Probing 3'UTRs as modular regulators of gene expression

A Preprint

Jamie Auxillos
School of Biological Sciences
University of Edinburgh
Scotland, EH9 3JW
jamie.auxillos@ed.ac.uk

Samuel Haynes
School of Biological Sciences
University of Edinburgh
Scotland, EH9 3JW
s.j.haynes@ed.ac.uk

Abhishek Jain
School of Biological Sciences
University of Edinburgh
Scotland, EH9 3JW

Clemence Alibert
School of Biological Sciences
University of Edinburgh
Scotland, EH9 3JW

Weronika Danecka
School of Biological Sciences
Univeristy of Edinburgh
Scotland, EH9 3JW
weronika.danecka@ed.ac.uk

Edward Wallace
School of Biological Sciences
Univeristy of Edinburgh
Scotland, EH9 3JW
edward.wallace@ed.ac.uk

August 11, 2021

Abstract

Genes are commonly abstracted into a coding sequence and cis-regulatory elements (CREs), such as promoter and terminator regions, and short sequence motifs within these regions. Modern cloning techniques allow easy assembly of synthetic genetic constructs from discrete cis-regulatory modules. However, it is unclear how much the contributions of CREs to gene expression depend on other CREs in the host gene. Using budding yeast, we probe the extent of composability, or independent effects, of distinct CREs. We confirm that the quantitative effect of a terminator on gene expression depends on both promoter and coding sequence. We then explore whether individual cis-regulatory motifs within terminator regions display similar context dependence, focusing on putative regulatory motifs inferred using transcriptome-wide datasets of mRNA decay. We construct a library of diverse reporter genes, consisting of different combinations of motifs within various terminator contexts, paired with different promoters, to test the extent of composability. Our results show that the effect of a motif on RNA abundance depends both on its host terminator, and also on the associated promoter sequence. Consequently, this emphasises the need for improved motif inference algorithms that include both local and global context effects, which in turn could aid researchers in the accurate use of diverse CREs for the engineering of synthetic genetic constructs.

1 Introduction

Our understanding of gene expression relies on the abstraction of a gene into discrete cis-regulatory elements (CRE), or parts, based on their most prominent function. The promoter coordinates the initiation of transcription. Transcription then progresses until it reaches the terminator, where cleavage and (in eukaryotes) polyadenylation of the nascent RNA transcript is coordinated. Extending the abstraction to the mature messenger RNA transcript, the coding sequence is flanked by untranslated regions (UTRs) at its 5' and 3' ends. The 5'UTR contributes to translation regulation and the 3'UTR contributes to transcript stability and localisation.

Abstracting specific characteristics to individual parts of a gene has facilitated the development of genetic toolkits that are based on the modular assembly of interchangeable parts. Furthermore, it is tempting to

assume that a part with a function validated in a single context, e.g. in a specific gene and under certain experimental conditions, would have the same function in a different context. However, this assumption blurs two concepts of modularity: modular assembly, how easy it is to construct combinations of parts; and composability, how function and regulation can be decomposed into contributions from individual parts. The assembly of parts whose functions are not composable will result in constructs with unpredictable functions or regulatory behaviours.

Modular assembly is now routine: modern DNA synthesis and cloning techniques allow the construction of complex gene pathways through the combinatorial assembly of chosen parts (1–4). Several groups have constructed standardised libraries comprising promoter (including 5'UTR) and terminator (including 3'UTR) parts, which can be combined to achieve desired expression of synthesized proteins. The characterisation of all possible promoter-terminator pairings is impractical, so standardised libraries generally characterise parts independently, for example testing many promoters with the same coding sequence and terminator (2). Similarly, even massively parallel approaches to characterise smaller CREs usually explore a large library of promoter elements with a single terminator (5), or conversely a library of terminators with a single promoter (6). Because these experiments do not measure interactions between different parts, they rely on the untested assumption that those parts have more or less identical effects across different contexts.

Computational methods for the discovery and quantification of short CREs, such as sequence motifs recognised by regulatory proteins, also tend to rely on models of independent effects, that is, composability. Prominent methods for mapping CRE sequence-function relationships include predicting functional data with short sequence features, often using linear models (7–9); or, directly comparing the sequences of genes with similar characteristics to determine the presence of short consensus sequence motifs (10, 11) using motif discovery software (12, 13). Both of these approaches make the implicit approximation that the contribution of a short CRE is independent of context, so that the effect of combining motifs is composed of a linear sum (on the appropriate scale) of the individual CRE contributions. The approximation that short CREs act independently helps to find elements that have clear contributions, and to simplify a vast search space that would be made exponentially larger by accounting for CRE interactions. However, framing the search for CREs around independent contributions from short motifs overlooks multi-part motifs, interactions between motifs, and that motifs may be active only in specific contexts.

The importance of interactions among CREs has been clear since the discovery of cooperative DNA binding by transcription factors (14). For example, whether post-transcriptional regulatory CREs are functional or not depends on the RNA sequence context (15). One large-scale study of synthetic combinations of bacterial promoters and ribosomal binding sites showed that expression was largely composed of independent contributions; 64% of constructs had protein expression levels with less than a 2-fold deviation from the predictions of a linear model without interactions (16). However, 5% of the constructs had, on average, a 13-fold deviation. Another study of promoter-terminator swaps in budding yeast revealed 2-fold changes due to promoter-terminator interactions (17), even though these CREs are separated by 700nt of coding sequence.

Here, we provide further evidence for the importance of interactions between regulatory regions, and show the limitations of assuming that gene expression regulation can be decomposed into independent contributions from short CREs. First, we use modular cloning to construct 120 combinations of promoter (including 5'UTR), coding sequence, and terminator (including 3'UTR) modules, showing that the quantitative effect of terminators on protein production depends on both promoter and coding sequence. We then investigate the effect of promoter and terminator context on the regulatory behaviour of short CREs in 3'UTRs. In order to select suitable 3'UTR motifs, we adapt a method to computationally predict mRNA log2-half-life as a linear combination of codon usage and motif counts (8). We then quantify the effect on mRNA and protein levels of selected 3'UTR motifs when inserted in, or removed from, diverse reporter constructs. Our results demonstrate that the incorporation of candidate decay motifs indeed results in a decrease in mRNA abundance. However, the quantitative contribution from motifs depends on both the local 3'UTR context and the promoter context. These results highlight limits to the composability of cis-regulatory elements.

2 Materials and methods

2.1 Strains and media

Saccharomyces cerevisiae strain BY4741 (*MATa his3*Δ1 *leu2*Δ0 *met15*Δ0 *ura3*Δ0) was used as the wild-type strain in this study, and the host for all yeast plasmid transformations. For all quantitative assays, plasmid-transformed strains were grown in synthetic complete medium without uracil (SC-Ura), containing 0.69% yeast nitrogen base without amino acids and with ammonium sulfate (Formedium, Norfolk, UK), 0.193%

amino acid drop-out supplement mixture (Formedium, Norfolk, UK) and 2% glucose. To prepare BY4741 for transformation, we grew it in YPDA medium, containing 2% peptone, 1% yeast extract, 2% glucose and 0.004% adenine.

2.2 Construction of chimeric reporter plasmids

All fluorescence reporter plasmids were constructed by Golden Gate assembly using the YeastFab system as described in (3). Promoters, coding sequences and terminators were either amplified from the yeast genome or synthesised by a commercial vendor (IDT) then cloned into a parts accepting plasmid (HcKan_P for promoters, HcKan_O for coding sequences and HcKan_T for terminators) by Golden Gate assembly using Bsa1-HFv2 (NEB). A detailed protocol for Golden Gate assembly is available at protocols.io, doi:10.17504/protocols.io.bkqrkvv6. Using these parts libraries, the promoters, coding sequences and terminators were assembled together into the transcription unit acceptor plasmid (POT1-ccdB) by Golden Gate assembly using Esp3I (NEB); these are low-copy centromeric plasmids with URA3 selection marker, as described in (3). Plasmid inserts were confirmed by Sanger sequencing (MRC PPU DNA Sequencing and Services, Dundee). DNA sequences used in this study are summarised in Supplementary Tables 2 and 3. Assembled plasmids were transformed into yeast BY4741 using lithium acetate transformation (18), and selected in SC-URA agar plates to isolate successful transformants.

The mCherry coding sequence is as used in (3), which in turn was amplified from the mCherry sequence in (5). The mTurquoise2 coding sequence is as used in (2).

2.3 Fluorescence measurements: Plate reader analysis of strain growth and fluorescence

Yeast with plasmids were grown in a 96-well deep well plate (VWR) containing 100μl of SC-Ura medium with 2% glucose and grown for ~12 hours at 30°C in a shaking incubator set at 250 rpm. The next day, the cultures were diluted to an OD of 0.2. For each sample, 3 technical replicates of 200μl were transferred to a 96-well black microtiter plate (Corning) and grown according to the protocol described in (19). The Tecan Infinity M200 series plate reader was set at the temperature of 29.9 (range of 29.4-30.4°C) with linear shaking (6 mm amplitude at 200-220 rpm). OD measurements were carried out at an absorbance wavelength of 595 nm with a measurement bandwidth of 9 nm with 15 reads. mCherry fluorescence measurements were carried out with an excitation wavelength at 585 nm and an emission wavelength of 620 nm (excitation bandwidth of 9 nm and emission bandwidth of 20 nm) with the gain set at 100. mTurquoise2 fluorescence measurements were carried out with an excitation wavelength at 434 nm and an emission wavelength of 474 nm (excitation bandwidth of 9 nm and emission bandwidth of 20 nm) with the gain set at 60.

Plate reader data were analysed using omniplate software (20). Omniplate accounts for autofluorescence and fits a model to the time series data using Gaussian processes to infer the time of maximum growth rate for each well. We minimised growth-dependent effects by using the fluorescence at maximum growth rate for all of our protein fluorescence experiments. Each fluorescence measurement was also normalised by OD to remove dependency on cell number, so every protein fluorescence measurement is recorded as fluorescence per OD at max growth rate. A detailed protocol for setting up and conducting the plate reader assay is available at protocols.io, doi:10.17504/protocols.io.bbicikaw.

2.4 RNA measurements: Strain growth, RNA extraction, RT-qPCR, and analysis

Yeast with plasmids were grown in a 24-well deep well plate plate (4titude) containing 1.5 ml of SC-Ura for ~20 hours at 30°C in a shaking incubator set at 250 rpm. The next day, the OD was diluted to a starting OD between 0.15-0.2 in a 12-column deep well reservoir plate (4titude) to a total volume of 7 ml. Diluted cultures were grown at 30°C in a shaking incubator set at 90 rpm to an OD of 0.5-0.7 then pelleted by centrifugation. Pelleted cells in the plate were stored in -80°C.

To extract RNA, we adapted a silica column DNA/RNA extraction protocol from Zymo Research (Irvine, California, USA). The pelleted cells were thawed and individually resuspended in 400 μl of RNA binding buffer (Zymo), then transferred to 2 ml screw cap tubes containing zirconia beads, lysed using the Precellys Evolution homogeniser then pelleted by centrifugation for 1.5 minutes. The supernatant was transferred to a Zymo Spin IIICG column (Zymo) then centrifuged. The flow through was mixed with 1 volume of ethanol then transferred to a Zymo Spin IIC column (Zymo) and centrifuged. This flow through was discarded and 1 volume of DNA/RNA Prep buffer (Zymo) was added then centrifuged. The column was washed with 700 μl of Zymo DNA/RNA Wash buffer (Zymo) then centrifuged. The column was washed a second time, but

with 400 μl of Zymo DNA/RNA Wash buffer (Zymo). The column was centrifuged once more to remove residual wash buffer in the column. Lastly, 30 μl of nuclease free water was added to the column then eluted. All centrifugation steps in the RNA extraction protocol were carried out at 12,000g for 1 minute unless otherwise stated. A detailed protocol for yeast growth and RNA extraction is available at protocols.io, doi:10.17504/protocols.io.beetjben.

The quantity and quality of the RNA was measured using both a spectrophotometer (DS-11, DeNovix, Wilmington, Delaware, USA) and Fragment Analyser (Agilent). 4 μg of RNA was treated with DNAse1 (Thermo) then inactivated using the RapidOut DNA removal kit (Thermo) according to the manufacturer's protocol. 2.5 μl of Random primer mix (NEB) was added to the mixture then separated into 2 PCR tubes (one for -RT and one for +RT) then denatured at 70°C followed by cooling on ice. Reverse transcription (RT) master mix was prepared, containing 2 μl of First Strand synthesis buffer, 0.75 μl of 10mM dNTP mix, 1.5 μl of nuclease free water, 0.25 μl of RNase inhibitor and 0.5 μl of SuperScript IV Reverse Transcriptase (Invitrogen) per reaction. 5 μl of the RT master mix was added to the denatured RNA then incubated at 25°C for 5 minutes then 55°C for 1 hour. The cDNA was diluted with 200 μl of nuclease free water.

Target cDNAs were measured by quantitative PCR with Brilliant II Fast SYBR Green QPCR Master Mix (Agilent) using a Lightcycler 480 qPCR machine (Roche). We measured all +RT reactions in technical triplicate, and negative control -RT samples using one replicate. We used the manufacturer's software to calculate quantification cycle (Cq) for each individual well using the fit points method, and exported both raw fluorescence and Cq data. All primer sets were thoroughly validated by serial dilution and by confirming amplicon size, sequences are available in Supplementary Table 4.

RT-qPCR data was analysed using our tidyqpcr R package v2.1.0-beta (https://github.com/ewallace/tidyqpcr). For each biological replicate, ΔCq values were calculated by normalising the median mCherry Cq values by the median Cq values of the three reference genes (RPS3, PGK1 and URA3). For the constructs with motif insertions in terminators, $\Delta\Delta$Cq values were calculated by normalising mCherry ΔCq by that of control construct mod_NNN strains (with the corresponding promoter) for tRPS3 and tTSA1 consructs. For the constructs with motif deletions in terminators, $\Delta\Delta$Cq values were calculated by normalising mCherry ΔCq by that of the WT terminator (with the corresponding promoter) for tPIR1 constructs. Complete scripts for qPCR analysis, quality control, and figure generation are available online at https://github.com/DimmestP/chimera_project_manuscript/.

2.5 Determining 3'UTR decay motifs

We initially selected 69 3'UTR motifs to investigate from three separate studies of cis-regulatory elements suspected to regulate mRNA decay (7, 8, 10). To select a short list of motifs to test for context dependence, we determined each motif's contribution to a linear model predicting half-life. Following (8), we quantified the effect of motifs on transcript half-life using a linear model predicting half-life on the basis of codon usage, 3'UTR length, and 3'UTR motif frequency.

$$log_2(\lambda_g^{1/2}) = \sum_c \beta_c p_{cg} + \sum_m \alpha_m n_{mg} + \gamma\omega_g + \epsilon$$

where $\lambda_g^{1/2}$ is the half-life of gene g, β_c is the coefficient associated with codon c, p_{cg} is the proportion of gene g's coding sequence that corresponds to codon c, γ is the coefficient associated with 3'UTR length, ω_g is the 3'UTR length of gene g, α_m is the coefficient associated with motif m, n_{mg} is number of occurrences of motif m in gene g's 3'UTR, and ϵ is the noise term. To choose 3'UTR lengths and to assess which sequence to use for 3'UTR motif search, we used the median 3'UTR length estimates (precisely, the median length of clustered major transcript isoforms) reported from the TIF-seq analysis in (21).

We removed motifs that did not significantly contribute to half-life by using a greedy model selection algorithm that minimises the Akaike information criterion (AIC).

$$AIC = 2k - 2ln(\hat{L})$$

where k is the number of parameters in the model and $\hat{L}$ is the maximum value of the likelihood function (22). We implement this motif comparison using the R function step (23, 24), to iteratively add the motif which reduces the model's AIC the most until the penalty for adding news terms overcomes the benefit of including a new motif. This procedure was run on each decay data set independently.

The variance explained by each predictor in the linear model trained on the (25) dataset is estimated in three ways. First, the variance explained by three individual linear models each containing one of the predicates is

calculated. Then, we determine the reduction in variance explained when one of the predicates is removed from the full model. Finally, the increase in variance explained when each predicate is added to the model sequentially, starting with codon usage and ending with 3'UTR length, is reported.

We selected the specific versions of the HWNCATTWY and TGTAHMNTA motifs by running two addition linear models predicting half-life that inferred separate coefficients for each version of its consensus sequence. Coefficients were reported for the significant motif versions (Supplementary Table 5, 6). We chose instances with similar effect size, statistical significance, and number of occurrences in native transcripts. We chose TTTCATTTC for HWNCATTWY and for TGTAHMNTA chose TGTACAATA over TGTATATTA specifically to avoid the 5nt stretch also found in ATATTC.

