAGGAGCTTCACC AGCACAGGAGACGGAGTGT
5U	GTCATGCGTCTCCGGATCGATAGTAGCGGTCTCCAGGTATTTGCCTTATCCACACATCCTTATGGCAAC CGAGTGATTAGAGAATCCTGGAGCACTGTCTCCCTGACCAGACACTCCCTATTTTAGAGGAGCTTCACC AGCACAGGAGACGGAGTGT
6A	GTCATGCGTCTCCACACAGAGCAGCTTGTACAGGATCAATATGGATGTTATGTAATCCAACATGTA GAGCAGGTCGTCCTGAGGATAAAAAGCAAAATTGTAGCAGAAATCCGAGGCAATGGGAGACGGAGTGT
6C	GTCATGCGTCTCCACACAGAGCAGCTTGTACAGGATCAATATGGAAGTTATGTAATCCGCCATGTA GAGCAGGTCGTCCTGAGGATAAAAAGCAAAATTGTAGCAGAAATCCGAGGCAATGGGAGACGGAGTGT
6G	GTCATGCGTCTCCACACAGAGCAGCTTGTACAGGATCAATATGGAAGTTATGTAATCGAACATGTA GGAGCAGGTCGTCCTGAGGATAAAAAGCAAAATTGTAGCAGAAATCCGAGGCAATGGGAGACGGAGT G
6U	GTCATGCGTCTCCACACAGAGCAGCTTGTACAGGATCAATATGGAAGTTATGTAATCCAACATGTA GGAGCAGGTCGTCCTGAGGATAAAAAGCAAAATTGTAGCAGAAATCCGAGGCAATGGGAGACGGAGT G
7A	GTCATGCGTCTCCAATGTA CTTGTATTGAGTCAGCACA AATTTGCATGCAATGTTGT GCAGAAGTGTGTT ACTCACGCCTACGTACGGAGCGCGCTGTGCTCATCGATGAGGTGTGCACCATGAACGACGGTCCCCAC AGTGCCTGGAGACGGAGTGT
7C	GTCATGCGTCTCCAATGTA CTTGTATTGAGTCAGCACA AATTTGCAAGCTATGTTGT GCGCAAGTGTGTT ACTCACGCCTACGTACGGAGCGCGCTGTGCTCATCGATGAGGTGTGCACCATGAACGACGGTCCCCAC AGTGCCTGGAGACGGAGTGT
7G	GTCATGCGTCTCCAATGTA CTTGTATTGAGTCAGCACA AATTTGCAAGCAATGTTGT GGAGAAGTGTGTT ACTCACGCCTACGTACGGAGCGCGCTGTGCTCATCGATGAGGTGTGCACCATGAACGACGGTCCCCAC AGTGCCTGGAGACGGAGTGT
7U	GTCATGCGTCTCCAATGTA CTTGTATTGAGTCAGCACA AATTTGCAAACAATGTTGT GCAGAAGTGTGTT ACTCACGCCTACGTACGGAGCGCGCTGTGCTCATCGATGAGGTGTGCACCATGAACGACGGTCCCCAC AGTGCCTGGAGACGGAGTGT

8A	GTCATGCGTCTCCGCCTTATACACCATGATGAAGGACCAGTATGCCTGCTACGTGGTCCAGAAGATGATT GACGTGGCGGAGCCAGGCCAGCGGAAGATCGTCATGCATAAGATCCGACCCGGAGACGGAGTGT
8C	GTCATGCGTCTCCGCCTTATACACCATGATGAAGGACCAGTATGCCAGCTACGTGGTCCGCAAGATGATT GACGTGGCGGAGCCAGGCCAGCGGAAGATCGTCATGCATAAGATCCGACCCGGAGACGGAGTGT
8G	GTCATGCGTCTCCGCCTTATACACCATGATGAAGGACCAGTATGCCAGCTACGTGGTCCGAGAAGATGAT TGACGTGGCGGAGCCAGGCCAGCGGAAGATCGTCATGCATAAGATCCGACCCGGAGACGGAGTGT
8U	GTCATGCGTCTCCGCCTTATACACCATGATGAAGGACCAGTATGCCAACTACGTGGTCCAGAAGATGAT TGACGTGGCGGAGCCAGGCCAGCGGAAGATCGTCATGCATAAGATCCGACCCGGAGACGGAGTGT
9	GTCATGCGTCTCCACCCACATCGCAACTCTTCGTAAGTACACCTATGGCAAGCACATTCTGGCCAAGCT GGAGAAGTACTACATGAAGAACGGTGTGACTTAGGCCGGACGCAGAGACCGGATGGCAGAGGATGGAG ACGGAGTGT

Table S20

The list of units (with Golden Gate cloning overhangs) used to assemble hexamers for Pumby.

