Determinants of Translational Efficiency in *Saccharomyces Cerevisiae*

By

Boris Zinshteyn

B.A. Biochemistry
M.S. Chemistry
University of Pennsylvania, 2009

Submitted to the Department of Biology In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

At the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2015

© 2015 Massachusetts Institute Of Technology. All rights reserved

Signature of Author: _____
Department of Biology
January 15, 2015

Certified by: _____
Wendy Gilbert
Associate Professor of Biology
Thesis Supervisor

Accepted by: _____
Amy Keating
Associate Professor of Biology
Chair, Graduate Committee

Determinants of Translational Efficiency in *Saccharomyces Cerevisiae*

By

Boris Zinshteyn

Submitted to the Department of Biology
On January 15, 2015 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Abstract

The goal of this thesis is to elucidate the mechanisms that govern translational efficiency (TE) - the amount of protein produced from each molecule of mRNA. While the mechanisms regulating the TE of a few specific messages are well understood, the general contribution of translational control to differences in cellular protein levels is currently unclear. Recent advances have enabled the direct measurement of protein levels and translation rates genome-wide, and studies in multiple organisms have found varying degrees of translation regulation, both at steady state, and in response to stress or developmental cues. Despite this influx of high-throughput data, the mechanisms underlying the differences in gene-specific and condition-dependent TE remain largely unknown.

In this thesis, I describe the roles of two different components of the translational machinery in regulating translational efficiency. In Chapter 1, I discuss the features of mRNA coding sequences that can affect TE, thereby introducing Chapter 2, in which I investigate the role of a conserved anticodon tRNA modification in determining the rate of translation elongation and the phenotypic consequences of its loss for budding yeast. In Chapter 3, I discuss the regulation of translation initiation to introduce Chapter 4, in which I explore how the RNA binding specificity of the core translation factor, yeast eukaryotic initiation factor 4G (eIF4G), contributes to genome-wide competition between mRNAs. Finally, I will discuss future directions for this work.

Thesis Supervisor: Wendy Gilbert
Title: Associate Professor of Biology

Acknowledgements

I would first like to thank my PhD advisor, Wendy Gilbert, for her guidance, encouragement and enthusiasm throughout this long process. I must acknowledge my thesis committee, Uttam RajBhandary and Dave Bartel, for providing sound scientific and career advice, and for always agreeing to meet with me, even on short notice. I am also grateful to Allan Jacobson for serving on my defense committee.

To all members of the Gilbert Lab, past and present: you made this a wonderful place to spend time and do science. Thank you to Mary Kay Thompson and Joshua Arribere, for paving the way as the lab's first grad students. A special thank you to Thomas Carlile, for being the best snoRNbAymate I could ask for. He saved me from both frostbite and eye injury, and answered all my questions about yeast many times over. Thank you Pavan Vaidyanathan, Maria Rojas-Duran, Kristen Bartoli, Audra Amasino, Julia Wang and Gina Mawla for all of the enjoyable scientific and social interactions we've had.

I am indebted to my truly amazing friends and classmates at MIT, who rarely hesitate to take the time to talk over an experiment, show me a new technique, or play some board games.

This endeavor would not have even begun if not for the support of my family, who had the courage to leave everything they knew behind and seek out a better life. Thank you Yan, Larisa, Elena, Misha, Tatyana and Mark. You have been a source of great inspiration and strength. I also thank my brother Daniel, who has engaged me in many hours of honest and unfiltered scientific discussion and debate.

Lastly, and most importantly, I must thank my beloved wife Laura for putting up with the odd hours, the missed dinners, and the terrible cell phone reception for all of these years; for being a finder of lost things, and the best partner I could ever ask for.

Biographical Note

EDUCATION

MIT, Department of Biology, Cambridge, MA
Ph.D. Candidate in Biology 2009-2015
University of Pennsylvania, Philadelphia, PA
B.A. with Distinction in Biochemistry 2009
Magna Cum Laude
Roy and Diana Vagelos Science Scholar
Minor in Mathematics
Minor in Computer & Information Science
M.S. in Chemistry 2009

RESEARCH EXPERIENCE

MIT Department of Biology, Cambridge, MA
Graduate Student with Professor Wendy Gilbert 2010-2015
The determinants of translational efficiency in *S. cerevisiae*

The Wistar Institute, Philadelphia, PA
Research Assistant with Professor Kazuko Nishikura 2005-2009
The relationship between Adenosine Deaminases that Act on RNA and microRNA processing and function.

PUBLICATIONS

Carlile TM, Rojas-Duran MF, Zinshteyn B, Shin H, Bartoli KM, and Gilbert WV. Pseudouridine profiling reveals widespread regulated mRNA pseudouridylation in yeast and human cells. Nature (2014) Nov 6;515(7525):143-6

Vaidyanathan, PP, Zinshteyn, B, Thompson, MK, and Gilbert, WV (2014). Protein kinase A regulates gene-specific translational adaptation in differentiating yeast. RNA.

Zinshteyn B, Gilbert WV (2013) Loss of a Conserved tRNA Anticodon Modification Perturbs Cellular Signaling. PLoS Genet 9(8): e1003675. doi:10.1371/journal.pgen.1003675

Valente L, Kawahara Y, Zinshteyn B, Iizasa H, Nishikura K. Posttranscriptional Gene Regulation by an Editor: ADAR and its Role in RNA Editing. *Post-transcriptional Gene regulation, RNA processing in Eukaryotes*. J Wu Ed. John Wiley & Sons, (2013). 41-

Zinshteyn, B., and Nishikura, K. Adenosine-to-inosine RNA editing. *Wiley Interdiscip Rev Syst Biol Med*. 2009 Sep-Oct;1(2):202-9.

Kawahara Y*, Zinshteyn B*, Sethupathy P, Iizasa H, Hatzigeorgiou AG, Nishikura K. Redirection of Silencing Targets by Adenosine-to-Inosine Editing of miRNAs. *Science* 315: 1137-1140 (2007) [* these authors contributed equally]

Kawahara Y, Zinshteyn B, Chemdrimada TP, Shiekhattar R, Nishikura, K. RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex. *EMBO Reports* 8: 763-769 (2007)

TEACHING EXPERIENCE

MIT, Cambridge, MA

Teaching Assistant, 7.91: Computational & Systems Biology	Spring 2013
Teaching Assistant, 7.012: Introduction to Biology	Fall 2010

AWARDS AND HONORS

Poster Prize MIT Biology, Building 68 retreat	2014
Henry and Francis Keany Rickard Fund Fellowship MIT Office of the Dean for Graduate Education	2013
Rose Award for Undergraduate Research University of Pennsylvania School of Arts and Sciences	2009
PGFI Excellence in Genomics Undergraduate Award Penn Genome Frontiers Institute	2009
John C. Makris Memorial Award University of Pennsylvania Department of Biochemistry	2009
Phi-Beta-Kappa	2009
Vagelos Award for Undergraduate Research University of Pennsylvania	2007

Table of Contents

Thesis Overview

Each step of eukaryotic gene expression – from transcription into messenger RNA (mRNA), RNA processing, export to the cytoplasm, and finally translation into protein – is regulated to ensure that cells have the right amount of each protein required to survive and function. The goal of this thesis is to elucidate the mechanisms that govern translational efficiency (TE) - the amount of protein produced from each molecule of mRNA.

While the mechanisms regulating the TE of a few specific messages are well understood, the general contribution of translational control to differences in cellular protein levels is currently unclear. Recent advances have enabled the direct measurement of protein levels and translation rates genome-wide (Ghaemmaghami et al., 2003; Ingolia et al., 2009), and studies in multiple organisms have found varying degrees of translation regulation, both at steady state (Albert et al., 2014; Li et al., 2014; McManus et al., 2014; Quax et al., 2013; Stadler and Fire, 2011), and in response to stress or developmental cues (Brar et al., 2012; Guo et al., 2010; Ingolia et al., 2011; Stadler and Fire, 2013; Stumpf et al., 2013; Subtelny et al., 2014; Vaidyanathan et al., 2014). Despite this influx of high-throughput data, the mechanisms underlying the differences in gene-specific and condition-dependent TE remain largely unknown.

In this thesis, I describe the roles of both elongation and initiation rates in regulating TE. In Chapter 1, I discuss the features of mRNA coding sequences that can affect TE, thereby introducing Chapter 2, in which I investigate the role of a conserved anticodon tRNA modification in determining the rate of translation elongation and the

phenotypic consequences of its loss for budding yeast. In Chapter 3, I discuss the regulation of translation initiation to introduce Chapter 4, in which I explore how the RNA binding specificity of the core translation factor, yeast eukaryotic initiation factor 4G (eIF4G), contributes to genome-wide competition between mRNAs. Finally, in chapter 5 I will discuss future directions for this work.

References

Albert, F.W., Muzzey, D., Weissman, J.S., and Kruglyak, L. (2014). Genetic influences on translation in yeast. PLoS Genet *10*, e1004692.

Brar, G.A., Yassour, M., Friedman, N., Regev, A., Ingolia, N.T., and Weissman, J.S. (2012). High-Resolution View of the Yeast Meiotic Program Revealed by Ribosome Profiling. Science *335*, 552–557.

Ghaemmaghami, 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.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835–840.

Ingolia, N.T., Ghaemmaghami, 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.

Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell *147*, 789–802.

Li, G.-W., Burkhardt, D., Gross, C., and Weissman, J.S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell *157*, 624–635.

McManus, C.J., May, G.E., Spealman, P., and Shteyman, A. (2014). Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. Genome Research *24*, 422–430.

Quax, T.E.F., Wolf, Y.I., Koehorst, J.J., Wurtzel, O., van der Oost, R., Ran, W., Blombach, F., Makarova, K.S., Brouns, S.J.J., Forster, A.C., et al. (2013). Differential Translation Tunes Uneven Production of Operon-Encoded Proteins. Cell Reports *4*, 938–944.

Stadler, M., and Fire, A. (2011). Wobble base-pairing slows in vivo translation elongation in metazoans. RNA *17*, 2063–2073.

Stadler, M., and Fire, A. (2013). Conserved Translatome Remodeling in Nematode Species Executing a Shared Developmental Transition. PLoS Genet *9*, e1003739.

Stumpf, C.R., Moreno, M.V., Olshen, A.B., Taylor, B.S., and Ruggero, D. (2013). Short Article. Molecular Cell *52*, 574–582.

Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H., and Bartel, D.P. (2014). Poly(A)-tail profiling reveals anembryonic switch in translational control. Nature *508*, 66-71

Vaidyanathan, P.P., Zinshteyn, B., Thompson, M.K., and Gilbert, W.V. (2014). Protein kinase A regulates gene-specific translational adaptation in differentiating yeast. RNA *20*, 912–922.

Chapter 1: Coding Sequence Determinants of Translational Efficiency

Differential translational efficiencies (TEs) of mRNAs are determined by the nucleotide sequences of the mRNAs themselves, and the interaction of these sequences with the translational machinery and regulatory proteins. All mRNAs have a 5' transcript leader (TL) and 3' untranslated region (UTR), which are often bound by regulatory factors that can affect the translatability or stability of the mRNA. In contrast, the primary role of the coding sequence is to determine the amino acid sequence of the encoded protein, but the exact nucleotides used to encode that sequence have major effects on the amount of functional protein produced. In this chapter, I will focus on how events during translation elongation influence the TE of an mRNA.

The Mechanism of Translation Elongation in Eukaryotes

Compared to eukaryotic translation initiation (see Chapter 3), elongation is a fairly simple process, requiring only two universally conserved elongation factors: eEF1A and eEF2. The elongation phase (Figure 1.1) begins once a ribosome has been assembled on the start codon at the end of translation initiation. During elongation, the ribosome must select the transfer RNA (tRNA) that correctly matches the codon in its acceptor (A) site while avoiding incorporation of non-cognate and near-cognate tRNAs, which are present in great excess. The free charged tRNAs are in complex with eEF1A, and recognition of the proper tRNA causes ribosome-stimulated hydrolysis of the

eEF1A-associated GTP, releasing eEF1A and allowing the tRNA be fully accommodated in the A site. The ribosome then catalyzes the transfer of the growing polypeptide chain to the A-site tRNA, and performs a 3-nucleotide translocation along the mRNA. Translocation is catalyzed by eEF2 binding and GTP hydrolysis and places the next codon into the A-site. The cycle is repeated until a stop codon is encountered, leading to peptide release and ribosome recycling (Dever and Green, 2012).

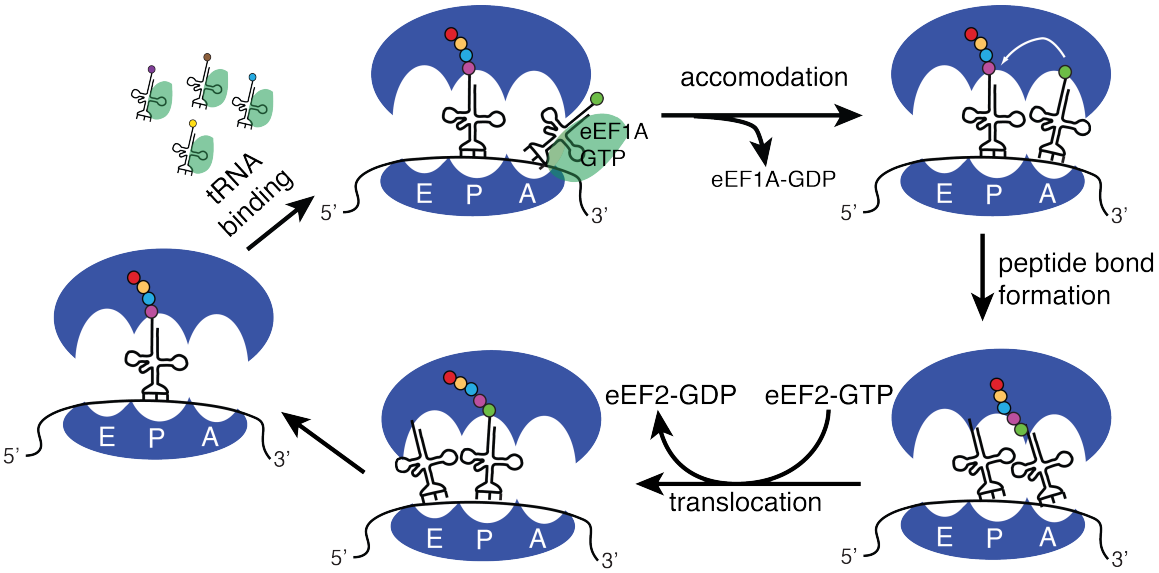

Figure 1.1: Mechanism of eukaryotic translation elongation.
This reaction scheme highlights the role of tRNA in the translation elongation cycle. Additional detail can be found in (Dever and Green, 2012; Kapp and Lorsch, 2004).

The Role of Translation Elongation in Determining TE

The effect on TE of inhibiting translation elongation depends on the relative rates of initiation and elongation for an mRNA. Low doses of elongation inhibitors do not affect overall synthesis of rabbit β-globin (Lodish, 1971) or reovirus proteins in infected human cells (Walden et al., 1981), indicating that elongation is not rate-limiting for these messages. Genome-wide experiments in yeast (Arava et al., 2003) and mouse embryonic stem cells (Ingolia et al., 2011) predict that ribosomes are so sparsely

packed on the majority of mRNAs that any single elongation cycle would need to be many fold slower in order to make elongation rate-limiting (Figure 1.2A). The exception to this rule are messages on which elongation occurs slowly at the 5' end due to rare codon usage (Chu et al., 2013) (see below), stable RNA structure (Doma and Parker, 2006), depletion or inhibition of elongation factors (Carlberg et al., 1990; Gutierrez et al., 2013), or stalling induced by peptide sequences, chaperones and regulatory proteins (Ingolia et al., 2011; Liu et al., 2013; Shalgi et al., 2012; Woolstenhulme et al., 2013). Slow elongation specifically at the start of an open reading frame can prevent the ribosome from vacating the start codon, effectively blocking initiation (Chu et al., 2013) (Figure 1.2B). Inhibition of initiation could also occur if an extreme ribosome pause later in the message was sufficiently long to cause queuing of ribosomes back to the start codon (Figure 1.2C).

Due to the wide range of initiation and elongation rates, different messages are likely to be initiation-limited to different extents and thus show different sensitivities to elongation inhibitors. Specifically, a message that is initiated highly efficiently would need a smaller decrease in elongation rate to reduce protein output than an inefficiently-initiated message. Consistent with this view, global inhibition of translation elongation preferentially inhibits production of specific proteins (Walden and Thach, 1986; Walden et al., 1981).

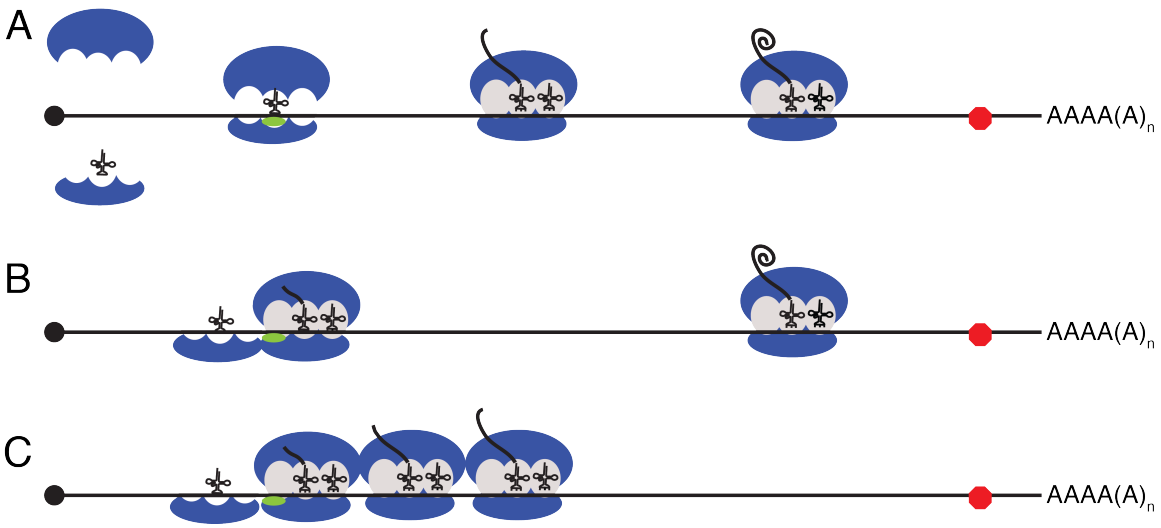

Figure 1.2: Models for rate-limiting translation elongation
(A) Initiation-limited translation leads to sparsely spaced ribosomes. This is thought to be the case for most yeast messages in rich media (Arava et al., 2003). Start and stop codons are indicated by green ovals and red octagons, respectively.
(B) Ribosomes are slow to leave the start codon, preventing new initiation events.
(C) An extremely slow elongation event in the middle of the ORF causes ribosomes to queue all the way to the start codon, preventing new initiation events.

Global Regulation Of Elongation Rate By eEF2 Phosphorylation

The best-characterized mechanism for regulating elongation is by phosphorylation of eEF2, which prevents this GTPase from binding ribosomes and thereby reduces the global rate of elongation in response to changes in pH or calcium in mammalian cells (Dorovkov et al., 2002; Ryazanov et al., 1988). This rate reduction is sufficient to render elongation rate-limiting and reduce overall protein output. eEF2 phosphorylation is thought to affect elongation of all mRNAs equally. In contrast, the choice of codons used to encode a particular amino acid sequence can have gene-specific effects on elongation rate.

The Role of tRNA Abundance in Modulating TE

Abundant and well-translated genes show a profound preference for use of specific codons when there is a choice of multiple codons for the same amino acid, a phenomenon referred to as codon bias. These codon preferences are often strikingly correlated with the abundance of the cognate tRNA (Gingold et al., 2014; Ikemura, 1981a; Tuller et al., 2010). Moreover, codon bias can dramatically affect protein production in certain contexts. Yields of recombinant protein are often greatly increased by optimization of the codon composition of the mRNA to match the codon bias of the host organism (Gustafsson et al., 2004; Plotkin and Kudla, 2010), suggesting that that elongation is rate-limiting for translation of these messages. Studies of frame-shifting efficiency, which is inversely proportional to elongation speed, showed that rare codons are on average decoded more slowly than preferred codons (Curran, 1989), raising the possibility that elongation is rate-limiting on some genes due to slow elongation at rare codons. Direct measurement of protein synthesis rates by incorporation of radiolabelled amino acids showed that genes with rare codons are elongated more slowly when transcribed at unnaturally high levels in *E. coli* (similar to methods used for recombinant protein production); however, these differences between codons disappeared at more physiological levels of transcription (Pedersen, 1984). Thus, the potential impact of codon bias on elongation rates and protein yields from ordinary cellular mRNAs is unclear.

An alternative explanation for codon bias is that highly expressed messages have co-evolved with the tRNA repertoire of the cell to prevent depletion of the cellular

tRNA pool (Ikemura, 1981b). Detailed analysis of reporter gene expression (Kudla et al., 2009; Qian et al., 2012) has shown that the primary effect of overexpression of an mRNA with rare codons is depletion of the ribosome or tRNA pools, which reduces protein output from all genes sharing the rare codons, and decreases the overall fitness of the host organism. This fitness effect is even detectable when preferred codons are overused in highly expressed mRNAs (Qian et al., 2012). These results suggest that codon usage and tRNA abundance has co-evolved with mRNA levels to prevent depletion of shared components of the translation machinery, and not primarily for fast translation of specific codons.

Different sub-steps of the elongation cycle may be rate-limiting for different codons, thereby causing codon-specific differences in the impact of mechanisms that regulate overall elongation rate. Earlier *in vitro* work found the tRNA binding and accommodation steps to be rate-limiting (Ledoux and Uhlenbeck, 2008; Pape et al., 1998), but recent studies with improved reaction conditions (Johansson et al., 2011) suggest that the chemistry of peptide bond formation may be rate-limiting for a subset of codons. Modern ribosome footprint profiling methods, which allow the detection of ribosome positions along mRNAs, have so far detected only a modest negative correlation between tRNA abundance and translation time of the cognate codon on cellular messages (Gardin et al., 2014), an effect far less than the several order of magnitude range predicted from models of codon adaptation (Chu et al., 2013; Tuller et al., 2010). It is possible that a technical limitation of the footprint profiling technique prevents detection of larger effects of tRNA abundance, but other studies have

successfully detected robust ribosome pausing in conditions of amino acid starvation (Gerashchenko and Gladyshev, 2014; Guydosh and Green, 2014; Li et al., 2012) and severe tRNA defects (Ishimura et al., 2014). In the latter case, ribosome accumulations were observed only in the absence of a quality control factor, suggesting that paused ribosomes are cleared *in vivo,* potentially preventing their detection in most experiments (Shoemaker and Green, 2012). In any case, it appears likely that factors other than cognate tRNA abundance determine the translation rate of specific codons. One of these factors is likely to be post-transcriptional modification of the cognate tRNAs for each codon.

The Role of tRNA Modifications in Determining Translation Elongation Rate

tRNAs play a central role in determining the speed and accuracy of translation elongation. Each tRNA must be recognized by only one of twenty or more aminoacyl-tRNA synthetases to ensure attachment of the correct amino acid. It must next base pair with one or more cognate codons, while rejecting a great excess of near-cognate and non-cognate codons (Agris et al., 2007; Crick, 1966). All tRNAs have a characteristic L-shaped structure (Ladner et al., 1975; Quigley and Rich, 1976) (Figure 1.3B) that must fit into the ribosome to catalyze peptidyl transfer. In order to enable specific and efficient decoding within the constraints of this common structure, tRNAs are extensively post-transcriptionally modified (Figure 1.3) – more than one hundred modified tRNA nucleotides are currently known, and the average yeast tRNA contains about a dozen distinct modifications (Agris et al., 2007; Phizicky and Hopper, 2010). Some, like the

thymidine and pseudouridine of the TψC loop, are found in every tRNA. Others, like the mcm^5s^2U nucleotide (Figure 1.3A, Chapter 2), are specific to a subset of tRNA species. The greatest variety of modified nucleotides is found in the anticodon loop and stem (Phizicky and Hopper, 2010), which are required for specific base-pairing interactions with the mRNA and for recognition by some tRNA synthetases (Rould et al., 1989; Ruff et al., 1991).

Extensive biochemical and genetic characterization has shown that tRNA modifications play crucial roles in tRNA charging (Madore et al., 1999; Sen and Ghosh, 1976), tRNA stability (Alexandrov et al., 2006), ribosome binding (Ashraf et al., 1999; Rezgui et al., 2013), decoding speed (Krüger et al., 1998), and fidelity (Johansson et al., 2008; Yarian et al., 2002). However, despite extensive evolutionary conservation, the majority of tRNA modifications are not required for yeast growth under laboratory conditions (Phizicky and Hopper, 2010), indicating that the effects of most individual tRNA modifications are likely to be modest *in vivo,* or only manifest under specific growth conditions. Indeed, numerous tRNA modifications are regulated in response to cellular stimuli, and may regulate translation of specific genes in response to stress (Chan et al., 2010; Patil et al., 2012; Söll and RajBhandary, 1995).

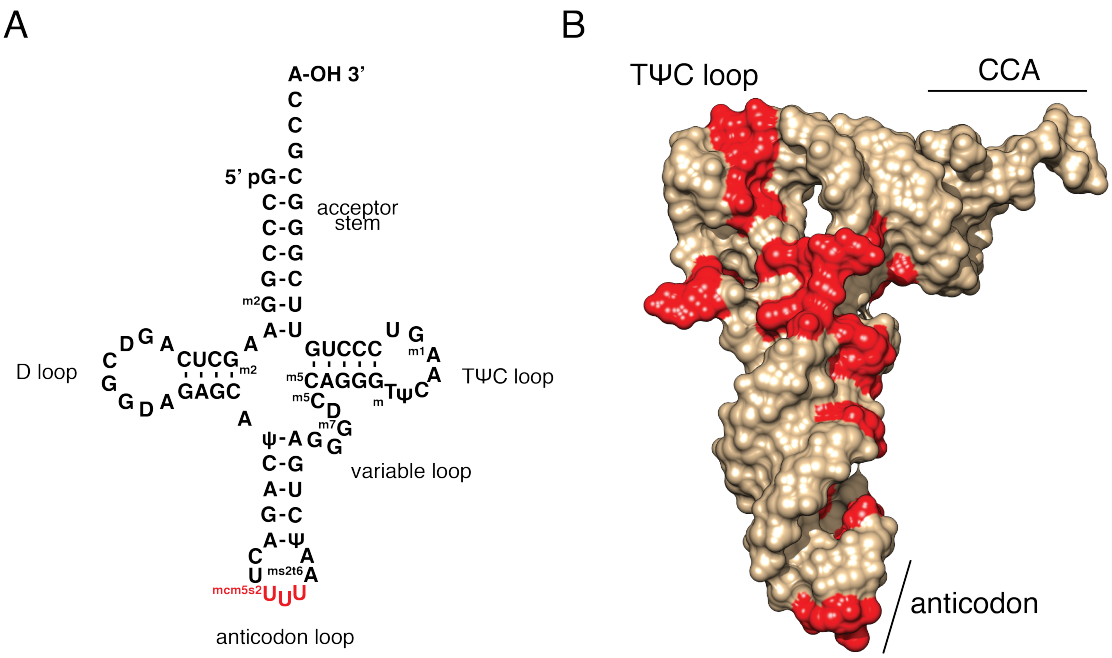

Figure 1.3: All tRNAs maintain a common structure, but are differentiated by multiple post-transcriptional modifications.
(A) Primary and secondary structure of mammalian tRNA lys. This sequence is unique to this tRNA, but the secondary structure is universal. Common elements are labeled, and the anticodon nucleotides are colored red. Note that the anticodon contains the mcm^{5}s^{2}U nucleotide, which is studied in Chapter 2.
(B) Crystal structure of the same tRNA in (A), showing the distinctive bent structure. Modified nucleotides are colored red. Sequence and structure data are from PDB ID 1FIR (Bénas et al., 2000). Structure visualization generated with UCSF Chimera (Pettersen et al., 2004).

Previous attempts to determine the *in vivo* effect of tRNA modifications on translation elongation rate have used artificial reporters with high mRNA expression or long stretches of cognate codons (Begley et al., 2007; Krüger et al., 1998), as it was not previously possible to measure elongation rates on individual native mRNAs. In Chapter 2, I use modern ribosome profiling methods to investigate the role in translation elongation of a conserved tRNA modification, mcm^{5}s^{2}U, which is required for survival of

budding yeast in numerous stress conditions. I find that loss of this modification causes slowed elongation at specific codons in native yeast mRNAs, as well as activation of stress response signaling.

References

Agris, P.F., Vendeix, F.A.P., and Graham, W.D. (2007). tRNA's wobble decoding of the genome: 40 years of modification. Journal of Molecular Biology *366*, 1–13.

Alexandrov, A., Chernyakov, I., Gu, W., Hiley, S.L., Hughes, T.R., Grayhack, E.J., and Phizicky, E.M. (2006). Rapid tRNA decay can result from lack of nonessential modifications. Molecular Cell *21*, 87–96.

Arava, Y., Wang, Y., Storey, J.D., Liu, C.L., Brown, P.O., and Herschlag, D. (2003). Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U.S.A. *100*, 3889–3894.

Ashraf, S.S., Sochacka, E., Cain, R., Guenther, R., Malkiewicz, A., and Agris, P.F. (1999). Single atom modification (O-->S) of tRNA confers ribosome binding. RNA *5*, 188–194.

Begley, U., Dyavaiah, M., Patil, A., Rooney, J.P., DiRenzo, D., Young, C.M., Conklin, D.S., Zitomer, R.S., and Begley, T.J. (2007). Trm9-catalyzed tRNA modifications link translation to the DNA damage response. Molecular Cell *28*, 860–870.

Bénas, P., Bec, G., Keith, G., Marquet, R., Ehresmann, C., Ehresmann, B., and Dumas, P. (2000). The crystal structure of HIV reverse-transcription primer tRNA(Lys,3) shows a canonical anticodon loop. RNA *6*, 1347–1355.

Carlberg, U., Nilsson, A., and Nygård, O. (1990). Functional properties of phosphorylated elongation factor 2. Eur. J. Biochem. *191*, 639–645.

Chan, C.T.Y., Dyavaiah, M., DeMott, M.S., Taghizadeh, K., Dedon, P.C., and Begley, T.J. (2010). A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. PLoS Genet *6*, e1001247.

Chu, D., Kazana, E., Bellanger, N., Singh, T., Tuite, M.F., and Haar, von der, T. (2013). Translation elongation can control translation initiation on eukaryotic mRNAs. Embo J. *33*, 21–34.

Crick, F.H. (1966). Codon--anticodon pairing: the wobble hypothesis. Journal of Molecular Biology *19*, 548–555.

Curran, J.F., and Yarus, M. (1989). Rates of aminoacyl-tRNA selection at 29 sense codons in vivo. Journal of Molecular Biology *209*, 65–77.

Dever, T.E., and Green, R. (2012). The elongation, termination, and recycling phases of translation in eukaryotes. Cold Spring Harbor Perspectives in Biology *4*, a013706.

Doma, M.K., and Parker, R. (2006). Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. Nature *440*, 561–564.

Dorovkov, M.V., Pavur, K.S., Petrov, A.N., and Ryazanov, A.G. (2002). Regulation of Elongation Factor-2 Kinase by pH †. Biochemistry *41*, 13444–13450.

Gardin, J., Yeasmin, R., Yurovsky, A., Cai, Y., Skiena, S., and Futcher, B. (2014). Measurement of average decoding rates of the 61 sense codons in vivo. Elife *3*.

Gerashchenko, M.V., and Gladyshev, V.N. (2014). Translation inhibitors cause abnormalities in ribosome profiling experiments. Nucleic Acids Research.

Gingold, H., Tehler, D., Christoffersen, N.R., Nielsen, M.M., Asmar, F., Kooistra, S.M., Christophersen, N.S., Christensen, L.L., Borre, M., Sørensen, K.D., et al. (2014). A Dual Programfor Translation Regulationin Cellular Proliferation and Differentiation. Cell *158*, 1281–1292.

Gustafsson, C., Govindarajan, S., and Minshull, J. (2004). Codon bias and heterologous protein expression. Trends Biotechnol. *22*, 346–353.

Gutierrez, E., Shin, B.-S., Woolstenhulme, C.J., Kim, J.-R., Saini, P., Buskirk, A.R., and Dever, T.E. (2013). eIF5A promotes translation of polyproline motifs. Molecular Cell *51*, 35–45.

Guydosh, N.R., and Green, R. (2014). Dom34 Rescues Ribosomes in 3' Untranslated Regions. Cell *156*, 950–962.

Ikemura, T. (1981a). Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system. Journal of Molecular Biology *151*, 389–409.

Ikemura, T. (1981b). Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes. Journal of Molecular Biology *146*, 1–21.

Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell *147*, 789–802.

Ishimura, R., Nagy, G., Dotu, I., Zhou, H., Yang, X.L., Schimmel, P., Senju, S.,

Nishimura, Y., Chuang, J.H., and Ackerman, S.L. (2014). Ribosome stalling induced by mutation of a CNS-specific tRNA causes neurodegeneration. Science *345*, 455–459.