2.6 Design of modified 3'UTRs for testing the effects of mutated motifs

RPS3 was chosen as the first 3'UTR for inserting motifs into as it was the only terminator in the characterized library that did not contain any of the 69 original motifs of interest. The tRPS3 3'UTR-terminator was modified to incorporate three 9 nt insertion sites for motifs (M1, M2 and M3). The M1 was inserted 24 nt downstream of the stop codon, M2 was inserted 15 nt downstream of M1 and the final insert M3 was inserted 4 nt downstream of M2 (Figure 4A). These positions were selected based on key design criteria, including; minimal perturbations of RNA secondary structure as predicted by RNAfold (26), position of motifs in native 3'UTRs and position of other CREs important for transcriptional control (Supplementary Table 7, Supplementary Figure 2). A control tRPS3 3'UTR mod_NNN was designed to incorporate random bases in each insertion site. Further modified 3'UTR-terminator designs were designed to incorporate individual motifs of interest previously identified, within the insertion sites described (Figure 4A).

We chose an alternative 3'UTR for screening the effects of inserting motifs of interest by searching for characteristics similar to RPS3. To this end, median-length 3'UTRs were extracted from the (21) dataset filtered for the following criteria; 1) does not contain any of the original 69 motifs of interest, 2) $<$ 300 nt in length, 3) from a highly expressed gene, 4) synthesisable as a gBlock by our manufacturer (IDT). The 3'UTR from TSA1 met these criteria.

Similar to the modified tRPS3 designs, in tTSA1 we designed three 9 nt motif insertion sites: M1 21 nt downstream of the stop codon, M2 20 nt downstream of M1, and M3 24 bp downstream of M2 (Figure 4A). The tTSA1 mod_NNN construct contained random bases in the M1, M2 and M3 sites, with the motif insertions in other modified constructs as for tRPS3.

To design deletion constructs we selected a native 3'UTR that contained the motifs of interest. Again, median-length 3'UTRs were extracted from the Pelechano et al. (2013) dataset filtered for the following criteria; 1) contains at least 3 shortlisted motifs of interest, 2) a highly expressed gene, 3) synthesisable. The PIR1 terminator chosen for motif deletion contains one copy each of the ATATTC and UGUAHMNU motifs, and 3 copies of the HWNCAUUWY motifs (Figure 5A), although did not contain the putative stability motif GTATACCTA.

The mutation of motifs for their removal from the PIR1 3'UTR was carried out so that: 1) at least 50% of the motif sequence (specifically the motif consensus sequence) was mutated to a base that does not correspond to the consensus sequence, 2) GC content was minimally altered, 3) Mutations that resulted in a limited change in the predicted secondary structure and minimum free energy (MFE) according to RNAfold (26), see Supplementary Table 7).

The native and modified candidate 3'UTRs were screened for the presence of Esp3I and BsaI sites within the sequence. For incorporation into the YeastFab system, the sequence 'agcgtgCGTCTCgTAGC' was added to the 5'-end of the 3'UTR and the sequence 'CCTCcGAGACGcagcac' was added to the 3'-end of the 3'UTR. To check if sequences were synthesisable, 100 nt downstream of the native 3'UTR was added to the candidate and the sequence checked at the IDT gBlock entry form (https://eu.idtdna.com/site/order/gblockentry).

2.7 Determining motif effect on abundance

A linear model predicting construct ΔCq's using the presence or absence of the four selected motifs was trained on each promoter-terminator pairing separately. The model included a term to account for interactions between the UGUAMNUA and HWNCAUUWY motifs The linear model also included a term for batch effects, between the 2 experimental batches of 3 biological replicates for each set of constructs, because this

improved the quality of model fit. The model was:

$$\Delta Cq = \sum_{m=1}^{4} \alpha_m n_m + \beta e_{rep} + \gamma n_{int} + \epsilon$$

where n_m is the copy number of motif m in the construct, e_{rep} is which experimental batch the construct was part of and n_{int} is the interaction term with value 1 if both UGUAMNUA and HWNCAUUWY motifs are present.

2.8 Predicting changes in transcript abundance from changes in half-life

A simple kinetic model of the production and decay of transcripts was used:

$$\varnothing \xrightarrow{k} m$$

$$m \xrightarrow{\delta} \varnothing$$

The steady state solution for the average number of transcripts, $\langle m \rangle$, is

$$\langle m \rangle = \frac{k}{\delta}$$

where k is the rate of transcription, which can include multiple active states, and δ is the rate of decay for the transcript (27).

Now consider a control transcript m_0, and a similar transcript with an altered terminator m_a. Assuming the alterations to the host gene's terminator have a minimal impact on the transcription rate, the above equation says that the ratio of predicted abundance $\langle m_a \rangle$ to the control transcript abundance, $\langle m_0 \rangle$, is the same as the ratio of their half-lives:

$$\frac{\langle m_a \rangle}{\langle m_0 \rangle} = \frac{\delta_0}{\delta_a} = \frac{\lambda_a^{1/2}}{\lambda_0^{1/2}}$$

This gives a linear effect on the log-scale abundance

$$\log\langle m_a \rangle = \Delta \log \lambda^{1/2} + \log\langle m_0 \rangle$$

and because PCR quantification cycle Cq is proportional to $\log 2(m)$, this directly leads to a linear effect on Cq.

All analysis made extensive use of the tidyverse and ggplot2 (28, 29).

3 Results

3.1 Terminator effects on gene expression depend on cis-regulatory context

To investigate the context dependence of cis-regulatory regions, we created a library of 120 constructs (Figure 1A), containing all combinations of 6 promoters (including 5'UTR), 2 coding sequences, and 10 terminators (including 3'UTR). We selected promoters and terminators from native yeast genes spanning a variety of different expression patterns and functions (Table 1). To choose specific sequence lengths for the terminators of our constructs, we referred to published measurements of median 3'UTR length (21) because sequences that are necessary and sufficient for efficient transcriptional termination are found upstream of the termination site (30). For terminators with measured length under 200nt, we used the standardised parts length of 200nt from the YeastFab library. The 2 coding sequences (CDS) expressed mCherry or mTurquoise2 fluorescent proteins, which are bright fluorophores with only 30% amino acid identity (31, 32).

Measuring fluorescence with a plate reader showed that, as expected, promoter choice dominated overall protein output. We observed 100-fold changes in fluorescence between the 4 highest expressing promoters (Figure 1B) and the 2 lowest expressing promoters (Figure 1C) across both coding sequences. Expression from the stress-induced pHSP26 was notably more variable than from pPGK1 and pRPS's, across 6 biological replicates, when combined with both coding sequences and a variety of terminators. We also confirmed that most differences in protein outputs are accounted for by changes in mRNA abundance by checking a subset of mCherry constructs using RT-qPCR ($R = 0.888$, Figure 1D).

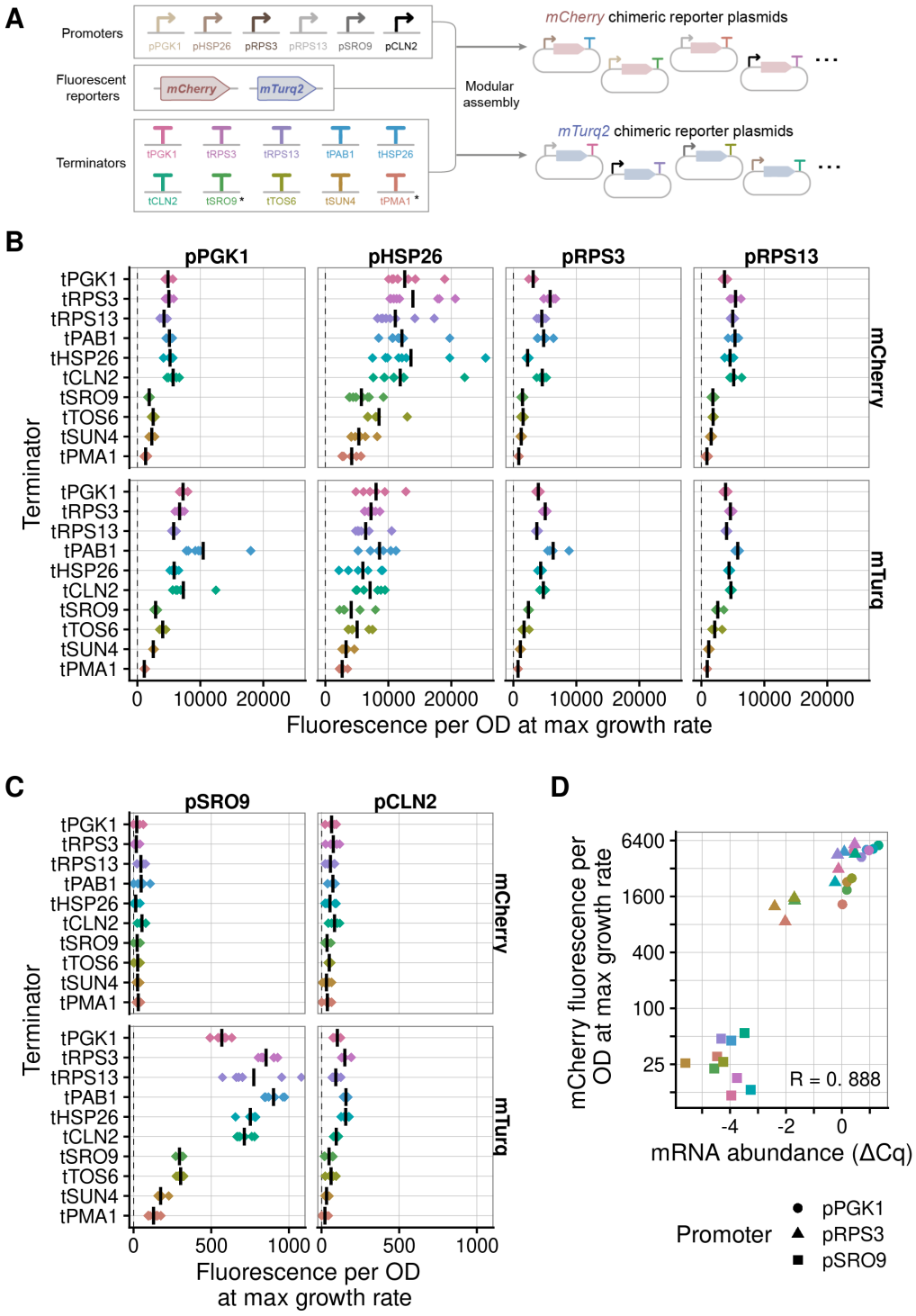


Figure 1: Both terminator and promoter contribute to gene expression. (**A**) Design of chimeric reporter plasmids with all combinations of 6 promoters, 2 fluorescent proteins, and 10 terminators. (**B**) Protein abundance estimated by mCherry and mTurquoise2 (mTurq) fluorescence measured at the time of each sample's maximum growth rate for 10 terminators paired with 4 high expressing promoters. Each diamond represents a biological replicate, averaged over 3 technical replicates. The vertical line is the mean of all 6 biological replicates. (**C**) Same as panel B, but for 2 low expressing promoters. Negative fluorescence values arising from instrument noise dominating measurements of constructs with negligible fluorescence are automatically set to 0. (**D**) mCherry fluorescence correlates with RT-qPCR mRNA abundance for 3 promoters paired with all 10 terminators. Note that both axes use a log_2 scale.

Table 1: Summary of the terminator library. The common gene name from which the terminator is extracted from is included alongside its systematic name. The median 3'UTR reported by (21) is a median over the lengths of each distinct isoform they detect. The usage of each motif is signified by S, for promoter-terminator swaps, or M, for motif insertion or deletion. A short summary of the protein function is included.

Gene Name	Systematic Name	Median 3'UTR Length	Construct Terminator Length	Usage	Function
PGK1	YCR012W	158	189	S	Glycolysis
RPS3	YNL178W	86	200	M&S	Ribosomal
RPS13	YDR064W	92	200	S	Ribosomal
PAB1	YER165W	150	200	S	RNA Binding
HSP26	YBR072W	164	200	S	Heat Shock
CLN2	YPL256C	203	200	S	Cell Cyclin
SRO9	YCL037C	543	545	S	RNA Binding
TOS6	YNL300W	256	256	S	Cell Wall
SUN4	YNL066W	198	198	S	Cell Wall
PMA1	YGL008C	421	421	S	Transmembrane ATPase
TSA1	YML028W	112	219	M	Redox Homestasis
PIR1	YKL164C	235	358	M	Cell Wall

Terminators also affect protein output with 5-fold changes in fluorescence seen within the same promoter-CDS sets, relative to the tPGK1 terminator of each group (Figure 2A). We focus on constructs with high-expression promoters due to the poor signal to noise ratio at low expression levels. The interaction of coding sequence and terminator is seen most clearly for tPAB1. tPAB1 is consistently the most highly expressed terminator in mTurq constructs, but is more variable in mCherry constructs. Meanwhile, tPGK1 highlights the interactions of promoter and terminator. tPGK1 is one of the most highly expressing terminators when paired with pPGK1 and pHSP26, but is up to 40% lower in expression when paired with pRPS3 or pRPS13. Overall, our results show that the contributions of terminators (including 3'UTRs) to gene expression depend on other parts within the gene.

We further investigated the effects of terminators on mRNA levels by comparing constructs with extended vs truncated terminators. It is known that disrupting the transcription termination signal lowers expression (6, 30). However, standardised parts libraries often assume a fixed terminator length for all genes, which is likely to omit the termination signal for genes with longer terminators. In the case of the YeastFab parts library (33) the fixed terminator length is 200bp. We compared a gene with a median 3'UTR length less than 200bp, RPS3 to a gene with a median 3'UTR length greater than 200bp, SRO9. The median length of the native RPS3 3'UTR is 86nt (21); truncating the terminator to 86bp or 59bp reduces protein output by almost 2-fold (Figure 2B). The median length of the native SRO9 3'UTR is 543nt (21); extending the terminator length to 543bp increases protein output by almost 2-fold (Figure 2B). This validates our assay's ability to detect known regulatory signals affecting transcription termination, while highlighting the importance of using well-informed annotations to construct parts libraries for synthetic biology. Note that we used the longer 543bp SRO9 terminator, and a similarly extended 421bp PMA1 terminator, for the main set of constructs (Figure 1).

3.2 Candidate cis-regulatory elements contribute to transcript decay rates

Next, we investigated how the regulatory effects of CREs contained within terminator regions depend on their context. First, 69 suitable CREs to test for context dependence were found through a literature search. All were suspected sequence motifs for mRNA binding proteins, several directly associated with proteins involved in mRNA degradation (7, 8, 10). Any motifs that were found in fewer than 6 gene 3'UTRs, as annotated by (21), were removed.

