

Global determinants of mRNA degradation rates in *Saccharomyces cerevisiae*

Benjamin Neymotin¹, Victoria Ettorre¹, and David Gresham^{1,2}

¹Center for Genomics and Systems Biology, Department of Biology, New York University,

²correspondence: dgresham@nyu.edu

Degradation of mRNA contributes to variation in transcript abundance. Studies of individual mRNAs show that *cis* and *trans* factors control mRNA degradation rates. However, transcriptome-wide studies have failed to identify global relationships between transcript properties and mRNA degradation. We investigated the contribution of *cis* and *trans* factors to transcriptome-wide degradation rate variation in the budding yeast, *Saccharomyces cerevisiae*, using multiple regression analysis. We find that multiple transcript properties are associated with mRNA degradation rates and that a model incorporating these factors explains ~50% of the genome-wide variance. Predictors of mRNA degradation rates include transcript length, abundance, ribosome density, codon adaptation index (CAI) and GC content of the third position in codons. To validate these factors we studied individual transcripts expressed from identical promoters. We find that decreasing ribosome density by mutating the translational start site of the *GAP1* transcript increases its degradation rate. Using variants of GFP that differ at synonymous sites, we show that increased GC content of the third position of codons results in decreased mRNA degradation rate. Thus, in steady-state conditions, a large fraction of genome-wide variation in mRNA degradation rates is determined by inherent properties of transcripts related to protein translation rather than specific regulatory mechanisms.

INTRODUCTION

Alterations in the abundance of mRNA result from changes in both the rate of transcript synthesis and the rate of transcript degradation. Synthesis and degradation of mRNAs is critical for control of gene expression and cell survival as ablation of either process results in rapid loss of viability (1, 2). The *cis* and *trans* factors that control rates of mRNA synthesis have been extensively studied in many systems (reviewed in (3)). By comparison, far less is known about factors that control rates of mRNA degradation. A complete understanding of gene expression regulation requires identification of the sources of variation in mRNA degradation.

Our understanding of the mechanisms by which mRNAs are degraded (reviewed in (4)) is largely the result of studies of specific transcripts (5-7). These studies have shown that mRNA degradation is controlled by *cis* factors, including sequence elements in the coding (8, 9) and untranslated (10, 11) regions, as well as factors in *trans*, including RNA binding proteins (12, 13) and non-coding RNAs (14). However, the extent to which these different factors impact global patterns of mRNA degradation remains unclear.

Genome-wide mRNA degradation rates have been determined for a number of organisms including bacteria (15), plants (16), flies (17), mouse (18) and human cell lines (19). In the budding yeast, *Saccharomyces cerevisiae*, genome-wide mRNA degradation rates have been measured using a variety of methods including transcriptional inhibition (20-22), genomic-run-on (23), and metabolic labeling (24, 25). In general, the concordance between different global studies of mRNA degradation rates is poor likely due to a combination of technical and biological sources of variation. Recently, we introduced RNA Approach to Equilibrium Sequencing (RATE-seq), which combines 4-thiouracil (4-tU) labeling and RNA-seq for determination of genome-wide *in-vivo* mRNA degradation rates (26). Using approach to equilibrium labeling kinetics and non-linear regression, RATE-seq overcomes several problems with existing methods providing improved accuracy of mRNA degradation rates estimates, and a measure of the precision with which the rates are measured, in steady-state conditions.

Despite discrepancies in estimates of mRNA degradation rates among different studies, three consistent features have been demonstrated across multiple transcriptome-wide datasets. First, there is variation in the rates at which different transcripts are degraded: some vary by as much as an order of magnitude. Second, transcripts for genes encoding functionally-related products have similar degradation rates (20, 26, 27). Third, no single property of transcripts explains the observed variation (20, 25, 24). This latter point suggests that either the causative factors are obscured in genome-wide studies, or a combination of different factors affect rates of degradation that have transcript specific effects. Potential properties of transcripts that might impact their rate of degradation include transcript length, GC content, transcript abundance, codon usage and folding properties. However, testing the effect of any single property of transcripts on global degradation rates is inherently challenging as each parameter can vary independently across transcripts. At the same time many transcript properties are correlated with each other making it difficult to identify causative factors. Thus, to identify determinants of

variation in mRNA degradation rates all known properties of transcripts must be considered simultaneously and experimental designs that modulate a single property are required for validation experiments.

Here, we analyzed mRNA degradation rates in *S. cerevisiae* using multiple regression analysis (28, 19) and genome-wide datasets that have defined gene specific parameters. These include protein levels (29), protein half-life (30), RNA abundance (31), transcription rates (32), UTR lengths (33), ribosome density (34), association with RNA binding proteins (35) and the function of the encoded product. A multiple regression model applied to mRNA degradation rates determined using RATE-seq accounts for 50% of the variation in mRNA degradation rates. Although less variation is explained for multiple regression models applied to other genome-wide mRNA degradation datasets, many predictors are significant in multiple datasets, suggesting that they are reproducible transcript properties that impact degradation rates. These features include ribosome density, codon adaptation index, and GC content of the wobble position in a codon (GC3) suggesting that protein translation is closely related to mRNA degradation.

Using experimental studies of individual transcripts, we show that changing ribosome density affects both the mRNA degradation rate and steady state levels. Using GFP coding sequence variants that differ only in their GC3 content we show that coding sequence affects mRNA degradation. Increasing the GC3 content increases the stability of mRNAs resulting in increased steady-state level. Our results suggest that mRNA degradation is determined by multiple factors, many of which are intimately linked to protein translation.

MATERIALS AND METHODS

Plasmid construction

Plasmid pCM188 (36) was used as the backbone for all plasmids. This CEN4 plasmid contains the *URA3* gene, a constitutively expressed tetracycline transactivator, and a multiple cloning site with a *CYC1* TATA region upstream, all under control of two copies of the tetracycline operator. Transcription of the gene is repressed in the presence of tetracycline or its derivatives, including doxycycline. Plasmids DGP147, DGP148, DGP149, and DGP231 are pCM188 with degenerate forms of GFP (37) ranging in GC content in the third position of each codon as 0.71, 0.38, 0.60, and 0.67 respectively (GFP4, GFP1, GFP2, GFP3). The coding sequence of each GFP was cloned into the BamHI and NotI sites. Plasmid DGP217 is pCM188 with the *GAP1* gene + 3'UTR cloned into the BamH I and Not I sites. Plasmid DGP218 is the same as DGP217, except the start codon of *GAP1* has been mutated to GTG.

Strains and Growth Conditions

The laboratory strain, FY3 (*MATa ura3-52*) which is isogenic to S288C was used for all experiments. DGY696, 697, 698, and 1281 have plasmids DGP147, 148, 149, and 231 respectively. DGY1193 and 1194 carry plasmids DGP217 and DGP218 respectively, with the *GAP1* locus deleted from start to stop with the KANMX4 cassette.

For all experiments, a single colony for each strain was inoculated in synthetic complete media without uracil, to maintain selection of the plasmids. Studies of the *GAP1* transcript were performed in nitrogen limiting media with proline as the limiting nitrogen source, as described in (38). Saturated cultures from overnight cultures were back-diluted 1:50 into media of the same composition. Cells were allowed to grow for 5.5 hrs (~2.5 doublings) before transcription was inhibited with doxycycline at a final concentration of 10µg/mL. Cells were collected by filtration on nitrocellulose membranes and snap frozen in liquid nitrogen.

RNA processing and qRT-PCR analysis

RNA was extracted using the hot phenol-chloroform method as in (26). Purified RNA was then treated with RQ1 DNase according to manufacturer recommendations. Reverse transcription was performed using random hexamers and MMLVRT enzyme. Quantitative Reverse Transcription PCR (qRT-PCR) was performed using the SYBR Green system and a Roche Light Cycler. RNA levels were quantified in comparison to the *HTA1* housekeeping gene, which is unaffected by Doxycycline addition. Ratios were calculated using the formula: $Y=2^{-(HTA1_{ct}-Gene_{ct})}$ with *ct* being the calculated cycle threshold. RNA levels from each time point were normalized to *t*=0, which was set to 1. All analyses with error bars are the mean +/- the standard error for 3-6 biological replicates. Values without error bars are the average of two replicates. Primer sequences had amplification efficiencies of at least 95% on RT products. The amplified product for all GFP strains begins between position 463 and 586 of the transcript. The amplified products are all between 80 and 120 base pairs long. The sequences used for qRT-PCR analysis are as follows: *HTA1* (Forward: 5'- GCTGGTAATGCTGCTAGGGATA-3', Reverse: 5'- TTACCCAATAGCTTGTTC AATT-3'), *GFP2*, *GFP4* (Forward: 5'-TTGCCGGATAACCACTACCT-3', Reverse: 5'- CCTGCTGCAGTCACAACTC-3'), *GFP3* (Forward: 5'-GCCGATAAGCAGAAGAATGG-3', Reverse: 5'-TGTTGATAATGGTCCGCAAG-3'), *GFP4* (Forward: 5'- CGACCATTACCAGCAGAACA-3', Reverse: 5'- GGGTCCTTTGACAGAGCAGA-3'), *GAP1-ATG*, *GAP1-GTG* (Forward: 5'-TTTGTCTGTCTTCGTCAC-3', Reverse: 5'- CTCTACGGATTCACTGGCAGCA-3')

Multiple Regression Analysis

For multiple regression analysis we used degradation rates rather than half-lives, which are typically log-normally distributed. To minimize the effects of extreme outliers we removed values more than 1.5 times the interquartile range. Degradation rates for most datasets were calculated as $\ln(2)/t_{\text{half-life}}$, except in (25) and (26) where the effects of dilution as a result of cellular growth was also considered. Transcript counts were from (31) and estimated based on an assumption of ~60,000 mRNA/cell (39). Protein per mRNA was calculated as the values from (29) divided by the values for counts. Codon adaptation index (40) was calculated for each transcript based on the codon frequency tables in the seqinr package in R (R core team). To normalize the data we log transformed the predictor variables ($\log_{10}(\text{Variable})$ or $\log_{10}(\text{Variable} + 1)$), except for GC content of each codon position, which is approximately normal in distribution, and ΔG , which is negative in

value. For categorical variables, transcripts were classified as bound by an RNABP based on data from (35), and as present or absent from a Gene Ontology group based on GO SLIM files downloaded from SGD.