Johansson, M., Ieong, K.-W., Trobro, S., Strazewski, P., Åqvist, J., Pavlov, M.Y., and Ehrenberg, M. (2011). pH-sensitivity of the ribosomal peptidyl transfer reaction dependent on the identity of the A-site aminoacyl-tRNA. Proceedings of the National Academy of Sciences *108*, 79–84.

Johansson, M.J.O., Esberg, A., Huang, B., Björk, G.R., and Byström, A.S. (2008). Eukaryotic wobble uridine modifications promote a functionally redundant decoding system. Mol Cell Biol *28*, 3301–3312.

Kapp, L.D., and Lorsch, J.R. (2004). The molecular mechanics of eukaryotic translation. Annu. Rev. Biochem. *73*, 657–704.

Krüger, M.K., Pedersen, S., Hagervall, T.G., and Sørensen, M.A. (1998). The modification of the wobble base of tRNAGlu modulates the translation rate of glutamic acid codons in vivo. Journal of Molecular Biology *284*, 621–631.

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.

Ladner, J.E., Jack, A., Robertus, J.D., Brown, R.S., Rhodes, D., Clark, B.F., and Klug, A. (1975). Structure of yeast phenylalanine transfer RNA at 2.5 A resolution. Proc. Natl. Acad. Sci. U.S.A. *72*, 4414–4418.

Ledoux, S., and Uhlenbeck, O.C. (2008). Different aa-tRNAs Are Selected Uniformly on the Ribosome. Molecular Cell *31*, 114–123.

Li, G.-W., Oh, E., and Weissman, J.S. (2012). The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature *484*, 538–541.

Liu, B., Han, Y., and Qian, S.-B. (2013). Cotranslational response to proteotoxic stress by elongation pausing of ribosomes. Molecular Cell *49*, 453–463.

Lodish, H.F. (1971). Alpha and beta globin messenger ribonucleic acid. Different amounts and rates of initiation of translation. J. Biol. Chem. *246*, 7131–7138.

Madore, E., Florentz, C., Giegé, R., Sekine, S., Yokoyama, S., and Lapointe, J. (1999). Effect of modified nucleotides on Escherichia coli tRNAGlu structure and on its aminoacylation by glutamyl-tRNA synthetase. Predominant and distinct roles of the mnm5 and s2 modifications of U34. Eur. J. Biochem. *266*, 1128–1135.

Pape, T., Wintermeyer, W., and Rodnina, M.V. (1998). Complete kinetic mechanism of elongation factor Tu-dependent binding of aminoacyl-tRNA to the A site of the E. coli ribosome. EMBO J. *17*, 7490–7497.

Patil, A., Dyavaiah, M., Joseph, F., Rooney, J.P., Chan, C.T.Y., Dedon, P.C., and Begley, T.J. (2012). Increased tRNA modification and gene-specific codon usage regulate cell cycle progression during the DNA damage response. Cell Cycle *11*, 3656–3665.

Pedersen, S. (1984). Escherichia coli ribosomes translate in vivo with variable rate. Embo J. *3*, 2895–2898.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem *25*, 1605–1612.

Phizicky, E.M., and Hopper, A.K. (2010). tRNA biology charges to the front. Genes & Development *24*, 1832–1860.

Plotkin, J.B., and Kudla, G. (2010). Synonymous but not the same: the causes and consequences of codon bias. Nature Publishing Group *12*, 32–42.

Qian, W., Yang, J.-R., Pearson, N.M., Maclean, C., and Zhang, J. (2012). Balanced Codon Usage Optimizes Eukaryotic Translational Efficiency. PLoS Genet *8*, e1002603.

Quigley, G.J., and Rich, A. (1976). Structural domains of transfer RNA molecules. Science *194*, 796–806.

Rezgui, V.A.N., Tyagi, K., Ranjan, N., Konevega, A.L., Mittelstaet, J., Rodnina, M.V., Peter, M., and Pedrioli, P.G.A. (2013). tRNA tKUUU, tQUUG, and tEUUC wobble position modifications fine-tune protein translation by promoting ribosome A-site binding. Proceedings of the National Academy of Sciences *110*, 12289–12294.

Rould, M.A., Perona, J.J., Söll, D., and Steitz, T.A. (1989). Structure of E. coli glutaminyl-tRNA synthetase complexed with tRNA(Gln) and ATP at 2.8 A resolution. Science *246*, 1135–1142.

Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J.C., and Moras, D. (1991). Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp). Science *252*, 1682–1689.

Ryazanov, A.G., Shestakova, E.A., and Natapov, P.G. (1988). Phosphorylation of elongation factor 2 by EF-2 kinase affects rate of translation. Nature *334*, 170–173.

Sen, G.C., and Ghosh, H.P. (1976). Role of modified nucleosides in tRNA: effect of modification of the 2-thiouridine derivative located at the 5'-end of the anticodon of yeast transfer RNA Lys2. Nucleic Acids Research *3*, 523–535.

Shalgi, R., Hurt, J.A., Krykbaeva, I., Taipale, M., Lindquist, S., and Burge, C.B. (2013).

Widespread regulation of translation by elongation pausing in heat shock. Molecular Cell *49*, 439–452.

Shoemaker, C.J., and Green, R. (2012). Translation drives mRNA quality control. Nature Structural & Molecular Biology *19*, 594–601.

Söll, D., and RajBhandary, U. (1995). tRNA (Amer Society for Microbiology).

Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., and Pilpel, Y. (2010). An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell *141*, 344–354.

Walden, W.E., and Thach, R.E. (1986). Translational control of gene expression in a normal fibroblast. Characterization of a subclass of mRNAs with unusual kinetic properties. Biochemistry *25*, 2033–2041.

Walden, W.E., Godefroy-Colburn, T., and Thach, R.E. (1981). The role of mRNA competition in regulating translation. I. Demonstration of competition in vivo. J. Biol. Chem. *256*, 11739–11746.

Woolstenhulme, C.J., Parajuli, S., Healey, D.W., Valverde, D.P., Petersen, E.N., Starosta, A.L., Guydosh, N.R., Johnson, W.E., Wilson, D.N., and Buskirk, A.R. (2013). Nascent peptides that block protein synthesis in bacteria. Proceedings of the National Academy of Sciences *110*, E878–E887.

Yarian, C., Townsend, H., Czestkowski, W., Sochacka, E., Malkiewicz, A.J., Guenther, R., Miskiewicz, A., and Agris, P.F. (2002). Accurate translation of the genetic code depends on tRNA modified nucleosides. J. Biol. Chem. *277*, 16391–16395.

Chapter 2: Loss of a Conserved tRNA Anticodon Modification Perturbs Cellular Signaling*

Abstract

Transfer RNA (tRNA) modifications enhance the efficiency, specificity and fidelity of translation in all organisms. The anticodon modification $mcm^5s^2U^{34}$ is required for normal growth and stress resistance in yeast; mutants lacking this modification have numerous phenotypes. Mutations in the homologous human genes are linked to neurological disease. The yeast phenotypes can be ameliorated by overexpression of specific tRNAs, suggesting that the modifications are necessary for efficient translation of specific codons. We determined the *in vivo* ribosome distributions at single codon resolution in yeast strains lacking mcm^5s^2U. We found accumulations at AAA, CAA, and GAA codons, suggesting that translation is slow when these codons are in the ribosomal A site, but these changes appeared too small to affect protein output. Instead, we observed activation of the *GCN4*-mediated stress response by a non-canonical pathway. Thus, loss of mcm^5s^2U causes global effects on gene expression due to perturbation of cellular signaling.

---

* This research was originally published in PLoS Genetics, and has been edited for presentation here. Zinshteyn B, Gilbert WV (2013) Loss of a Conserved tRNA Anticodon Modification Perturbs Cellular Signaling. PLoS Genet 9(8): e1003675.

Introduction

Transfer RNAs (tRNAs) from all domains of life contain numerous post-transcriptional modifications, many of which are highly conserved. These modifications enhance the efficiency, specificity and fidelity of translation(Agris et al., 2007; Johansson and Byström, 2005; Phizicky and Hopper, 2010). In the budding yeast *Saccharomyces cerevisiae,* three tRNAs ($tRNA^{Lys}_{UUU}, tRNA^{Gln}_{UUG}, tRNA^{Glu}_{UUC}$) are modified by addition of 5-methoxycarbonylmethyl (mcm^5) and 2-thio (s^2) groups to uridine at the 5' nucleotide of the tRNA anticodon (U34), resulting in an mcm^5s^2U nucleotide. The mcm^5s^2U modification (MSUM) and many of the responsible modifying enzymes are conserved across eukaryotes, having been identified in fungi (Huang et al., 2005; Leidel et al., 2009), plants(Mehlgarten et al., 2010), worms (Chen et al., 2009) and mammals (Chan et al., 1982). Despite widespread conservation, and extensive biochemical characterization, the physiological role of MSUM is unknown.

Genes required for MSUM are unusual among tRNA modification genes in the number and severity of their mutant phenotypes. Most yeast strains lacking tRNA modifications are viable and show no growth impairment (Johansson and Byström, 2005; Phizicky and Hopper, 2010), but *S. cerevisiae* and *C. elegans* double mutants lacking both mcm^5 and s^2 are not viable (Björk et al., 2007; Chen et al., 2009). In yeast, single mutants lacking either mcm^5 or s^2 have numerous phenotypes including temperature sensitivity, various chemical stress sensitivities, exocytosis defects, and transcriptional defects (Esberg et al., 2006; Krogan and Greenblatt, 2001). In *C. elegans,* mutants of the Elongator complex (comprised of *elp1* through *elp6*), which is

required to produce the mcm^5 modification, display neurological defects (Chen et al., 2009). In humans, mutations in IBKAP, the *elp1* homolog, cause familial dysautonomia (FD)(Slaugenhaupt et al., 2001), and mutations in *elp4* are associated with Rolandic epilepsy (Strug et al., 2009).

The molecular connection between these cellular/organismal phenotypes and the lack of specific tRNA anticodon modifications is currently unknown. Loss of either mcm^5 or s^2 impairs reading of both Watson-Crick (VAA) and wobble (VAG) cognate codons by the modified tRNAs (Johansson et al., 2008; Krüger et al., 1998), and chemical removal or modification of the s^2 moiety leads to a reduction in the rate of tRNA charging *in vitro* (Sen and Ghosh, 1976; Seno et al., 1974). The MSUM phenotypes were originally attributed to a proposed role of the Elongator complex in transcriptional elongation (Otero et al., 1999) before its function in tRNA modification was discovered (Huang et al., 2005). However, the phenotypes of yeast MSUM mutants, including the lethality in mutants lacking both mcm^5 and s^2, can be suppressed by overexpression of unmodified versions of two tRNAs that normally contain mcm^{5}s^{2}U – $tRNA_{UUG}^{Gln}$ and $tRNA_{UUU}^{Lys}$ (Esberg et al., 2006). These observations indicate that at least a subset of the yeast cellular phenotypes are tied to tRNA function. It has been argued that loss of MSUM leads to codon-specific translation defects leading to insufficient protein production, either from many genes, or from a few genes required to carry out particular cellular processes or stress responses, but this hypothesis has not been directly tested.

In this study, we examined codon level ribosome distributions genome-wide using ribosome footprint profiling (Ribo-seq). We found that loss of mcm^5 or s^2 leads to

slow translation elongation specifically at codons that Watson-Crick pair with MSUM tRNAs, but the magnitude of these changes appeared insufficient to affect protein output. Surprisingly, all of the MSUM strains showed gene expression signatures consistent with activation of the Gcn4p-mediated stress response pathway. We demonstrate that disruption of this pathway suppresses the MSUM mutant phenotypes independently of tRNA concentration.

Results

Ribosome Footprint Profiling Reveals Features of Translation for Specific Codons

We set out to determine whether MSUM mutants display codon-specific translation defects. Translational activity genome-wide was determined using Ribo-seq, which consists of isolating and sequencing ribosome-protected mRNA fragments from RNase-treated whole-cell lysates (Ingolia et al., 2009). This method reveals ribosome positions at single nucleotide resolution, and thus has the potential to identify translational defects affecting single codons (Ingolia et al., 2009; Stadler and Fire, 2011). Wild type (WT) yeast, as well as strains lacking the s^2 moiety (*ncs2Δ, ncs6Δ,* and *uba4Δ*), or mcm^5 (*elp3Δ*) (Figure 2.1A), were profiled by Ribo-seq, as well as RNA-seq. To assess the impact of these modifications on translation, the ribosome dwell time at specific codons was determined as follows. The positions of the A, P and E site codons within ribosome footprints of various lengths (25-31 nt) were determined by examining the 5' ends of footprints mapping to start codons, where initiating ribosomes

are expected to contain start codons in their P sites (Figure 2.1B)(Kapp and Lorsch, 2004). Next, to determine the genome-wide average ribosome dwell time for a given codon (Figure 2.1C, left), all instances of that codon in the genome were aligned, and 5' ends of reads mapping to the surrounding positions (Figure 2.1C) were summed (see Materials and Methods). The resulting metacodon plots show the relative number of ribosome footprints, and thus the relative amount of time the ribosome spends at each position, as the codon moves through the A, P and E sites. Codon identity is not expected to affect translation from the outer sites (±1, ±2), so the entire plot was normalized to the height of these peaks. The height of each peak is the bulk occupancy for that codon in that ribosomal site, similar to a previously described metric (Stadler and Fire, 2011). The metacodon distributions for ATG and stop codons indicated that the reads were properly assigned to the ribosomal sites (Figure 2.1C, right). We observed distinct and reproducible patterns of ribosome density for different codons in WT yeast (Figure 2.1C, 2.2A,B), consistent with the single-nucleotide resolution of this technique.

The metacodon plots of WT yeast provided insights into the determinants of translation rate for specific codons. Notably, all four proline codons spent over 2-fold more time in the P site than the average codon, while glycine codons spent ~40-50% more time in the A site (Figure 2.2C). This effect was additive for Pro-Gly pairs in the P and A sites, but not if the codon order was reversed (Figure 2.2D), indicating that the effects of Pro and Gly were specific to the P and A sites, respectively. This proline effect is reminiscent of the proline/glycine pausing recently discovered in bacteria lacking

elongation factor P (Doerfel et al., 2013; Ude et al., 2013; Woolstenhulme et al., 2013). The observed effects were consistent with *in vitro* data which showed that peptidyl transfer can be rate limiting for A-site glycine and proline codon translation at physiological pH (Johansson et al., 2011), and that proline induces particularly slow peptide bond formation when it is at the carboxyl terminus of the growing peptide chain (Pavlov et al., 2009) (Figure 2.2E). These results suggest that peptidyl transfer is rate limiting for certain Pro and Gly codons in yeast cells as well.

Experiments in recombinant systems have led to the strong expectation that translation times for codons should be inversely proportional to the concentrations of their cognate tRNAs (Pedersen, 1984; Tuller et al., 2010). To investigate potential sources of the distinctive metacodon distributions we observed, we performed unsupervised hierarchical clustering on them (Figure 2.3A). This analysis clustered many codons together based on their encoded amino acid or the first two nucleotides of the codon. Notably, codons did not cluster by tRNA adaptation index (tAI), a proxy for cognate tRNA abundance (Tuller et al., 2010). More directly, the bulk occupancies did not show a negative correlation with tAI in the A site (Figure 2.3B). There was also no correlation of codon occupancy with tRNA abundance measurements, genomic copy number, or a more recent codon usage metric which accounts for tRNA competition (Pechmann and Frydman, 2012) (data not shown). These results demonstrate that translation rates for particular yeast codons are not determined by the cellular concentrations of their cognate tRNAs, consistent with findings from Ribo-seq

experiments in mice and bacteria (Ingolia et al., 2011; Li et al., 2012) and from protein synthesis reporters (containing codon repeats) in yeast (Letzring et al., 2010).

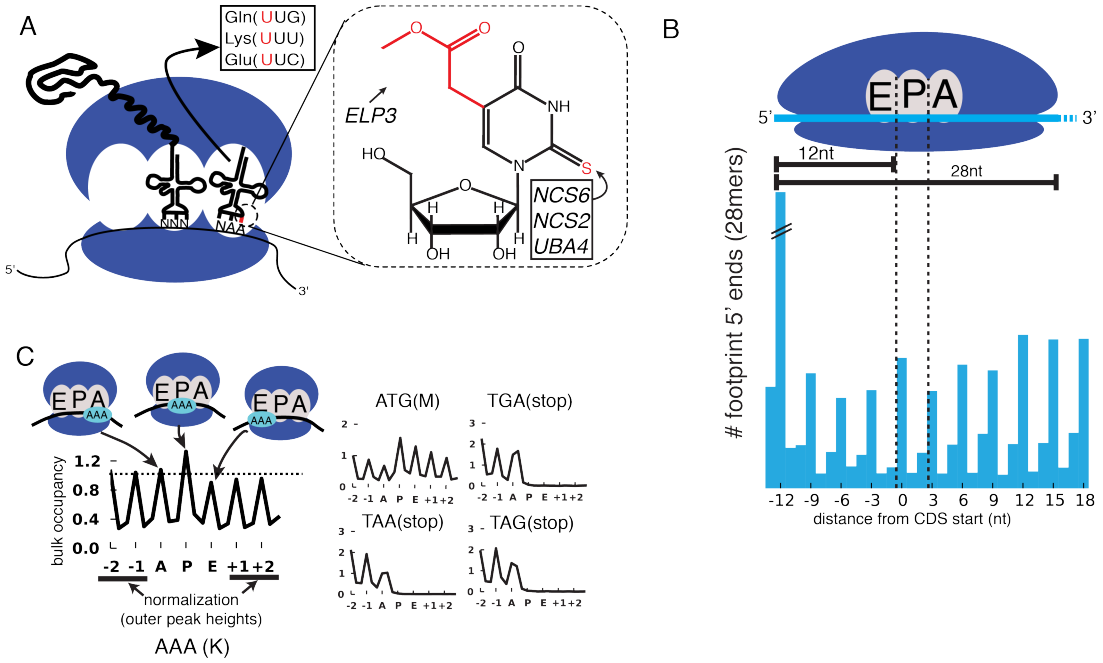

Figure 2.1: Method for bulk analysis of codon occupancy in MSUM strains.
(A) (left) mcm^{5}s^{2}U is found at the 5' nucleotide of the anticodon in three yeast tRNAs. (right) The structure of mcm^{5}s^{2}U, and the subset of modification genes whose mutants were profiled in this study are indicated.
(B) (top) Anatomy of a ribosome footprint, with P-site offset for 28mer reads indicated. (bottom) Metaplot of WT ribosome footprint reads summed across all start codons. The peak of upstream reads corresponds to ribosomes with start codons in their P site. The location of this peak is used to determine the location of A, P and E sites for each read length.
(C) (left) Explanation of metacodon plots. Similar to panel B, all in-frame instances of a given codon in the genome are aligned, and the reads mapping around those positions are summed. The resulting plot is then offset by the P-site distance, and normalized to the average peak height of the outer sites (±1, ±2). The peak heights for each site are the bulk codon occupancies, a proxy for the amount of time the ribosome spends with a given codon in each site, compared to its neighbors. (right) ATG codons and stop codons display the expected distributions with this metric. All plots are from WT yeast.

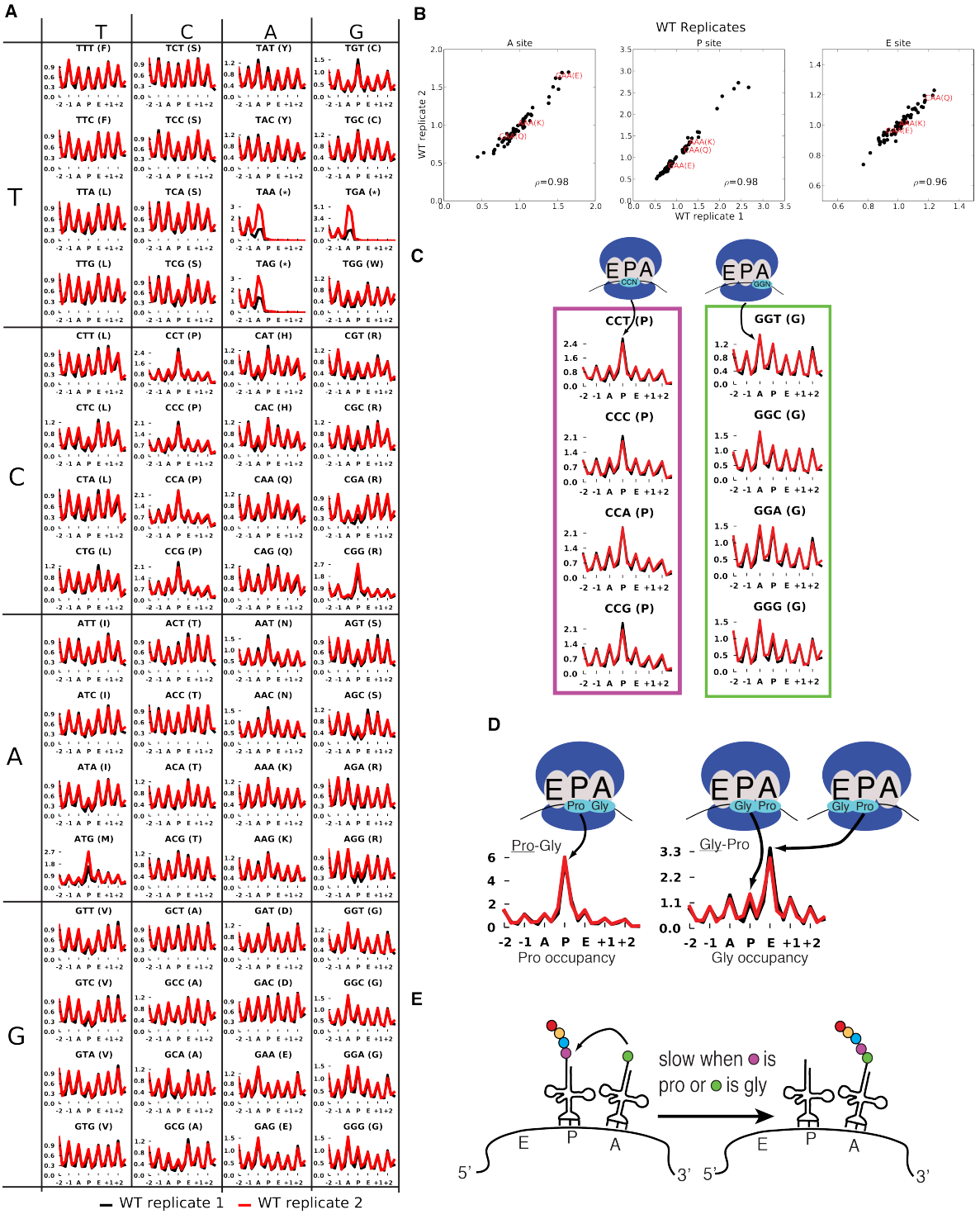

Figure 2.2: Metacodon plots provide information on translation kinetics at the codon level.

(A) Full set of metacodon plots, with superimposed WT replicates.

(B) Reproducibility of bulk codon occupancy metric. Spearman correlations are indicated.

(C) Details of metacodon plots for Gly and Pro.

(D) Metacodon plots for Pro-Gly and Gly-Pro pairs.
(E) Model for Pro and Gly metacodon plots. Peptidyl transfer is slow when Pro is in the P site, or Gly is in the A site, possibly making peptidyl-transfer rate-limiting for translocation, especially for Pro-Gly pairs.

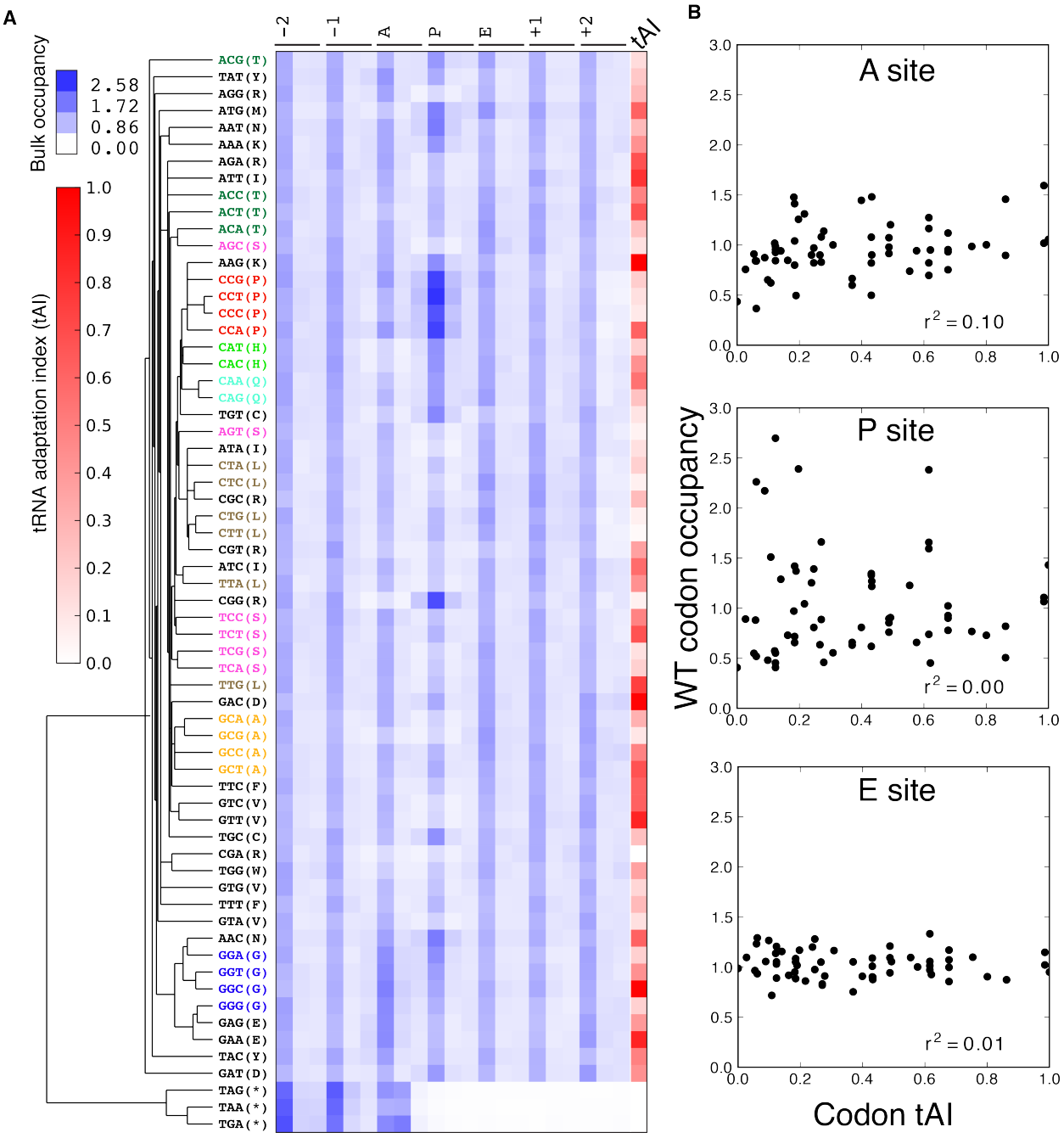

Figure 2.3: Codon occupancy is not determined by codon adaptation.
(A)Unsupervised hierarchical clustering of WT metacodon plots. Codons for the same amino acid that cluster together have been colored. The tRNA adaptation index (tAI) for each codon is indicated in red. The tAI is a proxy for cognate tRNA abundance.

(B) Correlations between WT codon occupancy and tAI for codons in each ribosome site.

Loss of MSUM Genes Reduces Translation Rate at AAA, CAA, GAA Codons

Having established the ability to detect differences in the translation of different codons, we next examined changes in codon-specific translation in the MSUM strains. Bulk occupancy for each codon in each ribosomal site (the height of the peaks in the metacodon plots) was determined for each mutant. All of the strains lacking the s^2 modification showed increases in ribosome density corresponding to CAA and AAA in the A site, while the *elp3Δ* strain showed an increase in the CAA and GAA codons (Figure 2.4A). The magnitude of the changes was largest when the affected codon was found in the ribosomal A-site. The magnitude and direction of change for the GAA codon was variable between mutants lacking the same modification, and even between biological replicates (Figure 2.4A), indicative of some underlying biological or technical noise in this measurement. Nonetheless, in all but one replicate, the largest increases in each mutant were for codons decoded by Watson-Crick pairing with MSUM tRNAs.

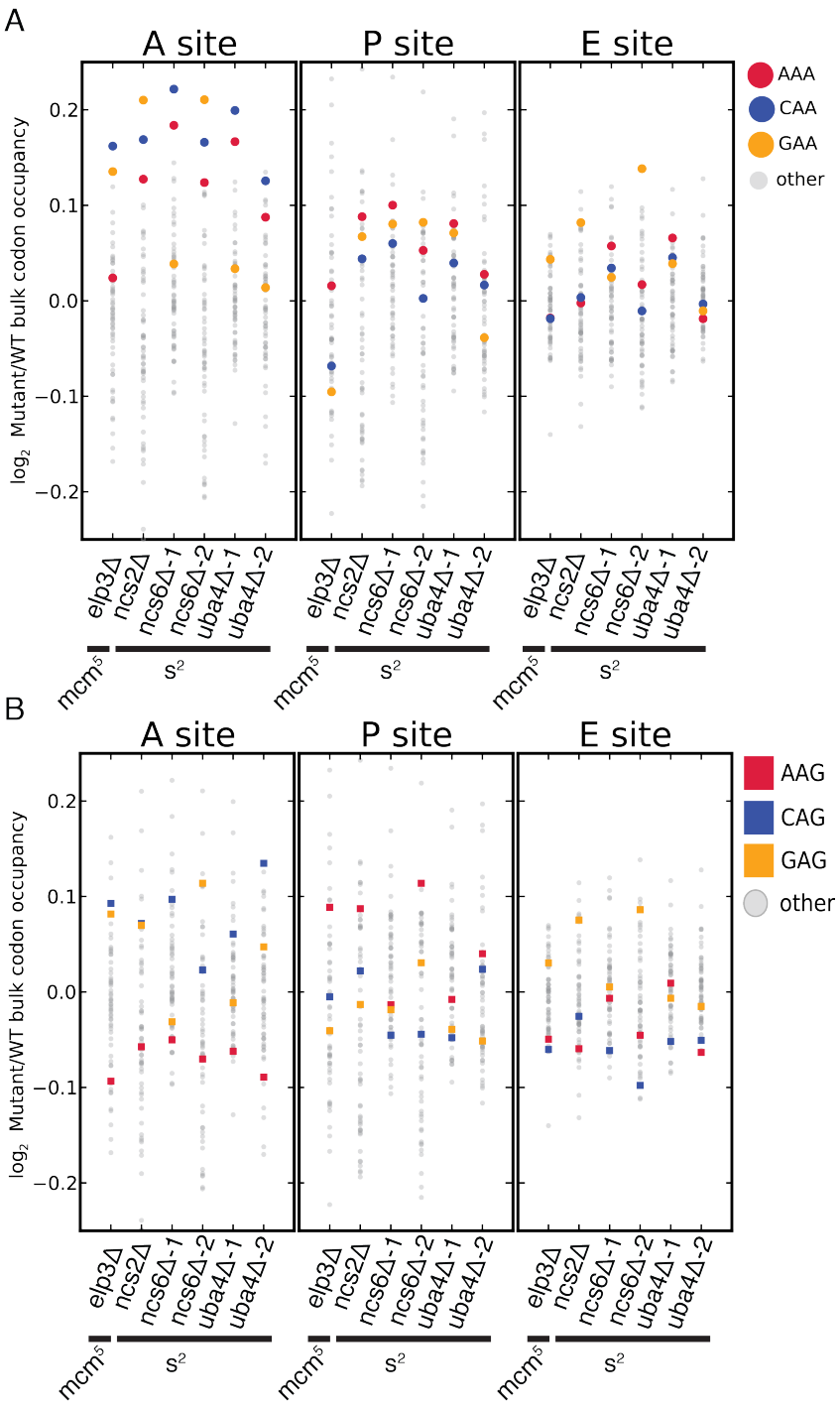

Figure 2.4: Genetic ablation of mcm^5 or s^2 leads to ribosome accumulation at specific codons.

(A and B) Changes in bulk codon occupancy in MSUM mutants. Both plots are the same, with different codons highlighted. Independent biological replicates were done for *ncs6Δ* and *uba4Δ*. All mutants are compared to a WT sample prepared and processed simultaneously.

mcm^{5}s^{2}U is Not Required for Wobble Decoding of AAG, CAG, and GAG Codons In Vivo

MSUM is necessary for wobble decoding of G-ending codons in strains that lack other cognate tRNAs(Johansson et al., 2008), but it is not clear whether the modified tRNAs contribute to decoding in the WT state where these other tRNAs are present. In our datasets AAG, CAG, and GAG codons showed smaller increases in bulk occupancy (and some net decreases) compared to their A-ending counterparts, suggesting that MSUM is mainly required for translation of VAA codons (Figure 2.4B). In order to assess the statistical significance of these changes, a metric for ribosome dwell time at individual codons was developed (Figure 2.5A). This metric normalizes the read counts at a particular codon by the mean read density of the open reading frame that contains it. The genome-wide distributions for all instances of each codon were compared between mutant and WT strains using the K-S test (Figure 2.5B, C). Due to the noise inherent in read sampling, many codons showed statistically significant changes. However, the VAA codons had p values many orders of magnitude smaller than all other codons, particularly in the *ncs6Δ* and *uba4Δ* datasets, which were from pooled biological replicates (Figure 2.5C). The pooled datasets provided data for approximately twice as many codons and may have averaged out biological and technical noise. Consistent with our analysis of bulk codon occupancy, the effect of MSUM loss was strongest in the A site for all 3 VAA codons. We did not see a corresponding statistical significance for the VAG codons (Figure 2.5C), indicating that mcm^{5}s^{2}U does not significantly contribute to the decoding of these codons *in vivo*. This

result does not contradict previous evidence that the modifications are required for translation of VAG codons by wobble pairing (Johansson et al., 2008), but indicates that $tRNAs_{UUB}$ contribute minimally to the translation of VAG codons *in vivo*, where $tRNAs_{CUB}$ with Watson-Crick complementarity are available.

