

ARTICLE

DOI: 10.1038/s41467-017-02745-0

OPEN

# PTRE-seq reveals mechanism and interactions of RNA binding proteins and miRNAs

Kyle A. Cottrell<sup>1</sup>, Hemangi G. Chaudhari<sup>2,3</sup>, Barak A. Cohen<sup>2,3</sup> & Sergej Djuranovic<sup>1</sup>

RNA binding proteins (RBP) and microRNAs (miRNAs) often bind sequences in 3' untranslated regions (UTRs) of mRNAs, and regulate stability and translation efficiency. With the identification of numerous RBPs and miRNAs, there is an urgent need for new technologies to dissect the function of the *cis*-acting elements of RBPs and miRNAs. We describe post-transcriptional regulatory element sequencing (PTRE-seq), a massively parallel method for assaying the target sequences of miRNAs and RBPs. We use PTRE-seq to dissect sequence preferences and interactions between miRNAs and RBPs. The binding sites for these effector molecules influenced different aspects of the RNA lifecycle: RNA stability, translation efficiency, and translation initiation. In some cases, post-transcriptional control is modular, with different factors acting independently of each other, while in other cases factors show specific epistatic interactions. The throughput, flexibility, and reproducibility of PTRE-seq make it a valuable tool to study post-transcriptional regulation by 3'UTR elements.

<sup>1</sup>Department of Cell Biology and Physiology, School of Medicine, Washington University, St. Louis, MO 63110, USA. <sup>2</sup>Department of Genetics, School of Medicine, Washington University, St. Louis, MO 63110, USA. <sup>3</sup>The Edison Family Center for Genome Sciences and Systems Biology, School of Medicine, Washington University, St. Louis, MO 63110, USA. Correspondence and requests for materials should be addressed to S.D. (email: [sergej.djuranovic@wustl.edu](mailto:sergej.djuranovic@wustl.edu))

Cellular factors post-transcriptionally regulate mRNA by altering its modification, localization, stability, and translation<sup>1</sup>. These *trans*-acting factors often bind to *cis* elements within the mRNA. Two important classes of *trans*-acting factors are RNA binding proteins (RBPs) and microRNAs (miRNAs).

miRNAs are short non-coding RNAs that mediate translational repression and destabilization of target mRNAs<sup>2–9</sup>. miRNAs recruit the Argonaute containing miRNA-induced silencing complex (miRISC) to specific mRNAs by base-pairing with complementary sequences within their 3'UTR<sup>2</sup>. Mammalian cells typically express many miRNAs, with the human genome currently thought to encode 2580 miRNAs<sup>10</sup>. Those miRNAs are predicted to target most human mRNAs<sup>11</sup>.

RBPs are a second prominent class of *trans*-acting factors that affect mRNAs through processes including: splicing, adenylation and deadenylation, degradation, localization, and translation<sup>1</sup>. Recent studies have sought to identify the complete set of RBPs in mammalian cells, and based on these studies the human genome contains >1000 RBPs, most of which have unknown functions<sup>12–15</sup>. Over 800 RBPs have been identified in cultured HeLa cells alone<sup>12</sup>. One well-characterized RBP is Pumilio, a member of the Puf family, which is conserved from yeast to humans and regulates translation and RNA-decay<sup>16–19</sup>. The RNA binding protein Smaug is also conserved from yeast (Vts1) to humans (SAM4A and SAM4B) and regulates both translation and RNA-decay<sup>20,21</sup>. Another well-known RBP family is the ELAVLs, homologs of the *Drosophila* embryonic lethal abnormal vision, *elav*<sup>22</sup>. These proteins bind AU-rich elements within mRNAs and either stabilize or destabilize mRNAs, as well as enhance or repress translation<sup>22</sup>. While the function and mechanism of action of some RBPs have been partially elucidated, for the majority of RBPs their functions remain unknown.

Most evidence for the function of RBPs, such as Pumilio and Smaug, has come from low-throughput experiments that study their targets during embryogenesis or from reporter experiments<sup>16,17,19–21</sup>. Given the large numbers of uncharacterized RBPs and miRNAs, we urgently need new approaches with higher throughput, which can be employed in diverse cell types and developmental stages.

Interactions between the miRISC and RBPs have been of great interest recently. With >2500 human miRNAs, that are predicted to target most mRNAs, and >1000 RBPs it is likely that many mRNAs are co-regulated by these factors<sup>10,12–15</sup>. Many RBP or miRNA binding sites have been shown to occur near predicted miRNA binding sites. In many cases these binding sites are immediately adjacent or even overlap<sup>23–28</sup>. Some RBPs cooperate with miRNAs in regulating the expression of specific genes. For example, Pumilio facilitates miRNA-mediated repression in both humans and *Drosophila*<sup>2,29,30</sup>. HuR, a RBP that binds AU-rich elements, can also modulate miRNA-mediated repression<sup>2,31,32</sup>. Understanding how mRNA *trans*-acting factors modulate the activity of one another is a major challenge. A tractable high-throughput approach would help unravel the interactions between different effectors of RNA regulation.

The widespread availability of high-throughput sequencing is powering the development of “omic” technologies to study miRNAs and RBPs. RNA-seq combined with ribosome profiling can reveal the effects of RBPs and miRNAs on target RNA expression and translation<sup>33–35</sup>. While these methods provide the throughput required to study the effects of miRNAs and RBPs across the genome, they do not provide the flexibility to construct and assay large numbers of reporters designed to dissect the effects of different combinations and affinities of RNA *cis*-regulatory elements.

In studies of transcriptional enhancers, Massively Parallel Reporter Gene Assays (MPRAs) are useful complements to

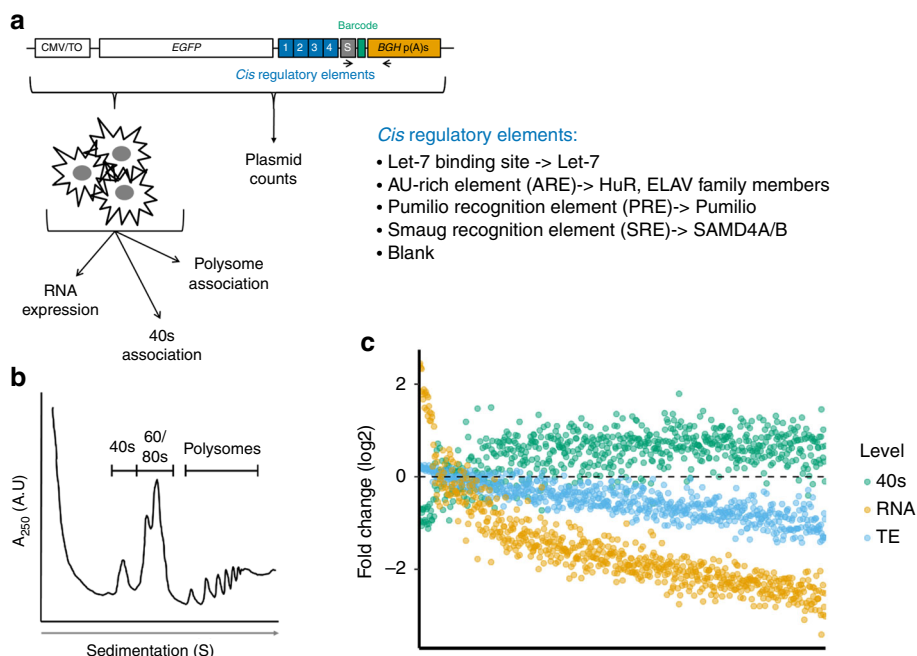
technologies that quantify the activity of endogenous genomic elements<sup>36–39</sup>. An analogous technology for assaying the activities of the *cis*-acting RNA sequences bound by RBPs and miRNAs would help unravel the network of interactions that underlies post-transcriptional regulation of mRNA. Such a system should provide the flexibility and throughput to dissect individual 3' UTR elements, assay the effects of changes in the strength and number of *cis*-acting RNA elements, and detect interactions between different types of *cis*-acting sequences.

Recently, several labs have employed plasmid or mRNA libraries to study endogenous 3'UTR elements<sup>40–43</sup>. These approaches generally rely on synthesizing or amplifying portions of 3'UTRs and fusing them to a reporter. While these techniques have identified 3'UTR motifs that have effects on RNA stability and protein amounts, none have been combined with polysome profiling to separate effects on RNA stability, translation efficiency, and translational initiation. In addition, naturally occurring 3' UTRs contain many different types of elements, making it difficult to deconvolve the effects of individual sites. A synthetic approach, in which large numbers of reporters with specific combinations of elements are designed and assayed, would provide the power necessary to isolate the effects of individual binding sites, as well as the interactions between sites. Because high-throughput methods for studying synthetic elements have proven to have great utility in dissecting interactions among transcription factors, we have extended this approach to post-transcriptional regulation<sup>44–47</sup>.

Here we report post-transcriptional regulatory element sequencing (PTRE-seq), an approach that uses a massively parallel reporter library to study the effects of synthetic 3'UTR elements on RNA stability, translation efficiency, and translation initiation. We use PTRE-seq to study the effects of known binding sites for RBPs and miRNAs, both individually and in combination. With this approach, we determine that the binding sites for these effector molecules influenced different aspects of the RNA lifecycle, including RNA stability, translation efficiency, and translation initiation. We observe *trans*-acting factors acting independently or in some cases epistatically. Finally, deploying PTRE-seq across multiple cell lines revealed the influence of the *trans* environment on post-transcriptional regulation by specific *trans*-acting factors. Altogether these results demonstrate the throughput, flexibility and reproducibility of PTRE-seq.