We quantified the regulatory effects of the remaining 38 candidate motifs by applying a linear model predicting half-life to 2 recent transcriptome-wide analyses of mRNA decay that used metabolic labeling, from Sun et al., (34) and Chan et al., (25). These datasets are loosely correlated in their half-life measurements across

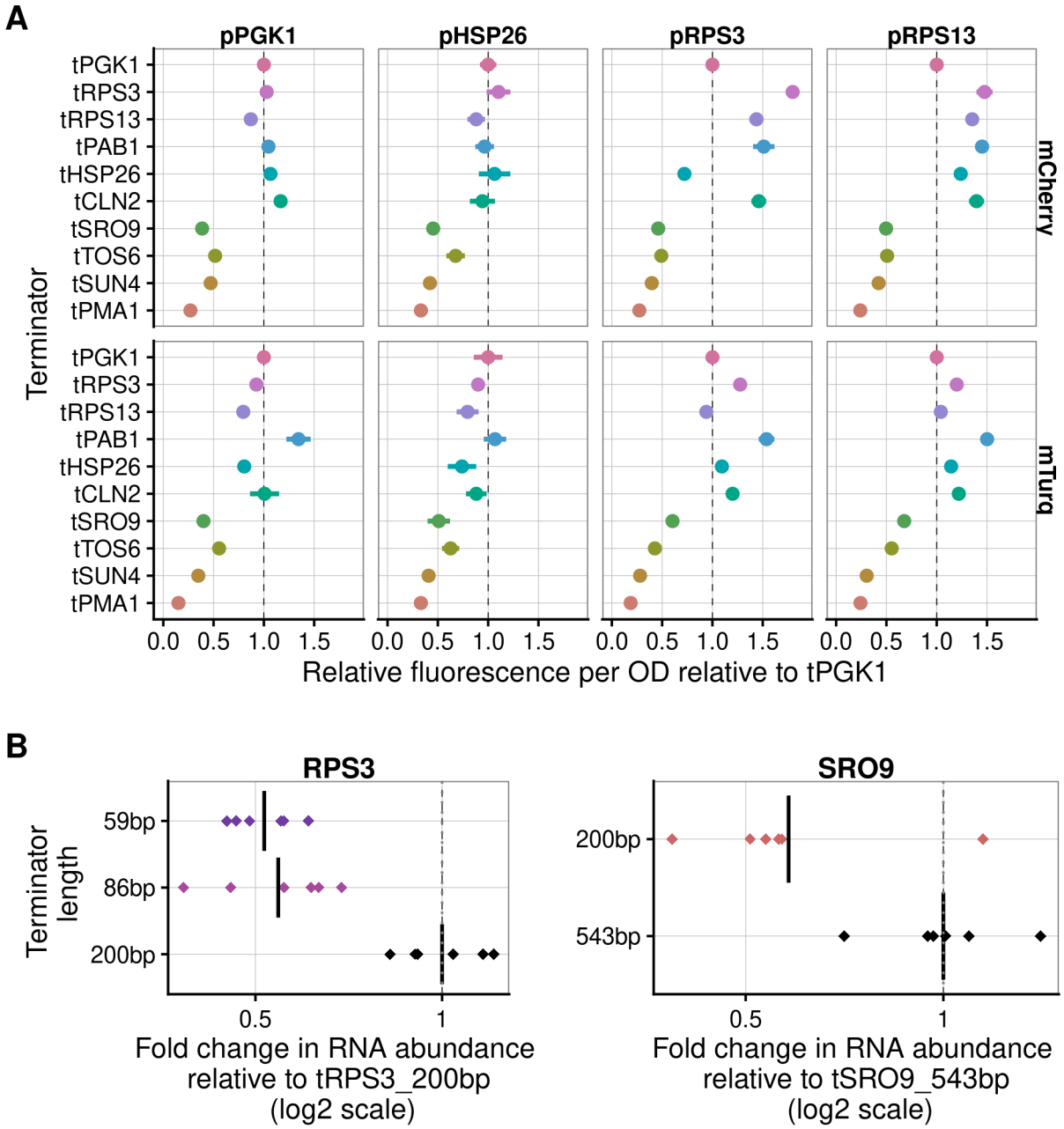


Figure 2: Terminator contributions to gene expression are promoter, coding sequence and length dependent. (**A**) Relative protein abundance from each terminator, normalized to a reference terminator tPGK1 for each matched promoter. The plot shows statistical summaries (mean and standard error) of 6 replicates for high-expression promoter data shown in Figure 1A. (**B**) RT-qPCR mRNA results targeting the mCherry ORF for pRPS3-mCherry-tRPS3 and pSRO9-mCherry-tSRO9 constructs with differing terminator lengths.

Table 2: Summary of shortlisted motif characteristics. The first 3 columns hold the consensus sequence for each motif and the exact versions deleted from or inserted into the host terminators. Then, we report the paper from which the motif was selected from; Shalgi et al., (7), Hogan et al., (10), or Cheng et al., (8). Next, we report each motif's maximum likelihood estimate by the linear model fit to two half-life datasets Chan et al., or Sun et al., (25, 34). Finally, the table includes notes on motif functions.

Consensus Seq	Inserted Motif	Deleted Motif	Source	Chan Coef	Sun Coef	Notes
ATATTC	ATATTC	ATATTC	Cheng	-0.075	-0.169	Decay motif
TGTAHMNTA	TGTACAATA	TGTACATTA	Hogan	-0.227	-0.056	Puf4p binding motif
GTATACCTA	GTATACCTA	GTATACCTA	Shalgi	0.502	0.283	Unknown
HWNCATTWY	TTTCATTTC	CTTCATTTC ATACATTAT AATCATTAT	Hogan	-0.084	-0.061	Khd1/Hek2 associated motif

4188 genes reported in both datasets, $R = 0.63$ (Figure 3A). However, Chan et al., used multiple time points to determine half-lives and estimated substantially smaller half-lives. Chan et al., also had greater coverage of genes in the yeast genome, 5529 vs 4304. Following Cheng et al., (8), we constructed a linear model to predict a transcript's half-life using the counts of motifs in its 3'UTR, the length of the 3'UTR, and the relative codon usage in each transcript's coding sequence (see Material and Methods). The linear model performed similarly on both datasets by explaining 44% and 41% of the variability in half-lives for the Chan et al., and Sun et al., datasets respectively (Figure 3C). This predictive power is comparable to the squared correlation between the datasets ($R^2 = 0.40$). Motifs that did not significantly contribute to the model were automatically filtered out using a greedy algorithm maximising the Akaike information criterion (AIC) during both training stages. Approximately 1.7% of the variance is explained by 7 significant motifs, with 42.0% explained by codon usage (Supplementary Table 1), consistent with previous analyses (8, 35). The top 7 most significant motifs from the Chan et al., data showed similar regulatory behaviour when tested on their own in the Sun et al., data, except for TGTAAATA which was stabilising in one dataset and destabilising in another, as we later discuss (Figure 3B).

We selected 4 motifs for exploring context dependence: TGTAHMNTA, GTATACCTA, HWNCATTWY, and ATATTC (Table 2). TGTAHMNTA and GTATACCTA were chosen as they had the largest coefficients amongst significant decay and stability motifs, respectively. HWNCATTWY was chosen due to its statistically significant effect in both datasets and, as it co-occurs with TGTAHMNTA in 68 native 3'UTRs, because it could be used for testing motif interactions. The final selected motif was ATATTC, as it is a statistically significant decay motif in both datasets, and it has been previously shown to lower mRNA abundance when inserted in reporter constructs (8). Functionally, TGTAHMNTA is the binding motif for Puf4p, and HWNCATTWY is associated with Khd1p/Hek2p-bound transcripts (10). However, it is not known how ATATTC and GTATACCTA affect mRNA decay.

3.3 Quantification of differential expression due to motif insertion or mutagenesis in multiple 3'UTRs

To quantify the effects and composability of selected motifs in different contexts, we designed a further set of reporter constructs (Figure 4A). Insertion reporter constructs contain combinations of motifs inserted in each of 2 host terminators, expressing mCherry from each of 3 different promoters. We first chose the ribosomal protein terminator tRPS3, as it was the only terminator in our initial library that did not contain any of the selected motifs. We selected thioredoxin peroxidase tTSA1 as the second host terminator because it also lacks selected motifs and has similar length to tRPS3. In each host terminator we chose 3 motif insertion sites, selecting for: minimum impact on transcript secondary structure, avoiding known transcription termination elements, and matching the positions of motifs in native genes. Having 3 insertion sites enabled us to quantify combinations of motifs, including duplicates of weaker motifs to increase the likelihood of detecting a clear effect on gene expression. We chose TGTACAATA and TTTCATTTC sequences as explicit versions of the TGTAHMNTA and HWNCATTWY consensus motifs respectively, and checked that these explicit versions have similar predicted effects on half-life transcriptome-wide (Supplementary Tables 5, 6). Altogether, 7 variant terminators were designed for these 2 host terminators: the wildtype terminator, a control to test the insertion sites with randomly generated sequences, 4 testing the effects of inserting each motif individually and a final variant to test interactions between the TGTAHMNTA and HWNCATTWY motifs. We created

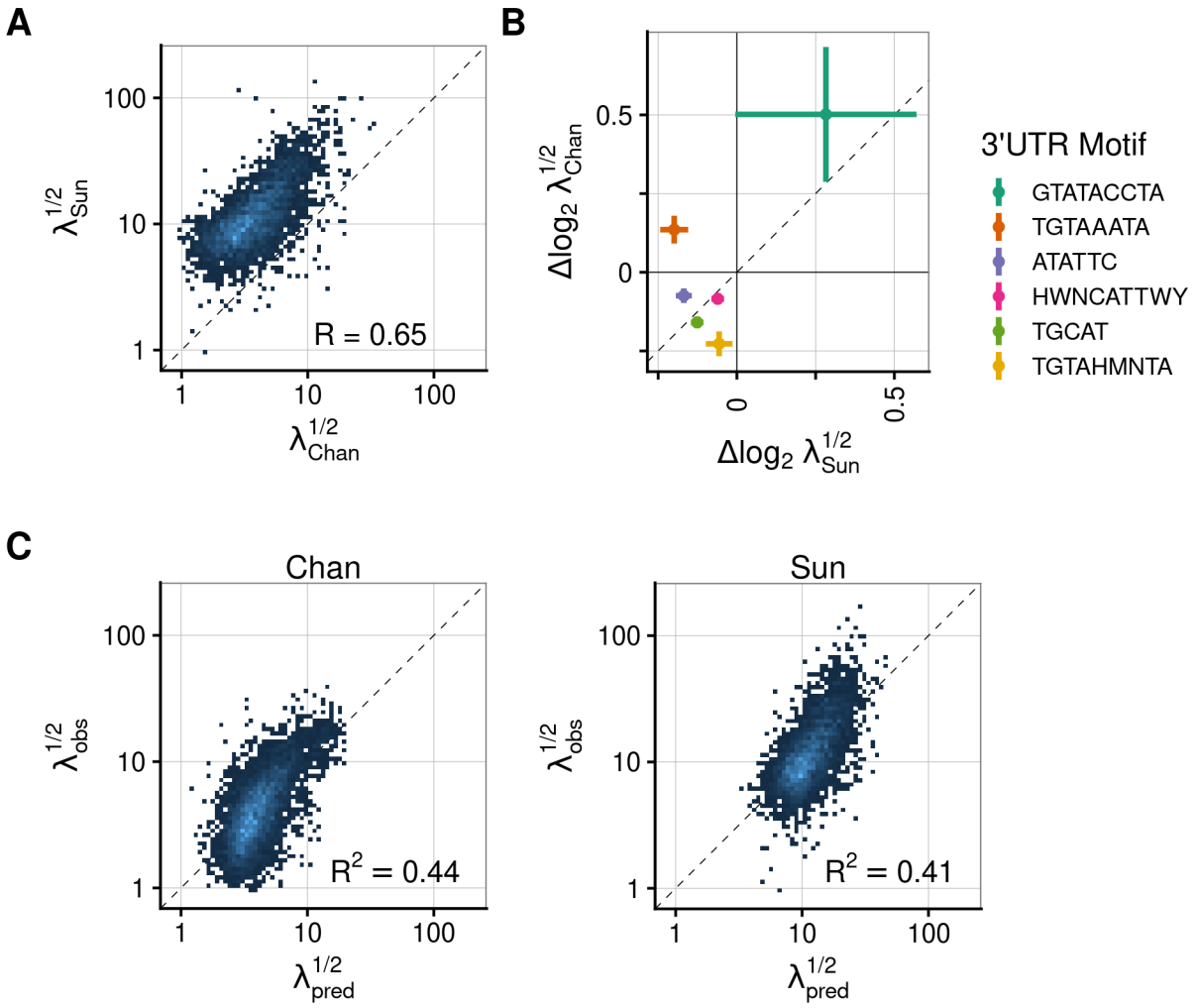


Figure 3: A linear model of transcript half-life quantifies the effect of candidate terminator motifs on half-life. (**A**) Correlation between the 2 transcript half-lives (λ), in minutes, reported in the two half-life datasets (25, 34). (**B**) Predicted contributions to log2 half-life for chosen motifs in the half-life datasets. (**C**) Predicted vs actual transcript half-lives calculated by a linear model of codon and motif usage trained on half-life datasets.

a construct library by pairing each terminator with three different promoters; its native promoter pairing (pRPS3 or pTSA1), the high-expression promoter pPGK1, and the low-expression promoter pSRO9.

Motifs predicted to affect half-life affect mRNA abundance when inserted into tRPS3 in reporter constructs (Figure 4B). We measured mRNA abundance by RT-qPCR across 6 biological replicates, each quantified in 3 technical replicates and normalised by the ΔCq method against values from 3 reference mRNAs (see methods). Insertion of 2 copies of ATATTC (mod_NAA) generally lowers the mRNA abundance, as much as 4-fold when paired with the pRPS3 promoter. Insertion of either TGTAHMNTA (mod_NTN), or 2 copies of HWNCATTWY (mod_HNH), tends to decrease mRNA abundance, and their combined insertion (mod_HTH) tends to decrease mRNA abundance even further. The putative stability motif GTATACCTA (mod_NGG) does not consistently or strongly affect mRNA abundance. However, comparison of the WT and control (mod_NNN) terminators does show that the creation of the insertion sites alone has an effect on mRNA levels.

Inserting the same motifs into our second host terminator gives qualitatively similar results (Figure 4C). Decay motifs generally lead to decay, although ATATTC (mod_NAA) has a weaker effect in tTSA1 than in tRPS3, and TGTAHMNTA (mod_NTN) has a stronger effect in tTSA1 than tRPS3. The putative stability motif GTATACCTA (mod_NGG) again has little effect.

We next quantified the effects of removing decay motifs from a native yeast terminator. We selected the cell wall protein PIR1 as our host terminator as it is a short and de-novo-synthesizable terminator that contains the ATATTC, TGTAHMNTA, and HWNCATTWY motifs. We designed 8 terminators in which the motif occurrences in tPIR1 were replaced by scrambled sequences (Figure 5A). We found that the removal of almost any decay motif from tPIR1 results in an increase in mRNA levels (Figure 5B).

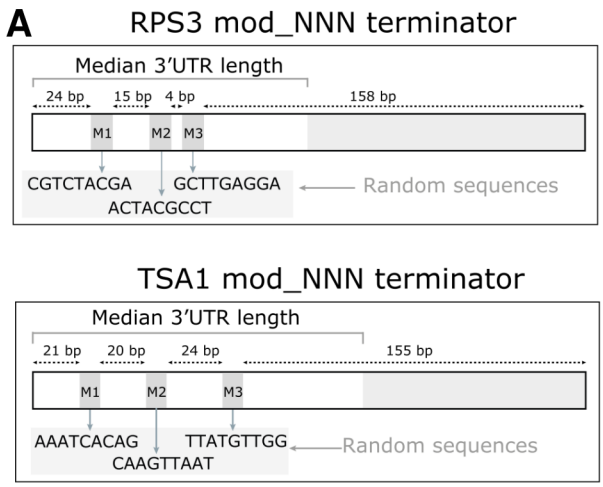
We confirmed that motif-dependent changes in mRNA abundance are reflected in protein abundance by measuring the fluorescence from a subset of reporter constructs with native promoter-terminator pairings (Supplementary Figure 1). The high correlation ($R = 0.96, 0.68, 0.96$ for tRPS3, tTSA1, tPIR1 constructs respectively) demonstrates that these combinations of decay motifs that change mRNA abundance also change the protein output, as expected.

Comparison of mRNA abundance across all constructs (Figure 4B, 4C, 5B) shows motif contributions change in magnitude but not direction depending on the context of the rest of the construct. The insertion of almost any decay motif into tTSA1 or tRPS3 results in a decrease in mRNA abundance, and removal of these from from tPIR1 results in an increase in mRNA abundance. However, the quantitative effects vary depending both on the immediate motif context in the host terminator and on the more distant context given by the promoter.

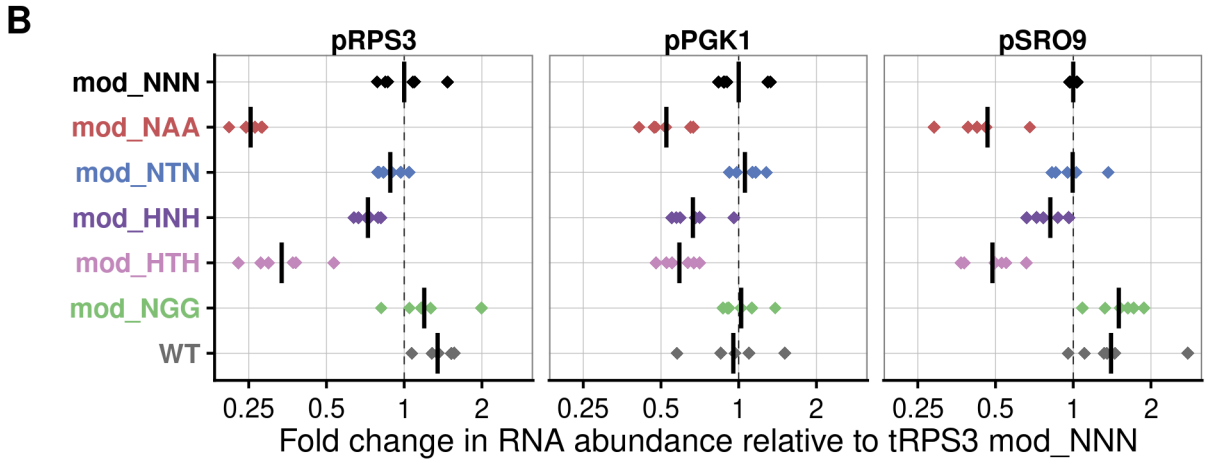
3.4 Motif effects on gene expression depend both on terminator context and promoter pairing

We compared the effects of cis-regulatory motifs on mRNA abundance to predicted effects inferred from the transcriptome-wide measurements of half-life. First, we trained a linear model using the RT-qPCR results to estimate the change in log2 mRNA abundance (i.e. ΔCq) due to the presence of a motif in each promoter and terminator combination. Using a simple model of transcript production and decay, we can argue that changes in mRNA abundance are directly proportional to changes in mRNA half-life (see methods). Therefore, we compared the motif's estimated change in log2 mRNA abundance to that predicted due to changes in log2 half-life, estimated from our transcriptome-wide analysis of the (25) dataset. The effect of motifs in reporter constructs is correlated with the predictive model, but the strength of the correlation depends on context (Figure 6A). Constructs with inserted motifs (i.e. tRPS3 and tTSA1 constructs) had a lower correlation with predicted effects than constructs with deleted motifs (i.e. tPIR1 constructs). Interestingly, motif effects on mRNA abundance appear to be greater than that predicted from their effect on half-life when their host terminator is paired with its native promoter.