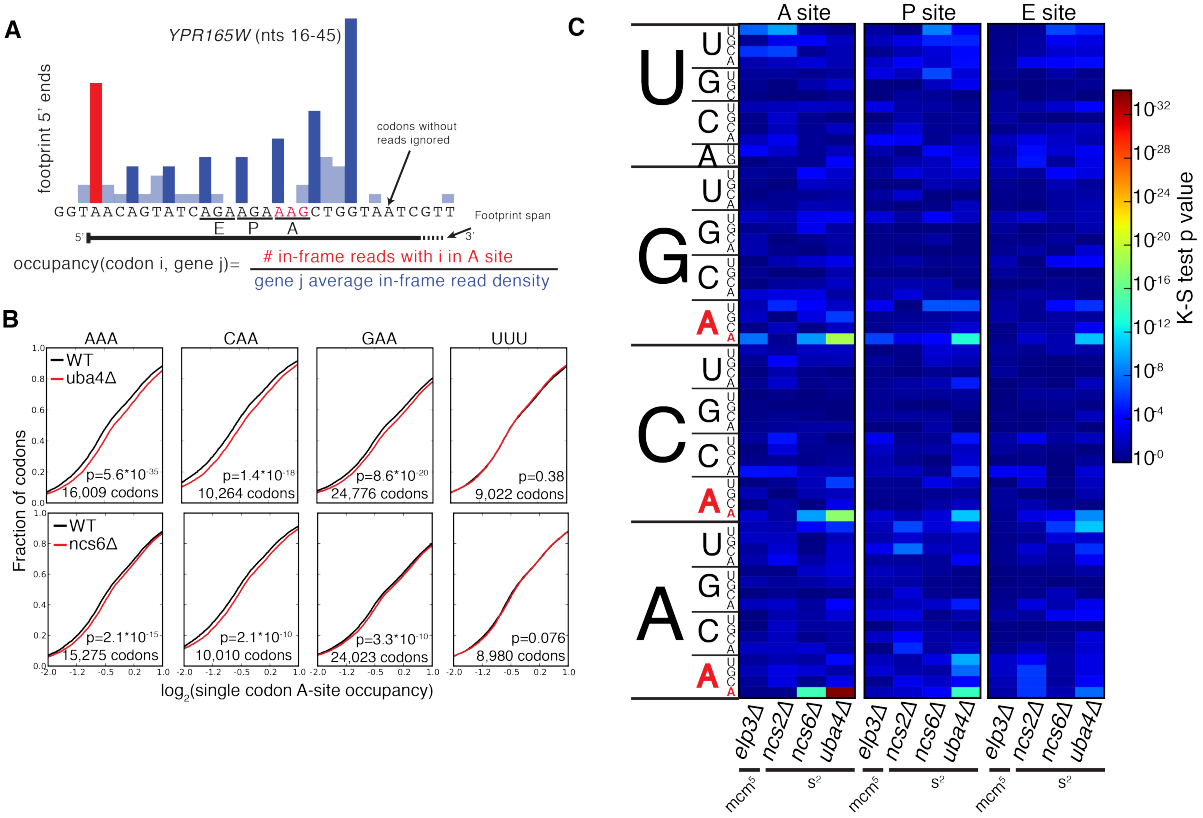

Figure 2.5: A single-codon occupancy metric shows that ribosome footprint accumulations at AAA, CAA, and GAA are statistically significant
(A) Description of the single codon occupancy metric. The occupancy for a given codon in a given site is the number of in-frame reads for that codon in that site, compared to the average in-frame read density for the parent gene.
(B) Cumulative distributions of single-codon occupancy for select codons in *ncs6Δ* and *uba4Δ*.
(C) Heatmap of K-S test p-values for all sense codons in all mutants. For *ncs6Δ* and *uba4Δ*, mutant and WT replicates were pooled to improve the accuracy of the metric.

The Elongation Defects in MSUM Strains Appear Insufficient to Affect Protein Levels

Despite the statistical significance of the increased ribosome dwell times at VAA codons in MSUM mutants, the magnitude of the changes does not seem to be large enough to generally affect protein output. Initiation, not elongation, is the rate-limiting step of eukaryotic translation in most circumstances (Lodish and Jacobsen, 1972; Walden et al., 1981), and the mean ribosome density is only 1 per 164 nts (Arava et al., 2003). Given this sparse spacing of ribosomes on yeast mRNAs, transcripts with mean ribosome density would require an elongation delay greater than the average translation time of 50 codons in order for an MSUM mutation to make elongation rate limiting. The most densely populated messages would require a 20-fold elongation delay. The average bulk increase observed for VAA codons was less than 17% (Figure 2.4D), and the largest confidently assigned (≥32 reads) single-codon change was less than 5-fold (Figure 2.6A,C). In the event of an elongation delay long enough to affect protein output, ribosome queuing should occur behind AAA and CAA codons with increased occupancy. However, no queuing was observed (Figure 2.6B, D). Codons with more read coverage display smaller changes than codons with low read coverage, indicating that the range of this metric is not being limited by sequencing depth (Figure 2.6A, 2.6C). We also did not observe increased ribosome density at stretches of 2 or more VAA codons (data not shown). These results were consistent with the polysome gradient profiles of the MSUM strains, which were indistinguishable from WT (data not shown), indicating that translation elongation in bulk was unaffected.

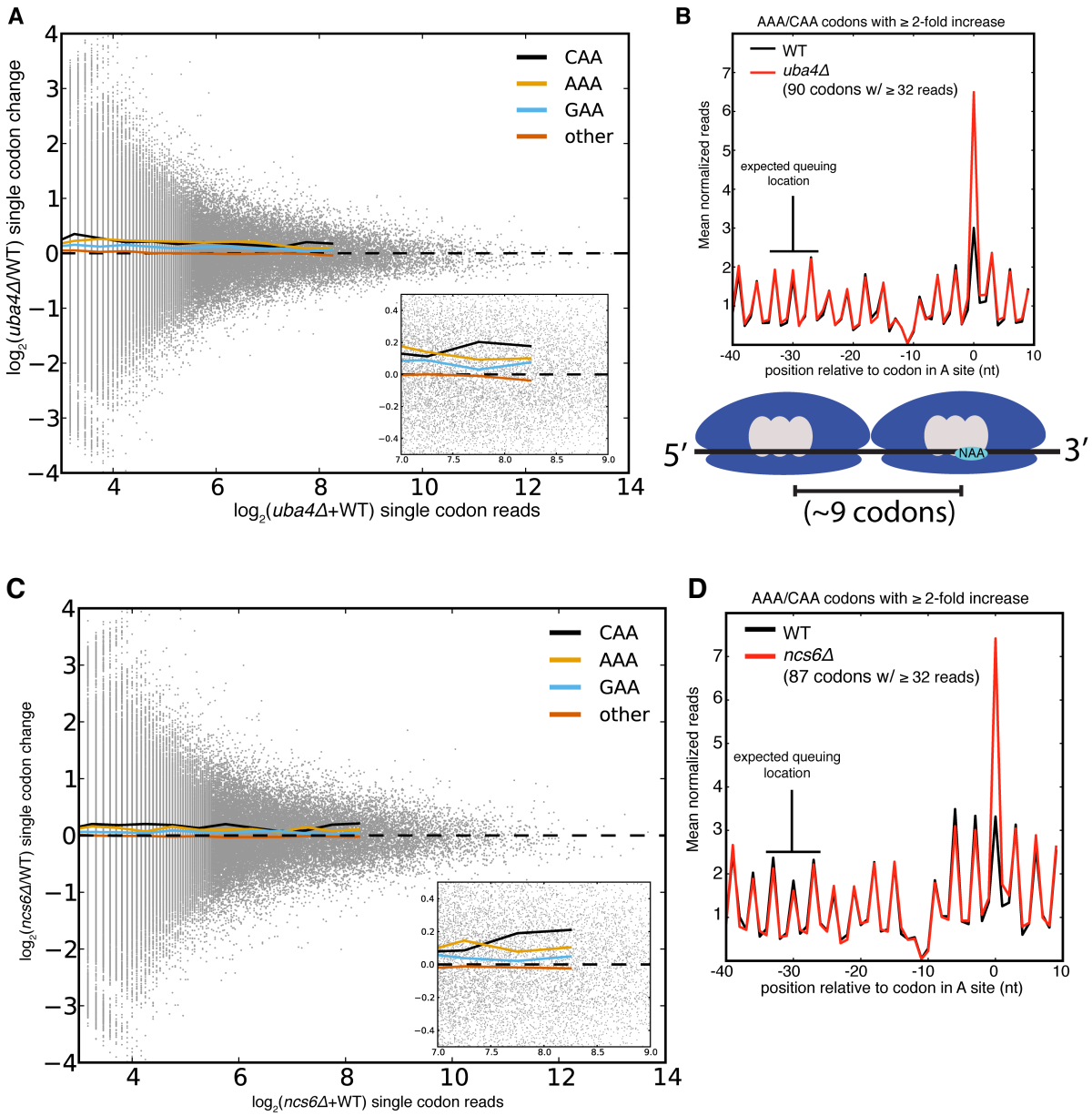

Figure 2.6: Single codon occupancy changes may be insufficient to affect protein output.

(A) Fold changes for all single codons in *uba4Δ* are plotted against their read density in grey. Colored lines are the mean fold changes for the specified codons over read-coverage bins of width 0.2 ($\log_2$ scaled). "Other" is a pool of all non-VAA codons.
(B) Metaplot of ribosome footprint density around all AAA and CAA codons with ≥2-fold change in *uba4Δ*, and ≥32 reads in both datasets. Reads at each position were normalized by the total number of reads for the parent gene, and averaged across all host genes that overlap that position. The plot is offset such that 0 corresponds to having the codon in the A site. The expected location of a ribosome queuing event is

indicated, and a diagram of such an event is shown below. The dip in ribosome footprint density at -10 is a computational artifact, due to an inability to determine read lengths of poly-adenylated fragments when they end in one or more adenosines.
(C, D)Same as in (A) and (B), but for *ncs6Δ.*

The GCN4-Mediated Stress Response Is Activated in MSUM Strains

In search of an alternative explanation for MSUM mutant phenotypes, we examined global ribosome footprint densities and transcript levels for perturbations in the MSUM mutant strains. Consistent with previous reports (Brar et al., 2012; Ingolia et al., 2009), gene expression values from Ribo-seq were highly reproducible (Figure 2.7A). Furthermore, all of the mutant strains showed similar RNA-seq and Ribo-seq changes when compared to WT strains (Figure 2.7B,C), indicating that these gene expression changes are likely to be downstream of a common defect. Replicate data for *ncs6Δ* and *uba4Δ* enabled us to assess the significance of particular changes using counting statistics (Robinson et al., 2009). This analysis identified a set of genes with significant changes in ribosome footprint densities, which were largely shared between *ncs6Δ* and *uba4Δ* (Figure 2.8A-C, 2.7B,C). The changes in ribosome footprint density were correlated with changes in transcript levels (pearson r=0.59 for *ncs6Δ*, 0.64 for *uba4Δ*), indicating that these gene expression changes were largely due to changes in the mRNA pool (Figure 2.8A,B). Intriguingly, a significant fraction (24/68) of the affected genes are known targets of the *GCN4* transcription factor (Natarajan et al., 2001) (Figure 2.8 A-C). To investigate the specificity of the observed induction of *GCN4* targets in MSUM mutants, we examined the behavior of *GCN4* targets in 1,924 yeast microarray studies using data from the SPELL curated yeast microarray compendium.

This compendium includes experiments sampling a broad range of environmental and genetic perturbations (Hibbs et al., 2007). We determined the significance of overlap between *GCN4* targets and the set of upregulated (≥2-fold) genes in each of these 1,924 datasets. Notably, the overlap between *GCN4* targets and induced genes in MSUM strains was more statistically significant than the overlap between *GCN4* targets and induced genes in 82% of the SPELL datasets. The datasets with a higher degree of overlap consisted mostly (at least 276/343) of gene deletions and stress conditions in which *GCN4* is known to play a role (e.g. heat, nutritional perturbation, osmotic stress and DNA damage) (Available online as Table S4 with the version of this work at http://www.plosgenetics.org/). Furthermore, *GCN4* targets as a whole showed increased ribosome footprint density in MSUM strains (Figure 2.8D). We further confirmed this enrichment for functional *GCN4* targets by examining the predicted Gcn4p binding affinity of the promoters for the affected genes (Nutiu et al., 2011). The promoter regions of the upregulated genes were enriched for Gcn4p binding motifs (Figure 2.8E). Using the same sets of upregulated genes from the SPELL compendium as above, less than 6% of these upregulated gene sets had a mean predicted Gcn4p occupancy greater than the genes upregulated in the MSUM strains (Table S4, available online). Thus, GCN4 target genes are transcriptionally upregulated in all MSUM strains.

To provide context for these gene expression changes, the same analyses were performed on Ribo-seq data from yeast subjected to amino acid (AA) starvation, a well-characterized *GCN4*-inducing condition (Ingolia et al., 2009). In this dataset, 20 minutes of amino acid starvation leads to a 4-fold increase in ribosome footprints on the *GCN4*

ORF. A larger number of genes displayed changes in AA starvation compared to MSUM ablation, and *GCN4* targets as a group had larger fold changes (median 2.0-fold induction vs. 1.2 and 1.1-fold for *uba4Δ* and *ncs6Δ* respectively). (Figure 2.9A, 2.9C). However, a smaller fraction of the significantly changing genes are GCN4 targets (13% in AA-starved cells, vs 29% and 30% for *uba4Δ* and *ncs6Δ* respectively) (Figure 2.8C, 2.9B). Furthermore, the starvation-induced genes had a smaller enrichment for predicted Gcn4p occupancy in their promoters compared to genes upregulated in the MSUM strains (Figure 2.8E). The limited induction of high-affinity Gcn4p targets in MSUM mutants is consistent with a weak but specific activation of the *GCN4* pathway.

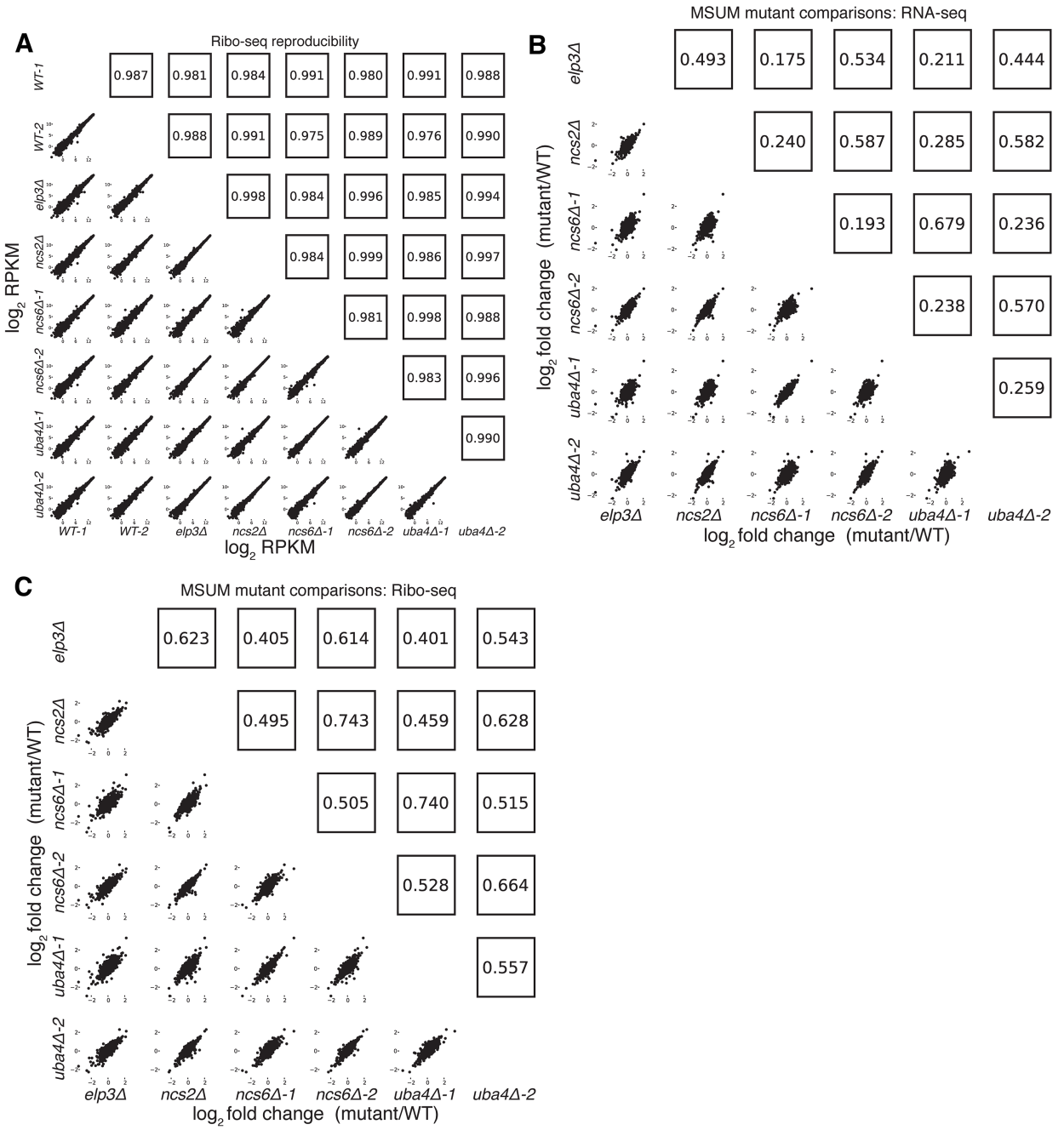

Figure 2.7: All MSUM mutants show similar gene expression changes compared to WT.

(A) Reproducibility of Ribo-seq data.

(B) Comparison of RNA-seq RPKM changes in mutant libraries.

(C) Comparison of Ribo-seq RPKM changes in mutant libraries. Pearson r^2 are presented.

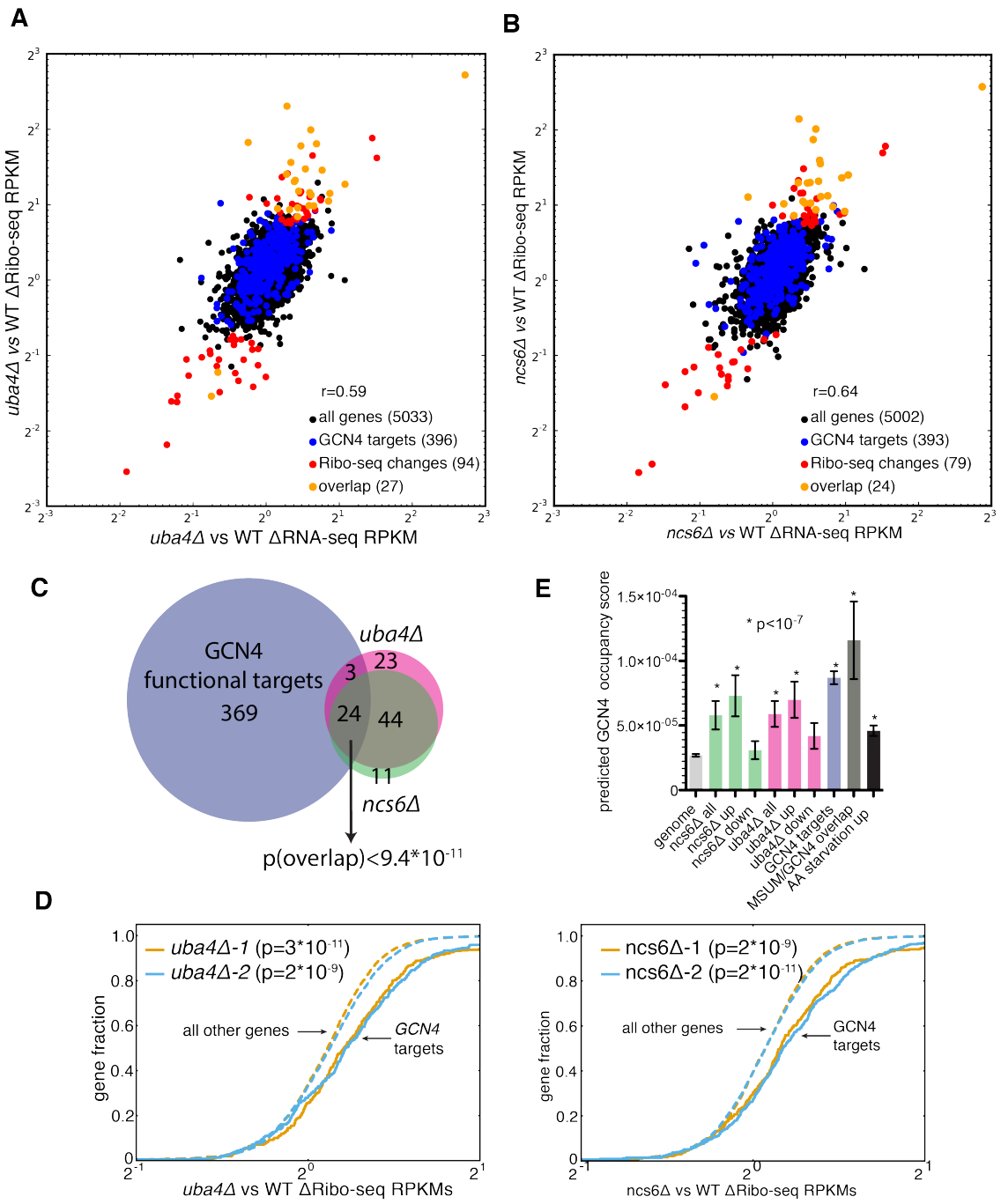

Figure 2.8: MSUM strains show the gene-expression signatures of *GCN4* activation

(A) Comparison of RNA-seq and Ribo-seq RPKM changes in *uba4Δ*. *GCN4* targets and statistically significant Ribo-seq changes are indicated. Values are the means of 2 biological replicates. Pearson R values shown.

(B) Comparison of RNA-seq and Ribo-seq RPKM changes in *ncs6Δ*, as in panel (A).

(C) Venn diagram of overlap between *GCN4* functional targets (blue) and significant Ribo-seq RPKM changes in *uba4Δ* (pink) and *ncs6Δ* (green). The significance of the overlap was computed using the hypergeometric distribution.
(D) Cumulative distribution plots of fold Ribo-seq changes for *GCN4* targets (solid lines) compared to all other genes (dashed lines) in *uba4Δ* (top) and *ncs6Δ*. P values are from a KS test of *GCN4* targets against the rest of the genome.
(E) Mean±SEM of predicted Gcn4p occupancy for groups of genes from panel B and figure S5, as determined by high-throughput *in vitro* binding assays (Nutiu et al., 2011). Bars are colored to match groups in panel B. P values are from t-tests comparing the indicated gene set against all genes in the genome.

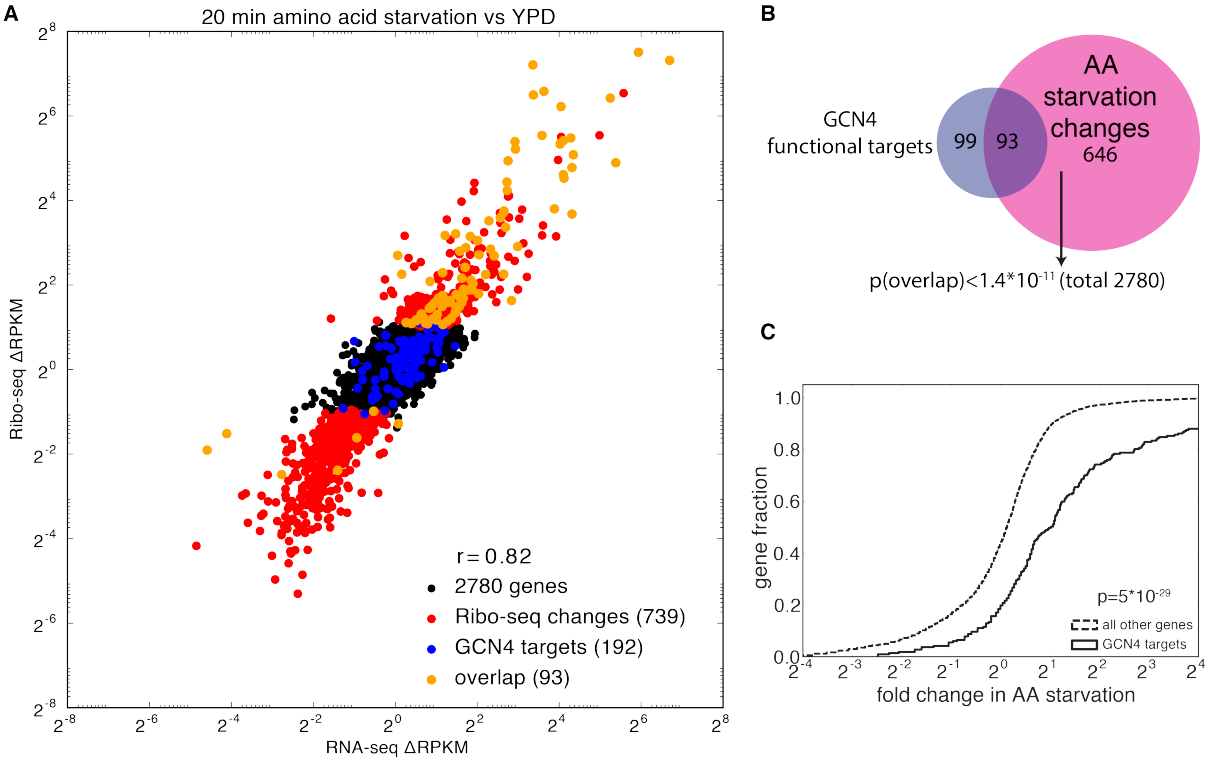

Figure 2.9: Amino acid starvation causes a stronger but less specific activation of GCN4 targets than MSUM ablation.
(A) Comparison of RNA-seq and Ribo-seq RPKM changes in amino acid (AA) starved yeast (data from (Ingolia et al., 2009)). *GCN4* targets and statistically significant Ribo-seq changes are indicated. Values are the means of 2 biological replicates.
(B) Venn diagram of overlap between *GCN4* functional targets (blue) and significant Ribo-seq changes upon AA starvation. The significance of the overlap was computed using the hypergeometric distribution.

(C) Cumulative distribution plots of fold Ribo-seq changes for *GCN4* targets (solid line) compared to all other genes (dashed line). P values are from a KS test of *GCN4* targets against the rest of the genome.

Induction of GCN4 Occurs Independently of GCN2

We next sought to identify the mechanism of *GCN4* pathway induction in MSUM strains. *GCN4* is known to be translationally regulated in response to a variety of insults, most notably by amino acid starvation (Hinnebusch, 2005). Translational repression of *GCN4* is mediated by four upstream open reading frames (uORFs), which prevent ribosomes from initiating on the protein-coding ORF. Conditions that decrease the efficiency of re-initiation allow some ribosomes to scan through the uORFs and initiate at the *GCN4* ORF. All four MSUM mutants showed ~2-fold translational upregulation of *GCN4*, as evidenced by increased ribosome footprint density in the ORF with no increase in mRNA levels (Figure 2.10A).

A reporter construct containing the transcript leader of *GCN4* fused to *lacZ* verified that the uORF-containing leader was sufficient to recapitulate the translational induction observed in MSUM strains (Figure 2.10B). The magnitude of this induction (2-4 fold) is consistent with a weak activation of the GCN pathway, as a 3hr shift to SC-Ura, and a constitutive *GCN2* allele (Ramirez et al., 1992) induced GCN4-lacz 7-fold and 50-fold, respectively. The best-characterized pathway of inducing *GCN4* involves the activation of the Gcn2p kinase by uncharged tRNA, leading to phosphorylation of eukaryotic initiation factor 2α (eIF2α) and reduced efficiency of initiation and re-initiation. We therefore tested the effect of *gcn2Δ* on *GCN4* induction by MSUM mutants.

Surprisingly, *GCN4-lacZ* was still induced in MSUM strains lacking GCN2 (Figure 2.10C). In addition, basal eIF2α phosphorylation levels were not increased in the MSUM strains, consistent with a *GCN2*-independent mechanism (Figure 2.10D). Thus, GCN4 translational induction in MSUM strains occurs by a non-canonical pathway.

In addition to the canonical *GCN2*-dependent response, some tRNA charging and modification defects have been shown to cause induction of *GCN4* by a *GCN2*-independent mechanism (Daugeron et al., 2011; de Aldana et al., 1994; Qiu et al., 2000). MSUM mutations may affect charging. *In vitro* experiments have shown that loss of the s^2 moiety of MSUM tRNAs reduces the efficiency of tRNA charging (Sen and Ghosh, 1976; Seno et al., 1974), although steady state tRNA charging levels are unaltered in MSUM mutants (Johansson et al., 2008). We reasoned that a kinetic defect in tRNA charging could lead to compensatory increases in tRNA synthetase gene expression (Frugier et al., 2005), which could suppress steady-state charging defects. We examined synthetase expression by unsupervised hierarchical clustering of mRNA abundance changes in all of the mutant strains. GlnRS, LysRS, GluRS and AspRS formed a cluster of increased expression in the MSUM mutants (Figure 2.11). Three of these synthetases (Gln, Lys, and Glu) have MSUM tRNAs as substrates. The specific upregulation of this set of tRNA synthetases, along with the global activation of *GCN4* targets, suggests that MSUM mutants have adjusted their cellular state to cope with the loss of the mcm^5s^2U modification (see Discussion).

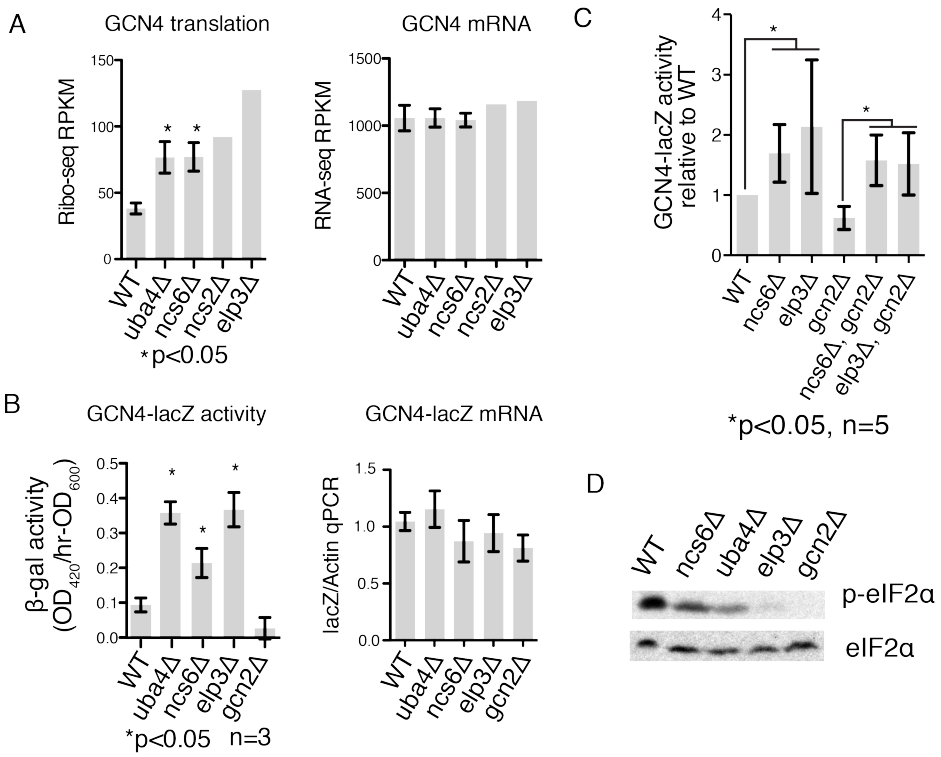

Figure 2.10: *GCN4* is induced independently of *GCN2* in MSUM strains.
(A) Ribo-seq and RNA-seq RPKMs for the *GCN4* open reading frame. Standard deviations are indicated for strains with replicate data.
(B) The indicated strains were transformed with a reporter containing the promoter and transcript leader of *GCN4* fused to *lacZ*. *LacZ* activity and mRNA levels were measured in log phase after overnight growth in YPD.
(C) *LacZ* assays were performed as in panel B, with the addition of double mutant strains. P values are for t-test against WT unless otherwise indicated.
(D) Western blot of total and phosphorylated eIF2α, in yeast cultures grown as in (B).