## Results

**Design and application of PTRE-seq.** We developed PTRE-seq to quantify the individual and combined effects of RBP and miRNA binding sites in 3'UTRs. We created 642 unique synthetic 3'UTRs composed of combinations of a let-7 binding site, the Pumilio recognition element (PRE), the Smaug recognition element (SRE), AU-rich elements (ARE), and a control sequence (“blank”). Bioinformatic analyses indicate that nearly 300 human transcripts contain a PRE, miRNA-binding site and an ARE (Supplementary Fig. 1). Over 1900 transcripts contain a PRE and an ARE, 698 contain an ARE and a miRNA-binding site and 653 contain a PRE and a miRNA-binding site. Between 13 and 15% of each of these pairs of regulatory elements occur within 150 nt of each other, with many overlapping or immediately adjacent (Supplementary Fig. 1). We arranged the regulatory elements in four positions within the 3'UTR (Fig. 1a), resulting in 200 bp long regulatory element section. The library included all possible combinations of the five elements within the four positions to generate the 625-unique synthetic 3'UTRs. The remaining seventeen synthetic 3'UTRs in our library contained variants of let-7 binding sites. Every unique synthetic 3'UTR is present 10 times in the library, each time associated with a different



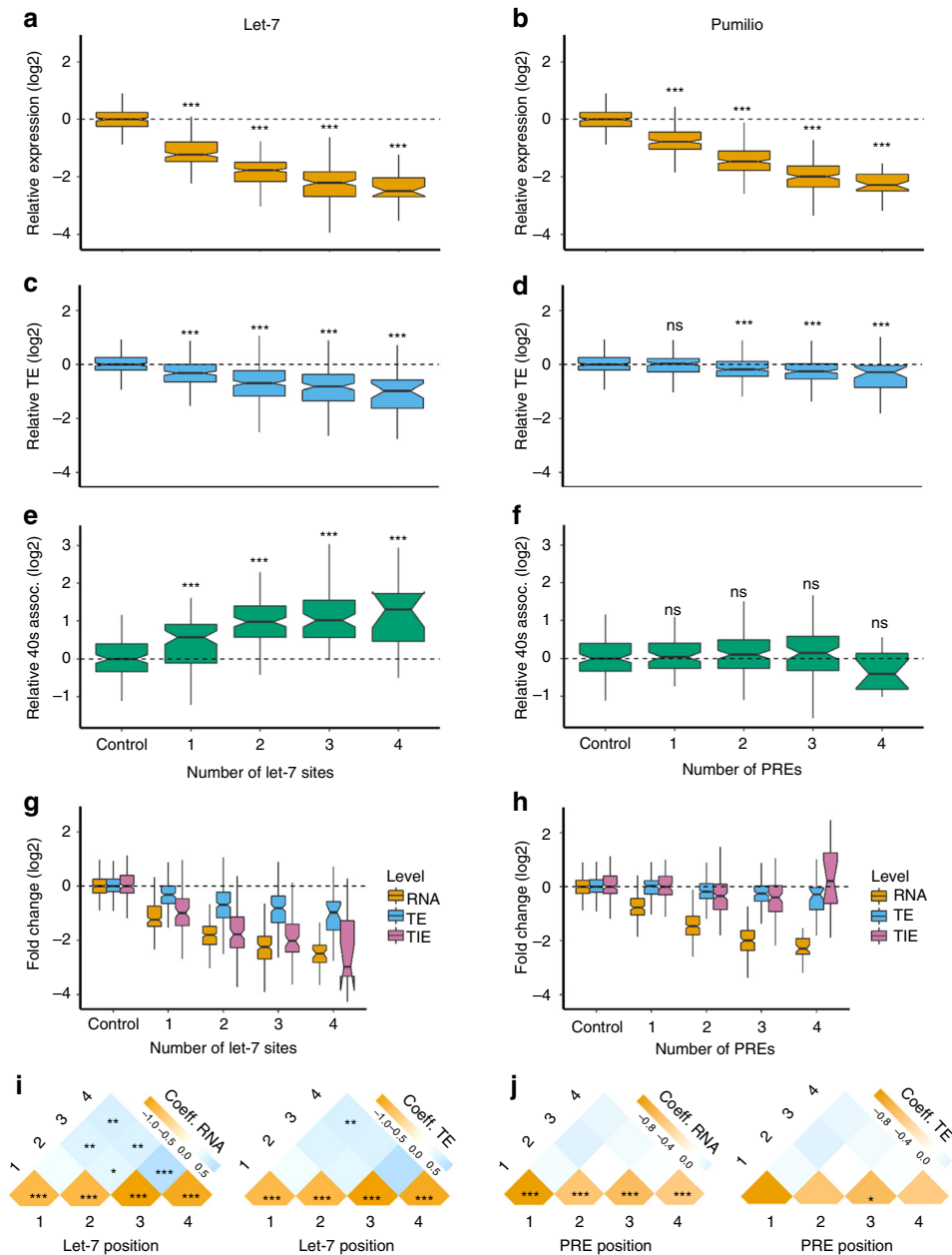
**Fig. 1** Design and application of PTRE-seq. **a** Schematic of the PTRE-seq library. Each *cis*-regulatory element (RE) within the library is inserted into an episomal reporter as shown. CMV/TO, cytomegalovirus promoter with the 5'UTR from the vector pCDNA5/FRT/TO. EGFP, enhanced green fluorescent protein. S, spacer sequence. BGH p(A)s, 3'UTR and polyadenylation signal from bovine growth hormone gene. Each unique synthetic 3'UTR, made up of binding sites for the REs shown, is represented by 10 barcodes. **b** Representative polysome profiling trace. mRNA was isolated from 40S, and polysome fractions. **c** Fold change of mRNA levels, translation efficiency, and 40S association for all reporters within the library. The reporters are arranged along the x-axis in decreasing order based on fold change

co-transcribed barcode. These provide replicate measurements when the barcodes are used to quantify the relative abundance of reporter mRNAs in total or ribosome associated fractions from transfected cells. Barcoded synthetic 3' UTR sequences were cloned downstream of a CMV promoter driven reporter gene to create a plasmid library.

We transfected HeLa cells with the library and collected the cells after 40 h. We isolated total RNA from a portion of the cells, and the remaining cells were lysed for polysome profiling to assay translational regulation. We collected mRNAs associated with the polysome fractions (translating ribosomes) and the 40S ribosome fractions (initiating ribosomes) (Fig. 1b and Supplementary Fig. 2). Messenger RNAs associated with the polysomal fractions are considered efficiently translated<sup>48</sup>. Since regulation of gene expression by various *cis* and *trans*-acting factors during mRNA translation often targets the translation initiation step, we separately analyzed the 40S fraction<sup>49</sup>. The 40S fraction contains mRNAs that are bound only by the small subunit of the ribosome during the translation initiation steps. We generated cDNA from the total RNA, polysome, and 40S associated mRNA and sequenced the barcodes to determine the relative abundance of every reporter in the library, in each fraction. Counts for every barcode in cDNA were normalized by counts determined by sequencing the input plasmid library. The Pearson correlation between replicate experiments ranged between 0.975 and 0.983 for total RNA, 0.703 and 0.787 for polysome associated RNA, and was 0.926 for 40S associated RNA (Supplementary Fig. 2), which allowed us to make quantitative comparisons between different synthetic 3'UTRs. To compute translation efficiency (TE), a measure of the reduction in translation beyond what is expected due to a reduction in mRNA levels, we normalized the barcode counts for each 3'UTR in the polysome fraction to its counts in total RNA. The same was done for the 40S associated RNAs to compute 40S association, which represents a proxy for the

engagement of the translation initiation complex with mRNAs. In all cases, we determined the relative effect by normalizing to the control reporter, which contains four “blank” sequences in the synthetic 3'UTR. For most reporters, we observed both reduced RNA expression and reduced TE, which was concomitant with an increase in 40S association (Fig. 1c). Correlations can be seen between each of these metrics (Supplementary Fig. 2e–h). Summary statistics for PTRE-seq measurements of RNA expression and TE are shown in Supplementary Figure 3. We validated our PTRE-seq findings using quantitative PCR (qPCR) and fluorescence measurements of GFP for several individual reporters from the library (Supplementary Fig. 4). The data we have obtained using PTRE-seq reveal the ability of this method to capture evidence for post transcriptional regulation at different steps.

**Linear regression and thermodynamic modeling of results.** The *cis*-elements in our library had strong effects in the data which we captured by fitting linear regression models for both RNA expression and TE to our data. For both the RNA expression model and the TE model, parameters included the identity of the element at each of the four positions and all pairwise interactions between elements at each position. The regression models captured the relationship between 3'UTR composition and relative RNA expression (five-fold cross-validation, Pearson correlation 0.87–0.93) (Supplementary Fig. 5) and the relationship between 3' UTR composition and TE (five-fold cross-validation, Pearson correlation 0.89–0.92) (Supplementary Fig. 6). The model predicted well the effects of individual elements and combinations of elements on RNA expression and TE (Supplementary Fig. 7). Interestingly, the models accurately predicted the RNA expression and TE of reporters containing three or four different binding sites using only the individual effect of each binding sites