Name	Sequence
module1- hex1 A	GTCATGCGTCTCCAGGTCGATAGTAGCGGTCTCCGAGGCGAACTTCACCAGCACACTGAA CAACTCGTGCAAGACCAGTATGGGTGCTATGTCATCCAACATGTCCTTGAGCACGGACGCC CGAAGACAAGTCAAAGATCGTGGCTGGAGACGGAGTGT
module1- hex1 C	GTCATGCGTCTCCAGGTCGATAGTAGCGGTCTCCGAGGCGAACTTCACCAGCACACTGAA CAACTCGTGCAAGACCAGTATGGGTCTATGTCATCCGGCATGTCCTTGAGCACGGACGCC CGAAGACAAGTCAAAGATCGTGGCTGGAGACGGAGTGT
module1- hex1 G	GTCATGCGTCTCCAGGTCGATAGTAGCGGTCTCCGAGGCGAACTTCACCAGCACACTGAA CAACTCGTGCAAGACCAGTATGGGTCTATGTCATCGAACATGTCCTTGAGCACGGACGCC CGAAGACAAGTCAAAGATCGTGGCTGGAGACGGAGTGT
module1- hex1 U	GTCATGCGTCTCCAGGTCGATAGTAGCGGTCTCCGAGGCGAACTTCACCAGCACACTGAA CAACTCGTGCAAGACCAGTATGGGAACTATGTCATCCAACATGTCCTTGAGCACGGACGCC CCGAAGACAAGTCAAAGATCGTGGCTGGAGACGGAGTGT
module1- hex2 A	GTCATGCGTCTCCAGGTCGATAGTAGCGGTCTCCGAACTTCACCAGCACACTGAACAACCTCG TGCAAGACCAGTATGGGTGCTATGTCATCCAACATGTCCTTGAGCACGGACGCCCGAAGA CAAGTCAAAGATCGTGGCTGGAGACGGAGTGT
module1- hex2 C	GTCATGCGTCTCCAGGTCGATAGTAGCGGTCTCCGAACTTCACCAGCACACTGAACAACCTCG TGCAAGACCAGTATGGGTCTATGTCATCCGGCATGTCCTTGAGCACGGACGCCCGAAGA CAAGTCAAAGATCGTGGCTGGAGACGGAGTGT
module1- hex2 G	GTCATGCGTCTCCAGGTCGATAGTAGCGGTCTCCGAACTTCACCAGCACACTGAACAACCTCG TGCAAGACCAGTATGGGTCTATGTCATCGAACATGTCCTTGAGCACGGACGCCCGAAGA CAAGTCAAAGATCGTGGCTGGAGACGGAGTGT
module1- hex2 U	GTCATGCGTCTCCAGGTCGATAGTAGCGGTCTCCGAACTTCACCAGCACACTGAACAACCTCG TGCAAGACCAGTATGGGAACTATGTCATCCAACATGTCCTTGAGCACGGACGCCCGAAGA CAAGTCAAAGATCGTGGCTGGAGACGGAGTGT
module1- hex3 A	GTCATGCGTCTCCAGGTCGATAGTAGCGGTCTCCGGCTGAACTTCACCAGCACACTGAACAA CTCGTGCAAGACCAGTATGGGTGCTATGTCATCCAACATGTCCTTGAGCACGGACGCCCGA AGACAAGTCAAAGATCGTGGCTGGAGACGGAGTGT
module1- hex3 C	GTCATGCGTCTCCAGGTCGATAGTAGCGGTCTCCGGCTGAACTTCACCAGCACACTGAACAA CTCGTGCAAGACCAGTATGGGTCTATGTCATCCGGCATGTCCTTGAGCACGGACGCCCGA AGACAAGTCAAAGATCGTGGCTGGAGACGGAGTGT
module1- hex3 G	GTCATGCGTCTCCAGGTCGATAGTAGCGGTCTCCGGCTGAACTTCACCAGCACACTGAACAA CTCGTGCAAGACCAGTATGGGTCTATGTCATCGAACATGTCCTTGAGCACGGACGCCCGA AGACAAGTCAAAGATCGTGGCTGGAGACGGAGTGT
module1- hex3 U	GTCATGCGTCTCCAGGTCGATAGTAGCGGTCTCCGGCTGAACTTCACCAGCACACTGAACAA CTCGTGCAAGACCAGTATGGGAACTATGTCATCCAACATGTCCTTGAGCACGGACGCCCG AAGACAAGTCAAAGATCGTGGCTGGAGACGGAGTGT
module1-	GTCATGCGTCTCCAGGTCGATAGTAGCGGTCTCCCTGAACTTCACCAGCACACTGAACAAC