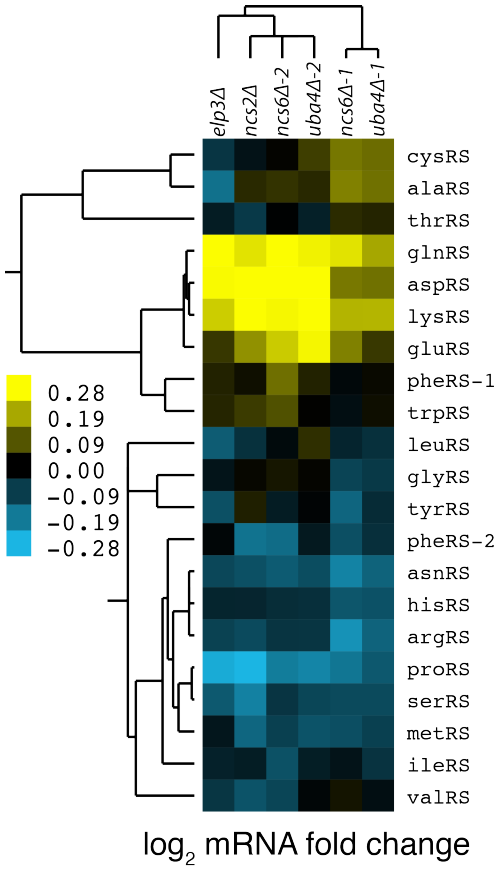

Figure 2.11: Aminoacyl-tRNA synthetases for MSUM tRNAs show a coordinated mRNA upregulation in MSUM strains.
Clustering of mRNA RPKM changes in MSUM strains clusters glnRS, lysRS, gluRS together. It is not clear why AspRS should be affected, but it has a unique regulatory mechanism (Frugier and Giegé, 2003), and clusters apart from the other synthetases in large scale microarray studies (Gasch et al., 2000).

Disruption of The GCN Pathway Partially Suppresses Some MSUM Phenotypes

To investigate the functional significance of *GCN4* misregulation in MSUM mutants, double mutants were constructed between *gcn2Δ* or *gcn4Δ* and *ncs6Δ* or *elp3Δ*, and tested for growth under conditions where MSUM mutants grow poorly. Under heat (40°C), caffeine and diamide stress, *gcnΔ*/MSUM double mutants showed

some increase in growth compared to the single MSUM mutants (Figure 2.12A, 2.13A). On rapamycin, the suppression by *gcn* deletion was similar in magnitude to the suppression by high-copy (hc)-tRNA (Figure 2.12A). We did not observe any rescue of slow growth on YPD at 30°C with either GCN deletion or hc-tRNA expression (Figure 2.12A, 2.13B). Expressing hc-tRNA in the double mutant strains conferred additional resistance in all stress conditions, indicating that the *GCN* pathway contributes to the MSUM phenotypes independently of the pathway affected by hc-tRNA expression (Figure 2.12B).

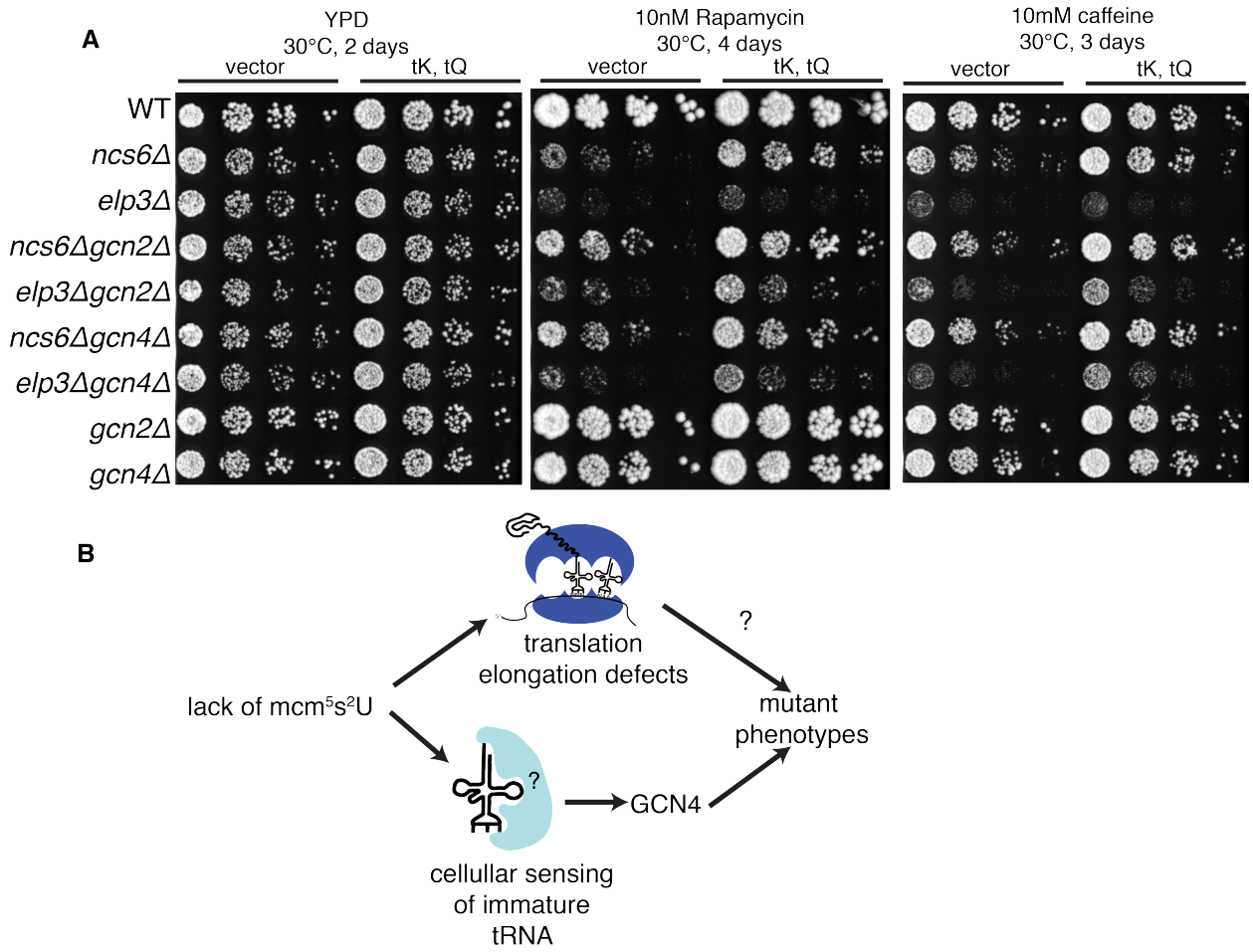

Figure 2.12: Disruption of the GCN pathway partially suppresses the stress sensitivity of MSUM strains, independently of tRNA overexpression.

(A) Yeast was grown to saturation in selective media. 5-fold serial dilutions were spotted onto YPD containing the indicated drug, and grown at the indicated temperature.
(B) The independent rescue of MSUM phenotypes by *gcnΔ* and hc-tRNA suggests that two independent pathways contribute to the mutant phenotypes.

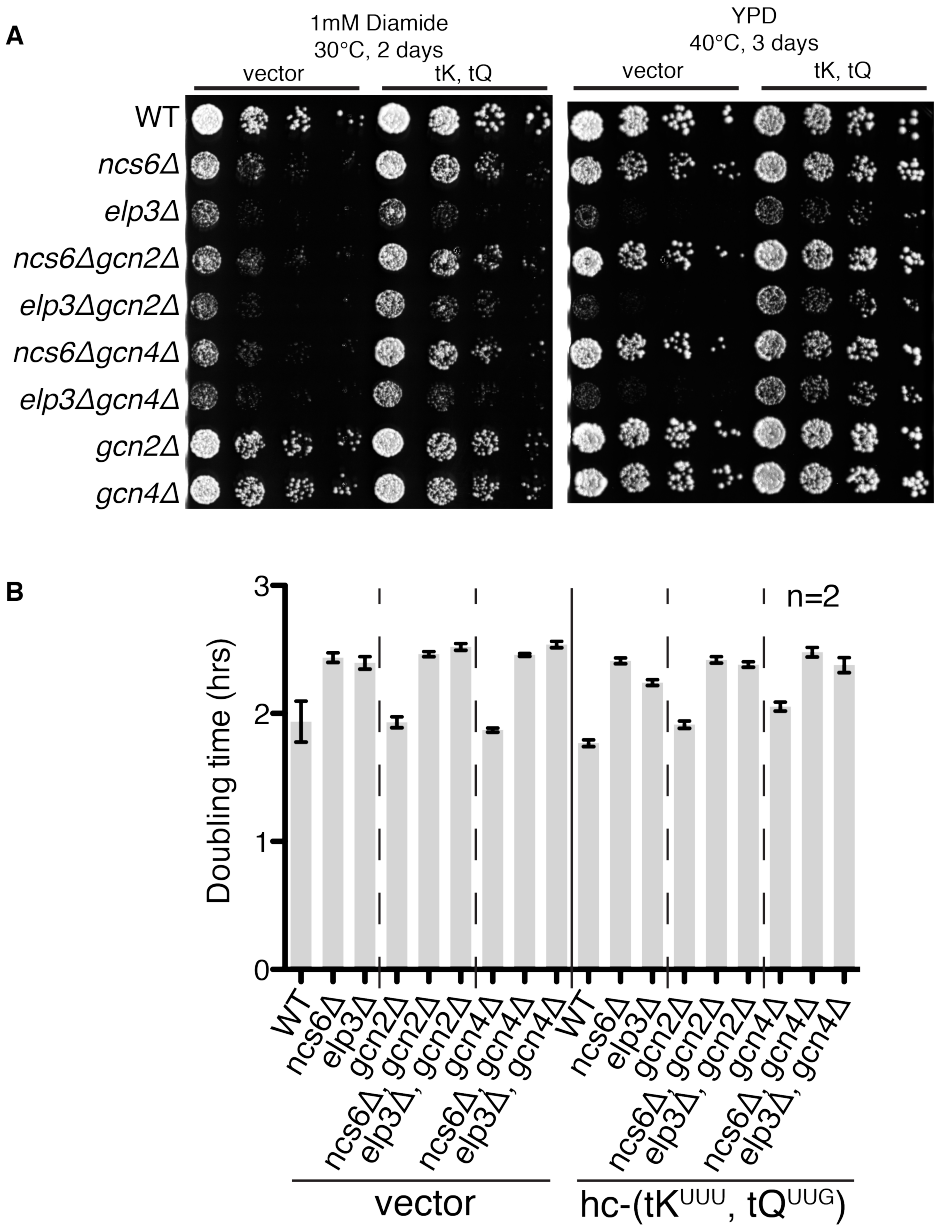

Figure 2.13: Effects of GCN disruption or hc-tRNA on MSUM phenotypes.

(A) Strains tested for growth in additional stress conditions. Yeast were grown to saturation in selective media. 5-fold serial dilutions were spotted onto YPD containing the indicated drug, and grown at the indicated temperature.
(B) Doubling times for strains grown in liquid media at 30°C. The means of two biological replicates, each with four technical replicates, is presented. The error bars indicate the propagated standard deviation of these measurements.

Discussion

MSUM tRNA modifications are conserved throughout eukarya and are required for organismal fitness in yeast, *C. elegans*, and humans. Due to the striking phenotypes of MSUM mutants, as well as the reported suppression by hc-tRNA (Esberg et al., 2006), we expected to find large increases in ribosome density at codons decoded by MSUM tRNAs. We did detect increased ribosome density at VAA codons, and the largest effects of MSUM ablation occurred in the ribosomal A-site, the only site where tRNA binding, and thus concentration, is expected to play a role (Kapp and Lorsch, 2004). Thus, our analysis was capable of detecting codon-level translation defects in these mutants. However, the small magnitude of the observed effect makes it unlikely that protein output is generally affected. Additionally, suppression by hc-tRNA was incomplete in our hands, and the extent of both phenotypes and suppression varied between *elp3Δ* and *ncs6Δ* mutants when they were directly compared, as opposed to examined separately as in previous studies (Esberg et al., 2006). This suggests that MSUM genes may play additional roles in the cell, or create tRNA defects that are not suppressible by tRNA overexpression.

Overall, we found complex and varied patterns of ribosome density surrounding the different codons of the genetic code. These patterns appear to be determined not by cognate tRNA concentrations, but by intrinsic properties of aminoacyl tRNAs or peptidyl

transfer kinetics, consistent with previous data showing that synonymous codon usage had little effect on protein output when mRNAs were expressed at physiological levels (Kudla et al., 2009; Pedersen, 1984). This overall result is also consistent with the results of a systematic study of protein output from codon-repeat reporters (Letzring et al., 2010). Our data do not recapitulate all of the findings of that study, most likely because the reporters contained unnaturally long stretches of rare codons and were expressed at levels high enough to deplete the native tRNA pool. Furthermore, unlike reporter gene assays, Ribo-seq is able to detect changes in translation rate that are too small to be detected in an assay for protein output.

Since tRNA concentrations vary over an order of magnitude (Tuller et al., 2010), yet had little effect on ribosome distributions at different codons, it is hard to understand how a ~2-3 fold overexpression of hypomodified tRNA (Bjork et al., 2007) could strongly affect the rate of ribosome movement. Our data do not rule out the possibility that one or more lowly expressed genes have elongation defects in MSUM mutants that are sufficient to reduce protein output. If so, there must be additional features that make codons in those genes unusually sensitive to the lack of the mcm^5s^2U modification. Indeed, loss of MSUM has been shown to cause a reduction in protein output in artificially sensitized conditions, such as the readthrough of stop codons by a suppressor tRNA (Chen et al., 2011; Huang et al., 2005). It is also possible that larger codon-specific translation defects were not manifest in our growth conditions, which would be consistent with the inability of hc-tRNA to rescue the slow growth of MSUM mutants on YPD. Our data also do not rule out the possibility that a slight increase in

ribosome dwell time could lead to amino acid misincorporation (Patil et al., 2012), misfolding of the protein product (Zhang et al., 2009), or degradation of the mRNA and/or protein by the mRNA surveillance machinery (Shoemaker and Green, 2012). Further experiments are needed to understand the mechanism(s) of phenotypic suppression by hc-tRNAs.

The largest changes detected in the MSUM mutants were transcriptional effects consistent with activation of the *GCN4* pathway. The gene expression signature of *GCN4* induction was noticed previously in *elpΔ* mutants (Krogan and Greenblatt, 2001), and was attributed to the presumed role of Elongator in transcription. However, the similarity of the *elp3Δ* gene expression changes to those of *ncs6Δ*, *ncs2Δ* and *uba4Δ,* which have clear roles in an independent tRNA modification pathway (Leidel et al., 2009; Noma et al., 2009; Schlieker et al., 2008), argues against this explanation. Instead, it appears that improperly modified tRNAs elicit a cellular stress response.

There is precedent for *GCN2*-independent activation of the *GCN4* pathway by perturbations of tRNAs. Nuclear aminoacylation of tRNAs facilitates export to the cytoplasm in yeast and Xenopus oocytes (Grosshans et al., 2000; Lund and Dahlberg, 1998), and disruption of this process can lead to nuclear accumulation of tRNA, as well as *GCN2*-independent *GCN4* induction (de Aldana et al., 1994; Qiu et al., 2000). Loss of the s^2 modification has been previously shown to reduce the rate of *in vitro* aminoacylation reactions for MSUM tRNAs (Sen and Ghosh, 1976; Seno et al., 1974). This charging defect could lead to nuclear accumulation of tRNA and the observed *GCN2*-independent induction of *GCN4*, despite the normal steady-state levels of

charged tRNA in MSUM strains (Johansson et al., 2008). The apparent transcriptional upregulation of all three synthetases that recognize MSUM tRNAs may reflect a cellular response to such a defect in tRNA charging. Consistent with a role for the *GCN* pathway in mediating physiologically relevant signaling in response to loss of MSUM, deletion of *GCN2* or *GCN4* partially suppressed the phenotypes of MSUM strains.

The observation that *GCN* deletion suppresses MSUM phenotypes independently of the phenotypic suppression conferred by hc-tRNA suggests that there are at least two independent pathways contributing to the MSUM phenotypes. This may have implications for Elongator complex mutants in higher eukaryotes. In *C. elegans,* rescue of MSUM phenotypes by hc-tRNA has not been demonstrated. Furthermore, the translational effects reported in *C. elegans* MSUM strains (Chen et al., 2009) are more consistent with a global decrease in translation initiation, as might be expected in conditions leading to *GCN4* activation, than with codon-specific elongation defects. Such secondary effects on gene expression may also play a role in the neurological symptoms of patients with mutations in *elp* genes. Indeed, induced pluripotent stem cells from FD patients with hypomorphic alleles of *elp1* display numerous transcriptional changes during differentiation compared to controls (Lee et al., 2009). It will be important to determine the extent to which tRNA-responsive signaling and transcriptional changes, in addition to codon-specific translation defects, contributes to the phenotypes of MSUM mutants in higher eukaryotes, and the severe and varied symptoms of FD patients.

Materials and Methods

Yeast Strains and Culture Conditions

All strains (Table 2.1) were in the s288c BY4742 background (*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*). MSUM and *GCN* deletions strains were constructed by PCR-mediated gene replacement as previously described (Longtine et al., 1998). All strains were grown in YPD (1% Yeast extract, 2% Peptone, 0.01% Adenine hemisulfate, 2% Dextrose) unless otherwise indicated. For growth assays with hc-tRNA plasmids, strains were grown in SC-Leu to maintain selection. Strains were then plated onto YPD.

Ribo-seq and RNA-seq

Yeast strains were grown from an OD600 of ~0.001-0.004 in aerated flasks at 30°C to mid-log phase (OD ~0.7), treated with 0.1mg/ml cycloheximide for 2 minutes, and harvested by centrifugation. Cells were lysed by vortexing with glass beads, and libraries were prepared essentially as described (Brar et al., 2012; Ingolia et al., 2009). For the WT-2, *elp3Δ*, *ncs2Δ*, *uba4Δ-2*, *ncs6Δ-2* libraries, triton was omitted until after lysis. For any analysis in which only 2 libraries are compared, the mutant was always compared to the WT sample processed identically. Sequencing data were deposited in the GEO database with the accession number GSE45366.

Read Mapping and Positional Assignment

Data analysis was performed using custom Python and Bash scripts developed in-house, unless otherwise indicated. Reads were mapped based on their 5' 21nt using Bowtie (Langmead et al., 2009). Reads were first mapped to *S. cerevisiae* rRNA, allowing up to 3 mismatches, and any mapping multiplicity. Any reads mapping to rRNA were discarded. Reads were then mapped to the *S. cerevisiae* genome downloaded from the saccharomyces genome database (SGD) on 5/26/2010, allowing up to 3 mismatches and requiring unique mapping. Read lengths were determined by comparing the original read sequence to the genomic sequence. Reads for which the beginning of the *in vitro* poly-A tail coincides with a genomic A have ambiguous length, and were excluded from length-specific analyses. Open reading frame (ORF) annotations downloaded from SGD were used to produce mappings of reads relative to the start codon for each ORF, which were used for all downstream calculations. For all codon-level analyses, reads of each length were processed separately, and 5' end mapping locations were subsequently pooled, and shifted 5' with the appropriate offsets (25:0, 26:0, 27:0, 28:0, 29:-1, 30:-1, 31:-2, negative numbers imply a 3' shift) to put them in frame with 28mer reads. When computing RPKMs (reads per kilobase of ORF sequence per million ORF reads) and read counts for each ORF, an unsplit pool of reads was used. The ORF positions are defined from 12nt upstream of the start codon to 14nt upstream of the stop codon. The first 8 codons of each ORF were excluded from all gene expression calculations to exclude possible artifacts from cycloheximide incubation.

Metacodon Plots and Bulk Occupancy Calculations

The value of position i in the metacodon vector for codon NNN is computed as follows:

$$reads(j,i) = \#\, read\ 5'\, ends\ mapping\ (i-21) nt\ upstream\ of\ 1st\ nt\ of\ codon\ j$$

Where the 21nt offset is the 28mer P-site offset (12nt) plus the distance from the p-site to the first nt in the metacodon plot.

$$raw\ metacodon(NNN,i) = \sum_{\substack{all\ codons\ j\ of \\ sequence\ NNN}} reads(j,i)$$

The normalized metacodon vector is computed by normalizing to the peak heights of the outer sites:

$$metacodon(NNN,i) = \frac{raw\ metacodon(NNN,i)}{mean(raw\ metacodon(NNN,j)\ for\ j\ in\ [0,3,15,18]}$$

The mapping of metacodon peaks to ribosomal sites is: (0:-2, 3:-1, 6:A, 9:P, 12:E,15:+1,18:+2). For Figure S1D, the summation is performed over all codon positions for the given amino-acid pair, using the position of the first nucleotide of the first codon in the pair.

Single Codon Occupancy Metric

The single codon occupancy for codon i in gene j in ribosomal site k is computed as:

$$occupancy(i,j,k) = \frac{\#\ read\ 5'\ ends\ corresponding\ to\ i\ in\ site\ k}{\frac{total\ \#\ reads\ in\ j}{\#\ codons\ with \geq 1\ read\ in\ j}}$$

For both the numerator and denominator, only in-frame reads (those whose 5' ends fall a multiple of 3 from the first nt of the site) were counted, and the first 4 codons, as well as codons with no in-frame reads were excluded.

Hierarchical Clustering

For Figure 2.3, the normalized metacodon vectors for each codon were used as inputs for cluster 3.0 (Eisen et al., 1998). Codons were clustered using spearman correlation and single linkage. Heatmaps were generated using Java Treeview (Saldanha, 2004). The tAI column was not used for clustering, and was only added afterwards for comparison. For Figure 2.11, centroid linkage was used for clustering.

Queuing Analysis

For each AAA and CAA codon with ≥ 2-fold increase, the reads at each surrounding position were normalized by the mean read density for the entire ORF. These values were summed relative to all of the codons analyzed, offset so that the 0 position corresponds to the codon in the A site, and the value at each position was divided by the total number of codons whose host gene overlapped the given position. A secondary ribosome pileup is expected to occur approximately one ribosome footprint width (~28nt) upstream of the slow codon. Due to the use of polyadenylation in library preparation, any read ending in an adenosine cannot be assigned a length, and is not included in this analysis. Because of this, there is a depletion of read density at ~-10 nts, corresponding to reads that end with 1 or more adenosines.

Gene Expression Analysis

Significant Ribo-seq changes were called using edgeR (Robinson et al., 2009). Significance was assessed using a Bonferroni-corrected p-value cutoff of 0.05. The significance of overlap with *GCN4* targets was assessed using the hypergeometric test, and the definition of target genes derived from Natarajan *et al* (Natarajan et al., 2001). The background for the hypergeometric test was defined as the set of genes with confident expression values for all datasets (5034 genes for MSUM datasets, 2780 for amino acid starvation).

β-galactosidase Assays

Starter cultures containing the GCN4-lacZ reporter plasmid (Table S2) were grown to saturation in SC-URA, then diluted into YPAD and grown in conditions identical to the Ribo-seq samples. At an OD600 of 0.7-0.8, 1ml aliquots each were taken for qPCR and β-galactosidase assays, spun down, media aspirated, and frozen. Pellets were resuspended in Z buffer and permeabilized as previously described (Amberg et al., 2006). Cell suspensions were transferred in triplicate to a transparent 96-well plate, and 1/5 volume of 4mg/ml ONPG was added. OD420 was measured every minute for 1 hour in a Bio-Tek synergy HT plate reader. β-galactosidase activity was defined as the slope of the linear portion of the OD420 vs. time graph, normalized by the OD600 of the culture at harvest.

Quantitative RNA Analysis

RNA was purified from yeast pellets as described (Collart and Oliviero, 2001). Reverse transcription and quantitative PCR was performed using Avian Myeloblastosis Virus Reverse Trancriptase (AMV-RT; Promega) and real-time reagents (Invitrogen) according to manufacturer's instructions using a Roche Lightcycler 480. See Table 2.3 for gene-specific primer sequences.

Western Blotting

Equal OD600 units of cell culture were TCA extracted and run on an SDS-PAGE gel. After transfer to nitrocellulose, blotting was performed with a 1:1000 dilution of phospho-eIF2α antibody (Cell Signaling 9721S). The membranes were stripped with One Minute western blot stripping buffer (GM Biosciences) and reprobed with a 1:1000 dilution of total-eIF2α antibody (a kind gift from Tom Dever).

Automated Liquid Growth Assays

Liquid growth assays were carried out as previously described (Toussaint and Conconi, 2006), except that saturated selective media starter cultures were diluted to an OD of 0.01 in YPD, then diluted 20-fold in YPD to a final volume of 100µl.

Strain	**Genotype**	**Plasmid**	**Source**
YWG11	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*		BY4742
YWG269	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ncs6::kan*		This study
YWG271	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 uba4::kan*		This study
YWG382	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 elp3::kan*		This study
YWG385	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ncs6::kan*		This study
YWG386	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ncs2::kan*		This study

YWG545	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*	p180	This study
YWG547	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ncs6::kan*	p180	This study
YWG557	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 elp3::kan*	p180	This study
YWG654	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 gcn2::kan*	p180	This study
YWG1003	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 elp3::kan gcn2::nat*	p180	This study
YWG1004	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ncs6::kan gcn2::nat*	p180	This study
YWG560	*MATα leu2-3,112 ura3-52 ino1 GCN2c-516*	p180	Ramirez *et al*. MCB, 1992
YWG683	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*	pWG445	This study
YWG684	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ncs6::kan*	pWG445	This study
YWG685	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 elp3::kan*	pWG445	This study
YWG1005	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ncs6::kan gcn2::nat*	pWG445	This study
YWG1006	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 elp3::kan gcn2::nat*	pWG445	This study
YWG1007	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ncs6::kan gcn4::nat*	pWG445	This study
YWG1008	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 elp3::kan gcn4::nat*	pWG445	This study
YWG686	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 gcn2::nat*	pWG445	This study
YWG699	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 gcn4::nat*	pWG445	This study
YWG678	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*	pWG449	This study
YWG679	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ncs6::kan*	pWG449	This study
YWG680	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 elp3::kan*	pWG449	This study
YWG1009	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ncs6::kan gcn2::nat*	pWG449	This study
YWG1010	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 elp3::kan gcn2::nat*	pWG449	This study
YWG1011	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 ncs6::kan gcn4::nat*	pWG449	This study
YWG1012	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 elp3::kan gcn4::nat*	pWG449	This study
YWG681	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 gcn2::nat*	pWG449	This study
YWG704	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 gcn4::nat*	pWG449	This study

Table 2.1: Yeast strains used in this study

Plasmid	**Contents**	**Source**
p180	GCN4-lacZ, URA3, CEN4	(Hinnebusch, 1985)
pWG445 (pRS425)	LEU2, 2μ	(Leidel et al., 2009)

pWG449 (pSZ64)	LEU2, tK(UUU), tQ(UUG), 2μ	(Leidel et al., 2009)

Table 2.2: Plasmids used in this study

Primer	sequence
oBZ47-lacZ_F	GAAAGCTGGCTACAGGAAGG
oBZ48-lacZ_R	GTTGCACCACAGATGAAACG
oTC10-actin_F	TTCTGAGGTTGCTGCTTTGG
oTC11-actin_R	CTTGGTGTCTTGGTCTACCG

Table 2. 3: Primers used for quantitative PCR

Acknowledgements

We thank A. Hinnebusch and S. Leidel for plasmids and strains; T. Dever for total-eIF2α antibody; M. Thompson for compiling the SPELL datasets; U. RajBhandary and D. Bartel for critical reading of the manuscript; and C. Burge, E. Wang and members of the Gilbert Lab for helpful discussions.

References

Agris, P.F., Vendeix, F.A.P., and Graham, W.D. (2007). tRNA's wobble decoding of the genome: 40 years of modification. Journal of Molecular Biology *366*, 1–13.

Amberg, D.C., Burke, D.J., and Strathern, J.N. (2006). Assay of β-Galactosidase in Yeast: Permeabilized Cell Assay. Cold Spring Harbor Protocols *2006*, 4158.

Arava, Y., Wang, Y., Storey, J.D., Liu, C.L., Brown, P.O., and Herschlag, D. (2003). Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U.S.A. *100*, 3889–3894.

Bjork, G.R., Huang, B., Persson, O.P., and Bystrom, A.S. (2007). A conserved modified wobble nucleoside (mcm5s2U) in lysyl-tRNA is required for viability in yeast. RNA *13*, 1245–1255.

Brar, G.A., Yassour, M., Friedman, N., Regev, A., Ingolia, N.T., and Weissman, J.S. (2012). High-Resolution View of the Yeast Meiotic Program Revealed by Ribosome Profiling. Science *335*, 552–557.

Chan, J.C., Yang, J.A., Dunn, M.J., Agris, P.F., and Wong, T.W. (1982). The nucleotide sequence of a glutamine tRNA from rat liver. Nucleic Acids Research *10*, 3755–3758.

Chen, C., Huang, B., Eliasson, M., Rydén, P., and Byström, A.S. (2011). Elongator Complex Influences Telomeric Gene Silencing and DNA Damage Response by Its Role in Wobble Uridine tRNA Modification. PLoS Genet *7*, e1002258.

Chen, C., Tuck, S., and Byström, A.S. (2009). Defects in tRNA modification associated with neurological and developmental dysfunctions in Caenorhabditis elegans elongator mutants. PLoS Genet *5*, e1000561.

Collart, M.A., and Oliviero, S. (2001). Preparation of yeast RNA. Curr Protoc Mol Biol *Chapter 13*, Unit13.12.

Daugeron, M.C., Lenstra, T.L., Frizzarin, M., Yacoubi, El, B., Liu, X., Baudin-Baillieu, A., Lijnzaad, P., Decourty, L., Saveanu, C., Jacquier, A., et al. (2011). Gcn4 misregulation reveals a direct role for the evolutionary conserved EKC/KEOPS in the t6A modification of tRNAs. Nucleic Acids Research *39*, 6148–6160.

de Aldana, C.R.V., Wek, R.C., Segundo, P.S., Truesdell, A.G., and Hinnebusch, A.G. (1994). Multicopy tRNA genes functionally suppress mutations in yeast eIF-2 alpha kinase GCN2: evidence for separate pathways coupling GCN4 expression to unchanged tRNA. Mol Cell Biol *14*, 7920–7932.

Doerfel, L.K., Wohlgemuth, I., Kothe, C., Peske, F., Urlaub, H., and Rodnina, M.V. (2013). EF-P is essential for rapid synthesis of proteins containing consecutive proline residues. Science *339*, 85–88.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. U.S.A. *95*, 14863–14868.

Esberg, A., Huang, B., Johansson, M.J.O., and Byström, A.S. (2006). Elevated Levels of Two tRNA Species Bypass the Requirement for Elongator Complex in Transcription and Exocytosis. Molecular Cell *24*, 139–148.

Frugier, M., and Giegé, R. (2003). Yeast Aspartyl-tRNA Synthetase Binds Specifically its Own mRNA. Journal of Molecular Biology *331*, 375–383.

Frugier, M., Ryckelynck, M., and Giegé, R. (2005). tRNA-balanced expression of a eukaryal aminoacyl-tRNA synthetase by an mRNA-mediated pathway. EMBO Rep. *6*, 860–865.

Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., and Brown, P.O. (2000). Genomic expression programs in the response of yeast cells to environmental changes. Mol. Biol. Cell *11*, 4241–4257.

Grosshans, H., Hurt, E., and Simos, G. (2000). An aminoacylation-dependent nuclear tRNA export pathway in yeast. Genes & Development *14*, 830–840.

Hibbs, M.A., Hess, D.C., Myers, C.L., Huttenhower, C., Li, K., and Troyanskaya, O.G. (2007). Exploring the functional landscape of gene expression: directed search of large microarray compendia. Bioinformatics *23*, 2692–2699.

Hinnebusch, A.G. (1985). A hierarchy of trans-acting factors modulates translation of an activator of amino acid biosynthetic genes in Saccharomyces cerevisiae. Mol Cell Biol *5*, 2349–2360.

Hinnebusch, A.G. (2005). Translational regulation of GCN4 and the general amino acid control of yeast. Annu. Rev. Microbiol. *59*, 407–450.

Huang, B., Johansson, M.J.O., and Byström, A.S. (2005). An early step in wobble uridine tRNA modification requires the Elongator complex. RNA *11*, 424–436.

Ingolia, N.T., Ghaemmaghami, 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.

Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell *147*, 789–802.

Johansson, M., and Byström, A. (2005). Transfer RNA modifications and modifying enzymes in Saccharomyces cerevisiae. Fine-Tuning of RNA Functions by Modification and Editing 87–120.

Johansson, M., Ieong, K.-W., Trobro, S., Strazewski, P., Åqvist, J., Pavlov, M.Y., and Ehrenberg, M. (2011). pH-sensitivity of the ribosomal peptidyl transfer reaction dependent on the identity of the A-site aminoacyl-tRNA. Proceedings of the National Academy of Sciences *108*, 79–84.