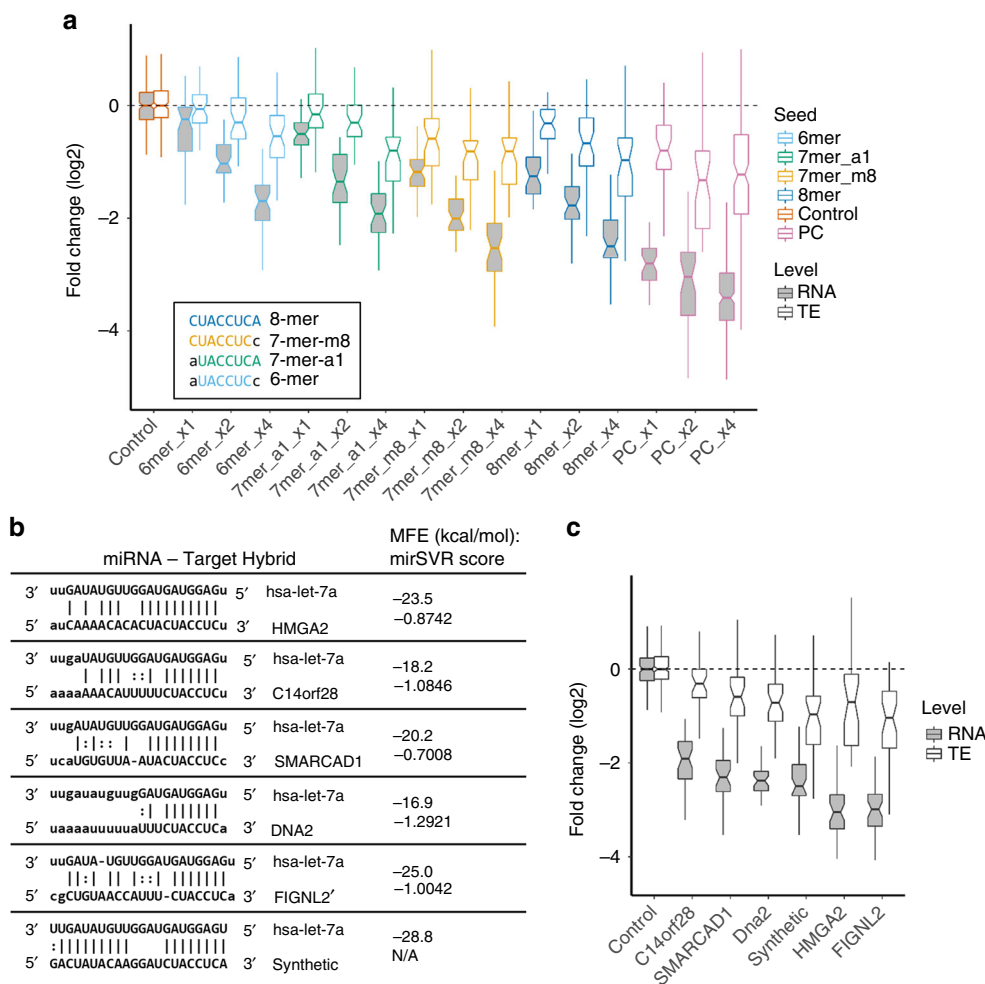


**Fig. 2** PTRE-seq reveals differences in the mechanism of repression by miRNAs and Pumilio. Fold change of RNA (a) TE (c), and 40S association (e) of let-7 binding site containing reporters within the PTRE-seq library. Fold change of RNA (b) TE (d), and 40S association (f), of PRE containing reporters within the PTRE-seq library. For **a–d** \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , *t*-test with Bonferroni correction. For panels **a–f** the results for all constructs containing one, two, three, or four sites is shown. The data for each site in positions one-four are shown in Supplementary Fig. 9. Panels **g** and **h** show composite boxplots with fold change of RNA, TE, and translation initiation efficiency (TIE) for let-7 and PRE respectively. TIE was calculated by normalizing polysome associated RNA/40S associated RNA. **i** The regression coefficients for linear models with parameters corresponding to let-7 alone or in combination with other let-7 sites at positions 1–4, or **j**, PREs alone or in combination with PREs at positions 1–4. In **i** and **j**, the left panels show the coefficients for RNA while the right panels show the coefficients for TE. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , *t*-test. Boxplot whiskers indicate the furthest datum that is 1.5\*Q1 (upper) or 1.5\*Q3 (lower). For clarity, outliers have been removed from boxplots but were used for statistical analysis

and pairwise interactions. Models fit with higher order interaction terms failed during cross-validation. This result, combined with the observation that models with individual effects and pairwise interactions perform well, suggests that higher-order interactions have, at most, only minimal effects on RNA expression and TE.

To gain mechanistic insights, we also fit a statistical thermodynamic model to our RNA expression and TE data<sup>50,51</sup>. Due to the position-dependent nature of ARE elements (described below), we excluded synthetic 3'UTRs containing ARE elements from this analysis. This model provides a formal biophysical

framework to capture saturation effects and cooperative interactions between *cis*-acting elements. Each 3'UTR is described as a collection of states, in which each state represents a particular configuration of bound and unbound elements on a 3'UTR. The model uses parameters that describe the free energies of interaction between RBP/miRNA–RBP/miRNA and RBP/miRNA–mRNA to compute the probability, or weight of each state. These interactions can be neighboring or non-neighboring, however, our implementation of the model does not explicitly model position of the RBPs. In each state bound factors either



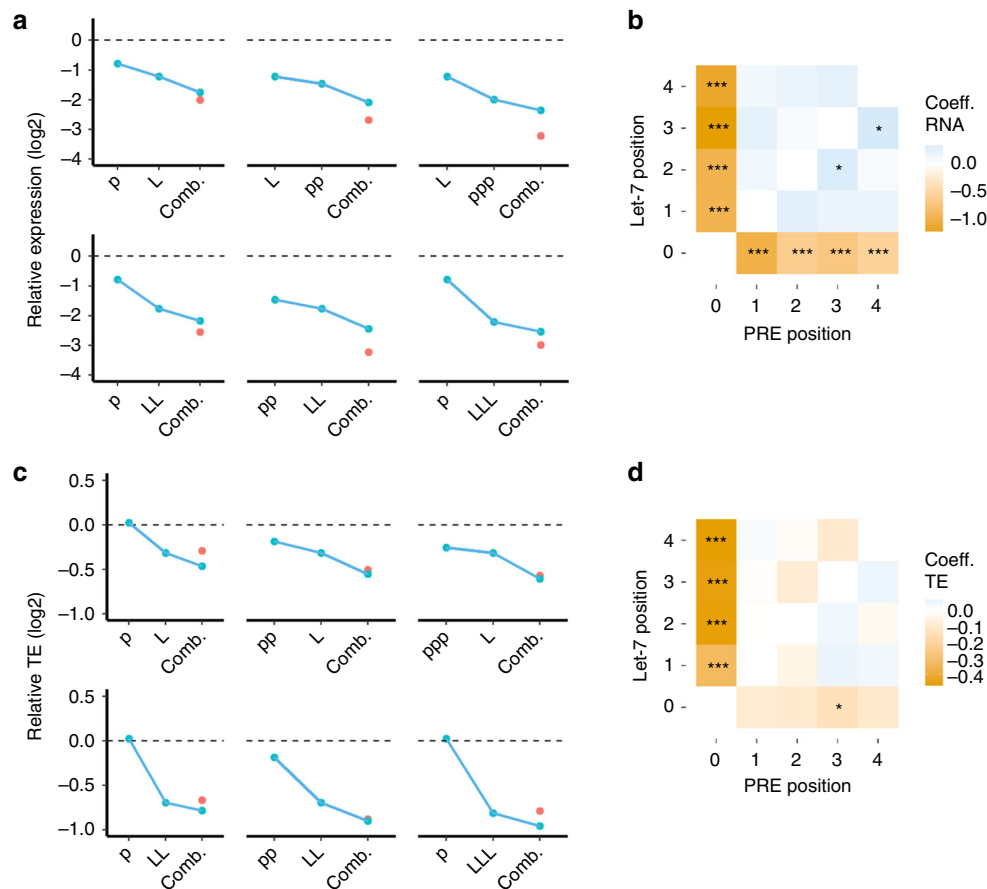
**Fig. 3** PTRE-seq reveals the effect of the let-7 binding site on repression. **a** Comparison of the fold change of reporters containing synthetic let-7 binding sites with altered seed binding. Also shown are reporters containing let-7 binding sites that have perfect complement (PC) binding to let-7. Each seed binding variant is present in either one, two or four copies. The inset describes the seed binding region of each seed-binding variant site. **b** Table describing the natural and synthetic let-7 binding sites used in this study. MFE, minimal free energy<sup>71</sup>. mirSVR, mirSVR score<sup>72</sup>. **c** Fold change of RNA and TE for reporters containing four copies each of natural or synthetic let-7 binding sites. Boxplot whiskers indicate the furthest datum that is 1.5\*Q1 (upper) or 1.5\*Q3 (lower). For clarity, outliers have been removed from boxplots but were used for statistical analysis

facilitate or inhibit the recruitment of mRNA decay machinery (or the ribosome for TE), and the weights of the different states are used to compute the probability that the mRNA decay machinery (or the ribosome for TE) is present at an mRNA. In the model, this probability is proportional to the output RNA expression or TE<sup>46,50</sup>.

A thermodynamic model with four independent parameters, one each for the interaction of the decay machinery with either let-7, PRE, SRE, or the “blank” site, predicted observed TE ( $R = 0.92$ ) and RNA expression well ( $R = 0.94$ ). The good performance of these models suggests that let-7, PRE, and SRE function mostly independently on UTRs. In most cases adding interaction terms did not improve the fit of these models to the data. This observation suggests that some of the self-interaction terms in the linear regression models (described below) are likely due to saturation of binding on UTRs with high copy numbers of *cis*-acting sites. The thermodynamic model naturally accounts for saturation without the need for interaction terms and describes the situation when saturation causes additional sites to have little or no effect. In two cases, the thermodynamic model for TE did improve with the addition of interaction terms, one for interaction between adjacent let-7 sites and another for interaction between adjacent PRE and let-7 sites ( $R = 0.93$ ,

Supplementary Fig. 8), which suggests epistatic interactions between these elements that cannot be accounted for by binding site saturation. The thermodynamic model for RNA expression also improved with the addition of five interaction terms ( $R = 0.94$ , Supplementary Fig. 9). We sought to identify the trends in our data that underlie the strong performance of these models.

**PTRE-seq reveals mechanisms of *trans*-acting factors.** For each RNA element in our library there are a series of constructs that contain only that element and the control ‘blank’ sequence. This allowed us to study the individual and copy-number-dependent effect of each RNA element. For the let-7 binding site we observed a reduction in both relative RNA expression and TE (Fig. 2a, c and Supplementary Fig. 10a, c). This suggests that not only is the abundance of the RNA reduced by the addition of let-7 sites, but also that the remaining RNAs are translated poorly relative to the control message. Both effects were dependent on the number of let-7 binding sites in the synthetic 3’UTR and the effects appear to saturate with additional sites (Fig. 2a). While our linear regression model captures well the effects of individual let-7 sites, it is easily influenced by saturation effects and thus cannot distinguish