In a linear multiple regression model (Crawley 2012), a parameter of interest is modeled as being dependent on two or more predictors. We use the degradation rate constant as the parameter of interest, and the other measurements as predictors. To build the model, we followed two separate approaches. In the first approach, we first determined the p-value of the pair wise correlation of each predictor to degradation rate. This indicates whether the regression coefficient is significantly different from zero, and whether or not the predictor has any effect on degradation rate. Next we included all predictors that have a p-value less than 0.05 into the multiple regression models. We then performed stepwise deletion of terms by removing sequentially the predictors with the highest p-value. The final model is then the reduced model where only significant terms remained. We obtained the same result using the step function in R, which reduces models based on Akaike's Information Criterion (AIC). In a second approach, we first calculated the significance of each predictor when it is the only one in the model, as above. We then started adding to the model based on the predictors with the lowest p-value. With each additional term we checked to see that all of the terms in the model were significant. If a new term was added and it was not significant, we removed it from the model. If a new term was added and a different predictor lost its significance, we tested a model with either the new predictor or the one that lost significance, and retained the one that explained more variation. We did not add terms that were insignificant in the pair wise correlation with degradation rate. Both approaches gave similar results. Model diagnostics suggest there is no obvious curvature or patterns in terms of increase or decrease in variance as a function of fitted values (**Figure S6**). There is also minimal curvature in the normal Q-Q plot, suggesting the model follows linearity (**Figure S6**).

R functions and packages

We performed all analyses using R (R Core Team 2013) and several open source packages. In addition to custom written functions in R, we also used functions from the following packages: *TeachingDemos*, *Biostrings*, *LSD*, *stringi*, *GeneRfold* and *seqinr*.

RESULTS

Multiple transcript properties affect global mRNA degradation rates

Previous studies have found evidence for the effect of specific properties of transcripts on the degradation rate of individual transcripts (10, 8, 11, 12). We tested the relationship between globally measured transcript features (**Table 1**) and genome-wide mRNA degradation rates determined using RATE-seq (**Table S1**). We find that several transcript features are significantly correlated with mRNA degradation rates (**Figure 1A** and **Table S2**). The most significant single feature predictive of mRNA degradation rates measured using RATE-seq is the length of the coding sequence, which explains almost 30% of the variance. The folding energy (ΔG) is also significantly associated with mRNA degradation

rates, which may be due to the fact that folding energy and coding sequence length are highly correlated. Several features related to the translation of transcripts are also significantly associated with mRNA degradation rates including ribosome density, the codon adaptation index and the GC3 content. We also tested whether the function of the encoded product is predictive of mRNA degradation rate and found that functional assignment using gene ontology (GO) terms explains a significant fraction of the variation (Figure 1A). This is consistent with the observation that transcripts encoding proteins in similar functional categories degrade with similar kinetics (20, 26, 27). In addition, association with specific mRNA binding proteins also explains a significant fraction of variation in mRNA degradation rates (Figure 1A). These results suggest a relationship between several transcript features and mRNA degradation rates. We observed similar relationships between these predictors and mRNA degradation rates measured using other methods (Figure S1) suggesting that some of these relationships are reproducible despite the poor agreement in mRNA degradation rates among different studies.

Although many transcript properties are correlated with each other (Figure S2), some properties show no correlation and therefore may exert independent and differential effects on the rate of mRNA degradation. Therefore, we used multiple regression analysis to model the contribution of multiple transcript features to variation in mRNA degradation rates simultaneously (methods). We initially built a model incorporating all factors and used sequential reduction to arrive at a minimal model (methods). We find that the explained variation when multiple transcript properties are included exceeds the variance explained by any single factor suggesting that degradation rates are determined by a combination of transcript features (“model” in Figure 1A). When the categorical factors of gene function and binding to specific proteins, are included 50% of the variance in mRNA degradation rates can be explained (“model++” in Figure 1A). Thus, the rates predicted by a multiple regression model are in good agreement with experimentally determined rates (Figure 1B). Models incorporating all features explain significant fractions of the variation reported in other mRNA degradation datasets, albeit with reduced explanatory power (Figure 1C). Interestingly, we find that models applied to mRNA degradation rates measured using transcriptional inhibition tend to explain much less variation than models applied to RNA degradation rates measured with different metabolic labeling methods.

Our model suggests an inverse relationship between mRNA degradation rate and translational elongation rates, as measured by Codon Adaptation Index (CAI) and ribosome density. Translation elongation rates are slowed during peptide bond formation for proline residues (41), and particularly when multiple prolines are encoded sequentially. To further investigate the role of translation elongation in mRNA degradation, we classified transcripts based on presence or absence of at least four sequential proline codons. Consistent with our multiple regression model, transcripts rich in proline degrade more rapidly than the rest of the transcriptome (Figure 2A). Interestingly, stretches of proline codons are also associated with overall lower protein expression levels (Figure 2B).

Ribosome density affects mRNA degradation rate

Our regression analysis suggests that different aspects of translation affect mRNA degradation rates. Ribosomes are generally thought to protect mRNAs from degradation (4). Consistent with previous analyses (42), we find that increased ribosome density is associated with decreased rates of mRNA degradation (**Figure S3**). To experimentally reduce the density of ribosomes on specific transcripts, we mutated the start codon of an endogenous transcript *GAP1*, which encodes the general amino acid permease, from ATG to GTG and placed it under control of doxycycline-repressible promoters (36). Addition of doxycycline has little effect on cellular physiology and no detectable effect on global gene expression (43). ATG start codons are required for the small ribosomal subunit to recruit the large ribosomal subunit for fully formed ribosomes. Mutation of ATG to GTG is expected to reduce the number of ribosomes bound to an mRNA, but not eliminate ribosome binding entirely as a downstream ATG may serve as a start codon for 80s ribosome formation.

We tested the *GAP1* transcripts for alteration in degradation kinetics as a function of start codon mutation. In the absence of a start codon, we find that the *GAP1* transcript is significantly decreased in steady state mRNA abundance (**Figure 3A**) and the transcript degrades more rapidly upon transcriptional inhibition (**Figure 3B** and **Figure 3C**). Our results suggest that a decrease in ribosome density increases the degradation of the *GAP1* transcript, consistent with the global trend detected in our regression model

Decreased GC content of the third codon position increases the rate of mRNA degradation

Our multiple regression model predicts that factors involved in translation including ribosome density, codon adaptation index and the GC3 content, contribute to variation in mRNA degradation rates (**Figure 1A**). GC3 content affects mRNA abundance and mRNA degradation in *E. coli* (37). By studying a small set of transcripts in mammalian cells, GC3 content was also found to affect mRNA levels, but not degradation rates (44) implying that mRNA synthesis or processing must underlie differences in mRNA levels. However, a more recent genome-wide study found evidence that decreased GC3 content is correlated with increased mRNA degradation rates (19).

To study the contribution of GC3 content to variation in mRNA degradation rates we used GFP constructs that differ in sequence at synonymous sites only (37). We studied four GFP transcripts that span a range of GC3 content (**Figure 4A**). Changes in GC3 content also results in overall variation in total GC content (**Figure 4A**). Coding sequences were placed under control of the identical doxycycline-regulated promoter and engineered to have the same UTRs (**Figure 4B**). We confirmed that all four constructs result in functional GFP expression (data not shown).

As all coding sequences are expressed from an identical promoter differences in steady state mRNA abundance must result from differences in degradation, synthesis or processing. We find that steady-state mRNA levels vary with GC3 content with the highest GC3 content resulting in the highest steady-state mRNA abundance (**Figure 4B**) consistent with observations in mammalian cells (44). Following addition of doxycycline to repress

transcription initiation, we confirmed that three of the four transcripts degrade differentially in a GC3-dependent manner (**Figure 4C and 4D**), consistent with our multiple regression prediction. Using the measured steady-state abundances and degradation rates for the three transcripts, we estimated synthesis rates. All three strains have similar estimated rates of synthesis consistent with differences in degradation rates being the primary determinant of differences in steady-state mRNA levels. We find that a fourth construct, which has much lower GC3 content, and the lowest steady-state abundance does not significantly differ in its mRNA degradation rate from the second lowest GC3 (**Figure S4**). This may reflect a limitation of the statistical power of our assay and the fact that other factors are likely to interact with the GC3 effect.

DISCUSSION

The abundance of mRNAs is determined by both the rate at which they are synthesized and the rate at which they are degraded. Previous studies of individual transcripts have determined *cis* and *trans* factors that affect mRNA degradation rates (4, 45). However, variation in genome-wide mRNA degradation rates cannot be explained by any single property of transcripts. In this study we sought to construct a comprehensive model that predicts mRNA degradation rates. Using multiple regression analysis, we find that 43% of variation in mRNA degradation rates determined using RATE-seq can be explained by considering multiple inherent properties of transcripts in a single model. By including association with specific RNA binding proteins and the function of the encoded product, ~50% of the genome-wide variation in mRNA degradation rates can be explained. Interestingly, we find that methods for measuring RNA degradation that use transcriptional inhibition tend to explain far less variation than less disruptive methods. This may reflect that fact that metabolic labeling methods, which minimally perturb the cell, yield more physiologically relevant degradation rates than transcriptional inhibition, which results in rapid cell death.

In our analysis of mRNA degradation rates measured using RATE-seq, coding sequence length is the strongest single predictor of mRNA degradation rates: in general, the longer a transcript the more rapidly it is degraded. Other genome-wide investigations have shown a positive relationship between the length of the mature mRNA and its rate of degradation (46, 19, 47, 48). Studies of individual transcripts have shown that increasing transcript length by addition of specific sequences containing “instability elements” enhances a transcript’s rate of degradation (49). It is possible that it is not the length of the transcript that affects its degradation, but the presence of additional regulatory elements, to which *trans* factors such as RNA binding proteins can bind. A second, though not mutually exclusive explanation for the observed relationship with mRNA length, relates to the abundance of a transcript. The highest transcript levels are achieved through rapid synthesis and slow degradation. Therefore, the most abundant transcripts are expected to have slow rates of degradation. Within the transcriptome, the most abundant transcripts include those encoding ribosomal proteins and histones, which also tend to have the shortest coding length. The observed effect of length on mRNA degradation rates is likely a combination of both regulatory elements and non-random relationships between the encoded function of the transcript and its length.