Johansson, M.J.O., Esberg, A., Huang, B., Björk, G.R., and Byström, A.S. (2008). Eukaryotic wobble uridine modifications promote a functionally redundant decoding system. Mol Cell Biol *28*, 3301–3312.

Kapp, L.D., and Lorsch, J.R. (2004). The molecular mechanics of eukaryotic translation. Annu. Rev. Biochem. *73*, 657–704.

Krogan, N.J., and Greenblatt, J.F. (2001). Characterization of a Six-Subunit Holo-Elongator Complex Required for the Regulated Expression of a Group of Genes in Saccharomyces cerevisiae. Mol Cell Biol *21*, 8203–8212.

Krüger, M.K., Pedersen, S., Hagervall, T.G., and Sørensen, M.A. (1998). The modification of the wobble base of tRNAGlu modulates the translation rate of glutamic acid codons in vivo. Journal of Molecular Biology *284*, 621–631.

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.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology *10*, R25.

Lee, G., Papapetrou, E.P., Kim, H., Chambers, S.M., Tomishima, M.J., Fasano, C.A., Ganat, Y.M., Menon, J., Shimizu, F., Viale, A., et al. (2009). Modelling pathogenesis and treatment of familial dysautonomia using patient-specific iPSCs. Nature *461*, 402–406.

Leidel, S., Pedrioli, P.G.A., Bucher, T., Brost, R., Costanzo, M., Schmidt, A., Aebersold, R., Boone, C., Hofmann, K., and Peter, M. (2009). Ubiquitin-related modifier Urm1 acts as a sulphur carrier in thiolation of eukaryotic transfer RNA. Nature *458*, 228–232.

Letzring, D.P., Dean, K.M., and Grayhack, E.J. (2010). Control of translation efficiency in yeast by codon-anticodon interactions. RNA *16*, 2516–2528.

Li, G.-W., Oh, E., and Weissman, J.S. (2012). The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature *484*, 538–541.

Lodish, H.F., and Jacobsen, M. (1972). Regulation of hemoglobin synthesis. Equal rates of translation and termination of - and -globin chains. J. Biol. Chem. *247*, 3622–3629.

Longtine, M.S., McKenzie, A., Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J.R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast *14*, 953–961.

Lund, E., and Dahlberg, J.E. (1998). Proofreading and Aminoacylation of tRNAs Before Export from the Nucleus. Science *282*, 2082–2085.

Mehlgarten, C., Jablonowski, D., Wrackmeyer, U., Tschitschmann, S., Sondermann, D., Jäger, G., Gong, Z., Byström, A.S., Schaffrath, R., and Breunig, K.D. (2010). Elongator function in tRNA wobble uridine modification is conserved between yeast and plants. Mol. Microbiol. *76*, 1082–1094.

Natarajan, K., Meyer, M.R., Jackson, B.M., Slade, D., Roberts, C., Hinnebusch, A.G., and Marton, M.J. (2001). Transcriptional profiling shows that Gcn4p is a master regulator of gene expression during amino acid starvation in yeast. Mol Cell Biol *21*, 4347–4368.

Noma, A., Sakaguchi, Y., and Suzuki, T. (2009). Mechanistic characterization of the sulfur-relay system for eukaryotic 2-thiouridine biogenesis at tRNA wobble positions. Nucleic Acids Research *37*, 1335–1352.

Nutiu, R., Friedman, R.C., Luo, S., Khrebtukova, I., Silva, D., Li, R., Zhang, L., Schroth, G.P., and Burge, C.B. (2011). Direct measurement of DNA affinity landscapes on a high-throughput sequencing instrument. Nat Biotechnol *29*, 659–664.

Otero, G., Fellows, J., Li, Y., de Bizemont, T., Dirac, A.M., Gustafsson, C.M., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J.Q. (1999). Elongator, a multisubunit component of a novel RNA polymerase II holoenzyme for transcriptional elongation. Molecular Cell *3*, 109–118.

Patil, A., Chan, C.T.Y., Dyavaiah, M., Rooney, J.P., Dedon, P.C., and Begley, T.J. (2012). Translational infidelity-induced protein stress results from a deficiency in Trm9-catalyzed tRNA modifications. RNA Biol *9*, 990–1001.

Pavlov, M.Y., Watts, R.E., Tan, Z., Cornish, V.W., Ehrenberg, M., and Forster, A.C. (2009). Slow peptide bond formation by proline and other N-alkylamino acids in translation. Proceedings of the National Academy of Sciences *106*, 50–54.

Pechmann, S., and Frydman, J. (2012). Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. Nature Structural & Molecular Biology *20*, 237–243.

Pedersen, S. (1984). Escherichia coli ribosomes translate in vivo with variable rate. Embo J. *3*, 2895–2898.

Phizicky, E.M., and Hopper, A.K. (2010). tRNA biology charges to the front. Genes & Development *24*, 1832–1860.

Qiu, H., Hu, C., Anderson, J., Bjork, G.R., Sarkar, S., Hopper, A.K., and Hinnebusch, A.G. (2000). Defects in tRNA processing and nuclear export induce GCN4 translation independently of phosphorylation of the alpha subunit of eukaryotic translation initiation factor 2. Mol Cell Biol *20*, 2505–2516.

Ramirez, M., Wek, R.C., Vazquez de Aldana, C.R., Jackson, B.M., Freeman, B., and Hinnebusch, A.G. (1992). Mutations activating the yeast eIF-2 alpha kinase GCN2: isolation of alleles altering the domain related to histidyl-tRNA synthetases. Mol Cell Biol *12*, 5801–5815.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2009). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Saldanha, A.J. (2004). Java Treeview--extensible visualization of microarray data. Bioinformatics *20*, 3246–3248.

Schlieker, C.D., Van der Veen, A.G., Damon, J.R., Spooner, E., and Ploegh, H.L. (2008). A functional proteomics approach links the ubiquitin-related modifier Urm1 to a

tRNA modification pathway. Proceedings of the National Academy of Sciences *105*, 18255–18260.

Sen, G.C., and Ghosh, H.P. (1976). Role of modified nucleosides in tRNA: effect of modification of the 2-thiouridine derivative located at the 5'-end of the anticodon of yeast transfer RNA Lys2. Nucleic Acids Research *3*, 523–535.

Seno, T., Agris, P.F., and Söll, D. (1974). Involvement of the anticodon region of Escherichia coli tRNAGln and tRNAGlu in the specific interaction with cognate aminoacyl-tRNA synthetase. Alteration of the 2-thiouridine derivatives located in the anticodon of the tRNAs by BrCN or sulfur deprivation. Biochim. Biophys. Acta *349*, 328–338.

Shoemaker, C.J., and Green, R. (2012). Translation drives mRNA quality control. Nature Structural & Molecular Biology *19*, 594–601.

Slaugenhaupt, S.A., Blumenfeld, A., Gill, S.P., Leyne, M., Mull, J., Cuajungco, M.P., Liebert, C.B., Chadwick, B., Idelson, M., Reznik, L., et al. (2001). Tissue-Specific Expression of a Splicing Mutation in the IKBKAP Gene Causes Familial Dysautonomia. The American Journal of Human Genetics *68*, 598–605.

Stadler, M., and Fire, A. (2011). Wobble base-pairing slows in vivo translation elongation in metazoans. RNA *17*, 2063–2073.

Strug, L.J., Clarke, T., Chiang, T., Chien, M., Baskurt, Z., Li, W., Dorfman, R., Bali, B., Wirrell, E., Kugler, S.L., et al. (2009). Centrotemporal sharp wave EEG trait in rolandic epilepsy maps to Elongator Protein Complex 4 (ELP4). European Journal of Human Genetics *17*, 1171–1181.

Toussaint, M., and Conconi, A. (2006). High-throughput and sensitive assay to measure yeast cell growth: a bench protocol for testing genotoxic agents. Nat Protoc *1*, 1922–1928.

Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., and Pilpel, Y. (2010). An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell *141*, 344–354.

Ude, S., Lassak, J., Starosta, A.L., Kraxenberger, T., Wilson, D.N., and Jung, K. (2013). Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. Science *339*, 82–85.

Walden, W.E., Godefroy-Colburn, T., and Thach, R.E. (1981). The role of mRNA competition in regulating translation. I. Demonstration of competition in vivo. J. Biol. Chem. *256*, 11739–11746.

Woolstenhulme, C.J., Parajuli, S., Healey, D.W., Valverde, D.P., Petersen, E.N.,

Starosta, A.L., Guydosh, N.R., Johnson, W.E., Wilson, D.N., and Buskirk, A.R. (2013). Nascent peptides that block protein synthesis in bacteria. Proceedings of the National Academy of Sciences *110*, E878–E887.

Zhang, G., Hubalewska, M., and Ignatova, Z. (2009). Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. Nature Structural & Molecular Biology *16*, 274–280.

Chapter 3: Regulation of Translational Efficiency By Translation Initiation Factors

Most known mechanisms of translational control regulate the initiation phase of translation (Hinnebusch and Lorsch, 2012), which is rate-limiting for the majority of mRNAs. There are both global and mRNA-specific mechanisms for regulating translation initiation, but the distinction between these two is not always clear-cut. Even mechanisms that globally repress translation differentially affect some messages (Lodish, 1974; Walden and Thach, 1986; Walden et al., 1981), or even induce translation of others (Hinnebusch, 2005; Vaidyanathan et al., 2014). In this chapter, I will review how sequence features of mRNAs interact with perturbations of the core initiation machinery to effect mRNA-specific regulation of translational efficiency.

The Mechanism of Ribosome Recruitment on mRNAs

Translation initiation in eukaryotes (Figure 3.1) is a complicated and dynamic process requiring many steps and factors. To prepare the mRNA for recruitment of the small ribosomal (40S) subunit, the 5' methylguanosine cap and poly-A tail are bound by the eukaryotic initiation factor (eIF) 4F complex and poly-A binding protein (PABP), respectively. eIF4F consists of the cap-binding protein eIF4E, the helicase eIF4A and the scaffold protein, eIF4G, which bridges the activities of eIF4E, eIF4A, PABP and other proteins (Aitken and Lorsch, 2012). In addition to multiple protein-protein interaction domains, eIF4G has three RNA binding domains capable of direct binding to

mRNA (Berset, 2003; Park et al., 2010). Although these RNA binding domains are essential for viability in *S. cerevisiae*, their role in translation is unclear.

eIF4F plays multiple incompletely defined roles in translation initiation. First, eIF4F is thought to melt the mRNA structure in the vicinity of the cap. This “activation” of mRNA allows recruitment of the 43S pre-initiation complex containing the small (40S) ribosomal subunit, initiator tRNA, and additional eIFs. However, eIF4F is not strictly required for 43S recruitment – eIF3 is sufficient (Figure 3.1) – although eIF4F greatly accelerates recruitment and is also important for downstream steps leading to 80S complex assembly (Jivotovskaya et al., 2006; Mitchell et al., 2010). These and other findings (Park et al., 2013; 2010) challenge the proposed rigid order of initiation complex assembly, and instead suggest that translation initiation complexes are assembled by a web of unordered and cooperative protein-protein and RNA-protein interactions.

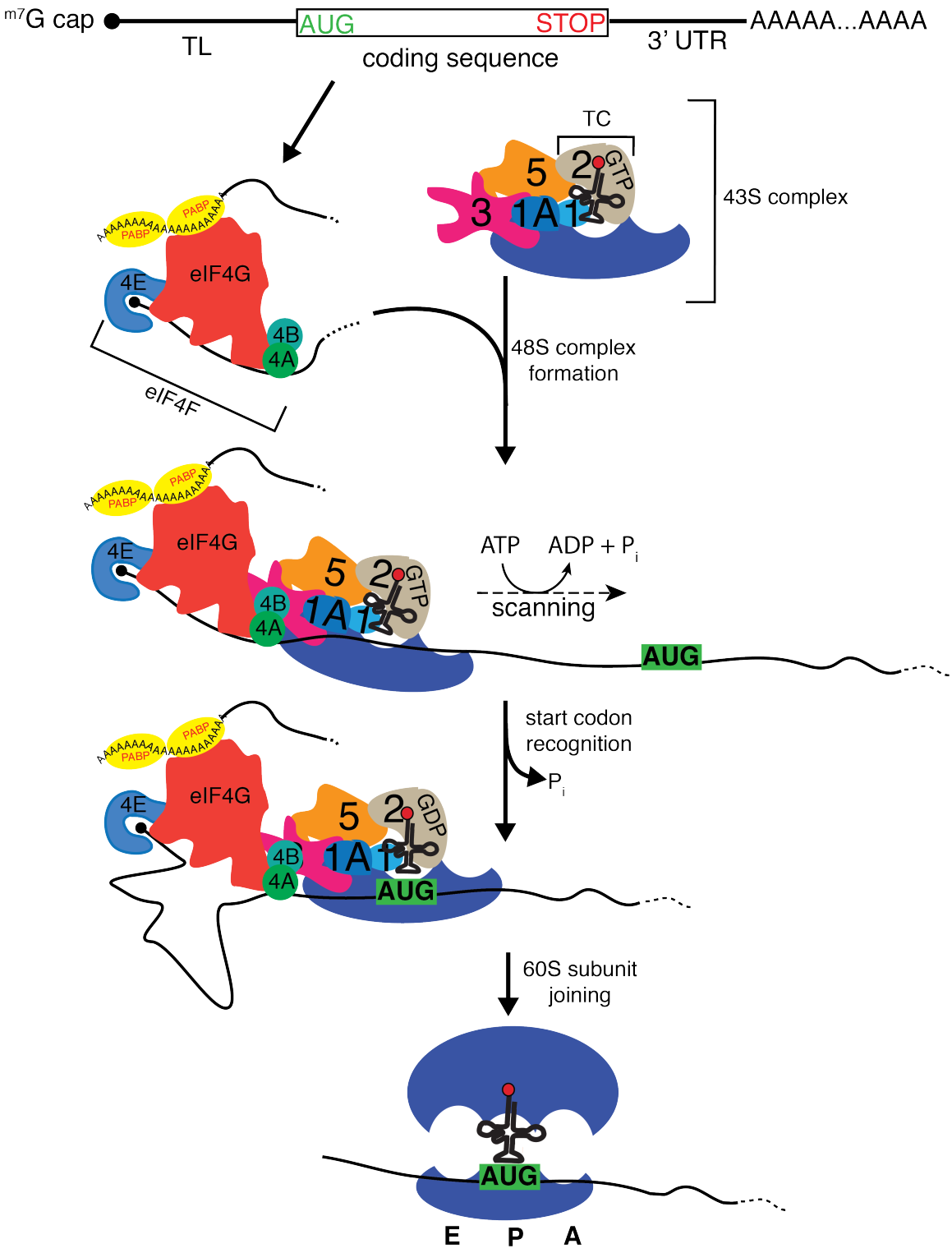

Figure 3.1: Mechanism of eukaryotic translation initiation.
This simplified diagram emphasizes the steps relevant to the regulatory modes detailed in this chapter. See text for details. eIF prefixes have been omitted from many factor names for clarity. TC – ternary complex of eIF2, initiator tRNA and GTP. See (Aitken

and Lorsch, 2012; Hinnebusch and Lorsch, 2012; Jackson et al., 2010) for further details.

Once the small ribosomal subunit has been recruited, the assembled 48S complex then scans 5' to 3' along the message, powered by the ATP-dependent helicase activity of eIF4A (Spirin, 2009; Vassilenko et al., 2011), until the first AUG codon in a proper context is recognized in the ribosome's peptidyltransfer (P) site (Kozak, 1980). Start codon recognition triggers a series of events resulting in a structural rearrangement of the 40S (and associated factors) that allows the 60S ribosomal subunit to join. 60S joining defines the reading frame and concludes initiation, allowing the elongation phase to begin.

The complexity of eukaryotic initiation provides many potential targets for regulation – nearly every step in the pathway is affected by interactions with other proteins or with the mRNA sequence, and these interactions can be modulated in response to a variety of signals. In the remainder of this chapter I will review some of the better-understood mechanisms for regulating translation initiation.

Phosphorylation of eIF2 Rapidly Inhibits Global Translation Initiation While Upstream Open Reading Frames Regulate the TE of Specific mRNAs

A common mechanism for globally inhibiting translation initiation is phosphorylation of the α subunit of eIF2 (eIF2α). eIF2 is a GTPase responsible for loading initiator tRNA on to the small ribosomal subunit. GTP hydrolysis by eIF2 reduces the affinity of eIF2 for the initiator tRNA and contributes to the fidelity of start

codon recognition. Phosphorylation of eIF2α prevents recycling of eIF2-GDP back into eIF2-GTP, thereby reducing the level of eIF2-GTP-$tRNA_i$ ternary compex (TC) and globally decreasing translation initiation. This phosphorylation is carried out by a family of eIF2α kinases, found in all eukaryotes, that are activated in response to diverse stresses including amino acid starvation, double-stranded RNA (a sign of viral infection), protein folding stress, and heme deficiency (Donnelly et al., 2013).

Paradoxically, the global repression of translation by eIF2α phosphorylation leads to a strong translational activation of certain messages including GCN4 and ATF4, which encode stress-responsive transcription factors. The TLs of GCN4 in yeast and ATF4 in metazoan species contain multiple short upstream open reading frames (uORFs). The spacing and sequence context of these uORFs serves as a sensor of TC levels – under normal conditions most ribosomes initiate on the uORFs, preventing translation of the genic ORF. In contrast, when eIF2 is phosphorylated the reduced levels of TC allow a fraction of 43S complexes to scan past the uORFs and initiate on the genic ORF, resulting in a large and rapid translational induction of GCN4 or ATF4 (Hinnebusch, 2005).

uORFs also play a more widespread role in regulating translation initiation. Approximately 20% of yeast transcripts contain a uORF, and the presence of these elements, which tend to be conserved, leads to substantial translational repression (Arribere and Gilbert, 2013; Pelechano et al., 2013). In mice and humans uORFs are found in ~40-50% of transcripts and translation of these transcripts is depressed compared to uORF-less messages (Calvo et al., 2009). There have also been reports of

widespread regulated translation of uORFs, often with non-canonical start codons, during cellular differentiation in yeast and mammals (Brar et al., 2012; Ingolia et al., 2011). Most uORFs permit substantial translation of the downstream ORF via re-initiation or leaky scanning (Kozak, 1989), and the factors that govern partitioning of initiating ribosomes between the uORFs and the genic ORF are only partially understood.

eIF4E Binding Proteins Inhibit Cap-Dependent Translation of Specific mRNAs

Another commonly regulated step in translation is mRNA activation by eIF4F. eIF4E binding proteins (4E-BPs) compete with eIF4G for binding to eIF4E, thereby reducing the amount of active eIF4F complex. In metazoans, the mTOR kinase phosphorylates 4E-BPs and prevents them from binding to eIF4E. Inhibition of mTOR kinase by amino acid starvation or chemical perturbation leads to dephosphorylation of 4E-BPs and a global reduction in protein synthesis (Harris and Lawrence, 2003; Shamji et al., 2003). Intriguingly, mTOR inhibition and subsequent dephosphorylation of 4E-BPs causes a particularly large reduction in translation of mRNAs with 5' terminal oligopyrimidine tracts (Thoreen et al., 2012), though the basis for this selectivity is not understood.

Budding yeast expresses two 4E-BPs, Caf20 and Eap1. Yeast strains lacking these factors show increased translation of specific mRNAs based on genome-wide polysome microarray studies (Cridge et al., 2010). Thus, like the mammalian 4E-BPs, Eap1 and Caf20 preferentially affect a subset of genes. Moreover, they seem to

regulate different sets of genes, most likely by interaction with specific RNA-binding proteins that target their repressing activity to particular messages (Cridge et al., 2010). The yeast 4E-BPs are required for various stress responses, including pseudohyphal growth, a differentiation program that diploid yeast undertake in response to nutrient deprivation (Ibrahimo et al., 2006). Together, these results suggest that both yeast and mammals may exploit the mRNA-specific effects of 4E-BPs to adapt to adverse conditions.

mRNA Competition for Limiting Initiation Factors Can Affect TE

Competition between mRNAs for limiting initiation factors is another mode of modulating initiation rates. In principle, if one mRNA can outcompete others for a limiting pool of some core initiation factor, then it should be preferentially translated. Indeed, several mammalian viruses use such a competition mechanism to hijack the host translational machinery for production of their own proteins (Jackson, 2005). Inter-mRNA competition also affects translation of native cellular messages. For example, in rabbit reticulocytes, α-globin mRNA is 40% more abundant than β-globin mRNA (Phillips et al., 1977), but equal numbers of α- and β-globin protein molecules are required to form functional hemoglobin. Stoichiometric quantities of α- and β-globin are produced by more frequent translation initiation on the β-globin mRNA (Lodish, 1974; Lodish and Jacobsen, 1972). This difference in TE was found to be due to competition of these mRNAs for a limited pool of eIF4F – raising the concentration of eIF4F lead to the production of excess α-globin (Kabat and Chappell, 1977; Ray et al., 1983).

Simple competition may be a more general determinant of the relative rates of mRNA translation than currently appreciated. In Chapter 4, I explore the mechanism of mRNA competition for eIF4G, the limiting member of the eIF4F complex, and one of the least abundant components of the yeast translational machinery (Firczuk et al., 2013; Haar and McCarthy, 2002). My findings suggest that the RNA binding specificity of eIF4G allows for competitive discrimination between yeast mRNAs, which may regulate mRNA-specific translational efficiencies in response to changing cellular conditions.

References

Aitken, C.E., and Lorsch, J.R. (2012). A mechanistic overview of translation initiation in eukaryotes. Nature Structural & Molecular Biology *19*, 568–576.

Arribere, J.A., and Gilbert, W.V. (2013). Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. Genome Research *23*, 977–987.

Berset, C., Zurbriggen, A., Djafarzadeh, S., Altmann, M., and Trachsel, H. (2003). RNA-binding activity of translation initiation factor eIF4G1 from Saccharomyces cerevisiae. RNA *9*, 871–880.

Brar, G.A., Yassour, M., Friedman, N., Regev, A., Ingolia, N.T., and Weissman, J.S. (2012). High-Resolution View of the Yeast Meiotic Program Revealed by Ribosome Profiling. Science *335*, 552–557.

Calvo, S.E., Pagliarini, D.J., and Mootha, V.K. (2009). Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proceedings of the National Academy of Sciences *106*, 7507–7512.

Cridge, A.G., Castelli, L.M., Smirnova, J.B., Selley, J.N., Rowe, W., Hubbard, S.J., McCarthy, J.E.G., Ashe, M.P., Grant, C.M., and Pavitt, G.D. (2010). Identifying eIF4E-binding protein translationally-controlled transcripts reveals links to mRNAs bound by specific PUF proteins. Nucleic Acids Research *38*, 8039–8050.

Donnelly, N., Gorman, A.M., Gupta, S., and Samali, A. (2013). The eIF2α kinases: their structures and functions. Cell. Mol. Life Sci. *70*, 3493–3511.

Firczuk, H., Kannambath, S., Pahle, J., Claydon, A., Beynon, R., Duncan, J., Westerhoff, H., Mendes, P., and McCarthy, J.E. (2013). An in vivo control map for the

eukaryotic mRNA translation machinery. Molecular Systems Biology *9*, 635.

Haar, von der, T., and McCarthy, J.E.G. (2002). Intracellular translation initiation factor levels in Saccharomyces cerevisiae and their role in cap-complex function. Mol. Microbiol. *46*, 531–544.

Harris, T.E., and Lawrence, J.C., Jr (2003). TOR signaling. Science Signaling *2003*, re15.

Hinnebusch, A.G. (2005). Translational regulation of GCN4 and the general amino acid control of yeast. Annu. Rev. Microbiol. *59*, 407–450.

Hinnebusch, A.G., and Lorsch, J.R. (2012). The mechanism of eukaryotic translation initiation: new insights and challenges. Cold Spring Harbor Perspectives in Biology *4*.

Ibrahimo, S., Holmes, L.E.A., and Ashe, M.P. (2006). Regulation of translation initiation by the yeast eIF4E binding proteins is required for the pseudohyphal response. Yeast *23*, 1075–1088.

Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell *147*, 789–802.

Jackson, R.J. (2005). Alternative mechanisms of initiating translation of mammalian mRNAs. Biochem. Soc. Trans. *33*, 1231–1241.

Jackson, R.J., Hellen, C.U., and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. Nature Reviews Molecular Cell Biology *11*, 113–127.

Jivotovskaya, A.V., Valasek, L., Hinnebusch, A.G., and Nielsen, K.H. (2006). Eukaryotic Translation Initiation Factor 3 (eIF3) and eIF2 Can Promote mRNA Binding to 40S Subunits Independently of eIF4G in Yeast. Mol Cell Biol *26*, 1355–1372.

Kabat, D., and Chappell, M.R. (1977). Competition between globin messenger ribonucleic acids for a discriminating initiation factor. J. Biol. Chem. *252*, 2684–2690.

Kozak, M. (1980). Evaluation of the “scanning model” for initiation of protein synthesis in eucaryotes. Cell *22*, 7–8.

Kozak, M. (1989). The scanning model for translation: an update. J Cell Biol *108*, 229–241.

Lodish, H.F. (1974). Model for the regulation of mRNA translation applied to haemoglobin synthesis. Nature *251*, 385–388.

Lodish, H.F., and Jacobsen, M. (1972). Regulation of hemoglobin synthesis. Equal rates

of translation and termination of α and β-globin chains. J. Biol. Chem. *247*, 3622–3629.

Mitchell, S.F., Walker, S.E., Algire, M.A., Park, E.-H., Hinnebusch, A.G., and Lorsch, J.R. (2010). The 5'-7-Methylguanosine Cap on Eukaryotic mRNAs Serves Both to Stimulate Canonical Translation Initiation and to Block an Alternative Pathway. Molecular Cell *39*, 950–962.

Park, E.H., Walker, S.E., Zhou, F., Lee, J.M., Rajagopal, V., Lorsch, J.R., and Hinnebusch, A.G. (2013). Yeast Eukaryotic Initiation Factor 4B (eIF4B) Enhances Complex Assembly between eIF4A and eIF4G in Vivo. Journal of Biological Chemistry *288*, 2340–2354.

Park, E.-H., Walker, S.E., Lee, J.M., Rothenburg, S., Lorsch, J.R., and Hinnebusch, A.G. (2011). Multiple elements in the eIF4G1 N-terminus promote assembly of eIF4G1•PABP mRNPs in vivo. Embo J. *30*, 302–316.

Pelechano, V., Wei, W., and Steinmetz, L.M. (2013). Extensive transcriptional heterogeneity revealed by isoform profiling. Nature *497*, 127–131.

Phillips, J.A., Snyder, P.G., and Kazazian, H.H. (1977). Ratios of alpha-to beta-globin mRNA and regulation of globin synthesis in reticulocytes. Nature *269*, 442–445.

Ray, B.K., Brendler, T.G., Adya, S., Daniels-McQueen, S., Miller, J.K., Hershey, J.W., Grifo, J.A., Merrick, W.C., and Thach, R.E. (1983). Role of mRNA competition in regulating translation: further characterization of mRNA discriminatory initiation factors. Proc. Natl. Acad. Sci. U.S.A. *80*, 663–667.

Shamji, A.F., Nghiem, P., and Schreiber, S.L. (2003). Integration of growth factor and nutrient signaling: implications for cancer biology. Molecular Cell *12*, 271–280.

Spirin, A.S. (2009). How Does a Scanning Ribosomal Particle Move along the 5′-Untranslated Region of Eukaryotic mRNA? Brownian Ratchet Model. Biochemistry *48*, 10688–10692.

Thoreen, C.C., Chantranupong, L., Keys, H.R., Wang, T., Gray, N.S., and Sabatini, D.M. (2012). A unifying model for mTORC1-mediated regulation of mRNA translation. Nature *486*, 109–113.

Vaidyanathan, P.P., Zinshteyn, B., Thompson, M.K., and Gilbert, W.V. (2014). Protein kinase A regulates gene-specific translational adaptation in differentiating yeast. RNA *20*, 912–922.

Vassilenko, K.S., Alekhina, O.M., Dmitriev, S.E., Shatsky, I.N., and Spirin, A.S. (2011). Unidirectional constant rate motion of the ribosomal scanning particle during eukaryotic translation initiation. Nucleic Acids Research *39*, 5555–5567.

Walden, W.E., and Thach, R.E. (1986). Translational control of gene expression in a normal fibroblast. Characterization of a subclass of mRNAs with unusual kinetic properties. Biochemistry *25*, 2033–2041.

Walden, W.E., Godefroy-Colburn, T., and Thach, R.E. (1981). The role of mRNA competition in regulating translation. I. Demonstration of competition in vivo. J. Biol. Chem. *256*, 11739–11746.

Chapter 4: Intrinsic RNA-Binding Preferences of Eukaryotic Translation Initiation Factor eIF4G Contribute to Competitive Discrimination of Different mRNAs

Abstract

Translational control of gene expression plays essential roles in cellular stress responses and organismal development by enabling rapid, selective, and localized control of protein production. Translational regulation depends on context-dependent differences in the translational efficiencies of mRNAs, but the key mRNA features that distinguish efficiently translated mRNAs are largely unknown. Here we comprehensively determined the RNA-binding preferences of the central initiation factor eIF4G to assess whether core translation initiation factors have intrinsic sequence preferences that contribute to preferential translation of specific mRNAs. We identified a simple sequence motif – oligo-uridine – that mediates high-affinity binding to eIF4G. Oligo(U) motifs occur naturally in the transcript leaders of hundreds of yeast genes and are conserved between yeast species. Notably, mRNAs containing oligo(U) motifs resist translational repression upon depletion of eIF4G, demonstrating the impact of these motifs in vivo. Together, our data suggest a mechanism for selective translational control mediated by core initiation factors.

Introduction

Eukaryotic cells in different tissues and developmental stages require vastly different protein complements to achieve their form and function, despite having identical genomes. The contribution of translational control to regulated differences in protein production is well established in specific cases. However, despite decades of detailed characterization of the factors involved in translation, the general mechanisms governing the rate at which protein is produced from a particular species of mRNA are poorly understood.

Transcript leaders (TLs, also known as 5' UTRs) directly contact the translation initiation machinery and can strongly influence the rate of translation. Different TL sequences from yeast are sufficient to confer a thousand-fold range of translational efficiencies (TEs) both *in vivo* and in lysates (Rojas-Duran and Gilbert, 2012), but the functional elements within these TLs have not been determined. Several well-characterized TL features can cause varying degrees of translational repression including upstream open reading frames, which divert ribosomes from the functional open reading frame of an mRNA (Arribere and Gilbert, 2013; Hinnebusch, 2005; Pelechano et al., 2013); stable RNA secondary structures, which block or impede translation initiation (Ding et al., 2014; Kozak, 1990)**;** and target sites for certain RNA-binding proteins, which lead to translational repression and mRNA decay of specific sets of messages (Beckmann et al., 2005; Hentze et al., 2004). In contrast, almost nothing is known about TL features that can act as translational enhancers, though such elements

could contribute substantially to the wide range of translational efficiencies observed in eukaryotic cells (Gilbert, 2010; Shatsky et al., 2014).

Eukaryotic initiation factor 4G (eIF4G) is a prime candidate to mediate the effects of TL enhancers on mRNA-specific translational efficiencies. eIF4G is the largest subunit of eIF4F, the complex responsible for initial recognition of translation-competent mRNAs by the translation machinery (Hinnebusch and Lorsch, 2012; Jackson et al., 2010). eIF4G bridges the cap-binding activity of eIF4E, the helicase activity of eIF4A, and the poly-A binding activity of PABP/Pab1. eIF4G also contains three RNA binding domains that are capable of directly interacting with mRNA and are essential for yeast growth (Berset, 2003; Park et al., 2010), although specific functional eIF4G-mRNA interactions have not yet been characterized in yeast. In contrast, in the context of some viruses, direct RNA binding by eIF4G is both necessary and sufficient to confer efficient translation, even in the absence of a cap or a poly-A tail (Pestova et al., 1996a; 1996b). Thus, TLs that bind to eIF4G with high affinity could facilitate preferential translation of specific mRNAs under conditions of limiting eIF4F activity such as viral infection (Castelló et al., 2006; 2011), nutrient depletion (Thoreen et al., 2012), and heat shock (Cuesta et al., 2000).

Here we have comprehensively determined the RNA binding preferences of eIF4G and tested the effect of these preferences on yeast translation in vitro and in vivo. Using RNA Bind-n-Seq, a quantitative high-throughput technique to measure RNA binding affinities in vitro (Lambert et al., 2014), we show that recombinant eIF4G1 from *S. cerevisiae* preferentially binds to unstructured RNA sequences containing oligo-

uridine (U). Hundreds of yeast TLs contain oligo(U) sequences, which are evolutionarily conserved among budding yeast species and are enriched in genes with regulatory roles. We find that eIF4G-oligo(U) interactions are functional in translation. RNA containing oligo(U) is a potent inhibitor of eIF4G-dependent translation in yeast extracts, and genes with oligo(U) in their TL are resistant to eIF4G depletion *in vivo.* Thus, the intrinsic RNA-binding activity of eIF4G contributes to competitive discrimination of different mRNAs by the translational machinery

Results

eIF4G1 binds oligo(U) sequences with high affinity

To determine the RNA binding specificity of eIF4G, we used the recently-developed RNA Bind-N-Seq (RBNS) technique (Lambert et al., 2013). This competitive *in vitro* binding assay consists of mixing randomized RNA libraries with different concentrations of protein, and sequencing the bound RNA. RBNS reports directly and quantitatively on the innate RNA-binding preferences of a protein of interest, is not biased by the sequences present in a genome, and is not affected by processes that could indirectly alter protein-RNA association *in vivo*. We performed RBNS with recombinant eIF4G1, the more abundant and better characterized of two *S. cerevisiae* paralogs of eIF4G (Ghaemmaghami et al., 2003; Goyer et al., 1993), and a library of random 20mer RNA (Experimental Procedures).