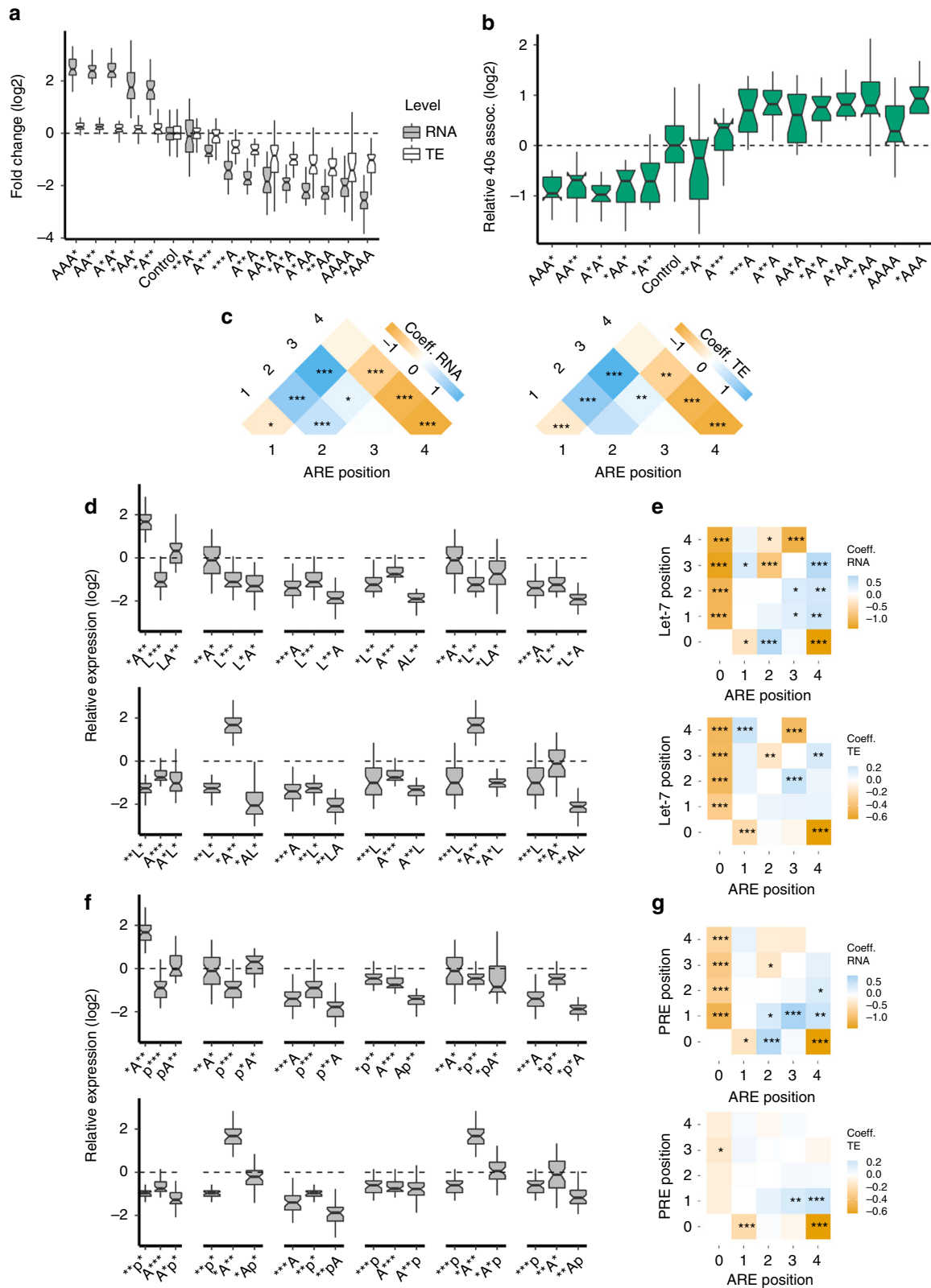
**Fig. 4** Pumilio and miRNAs function independently. **a** The effect of a let-7, PRE or a combination of the two elements on relative expression, and **c** relative TE. The median relative expression or TE is plotted across all barcodes and replicates. Red dot, the product of each individual effect, the expected result assuming independence. The regression coefficients from the linear regression model for RNA expression (**b**), and TE (**d**), for the parameters corresponding to let-7 or PREs alone or interactions between positions containing let-7 or PREs. For **a** and **c** L = let-7 binding site, p = PRE, \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , *t*-test

between saturation effects and true epistatic interactions (Fig. 2i). To counter this we employed our thermodynamic model. The thermodynamic model requires an interaction term between let-7 binding sites that stabilizes RNA for a good fit (Supplementary Fig. 9). Since the thermodynamic model is robust to saturation effects, this interaction term suggests epistatic antagonism between let-7 binding sites.

The Pumilio recognition element (PRE) behaved differently than the let-7 binding site. Like the let-7 binding site, PREs also decreased RNA expression, but had a much more modest effect on TE (Fig. 2b, d and Supplementary Fig. 11b, d). Additional PREs appeared to show a modest saturation effect on RNA expression (Fig. 2b). However, our thermodynamic model for RNA or TE revealed no significant epistatic interactions between PREs and therefore we have excluded this term from the model. While both the let-7 binding site and PREs reduced TE, only the let-7 binding site affected the association with the 40S ribosomal subunit (Fig. 2e, f). Let-7 binding sites increased 40S association in a manner dependent on the number of sites. This could be caused by slowed scanning of the 5'UTR or recruitment of the large subunit of the ribosome. These data show a clear distinction between the mechanisms of post-transcriptional regulation by Pumilio and the miRISC. These differences are illustrated by normalizing polysome associated RNA counts by 40S associated RNA counts, which produces a value we term translation initiation efficiency (TIE). TIE is reduced substantially, and in a copy-number-dependent manner by let-7 and only modestly by

Pumilio (Fig. 2g, h). Taken together our results suggest that Pumilio works mostly by destabilizing its target mRNAs, while the miRISC functions both by affecting mRNA stability and by inhibiting translation initiation. This mechanistic difference was clearly captured by PTRE-seq.

**PTRE-seq unveils the effect of miRNA-target base-pairing.** The efficiency of miRNA-mediated repression depends on the number and quality of binding sites in its target<sup>34,35,52,53</sup>. Nucleotides 2–7 of the miRNA constitute the “seed” sequence, and weak seed pairing reduces the effectiveness of miRNA-mediated repression<sup>34,35,52,53</sup>. In addition to the 625 combinations of regulatory elements described above, we included constructs in the library to study the effect of base-pairing between the miRNA and its target on repression. This included a series of constructs with one, two, or four binding sites for let-7 with either 6-mer, 7-mer-a1, 7-mer-m8 or 8-mer base-pairing in the seed region<sup>52</sup>, as well as binding sites for let-7 that have perfect base-pairing with the target. We observed a clear copy-number-dependent and seed-pairing dependent effect on RNA expression and TE for these reporters (Fig. 3a, Supplementary Figure 12). The repression at the level of RNA and TE was greatest for target sites with 8-mer or 7-mer-m8 pairing. A single copy of the perfect complement let-7 binding site was more effective at reducing RNA expression than four copies of the binding site with a mispairing bulge (Fig. 3a). In addition to studying the effect of seed-pairing alone,



**Fig. 5** AU-rich elements modulate repression by Pumilio and miRNAs. **a** The position of an ARE within the synthetic 3'UTR determines the relative TE or RNA expression. **b** The relative 40s association of ARE containing reporters. **c** Heatmap of the regression coefficients for the parameters corresponding to AREs alone. Left panel shows coefficients for RNA expression and the right panel shows coefficients for TE. **d** AREs modulate repression by miRNAs in a position-dependent manner. The green box highlights an example of stimulation of miRNA-mediated RNA destabilization by an ARE. **e** The regression coefficients for the parameters corresponding to let-7 or AREs alone or interactions between positions containing let-7 or ARE. **f** AREs modulate repression by PREs in a position dependent manner. **g** The regression coefficients for the parameters corresponding to PREs or AREs alone or interactions between positions containing PRE or ARE. For **a, b, d** and **f**, \* = Blank, A = ARE, L = let-7 and p = PRE. For **c, e** and **g**, \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001, *t*-test. Boxplot whiskers indicate the furthest datum that is 1.5\*Q1 (upper) or 1.5\*Q3 (lower). For clarity, outliers have been removed from boxplots but were used for statistical analysis

we also studied the effect of endogenous let-7 binding sites. For this we made constructs containing four copies of a let-7 binding site from the 3'UTR for *HMG2*, *SMARCA1*, *DNA2*, *C14orf28* and *FIGNL2*. While, the synthetic binding site is predicted to have the most favorable binding (Fig. 3b), the sequences from two of the natural 3'UTRs (*HMG2* and *FIGNL2*) reduced RNA expression to a greater extent (Fig. 3c). We suspect that secondary structure around the let-7 binding sites in these reporters is contributing to let-7 binding. This can be seen by making a simple linear regression model for fold change of RNA expression with base-pairing minimal free energy (MFE) and 3'UTR secondary structure MFE as parameters. A model that includes each parameter and an interaction term gave a better fit ( $R = 0.81$ ) than base-pairing MFE ( $R = 0.56$ ) or secondary structure MFE ( $R = 0.12$ ) alone. Because this model was made with only a few data points it is only suggestive. This secondary structure of the 3'UTR could explain the observation that some binding sites, even with better thermodynamics, were not as well repressed.

**Pumilio does not enhance miRISC function.** Enhancement of miRNA-mediated repression by the RBP Pumilio has been observed for a handful of targeted mRNAs<sup>29,30</sup>. In our library, the combination of a let-7 binding site and a PRE resulted in a reduction of RNA expression that was slightly less than the product of their individual effects (Fig. 4a). This was true for every combination we tested. The coefficients from our linear regression model for RNA levels are positive for all combinations of PRE and let-7, while the coefficients for each alone is negative (Fig. 4b). For TE we observed a modest enhancement of repression for some combinations and no effect for others (Fig. 4c), this was captured by our linear regression model which showed a mix of positive and negative coefficients for the combinations of PRE and let-7 (Fig. 4d). The pairwise arrangement of let-7 binding sites and PREs had no effect on repression (Supplementary Fig. 12). These data suggest a slight antagonism between the two elements in regard to their effects on RNA stability, and is reminiscent of the saturation we observed with additional let-7 or Pumilio binding sites. The thermodynamic model for TE includes a statistically non-significant anti-cooperative interaction between let-7 and PRE sites, while the model for RNA decay includes anti-cooperative interaction terms for a subset of let-7 and PRE binding site combinations (Supplementary Fig. 8 and 9). Thus, the miRISC and Pumilio function independently in most UTRs and reduced repression seen with combinations of sites is mostly because of saturation effects. As miRNAs and Pumilio are thought to promote mRNA decay using the same pathway it is not surprising that when both are bound to the same message there is no enhanced degradation<sup>6-8,16,17</sup>.

**Position dependent effects of AU-rich elements.** Several RBPs can bind AU-rich elements. ARE binding proteins, such as HuR (ELAVL1), can either stabilize or destabilize target mRNAs, and can enhance or repress translation<sup>22</sup>. Other RBPs, such as tristetraprolin (TTP) and AUF1 (hnRNPD), are also ARE-binding proteins (ARE-BP)<sup>22</sup>. In our library AREs either enhanced or repressed RNA expression and TE depending on their position in the 3'UTR (Fig. 5a). In the first or fourth position in our synthetic 3'UTR, ARE reduces both RNA expression and TE, while an ARE in the second position increases RNA expression and TE, and an ARE in the third position has no effect on either metric. The combination of multiple AREs altered this position-dependent effect. Generally, any combination with an ARE in position four had reduced RNA and TE while all other combinations had increased RNA and TE. We observed similar effects on 40S association where any combination of ARE that reduced TE

resulted in increased 40S association and vice versa for those that increased TE (Fig. 5b). These observations were captured by our linear regression model for RNA expression, which showed an ARE at position four to have a negative coefficient and an ARE at position two to have a positive coefficient (Fig. 5c). Since the linear regression model cannot distinguish between saturation effects and epistatic interactions it is difficult to assign a cause. However, the results clearly show instances where AREs in specific arrangements lead to increased or decreased RNA expression and/or TE.