We next directly compared the estimated coefficients for each motif's effect on mRNA abundance across promoter-terminator pairing (Figure 6B). The effect of a motif depends on terminator context. For example, ATATTCA reduces mRNA abundance substantially more when inserted in tRPS3 than in tTSA1. Meanwhile, TGTAHMNTA significantly reduces mRNA abundance when inserted in tTSA1, but not tRPS3, whichever promoter is chosen. Promoter choice also influences the magnitude of a motif's contribution to mRNA levels. For the ATATTC, TGTAHMNTA and HWNCAUUWY motifs the greatest reduction in mRNA abundance occurred when native promoter-terminator pairings are measured. This is true for all 3 decay motifs across all 3 host terminators, except for HWNCAUUWY in pRPS3-tRPS3 constructs.



Terminator	Motif 1	Motif 2	Motif 3
mod_NNN	Random	Random	Random
mod_NAA	Random	ATATTC	ATATTC
mod_NTN	Random	TGTAHMNTA	Random
mod_HNH	HWNCATTWY	Random	HWNCATTWY
mod_HTH	HWNCATTWY	TGTAHMNTA	HWNCATTWY
mod_NGG	Random	GTATACCTA	GTATACCTA
WT	NA	NA	NA



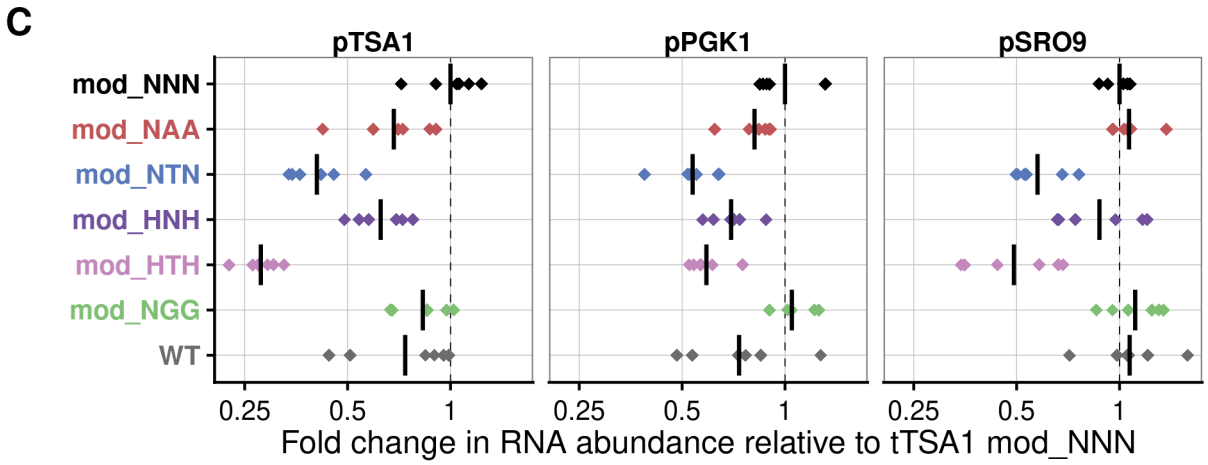
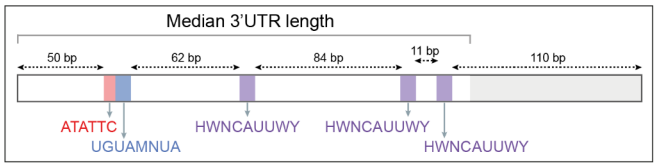


Figure 4: Motifs inserted into RPS3 and TSA1 host terminators change transcript abundance in RT-qPCR measurements. (**A**) Design of motif insertion sites in native RPS3 and TSA1 terminators, highlighting random insertion used as a negative control. (**B**) Fold changes in transcript abundance for tRPS3 constructs paired with 3 promoters; pRPS3, pPGK1 and pSRO9. (**C**) Fold changes in transcript abundance for tTSA1 constructs paired with three promoters; pTSA1, pPGK1 and pSRO9. Each diamond represents a biological replicate, averaged over 3 technical replicates. The vertical line represents the mean of all 6 biological replicates. Fold changes are relative to the abundance of the mod_NNN construct, i.e. $2^{\Delta\Delta Cq}$ (see methods).

A



B

Terminator	Motif 1	Motif 2	Motif 3	Motif 4	Motif 5
WT	ATATTC	TGTAHMNTA	HWNCATTWY	HWNCATTWY	HWNCATTWY
mod_NTHHH	Random	TGTAHMNTA	HWNCATTWY	HWNCATTWY	HWNCATTWY
mod_ANHHH	ATATTC	Random	HWNCATTWY	HWNCATTWY	HWNCATTWY
mod_ATNHH	ATATTC	TGTAHMNTA	Random	HWNCATTWY	HWNCATTWY
mod_ATHNH	ATATTC	TGTAHMNTA	HWNCATTWY	Random	HWNCATTWY
mod_ATNNN	ATATTC	TGTAHMNTA	Random	Random	Random
mod_ANNNN	ATATTC	Random	Random	Random	Random
mod_NTNNN	Random	TGTAHMNTA	Random	Random	Random

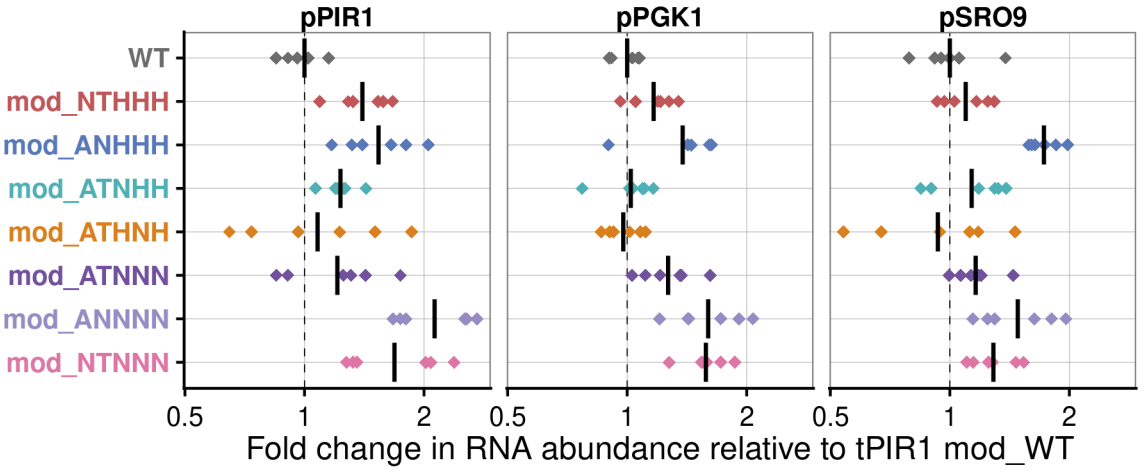


Figure 5: Motifs removed from PIR1 host terminators change transcript abundance in RT-qPCR measurements. (**A**) Design of PIR1 constructs with combinations of motifs replaced by random nucleotide sequences. (**B**) Fold changes in transcript abundance for tPIR1 constructs paired with 3 promoters; pPIR1, pPGK1 and pSRO9. Each diamond represents a biological replicate, averaged over 3 technical replicates. The vertical line represents the mean of all 6 biological replicates. Fold changes are relative to the abundance of the WT construct, i.e. $2^{\Delta\Delta Cq}$ (see methods).

Regulatory interactions between different motifs also change depending on host terminator and promoter context. We included an interaction term that quantifies how the effect of including both HWNCAUUWY and TGTAHMNTA together differs from the sum of the effects of including these motifs individually. The combination of TGTAHMNTA and HWNCAUUWY in tRPS3 has no additional effect beyond a simple sum of their individual effects when paired with pPGK1. However, when tRPS3 is paired with pRPS3 or pSRO9, the combination has a greater effect than expected from the sum of each motif's individual effect. The combination of TGTAHMNTA and HWNCAUUWY in tTSA1 has no additional effect, except when paired with pPGK1, where it has a lesser effect than expected. Finally, the combination of TGTAHMNTA and HWNCAUUWY in tPIR1 always has an additional effect across the three promoters. The combination has a greater effect than expected when paired with pPIR1 but a lesser effect when paired with pPGK1 or pTSA1.

4 Discussion

Our measurements of protein output from promoter-terminator combinations highlight the limitations of assuming that CREs are quantitatively composable. Unsurprisingly, fluorescent protein expression levels are dominated by promoter choice, up to 100 fold, and then terminator choice, up to 5 fold. However, our observations of 1.5 fold change in the relative effect of terminators depending on coding sequence and promoter choice highlight the quantitative limitations of the assumption of composability. Previous work by (16) suggests that screening a larger library of CREs would reveal a subset with stronger quantitative interactions.

Terminators taken from co-regulated genes have similar behaviour in our reporter constructs. Terminators that have consistently low expression include tSUN4 and tTOS6, taken from cell wall proteins whose transcripts are bound and regulated by shared RNA-binding proteins (10). Terminators have similar effects when paired with either of the two ribosomal protein promoters pRPS3 or pRPS13, which is consistent with the co-regulation of ribosomal proteins in cells. Further investigations are required into the shared sequence signals and proteins that explain these regulatory patterns and affect the composability of these elements.

We find that the effect of short CREs in 3'UTRs depends on the context of their host construct. The contribution of our selected cis-regulatory motifs to mRNA stability, as inferred from transcriptome-wide analysis, correlates with their effect on the abundance of reporter constructs. However, their contributions to mRNA abundance were inconsistent and dependent on context: TGTAHMNTA has no measurable effect when inserted into tRPS3 but the expected effect in tTSA1, ATATTC has the expected effect in tRPS3 but little effect in tTSA1, and HWNCATTWY can either decrease or increase mRNA levels when removed from tPIR1, depending on the promoter. Also, the two tPIR1 constructs with different mutated HWNCATTWY motifs had different expression levels suggesting HWNCATTWY has a position dependent effect. Although, the exact sequences of the two motifs did also differ by 4 nucleotides. In addition, our original motif search predicted a TGTAAATA motif to be stabilising in the (34) dataset but destabilising in the (25) dataset. TGTAAATA is a known binding motif for Puf3 and carbon deficient conditions trigger Puf3 to degrade its targets (36). Carbon deficient conditions were likely present due to the bespoke CSM-lowURA media used by (25), which is similar to the SC-URA media used in our assays, contrasting with carbon-rich YPD media used by (34). Thus, carbon-dependent regulation by Puf3 could explain the different behaviour of TGTAAATA in the different datasets. These disparities in predicted and experimentally verified regulatory behaviour raises questions about current methods for discovering novel cis-regulatory elements and, more generally, for understanding and manipulating regulatory pathways.

4.1 Mechanisms for motif terminator context dependence

This study documents CRE variability with local context, but CRE behavior is also known to vary temporally and spatially in cells, further supporting that CRE functions are not generally decomposable. Our results support long-standing mechanistic observations that CRE contributions depend on the presence of other CREs in the host gene. However, CREs are often regulated by cognate trans-regulatory elements (TREs) that are spatiotemporally regulated themselves. These effects are not separable, because spatiotemporal regulation of gene expression conferred by one CRE can change the interaction of a second CRE with its relevant TREs. For example, a motif may act as the regulatory binding site of a TRE only after the transcript is localised to a time and place where that TRE is present, as we discuss below.

Temporal regulation by successive CREs is common in developmental processes. For example, in *C. elegans* lin-4 is a non-coding RNA gene crucial for regulating cell fates during the early stages of larval development (37). Lin-4 is a small RNA that acts as a TRE by binding to a repeated motif in the 3'UTR of its target

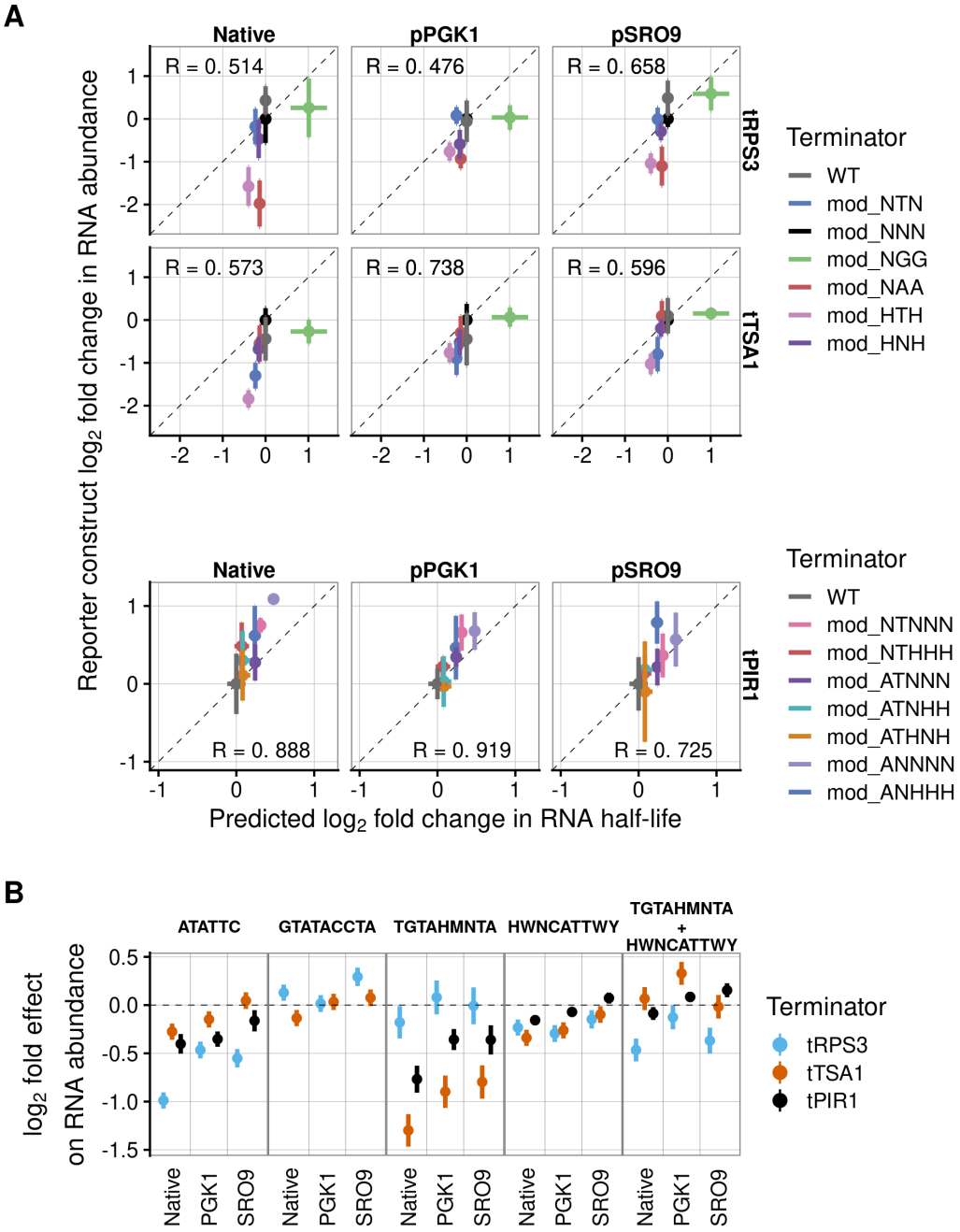


Figure 6: Promoter and terminator context alter the regulatory effect of motifs. (**A**) Predicted change in transcript abundance inferred from transcriptome-wide motif contributions to half-life, compared to RT-qPCR measurements of reporter transcript abundance. The y-axis shows statistical summaries (mean and standard error) of 6 replicates for data shown in figures 4 and 5. The fold change is relative to the mod0 construct in each promoter-terminator pairing for the tRPS3 and tTSA1 sets, and relative to the WT constructs for tPIR1 sets. Native promoter panels show the promoter paired with the terminator from the same set, e.g. pRPS3 with tRPS3. (**B**) Motif contributions to fold changes in mRNA abundance for reporter constructs with different promoter and terminator contexts. This is calculated by a linear model with a coefficient for the effect of each motif in each set, applied to ΔCq against 3 reference genes. The last column shows the interaction term between TGTAHMNTA and HWNCATTWY.

mRNA lin-14, and inhibiting the translation of lin-14 (38). Since lin-4 is only expressed at the end of the first larval development stage, the 3'UTR motifs in lin-14 only begin to inhibit its translation late into stage 1, initiating the start of stage 2 (39). The regulation of lin-4 by lin-14 binding to the 3'UTR thus depends on the temporal expression pattern encoded by lin-4's promoter. Similarly, to establish meiotic chromosome segregation in budding yeast, mRNA encoding cyclin CLB3 is transcribed in stage I of meiosis, but is translationally repressed until stage II of meiosis. CLB3 is translationally repressed via its 5'UTR by the RNA-binding protein Rim4. During the transition to meiosis II, Rim4 is phosphorylated which inhibits binding to CLB3 and enables CLB3 to be translated (40). Here, 5'UTR-specified translational control of CLB3 by Rim4 depends on the timing of promoter-specified transcriptional control.