Analysis of the RNA sequences bound by eIF4G1 revealed a strong enrichment for sequences containing oligo(U), and a weaker enrichment for sequences containing

five or more consecutive Gs. These enrichments showed the expected concentration dependence; they peaked at 320nM, and then decreased at higher protein concentrations where binding ceases to be competitive and all sequences are bound to a similar extent (Figure 4.1A). The enrichment for oligo(U) was comparable in magnitude to that observed previously for known binding sites of mammalian splicing factors (Figure 4.1B, 4.2, (Lambert et al., 2013)). To control for over-counting of shorter oligomer stretches within longer ones, the enrichment for homopolymers of various lengths was re-computed at each eIF4G concentration, excluding sub-polymers (Figure 4.2). The enrichment for oligo(U) increased with the length of the homopolymer stretch, indicating that eIF4G preferentially binds to longer oligo(U) stretches.

Next we confirmed the apparent preference of eIF4G1 for binding oligo(U) using electrophoretic mobility shift assays (EMSAs). Homopolymeric U and A sequences ~40 nucleotides long were generated with polynucleotide phosphorylase in the presence of a single nucleotide diphosphate to eliminate the possibility of misincorporated nucleotides (Milligan and Uhlenbeck, 1989). eIF4G1 bound to U_{40} with a 17-fold tighter affinity than A_{40} (Figure 4.1C), confirming the RNA sequence preference observed by RBNS. To see if short stretches of the homopolymers were sufficient for tight binding in the context of a longer sequence, U_{10} or G_{10} were embedded into a sequence of 25 CA repeats and tested by EMSA with eIF4G1. G_{10} reduced the K_d two-fold compared to the poly(CA) control, and addition of U_{10} resulted in a K_d comparable to the U_{40} homopolymer (Figure 4.1D), recapitulating the results from RBNS. We have thus tested the eIF4G affinity of

all 87,380 possible sequences up to length 8 and determined that oligo(U) is the preferred binding site for eIF4G1.

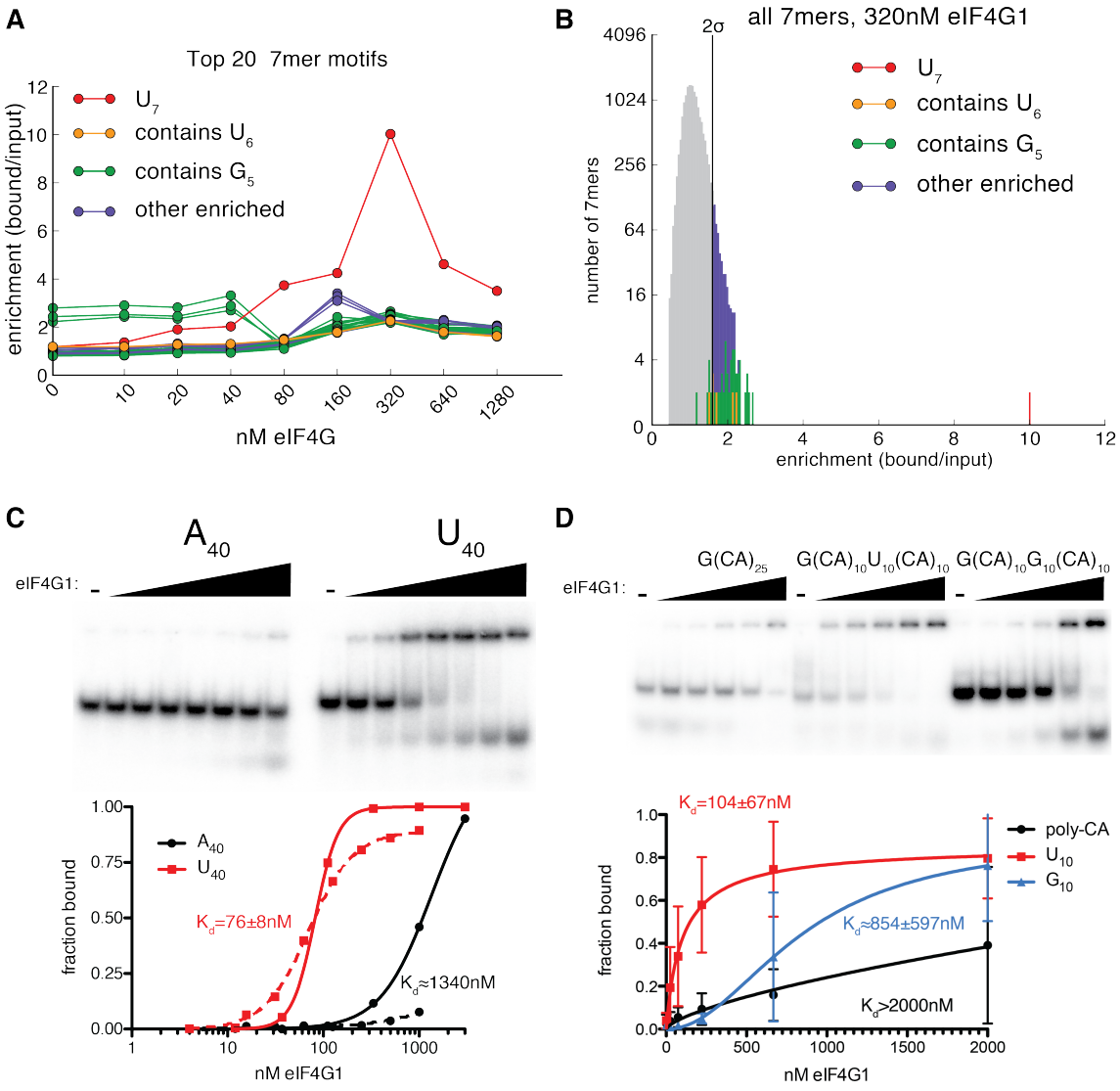

Figure 4.1: eIF4G1 preferentially binds to oligo(U) sequences
(A) RNA bind-n-seq enrichment scores at all concentrations, for the 20 7nt sequences that were most enriched at 320nM eIF4G1.
(B) Log-scaled histogram of RNA bind-n-seq enrichment values, at 320nM eIF4G1, for all 7nt sequences. Vertical line indicates 2 standard deviations from the mean of the distribution.
(C) Electrophoretic mobility shift assay (EMSA) to measure binding affinity of eIF4G1 homopolymer sequences. (Top) Representative gel. Binding reactions were performed

with 2-fold dilutions of eIF4G1 ranging from 15-1000nM. (Bottom) Quantification of homopolymer binding experiments. K_ds are mean±SD from 2 independent replicates. A_{40} data did not yield a reliable fit in one replicate.
(D) Electrophoretic mobility shift assay (EMSA) on model TL sequences. (Top) Representative gel. Binding reactions were performed with 3-fold dilutions of eIF4G1 ranging from 25-2000nM. (Bottom) Quantification of binding experiments. Each data point is mean±SD from 2 (G_{10}) or 3 (CA, U_{10}) independent replicates.

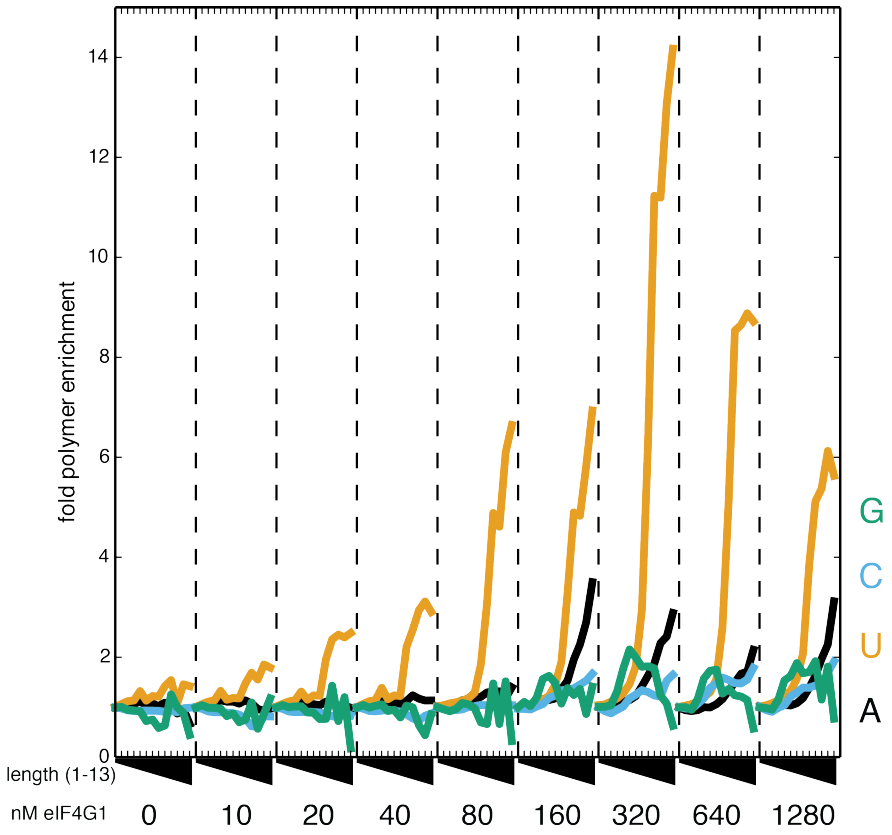

Figure 4.2: eIF4G1's affinity for oligo-(U) increases with oligomer length.
For each concentration of eIF4GI, the enrichment relative to the input library for each length of oligo-U (1-13nt) is shown. For the purposes of this analysis, homopolymers that are part of a larger homopolymer sequence were excluded.

eIF4G1 preferentially binds unstructured oligo(U) sequences

The known high-affinity binding sites for mammalian eIF4G in viral mRNAs are highly structured (Jackson et al., 2010; Pestova et al., 1996b), suggesting that yeast eIF4G1 may have structural preferences as well. We determined the structural contexts of the sequence motifs bound by eIF4G1 by computing single-nucleotide pairing

probabilities for each 20mer containing a U_7 sequence. The probabilities around these motifs were averaged into pairing probability profiles for each concentration, and normalized by the equivalent profile for the input sequences to determine the pairing preferences of the bound motifs (Figure 4.3A). The sequences bound by eIF4G show a sharp drop in pairing probability at the bound oligo(U) motif (Figure 4.3B). Furthermore, the preference for single-strandedness showed the same concentration dependence as the enrichment for oligo(U), with the strongest enrichment for oligo(U) coinciding with the strongest preference for unpaired sequences at 320nM (Compare Figure 4.3B with Figure 4.1A). Thus eIF4G1 preferentially binds oligo(U) in unstructured contexts.

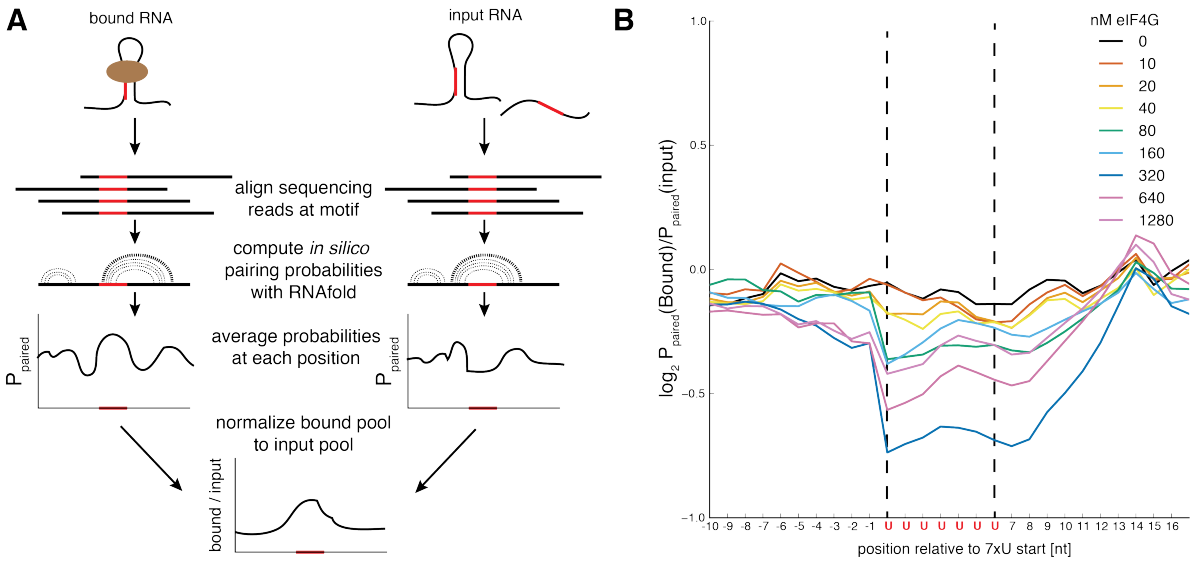

Figure 4.3: eIF4G1 binds oligo(U) in unstructured contexts
(A) Overview of RNA structure analysis. All instances of a given motif in the bound and input RNA pools are aligned, and folded *in silico* to generate a pairing probability for each position. In this cartoon example, darker arcs indicate larger pairing probabilities. These probabilities are averaged across instances of the motif to generate a pairing probability profile, indicating how likely each base in and the around the motif is to be

paired. The bound profile is then normalized to the input profile, to determine the differences in structural propensity between the bound and input RNA populations. (B) Pairing probabilities for oligo(U) and surrounding sequence for each eIF4G1 concentration, normalized to the input sequences. The sequences are aligned at the 5' U of the homopolymer, but many extend more than 7 nucleotides, explaining the sharp drop in pairing probability at the 5' end of the sequence, and the gradual rise at the 3' end.

Oligo(U) sequences are conserved and enriched in genes with regulatory functions

To investigate the possible physiological role of oligo(U) sequences, we searched the *S. cerevisiae* transcriptome for TLs with continuous stretches of U's. 456 genes were identified that contain oligo(U) motifs ≥7 nucleotides within the longest annotated TL (Xu et al., 2009) (Table 4.1) These oligo(U) elements are significantly conserved above the background conservation for the TLs that contain them (Figure 4.4A), and 47 U_7 motifs are completely conserved between five yeast species that shared their last common ancestor ~10 million years ago (Budovskaya et al., 2005) (Table 4.2). Remarkably, the conservation of oligo(U) is similar to that observed for uAUG elements (Figure 4.4A, 4.5), which have well characterized molecular and physiological functions in translational control (Arribere and Gilbert, 2013; Sonenberg and Hinnebusch, 2009). This high degree of conservation of oligo(U) sequences in TLs implies that these elements play a physiologically relevant role subject to selective pressure.

The previously described mechanisms of eIF4G-mediated translational control affect specific functional groups of messages, such as those that encode ribosomal proteins (Thoreen et al., 2012) or mediate invasive growth (Gilbert et al., 2007). To see if oligo(U) elements also regulate specific cellular pathways, we performed gene

ontology analysis (Balakrishnan et al., 2012) on genes with TLs containing U_7. We found a strong enrichment for genes with regulatory functions and for genes whose protein products localize to the cell periphery or cellular bud (Figure 4.4B). Several of the genes with conserved U_7 motifs in their TL are involved in invasive or pseudohyphal growth (Table 4.2) (Ryan et al., 2012; Shively et al., 2013), yeast differentiation programs in which regulation of eIF4G activity towards specific sets of messages has previously been shown to be essential (Gilbert et al., 2007; Ibrahimo et al., 2006). Together, these results raise the possibility that yeast cells utilize mRNA affinity for eIF4G to tune the translation of genes necessary for response to nutrient deprivation.

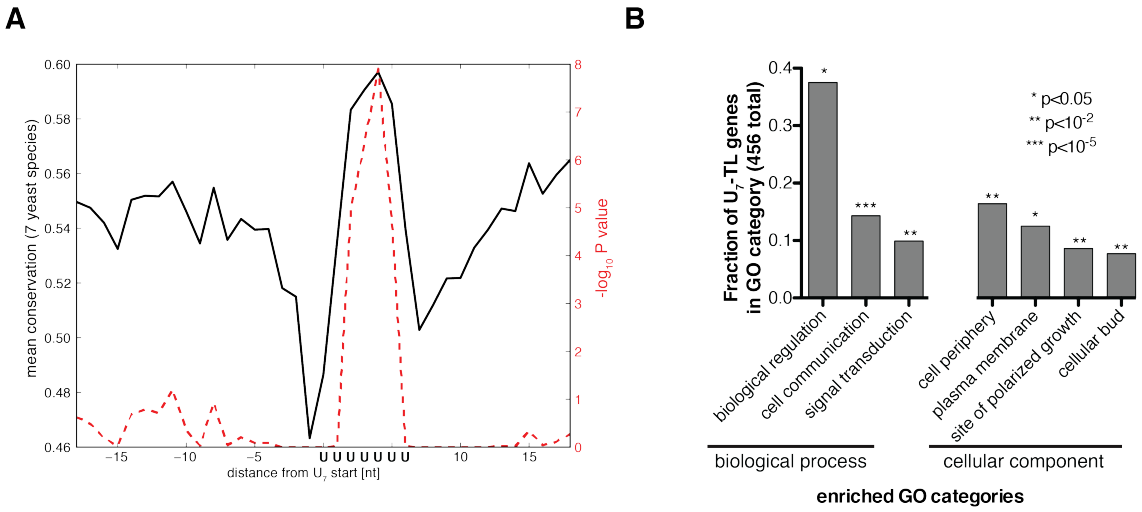

Figure 4.4: Oligo(U) sequences in TLs are conserved, and enriched for regulatory functions
(A) Average sequence conservation score between 7 yeast species for all U_7 sequences in yeast TLs.
(B) Fraction of genes with U_7 in their TL that belong to each significantly enriched gene ontology category

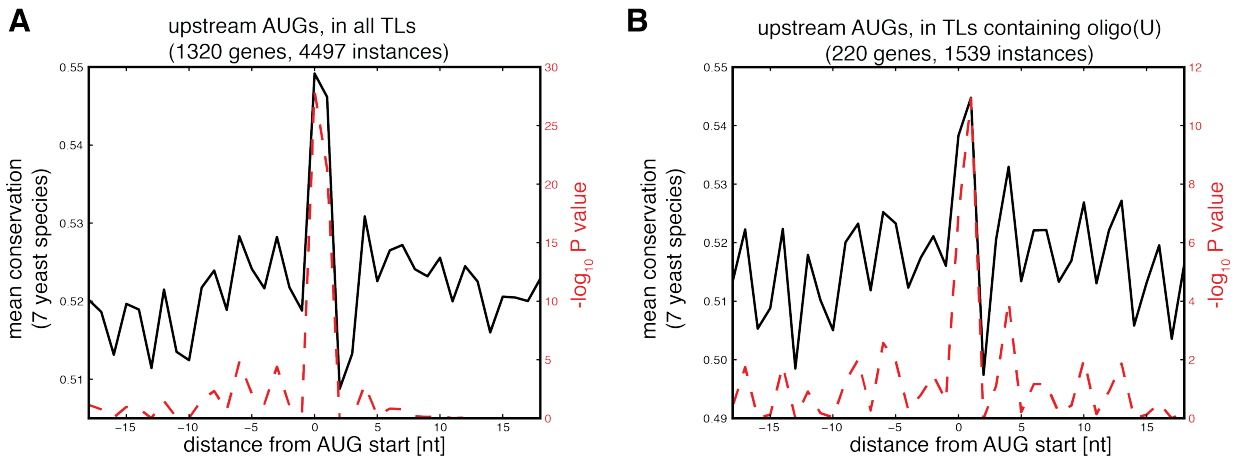

Figure 4.5: Conservation of oligo(U) is comparable to conservation of upstream AUG sequences.
(A) Average conservation for all upstream AUG sequences
(B) Average conservation for upstream AUG sequences in genes containing ≥7xU in their TLs.

Oligo(U) motifs affect eIF4G-dependent translation in vitro and in vivo

To further investigate the physiological relevance of eIF4G1's affinity for oligo(U), we tested the ability of U_{10}-containing RNA to inhibit eIF4G-dependent translation in lysate (Figure 4.6A). Yeast translation extracts were programmed with uncapped, polyadenylated firefly luciferase mRNAs containing TLs from different genes: PRE2, a well-translated cellular mRNA (Rojas-Duran and Gilbert, 2012), YMR181C, which contains an eIF4G-dependent cellular internal ribosome entry site (IRES) (Gilbert et al., 2007), and the Cricket Paralysis Virus (CrPV) IRES, which confers factor-independent translation initiation. Addition of poly(CA) RNA had little effect on the translation of any mRNA, while high concentrations of U_{10} specifically inhibited the eIF4G-dependent PRE2 and YMR181C transcripts, with no effect on CrPV IRES-dependent translation (Figure 4.6B). This result indicates that oligo(U) sequences are capable of sequestering translation-competent eIF4G from yeast lysates.

Inhibition of translation by oligo(U) in trans suggests that eIF4G in the context of the eIF4F complex also has a high affinity for oligo(U). Because the affinity of eIF4G for eIF4E is ~15nM (Mitchell et al., 2010), and the concentrations of both proteins is >> 15nM in yeast cells (Haar and McCarthy, 2002), eIF4G and eIF4E are likely to be in a complex in our concentrated translation extracts. Thus RNA competition for eIF4F can affect translation *in vitro*.

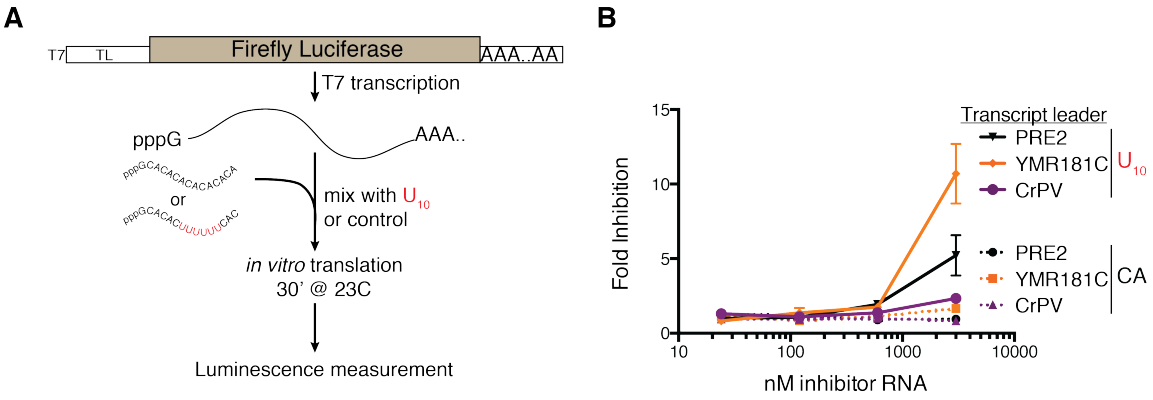

Figure 4.6: Oligo-U inhibits eIF4G-dependent translation in trans
(A) Overview of reporter constructs and translation assay.
(B) Fold inhibition (-inhibitor/+inhibitor luminescence) of mRNAs with three different TLs by the poly(CA) control ($G(CA)_{25}$) or U_{10} ($G(CA)_{10}U_{10}(CA)_{10}$). The inhibitor RNAs are the same as those in Figure 4.1D. Error bars are the standard deviation of 3 technical replicates.

To determine the impact of oligo(U) sequences on translational efficiency *in vivo,* we first compared published TE values (Park et al., 2011) for genes containing oligo(U) in their TLs or 3'UTRs to the rest of the genome. Genes with TL oligo(U) motifs did not show an increase in translation in steady-state rich media conditions, and in fact had slightly reduced TEs compared to all other genes (Figure 4.7). The lack of any observed translational advantage for messages containing oligo(U) sequences under these growth conditions is consistent with the high cellular concentration of eIF4G (~1uM,

Experimental Procedures), at which eIF4G1 does not effectively discriminate oligo(U) from other sequences (Figure 4.1A).

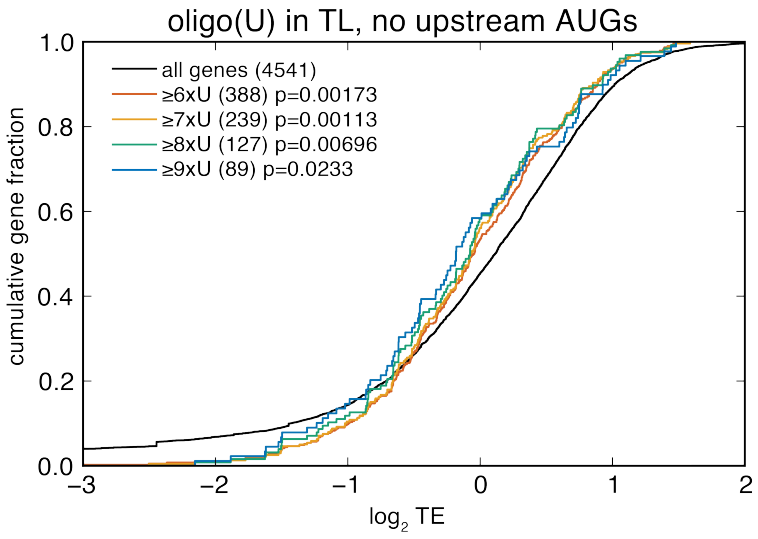

Figure 4.7: Oligo(U) sequences do not stimulate basal levels of translation in rich media.
Cumulative distributions of translational efficiencies for all genes upstream AUGs, compared to subsets of genes with different lengths of oligo-U in their TL. TE data from (Subtelny et al., 2014)(GEO accession GSE53313).

We next considered whether oligo(U) sequences might confer improved translation in conditions where eIF4G is limiting. A previous study (Park et al., 2011) examined the effect of eIF4G depletion on TE genome-wide using polysome microarrays. We re-analyzed this dataset, focusing on genes with homopolymer sequences in their TLs. Genes with oligo(U), but not oligo(A), in their TL show higher TE upon eIF4G depletion compared to all other genes (Figure 4.8A, B). The magnitude of this effect is increased if the gene set is limited to genes without uAUGs (Figure 4.8C), or to conserved oligo(U) sequences (Figure 4.8D), and persists when the gene sets are controlled for length and AU content biases (Figure 4.9). We interpret the relative

increase in translation of oligo(U)-containing messages as a maintenance of previous translational activity in the context of a global reduction (Park et al., 2011). Together, these results suggest that when eIF4G levels are limiting, translational efficiencies are partially determined by competition between mRNAs for eIF4G.

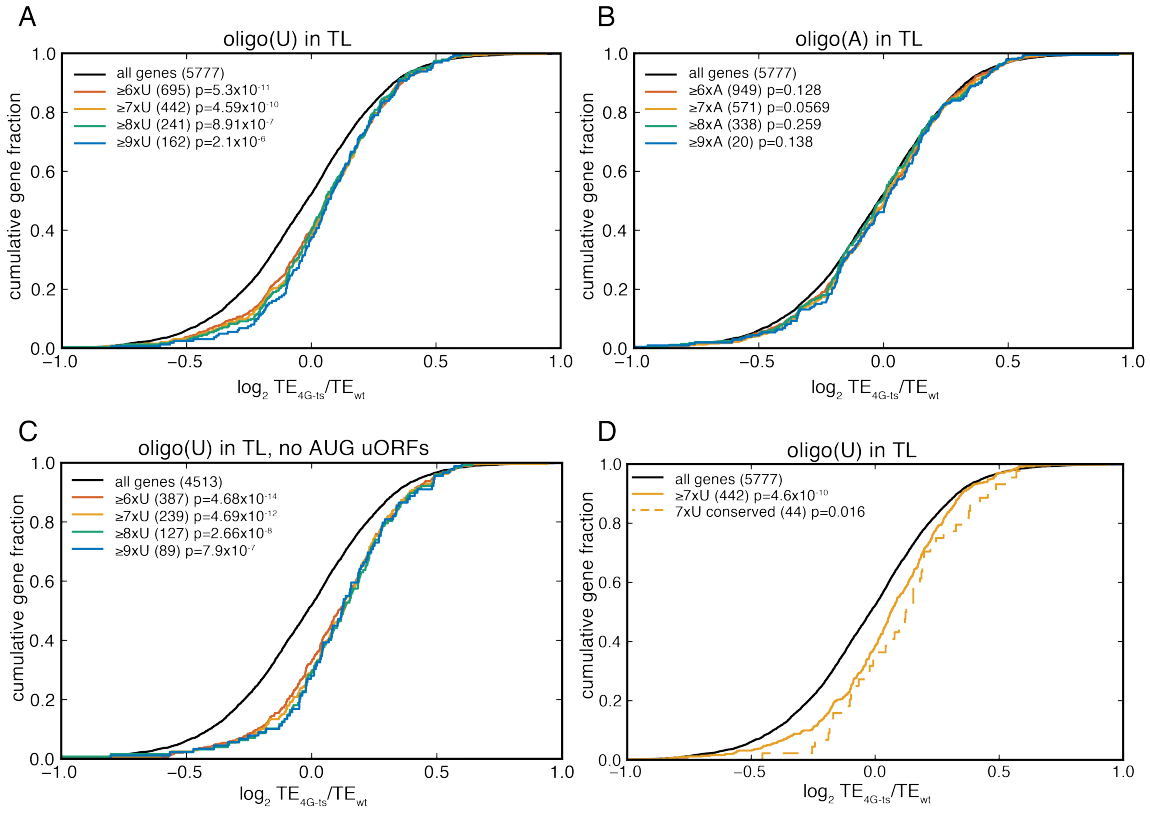

Figure 4.8: Translation of mRNAs with oligo-U in their TL is resistant to eIF4G depletion

(A) TE change upon eIF4G depletion for all genes, and genes containing at least 6 continuous U's in their longest annotated TL.
(B) TE change upon eIF4G depletion for all genes, and genes containing at least 6 continuous A's in their longest annotated TL.
(C) Same as in A, but limited to genes without an AUG triplet in their TL.
(D) TE change upon eIF4G depletion for all genes, and genes containing at least 7 continuous U's in their longest annotated TL, and genes with a conserved 7xU in their TL.
The number of genes for each set is indicated in parentheses, followed by a KS test p value against the "all genes" set for that plot. TE values are from (Park et al., 2011).

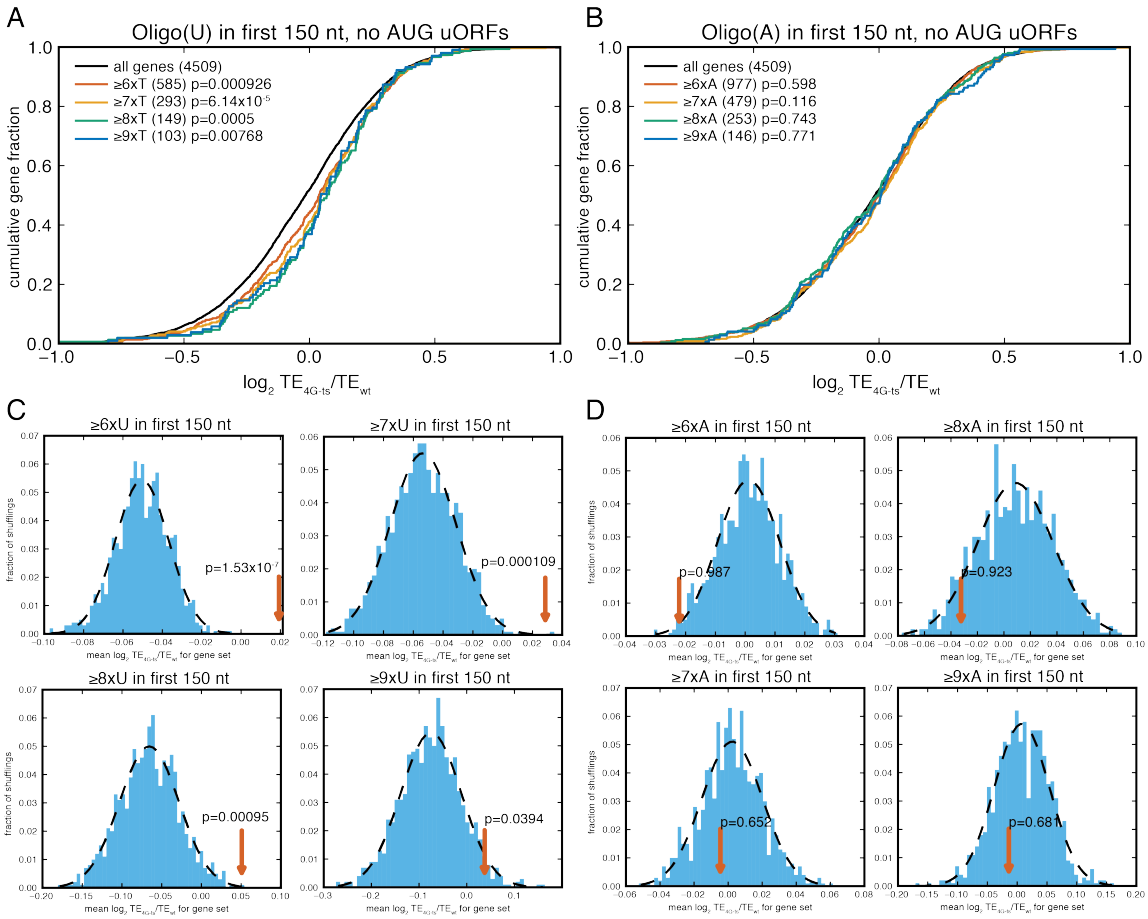

Figure 4.9: Transcripts containing oligo-U at their 5' end are resistant to eIF4G depletion.