Association with ribosomes increases mRNA stability

Our multiple regression analysis provided evidence that protein translation impacts mRNA degradation rates. Previous studies regarding the role of translation on mRNA degradation have shown differing results. In a detailed study of the *CYC1* transcript, all start codons were removed from its coding region thereby preventing ribosome binding and translation (50). Following global transcription inhibition with thiolutin, the translationally impaired transcript degraded with similar kinetics to the translated transcript. Similarly, when translation of the *MFA2* transcript was inhibited by introducing a strong secondary structure in the 5' region of the transcript it was not found to alter the degradation kinetics following transcriptional inhibition using the GAL system (51). By contrast, using the same method to reduce translation of the *PGK1* transcript, results in an increased mRNA degradation rate (6). Studies based on individual transcripts may be limited due to both interaction with additional factors and issues of statistical power. Our analysis shows that genome-wide, increased ribosome density is correlated with decreased mRNA rates (**Figure 1** and **Figure S3**) as recently reported (42).

To study the effect of ribosome density on the rate of mRNA degradation we mutated the start codon of *GAP1* from an ATG to GTG. Consistent with results from our genome-wide study, loss of the first start codon results in reduced mRNA levels and an increased rate of mRNA degradation. Using a functional assay, we found that mutation of the start codon does not result in a loss of protein function (**Figure S5**). The next potential start codon is 288 nucleotides (96 amino acids) downstream of the wildtype start codon and thus we postulate that its use may result in a protein product that retains much of the wildtype GAP1 activity. Within the first 96 amino acids are residues known to affect the extent of functionality and localization of the encoded permease (52). Therefore, in addition to a decrease in ribosome density, we cannot exclude the possibility that alteration in protein function and/or possibly its production contributes to the decreased stability of the transcript.

Synonymous coding mutations affect mRNA stability

Regression analysis suggested genome-wide relationships between codon usage and mRNA degradation rates. We find a negative correlation between codon adaptation index and mRNA degradation rate. Bias towards more frequent codons in a transcript may increase the rate of elongation during protein translation (53). Therefore, this negative correlation suggests that faster elongation by ribosomes may result in decreased rates of degradation. Consistent with this possibility we find that the presence of multiple sequential proline codons, which greatly slow the elongation rate, results in faster mRNA degradation rates.

We also find genome-wide evidence that the GC3 content is negatively correlated with rates of mRNA degradation. These results are consistent with an earlier study that found a positive correlation between synonymous A|T dinucleotides spanning codon boundaries and mRNA degradation rates (54). To validate this result, we experimentally tested the effect of GC3 content on mRNA degradation using GFP-encoding transcripts that differ in

GC3 content. Consistent with our genome-wide analysis we find that decreasing the GC3 content results in increased mRNA degradation rates and lowered steady-state abundances. Recently, increased GC3 content has been suggested to decrease mRNA degradation rates in human lymphoblastoid cells and possibly explain some variation in mRNA degradation rates between different individuals (19). Thus, the relationship between GC3 content and mRNA degradation rates may be widely conserved in eukaryotes.

Conclusion

Our study shows that genome-wide variation in mRNA degradation rates is best explained by a combination of different transcript features as suggested more than two decades ago (49). The fact that many properties of transcripts are correlated makes it difficult to identify causative relationships; however, through careful experimentation it is possible to confirm genome-wide principles as we have shown in this study. Many of the factors that affect genome-wide patterns of mRNA degradation rates are related to protein production highlighting the close relationship between mRNA degradation and translation, which warrants further investigation.

FUNDING

This work was supported by the National Institutes of Health (GM107466 to D.G.); the National Science Foundation (MCB-1244219 to D.G.); and a DuPont Young Professor award (to D.G.).

ACKNOWLEDGEMENTS

We thank Joshua Plotkin and Grzegorz Kudla for providing the degenerate GFP constructs for our studies. We thank members of the Gresham Lab for helpful discussions and suggestions.

AUTHOR CONTRIBUTIONS

B.N. performed all computational analyses and biological validation of GC content. V.E. performed biological validation of ribosome density. B.N. and D.G. conceived of the project and wrote the paper.

REFERENCES

1. Nonet,M., Scafe,C., Sexton,J. and Young,R. (1987) Eucaryotic RNA polymerase conditional mutant that rapidly ceases mRNA synthesis. *Mol Cell Biol*, **7**, 1602–1611.
2. Anderson,J.S. and Parker,R.P. (1998) The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex. *EMBO J.*, **17**, 1497–1506.
3. Hager,G.L., McNally,J.G. and Misteli,T. (2009) Transcription Dynamics. *Mol Cell*, **35**, 741–753.
4. Parker,R. (2012) RNA degradation in *Saccharomyces cerevisiae*. *Genetics*, **191**, 671–702.
5. Decker,C.J. and Parker,R. (1993) A turnover pathway for both stable and unstable mRNAs in yeast: evidence for a requirement for deadenylation. *Genes Dev*, **7**, 1632–1643.
6. Muhlrاد,D., Decker,C.J. and Parker,R. (1995) Turnover mechanisms of the stable yeast PGK1 mRNA. *Mol Cell Biol*, **15**, 2145–2156.
7. Beelman,C.A., Stevens,A., Caponigro,G., LaGrandeur,T.E., Hatfield,L., Fortner,D.M. and Parker,R. (1996) An essential component of the decapping enzyme required for normal rates of mRNA turnover. *Nature*, **382**, 642–646.
8. Parker,R. and Jacobson,A. (1990) Translation and a 42-nucleotide segment within the coding region of the mRNA encoded by the MAT alpha 1 gene are involved in promoting rapid mRNA decay in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, **87**, 2780–2784.
9. Wisdom,R. and Lee,W. (1991) The protein-coding region of c-myc mRNA contains a sequence that specifies rapid mRNA turnover and induction by protein synthesis inhibitors. *Genes Dev*, **5**, 232–243.
10. Shaw,G. and Kamen,R. (1986) A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell*, **46**, 659–667.
11. Muhlrاد,D. and Parker,R. (1992) Mutations affecting stability and deadenylation of the yeast MFA2 transcript. *Genes Dev*, **6**, 2100–2111.
12. Olivas,W. and Parker,R. (2000) The Puf3 protein is a transcript-specific regulator of mRNA degradation in yeast. *EMBO J.*, **19**, 6602–6611.
13. Chen,C.Y., Gherzi,R., Ong,S.E., Chan,E.L., Raijmakers,R., Pruijn,G.J., Stoecklin,G., Moroni,C., Mann,M. and Karin,M. (2001) AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. *Cell*, **107**, 451–464.
14. Jing,Q., Huang,S., Guth,S., Zarubin,T., Motoyama,A., Chen,J., Di Padova,F., Lin,S.-C.,

- Gram,H. and Han,J. (2005) Involvement of microRNA in AU-rich element-mediated mRNA instability. *Cell*, **120**, 623–634.
15. Selinger,D.W., Saxena,R.M., Cheung,K.J., Church,G.M. and Rosenow,C. (2003) Global RNA half-life analysis in Escherichia coli reveals positional patterns of transcript degradation. *Genome Res*, **13**, 216–223.
16. Narsai,R., Howell,K.A., Millar,A.H., O'Toole,N., Small,I. and Whelan,J. (2007) Genome-wide analysis of mRNA decay rates and their determinants in Arabidopsis thaliana. *Plant Cell*, **19**, 3418–3436.
17. Thomsen,S., Anders,S., Janga,S.C., Huber,W. and Alonso,C.R. (2010) Genome-wide analysis of mRNA decay patterns during early Drosophila development. *Genome Biol*, **11**, R93.
18. Rabani,M., Levin,J.Z., Fan,L., Adiconis,X., Raychowdhury,R., Garber,M., Gnirke,A., Nusbaum,C., Hacohen,N., Friedman,N., et al. (2011) Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat Biotechnol*, **29**, 436–442.
19. Duan,J., Shi,J., Ge,X., Dölken,L., Moy,W., He,D., Shi,S., Sanders,A.R., Ross,J. and Gejman,P.V. (2013) Genome-wide survey of interindividual differences of RNA stability in human lymphoblastoid cell lines. *Sci Rep*, **3**, 1318.
20. Wang,Y., Liu,C.L., Storey,J.D., Tibshirani,R.J., Herschlag,D. and Brown,P.O. (2002) Precision and functional specificity in mRNA decay. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 5860–5865.
21. Grigull,J., Mnaimneh,S., Pootoolal,J., Robinson,M.D. and Hughes,T.R. (2004) Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays reveals posttranscriptional control of ribosome biogenesis factors. *Mol Cell Biol*, **24**, 5534–5547.
22. Shalem,O., Dahan,O., Levo,M., Martinez,M.R., Furman,I., Segal,E. and Pilpel,Y. (2008) Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. *Mol Syst Biol*, **4**, 223.
23. García-Martínez,J., Aranda,A. and Pérez-Ortín,J.E. (2004) Genomic run-on evaluates transcription rates for all yeast genes and identifies gene regulatory mechanisms. *Mol Cell*, **15**, 303–313.
24. Miller,C., Schwalb,B., Maier,K., Schulz,D., Dümcke,S., Zacher,B., Mayer,A., Sydow,J., Marcinowski,L., Dölken,L., et al. (2011) Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Mol Syst Biol*, **7**, 458.
25. Munchel,S.E., Shultzaberger,R.K., Takizawa,N. and Weis,K. (2011) Dynamic profiling of mRNA turnover reveals gene-specific and system-wide regulation of mRNA decay. *Mol*

Biol Cell, 10.1091/mbc.E11-01-0028.

26. Neymotin,B., Athanasiadou,R. and Gresham,D. (2014) Determination of in vivo RNA kinetics using RATE-seq. *RNA*, **20**, 1645–1652.
27. Yang,E., van Nimwegen,E., Zavolan,M., Rajewsky,N., Schroeder,M., Magnasco,M. and Darnell,J.E. (2003) Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. *Genome Res*, **13**, 1863–1872.
28. Eck,S. and Stephan,W. (2008) Determining the relationship of gene expression and global mRNA stability in *Drosophila melanogaster* and *Escherichia coli* using linear models. *Gene*, **424**, 102–107.
29. Ghaemmamghami,S., Huh,W.-K., Bower,K., Howson,R.W., Belle,A., Dephoure,N., O'Shea,E.K. and Weissman,J.S. (2003) Global analysis of protein expression in yeast. *Nature*, **425**, 737–741.
30. Belle,A., Tanay,A., Bitincka,L., Shamir,R. and O'Shea,E.K. (2006) Quantification of protein half-lives in the budding yeast proteome. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 13004–13009.
31. Lipson,D., Raz,T., Kieu,A., Jones,D.R., Giladi,E., Thayer,E., Thompson,J.F., Letovsky,S., Milos,P. and Causey,M. (2009) Quantification of the yeast transcriptome by single-molecule sequencing. *Nat Biotechnol*, **27**, 652–658.
32. Pelechano,V., Chávez,S. and Pérez-Ortín,J.E. (2010) A complete set of nascent transcription rates for yeast genes. *PLoS ONE*, **5**, e15442.
33. Nagalakshmi,U., Wang,Z., Waern,K., Shou,C., Raha,D., Gerstein,M. and Snyder,M. (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, **320**, 1344–1349.
34. Ingolia,N.T., Ghaemmamghami,S., Newman,J.R.S. and Weissman,J.S. (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, **324**, 218–223.
35. 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 Biol*, **6**, e255.
36. Garí,E., Piedrafita,L., Aldea,M. and Herrero,E. (1997) A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in *Saccharomyces cerevisiae*. *Yeast*, **13**, 837–848.
37. Kudla,G., Murray,A.W., Tollervey,D. and Plotkin,J.B. (2009) Coding-sequence determinants of gene expression in *Escherichia coli*. *Science*, **324**, 255–258.