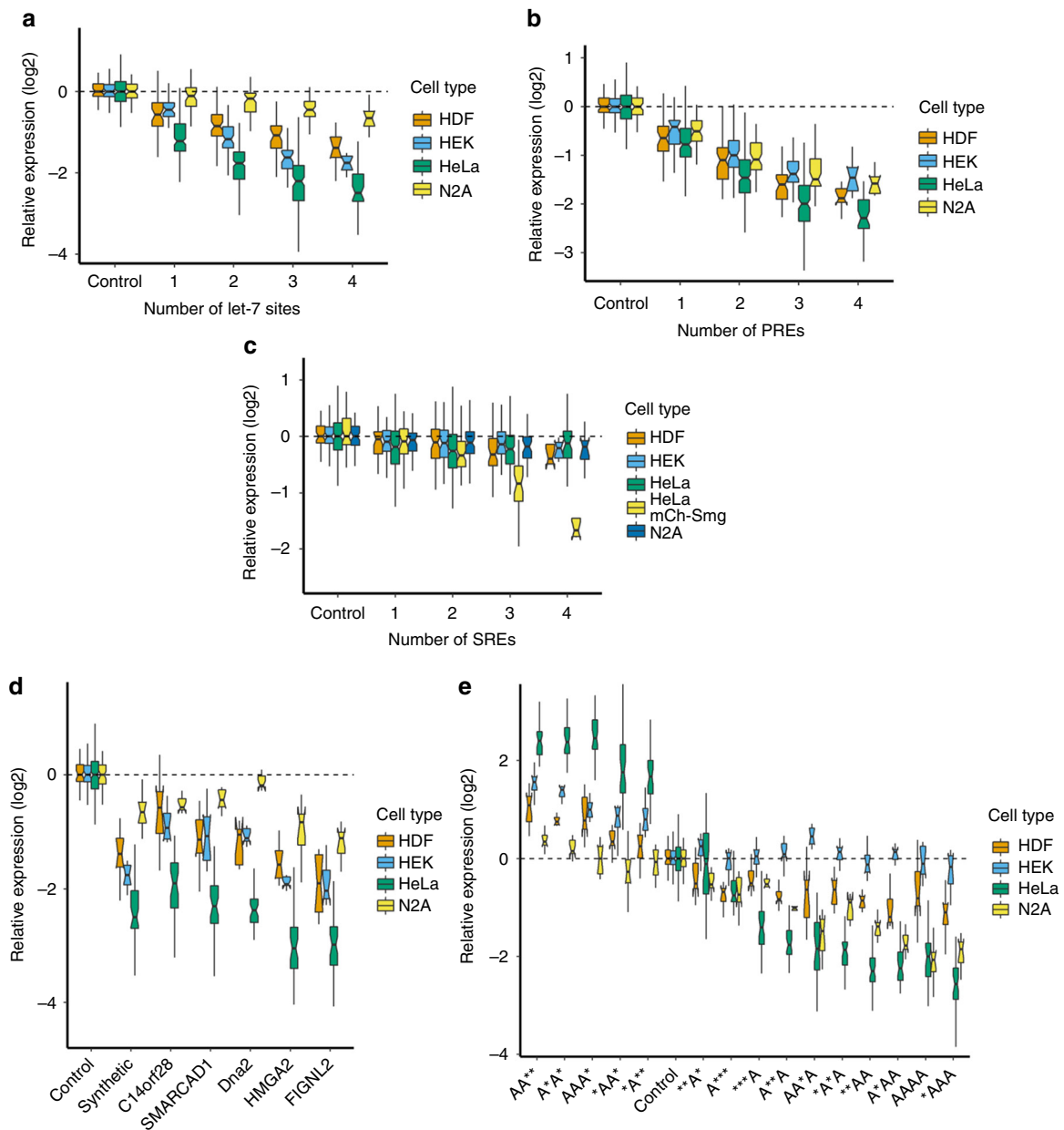
#### AU-rich elements modulate activity of miRNAs and Pumilio.

The AU-rich element binding protein HuR has been shown to both activate and inhibit miRNA-mediated repression<sup>2</sup>. During recovery from stress HuR relieves miRNA-mediated repression of the catalase mRNA (*Cat1*)<sup>32</sup>. In contrast, HuR binding to the *c-Myc* 3'UTR activates miRNA-mediated repression<sup>31</sup>. We observed AREs in our library either enhancing or suppressing miRNA-mediated repression in a position dependent manner (Fig. 5d and Supplementary Fig. 13). For example, a let-7 binding site at position three reduces RNA expression while an ARE at position two increases RNA expression, but the combination of ARE and let-7 reduces RNA expression more than the let-7 binding site alone. Our linear regression likely captured some of these effects but because it cannot distinguish between saturation effects and epistatic interactions we cannot assign a cause (Fig. 5e). However, it is clear that in some cases specific combinations of let-7 binding sites and AREs resulted in obvious changes to RNA expression or TE, for instance changing from a message that is stable to one that is unstable, see above example. We observed similar position-dependent modulation of Pumilio activity by AREs (Fig. 5f and g). Together these data show that AU-rich element binding proteins can modulate the repression by the miRISC and Pumilio in a position-dependent manner, even though, as shown above, Pumilio and let-7 utilize different mechanisms to repress RNA expression and TE.

#### Post-transcriptional regulation varies across cell types.

In addition to HeLa cells, we also transfected our PTRE-seq library into three other cell types: human embryonic kidney (HEK293), human neonatal dermal fibroblast (HDF), and a mouse neuroblastoma (N2A). We observed wide variation in the effect of each regulatory element tested across the four cell lines. It is possible that some of these changes could be caused by differences in transfection efficiency or transcription across the cell lines tested. The let-7 binding site caused robust reduction of RNA expression in HeLa cells, but this effect was smaller in magnitude in HEK293, HDF and N2A (Fig. 6a, d and Supplementary Fig. 14). Neuroblastoma cells are thought to have very little expression of let-7<sup>54</sup>. In contrast, we observed modest variations in the magnitude of repression by PREs across the four cell lines (Fig. 6b). PREs were most effective in HeLa and least effective in N2A or HEK293 cells. Interestingly, we observed only a very modest reduction in RNA expression and no effect on TE (Supplementary Fig. 15) for reporters containing SREs across all cell-lines. Only when we overexpressed the *Drosophila* homolog of SAMD4A and SAMD4B, Smaug (mCh-Smg) did we see a substantial reduction in RNA expression (Fig. 6c). For AREs, the cell type not only altered the magnitude of the effect but could also abrogate the effect entirely (Fig. 6e). For example, in HeLa the AREs could both reduce or increase RNA expression, while in HEK293 we only observed increased RNA expression by AREs. Conversely, in N2A we observed robust reductions in RNA expression by AREs but very modest increases in RNA expression. As AREs are known to be bound by multiple RBPs this





**Fig. 6** The regulatory capacity of miRNAs and AU-rich elements vary across cell types The relative expression of reporters containing let-7 binding sites (a), PREs (b), SREs (c), natural binding sites for let-7 (d), or AREs (e). In panel c HeLa-mCh-Smg refers to HeLa cells that were cotransfected with the PTRE-seq library and a plasmid for expression of mCherry-Smaug. HDF, neonatal human dermal fibroblasts. HEK, human embryonic kidney. N2A, mouse neuro2A. For e \* =Blank and A=ARE. Boxplot whiskers indicate the furthest datum that is 1.5\*Q1 (upper) or 1.5\*Q3 (lower). For clarity, outliers have been removed from boxplots but were used for statistical analysis

finding suggests the presence of a different profile of active ARE-binding RBPs in each cell type.

**Discussion**

To better understand the post-transcriptional regulation of mRNAs, we must determine how regulatory factors function both independently and in combination with each other. Towards this end we developed PTRE-seq, a powerful new high-throughput tool for interrogating the additive and combined effects of binding sites for RBPs and miRNAs on RNA stability and translation. As PTRE-seq is extended to additional RBPs and miRNAs, we will better understand the network of molecular interactions that comprise post-transcriptional regulatory systems.

Using PTRE-seq we observed decreased RNA levels and decreased association with polysomes mediated by the let-7 miRNA. By fractionating the cell lysates before analysis, we determined that the reduction in polysome associated RNA was more than could be accounted for by the decrease in RNA levels alone, indicating that let-7 reduces RNA levels and reduces the efficiency with which the remaining RNA is translated. This decrease in translational efficiency also correlated with an increase in 40S association of mRNAs targeted by the miRNA let-7. These results are consistent with a proposed model in which the miRISC inhibits translation initiation at the scanning step by induced dissociation of the helicase subunit eIF4A of the eIF4F complex<sup>3-5,9,55</sup>. The reduced rate of scanning increases the time that the 40S ribosome is bound to the message prior to identification of the start codon and recruitment of the 60S ribosome.

This delay in subunit joining would increase the time mRNAs spend bound by the 40S ribosome while reducing translation efficiency. The ability of PTRE-seq to separate effects on RNA levels from effects on different steps of translation is an important advantage of this method.

Beyond identifying the mechanism of miRNA-mediated repression another major challenge in the field remains defining binding sites with gene regulatory activity of the thousands of miRNAs in the cell<sup>56–60</sup>. Using PTRE-seq we were able to study the efficacy of target sites for let-7. Our results are consistent with previous studies: miRNA efficacy depends on thermodynamics of binding and 3'UTR structure<sup>34,35,52</sup>. We observed stronger repression for messages containing more base-pairing within the seed sequence: 8-mer >7-mer >6-mer. This finding was consistent with previous studies of endogenous miRNA targets<sup>34,35</sup>. Furthermore, reporters containing four copies each of endogenous let-7 binding sites showed variable repression that was not always dependent on thermodynamics of base-pairing. A simple linear regression model revealed that the secondary structure around the let-7 binding sites contributed to the magnitude of repression. This finding is consistent with a model for miRNA target prediction which incorporates the thermodynamics of miRNA binding and secondary structure near the binding site<sup>56</sup>. This type of analysis could be used for other miRNAs to empirically define their binding sites with largest impact on gene regulation.

In contrast to let-7, Pumilio decreased RNA levels with only very modest effects on polysome association and no effect on 40S subunit binding. Our results are consistent with the findings that Pumilio and miRNAs inhibit translation at different steps<sup>3–5,9,16,19</sup>. It is also possible that the differences we observed between let-7 binding sites and PREs could reflect differences in the kinetics of repression by the miRISC or Pumilio. Besides their individual roles in regulation of gene expression, miRNAs and Pumilio have been shown to function together. In some cases, Pumilio can activate miRNA-mediated repression of specific mRNAs<sup>29,30</sup>. However, in our experiments the effects of let-7 and Pumilio were largely independent. Pumilio may only activate particular miRNA targets by opening certain secondary structures and enabling miRNA binding<sup>30</sup>. To test this model PTRE-seq could be performed on synthetic messages carrying miRNA binding sites and PREs in the context of varying secondary structures.