Spatial regulation by successive CREs is also important, as one CRE can facilitate transcript localisation whilst other CREs and/or TREs confer localisation-dependent regulation. In yeast, the Ash1 protein represses mating-type switching only in daughter cells (41), due to localisation of the ASH1 transcript at the bud tip and subsequent localised translation (42). It is thought that co-transcriptional recruitment of She2 to the ASH1 transcript in the nucleus enables the later recruitment of cytoplasmic factors Khd1/Hek2 and Puf6, factors known for translational repression. Furthermore, successful transport of ASH1 to the bud tip by She2-She3-Myo4 complexes depends on translational repression by Khd1 and Puf6. Later, phosphorylation of Khd1 and Puf6 by bud-membrane-localised kinases leads to localised translational activation of the ASH1 mRNP (43, 44).

Another example where the effect of a CRE on a transcript depends on co-localisation with a regulatory kinase comes from the fungal RNA-binding protein Ssd1. Yeast cell wall proteins such as Sun4 and Tos6 are translationally repressed by Ssd1 (45), which binds the 5'UTRs of their transcripts (10, 46). It is thought that these transcripts are translationally activated at bud sites after the phosphorylation of Ssd1 by a localised kinase, Cbk1 (45). There is no evidence that Ssd1 directly acts to transport RNA, so this localised activation presumably depends on localisation that is conferred by other CREs, that recruit other RNA-binding proteins to those transcripts (10), that then in turn recruit transport machinery. We note that post-transcriptional regulation of SUN4 and TOS6 might explain the repressive effect of their terminators when paired with high-expression promoters (Figure 1).

Secondary structures can be integral to motif function, and secondary structure depends on sequences beyond a motif. Again for ASH1, key cis-regulatory elements depend on secondary structure (47). Secondary structure is also essential for Smg recognition elements found in transcript targets of Vts1p (48). In mammalian cells, secondary structures have an overall inhibitory effect on motif functionality (15). In yeast, the presence of stable stem loops within the 3'UTR reduces the binding of known destabilising RBPs (49).

Motif behaviour also depends on CREs across the entire host transcript. Regulation of the ASH1 transcript depends on the concerted action of multiple CREs (50, 51). In other cases, multiple motifs, or multiple copies of the same motif, must occur together to enable recruitment of the effector protein. Recruitment of Puf3 to promote mRNA transcript decay requires two Puf3 motifs in the 3'UTR (52). Indeed, clusters of short degenerate motifs have been proposed as a general regulatory feature of RNA-binding proteins (53)

4.2 Implications for novel cis-regulatory element searches

Motif detection algorithms need to progress beyond the assumption that the presence of a motif alone ensures its functional effect. The effect of a motif can be influenced by factors such as growth conditions and time of measurement. Uncertainty around the active state of a motif could be introduced using latent variables, as used in transcription factor (TF) binding site prediction algorithms to represent the proportion of active TF sites (54, 55). Alternatively, non-linear relationships between factors affecting motif activity can be modeled by partition regression techniques, such as random forests. Again looking at TF algorithms, partitioning algorithms could be used to predict functional datasets using short sub-sequences of transcripts (56).

Reductionist search algorithms that attempt to discover the shortest possible sequence contributing to expression may limit the discovery of novel CREs. High resolution maps of protein-RNA interactions are revealing RBPs with gapped, multi-partite motifs (46) or motifs that must be repeated in the same transcript to be effective (52, 57). Common motif searching approaches, such as MEME Suite (13), have extended beyond fixed length, gapless motifs but may still assume motifs occur independently and only once per transcript (58). Although reducing the search constraints to allow longer, gapped motifs could increase the false positive rate, the availability of precise interactome datasets suggests a need for more flexible search algorithms that effectively discover complex motifs.

4.3 Implications for engineering biology

Our results support previous studies showing unpredictability in the expression of combinations of otherwise well-characterised regulatory elements. In the creation of biological circuits, for example, time-consuming directed evolution assays have been implemented to overcome mis-matches in component expression levels (59). The availability of extensively characterised parts libraries now enables large scale searches for CREs with suitable expression. Together with high-throughput assays, pools of combinations of suitable CREs can be quickly screened for the correct expression levels (16).

Our conceptual model of a gene influences the tools we make, and can limit our ability to predict and manipulate gene expression (60). Conceptual models of CREs as independently acting composable parts exemplify these limits. The characterisation of an element's function in a single context is not enough to make generalisable conclusions. Both small- and large-scale studies of CREs should ideally be carried out with multiple combinations of other regulatory elements, in cells in multiple growth conditions, to build a comprehensive understanding of transcriptional and post-transcriptional regulation.

5 Funding

This work was supported by The Wellcome Trust [grant number 208779/Z/17/Z to E.W.J.W.].

6 Acknowledgments

We thank Naomi Nakayama, Erika Szymanski, Filippo Menolascina, and members of the Wallace lab for discussions on the manuscript.

7 References

1. Curran, K.A., Karim, A.S., Gupta, A. and Alper, H.S. (2013) Use of expression-enhancing terminators in saccharomyces cerevisiae to increase mRNA half-life and improve gene expression control for metabolic engineering applications. *Metabolic Engineering*, **19**, 88–97.
2. Lee, M.E., DeLoache, W.C., Cervantes, B. and Dueber, J.E. (2015) A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synthetic Biology*, **4**, 975–986.
3. Garcia-Ruiz, E., Auxillos, J., Li, T., Dai, J. and Cai, Y. (2018) YeastFab: High-throughput genetic parts construction, measurement, and pathway engineering in yeast. *Methods in enzymology*, **608**, 277–306.
4. Andreou, N., Andreas I. AND Nakayama (2018) Mobius assembly: A versatile golden-gate framework towards universal DNA assembly. *PLOS ONE*, **13**, 1–18.
5. Sharon, E., Kalma, Y., Sharp, A., Raveh-Sadka, T., Levo, M., Zeevi, D., Keren, L., Yakhini, Z., Weinberger, A. and Segal, E. (2012) Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nature biotechnology*, **30**, 521–530.
6. Shalem, E.A.L., Ophir AND Sharon (2015) Systematic dissection of the sequence determinants of gene 3' end mediated expression control. *PLOS Genetics*, **11**, 1–21.
7. Shalgi, R., Lapidot, M., Shamir, R. and Pilpel, Y. (2005) A catalog of stability-associated sequence elements in 3' UTRs of yeast mRNAs. *Genome biology*, **6**.
8. Cheng, J., Maier, K.C., Avsec, Ž., Petra, R.U.S. and Gagneur, J. (2017) Cis-regulatory elements explain most of the mRNA stability variation across genes in yeast. *RNA*, **23**, 1648–1659.
9. Vijayabaskar, M.S., Goode, D.K., Obier, N., Lichtinger, M., Emmett, A.M.L., Abidin, F.N.Z., Shar, N., Hannah, R., Assi, S.A., Lie-A-Ling, M., *et al.* (2019) Identification of gene specific cis-regulatory elements during differentiation of mouse embryonic stem cells: An integrative approach using high-throughput datasets. *PLoS Computational Biology*, **15**, e1007337.
10. Hogan, D.J., Riordan, D.P., Gerber, A.P., Herschlag, D. and Brown, P.O. (2008) Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *PLoS Biology*, **6**, e255.
11. Kretz, M., Siprashvili, Z., Chu, C., Webster, D.E., Zehnder, A., Qu, K., Lee, C.S., Flockhart, R.J., Groff, A.F., Chow, J., *et al.* (2013) Control of somatic tissue differentiation by the long non-coding RNA TINCR. *Nature*, **493**, 231–235.
12. Elemento, O., Slonim, N. and Tavazoie, S. (2007) A universal framework for regulatory element discovery across all genomes and data types. *Molecular Cell*, **28**, 337–350.
13. Bailey, T.L., Johnson, J., Grant, C.E. and Noble, W.S. (2015) The MEME suite. *Nucleic Acids Research*, **43**, W39–W49.

14. Ptashne, M., Jeffrey, A., Johnson, A.D., Maurer, R., Meyer, B.J., Pabo, C.O., Roberts, T.M. and Sauer, R.T. (1980) How the lambda repressor and cro work. *Cell*, **19**, 1–11.
15. Taliaferro, J.M., Lambert, N.J., Sudmant, P.H., Dominguez, D., Merkin, J.J., Alexis, M.S., Bazile, C.A. and Burge, C.B. (2016) RNA sequence context effects measured in vitro predict in vivo protein binding and regulation. *Molecular cell*, **64**, 294–306.
16. Kosuri, S., Goodman, D.B., Cambray, G., Mutalik, V.K., Gao, Y., Arkin, A.P., Endy, D. and Church, G.M. (2013) Composability of regulatory sequences controlling transcription and translation in escherichia coli. *Proceedings of the National Academy of Sciences*, **110**, 14024–14029.
17. Ito, Y., Yamanishi, M., Ikeuchi, A., Imamura, C., Tokuhiro, K., Kitagawa, T. and Matsuyama, T. (2013) Characterization of five terminator regions that increase the protein yield of a transgene in saccharomyces cerevisiae. *Journal of Biotechnology*, **168**, 486–492.
18. Gietz, R.D. and Woods, R.A. (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods in enzymology*, **350**, 87–96.
19. Lichten, C.A., White, R., Clark, I.B. and Swain, P.S. (2014) Unmixing of fluorescence spectra to resolve quantitative time-series measurements of gene expression in plate readers. *BMC biotechnology*, **14**, 1–11.
20. Swain, P.S., Stevenson, K., Leary, A., Montano-Gutierrez, L.F., Clark, I.B.N., Vogel, J. and Pilizota, T. (2016) Inferring time derivatives including cell growth rates using gaussian processes. *Nature Communications*, **7**, 1–8.
21. Pelechano, V., Wei, W. and Steinmetz, L.M. (2013) Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature*, **497**, 127–131.
22. Akaike, H. (1998) Information theory and an extension of the maximum likelihood principle. 10.1007/978-1-4612-1694-0_15.
23. R Core Team (2020) R: A language and environment for statistical computing R Foundation for Statistical Computing, Vienna, Austria.
24. Ripley, B. and Venables, B. (2002) Modern applied statistics with s 4th ed. Springer.
25. Chan, L.Y., Mugler, C.F., Heinrich, S., Vallotton, P. and Weis, K. (2018) Non-invasive measurement of mRNA decay reveals translation initiation as the major determinant of mRNA stability. *eLife*, **7**.
26. Lorenz, R., Bernhart, S.H., Siederdissen, C.H. zu, Tafer, H., Flamm, C., Stadler, P.F. and Hofacker, I.L. (2011) ViennaRNA package 2.0. *Algorithms for Molecular Biology*, **6**, 26.
27. Sánchez, Á. and Kondev, J. (2008) Transcriptional control of noise in gene expression. *Proceedings of the National Academy of Sciences*, **105**, 5081–5086.
28. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., *et al.* (2019) Welcome to the tidyverse. *Journal of Open Source Software*, **4**, 1686.
29. Wickham, H. (2016) ggplot2: Elegant graphics for data analysis Springer-Verlag New York.
30. Guo, Z. and Sherman, F. (1996) 3'-end-forming signals of yeast mRNA. *Trends in biochemical sciences*, **21**, 477–481.
31. Shaner, N.C., Campbell, R.E., Steinbach, P.A., Giepmans, B.N., Palmer, A.E. and Tsien, R.Y. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. Red fluorescent protein. *Nature biotechnology*, **22**, 1567–1572.
32. Goedhart, J., Von Stetten, D., Noirclerc-Savoye, M., Lelimousin, M., Joosen, L., Hink, M.A., Van Weeren, L., Gadella, T.W. and Royant, A. (2012) Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nature communications*, **3**, 1–9.
33. Guo, Y., Dong, J., Zhou, T., Auxillos, J., Li, T., Zhang, W., Wang, L., Shen, Y., Luo, Y., Zheng, Y., *et al.* (2015) YeastFab: The design and construction of standard biological parts for metabolic engineering in saccharomyces cerevisiae. *Nucleic acids research*, **43**, e88–e88.
34. Sun, M., Schwalb, B., Pirkl, N., Maier, K.C., Schenk, A., Failmezger, H., Tresch, A. and Cramer, P. (2013) Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. *Molecular Cell*, **52**, 52–62.
35. Presnyak, V., Alhusaini, N., Chen, Y.H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K.E., Graveley, B.R., *et al.* (2015) Codon optimality is a major determinant of mRNA stability. *Cell*, **160**, 1111–1124.
36. Miller, M.A., Russo, J., Fischer, A.D., Leban, F.A.L. and Olivas, W.M. (2014) Carbon source-dependent alteration of Puf3p activity mediates rapid changes in the stabilities of mRNAs involved in mitochondrial function. *Nucleic Acids Research*, **42**, 3954–3970.
37. Wightman, B., Ha, I. and Ruvkun, G. (1993) Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell*, **75**, 855–862.
38. Lee, R.C., Feinbaum, R.L. and Ambros, V. (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, **75**, 843–854.