38. Hong,J. and Gresham,D. (2014) Molecular specificity, convergence and constraint shape adaptive evolution in nutrient-poor environments. *PLoS Genet.*, **10**, e1004041.
39. Zenklusen,D., Larson,D.R. and Singer,R.H. (2008) Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol.*, **15**, 1263–1271.
40. Sharp,P.M. and Li,W.H. (1987) The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res*, **15**, 1281–1295.
41. Gardin,J., Yeasmin,R., Yurovsky,A., Cai,Y., Skiena,S. and Fitcher,B. (2014) Measurement of average decoding rates of the 61 sense codons in vivo. *Elife*, **3**.
42. Edri,S. and Tuller,T. (2014) Quantifying the effect of ribosomal density on mRNA stability. *PLoS ONE*, **9**, e102308.
43. Wishart,J.A., Hayes,A., Wardleworth,L., Zhang,N. and Oliver,S.G. (2005) Doxycycline, the drug used to control the tet-regulatable promoter system, has no effect on global gene expression in *Saccharomyces cerevisiae*. *Yeast*, **22**, 565–569.
44. Kudla,G., Lipinski,L., Caffin,F., Helwak,A. and Zylicz,M. (2006) High guanine and cytosine content increases mRNA levels in mammalian cells. *PLoS Biol*, **4**, e180.
45. Schoenberg,D.R. and Maquat,L.E. (2012) Regulation of cytoplasmic mRNA decay. *Nat Rev Genet*, **13**, 246–259.
46. Feng,L. and Niu,D.-K. (2007) Relationship between mRNA stability and length: an old question with a new twist. *Biochem. Genet.*, **45**, 131–137.
47. Dressaire,C., Picard,F., Redon,E., Loubière,P., Queinnec,I., Girbal,L. and Coccagn-Bousquet,M. (2013) Role of mRNA stability during bacterial adaptation. *PLoS ONE*, **8**, e59059.
48. Geisberg,J.V., Moqtaderi,Z., Fan,X., Oszolak,F. and Struhl,K. (2014) Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. *Cell*, **156**, 812–824.
49. Caponigro,G., Muhlrud,D. and Parker,R. (1993) A small segment of the MAT alpha 1 transcript promotes mRNA decay in *Saccharomyces cerevisiae*: a stimulatory role for rare codons. *Mol Cell Biol*, **13**, 5141–5148.
50. Yun,D.-F. and Sherman,F. (1996) Degradation of CYC1 mRNA in the yeast *Saccharomyces cerevisiae* does not require translation. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 8895–8900.
51. Beelman,C.A. and Parker,R. (1994) Differential effects of translational inhibition in cis and in trans on the decay of the unstable yeast MFA2 mRNA. *J Biol Chem*, **269**, 9687–

9692.

52. Merhi,A., Gérard,N., Lauwers,E., Prévost,M. and André,B. (2011) Systematic mutational analysis of the intracellular regions of yeast Gap1 permease. *PLoS ONE*, **6**, e18457.
53. Plotkin,J.B. and Kudla,G. (2011) Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet*, **12**, 32–42.
54. Carlini,D.B. (2005) Context-dependent codon bias and messenger RNA longevity in the yeast transcriptome. *Mol. Biol. Evol.*, **22**, 1403–1411.

FIGURE LEGENDS

Figure 1. Multiple factors are associated with variation in mRNA degradation rates. **(a)** Individual predictors explain different amounts of the variation in mRNA degradation rates determined using RATE-seq. Multiple regression models including continuous variables (model) and continuous and categorical variables (model++) explain around half the variation in mRNA degradation rates. Positive correlations are indicated in red whereas negative correlations are indicated in green. Values corresponding to ANOVA and multiple regression are in blue. **(b)** A comparison of modeled mRNA degradation rates with measured mRNA degradation rates shows that the model behaves similarly across the entire range of mRNA degradation rates. **(c)** Multiple regression models applied to published mRNA degradation rates explain different amounts of the variation. Less variance can be explained for mRNA degradation rates that rely on transcriptional inhibition. Genomic Run On is indicated with GRO. Dynamic Transcriptome Analysis is indicated with DTA. Labeling studies using thiouracil or thiouridine are indicated in parentheses with 4tU and 4sU respectively.

Figure 2. The presence of multiple proline codons affects degradation rates and protein production. **(a)** Proline-rich proteins, which are translated more slowly, tend to have more stable transcripts than proline-poor proteins. **(b)** The steady-state abundance of poly-proline containing proteins is reduced compared to the global distribution of protein abundances.

Figure 3. The effect of ribosome density on *GAP1* mRNA degradation rate. **(a)** Mutation of the start codon in *GAP1* from ATG to GTG results in a reduced steady state transcript level **(a)** and **(b)** increased mRNA degradation. In **(b)** we show the average value for each time point \pm SEM. In **(c)** we show bootstrapped confidence intervals for the regression of all data points. Solid lines indicate the line of best fit and dotted lines indicate confidence intervals. In blue is the transcript with a normal start codon. In red is the transcript with a mutated start codon.

Figure 4: GC3 content affects degradation kinetics and steady state levels. **(a)** Four GFP transcripts containing synonymous mutations span a range of GC3 content (left) and overall GC (right). Alignment of each GFP relative to GFP2. Positions of similarity in sequence are depicted by gray line and differences are in triangles. Colors indicate different GFP constructs. **(b)** Differences in GC3 content affect steady state levels of transcripts. **(c,d)** GFP2, GFP3 and GFP4 degrade in a GC3-dependent manner. In **(c)** we show the average value for each time point \pm SEM. In **(d)** we show bootstrapped 95% confidence intervals for the regression of all data points. Solid lines indicate the line of best fit and dotted lines indicate confidence intervals.

Table 1. Parameters included in model.

Predictor	N	Units	Transformation	Reference
Coding Length	5850	nucleotides	$\log_{10}(\text{nucleotides})$	SGD
3'UTR length	4950	nucleotides	$\log_{10}(\text{nucleotides}+1)$	SGD/ (33)
5'UTR length	4345	nucleotides	$\log_{10}(\text{nucleotides}+1)$	SGD/ (33)
5'UTR GC content	4345	proportion	none	SGD/(33)
3'UTR GC content	4911	proportion	none	SGD/(33)
Abundance	5488	rpM	$\log_{10}((\text{rpM}/\text{total reads}) * 60,000)$	(31)
Protein/ cell	3818	protein/cell	$\log_{10}(\text{protein/cell})$	(29)
Ribosome density	5269	rpkM	$\log_{10}(\text{rpkM}+1)$	(34)
Transcription rate	4346	molecules/min	$\log_{10}(\text{molecules/min})$	(32)
Coding GC position 1	5850	proportion	none	SGD
Coding GC position 2	5850	proportion	none	SGD
Coding GC position 3	5850	proportion	none	SGD
Protein half-life	3164	min	$\log_{10}(\text{min})$	(30)
deltaG	5850	kcal/mol	none	SGD/GeneRFold
CAI	5850	relative scale	$\log_{10}(\text{relative scale})$	CAI function in R
Protein/ mRNA	3583	protein/cell/ transcript	$\log_{10}(\text{protein/cell/transcript}+1)$	(29, 31)

SUPPLEMENTAL FIGURES

Figure S1. Multiple predictors best explain variation in mRNA degradation. We show the adjusted R^2 for individual predictors in each model. Predictors are as follows: 1=Coding length, 2=3'UTR length, 3=5'UTR length, 4=5'UTR GC content, 5=3'UTR GC content, 6=mRNA abundance, 7=Protein per cell, 8=Ribosome density, 9=Transcription rate, 10=GC1, 11=GC2, 12=GC3, 13=Protein half-life, 14=deltaG, 15=CAI, 16=Protein per mRNA, 17=GO, 18=RNABP, 19=model with continuous variables, and 20=model with continuous variables +GO +RNABP.

Figure S2. Predictors of mRNA degradation variation are highly correlated. We show that a number of predictors are highly correlated.

Figure S3. Degradation rate decreases with increased ribosome density. The ribosome density of a transcript is inversely correlated with degradation rate, showing that the more ribosomes present, the more stable the transcript.

Figure S4. GFP1 and GFP2 are similar in their degradation kinetics.

Figure S5. Effect of mutated start codon on GFP protein. We grew yeast in minimal media containing D-histidine and D-serine. D-his is toxic to cells and leads to cell growth inhibition and death. In the absence of a functional GAP1 protein, cells do not have the permease and do not take up D-his. We show five strains: GAP1-WT=wild type strain, GAP1-KO=strain with complete knockout of coding sequence of GAP1, GAP1-NF=Derived strain with non-functional GAP1, GAP1-ATG=plasmid GAP1 with normal start codon, and GAP1-GTG=plasmid GAP1 with GTG in place of start codon.