(A) TE change upon eIF4G depletion for all genes without upstream AUGs compared to transcripts containing at least 6 continuous Us in their first 150 nucleotides. Number of genes in each set is in parentheses, followed by the KS test p-value for that set against "all genes".

(B) Same as (A), except with oligo-A sequences

(C, D) Sequence randomization controls for A and B, respectively. To control for TL length and AU-richness bias, distributions were generated by shuffling the first 150 nucleotides of all yeast transcripts 1000 times, and selecting all transcripts with a homopolymer of at least the indicated length. Histogram of the mean TE change for all randomized sets is indicated. The mean of the real set of genes is indicated by the arrow. The p-values were computed by fitting a normal distribution (dotted line) to the histogram, and integrating it from the real mean TE to infinity. TE data from (Park et al., 2011)

Discussion

The RNA binding activity of eIF4G is essential for yeast growth (Berset, 2003), yet the role of this activity in translation initiation is not currently understood. Although previous work suggested that eIF4G binds most mRNAs non-specifically (Berset, 2003; Goyer et al., 1993), certain viral mRNAs rely on high-affinity binding to eIF4G for their translation (Pestova et al., 1996b), raising the possibility that specific cellular mRNAs might also exploit the intrinsic RNA-binding preferences of eIF4G for their translation. Here we have systematically determined the RNA-binding preferences of eIF4G1 in vitro and discovered a strong affinity preference of eIF4G1 for oligo(U) sequences, which are present in the transcript leaders of hundreds of yeast mRNAs. Finally, we show that oligo(U) motifs confer preferential translation in vivo under conditions of limiting eIF4G activity, thus demonstrating the physiological significance of these eIF4G binding elements.

The RNA sequence preferences of eIF4G could provide the basis for competition between mRNAs for limiting components of the translation machinery and a possible mode for regulating translation rate, both at steady state and during translational responses to cellular stimuli. Oligo(U) motifs did not detectably affect basal translation rates in rich media, but they conferred preferential translation upon eIF4G depletion. There are several potential explanations for these effects. First, there may be sufficient eIF4G in cells growing in rich media to eliminate competition between mRNAs. There is also limited evidence suggesting that steady-state translation, especially on heavy polysomes, is reliant on intra-polysomal ribosome recycling rather than de novo

ribosome recruitment (Adamson et al., 1969; Kopeina et al., 2008), which would limit the effect of eIF4G on steady-state ribosome density to the pool of messages that have not matured into polysomes. When eIF4G is depleted, there is more competition for the protein, and the resulting reduction in polysomes (Park et al., 2011) would increase the relative ratio of de novo ribosome recruitment to recycling, leading to an increased contribution of mRNA competition to translational efficiency.

Although the conditions used here to deplete eIF4G were genetic rather than environmental, many cellular stress responses involve global translational changes that facilitate reprogramming of the translatome faster or more efficiently than is possible by a transcriptional response alone (Ashe et al., 2000; Hinnebusch, 2005; Shalgi et al., 2012; Vaidyanathan et al., 2014). Notably, several stress responses require mRNA-specific modulation of eIF4G activity, either by eIF4G recruitment (Gilbert et al., 2007), or through inhibition of eIF4G-dependent translation by eIF4E-binding proteins (Cridge et al., 2010; Ibrahimo et al., 2006; Thoreen et al., 2012). Thus, the translational capacity of mRNAs may be more generally controlled by their direct affinity for eIF4G during the response to or recovery from cellular stressors. The evolutionary conservation of oligo(U) sequences within TLs, as well as the cellular roles of the genes containing them, suggest that differential mRNA recognition by eIF4G plays a role in responses to cellular stimuli.

The portions of eIF4G1 responsible for specific binding to unstructured oligo(U) motifs are currently unknown. Although eIF4G was initially assigned an RNA recognition motif based on sequence (Goyer et al., 1993), this domain was found not to bind RNA

(Berset, 2003), and the empirically determined RNA binding domains of eIF4G1 do not resemble any well-defined RNA binding motifs at the sequence level (Burd and Dreyfuss, 1994). However, two of these domains are arginine-serine rich (RS), resembling the unstructured RS domains commonly found in splicing factors. Although splicing factor RS domains are thought to mediate protein-protein interactions, not RNA binding (Graveley, 2000), in eIF4G, these domains appear to mediate both RNA and protein interactions (Singh et al., 2012). It will be interesting to determine which of the domains of eIF4G are required to confer RNA binding specificity, and whether this specificity is essential for yeast growth.

The factors that influence the rate of translation initiation are just beginning to be described, and, until now, have not included the intrinsic RNA-binding preferences of the core initiation factors. Several other translation initiation factors, including eIF4A, eIF4B, and multiple subunits of eIF3, have RNA binding activity and may also have differential affinity for mRNAs. Our results with eIF4G1 suggest it will be important to determine the RNA binding preferences of these other initiation factors and assess their contributions to the translation of specific mRNAs. With a complete affinity profile of the translation initiation machinery, we can begin to quantitatively predict the protein output of a transcriptome from the sequence of its mRNAs.

Experimental Procedures

Purification of Yeast eIF4G1

Full-length eIF4G fused with N-terminal GST and C-terminal 6xHis tags was purified as described (Mitchell et al., 2010), with the following modifications. [Now only describe things you did differently.] The pGEX4T1-GST-eIF4G1-His6, or pGEX4T1-GST-eIF4G2-His6 plasmid was transformed into BL21 CodonPlus-RIL competent cells (agilent) and grown in 4-6 liters of super broth (3.2% Tryptone, 2.0% yeast extract, 0.5% NaCl, pH 7.5.) to an OD600 of ~2.5 at 37°C. After 4 hours of induction by 1mM IPTG at 30°C, cultures were pelleted, resuspended in lysis buffer (20mM HEPES-KOH, pH 7.4, 100mM KCl, 10% glycerol, 0.1% Triton x100, 20mM Imidazole, 10mM 2-mercaptoethanol, 0.1mM PMSF, 1µg/ml pepstatin, aprotinin, leupeptin, 1 Roche EDTA-free protease inhibitor tab per 50ml), and frozen by dripping into liquid nitrogen. Cells were lysed by cryogenic ball milling in a Retsch Cryomill. Lysates were treated with DNase I (New England Biolabs), centrifuged to remove debris, brought to 0.5M KCl, and then passed through 3-micron (Sterlitech), then 0.2 micron (VWR) filters. The protein was enriched from lysate on a HisTrap nickel column (GE Healthcare), eluted with 250mM imidazole, and exchanged into heparin buffer A (150mM NaCl, phosphate buffer pH7.3, 10% glycerol, 2mM DTT). The major fractions were brought to 2mM $CaCl_2$ and 1U/µl micrococcal nuclease (New England Biolabs), incubated for 30 minutes at room temperature, and then brought to 5mM EGTA. The solution was filtered, passed over a HiTrap Heparin HP column (GE Healthcare), and eluted with a gradient of 0.5-2M NaCl.

The eluate was loaded onto a GSTrap HP column (GE Healthcare), and eluted with 10mM glutathione in 50mM Tris-Cl, 10% glycerol, pH7.5. The eluate was concentrated with Amicon Ultra 3.5kDa cutoff spin columns (Millipore) and passed over an S200 sizing column into storage buffer (20mM HEPES, pH 7.4, 100mM KCl, 10% glycerol, 2mM DTT). The protein eluted from the column in two distinct peaks. The early-eluting (higher molecular weight) peak had a higher A260/A280 ratio, suggesting that it consists of RNA-bound protein (Figure 4.10A). When run on SDS-PAGE gels, both peaks consisted of mostly full-length protein, with some proteolysis products (Figures 4.10B, 4.10C). The RNA-free peak was spin-column concentrated and used for subsequent experiments.

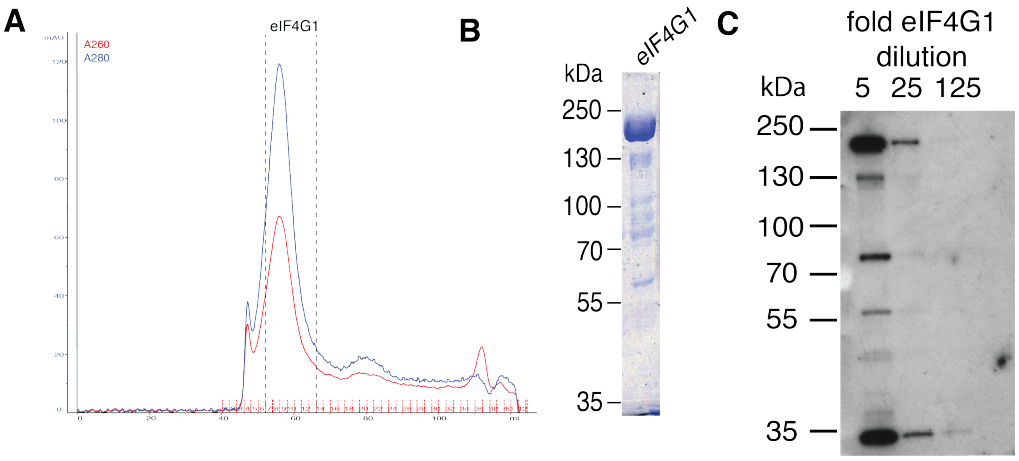

Figure 4.10: Purification of eIF4G1
(A) A260 and A280 traces from size exclusion chromatography of purified eIF4G1 (left) and eIF4G2 (right). The dotted lines indicate the fractions that were collected and used for subsequent experiments. The RNA-containing complex is visible at ~50ml elution volume.
(B) Coomassie stained 7% SDS-PAGE gels of pooled eIF4G1 fractions from (A).
(C) Western blot of recombinant eIF4G1 using a polyclonal eIF4G1 antibody (Clarkson et al., 2010). Most bands visible in (B) are reactive to the antibody, indicating that the extra bands are eIF4G1 cleavage products.

RNA Library Synthesis

Randomized RNA libraries consisted of 20 nucleotides of randomized RNA sequence, followed by an adaptor sequence used for priming reverse transcription and Illumina library preparation. Libraries were ordered from Integrated DNA Technologies as single-stranded DNA oligos with a 5' T7 promoter sequence (oBZ52), and gel purified. A complementary oligo (oBZ49) was annealed to the T7 promoter sequence, and transcription carried out in as described in (Rio et al., 2011). The transcription products were filtered through G-50 size exclusion spin columns (GE healthcare), precipitated, and gel purified.

RNA Bind-n-Seq

For randomized pools, 1µM RNA pool was incubated with concentrations of eIF4G ranging from 0-1280nM for 30' @ 22°C in a 50ul volume of PBS, 0.1% TWEEN, 1mg/ml BSA, 3mM $MgCl_2$, 1mM DTT. The solution was then incubated with 50µl of MagneGST glutathione particle slurry (Promega, pelleted and washed 5x with PBST) for 10' @ 22C. The supernatant was removed and washed once with 1ml PBST, then eluted by adding 100ul elution buffer (50mM Tris 7.5, 1mM EDTA, 0.5% SDS) and heating at 70°C for 10 minutes. Eluates were phenol-chloroform extracted and ethanol precipitated. RNA was reverse transcribed at 48°C using SuperScript III (life technologies) and primer oTC225 (Carlile et al., 2014), and extended products were gel-purified. Circularization and PCR was performed as previously described in (Carlile

et al., 2014), with the addition of 1M betaine to the PCR reaction. Libraries were sequenced with 40-50nt single end reads on an Illumina HiSeq 2000.

Sequence Enrichment Analysis

Sequencing data were analyzed using the code and algorithms developed in (Lambert et al., 2013), as well as custom Python scripts developed in-house, unless otherwise indicated.

RNA Folding Analysis

The RNA pairing probability analysis was modified from (Lambert et al., 2014). All reads in a given sequencing library containing a sequence of interest (such as U_7 for Figure 2B) were folded by calling the partition fold function of the Vienna RNAfold package (Lorenz et al., 2011) from a custom Perl script. Calling RNA::pf_fold($seq) and RNA::get_pr($k,$m) computes the pairing probability between positions k and m in the sequence, where $seq is the full RNA sequence, including the common 3' adaptor. The total pairing probability for a given position in the read is the sum of its pairing probabilities with all other positions in the RNA sequence. Average probability profiles were generated by averaging the pairing probabilities for a given position within the motif or surrounding sequence across all reads, and normalized by the average pairing probability at the same motif position in the input library.

Conservation Analysis

Transcript leader and 3'UTR annotations were defined based on (Xu et al., 2009). 7-yeast alignment data containing whole-genome alignments between *S. cerevisiae (sacCer3), S. paradoxus, S. Mikatae, S. kudriavzevii, S. Bayanusm, S. castelli,* and *S. kluyveri* was downloaded from the UCSC genome browser. For meta-plots (Figure 4.4A, 4.5), conservation score was defined as the fraction of 7 yeast species that shared the *S. cerevisiae* sequence at the given position, and averaged over all instances of the motif. P values were calculated by choosing a set of random positions in the same group of transcript leaders 1000 times, computing the z-score for the real average value against the averages of this distribution, and integrating the normal distribution from that Z value to infinity. To determine the extent of conservation for the full oligomer, a homopolymer was considered conserved in another species if the same homopolymer was present at the aligned position, plus an additional 5 nucleotides on either side, to account for insertions and deletions that commonly occur during alignment of homopolymer sequences.

Estimation of Cellular mRNA and eIF4G Content

In order to estimate the number of mRNA molecules in a rapidly-dividing yeast cell, we re-scaled existing RNA-seq RPKM data from (Subtelny et al., 2014), prepared by rRNA depletion, (GEO accession GSE53313) into per-cell mRNA counts using a linear regression against single-molecule FISH counts for 13 mRNAs (DOA1, KAP104, MDN1, POL1, PRE3, PRE7, PRP8, PUP1, RPB1, RPB3, TAF12, TAF5, TAF6) from

(Gandhi et al., 2010; Zenklusen et al., 2008). PDR5 was also available, but omitted as an outlier. A linear fit of Y=(0.02538±0.00728)X+(1.228±1.15156) was obtained. Rescaling and summing all RNA-seq RPKM values results in an mRNA content of ~ 26,400±11,400 (ranges indicate 95% confidence intervals). Total eIF4G content is estimated at 24,000 molecules per cell, based on the low-throughput value of 17,500 molecules per cell for eIF4G1 from (Haar and McCarthy, 2002), and assuming that eIF4G2 is present at 35% the level of eIF4G1 based on the high-throughput measurements of (Ghaemmaghami et al., 2003). Concentrations were computed assuming a cell volume of 42 femtoliters (Jorgensen et al., 2002).

In Vitro Translation

Yeast translation extracts and uncapped luciferase mRNA were made exactly as described in (Rojas-Duran and Gilbert, 2012). PRE2, CrPV and YMR181C RNAs were transcribed from plasmids pWG476, 502 and 499 respectively. Poly-CA TLs were ordered as complementary oligos with ends matching HindIII and NcoI restriction sites (oBZ118-121), annealed, treated with polynucleotide kinase (New England Biolabs) and cloned into pWG505(Rojas-Duran and Gilbert, 2012) at the HindIII and NcoI sites.

Electrophoretic Mobility Shift Assays

Poly-A and poly-U RNA substrates were generated by incubating 1 unit of polynucleotide phosphorylase (Sigma) in 1 ml of 10mM UDP/ADP, 50mM Tris pH9.0, 2mM $MgCl_2$. Reactions were concentrated with the Zymo DNA clean and Concentrator

25 kit, and fragmented by incubation in 10mM $ZnCl_2$ for 4 minutes at 94°C. Fragmentation reactions were stopped by addition of EDTA to 40mM, and ethanol precipitated. Fragmented RNA was run on a urea-PAGE gel to size-select the desired size. Poly-CA sequences (and those with homopolymer inserts) were created by T7 transcription (same as RNA libraries above) from oligos oBZ70-72. T7-transcribed RNA was dephosphorylated with calf intestinal phosphatase (New England Biolabs) and cleaned up with oligo clean & concentrator kit (Zymo Research). RNA was end-labeled by reaction of T4 polynucleotide kinase (New England Biolabs) in the presence of γ-^{32}P ATP (Perkin Elmer), and separated from unincorporated ATP and phosphate with G-25 or G-50 spin columns (GE healthcare).

Approximately 60 femtomoles of radiolabeled RNA was equilibrated at 22°C in 5ul 2xTHEM (35mM Tris, 57mM HEPES, 0.1mM EDTA, 2.5mM $MgCl_2$), 4mM DTT, 0.1-1mg/ml BSA, 10 units SUPERase•in (Life Technologies). eIF4G1 was diluted to 2x the assay concentration in enzyme storage buffer at 22°C. 5µl of protein mix was added to the RNA and equilibrated for 30' at 22°C. Samples were loaded onto a 1xTHEM 7% acrylamide MIDI gel (pre-run for 1hr at 140v at 4°C) and run for 40-120' at 140v. In some experiments, 2ul loading dye (1x THEM, 60% sucrose, xylene cyanol, bromophenol blue) was added before loading, and did not affect results. Gels were wrapped in plastic, and exposed to phosphor screens for 5'-50', and scanned on a Typhoon phosphorimager. Bands were quantified using ImageQuant TL software (GE Healthcare). Fraction bound was defined as the ratio of the upper band intensity to the total intensity of upper and lower bands. A γ-^{32}P ATP-only control was run on each gel

to differentiate full-length RNA from degradation products or unincorporated radioactive ATP or phosphate. Plots were generated in Prism 5 (GraphPad software) and fit to the hill equation with K_d, hill coefficient, and maximum fraction bound as free parameters.

Translational Efficiency Analyses

The TE changes upon eIF4G depletion were computed from supplementary table 1 of (Park et al., 2011), by taking the median of three replicate TE values (heavy polysome / total) for WT and 4Gts. Basal TEs in YPD are from GEO accession GSE53313, with no further processing. Transcripts were defined based on the TL lengths from (Xu et al., 2009). Sequences were shuffled with the uShuffle package for Python (Jiang et al., 2008), preserving trinucleotide content.

Acknowledgements

We thank Sarah Walker and John Lorsch for providing protocols, plasmids and advice for eIF4G purification; Nicole Lambert, Alex Robertson and Chris Burge for providing RBNS methods and analysis algorithms prior to publication; Oksana Sergeeva and the King lab for sharing protein purification expertise and equipment; Maria Rojas-Duran for preparing sequencing libraries, and David Bartel for discussions. Sequencing was performed by the MIT BioMicro Center. This work was supported by grants from the National Institutes of Health (GM094303, GM081399) to WVG. BZ was supported by the Henry and Francis Keany Rickard fund fellowship and by the NIH Pre-Doctoral Training Grant T32GM007287.

References

Adamson, S.D., Howard, G.A., and Herbert, E. (1969). The ribosome cycle in a reconstituted cell-free system from reticulocytes. Cold Spring Harb. Symp. Quant. Biol. *34*, 547–554.

Arribere, J.A., and Gilbert, W.V. (2013). Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. Genome Research *23*, 977–987.

Ashe, M.P., De Long, S.K., and Sachs, A.B. (2000). Glucose depletion rapidly inhibits translation initiation in yeast. Mol. Biol. Cell *11*, 833–848.

Balakrishnan, R., Park, J., Karra, K., Hitz, B.C., Binkley, G., Hong, E.L., Sullivan, J., Micklem, G., and Cherry, J.M. (2012). YeastMine--an integrated data warehouse for Saccharomyces cerevisiae data as a multipurpose tool-kit. Database (Oxford) *2012*, bar062.

Beckmann, K., Grskovic, M., Gebauer, F., and Hentze, M.W. (2005). A Dual Inhibitory Mechanism Restricts msl-2 mRNA Translation for Dosage Compensation in Drosophila. Cell *122*, 529–540.

Berset, C., Zurbriggen, A., Djafarzadeh, S., Altmann, M., and Trachsel, H. (2003). RNA-binding activity of translation initiation factor eIF4G1 from Saccharomyces cerevisiae. RNA *9*, 871–880.

Budovskaya, Y.V., Stephan, J.S., Deminoff, S.J., and Herman, P.K. (2005). An evolutionary proteomics approach identifies substrates of the cAMP-dependent protein kinase. Proc. Natl. Acad. Sci. U.S.A. *102*, 13933–13938.

Burd, C.G., and Dreyfuss, G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. Science *265*, 615–621.

Carlile, T.M., Rojas-Duran, M.F., Zinshteyn, B., Shin, H., Bartoli, K.M., and Gilbert, W.V. (2014). Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. Nature 515, 143-146.

Castelló, A., Álvarez, E., and Carrasco, L. (2006). Differential Cleavage of eIF4GI and eIF4GII in Mammalian Cells EFFECTS ON TRANSLATION. J. Biol. Chem. *281*, 33206–33216.

Castelló, A., Álvarez, E., and Carrasco, L. (2011). The Multifaceted Poliovirus 2A Protease: Regulation of Gene Expression by Picornavirus Proteases. Journal of Biomedicine and Biotechnology *2011*, 1–23.

Clarkson, B.K., Gilbert, W.V., and Doudna, J.A. (2010). Functional overlap between

eIF4G isoforms in Saccharomyces cerevisiae. PLoS ONE *5*, e9114.

Cridge, A.G., Castelli, L.M., Smirnova, J.B., Selley, J.N., Rowe, W., Hubbard, S.J., McCarthy, J.E.G., Ashe, M.P., Grant, C.M., and Pavitt, G.D. (2010). Identifying eIF4E-binding protein translationally-controlled transcripts reveals links to mRNAs bound by specific PUF proteins. Nucleic Acids Research *38*, 8039–8050.

Cuesta, R., Laroia, G., and Schneider, R.J. (2000). Chaperone hsp27 inhibits translation during heat shock by binding eIF4G and facilitating dissociation of cap-initiation complexes. Genes & Development *14*, 1460–1470.

Ding, Y., Tang, Y., Kwok, C.K., Zhang, Y., Bevilacqua, P.C., and Assmann, S.M. (2014). In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. Nature *505*, 696–700.

Gandhi, S.J., Zenklusen, D., Lionnet, T., and Singer, R.H. (2010). Transcription of functionally related constitutive genes is not coordinated. Nature Structural & Molecular Biology *18*, 27–34.

Ghaemmaghami, 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.

Gilbert, W.V. (2010). Alternative Ways to Think about Cellular Internal Ribosome Entry. Journal of Biological Chemistry *285*, 29033–29038.

Gilbert, W.V., Zhou, K., Butler, T.K., and Doudna, J.A. (2007). Cap-Independent Translation Is Required for Starvation-Induced Differentiation in Yeast. Science *317*, 1224–1227.

Goyer, C., Altmann, M., Lee, H.S., Blanc, A., Deshmukh, M., Woolford, J.L., Trachsel, H., and Sonenberg, N. (1993). TIF4631 and TIF4632: two yeast genes encoding the high-molecular-weight subunits of the cap-binding protein complex (eukaryotic initiation factor 4F) contain an RNA recognition motif-like sequence and carry out an essential function. Mol Cell Biol *13*, 4860–4874.

Graveley, B.R. (2000). Sorting out the complexity of SR protein functions. RNA *6*, 1197–1211.

Haar, von der, T., and McCarthy, J.E.G. (2002). Intracellular translation initiation factor levels in Saccharomyces cerevisiae and their role in cap-complex function. Mol. Microbiol. *46*, 531–544.

Hentze, M.W., Muckenthaler, M.U., and Andrews, N.C. (2004). Balancing acts: molecular control of mammalian iron metabolism. Cell *117*, 285–297.

Hinnebusch, A.G. (2005). Translational regulation of GCN4 and the general amino acid control of yeast. Annu. Rev. Microbiol. *59*, 407–450.

Hinnebusch, A.G., and Lorsch, J.R. (2012). The mechanism of eukaryotic translation initiation: new insights and challenges. Cold Spring Harbor Perspectives in Biology *4*.

Ibrahimo, S., Holmes, L.E.A., and Ashe, M.P. (2006). Regulation of translation initiation by the yeast eIF4E binding proteins is required for the pseudohyphal response. Yeast *23*, 1075–1088.

Jackson, R.J., Hellen, C.U., and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. Nature Reviews Molecular Cell Biology *11*, 113–127.

Jiang, M., Anderson, J., Gillespie, J., and Mayne, M. (2008). uShuffle: a useful tool for shuffling biological sequences while preserving the k-let counts. BMC Bioinformatics *9*, 192.

Jin, R., Dobry, C.J., McCown, P.J., and Kumar, A. (2008). Large-scale analysis of yeast filamentous growth by systematic gene disruption and overexpression. Mol. Biol. Cell *19*, 284–296.

Jorgensen, P.P., Nishikawa, J.L.J., Breitkreutz, B.-J.B., and Tyers, M.M. (2002). Systematic identification of pathways that couple cell growth and division in yeast. Science *297*, 395–400.

Kopeina, G.S., Afonina, Z.A., Gromova, K.V., Shirokov, V.A., Vasiliev, V.D., and Spirin, A.S. (2008). Step-wise formation of eukaryotic double-row polyribosomes and circular translation of polysomal mRNA. Nucleic Acids Research *36*, 2476–2488.

Kozak, M. (1990). Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. Proc. Natl. Acad. Sci. U.S.A. *87*, 8301–8305.

Lambert, N., Robertson, A., Jangi, M., McGeary, S., Sharp, P.A., and Burge, C.B. (2014). RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. Molecular Cell *54*, 887–900.

Lorenz, R., Bernhart, S.H., Höner Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithms Mol Biol *6*, 26.

Milligan, J.F., and Uhlenbeck, O.C. (1989). Synthesis of small RNAs using T7 RNA polymerase. Methods in Enzymology *180*, 51.

Mitchell, S.F., Walker, S.E., Algire, M.A., Park, E.-H., Hinnebusch, A.G., and Lorsch, J.R. (2010). The 5'-7-Methylguanosine Cap on Eukaryotic mRNAs Serves Both to Stimulate Canonical Translation Initiation and to Block an Alternative Pathway.

Molecular Cell *39*, 950–962.

Park, E.-H., Walker, S.E., Lee, J.M., Rothenburg, S., Lorsch, J.R., and Hinnebusch, A.G. (2011a). Multiple elements in the eIF4G1 N-terminus promote assembly of eIF4G1•PABP mRNPs in vivo. Embo J. *30*, 302–316.

Park, E.-H., Zhang, F., Warringer, J., Sunnerhagen, P., and Hinnebusch, A.G. (2011b). Depletion of eIF4G from yeast cells narrows the range of translational efficiencies genome-wide. BMC Genomics *12*, 68.

Pelechano, V., Wei, W., and Steinmetz, L.M. (2013). Extensive transcriptional heterogeneity revealed by isoform profiling. Nature *497*, 127–131.

Pestova, T.V., Hellen, C.U., and Shatsky, I.N. (1996a). Canonical eukaryotic initiation factors determine initiation of translation by internal ribosomal entry. Mol Cell Biol *16*, 6859–6869.

Pestova, T.V., Shatsky, I.N., and Hellen, C.U. (1996b). Functional dissection of eukaryotic initiation factor 4F: the 4A subunit and the central domain of the 4G subunit are sufficient to mediate internal entry of 43S preinitiation complexes. Mol Cell Biol *16*, 6870–6878.

Rio, D.C., Ares, M.J., Hannon, G.J., Nilsen, T.W., Rio, D.C., Ares, M.J., Hannon, G.J., and Nilsen, T.W. (2011). RNA: a laboratory manual. (Cold Spring Harbour Laboratory Press).

Rojas-Duran, M.F., and Gilbert, W.V. (2012). Alternative transcription start site selection leads to large differences in translation activity in yeast. RNA *18*, 2299–2305.

Ryan, O., Shapiro, R.S., Kurat, C.F., Mayhew, D., Baryshnikova, A., Chin, B., Lin, Z.-Y., Cox, M.J., Vizeacoumar, F., Cheung, D., et al. (2012). Global gene deletion analysis exploring yeast filamentous growth. Science *337*, 1353–1356.

Shalgi, R., Hurt, J.A., Krykbaeva, I., Taipale, M., Lindquist, S., and Burge, C.B. (2013). Widespread regulation of translation by elongation pausing in heat shock. Molecular Cell *49*, 439–452.

Shatsky, I.N., Dmitriev, S.E., Andreev, D.E., and Terenin, I.M. (2014). Transcriptome-wide studies uncover the diversity of modes of mRNA recruitment to eukaryotic ribosomes. Critical Reviews in Biochemistry and Molecular Biology 1–14.

Shively, C.A., Eckwahl, M.J., Dobry, C.J., Mellacheruvu, D., Nesvizhskii, A., and Kumar, A. (2013). Genetic networks inducing invasive growth in Saccharomyces cerevisiae identified through systematic genome-wide overexpression. Genetics *193*, 1297–1310.

Singh, C.R., Watanabe, R., Chowdhury, W., Hiraishi, H., Murai, M.J., Yamamoto, Y.,

Miles, D., Ikeda, Y., Asano, M., and Asano, K. (2012). Sequential Eukaryotic Translation Initiation Factor 5 (eIF5) Binding to the Charged Disordered Segments of eIF4G and eIF2 Stabilizes the 48S Preinitiation Complex and Promotes Its Shift to the Initiation Mode. Mol Cell Biol *32*, 3978–3989.

Sonenberg, N., and Hinnebusch, A.G. (2009). Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. Cell *136*, 731–745.

Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H., and Bartel, D.P. (2014). Poly(A)-tail profiling reveals anembryonic switch in translational control. Nature *508*, 66-71

Thoreen, C.C., Chantranupong, L., Keys, H.R., Wang, T., Gray, N.S., and Sabatini, D.M. (2012). A unifying model for mTORC1-mediated regulation of mRNA translation. Nature *486*, 109–113.

Vaidyanathan, P.P., Zinshteyn, B., Thompson, M.K., and Gilbert, W.V. (2014). Protein kinase A regulates gene-specific translational adaptation in differentiating yeast. RNA *20*, 912–922.

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L.M. (2009). Bidirectional promoters generate pervasive transcription in yeast. Nature *457*, 1033–1037.

Zenklusen, D., Larson, D.R., and Singer, R.H. (2008). Single-RNA counting reveals alternative modes of gene expression in yeast. Nature Structural & Molecular Biology *15*, 1263–1271.