39. Olsen, P.H. and Ambros, V. (1999) The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. *Developmental Biology*, **216**, 671–680.
40. Berchowitz, L.E., Gajadhar, A.S., Werven, F.J. van, Rosa, A.A.D., Samoylova, M.L., Brar, G.A., Xu, Y., Xiao, C., Futcher, B., Weissman, J.S., *et al.* (2013) A developmentally regulated translational control pathway establishes the meiotic chromosome segregation pattern. *Genes and Development*, **27**, 2147–2163.
41. Sil, A. and Herskowitz, I. (1996) Identification of an asymmetrically localized determinant, Ash1p, required for lineage-specific transcription of the yeast HO gene. *Cell*, **84**, 711–722.
42. Niednery, A., Edelmanny, F.T. and Niessing, D. (2014) Of social molecules: The interactive assembly of ASH1 mRNA-transport complexes in yeast. *RNA Biology*, **11**, 998–1009.
43. Paquin, N., Ménade, M., Poirier, G., Donato, D., Drouet, E. and Chartrand, P. (2007) Local activation of yeast ASH1 mRNA translation through phosphorylation of Khd1p by the casein kinase Yck1p. *Molecular Cell*, **26**, 795–809.
44. Deng, Y., Singer, R.H. and Gu, W. (2008) Translation of ASH1 mRNA is repressed by Puf6p-Fun12p/eIF5B interaction and released by CK2 phosphorylation. *Genes and Development*, **22**, 1037–1050.
45. Jansen, J.M., Wanless, A.G., Seidel, C.W. and Weiss, E.L. (2009) Cbk1 regulation of the RNA-binding protein Ssd1 integrates cell fate with translational control. *Current Biology*, **19**, 2114–2120.
46. Bayne, R.A., Jayachandran, U., Kasprowicz, A., Bresson, S., Tollervey, D., Wallace, E.W.J. and Cook, A.G. (2020) Yeast Ssd1 is a non-enzymatic member of the RNase II family with an alternative RNA recognition interface. *bioRxiv*, 10.1101/2020.10.22.350314.
47. Olivier, C., Poirier, G., Gendron, P., Boisgontier, A., Major, F. and Chartrand, P. (2005) Identification of a conserved RNA motif essential for She2p recognition and mRNA localization to the yeast bud. *Molecular and Cellular Biology*, **25**, 4752–4766.
48. Aviv, T., Lin, Z., Ben-Ari, G., Smibert, C.A. and Sicheri, F. (2006) Sequence-specific recognition of RNA hairpins by the SAM domain of Vts1p. *Nature Structural and Molecular Biology*, **13**, 168–176.
49. Geisberg, J.V., Moqtaderi, Z. and Struhl, K. (2020) The transcriptional elongation rate regulates alternative polyadenylation in yeast. *eLife*, **9**, 1–55.
50. Chartrand, P., Meng, X.H., Singer, R.H. and Long, R.M. (1999) Structural elements required for the localization of ASH1 mRNA and of a green fluorescent protein reporter particle in vivo. *Current Biology*, **9**, 333–338.
51. Gonzalez, I., Buonomo, S.B.C., Nasmyth, K. and Ahsen, U.V. (1999) ASH1 mRNA localization in yeast involves multiple secondary structural elements and ASH1 protein translation. *Current Biology*, **9**, 337–340.
52. Jackson, J.S., Houshmandi, S.S., Leban, F.L. and Olivas, W.M. (2004) Recruitment of the Puf3 protein to its mRNA target for regulation of mRNA decay in yeast. *RNA*, **10**, 1625–1636.
53. Cereda, M., Pozzoli, U., Rot, G., Juvan, P., Schweitzer, A., Clark, T. and Ule, J. (2014) RNAmotifs: Prediction of multivalent RNA motifs that control alternative splicing. *Genome Biology*, **15**, R20.
54. Dai, Z., Iqbal, M., Lawrence, N.D. and Rattray, M. (2017) Efficient inference for sparse latent variable models of transcriptional regulation. *Bioinformatics*, **33**, 3776–3783.
55. Sanguinetti, G., Lawrence, N.D. and Rattray, M. (2006) Probabilistic inference of transcription factor concentrations and gene-specific regulatory activities. *Bioinformatics*, **22**, 2775–2781.
56. Behjati Ardakani, F., Schmidt, F. and Schulz, M. (2019) Predicting transcription factor binding using ensemble random forest models [version 2; peer review: 2 approved]. *F1000Research*, **7**.
57. Gu, W., Deng, Y., Zenklusen, D. and Singer, R.H. (2004) A new yeast PUF family protein, Puf6p, represses ASH1 mRNA translation and is required for its localization. *Genes and Development*, **18**, 1452–1465.
58. Frith, M.C., Saunders, N.F.W., Kobe, B. and Bailey, T.L. (2008) Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Computational Biology*, **4**, e1000071.
59. Yokobayashi, Y., Weiss, R. and Arnold, F.H. (2002) Directed evolution of a genetic circuit. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 16587–16591.
60. Szymanski, E. and Scher, E. (2019) Models for DNA design tools: The trouble with metaphors is that they don't go away. *ACS Synthetic Biology*, **8**, 2635–2641.

8 Supplementary Data

Supplementary Table 1: Variance explained by each type of CRE in the half-life model applied to data from Chan et al. [@Chan2018] Variance explained was estimated in 3 different ways. The individual variance explained by a linear model containing each type of CRE on its own. The drop in variance explained when one type of CRE is removed from the full model. The cumulative variance explained when each CRE is added to the linear model in sequence; starting with codon usage, then adding motifs and finally 3'UTR length.

CRE	Individual	Drop	Cumulative
Codon Usage	42.0%	-40.5%	42.0%
Motifs	3.2%	-1.6%	43.7%
3'UTR Length	0.6%	0.0%	43.7%

Supplementary Table 2: Table showing the DNA sequences for all mCherry reporter constructs. mCherry construct sequences are available on the manuscript github repo. https://github.com/DimmestP/chimera_project_manuscript/supplementary_data_chapter/data/mCherry_fluorescence_reporter_dna_sequences.csv

Supplementary Table 3: Table showing the DNA sequences for all mCherry reporter constructs. mTurq construct sequences are available on the manuscript github repo https://github.com/DimmestP/chimera_project_manuscript/supplementary_data_chapter/data/mTurq_fluorescence_reporter_dna_sequences.csv

Supplementary Table 4: Primer sequences created for all qPCR experiments.

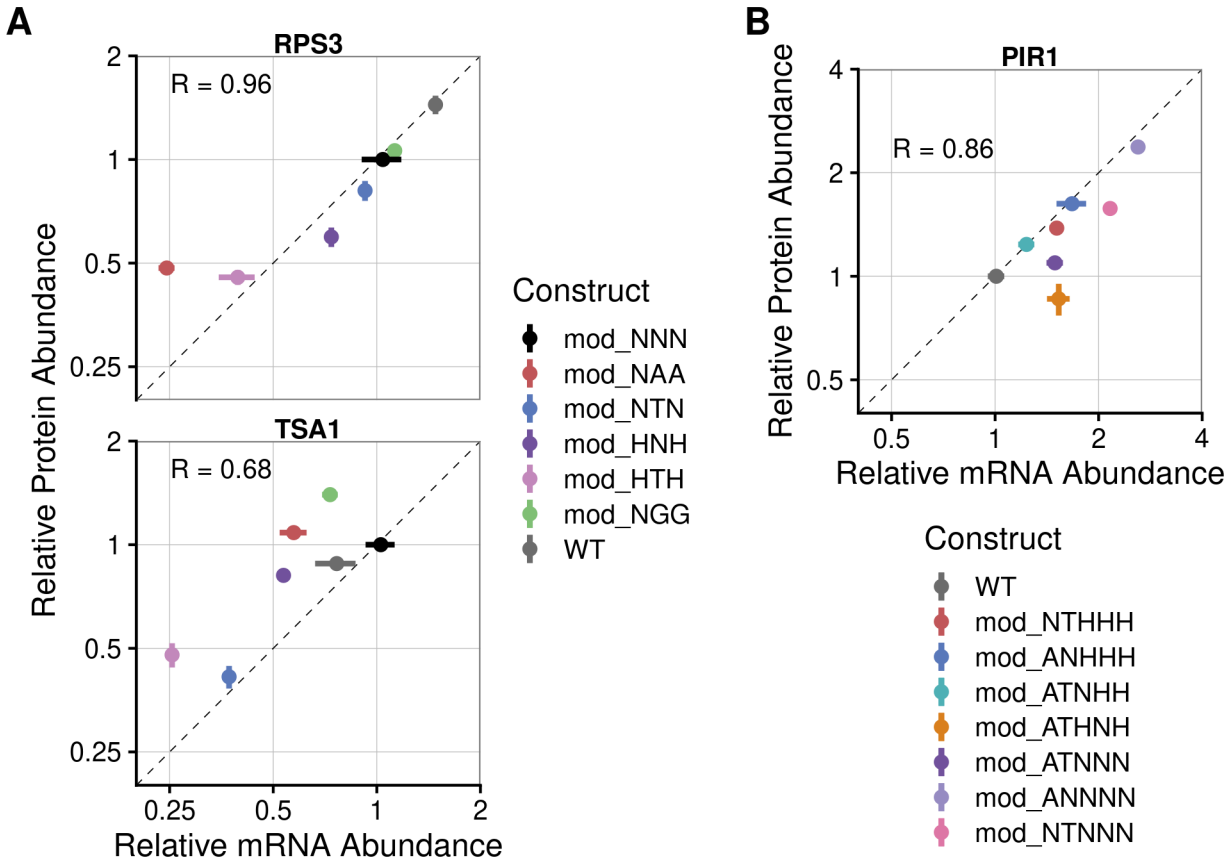
Primer name	Primer sequence	Purpose
mCh_F7	AGGACGGCGAGTTCATCTA	qPCR primer for mCherry ORF
mCh_R7	CCCATGGTCTTCTTCTGCATTA	qPCR primer for mCherry ORF
RPS3_F1	TCGCTGACGGTGTCTTCTACG	qPCR primer for RPS3 ORF
RPS3_R1	TCGGTCTTGGTTGGAGTGACA	qPCR primer for RPS3 ORF
RPS13_EF	CTAGAAATGCTCCAGCTTGGTTCAA	qPCR primer for RPS13 ORF
RPS13_ER	TCAAACCCTTTCTCGCGTACTTG	qPCR primer for RPS13 ORF
PGK1_F2	GCTGCTTTGCCAACCATCAA	qPCR primer for PGK1 ORF
PGK1_R2	TCGTTTCTTTCACCGTTTGGTC	qPCR primer for PGK1 ORF
mTu_F2	TTGGGGTGTTCAATGTTTTGC	qPCR primer for mTurq2 ORF
mTu_R2	TGAACATAACCTTCTGGCATGG	qPCR primer for mTurq2 ORF

Supplementary Table 5: Selecting the HWNCAUUWY motif, TTTCATTTC. Summary of the number of occurences and contributions to a linear model predecting half-life for each possible version of the HWNCAUUWY motif.

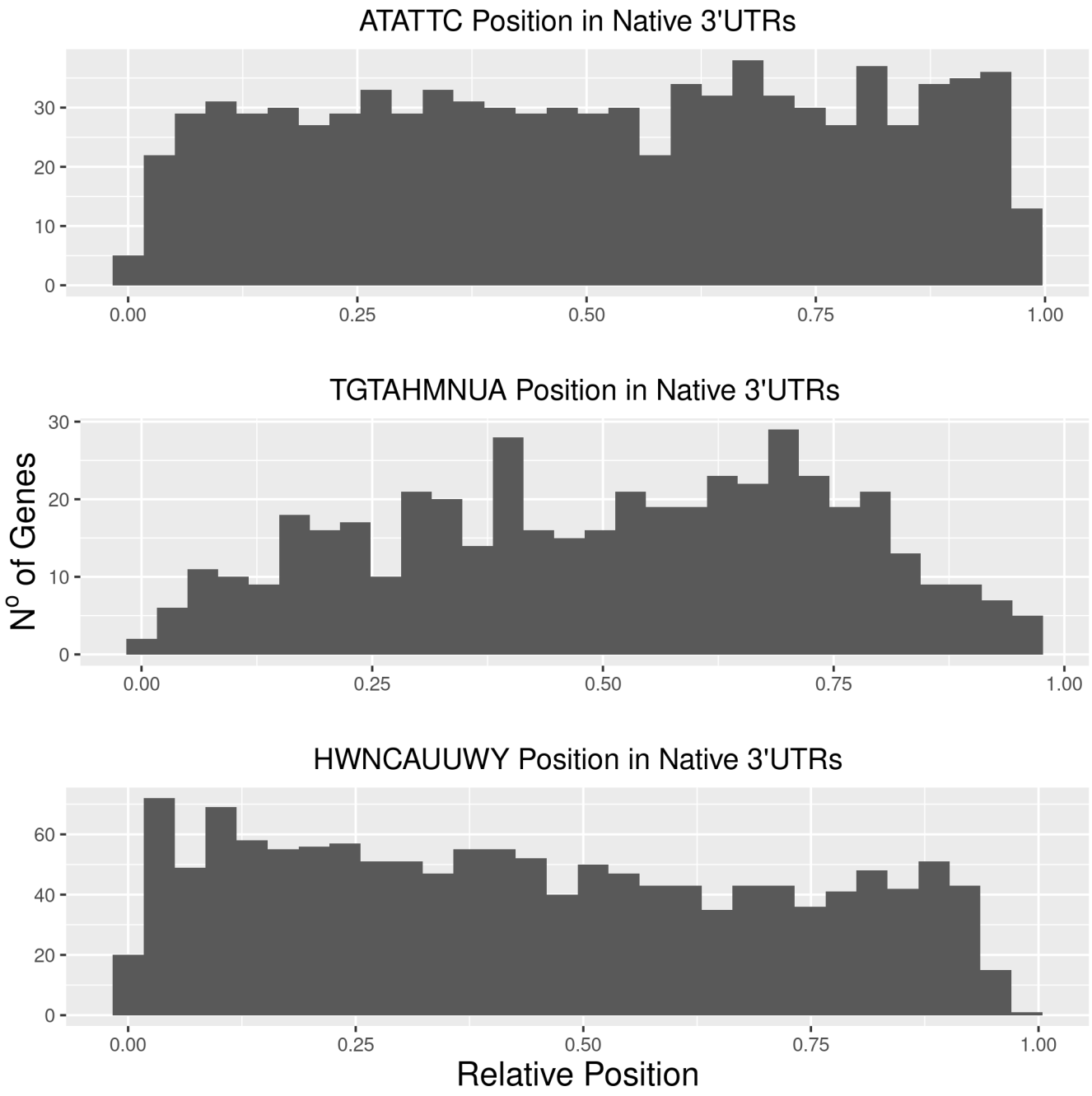
Motif	Total Count	Coef	p.value
CTTCATTTC	14	-0.52197	0.00474
ATTCATTTC	22	-0.44858	0.00577
TAGCATTTT	19	-0.43109	0.00934
TTGCATTTT	46	-0.27615	0.01088
CTGCATTTT	15	-0.47916	0.01162
TTTCATTTC	42	-0.27049	0.01421
AAACATTTC	13	-0.46682	0.01904
CTGCATTAT	10	-0.53910	0.02027
TTTCATTTT	103	-0.14469	0.03671
CTGCATTTC	6	-0.63943	0.03775
TTACATTAC	18	-0.43018	0.03918
TTCCATTAT	15	0.37314	0.04106
ATGCATTTT	31	-0.26015	0.04115
CATCATTAT	16	-0.38091	0.04879
ATTCATTAT	39	-0.22684	0.04928

Supplementary Table 6: Selecting the UGUAHMNUA motif, TGTACAATA. Summary of the number of occurences and contributions to a linear model predecting half-life for each possible version of the UGUAHMNUA motif.

Motif	Total Count	Coef	p.value
TGTATATTA	83	-0.51920	0.00000
TGTATCATA	17	-0.67490	0.00004
TGTATAATA	72	-0.27914	0.00066
TGTACCATA	6	-0.98705	0.00115
TGTACACTA	16	-0.56750	0.00373
TGTACAATA	27	-0.37788	0.00551
TGTAACATA	19	-0.44503	0.00687
TGTATACTA	36	-0.30274	0.00863
TGTACATTA	26	-0.29426	0.02791



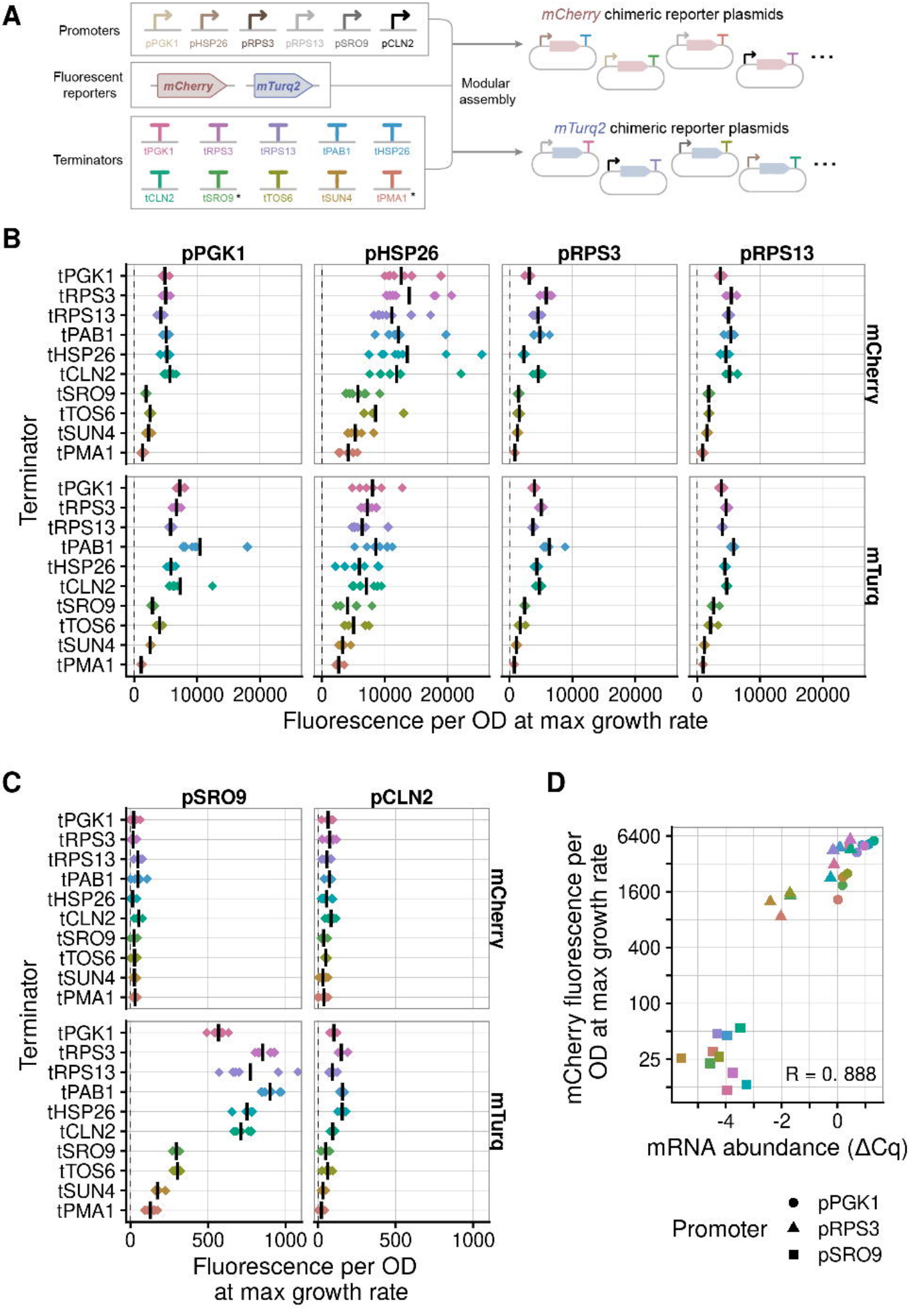
Supplementary Figure 1: Relative protein abundance correlates with relative mRNA abundance for reporter constructs with modified 3'UTRs. (**A**) Fold changes in RT-qPCR vs fold changes in mCherry fluorescence for all terminator constructs in the pRPS3-mCherry-tRPS3 and pTSA1-mCherry-tTSA1 pairings. Transcript abundance is relative to the mod0 construct of each promoter-terminator pairing. (**B**) Fold changes in RT-qPCR vs fold changes in mCherry fluorescence for all terminator constructs in the pPIR1-mCherry-tPIR1 pairing. Transcript abundance is relative to WT construct of the promoter-terminator pairing. The mean and standard error calculated over 6 biological replicates is plotted for each construct.