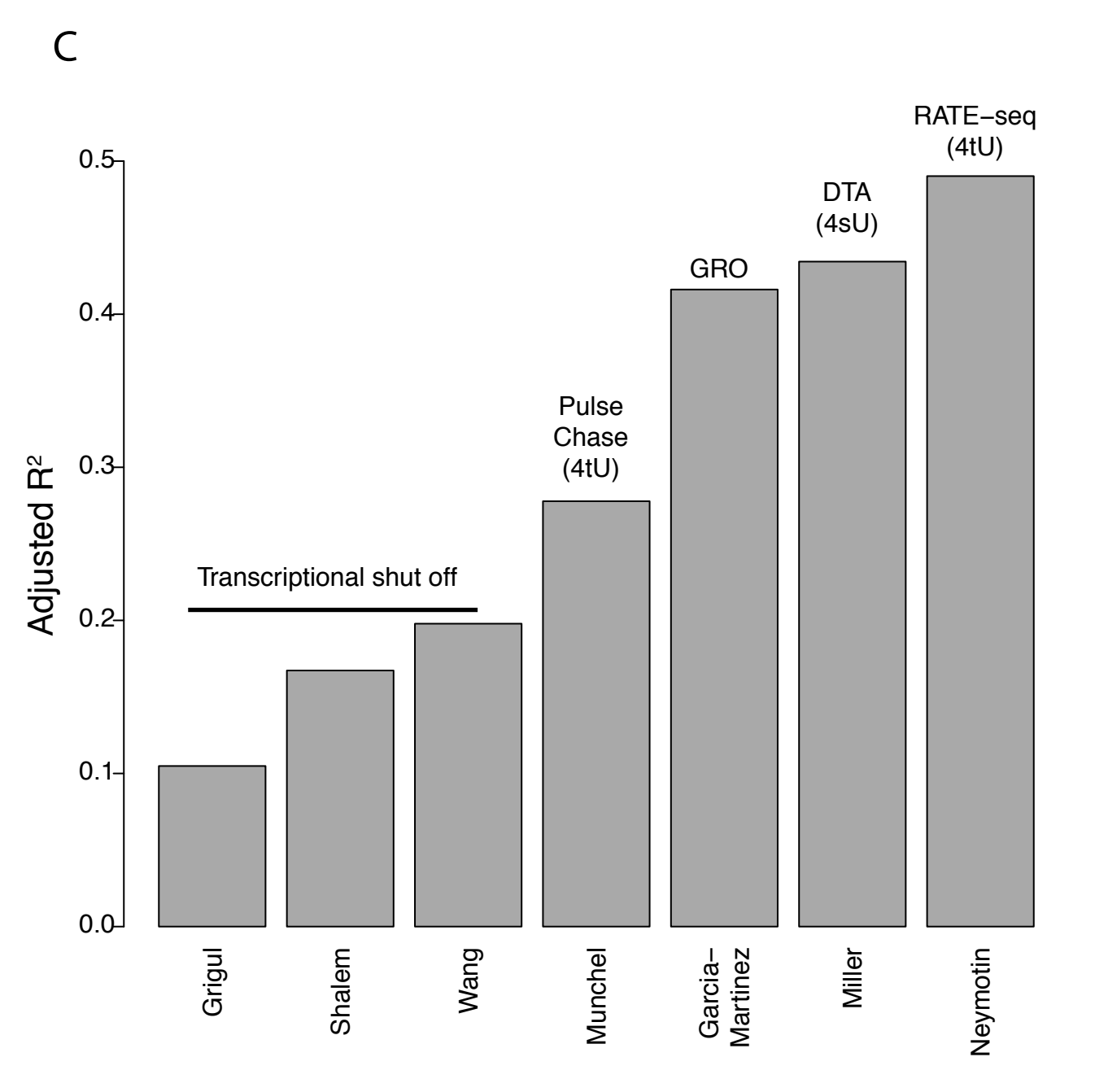
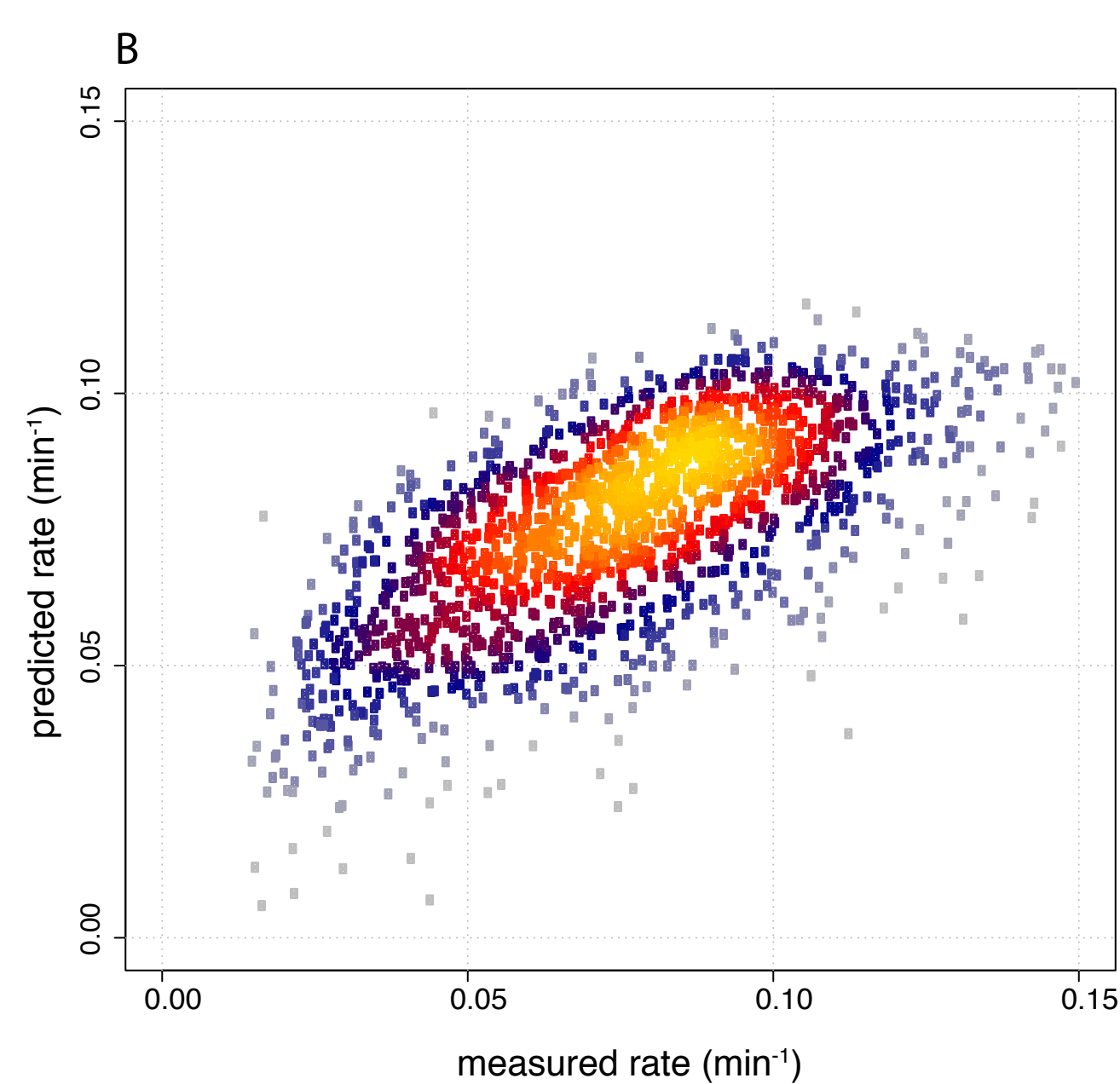
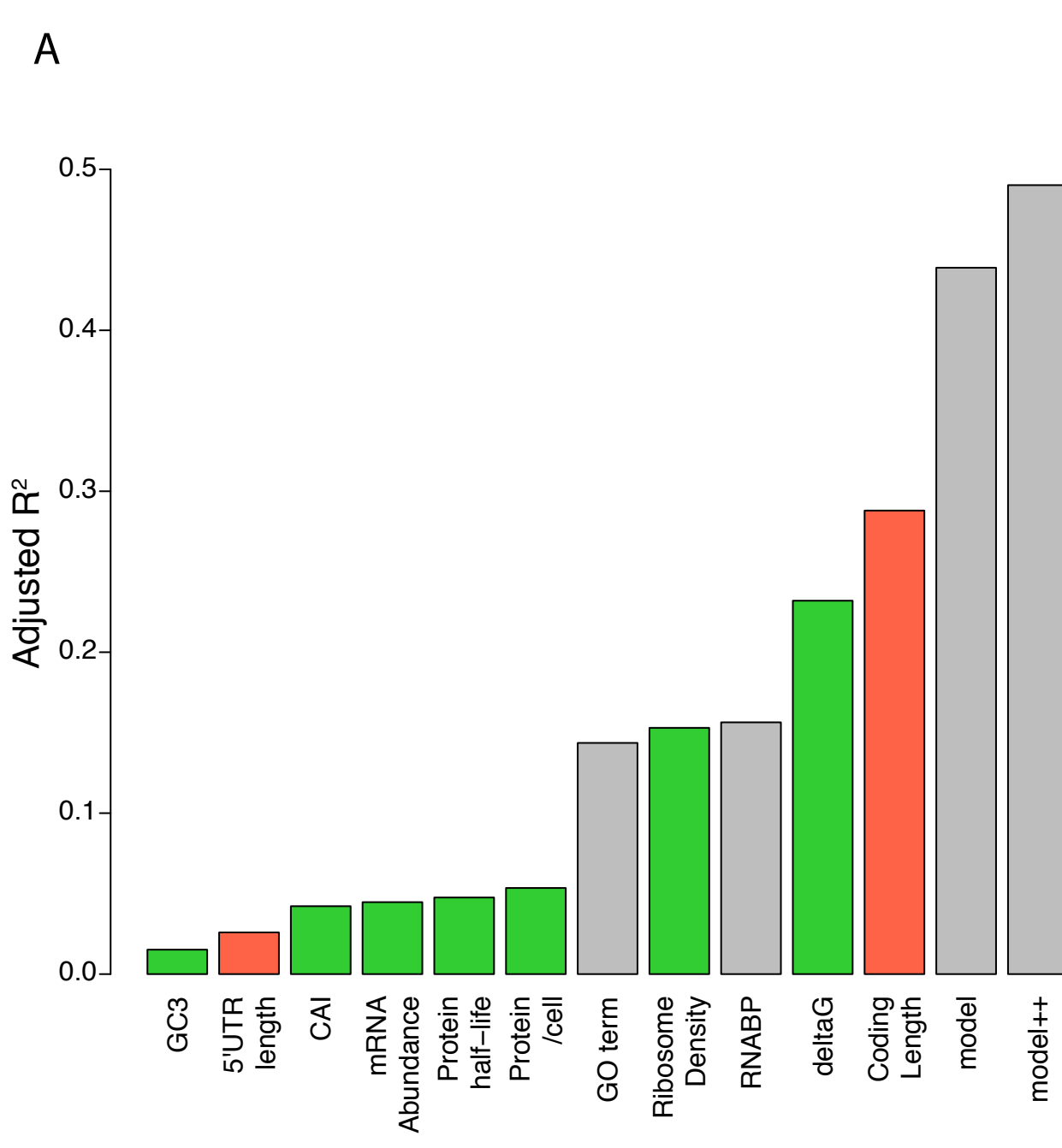
Figure S6. Diagnostics of the model fit.