The effects of let-7 or Pumilio sites showed almost no dependency on their position in the 3'UTR. In contrast, we observed a strong positional effect for AREs. AREs are bound by several RBPs including the ELAVLs, Auf1, and TTP, and different ARE binding proteins can stabilize or destabilize mRNA targets, as well as repress or enhance translation. The dependency of ARE on position could be explained if different ARE binding proteins are binding at different positions in the 3'UTR. Although the sequence of the ARE is the same at each position, the flanking sequence context is different, and RNA secondary structure may vary based on the position of the ARE (Supplementary Fig. 16). This altered structure might affect which ARE-BP bind to the sequence. The varied effects of AREs across cell-lines are consistent with this hypothesis. While AREs both increased and decreased RNA expression in most cell lines tested, in HEK293 we only observed increased RNA expression. As the expression of ARE-BPs is known to vary across cell and tissue types, this finding suggests that the ARE-BPs with different effects are binding to the same reporters in different cell-lines<sup>22</sup>. In any given cell-line, the cumulative effect of multiple ARE-BPs determines the overall activity of AREs.

Our experiments revealed strong epistatic interactions between AREs and sites for let-7 and Pumilio. The AREs either enhanced or suppressed miRNA- and Pumilio-mediated repression,

depending on their position in the 3'UTR. These effects were not dependent on the proximity in linear sequence space between the two binding sites. These results demonstrate that AREs can modulate repression by both miRISC and RBPs such as Pumilio. Alternatively, the presence of the PRE or miRNA binding site may modulate the effect of the ARE by changing the secondary structure of the mRNA. These observations warrant future studies into which ARE-BP are responsible for these epistatic interactions, and whether the effects are mediated through mRNA secondary structure or potentially through interactions between *trans*-acting factors. An intriguing possibility is that the ARE may be bound by cytoplasmic polyadenylation element binding protein, CPEB. The consensus binding site for CPEB is UUUUUAU<sup>61</sup>, but CPEB also binds the sequence UUUUUAU<sup>62</sup>, which appears once in the ARE used in our library. CPEB has been previously shown to work with Pumilio in regulating mRNA translation<sup>63</sup>.

While we observed robust effects on the RNA expression and translation of reporters containing let-7 binding sites, PREs and AREs, in our experiments, the SRE caused only modest changes in RNA expression and no change in TE or 40S association. When we overexpressed *Drosophila* Smaug in HeLa cells, we observed a reduction in the RNA levels of SRE containing reporters. This suggests that at least in the cell-lines we tested the mammalian Smaug homologs, *SAMD4A* and *SAMD4B*, are expressed at low levels or are not efficacious.

Our results provide further evidence for the mechanisms of post-transcriptional regulation by miRNAs, Pumilio and AREs. PTRE-seq will serve as a valuable tool studying the effects of multiple *cis*-acting elements, both individually and in combination, and for unraveling their effects on different aspects of RNA stability and translational control.

## Methods

**Construction of library.** To create the PTRE-seq library we first generated the plasmid pCDNA5/FRT/TO-EGFP-RE. EGFP was PCR amplified with EGFP-F and RE-R primers, Supplementary Table 1. This appended a single *NheI*, *EcoRV*, and *KpnI* sites downstream of EGFP. The PCR product was ligated into pCDNA5/FRT/TO that had been previously cut with *PmeI*.

A pool of 6500 unique 200-mer oligonucleotides was ordered from Agilent Technologies™. Oligonucleotides were designed to contain all combinations of either a let-7 binding site, PRE, SRE, AU-rich element or a “blank” control sequence. The sequence for each of these elements is described in Supplementary Table 2. Each of these unique combinations was synthesized with 10 different 9 bp barcodes. This provided a total of 6250 oligonucleotides. The remaining oligonucleotides consisted of 40 additional copies of the control sequence (4 place holders, “blanks”), 50 copies of a low expression control (4×let-7 perfect complement) and a series of constructs containing natural or synthetic let-7 sites. In total, the library consisted of 642 unique ‘synthetic 3'UTRs’ each with 10 unique barcodes, except for the controls described above. The sequences of each of these “synthetic 3'UTRs” are in the Supplementary Data 1. Each oligo has a 5' and 3' priming region which are identical across all oligonucleotides. The oligonucleotides also contained a restriction enzyme sites for subsequent cloning. A generic oligonucleotide appears as follows: 5' – GTAGCATCTGTCCGCTAGC-132nt regulatory element-ATGCATcGATATCaCTCGAGxxxxxxxxGGTACC CGACTACTACTACG – 3'. The restriction enzymes are underlined and are from 5' to 3': *NheI*, *NsiI*, *EcoRV*, *XhoI*, and *KpnI*.

The library was PCR amplified for four cycles using Phusion High-Fidelity polymerase (NEB) and primers Lib\_F and Lib\_R. We cloned the amplicon into pCDNA5/FRT/TO-EGFP-RE using *NheI* and *KpnI*. We prepared plasmid DNA from ~40,000 colonies to generate library RE\_Array\*. We then cloned a “spacer” sequence in between the regulatory elements and the barcode. This “spacer” is the reverse of a sequence within the BGH 3'UTR and is used for amplification of the barcodes from cDNA or plasmid. The “spacer” was ordered as a pair of oligonucleotides that were annealed to form a dsDNA oligonucleotide with overhangs compatible with DNA cleaved by *NsiI* and *XhoI*. The “spacer” was cloned into the RE\_Array\* using *NsiI* and *XhoI*. We collected plasmid DNA from ~250,000 colonies to generate the library RE\_Array\_1.

To clone individual reporters from the library we sequenced 94 colonies from the RE\_Array\* library. We chose from those clones seven reporters of interest. For the control reporter and three reporters targeted by let-7 (\*7\*, 7777, and 7pc-x2), we ordered oligonucleotides that were ligated into the vector pCDNA5/FRT/TO-

EGFP-RE as described above. The “spacer” was ligated into the plasmids containing the reporters, as described above.

**Cell culture and transfection.** HeLa (CCL-2.2, ATCC), HDFn (C0045C, Thermo Fisher), N2A (CCL-131, ATCC) and T-Rex<sup>TM</sup>-293 cells (R71007, Thermo Fisher) were grown in DMEM (Gibco) supplemented with 10% heat-inactivated FBS (Gibco), 1×Penicillin streptomycin and glutamine (Gibco) and 1×MEM Non-Essential Amino Acids (Gibco). Transfection was carried using the Neon Transfection System (Invitrogen) per manufacturer protocol. For each transfection,  $2.5 \times 10^6$  cells were electroporated with 8  $\mu\text{g}$  of RE\_Array\_1. For transfection of mCh-Smg, we electroporated 8  $\mu\text{g}$  of pCDNA-D40-mCh-Smg along with 8  $\mu\text{g}$  of RE\_Array\_1 into HeLa cells as described above. The mCh-Smg plasmid was made by PCR amplifying the Smaug coding sequence (CDS) from *Drosophila* S2 cell cDNA using the primers described in Supplementary Table 1. The Smaug CDS was fused to mCherry through overlap PCR using primers described in Supplementary Table 1. This PCR product was cloned into pENTR-D-TOPO (Invitrogen) and subsequently recombined into pcDNA-D40 (Invitrogen) using LR Clonase (Invitrogen).

For transfection of the individual reporters we used Effectene (Promega). The cells were transfected in a 12-well plate with 500 ng each of the EGFP reporter and pCDNA-mCherry<sup>64</sup> per the manufacturers protocol. The cells were split 24 h later into two separate 12-well plates and a 96-well plate. Forty hours after transfection the fluorescence was measured using a Synergy H4 plate reader (BioTek), at the same time RNA and protein was isolated from the 12-well plates.

**RNA isolation and polysome profiling.** Total RNA was isolated using Qiagen RNA mini-prep per manufacturer’s protocol. For HeLa cells we collected total RNA from four biological replicates, when testing the library in other cell types we collected one biological replicate. For polysome profiling, cells were treated with 10  $\mu\text{g ml}^{-1}$  cycloheximide for 5 min prior to harvesting and counting. A total of  $3 \times 10^6$  cells were lysed and the lysate was subjected to ribosome fractionation using 7 to 47% sucrose gradient (Teledyne ISCO) as described previously<sup>65</sup>. In brief, cells were treated with 10  $\mu\text{g/ml}$  cycloheximide for 5 min prior to harvesting and counting. A total of  $3 \times 10^6$  cells were lysed in 250  $\mu\text{L}$  of polysome lysis buffer (20 mM Tris pH 7.2, 130 mM KCl, 10 mM MgCl<sub>2</sub>, 2.5 mM DTT, 0.5% NP-40, 0.2 mg/ml Heparin, 0.5% Sodium Deoxycholate, 10  $\mu\text{g/ml}$  cycloheximide and 200 U/ml RNase Inhibitor) on ice for 20 min prior to clearing at  $8000 \times g$  for 10 min at 4 °C. The lysate was layered over a 7–47% sucrose gradient and subjected to centrifugation at  $160,000 \times g$  for 3 h at 4 °C. The gradient was fractionated by upward displacement with constant measurement of absorbance at 254 nm using a Teledyne ISCO fractionator. RNA was isolated from 40S and polysome fractions using Ribozol (Amresco). We collected polysome fractions from four biological replicates and 40S fractions from two. Isolated RNA was treated with Turbo DNase (Ambion). For qPCR of rRNA from total, 40S and polysome fractions, first strand cDNA synthesis was carried out using Superscript IV reverse transcriptase (Invitrogen) with random hexamer priming. qPCR was performed with iQ<sup>TM</sup> SYBR Green master mix with the 18S and 28S rRNA primers described in Supplementary Table 1.

For qPCR of the individual reporters: RNA was isolated using the Qiagen RNA mini-prep per manufacturer protocol. Isolated RNA was treated with Turbo DNase (Ambion) prior to first strand cDNA synthesis using Superscript Vilo (Invitrogen). Quantitative PCR was performed with EGFP and mCherry primers, Supplementary Table 1.