hex4 A	CGTGCAAGACCAGTATGGGTGCTATGTCATCCAACATGTCCTTGAGCACGGACGCCCCGAA GACAAGTCAAAGATCGTGGCTGGAGACGGAGTGT
module1- hex4 C	GTCATGCGTCTCCAGGTCGATAGTAGCGGTCTCCCTGAACTTCACCAGCACACTGAACAACT CGTGCAAGACCAGTATGGGTCTATGTCATCCGGCATGTCCTTGAGCACGGACGCCCCGAA GACAAGTCAAAGATCGTGGCTGGAGACGGAGTGT
module1- hex4 G	GTCATGCGTCTCCAGGTCGATAGTAGCGGTCTCCCTGAACTTCACCAGCACACTGAACAACT CGTGCAAGACCAGTATGGGTCTATGTCATCGAACATGTCCTTGAGCACGGACGCCCCGAA GACAAGTCAAAGATCGTGGCTGGAGACGGAGTGT
module1- hex4 U	GTCATGCGTCTCCAGGTCGATAGTAGCGGTCTCCCTGAACTTCACCAGCACACTGAACAACT CGTGCAAGACCAGTATGGGAACTATGTCATCCAACATGTCCTTGAGCACGGACGCCCCGAA GACAAGTCAAAGATCGTGGCTGGAGACGGAGTGT
module2 A	GTCATGCGTCTCCGGCTGAACTTCACCAGCACACTGAACAACTCGTGCAAGACCAGTATGG GTGCTATGTCATCCAACATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTG GCTGAGGAGACGGAGTGT
module2 C	GTCATGCGTCTCCGGCTGAACTTCACCAGCACACTGAACAACTCGTGCAAGACCAGTATGG GTCCTATGTCATCCGGCATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTG GCTGAGGAGACGGAGTGT
module2 G	GTCATGCGTCTCCGGCTGAACTTCACCAGCACACTGAACAACTCGTGCAAGACCAGTATGG GTCCTATGTCATCGAACATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTG GCTGAGGAGACGGAGTGT
module2 U	GTCATGCGTCTCCGGCTGAACTTCACCAGCACACTGAACAACTCGTGCAAGACCAGTATGG GAACTATGTCATCCAACATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTG GCTGAGGAGACGGAGTGT
module3 A	GTCATGCGTCTCCCTGAACTTCACCAGCACACTGAACAACTCGTGCAAGACCAGTATGGGTG CTATGTCATCCAACATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTGGCT GAACGGAGACGGAGTGT
module3 C	GTCATGCGTCTCCCTGAACTTCACCAGCACACTGAACAACTCGTGCAAGACCAGTATGGGTG CTATGTCATCCGGCATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTGGCT GAACGGAGACGGAGTGT
module3 G	GTCATGCGTCTCCCTGAACTTCACCAGCACACTGAACAACTCGTGCAAGACCAGTATGGGTG CTATGTCATCGAACATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTGGCT GAACGGAGACGGAGTGT
module3 U	GTCATGCGTCTCCCTGAACTTCACCAGCACACTGAACAACTCGTGCAAGACCAGTATGGGA ACTATGTCATCCAACATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTGGC TGAACGGAGACGGAGTGT
module4 A	GTCATGCGTCTCCGAACTTCACCAGCACACTGAACAACTCGTGCAAGACCAGTATGGGTGCT ATGTCATCCAACATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTGGGAGA CGGAGTGT
module4 C	GTCATGCGTCTCCGAACTTCACCAGCACACTGAACAACTCGTGCAAGACCAGTATGGGTCT ATGTCATCCGGCATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTGGGAGA CGGAGTGT