SUPPLEMENTAL TABLES

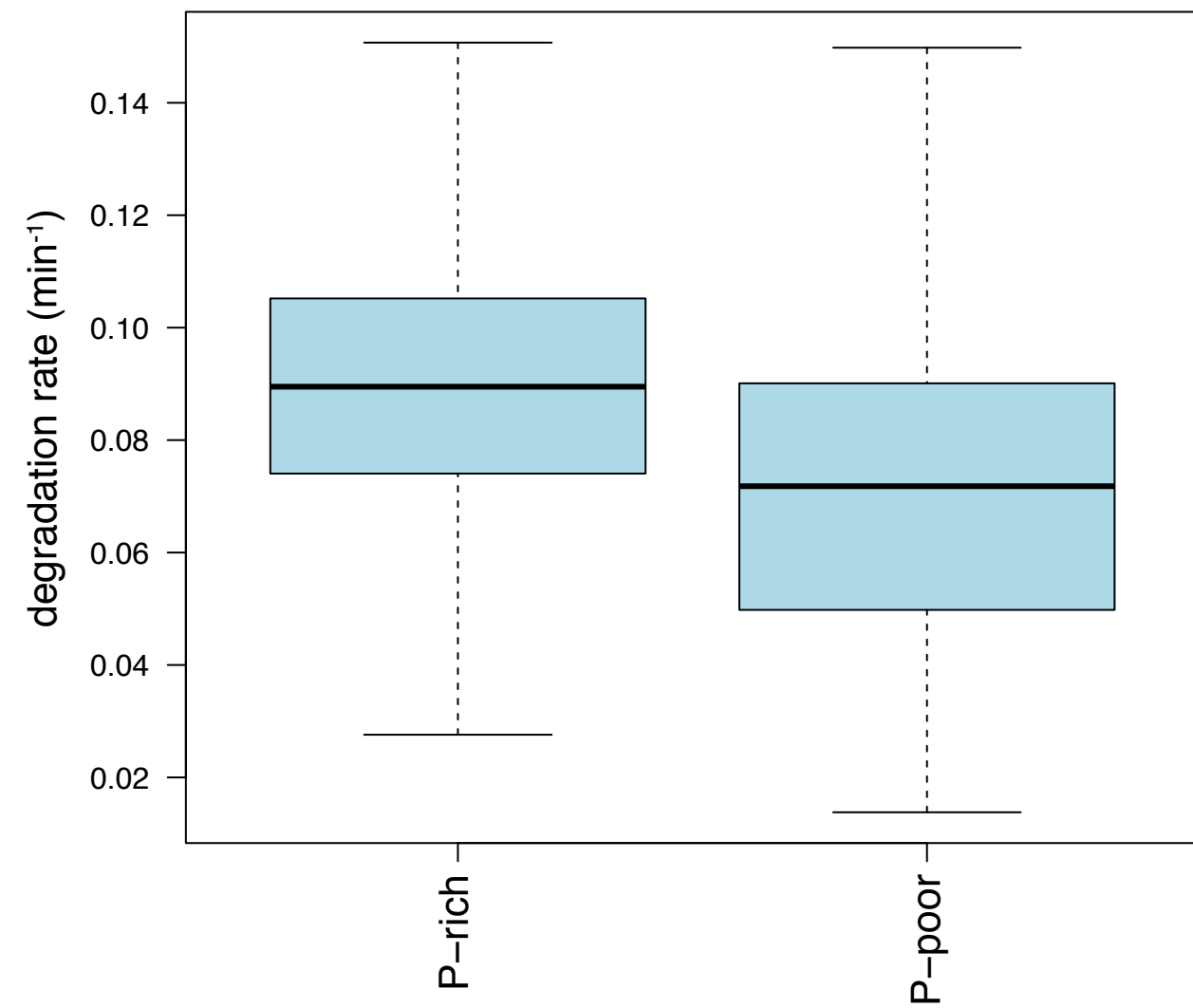
Table S1. Matrix with all parameters included in regression analysis.

Table S2. Matrix with p-value, adjusted R^2 , and correlation for each predictor and reported mRNA degradation rate.

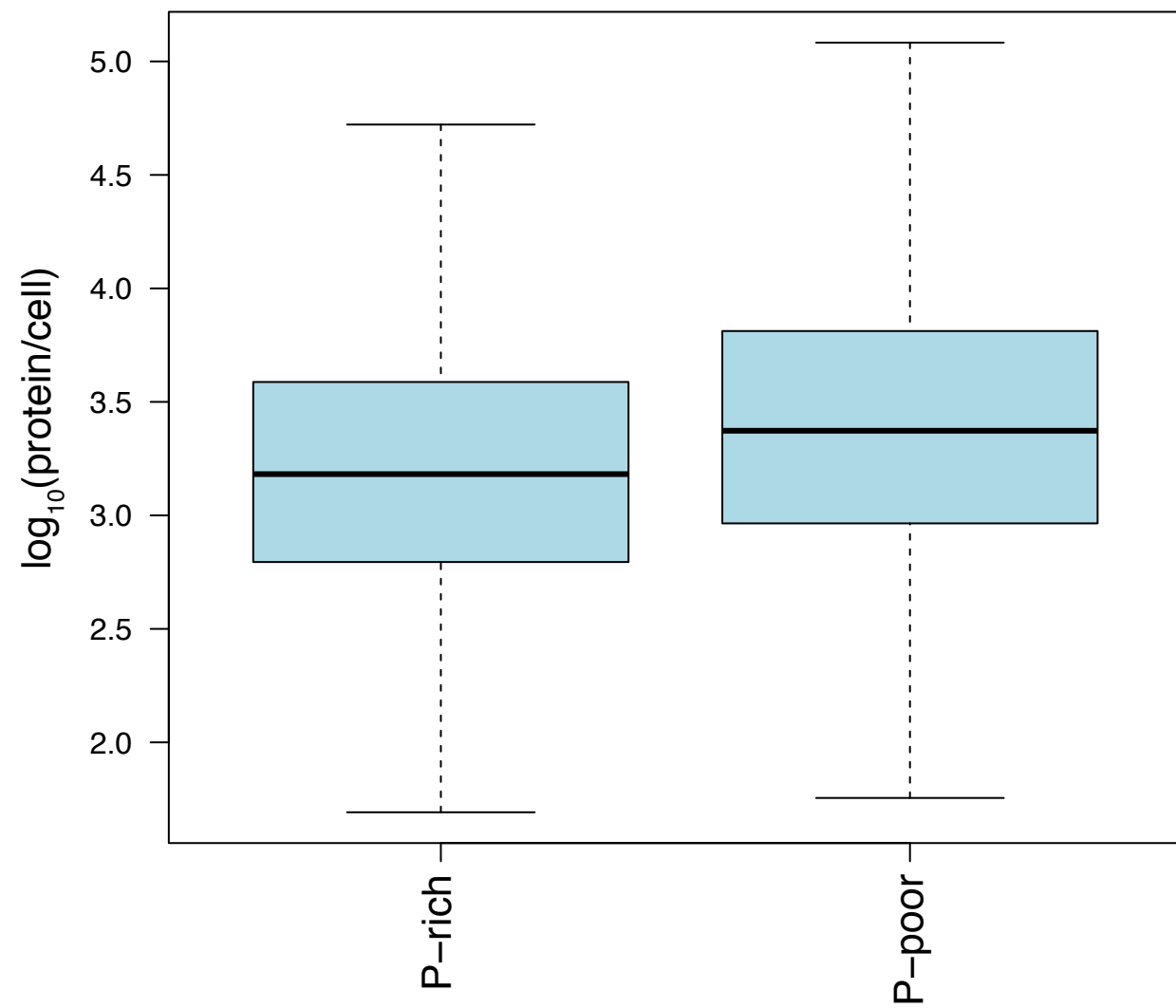
Table S3. Data for qPCR experiments.