**Illumina library preparation.** First strand cDNA synthesis for ribosome associated RNA or total RNA was carried out using Superscript IV reverse transcriptase (Invitrogen) with random hexamer priming. The barcode was amplified from cDNA or plasmid using RE\_Amp\_F and RE\_Amp\_R primers with Phusion-HF MM (NEB): 98 °C for 1 min, 22 cycles: 98 °C for 10 s, 55 °C for 30 s, 72 °C for 30 s, and 72 °C for 5 min. The amplicon was purified using Nucleospin Gel and PCR cleanup kit (Macherey Nagel) and subsequently digested with *XhoI* and *SpeI*. The digestion product was purified as before and ligated to the Illumina adapters described in Supplementary Table 1. This product was amplified using IL\_Enrich\_F and IL\_Enrich\_R with Phusion HF MM (NEB): 98 °C for 1 min, 21 cycles: 98 °C for 10 s, 66 °C for 30 s, 72 °C for 30 s, and 72 °C for 5 min. This product was resolved by agarose gel electrophoresis and the appropriate sized band was excised and purified using Nucleospin Gel and PCR cleanup kit (Macherey Nagel).

The Illumina library was multiplexed and run on four lanes of Illumina NextSeq machine. Barcodes counts were determined. Only barcodes with greater than >10 counts in the cDNA and plasmid pools were used for analysis.

**Western blot analysis.** Cells that were transfected with individual reporters were lysed with Lysis Buffer (50 mM Tris pH 7.8, 150 mM NaCl, 1% NP-40). Western blot analysis was performed, as described previously<sup>64</sup>. The following primary antibodies were used in western analysis at the given dilution: GFP, 1:2000 (Clontech, 632381);  $\beta$ -actin-HRP, 1:2000, (Cell Signaling, 12262); Anti-mouse IgG HRP, 1:10,000 (Cell Signaling, 7076S).

**Data analysis.** Relative RNA expression for each regulatory element was calculated as described below. In brief cDNA counts for each barcode were normalized by the plasmid counts for the same barcode. The normalized expression was set relative to the median normalized expression of the control, 4×Blank.

$$\text{Relative RNA Expression} = \log_2 \left[ \frac{\text{cDNA}_x / \text{plasmid}_x}{\text{median} \left( \frac{\text{cDNA}_{\text{control}}}{\text{plasmid}_{\text{control}}} \right)} \right]$$

Relative TE for each regulatory element was calculated as described below. In brief polysome associated cDNA (pRNA) counts for each barcode were normalized by the plasmid counts for the same barcode. The normalized expression was set relative to the median normalized TE of the control, 4×Blank.

$$\text{Relative TE} = \log_2 \left[ \frac{\text{pRNA}_x / \text{cDNA}_x}{\text{median} \left( \frac{\text{pRNA}_{\text{control}}}{\text{cDNA}_{\text{control}}} \right)} \right]$$

Relative 40S association for each regulatory element was calculated as described below. In brief, 40S associated cDNA (srRNA) counts for each barcode were normalized by the plasmid counts for the same barcode. The normalized expression was set relative to the median normalized 40S association of the control, 4×Blank.

$$\text{Relative 40s Association} = \log_2 \left[ \frac{\text{srRNA}_x / \text{cDNA}_x}{\text{median} \left( \frac{\text{srRNA}_{\text{control}}}{\text{cDNA}_{\text{control}}} \right)} \right]$$

Relative TIE for each regulatory element was calculated as described below. In brief, cDNA from polysome associated RNA (pRNA) counts for each barcode were normalized by the cDNA from 40S associated RNA (srRNA) The normalized expression was set relative to the median normalized TIE of the control, 4 × Blank.

$$\text{Relative TIE} = \log_2 \left[ \frac{\text{pRNA}_x / \text{srRNA}_x}{\text{median} \left( \frac{\text{pRNA}_{\text{control}}}{\text{srRNA}_{\text{control}}} \right)} \right]$$

**Linear regression model.** For each synthetic 3’UTR, we calculate median fold change across all 10 barcodes. Median fold change values are fit to linear model with interacting terms for the let-7 binding site, PRE, SRE, AU-rich element or space-holding sequence, at four positions using the lm function in R<sup>66,67</sup>. Coefficients are obtained in reference to “blank” sequence at each position. For cross-validation, we randomly divide the data into five parts and use 80% of the data to train and tested on the remaining 20%. This procedure is repeated five times. The parameters for our linear regression model are shown below:

$$\text{Relative RNA Expression} \sim P_1 + P_2 + P_3 + P_4 + \text{sum}(P_i * P_j)$$

$$\text{Relative TE} \sim P_1 + P_2 + P_3 + P_4 + \text{sum}(P_i * P_j)$$

where  $j = 1$  to 4 and  $i \neq j$ , \* = Interactions

**Thermodynamic model.** To model reporter mRNA levels as a function of the factors bound to their synthetic 3’UTRs we implemented a modified version of a thermodynamic model of transcriptional regulation proposed previously<sup>68</sup> and then later modified<sup>45,46,51</sup>. We modified this framework slightly to model post-transcriptional regulation of mRNA by RNA binding proteins and miRNAs as described below.

In our model we specify that each 3’UTR has four binding sites for factors that bind RNA (RNA binding proteins and miRNAs). When a spacer sequence is present at one of the sites we assume that no factor is bound to the site. We also specify that the 3’UTR has one additional binding site for a rate-limiting factor that controls the regulation of the mRNA. RBPs and miRNAs bound on the 3’UTR either facilitate or inhibit the binding of this rate-limiting factor. The primary assumption of this model is that the equilibrium binding of RBPs and miRNAs determine the occupancy of the rate-limiting factor on the 3’UTR. The equilibrium assumption is justified because the timescales at which RBPs and miRNA bind and unbind their sites are much faster than the timescale at which regulation of mRNA levels occur<sup>69,70</sup>.

In our model of synthetic 3’UTRs we specify that there are five ( $N$ ) total binding sites, four for RBPs and miRNAs, and one for the rate-limiting regulator. Each 3’UTR can exist in many possible states, where a state is a particular configuration of bound factors. A state of the 3’UTR is specified by giving the occupancy  $\sigma_i$ , either 0 or 1, of each site. The model calculates a Boltzmann weight for each state of the 3’

UTR using:

$$W(\sigma_1, \dots, \sigma_N, \sigma_{1,1}, \dots, \sigma_{N,N}) = \exp\left(-\sum_{i=1}^N q_i \sigma_i - \sum_{j=1}^N \omega_{i,j} \sigma_{i,j}\right)$$

where  $\sigma_{i,j}$  represents the number of instances of factor  $i$  and factor  $j$  bound to each other in that configuration (see below),  $q_i$  represents the change in free energy of RNA binding factor  $i$  binding to its respective site plus the natural log of the concentration of that factor, and  $\omega_{i,j}$  is the change in free energy of factor  $i$  and factor  $j$  interacting, which is a measure of cooperativity between factor  $i$  and  $j$ . We assume that no factor binds to the spacer sequence and that all of the RNA-binding factors are present in the cell at roughly equal concentrations and bind 3'UTR with same affinity. We also assume that the rate limiting factor binds the 3'UTR with a particular affinity, set to two units for the PTRE-seq data. We model different factor-factor interactions by allowing cooperativity in particular configurations. In some cases, we allow cooperativity only when two factors are simultaneously bound to directly adjacent sites. In other cases, we allow cooperativity as long as the two factors are bound to the same 3'UTR at any two sites. The rate-limiting factor is allowed to interact with all factors bound to the 3'UTR. In our notation  $q_{\text{Factor}}$  and  $\omega_{\text{Factor}}$  are equal, respectively, to  $\ln(q_{\text{Factor}})$  and  $\ln(\omega_{\text{Factor}})$  described previously<sup>51</sup>. To calculate the probability that the rate-limiting factor is bound to the 3'UTR, and therefore the predicted level of the mRNA, we use the equation:

$$\text{mRNA level} = \alpha \left( \frac{\sum W \times \delta(\text{RLF})}{\sum W} \right)$$

where  $\alpha$  is a scaling factor whose value is the least squares estimate, and  $\delta$ (rate limiting factor) is a delta function that is zero when the rate limiting factor is not bound and one when it is bound. The summations are over all possible states of the 3'UTR. To fit the model to data we search for values of  $\alpha$ , all  $q$ , and all  $\omega$  that best fit the observed mRNA levels measured by PTRE-seq. The parameters are fit with custom Python scripts using SciPy to minimize the objective function using constrained minimization optimization algorithms L-BFGS-B and SLSQP in alternating fashion until the parameter values converge. The asymptotic normal distribution for the parameter estimate is used to calculate the 95% confidence intervals for the parameter values<sup>46</sup>. After fitting, any non-zero values of a particular  $\omega_{i,j}$  are interpreted as cooperativity between the two factors.

**Data availability.** The authors declare that the data supporting the findings described here are available within the article, the Supplementary Information or Supplementary Data 1–7. Extra data are available from the corresponding author upon request. Scripts used for analysis and model fitting are available at the Github repository under MIT license ([https://github.com/hemangichaudhari/Cottrell\\_PTRE-seq\\_scripts](https://github.com/hemangichaudhari/Cottrell_PTRE-seq_scripts)). Sequencing data is available at the Sequence Read Archive under accession code SRP127467.