module4 G	GTCATGCGTCTCCGAACTTCACCAGCACACTGAACAACCTCGTGCAAGACCAGTATGGGTCCT ATGTCATCGAACATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTGGGAGA CGGAGTGT
module4 U	GTCATGCGTCTCCGAACTTCACCAGCACACTGAACAACCTCGTGCAAGACCAGTATGGGAAC TATGTCATCCAACATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTGGGAG ACGGAGTGT
module5 A	GTCATGCGTCTCCCGTGGCTGAACTTCACCAGCACACTGAACAACCTCGTGCAAGACCAGTAT GGGTGCTATGTCATCCAACATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCG TGGCGGAGACGGAGTGT
module5 C	GTCATGCGTCTCCCGTGGCTGAACTTCACCAGCACACTGAACAACCTCGTGCAAGACCAGTAT GGGTCCTATGTCATCCGGCATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCG TGGCGGAGACGGAGTGT
module5 G	GTCATGCGTCTCCCGTGGCTGAACTTCACCAGCACACTGAACAACCTCGTGCAAGACCAGTAT GGGTCCTATGTCATCGAACATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCG TGGCGGAGACGGAGTGT
module5 U	GTCATGCGTCTCCCGTGGCTGAACTTCACCAGCACACTGAACAACCTCGTGCAAGACCAGTAT GGGAACTATGTCATCCAACATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCG TGGCGGAGACGGAGTGT
module6- hex1 A	GTCATGCGTCTCCTGGCTGAACTTCACCAGCACACTGAACAACCTCGTGCAAGACCAGTATGG GTGCTATGTCATCCAACATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTG GCTGAACAGAGACCGGATGGCAGAAGGTGGAGACGGAGTGT
module6- hex1 C	GTCATGCGTCTCCTGGCTGAACTTCACCAGCACACTGAACAACCTCGTGCAAGACCAGTATGG GTCCTATGTCATCCGGCATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTG GCTGAACAGAGACCGGATGGCAGAAGGTGGAGACGGAGTGT
module6- hex1 G	GTCATGCGTCTCCTGGCTGAACTTCACCAGCACACTGAACAACCTCGTGCAAGACCAGTATGG GTCCTATGTCATCGAACATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTG GCTGAACAGAGACCGGATGGCAGAAGGTGGAGACGGAGTGT
module6- hex1 U	GTCATGCGTCTCCTGGCTGAACTTCACCAGCACACTGAACAACCTCGTGCAAGACCAGTATGG GAACTATGTCATCCAACATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTG GCTGAACAGAGACCGGATGGCAGAAGGTGGAGACGGAGTGT
module6- hex2 A	GTCATGCGTCTCCTGGCTGAACTTCACCAGCACACTGAACAACCTCGTGCAAGACCAGTATGG GTGCTATGTCATCCAACATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTG GCTAGAGACCGGATGGCAGAAGGTGGAGACGGAGTGT
module6- hex2 C	GTCATGCGTCTCCTGGCTGAACTTCACCAGCACACTGAACAACCTCGTGCAAGACCAGTATGG GTCCTATGTCATCCGGCATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTG GCTAGAGACCGGATGGCAGAAGGTGGAGACGGAGTGT
module6- hex2 G	GTCATGCGTCTCCTGGCTGAACTTCACCAGCACACTGAACAACCTCGTGCAAGACCAGTATGG GTCCTATGTCATCGAACATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTG GCTAGAGACCGGATGGCAGAAGGTGGAGACGGAGTGT
module6- hex2 U	GTCATGCGTCTCCTGGCTGAACTTCACCAGCACACTGAACAACCTCGTGCAAGACCAGTATGG GAACTATGTCATCCAACATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTG

	GCTAGAGACCGGATGGCAGAAGGTGGAGACGGAGTGT
module6- hex3 A	GTCATGCGTCTCCTGGCTGAACTTCACCAGCACACTGAACAACCTCGTGCAAGACCAGTATGG GTGCTATGTCATCCAACATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTG GCTGAAGAGACCGGATGGCAGAAGGTGGAGACGGAGTGT
module6- hex3 C	GTCATGCGTCTCCTGGCTGAACTTCACCAGCACACTGAACAACCTCGTGCAAGACCAGTATGG GTCCTATGTCATCCGGCATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTG GCTGAAGAGACCGGATGGCAGAAGGTGGAGACGGAGTGT
module6- hex3 G	GTCATGCGTCTCCTGGCTGAACTTCACCAGCACACTGAACAACCTCGTGCAAGACCAGTATGG GTCCTATGTCATCGAACATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTG GCTGAAGAGACCGGATGGCAGAAGGTGGAGACGGAGTGT
module6- hex3 U	GTCATGCGTCTCCTGGCTGAACTTCACCAGCACACTGAACAACCTCGTGCAAGACCAGTATGG GAACTATGTCATCCAACATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTG GCTGAAGAGACCGGATGGCAGAAGGTGGAGACGGAGTGT
module6- hex4 A	GTCATGCGTCTCCTGGCTGAACTTCACCAGCACACTGAACAACCTCGTGCAAGACCAGTATGG GTGCTATGTCATCCAACATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTG GCTGGACGCAGAGACCGGATGGCAGAAGGTGGAGACGGAGTGT
module6- hex4 C	GTCATGCGTCTCCTGGCTGAACTTCACCAGCACACTGAACAACCTCGTGCAAGACCAGTATGG GTCCTATGTCATCCGGCATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTG GCTGGACGCAGAGACCGGATGGCAGAAGGTGGAGACGGAGTGT
module6- hex4 G	GTCATGCGTCTCCTGGCTGAACTTCACCAGCACACTGAACAACCTCGTGCAAGACCAGTATGG GTCCTATGTCATCGAACATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTG GCTGGACGCAGAGACCGGATGGCAGAAGGTGGAGACGGAGTGT
module6- hex4 U	GTCATGCGTCTCCTGGCTGAACTTCACCAGCACACTGAACAACCTCGTGCAAGACCAGTATGG GAACTATGTCATCCAACATGTCCTTGAGCACGGACGCCCCGAAGACAAGTCAAAGATCGTG GCTGGACGCAGAGACCGGATGGCAGAAGGTGGAGACGGAGTGT