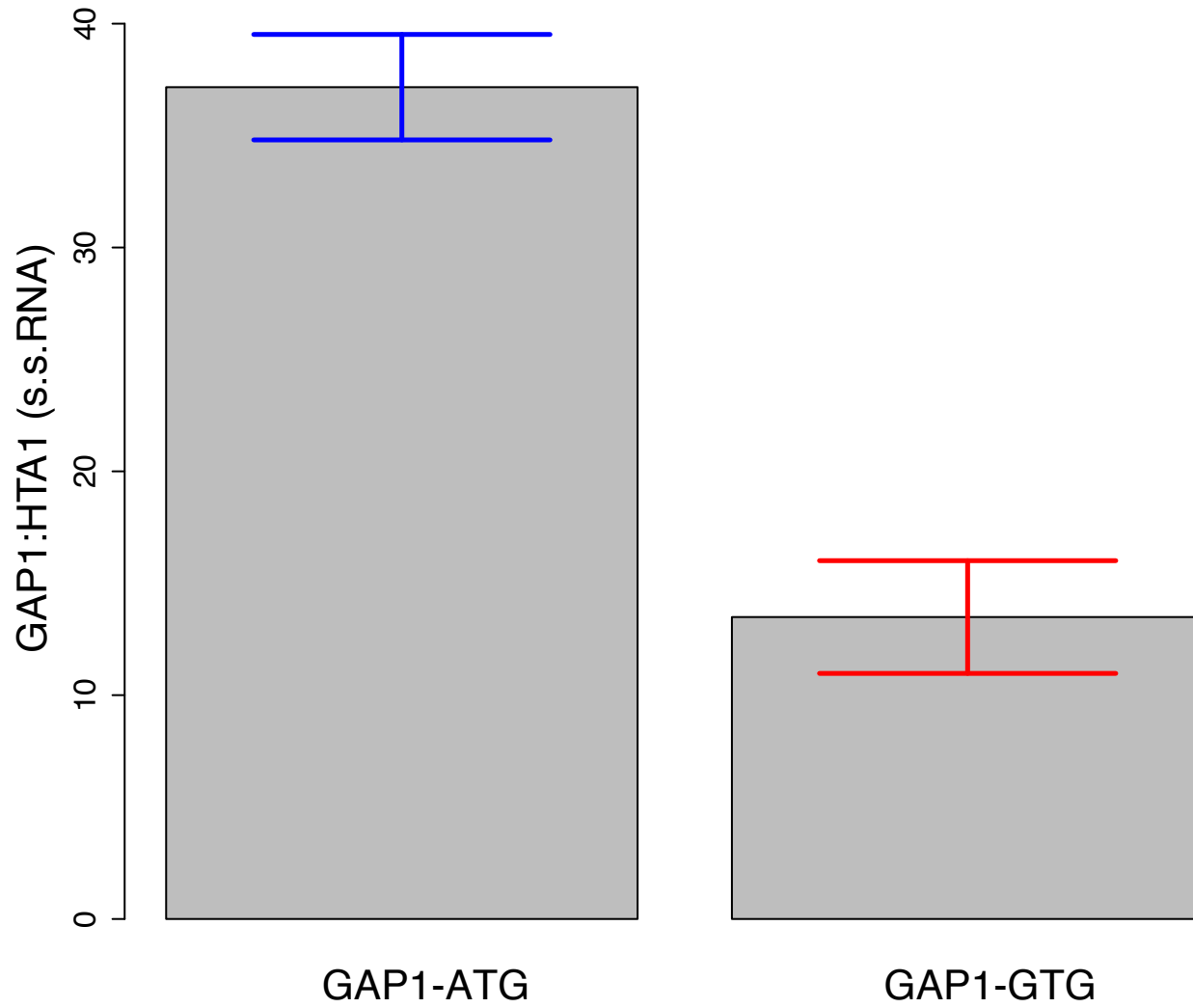
A



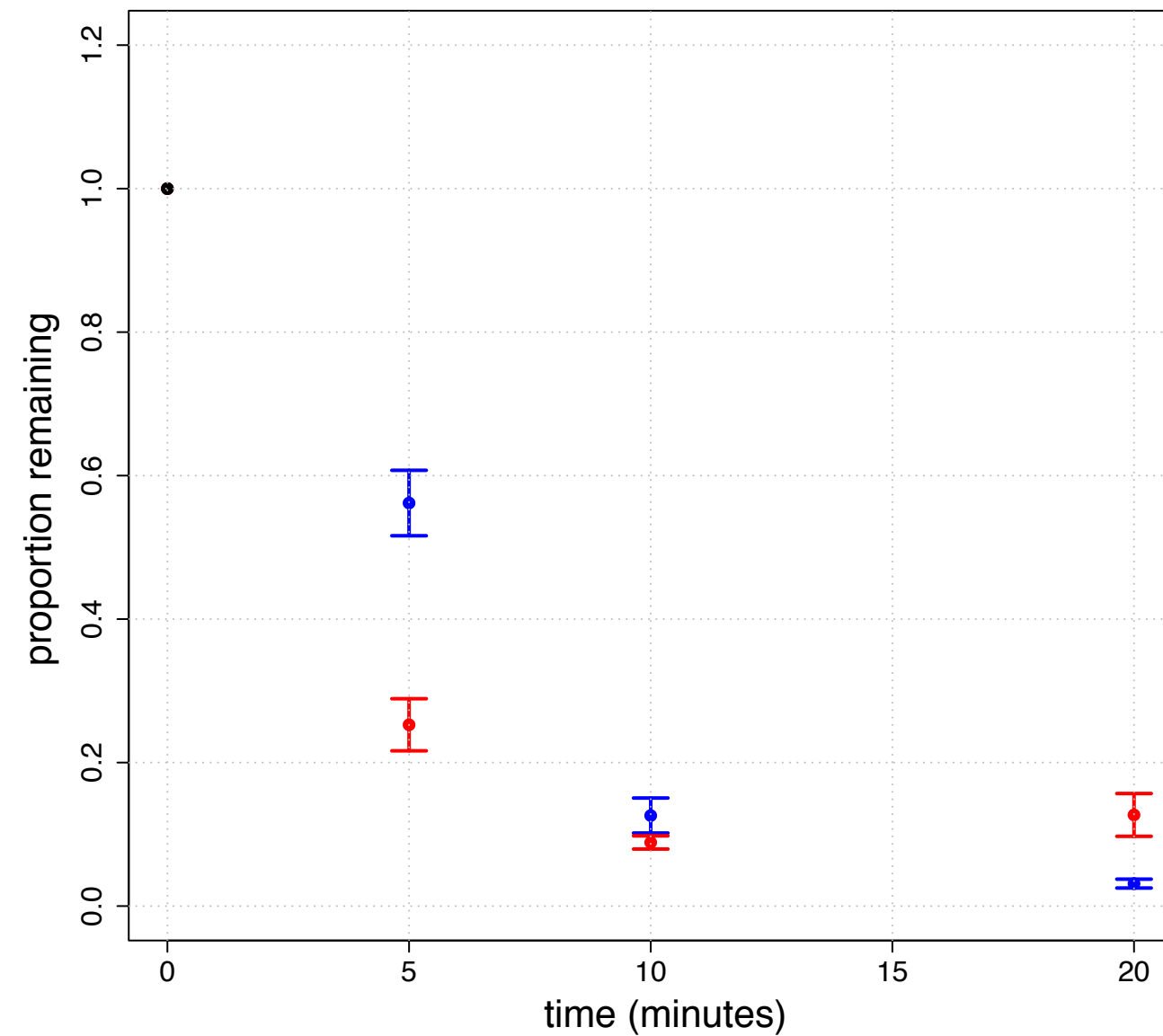
B



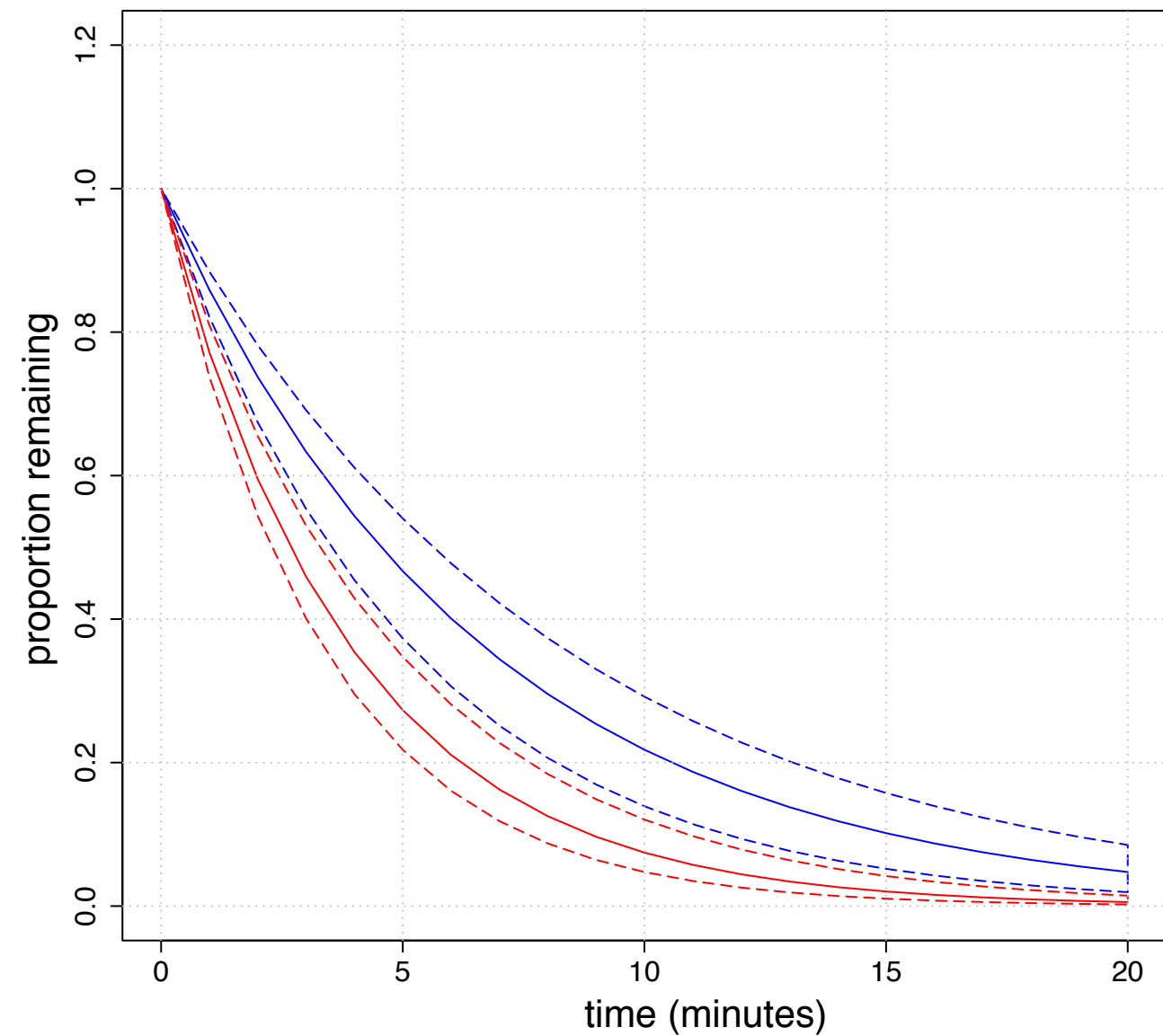
A