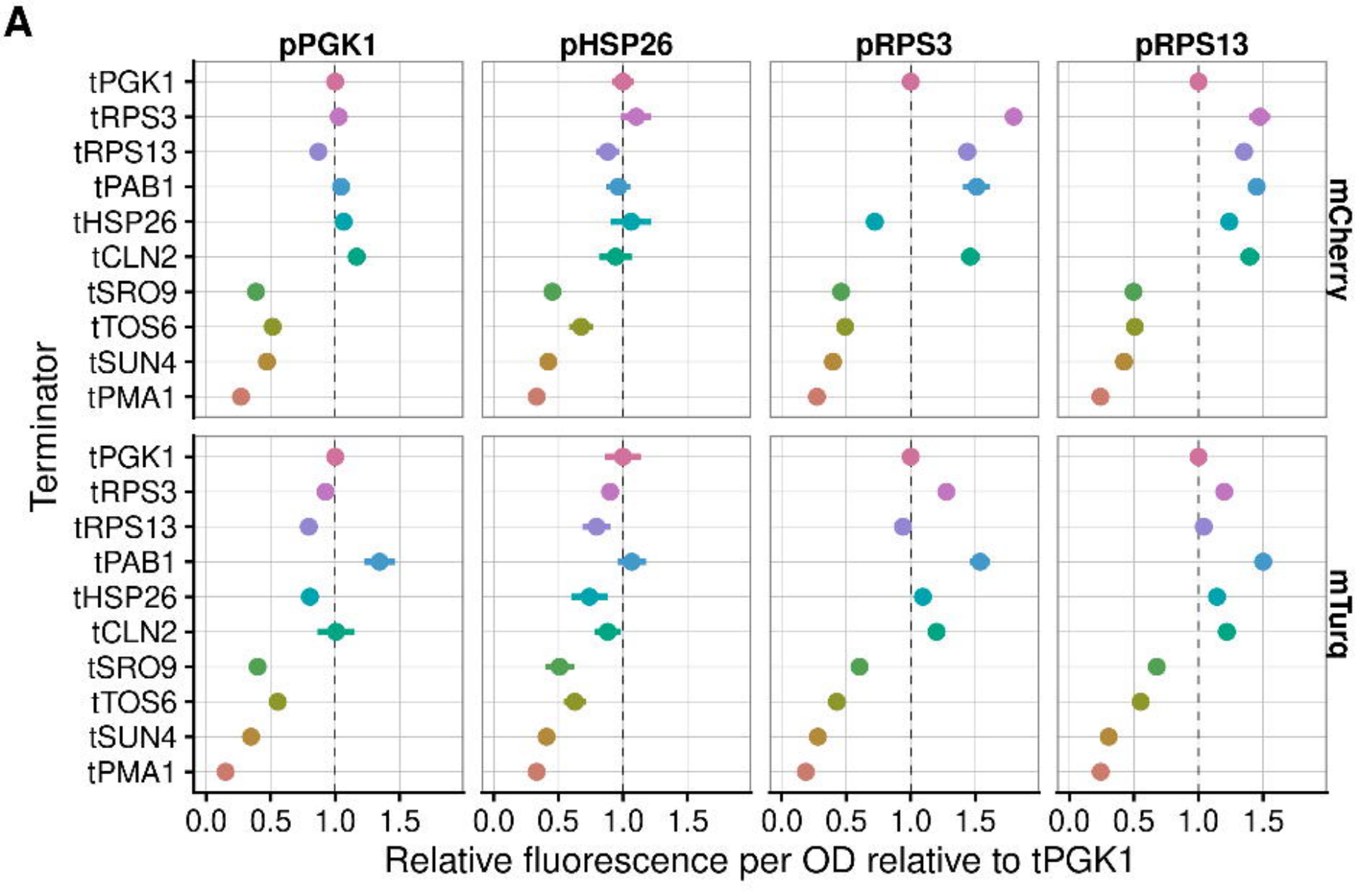


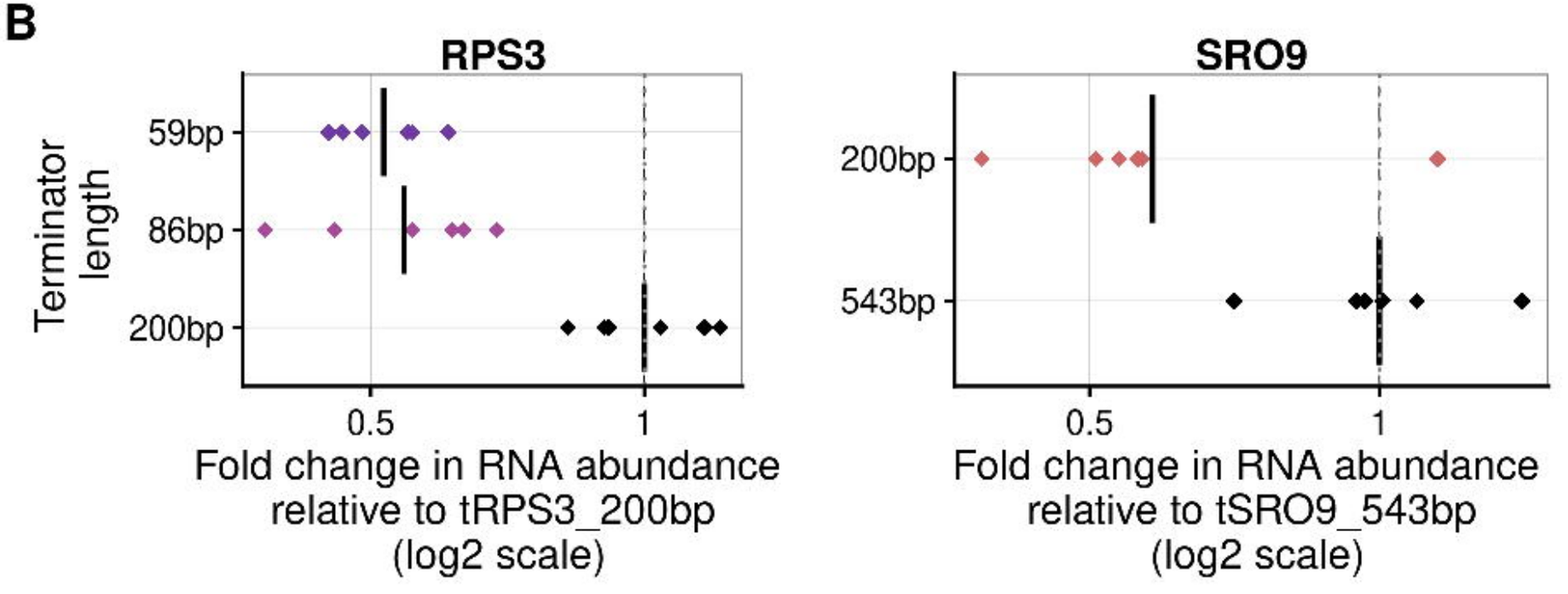
Supplementary Figure 2: Relative positions of ATATTC, TGTAHMNUA and HWNCAUUWY motifs in native 3'UTRs.

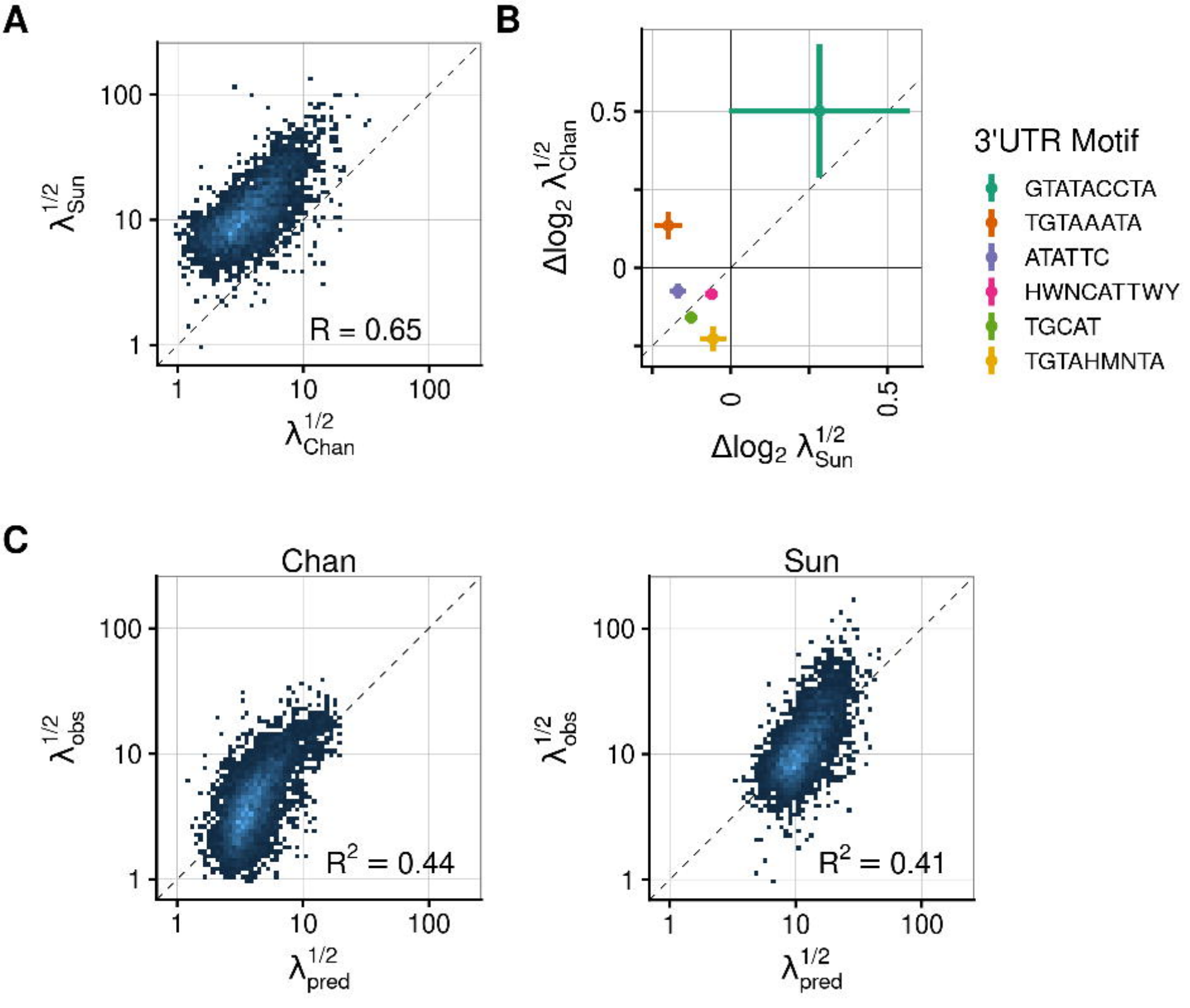
Supplementary Table 7: Table showing minimum free energies of 3'UTR constructs with inserted/deleted motifs.

Terminator	Construct	Minimum free energy (kcal/mol)
RPS3	WT	-6.10
RPS3	mod_NNN	-11.50
RPS3	mod_NTN	-13.00
RPS3	mod_NAA	-7.40
RPS3	mod_NGG	-10.80
RPS3	mod_HNH	-5.50
RPS3	mod_HTH	-7.40
TSA1	WT	-6.90
TSA1	mod_NNN	-10.14
TSA1	mod_NTN	-10.60
TSA1	mod_NAA	-8.90
TSA1	mod_NGG	-9.80
TSA1	mod_HNH	-6.10
TSA1	mod_HTH	-6.67
PIR1	WT	-28.00
PIR1	mod_ANHHH	-30.60
PIR1	mod_NTHHH	-27.70
PIR1	mod_ATNHH	-28.40
PIR1	mod_ATHNH	-27.00
PIR1	mod_ATNNN	-30.40
PIR1	mod_ANNNN	-32.60
PIR1	mod_NTNNN	-30.10


A
Promoters
pPGK1
pHSP26
pRPS3
pRPS13
pSRO9
pCLN2
Fluorescent reporters
mCherry
mTurq2
Terminators
tPGK1
tRPS3
tRPS13
tPAB1
tHSP26
tCLN2
tSRO9 *
tTOS6
tSUN4
tPMA1 *
Modular assembly
mCherry chimeric reporter plasmids
mTurq2 chimeric reporter plasmids
B
pPGK1
pHSP26
pRPS3
pRPS13
mCherry
mTurq
Terminator
0
10000
20000
Fluorescence per OD at max growth rate
C
pSRO9
pCLN2
0
500
1000
Fluorescence per OD at max growth rate
D
mCherry fluorescence per OD at max growth rate
6400
1600
400
100
25
R = 0. 888
-4
-2
0
mRNA abundance (ΔCq)
Promoter
pPGK1
pRPS3
pSRO9

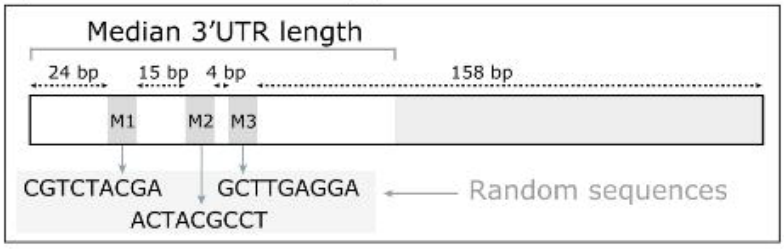

A
pPGK1
pHSP26
pRPS3
pRPS13
mCherry
mTurq
Terminator
tPGK1
tRPS3
tRPS13
tPAB1
tHSP26
tCLN2
tSRO9
tTOS6
tSUN4
tPMA1
0.0
0.5
1.0
1.5
Relative fluorescence per OD relative to tPGK1


B
RPS3
SRO9
Terminator length
59bp
86bp
200bp
543bp
0.5
1
Fold change in RNA abundance relative to tRPS3_200bp (log2 scale)
Fold change in RNA abundance relative to tSRO9_543bp (log2 scale)