Table S21

List of all the combinations of PumHD units and stacking amino acids that we tested as potential Pumby modules.

Unit	Stacking amino acid	Protein sequences tested
3	Y	AUAGAUGU; GCGAGCAC; AUAGAUCU; AUAUAUGU; AUAUAU
3	R	AUAGAUGU; GCGAGCAC; AUAGAUCU; AUAUAUGU; AUAUAU
6	R	AUAGAUGU; GCGAGCAC; AUAGAUCU; AUAUAUGU; AUAUAU

References

1. Abil Z, Denard CA, Zhao H (2014) Modular assembly of designer PUF proteins for specific post-transcriptional regulation of endogenous RNA. *J Biol Eng* 8(1):7.
2. Sanjana NE, et al. (2012) A transcription activator-like effector toolbox for genome engineering. *Nat Protoc* 7(1):171–92.
3. Dey S, et al. (2010) Both transcriptional regulation and translational control of ATF4 are central to the integrated stress response. *J Biol Chem* 285(43):33165–74.
4. Whitney ML, Jefferson LS, Kimball SR (2009) ATF4 is necessary and sufficient for ER stress-induced upregulation of REDD1 expression. *Biochem Biophys Res Commun* 379(2):451–5.
5. Kertesz M, et al. (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature* 467(7311):103–7.
6. Shabalina SA, Ogurtsov AY, Spiridonov NA (2006) A periodic pattern of mRNA secondary structure created by the genetic code. *Nucleic Acids Res* 34(8):2428–37.
7. Choudhury R, Tsai YS, Dominguez D, Wang Y, Wang Z (2012) Engineering RNA endonucleases with customized sequence specificities. *Nat Commun* 3:1147.
8. Lam SS, et al. (2014) Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat Methods* 12(1):51–54.
9. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31(13):3406–3415.
10. Ozawa T, Natori Y, Sato M, Umezawa Y (2007) Imaging dynamics of endogenous mitochondrial RNA in single living cells. *Nat Methods* 4(5):413–419.
11. Cheong C-G, Hall TMT (2006) Engineering RNA sequence specificity of Pumilio repeats. *Proc Natl Acad Sci U S A* 103(37):13635–13639.
12. Filipovska A, Razif MFM, Nygård KKA, Rackham O (2011) A universal code for RNA recognition by PUF proteins. *Nat Chem Biol* 7(7):425–427.
13. Dong S, et al. (2011) Specific and modular binding code for cytosine recognition in Pumilio/FBF (PUF) RNA-binding domains. *J Biol Chem* 286(30):26732–26742.
14. Yamada T, Yoshimura H, Inaguma A, Ozawa T (2011) Visualization of nonengineered single mRNAs in living cells using genetically encoded fluorescent probes. *Anal Chem* 83(14):5708–5714.
15. Wang Y, Cheong C-G, Hall TMT, Wang Z (2009) Engineering splicing factors with designed specificities. *Nat Methods* 6(11):825–30.

16. Cao J, Arha M, Sudrik C, Schaffer D V., Kane RS (2014) Bidirectional regulation of mRNA translation in mammalian cells by using PUF domains. *Angew Chemie - Int Ed* 53(19):4900–4904.
17. Tilsner J, et al. (2009) Live-cell imaging of viral RNA genomes using a Pumilio-based reporter. *Plant J* 57(4):758–770.
18. Yoshimura H, Inaguma A, Yamada T, Ozawa T (2012) Fluorescent probes for imaging endogenous beta-actin mRNA in living cells using fluorescent protein-tagged pumilio. *ACS Chem Biol* 7(6):999–1005.