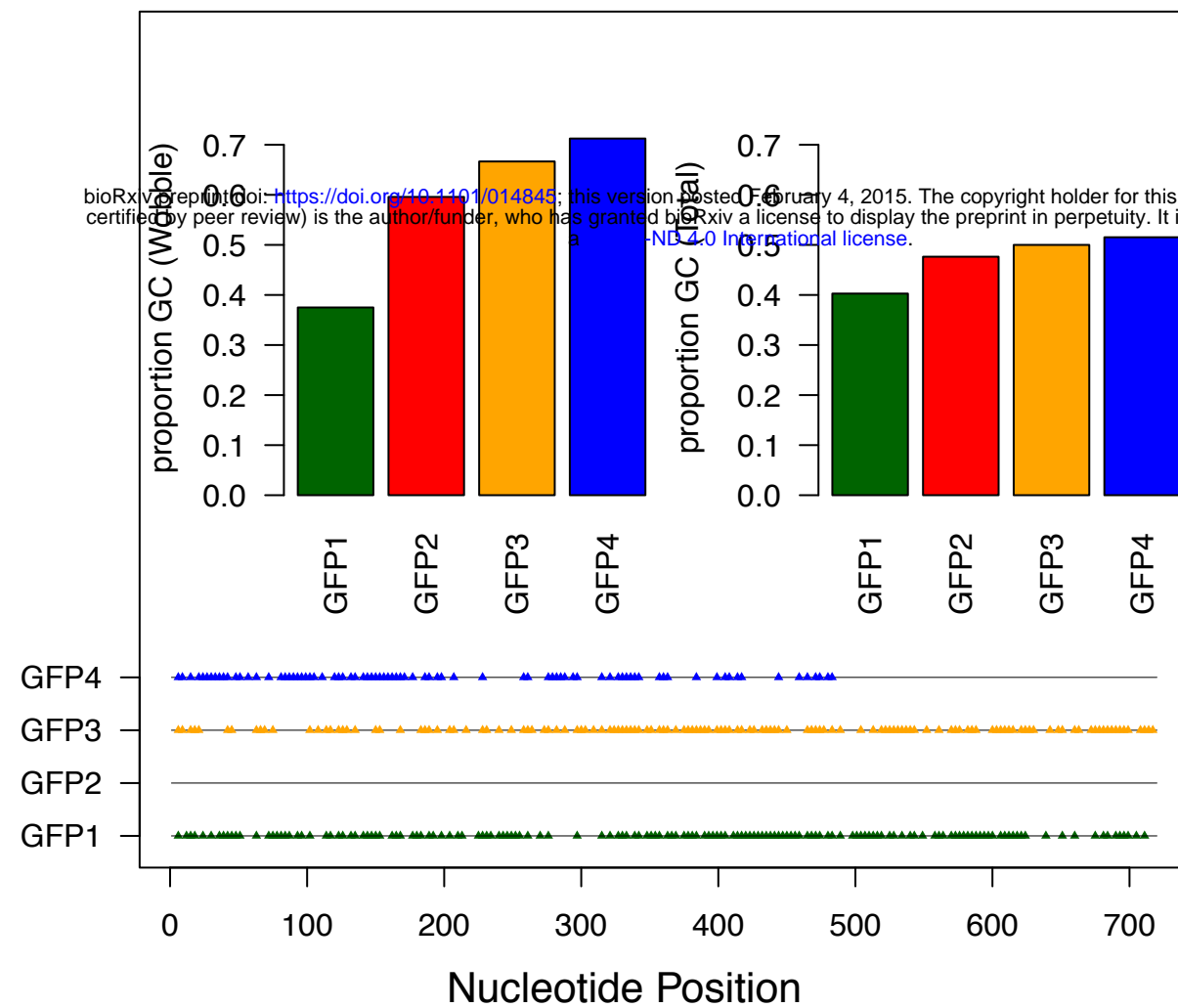
B



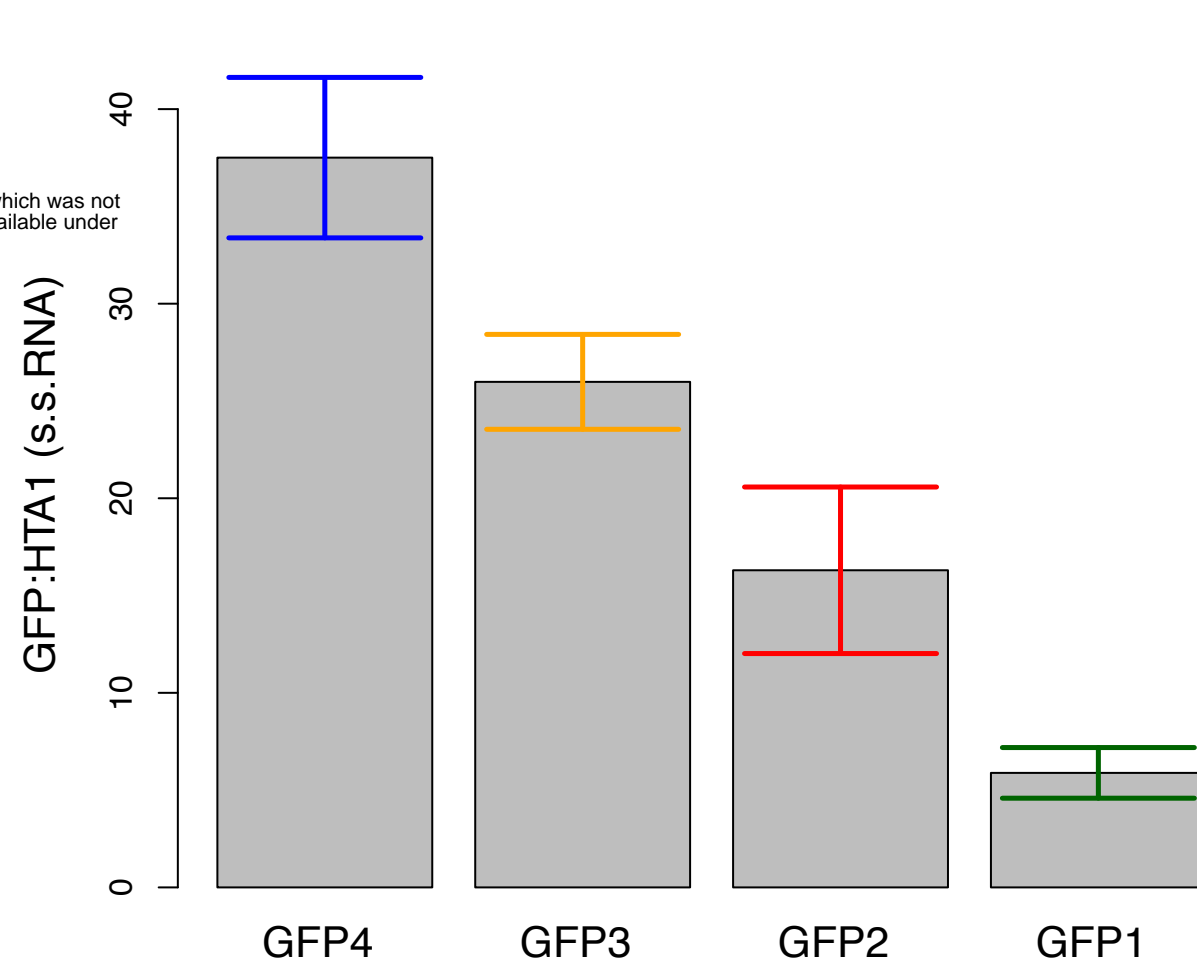
C



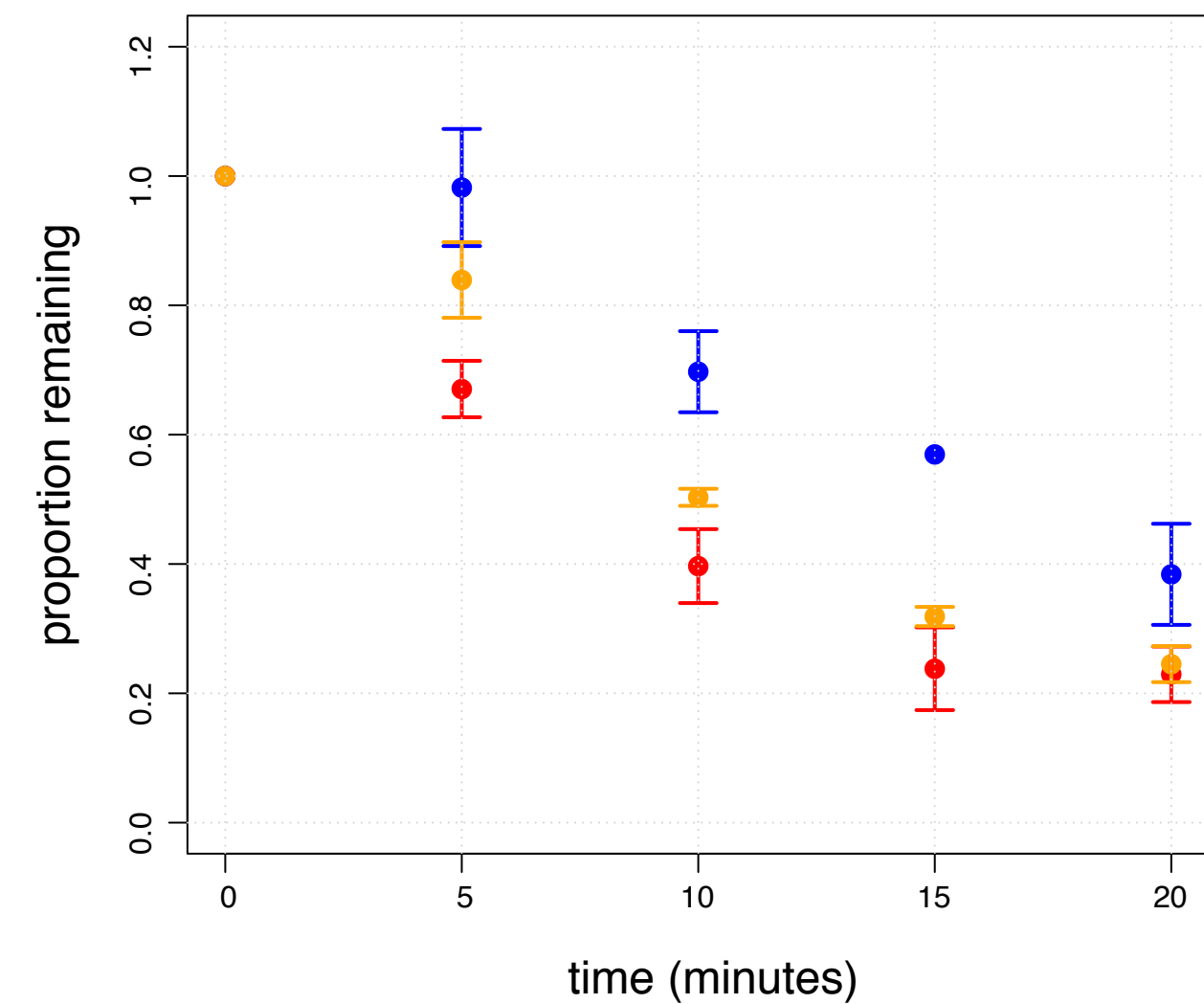
A



B



C



D