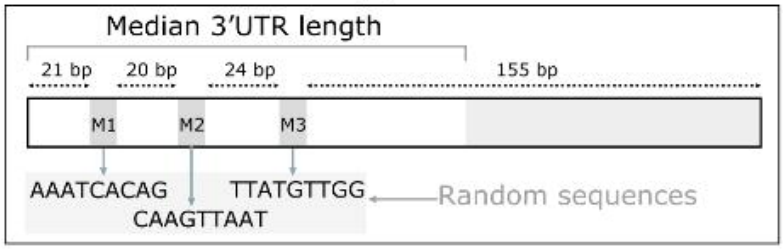

A
100
10
1
R = 0.65
1
10
100
$\lambda^{1/2}_{Sun}$
$\lambda^{1/2}_{Chan}$
B
0.5
0
0
0.5
$\Delta\log_2 \lambda^{1/2}_{Chan}$
$\Delta\log_2 \lambda^{1/2}_{Sun}$
3'UTR Motif
GTATACCTA
TGTAAATA
ATATTC
HWNCATTWY
TGCAT
TGTAHMNTA
C
Chan
Sun
$R^2 = 0.44$
$R^2 = 0.41$
$\lambda^{1/2}_{obs}$
$\lambda^{1/2}_{pred}$

A

RPS3 mod_NNN terminator

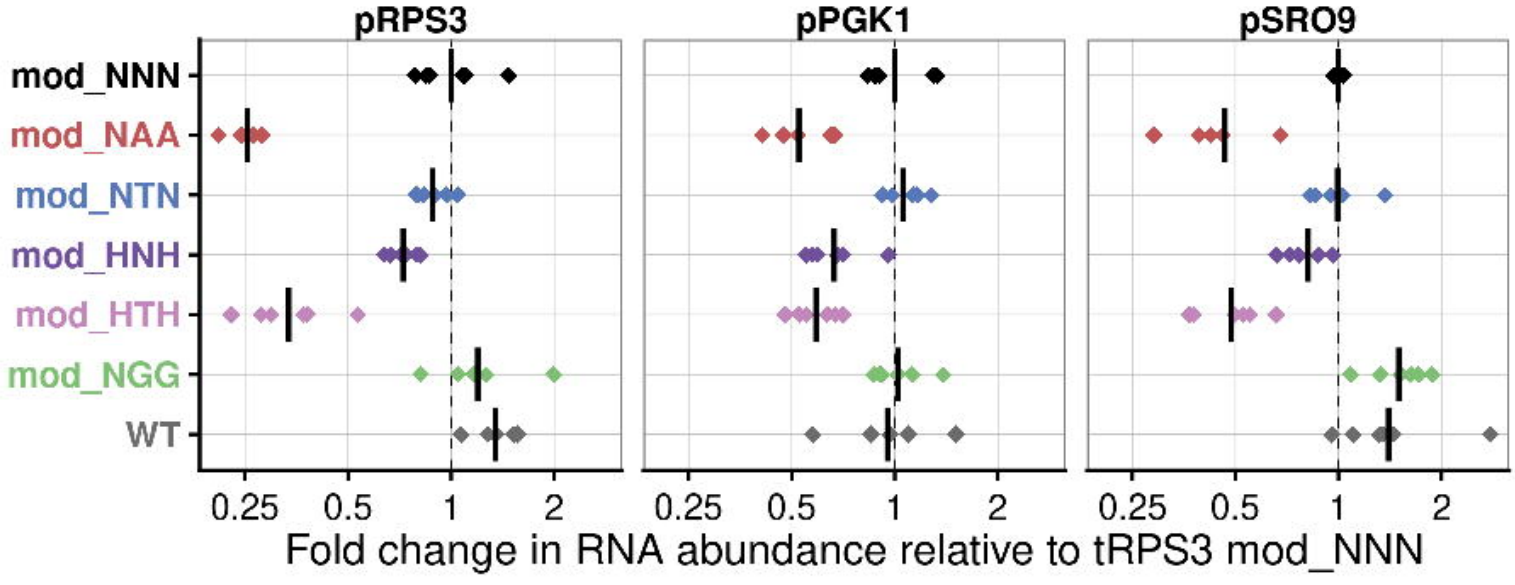


TSA1 mod_NNN terminator

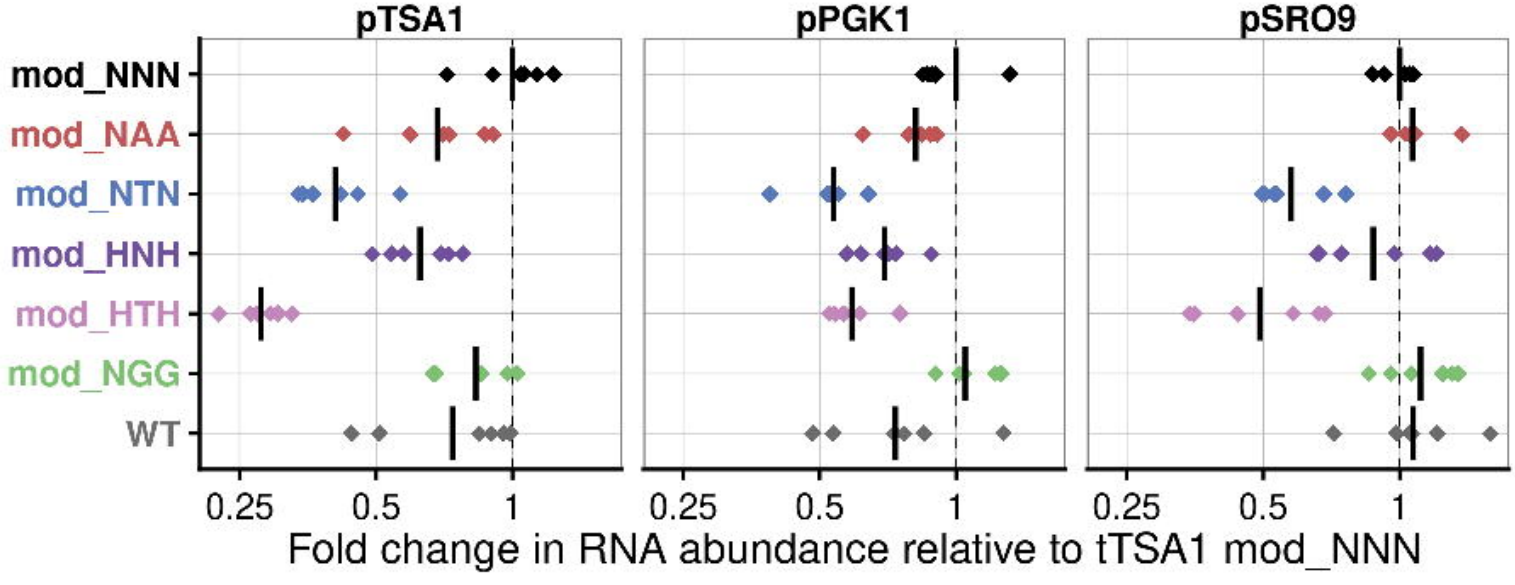


Terminator	Motif 1	Motif 2	Motif 3
mod_NNN	Random	Random	Random
mod_NAA	Random	ATATTC	ATATTC
mod_NTN	Random	TGTAHMNTA	Random
mod_HNH	HWNCATTWY	Random	HWNCATTWY
mod_HTH	HWNCATTWY	TGTAHMNTA	HWNCATTWY
mod_NGG	Random	GTATACCTA	GTATACCTA
WT	NA	NA	NA

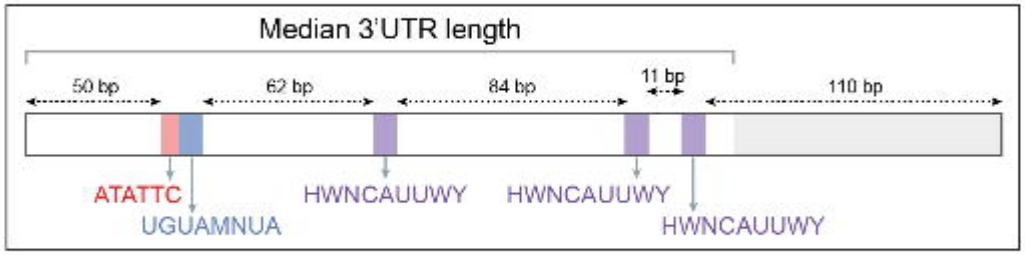
B



C

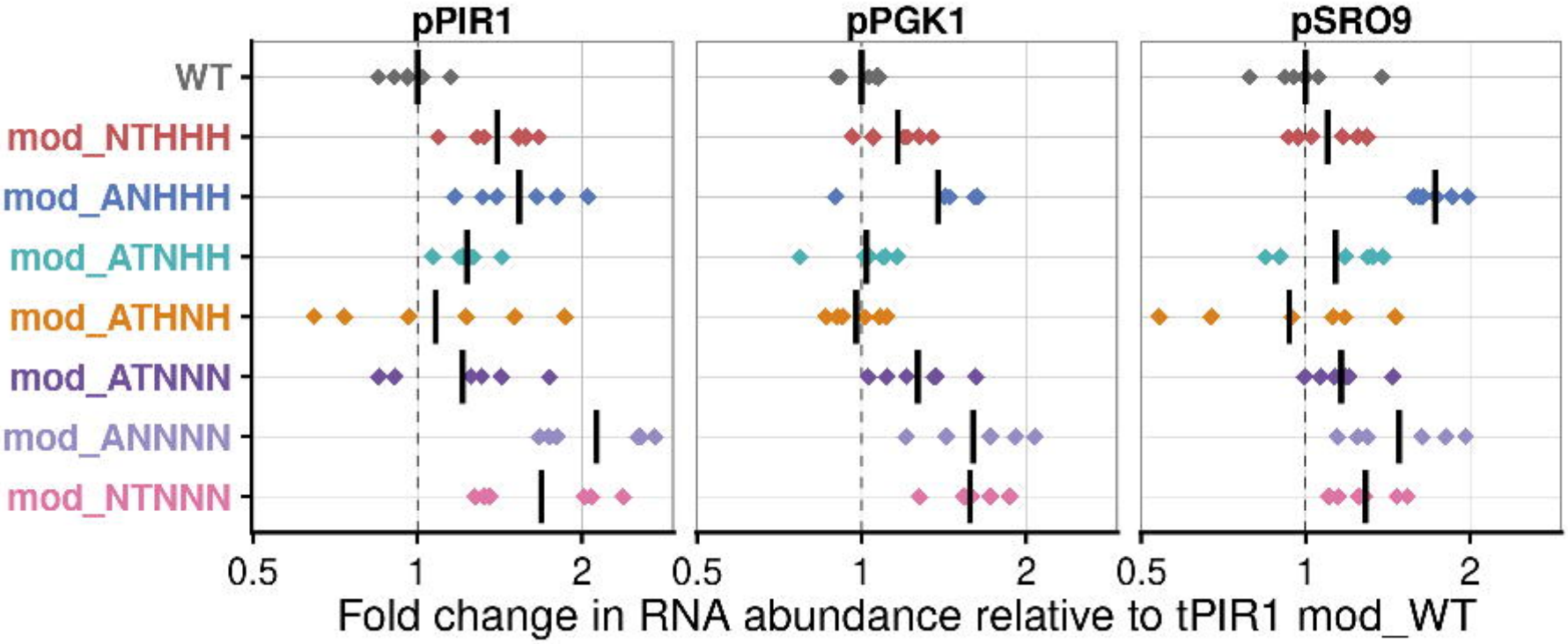


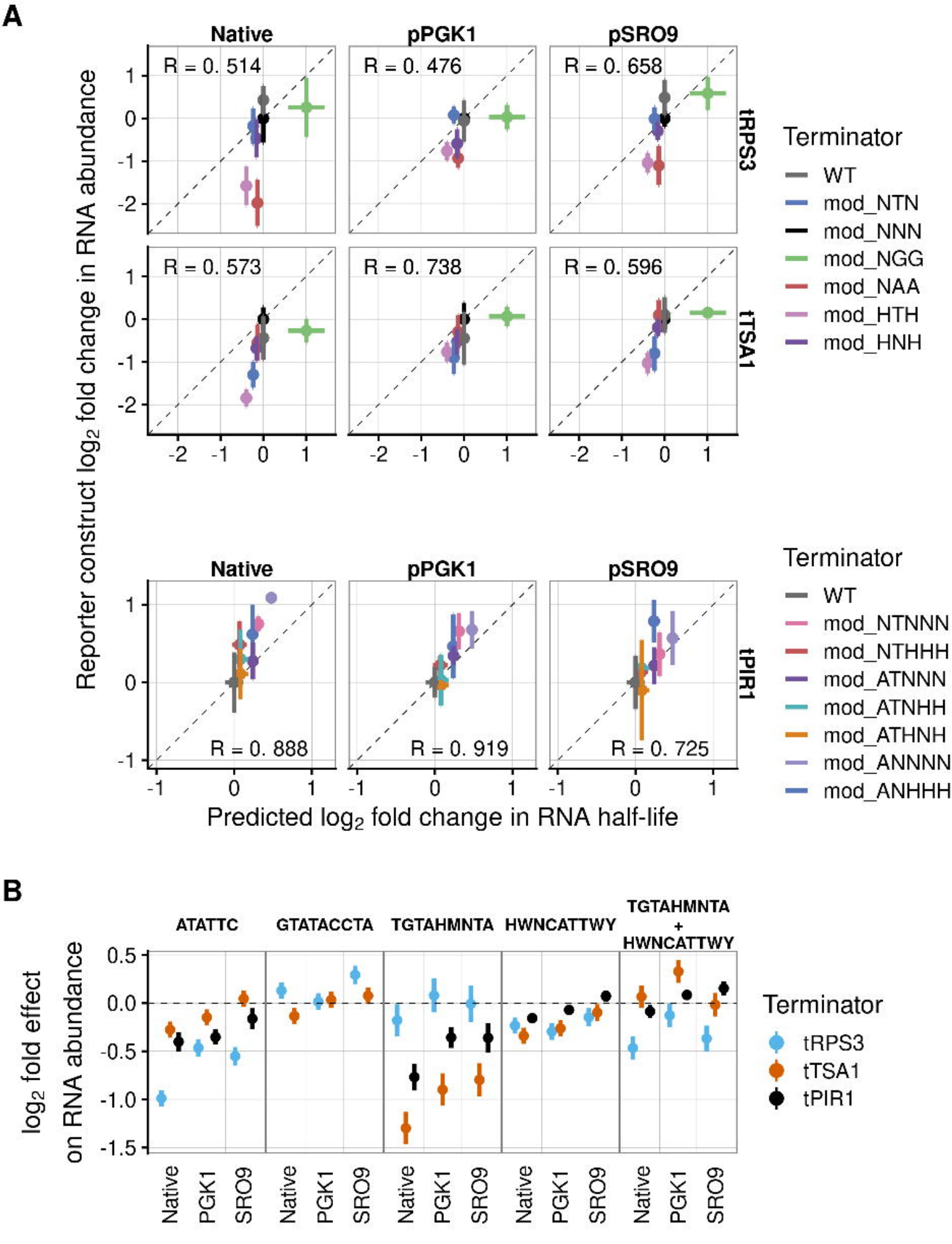
A



B

Terminator	Motif 1	Motif 2	Motif 3	Motif 4	Motif 5
WT	ATATTC	TGTAHMNTA	HWNCATTWY	HWNCATTWY	HWNCATTWY
mod_NTHHH	Random	TGTAHMNTA	HWNCATTWY	HWNCATTWY	HWNCATTWY
mod_ANHHH	ATATTC	Random	HWNCATTWY	HWNCATTWY	HWNCATTWY
mod_ATNHH	ATATTC	TGTAHMNTA	Random	HWNCATTWY	HWNCATTWY
mod_ATHNH	ATATTC	TGTAHMNTA	HWNCATTWY	Random	HWNCATTWY
mod_ATNNN	ATATTC	TGTAHMNTA	Random	Random	Random
mod_ANNNN	ATATTC	Random	Random	Random	Random
mod_NTNNN	Random	TGTAHMNTA	Random	Random	Random




A
Native
pPGK1
pSRO9
R = 0. 514
R = 0. 476
R = 0. 658
R = 0. 573
R = 0. 738
R = 0. 596
tRPS3
tTSA1
Terminator
WT
mod_NTN
mod_NNN
mod_NGG
mod_NAA
mod_HTH
mod_HNH
Reporter construct log2 fold change in RNA abundance
R = 0. 888
R = 0. 919
R = 0. 725
tPIR1
Terminator
WT
mod_NTNNN
mod_NTHHH
mod_ATNNN
mod_ATNHH
mod_ATHNH
mod_ANNNN
mod_ANHHH
Predicted log2 fold change in RNA half-life
B
ATATTC
GTATACCTA
TGTAHMNTA
HWNCATTWY
TGTAHMNTA + HWNCATTWY
log2 fold effect on RNA abundance
Native
PGK1
SRO9
Terminator
tRPS3
tTSA1
tPIR1