Received: 12 June 2017 Accepted: 22 December 2017

Published online: 19 January 2018

## References

- Halbeisen, R. E., Galgano, A., Scherrer, T. & Gerber, A. P. Post-transcriptional gene regulation: from genome-wide studies to principles. *Cell Mol. Life. Sci.* **65**, 798–813 (2008).
- Fabian, M. R. & Sonenberg, N. The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. *Nat. Struct. Mol. Biol.* **19**, 586–593 (2012).
- Fukaya, T., Iwakawa, H. O. & Tomari, Y. MicroRNAs block assembly of eIF4F translation initiation complex in *Drosophila*. *Mol. Cell* **56**, 67–78 (2014).
- Meijer, H. A. et al. Translational repression and eIF4A2 activity are critical for microRNA-mediated gene regulation. *Science* **340**, 82–85 (2013).
- Mathonnet, G. et al. MicroRNA inhibition of translation initiation in vitro by targeting the cap-binding complex eIF4F. *Science* **317**, 1764–1767 (2007).
- Braun, J. E., Huntzinger, E., Fauser, M. & Izaurralde, E. GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. *Mol. Cell* **44**, 120–133 (2011).
- Eulalio, A. et al. Deadenylation is a widespread effect of miRNA regulation. *RNA* **15**, 21–32 (2009).
- Behm-Ansmant, I. et al. mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes Dev.* **20**, 1885–1898 (2006).
- Fukao, A. et al. MicroRNAs trigger dissociation of eIF4AI and eIF4AII from target mRNAs in humans. *Mol. Cell* **56**, 79–89 (2014).
- Kozomara, A. & Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* **39**, D152–D157 (2011).
- Friedman, R. C., Farh, K. K., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **19**, 92–105 (2009).
- Castello, A. et al. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393–1406 (2012).
- Brannan, K. W. et al. SONAR discovers RNA-binding proteins from analysis of large-scale protein–protein interactomes. *Mol. Cell* **64**, 282–293 (2016).
- Baltz, A. G. et al. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol. Cell* **46**, 674–690 (2012).
- Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat. Rev. Genet.* **15**, 829–845 (2014).
- Weidmann, C. A., Raynard, N. A., Blewett, N. H., Van Etten, J. & Goldstrohm, A. C. The RNA binding domain of Pumilio antagonizes poly-adenosine binding protein and accelerates deadenylation. *RNA* **20**, 1298–1319 (2014).
- Van Etten, J. et al. Human Pumilio proteins recruit multiple deadenylases to efficiently repress messenger RNAs. *J. Biol. Chem.* **287**, 36370–36383 (2012).
- Gerber, A. P., Luschnig, S., Krasnow, M. A., Brown, P. O. & Herschlag, D. Genome-wide identification of mRNAs associated with the translational regulator PUMILIO in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **103**, 4487–4492 (2006).
- Cao, Q., Padmanabhan, K. & Richter, J. D. Pumilio 2 controls translation by competing with eIF4E for 7-methyl guanosine cap recognition. *RNA* **16**, 221–227 (2010).
- Baez, M. V. & Boccaccio, G. L. Mammalian Smaug is a translational repressor that forms cytoplasmic foci similar to stress granules. *J. Biol. Chem.* **280**, 43131–43140 (2005).
- Smibert, C. A., Wilson, J. E., Kerr, K. & Macdonald, P. M. Smaug protein represses translation of unlocalized nanos mRNA in the *Drosophila* embryo. *Genes Dev.* **10**, 2600–2609 (1996).
- Barreau, C., Paillard, L. & Osborne, H. B. AU-rich elements and associated factors: are there unifying principles? *Nucleic Acids Res.* **33**, 7138–7150 (2005).
- Plass, M., Rasmussen, S. H. & Krogh, A. Highly accessible AU-rich regions in 3' untranslated regions are hotspots for binding of regulatory factors. *PLoS Comput. Biol.* **13**, e1005460 (2017).
- Jiang, P., Singh, M. & Collier, H. A. Computational assessment of the cooperativity between RNA binding proteins and MicroRNAs in Transcript Decay. *PLoS Comput. Biol.* **9**, e1003075 (2013).
- Galgano, A. et al. Comparative analysis of mRNA targets for human PUF-family proteins suggests extensive interaction with the miRNA regulatory system. *PLoS ONE* **3**, e3164 (2008).
- Mukherjee, N. et al. Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. *Mol. Cell* **43**, 327–339 (2011).
- Rinck, A. et al. The human transcriptome is enriched for miRNA-binding sites located in cooperativity-permitting distance. *RNA Biol.* **10**, 1125–1135 (2013).
- Preusse, M. et al. SimiRa: A tool to identify coregulation between microRNAs and RNA-binding proteins. *RNA Biol.* **12**, 998–1009 (2015).
- Miles, W. O., Tschöp, K., Herr, A., Ji, J. Y. & Dyson, N. J. Pumilio facilitates miRNA regulation of the E2F3 oncogene. *Genes Dev.* **26**, 356–368 (2012).
- Kedde, M. et al. A Pumilio-induced RNA structure switch in p27-3' UTR controls miR-221 and miR-222 accessibility. *Nat. Cell Biol.* **12**, 1014–1020 (2010).
- Kim, H. H. et al. HuR recruits let-7/RISC to repress c-Myc expression. *Genes Dev.* **23**, 1743–1748 (2009).
- Bhattacharyya, S. N., Habermacher, R., Martine, U., Closs, E. I. & Filipowicz, W. Relief of microRNA-mediated translational repression in human cells subjected to stress. *Cell* **125**, 1111–1124 (2006).
- Cai, Y. & Futcher, B. Effects of the yeast RNA-binding protein Whi3 on the half-life and abundance of CLN3 mRNA and other targets. *PLoS ONE* **8**, e84630 (2013).
- Bazzini, A. A., Lee, M. T. & Giraldez, A. J. Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* **336**, 233–237 (2012).
- Guo, H., Ingolia, N. T., Weissman, J. S. & Bartel, D. P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* **466**, 835–840 (2010).
- Melnikov, A. et al. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.* **30**, 271–277 (2012).
- Patwardhan, R. P. et al. Massively parallel functional dissection of mammalian enhancers in vivo. *Nat. Biotechnol.* **30**, 265–270 (2012).
- Sharon, E. et al. Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat. Biotechnol.* **30**, 521–530 (2012).
- Kwasnieski, J. C., Mogno, I., Myers, C. A., Corbo, J. C. & Cohen, B. A. Complex effects of nucleotide variants in a mammalian cis-regulatory element. *Proc. Natl Acad. Sci. USA* **109**, 19498–19503 (2012).
- Zhao, W. et al. Massively parallel functional annotation of 3' untranslated regions. *Nat. Biotechnol.* **32**, 387–391 (2014).

41. Oikonomou, P., Goodarzi, H. & Tavazoie, S. Systematic identification of regulatory elements in conserved 3' UTRs of human transcripts. *Cell Rep.* **7**, 281–292 (2014).
42. Wissink, E. M., Fogarty, E. A. & Grimson, A. High-throughput discovery of post-transcriptional cis-regulatory elements. *BMC Genom.* **17**, 177 (2016).
43. Yartseva, V., Takacs, C. M., Vojnar, C. E., Lee, M. T. & Giraldez, A. J. RESA identifies mRNA-regulatory sequences at high resolution. *Nat. Methods* **14**, 201–207 (2017).
44. Levo, M. et al. Unraveling determinants of transcription factor binding outside the core binding site. *Genome Res.* **25**, 1018–1029 (2015).
45. Gertz, J., Siggia, E. D. & Cohen, B. A. Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature* **457**, 215–218 (2009).
46. Fiore, C. & Cohen, B. A. Interactions between pluripotency factors specify cis-regulation in embryonic stem cells. *Genome Res.* **26**, 778–786 (2016).
47. White, M. A. Understanding how cis-regulatory function is encoded in DNA sequence using massively parallel reporter assays and designed sequences. *Genomics* **106**, 165–170 (2015).
48. Warner, J. R., Knopf, P. M. & Rich, A. A multiple ribosomal structure in protein synthesis. *Proc. Natl Acad. Sci. USA* **49**, 122–129 (1963).
49. Sonenberg, N. & Hinnebusch, A. G. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136**, 731–745 (2009).
50. Sherman, M. S. & Cohen, B. A. Thermodynamic state ensemble models of cis-regulation. *PLoS Comput. Biol.* **8**, e1002407 (2012).
51. Buchler, N. E., Gerland, U. & Hwa, T. On schemes of combinatorial transcription logic. *Proc. Natl Acad. Sci. USA* **100**, 5136–5141 (2003).
52. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).
53. Doench, J. G. & Sharp, P. A. Specificity of microRNA target selection in translational repression. *Genes Dev.* **18**, 504–511 (2004).
54. Powers, J. T. et al. Multiple mechanisms disrupt the let-7 microRNA family in neuroblastoma. *Nature* **535**, 246–251 (2016).
55. Ricci, E. P. et al. miRNA repression of translation in vitro takes place during 43S ribosomal scanning. *Nucleic Acids Res.* **41**, 586–598 (2013).
56. Agarwal, V., Bell, G. W., Nam, J. W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *Life* **4**, e05005 (2015).
57. Garcia, D. M. et al. Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. *Nat. Struct. Mol. Biol.* **18**, 1139–1146 (2011).
58. Grimson, A. et al. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* **27**, 91–105 (2007).
59. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15–20 (2005).
60. Hamzeiy, H., Allmer, J. & Yousef, M. Computational methods for microRNA target prediction. *Methods Mol. Biol.* **1107**, 207–221 (2014).
61. Hake, L. E., Mendez, R. & Richter, J. D. Specificity of RNA binding by CPEB: requirement for RNA recognition motifs and a novel zinc finger. *Mol. Cell Biol.* **18**, 685–693 (1998).
62. Stebbins-Boaz, B., Hake, L. E. & Richter, J. D. CPEB controls the cytoplasmic polyadenylation of cyclin, Cdk2 and c-mos mRNAs and is necessary for oocyte maturation in *Xenopus*. *EMBO J.* **15**, 2582–2592 (1996).
63. Piqué, M., López, J. M., Foissac, S., Guigó, R. & Méndez, R. A combinatorial code for CPE-mediated translational control. *Cell* **132**, 434–448 (2008).
64. Cottrell, K. A. & Djuranovic, S. Urb-RIP—an adaptable and efficient approach for immunoprecipitation of RNAs and associated RNAs/Proteins. *PLoS ONE* **11**, e0167877 (2016).
65. Kuchenreuther, M. J. & Weber, J. D. The ARF tumor-suppressor controls Drosha translation to prevent Ras-driven transformation. *Oncogene* **33**, 300–307 (2014).
66. Wilkinson, G. N. & Rogers, C. E. Symbolic descriptions of factorial models for analysis of variance. *Appl. Stat.* **22**, 392–399 (1973).
67. Chambers, J. M. *Statistical Models in S* (Wadsworth & Brooks/Cole, 1992).
68. Shea, M. A. & Ackers, G. K. The OR control system of bacteriophage lambda. A physical-chemical model for gene regulation. *J. Mol. Biol.* **181**, 211–230 (1985).
69. Garcia, H. G., Kondev, J., Orme, N., Theriot, J. A. & Phillips, R. Thermodynamics of biological processes. *Methods Enzymol.* **492**, 27–59 (2011).
70. Phillips, R. Napoleon is in equilibrium. *Annu Rev. Condens Matter Phys.* **6**, 85–111 (2015).
71. Rehmsmeier, M., Steffen, P., Hochsmann, M. & Giegerich, R. Fast and effective prediction of microRNA/target duplexes. *RNA* **10**, 1507–1517 (2004).
72. Betel, D., Koppal, A., Agius, P., Sander, C. & Leslie, C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.* **11**, R90 (2010).

## Acknowledgements

We are thankful to Weber J. and Maggi L. for assistance and equipment used in poly-some profiling. We thank Jovanovic M. and Szczesny P. for valuable comments. H.G.C. and B.A.C. supported by GM092910 and R01-HG008687. This work was funded by NIH GM007067 to K.A.C. and R01-GM112824 to S.D.

## Author contributions

K.A.C. performed all experiments, designed the library, performed data analysis and wrote the manuscript. H.G.C. assisted in the library design, data analysis and model generation. B.A.C. assisted in library design and provided comments on data analysis. S.D. conceived the project and provided funding. All authors were involved in editing the manuscript.

## Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-017-02745-0>.

**Competing interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017