`

ID	gene name	chr	TL length	Oligo(U) length	Oligo(U) position relative to AUG
YGR127W		VII	224	25	-222
YHR009C	TDA3	VIII	251	20	-172
YMR182W-A		XIII	2377	20	-102
YJL089W	SIP4	X	113	18	-53
YHR032C-A		VIII	1475	17	-1468
YDR313C	PIB1	IV	61	17	-51
YLR455W		XII	427	17	-335
YBR169C	SSE2	II	69	16	-50
YIL083C	CAB2	IX	339	16	-144
YER132C	PMD1	V	448	15	-125
YDR310C	SUM1	IV	289	15	-220
YOR324C	FRT1	XV	154	15	-47
YIL055C		IX	498	15	-93
YOR213C	SAS5	XV	116	14	-39
YGR138C	TPO2	VII	276	14	-182
YGL167C	PMR1	VII	421	14	-324
YMR043W	MCM1	XIII	349	14	-192
YMR300C	ADE4	XIII	259	13	-156
YOR198C	BFR1	XV	180	13	-150
YMR124W		XIII	270	13	-162
YER033C	ZRG8	V	215	13	-120
YDL088C	ASM4	IV	106	13	-48
YGL216W	KIP3	VII	258	13	-121
YLR055C	SPT8	XII	84	13	-26
YGR180C	RNR4	VII	221	13	-142
YIL119C	RPI1	IX	187	13	-42
YBL102W	SFT2	II	105	12	-44
YKL042W	SPC42	XI	236	12	-19
YBL081W		II	302	12	-40
YJR153W	PGU1	X	361	12	-278
YPL189W	GUP2	XVI	1024	12	-324
YDR096W	GIS1	IV	556	12	-44
YOL100W	PKH2	XV	243	12	-24
YDR103W	STE5	IV	559	12	-236
YCR008W	SAT4	III	458	12	-341
YJL084C	ALY2	X	135	12	-48
YNR004W	SWM2	XIV	461	12	-225
YOR245C	DGA1	XV	225	12	-135
YER045C	ACA1	V	257	12	-89

YOR344C	TYE7	XV	245	12	-92
YER059W	PCL6	V	157	12	-65
YKR092C	SRP40	XI	203	12	-117
YJR151C	DAN4	X	1846	12	-1507
YNL066W	SUN4	XIV	227	12	-27
YBR140C	IRA1	II	270	12	-215
YBL067C	UBP13	II	282	12	-164
YIR017C	MET28	IX	305	12	-208
YLL045C	RPL8B	XII	73	12	-56
YER111C	SWI4	V	317	12	-279
YBR212W	NGR1	II	433	11	-295
YPL036W	PMA2	XVI	566	11	-67
YML106W	URA5	XIII	300	11	-243
YPL184C	MRN1	XVI	1074	11	-172
YPL184C	MRN1	XVI	1074	11	-47
YJL038C	LOH1	X	1195	11	-529
YNR051C	BRE5	XIV	743	11	-264
YBR285W		II	142	11	-120
YDR275W	BSC2	IV	659	11	-154
YDR275W	BSC2	IV	659	11	-141
YDR295C	HDA2	IV	262	11	-115
YKL135C	APL2	XI	121	11	-49
YDL053C	PBP4	IV	189	11	-133
YNL242W	ATG2	XIV	138	11	-23
YGL014W	PUF4	VII	407	11	-329
YPL089C	RLM1	XVI	260	11	-118
YJL100W	LSB6	X	443	11	-382
YJR103W	URA8	X	179	11	-172
YFL033C	RIM15	VI	220	11	-69
YKR093W	PTR2	XI	246	11	-46
YKR102W	FLO10	XI	3396	11	-1628
YDR501W	PLM2	IV	116	11	-93
YER073W	ALD5	V	514	11	-505
YOR222W	ODC2	XV	489	11	-227
YBR157C	ICS2	II	1292	10	-407
YGR241C	YAP1802	VII	516	10	-311
YOL001W	PHO80	XV	832	10	-184
YKR100C	SKG1	XI	372	10	-40
YKR103W	NFT1	XI	2600	10	-730
YJR127C	RSF2	X	671	10	-597

YNL007C	SIS1	XIV	177	10	-56
YDR464W	SPP41	IV	123	10	-56
YPR192W	AQY1	XVI	101	10	-87
YKR003W	OSH6	XI	164	10	-160
YDR232W	HEM1	IV	251	10	-132
YOR355W	GDS1	XV	524	10	-96
YFL054C		VI	5098	10	-3048
YDR251W	PAM1	IV	301	10	-201
YPL189W	GUP2	XVI	1024	10	-215
YOR086C	TCB1	XV	193	10	-105
YBL041W	PRE7	II	259	10	-199
YJL145W	SFH5	X	85	10	-60
YNR051C	BRE5	XIV	743	10	-94
YBR283C	SSH1	II	193	10	-113
YGL096W	TOS8	VII	99	10	-65
YDR283C	GCN2	IV	236	10	-49
YPR138C	MEP3	XVI	230	10	-48
YDR096W	GIS1	IV	556	10	-238
YDL076C	RXT3	IV	141	10	-41
YKR096W		XI	309	10	-166
YDR122W	KIN1	IV	293	10	-65
YNL190W		XIV	273	10	-91
YJL164C	TPK1	X	229	10	-126
YMR009W	ADI1	XIII	1012	10	-412
YFL005W	SEC4	VI	184	10	-102
YGL215W	CLG1	VII	474	10	-394
YGL215W	CLG1	VII	474	10	-89
YPL257W		XVI	54	10	-48
YPL032C	SVL3	XVI	365	10	-106
YOL002C	IZH2	XV	265	10	-134
YBR083W	TEC1	II	1828	10	-1434
YGL190C	CDC55	VII	556	10	-476
YGR041W	BUD9	VII	344	10	-81
YER054C	GIP2	V	444	10	-157
YMR182W-A		XIII	2377	10	-1731
YMR182W-A		XIII	2377	10	-177
YNL271C	BNI1	XIV	287	10	-148
YLR034C	SMF3	XII	207	10	-110
YKR099W	BAS1	XI	307	10	-34
YDL140C	RPO21	IV	603	10	-262
YER073W	ALD5	V	514	10	-284

YMR240C	CUS1	XIII	406	10	-323
YNL053W	MSG5	XIV	115	10	-79
YLR139C	SLS1	XII	915	10	-146
YPL272C		XVI	1657	10	-1450
YLR371W	ROM2	XII	194	10	-141
YLL028W	TPO1	XII	154	10	-44
YLR296W		XII	275	9	-142
YLR307C-A		XII	492	9	-190
YDL146W	LDB17	IV	182	9	-71
YBR208C	DUR1	II	463	9	-162
YDR044W	HEM13	IV	229	9	-139
YDR232W	HEM1	IV	251	9	-13
YMR102C		XIII	294	9	-294
YEL013W	VAC8	V	240	9	-178
YER053C	PIC2	V	247	9	-117
YOR132W	VPS17	XV	70	9	-50
YER129W	SAK1	V	508	9	-91
YPL189W	GUP2	XVI	1024	9	-225
YKR015C		XI	674	9	-13
YNR051C	BRE5	XIV	743	9	-648
YDR275W	BSC2	IV	659	9	-169
YDR089W		IV	121	9	-62
YKR014C	YPT52	XI	150	9	-70
YGR143W	SKN1	VII	451	9	-426
YKL165C-A		XI	1636	9	-534
YAL029C	MYO4	I	244	9	-168
YGL067W	NPY1	VII	93	9	-34
YDL241W		IV	442	9	-402
YGL008C	PMA1	VII	240	9	-197
YDR119W	VBA4	IV	84	9	-44
YGR157W	CHO2	VII	242	9	-201
YDR477W	SNF1	IV	288	9	-46
YPL233W	NSL1	XVI	539	9	-416
YNL182C	IPI3	XIV	490	9	-145
YOR181W	LAS17	XV	482	9	-217
YLR417W	VPS36	XII	103	9	-15
YGR197C	SNG1	VII	222	9	-105
YGL215W	CLG1	VII	474	9	-225
YLR220W	CCC1	XII	210	9	-71
YLR220W	CCC1	XII	210	9	-39
YIL056W	VHR1	IX	292	9	-203

YDL171C	GLT1	IV	292	9	-101
YKR028W	SAP190	XI	360	9	-312
YGL180W	ATG1	VII	548	9	-310
YGL173C	XRN1	VII	400	9	-260
YDR223W	CRF1	IV	458	9	-223
YBR118W	TEF2	II	105	9	-61
YBR103W	SIF2	II	152	9	-101
YPL145C	KES1	XVI	283	9	-195
YGR041W	BUD9	VII	344	9	-37
YLR150W	STM1	XII	155	9	-124
YOR341W	RPA190	XV	117	9	-45
YOR348C	PUT4	XV	92	9	-76
YMR180C	CTL1	XIII	177	9	-101
YDL140C	RPO21	IV	603	9	-432
YER073W	ALD5	V	514	9	-372
YPR102C	RPL11A	XVI	141	9	-54
YDL133W	SRF1	IV	146	9	-24
YBR162C	TOS1	II	271	9	-181
YLR133W	CKI1	XII	69	9	-62
YFR026C	ULI1	VI	1016	9	-703
YDR092W	UBC13	IV	69	9	-54
YGR230W	BNS1	VII	141	9	-68
YIL119C	RPI1	IX	187	9	-173
YHR079C	IRE1	VIII	260	8	-129
YHR082C	KSP1	VIII	943	8	-720
YNL201C	PSY2	XIV	267	8	-69
YML018C		XIII	149	8	-50
YDR280W	RRP45	IV	130	8	-98
YMR070W	MOT3	XIII	208	8	-128
YOR216C	RUD3	XV	58	8	-15
YER132C	PMD1	V	448	8	-310
YKL033W-A		XI	104	8	-16
YEL025C		V	786	8	-256
YPL036W	PMA2	XVI	566	8	-375
YBR201C-A		II	352	8	-124
YOR137C	SIA1	XV	396	8	-217
YER148W	SPT15	V	201	8	-160
YPL224C	MMT2	XVI	198	8	-158
YMR083W	ADH3	XIII	658	8	-151

YEL017W	GTT3	V	464	8	-451
YKL051W	SFK1	XI	481	8	-84
YER155C	BEM2	V	274	8	-150
YOR355W	GDS1	XV	524	8	-345
YOR355W	GDS1	XV	524	8	-106
YMR102C		XIII	294	8	-260
YEL009C	GCN4	V	602	8	-515
YER167W	BCK2	V	274	8	-137
YER129W	SAK1	V	508	8	-437
YBR253W	SRB6	II	49	8	-16
YPL184C	MRN1	XVI	1074	8	-224
YOR010C	TIR2	XV	1144	8	-290
YJL038C	LOH1	X	1195	8	-464
YJL038C	LOH1	X	1195	8	-110
YHR007C	ERG11	VIII	168	8	-62
YDR062W	LCB2	IV	396	8	-149
YDR069C	DOA4	IV	58	8	-31
YOR267C	HRK1	XV	573	8	-322
YOL125W	TRM13	XV	336	8	-169
YDR017C	KCS1	IV	118	8	-43
YDR073W	SNF11	IV	241	8	-195
YDR452W	PPN1	IV	137	8	-104
YKL176C	LST4	XI	561	8	-411
YOL103W	ITR2	XV	199	8	-18
YGL075C	MPS2	VII	212	8	-68
YOR335C	ALA1	XV	141	8	-135
YKL086W	SRX1	XI	218	8	-189
YCR008W	SAT4	III	458	8	-416
YLR381W	CTF3	XII	36	8	-14
YKL152C	GPM1	XI	66	8	-54
YIL100W		IX	2113	8	-1582
YOR178C	GAC1	XV	348	8	-196
YLR224W		XII	329	8	-222
YGL041W-A		VII	260	8	-207
YMR171C	EAR1	XIII	198	8	-24
YGL039W		VII	92	8	-21
YJL164C	TPK1	X	229	8	-12
YMR042W	ARG80	XIII	89	8	-50
YGR192C	TDH3	VII	88	8	-61
YJL153C	INO1	X	386	8	-28
YFR001W	LOC1	VI	120	8	-107

YBR054W	YRO2	II	523	8	-203
YGL014W	PUF4	VII	407	8	-343
YML011C	RAD33	XIII	49	8	-10
YLR024C	UBR2	XII	211	8	-87
YIL056W	VHR1	IX	292	8	-157
YPR028W	YOP1	XVI	537	8	-118
YJR137C	MET5	X	197	8	-96
YCL030C	HIS4	III	56	8	-15
YJR103W	URA8	X	179	8	-100
YOL017W	ESC8	XV	88	8	-54
YMR107W	SPG4	XIII	60	8	-60
YCR067C	SED4	III	112	8	-42
YBR083W	TEC1	II	1828	8	-635
YGL190C	CDC55	VII	556	8	-410
YER045C	ACA1	V	257	8	-211
YEL063C	CAN1	V	243	8	-155
YPL141C	FRK1	XVI	257	8	-101
YER052C	HOM3	V	136	8	-17
YPL133C	RDS2	XVI	138	8	-56
YMR182W-A		XIII	2377	8	-80
YNL080C	EOS1	XIV	174	8	-72
YHR026W	VMA16	VIII	139	8	-54
YER146W	LSM5	V	659	8	-300
YAL017W	PSK1	I	363	8	-100
YOR381W	FRE3	XV	1236	8	-168
YKL109W	HAP4	XI	325	8	-113
YBL067C	UBP13	II	282	8	-182
YGL114W		VII	244	8	-96
YDL138W	RGT2	IV	287	8	-132
YDR335W	MSN5	IV	357	8	-32
YHR046C	INM1	VIII	88	8	-56
YPR041W	TIF5	XVI	122	8	-60
YIL160C	POT1	IX	537	8	-161
YDR242W	AMD2	IV	86	8	-80
YML066C	SMA2	XIII	464	8	-198
YMR030W	RSF1	XIII	159	8	-57
YDL112W	TRM3	IV	66	8	-17
YGR229C	SMI1	VII	338	8	-140
YMR043W	MCM1	XIII	349	8	-172
YMR010W		XIII	218	8	-92
YKL062W	MSN4	XI	165	8	-34

YER114C	BOI2	V	313	8	-117
YPL262W	FUM1	XVI	183	8	-63
YLR004C	THI73	XII	309	8	-248
YIL119C	RPI1	IX	187	8	-55
YGR233C	PHO81	VII	202	7	-149
YGR240C	PFK1	VII	275	7	-262
YBR158W	AMN1	II	263	7	-177
YGR184C	UBR1	VII	401	7	-142
YHR082C	KSP1	VIII	943	7	-28
YBR203W	COS11 1	II	227	7	-138
YPL058C	PDR12	XVI	425	7	-87
YNL064C	YDJ1	XIV	214	7	-99
YKL046C	DCW1	XI	227	7	-54
YMR246W	FAA4	XIII	245	7	-53
YEL036C	ANP1	V	229	7	-57
YML018C		XIII	149	7	-36
YLL013C	PUF3	XII	292	7	-187
YPL242C	IQG1	XVI	64	7	-59
YOR209C	NPT1	XV	151	7	-105
YPL049C	DIG1	XVI	355	7	-95
YDR206W	EBS1	IV	334	7	-244
YDL146W	LDB17	IV	182	7	-109
YBR211C	AME1	II	109	7	-72
YBR212W	NGR1	II	433	7	-174
YKR103W	NFT1	XI	2600	7	-903
YJR124C		X	252	7	-203
YDR539W	FDC1	IV	103	7	-77
YJR127C	RSF2	X	671	7	-560
YER132C	PMD1	V	448	7	-321
YER132C	PMD1	V	448	7	-295
YMR265C		XIII	417	7	-323
YOR043W	WHI2	XV	580	7	-297
YOR043W	WHI2	XV	580	7	-170
YOR043W	WHI2	XV	580	7	-127
YOR044W	IRC23	XV	301	7	-15
YPL227C	ALG5	XVI	262	7	-233
YMR304W	UBP15	XIII	201	7	-45
YPL042C	SSN3	XVI	239	7	-239
YEL030W	ECM10	V	435	7	-66
YEL028W		V	595	7	-292

YEL025C		V	786	7	-425
YEL023C		V	565	7	-404
YBR201C-A		II	352	7	-47
YPL228W	CET1	XVI	133	7	-48
YDR515W	SLF1	IV	216	7	-131
YDR389W	SAC7	IV	524	7	-288
YDR389W	SAC7	IV	524	7	-131
YGR144W	THI4	VII	233	7	-187
YBL081W		II	302	7	-184
YPR192W	AQY1	XVI	101	7	-70
YGR146C	ECL1	VII	296	7	-97
YMR273C	ZDS1	XIII	186	7	-111
YOR233W	KIN4	XV	197	7	-165
YOR233W	KIN4	XV	197	7	-150
YLR258W	GSY2	XII	69	7	-52
YDR044W	HEM13	IV	229	7	-85
YHR112C		VIII	515	7	-162
YBL099W	ATP1	II	321	7	-107
YOR064C	YNG1	XV	50	7	-18
YGR244C	LSC2	VII	106	7	-91
YEL009C	GCN4	V	602	7	-234
YPL160W	CDC60	XVI	68	7	-50
YBR250W	SPO23	II	700	7	-479
YBR250W	SPO23	II	700	7	-222
YGR249W	MGA1	VII	180	7	-150
YBR194W	AIM4	II	121	7	-106
YFL054C		VI	5098	7	-4935
YER168C	CCA1	V	53	7	-11
YKR002W	PAP1	XI	239	7	-16
YBL052C	SAS3	II	262	7	-254
YDR247W	VHS1	IV	348	7	-125
YPL184C	MRN1	XVI	1074	7	-1014
YPL184C	MRN1	XVI	1074	7	-788
YBL055C		II	387	7	-32
YOR010C	TIR2	XV	1144	7	-966
YOR010C	TIR2	XV	1144	7	-649
YOR091W	TMA46	XV	336	7	-109
YPL243W	SRP68	XVI	100	7	-46
YDL006W	PTC1	IV	337	7	-264
YDR420W	HKR1	IV	374	7	-351
YDR422C	SIP1	IV	68	7	-67

YNR051C	BRE5	XIV	743	7	-339
YMR139W	RIM11	XIII	219	7	-119
YKL182W	FAS1	XI	552	7	-179
YOR267C	HRK1	XV	573	7	-286
YBL017C	PEP1	II	209	7	-112
YBL015W	ACH1	II	94	7	-29
YOR116C	RPO31	XV	268	7	-132
YER032W	FIR1	V	69	7	-13
YIL118W	RHO3	IX	132	7	-103
YDR275W	BSC2	IV	659	7	-54
YBR292C		II	783	7	-89
YHL016C	DUR3	VIII	797	7	-730
YLR353W	BUD8	XII	472	7	-428
YMR148W	OSW5	XIII	244	7	-11
YGL094C	PAN2	VII	195	7	-98
YDR452W	PPN1	IV	137	7	-70
YLR378C	SEC61	XII	218	7	-131
YOL109W	ZEO1	XV	79	7	-69
YDR292C	SRP101	IV	154	7	-43
YDR293C	SSD1	IV	853	7	-381
YDR293C	SSD1	IV	853	7	-172
YLR355C	ILV5	XII	87	7	-39
YDL117W	CYK3	IV	380	7	-134
YIL108W		IX	227	7	-148
YDR096W	GIS1	IV	556	7	-119
YOR313C	SPS4	XV	568	7	-253
YMR158W	MRPS8	XIII	54	7	-29
YOR133W	EFT1	XV	81	7	-70
YNL011C		XIV	261	7	-212
YDL084W	SUB2	IV	164	7	-20
YML038C	YMD8	XIII	134	7	-119
YKL165C-A		XI	1636	7	-306
YKL161C	KDX1	XI	212	7	-190
YMR223W	UBP8	XIII	217	7	-24
YOR152C		XV	605	7	-70
YCR008W	SAT4	III	458	7	-121
YAL029C	MYO4	I	244	7	-237
YOL085W-A		XV	689	7	-647
YIL106W	MOB1	IX	99	7	-86
YIL106W	MOB1	IX	99	7	-57

YPR159W	KRE6	XVI	296	7	-225
YBL002W	HTB2	II	439	7	-116
YGL008C	PMA1	VII	240	7	-223
YER065C	ICL1	V	333	7	-77
YJL084C	ALY2	X	135	7	-69
YGR149W		VII	393	7	-176
YPR072W	NOT5	XVI	101	7	-39
YMR320W		XIII	1303	7	-1086
YMR320W		XIII	1303	7	-663
YNL046W		XIV	181	7	-113
YDR259C	YAP6	IV	215	7	-92
YMR088C	VBA1	XIII	144	7	-86
YPR026W	ATH1	XVI	581	7	-133
YMR271C	URA10	XIII	570	7	-356
YMR271C	URA10	XIII	570	7	-326
YDR121W	DPB4	IV	130	7	-82
YML075C	HMG1	XIII	235	7	-108
YNR016C	ACC1	XIV	1080	7	-370
YNR016C	ACC1	XIV	1080	7	-54
YJR036C	HUL4	X	292	7	-247
YGL255W	ZRT1	VII	2073	7	-1470
YGL255W	ZRT1	VII	2073	7	-929
YOR175C	ALE1	XV	531	7	-236
YLR407W		XII	238	7	-231
YOR176W	HEM15	XV	232	7	-200
YKL139W	CTK1	XI	247	7	-126
YKL139W	CTK1	XI	247	7	-115
YML007W	YAP1	XIII	191	7	-116
YPL020C	ULP1	XVI	127	7	-41
YGL242C		VII	64	7	-11
YLR224W		XII	329	7	-173
YLR224W		XII	329	7	-57
YPL016W	SWI1	XVI	384	7	-273
YLR138W	NHA1	XII	157	7	-46
YDR495C	VPS3	IV	59	7	-33
YDR357C	CNL1	IV	169	7	-75
YJR060W	CBF1	X	160	7	-153
YDL053C	PBP4	IV	189	7	-117
YNL283C	WSC2	XIV	464	7	-174
YNL278W	CAF120	XIV	108	7	-88

YOR161C	PNS1	XV	159	7	-93
YML091C	RPM2	XIII	298	7	-85
YLR422W		XII	150	7	-29
YNL161W	CBK1	XIV	180	7	-80
YGR191W	HIP1	VII	230	7	-75
YJL153C	INO1	X	386	7	-98
YPR194C	OPT2	XVI	314	7	-90
YDR153C	ENT5	IV	136	7	-102
YNL095C		XIV	165	7	-62
YOL039W	RPP2A	XV	198	7	-51
YMR216C	SKY1	XIII	511	7	-271
YNR048W		XIV	36	7	-14
YBR054W	YRO2	II	523	7	-133
YIL083C	CAB2	IX	339	7	-22
YLR332W	MID2	XII	437	7	-264
YPL116W	HOS3	XVI	212	7	-181
YLR028C	ADE16	XII	217	7	-115
YJL137C	GLG2	X	64	7	-42
YLR219W	MSC3	XII	288	7	-285
YLR220W	CCC1	XII	210	7	-122
YDR173C	ARG82	IV	43	7	-10
YNL144C		XIV	115	7	-55
YER026C	CHO1	V	84	7	-84
YPL221W	FLC1	XVI	233	7	-170
YLR041W		XII	737	7	-261
YER184C		V	378	7	-148
YGR281W	YOR1	VII	236	7	-149
YGL193C		VII	2262	7	-1627
YIL053W	RHR2	IX	136	7	-89
YKR019C	IRS4	XI	102	7	-34
YJL129C	TRK1	X	277	7	-18
YPR028W	YOP1	XVI	537	7	-254
YPR035W	GLN1	XVI	442	7	-217
YOL015W	IRC10	XV	653	7	-73
YLR455W		XII	427	7	-81
YLR455W		XII	427	7	-66
YPL172C	COX10	XVI	129	7	-32
YGR014W	MSB2	VII	264	7	-209
YGR014W	MSB2	VII	264	7	-201
YFL050C	ALR2	VI	741	7	-698
YCR084C	TUP1	III	253	7	-60

YMR108W	ILV2	XIII	242	7	-162
YDR007W	TRP1	IV	190	7	-119
YGL190C	CDC55	VII	556	7	-401
YGL167C	PMR1	VII	421	7	-267
YDL188C	PPH22	IV	182	7	-57
YNL103W	MET4	XIV	166	7	-19
YIL030C	SSM4	IX	61	7	-56
YOL140W	ARG8	XV	69	7	-42
YGR041W	BUD9	VII	344	7	-336
YER047C	SAP1	V	199	7	-108
YJL089W	SIP4	X	113	7	-30
YOR340C	RPA43	XV	76	7	-52
YBR218C	PYC2	II	128	7	-81
YFL026W	STE2	VI	505	7	-155
YPR119W	CLB2	XVI	392	7	-281
YPR119W	CLB2	XVI	392	7	-76
YDR028C	REG1	IV	514	7	-206
YNL287W	SEC21	XIV	191	7	-129
YIL022W	TIM44	IX	106	7	-50
YLR083C	EMP70	XII	174	7	-127
YDR277C	MTH1	IV	183	7	-178
YHR019C	DED81	VIII	63	7	-12
YJL079C	PRY1	X	155	7	-79
YNR075W	COS10	XIV	792	7	-164
YPL019C	VTC3	XVI	327	7	-77
YHR206W	SKN7	VIII	280	7	-70
YMR182W-A		XIII	2377	7	-1276
YER063W	THO1	V	75	7	-48
YPL012W	RRP12	XVI	109	7	-86
YNL277W	MET2	XIV	98	7	-82
YMR041C	ARA2	XIII	225	7	-147
YOR370C	MRS6	XV	335	7	-83
YJR151C	DAN4	X	1846	7	-400
YIR006C	PAN1	IX	172	7	-146
YHR071W	PCL5	VIII	381	7	-221
YER146W	LSM5	V	659	7	-407
YNL065W	AQR1	XIV	122	7	-43
YKR102W	FLO10	XI	3396	7	-2977
YKR102W	FLO10	XI	3396	7	-1194
YKR102W	FLO10	XI	3396	7	-781
YDL145C	COP1	IV	284	7	-46

YKL109W	HAP4	XI	325	7	-304
YGR282C	BGL2	VII	68	7	-59
YDL213C	NOP6	IV	278	7	-102
YKR090W	PXL1	XI	255	7	-85
YPR104C	FHL1	XVI	288	7	-169
YPR104C	FHL1	XVI	288	7	-103
YNL057W		XIV	1248	7	-728
YHR074W	QNS1	VIII	50	7	-22
YDR528W	HLR1	IV	301	7	-40
YNL251C	NRD1	XIV	283	7	-110
YDL133W	SRF1	IV	146	7	-49
YOR222W	ODC2	XV	489	7	-197
YPL186C	UIP4	XVI	252	7	-117
YOR049C	RSB1	XV	209	7	-118
YPR115W	RGC1	XVI	192	7	-163
YBR165W	UBS1	II	79	7	-33
YBR167C	POP7	II	319	7	-122
YER087W	AIM10	V	59	7	-32
YGL103W	RPL28	VII	47	7	-44
YFR020W		VI	957	7	-33
YIL160C	POT1	IX	537	7	-363
YDR237W	MRPL7	IV	150	7	-48
YBR242W		II	153	7	-102
YPR043W	RPL43A	XVI	64	7	-53
YOL105C	WSC3	XV	278	7	-178
YOL105C	WSC3	XV	278	7	-25
YFR030W	MET10	VI	107	7	-46
YPR053C		XVI	1191	7	-925
YPL088W		XVI	655	7	-462
YML066C	SMA2	XIII	464	7	-167
YPL272C		XVI	1657	7	-1622
YMR037C	MSN2	XIII	465	7	-332
YMR037C	MSN2	XIII	465	7	-79
YFR040W	SAP155	VI	300	7	-172
YGR286C	BIO2	VII	108	7	-104
YGL201C	MCM6	VII	64	7	-9
YMR043W	MCM1	XIII	349	7	-147
YNL268W	LYP1	XIV	372	7	-199
YLR371W	ROM2	XII	194	7	-116
YPL262W	FUM1	XVI	183	7	-106

YFR039C		VI	138	7	-89
YIL119C	RPI1	IX	187	7	-150
YIL119C	RPI1	IX	187	7	-135
YEL044W	IES6	V	708	7	-123
YDR005C	MAF1	IV	137	7	-25
YLL027W	ISA1	XII	209	7	-50
YOR020W-A		XV	515	7	-325
YOR020W-A		XV	515	7	-205
YLL025W	PAU17	XII	1977	7	-1465
YLL025W	PAU17	XII	1977	7	-980

Table 4.1: All instances of 7 or more consecutive U in yeast TLs

gene	chr	start	end	sacCer3 oligo(U) length	species conserved	species aligned	Null mutant Invasion Phenotypes (yeastgenome.org)
YLR455W	chrXII	1053294	1053310	17	5	5	
YIL083C	chrIX	204782	204797	16	5	5	
YDR310C	chrIV	1084522	1084536	15	5	5	
YIL119C	chrIX	137907	137919	13	5	5	FLO11 transcription deregulated
YGR180C	chrVII	856431	856443	13	5	5	
YOR344C	chrXV	978150	978161	12	5	5	
YKR092C	chrXI	613991	614002	12	5	5	
YBL067C	chrII	96035	96046	12	5	5	biofilm formation decreased
YOR222W	chrXV	758103	758113	11	5	5	
YGL014W	chrVII	465812	465822	11	5	5	No Invasive growth
YFL033C	chrVI	74486	74496	11	5	5	establishment of stationary phase
YPR138C	chrXVI	812492	812501	10	5	5	Filamentous growth decreased
YOL002C	chrXV	324489	324498	10	5	5	Decreased invasion (on overexpression)
YNR051C	chrXIV	718412	718421	10	5	5	No Invasive growth
YNL190W	chrXIV	282304	282313	10	5	5	
YNL007C	chrXIV	619612	619621	10	5	6	
YKR100C	chrXI	639999	640008	10	5	5	No Invasive growth
YKR003W	chrXI	445221	445230	10	5	5	
YJR127C	chrX	663647	663656	10	5	5	
YDR232W	chrIV	927320	927329	10	5	6	

YBR283C	chrII	770519	770528	10	5	5	
YOR341W	chrXV	960942	960950	9	5	5	Increased invasive growth
YNR051C	chrXIV	718967	718975	9	5	5	
YMR102C	chrXIII	472638	472646	9	5	5	Increased invasion (on overexpression)
YKR028W	chrXI	493945	493953	9	5	7	
YGR143W	chrVII	774767	774775	9	5	5	
YGL215W	chrVII	87756	87764	9	5	5	
YGL008C	chrVII	482855	482863	9	5	5	
YER129W	chrV	417190	417198	9	5	5	No pseudohyphal growth
YDR089W	chrIV	622050	622058	9	5	5	
YAL029C	chrI	92430	92438	9	5	5	
YPL262W	chrXVI	47273	47280	8	5	5	
YOR355W	chrXV	1005031	1005038	8	5	6	
YOL125W	chrXV	83665	83672	8	5	5	
YMR102C	chrXIII	472605	472612	8	5	5	Increased invasion (on overexpression)
YMR010W	chrXIII	285008	285015	8	5	5	Increased invasion on overexpression. 2nd ORF of a bictistronic transcript.
YLR024C	chrXII	193361	193368	8	5	5	
YHR026W	chrVIII	160790	160797	8	5	6	
YER155C	chrV	482991	482998	8	5	5	
YOR222W	chrXV	758133	758139	7	5	5	
YNL251C	chrXIV	174419	174425	7	5	5	Increased invasion (on overexpression)
YMR043W	chrXIII	353724	353730	7	5	5	
YLR219W	chrXII	573866	573872	7	5	5	
YKL182W	chrXI	100492	100498	7	5	5	
YGL190C	chrVII	147784	147790	7	5	6	No pseudohyphal growth
YGL008C	chrVII	482883	482889	7	5	5	
YBR158W	chrII	556372	556378	7	5	5	No Invasive growth
YBL015W	chrII	194093	194099	7	5	7	No pseudohyphal growth

Table 4.2: All instances of U_7 in *S cerevisiae* TLs conserved through sacBay, sacCer3, sacKud, sacMik, sacPar.

Gene phenotypes related to pseudohyphal or invasive growth are shown. See Methods for details

Oligo Name	Use	Sequence (5'->3')
oBZ49	T7 antisense primer	TAATACGACTCACTATA
oBZ52	20nt random library	CCTTGGCACCCGAGAATTCCANNNNNNNNNNNNNNNNNNNCTATAGTGAGTCGTATTA
oBZ70	Poly-CA oligo for EMSA	TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGCTATAGTGAGTCGTATTA
oBZ71	U_{10} oligo for EMSA	TGTGTGTGTGTGTGTGTGTGAAAAAAAAAATGTGTGTGTGTGTGTGTGCTATAGTGAGTCGTATTA
oBZ72	G_{10} oligo for EMSA	TGTGTGTGTGTGTGTGTGTGCCCCCCCCCCTGTGTGTGTGTGTGTGTGCTATAGTGAGTCGTATTA

Table 4.3: DNA oligos used in this study.
All oligos except for oBZ49 were gel-purified prior to use.

Chapter 5: Future Directions

Summary

The focus of this thesis is the elucidation of mechanisms by which sequence features determine the translational efficiency of an mRNA. In Chapter 2 I determined the codon-level effects of loss of the mcm^5s^2U tRNA modification (MSUM). Based on previous genetic and biochemical evidence, yeast mutants were strongly expected to show rate-limiting translation elongation defects at specific codons. However, direct measurement of codon-specific ribosome densities by Ribo-seq suggested that this was not the case. Instead, these strains appear to have activated a GCN4-mediated response to hypomodified tRNA, which accounts for part of the mutant phenotypes. In Chapter 4, I explored the hypothesis that mRNA sequences and mediate competition for translation initiation factors. I used *in vitro* binding to a pool of randomized RNA to determine the innate RNA binding preferences of eIF4G, and found that it preferentially binds oligo(U) sequences. Oligo(U) sequences are found in hundreds of yeast transcript leaders (TLs), and conserved among yeast species. mRNAs that contain oligo(U) in their TL are preferentially translated when eIF4G is depleted, suggesting that these mRNAs are outcompeting others for limiting eIF4G. In this chapter, I review the implications of these findings and propose future directions for this research.

The link between loss of the mcm^{5}s^{2}U tRNA modification and organismal phenotypes

Amino acid misincorporation and protein misfolding

The small codon-specific changes seen in the MSUM-deficient strains suggests that the MSUM phenotypes are not the result of rate-limiting elongation defects on specific mRNAs. However, the slightly slowed translation at MSUM codons could still reduce functional protein levels without making elongation rate-limiting. Slowed translation might lead to occasional misincorporation of amino acids at MSUM codons, as near-cognate tRNAs would now have more chances to sample these codons. As the MSUM codons are commonly used in yeast genes, even a small rate of misincorporation would lead to a significant number proteins with one or more misincorporated amino acids. Precise regulation of translation elongation rates appears to be important for proper folding of some *E. coli* proteins (Zhang et al., 2009) *in vitro*, as well as recognition of membrane proteins by the signal recognition particle in yeast (Pechmann and Frydman, 2012). The small perturbations in decoding rate at MSUM codons could be sufficient to cause protein misfolding if they occur in a particularly sensitive part of a message. Both of these mechanisms could lead to a lack of functional protein, or even the accumulation of toxic misfolded peptides. Such accumulations of unfolded proteins would likely lead to activation of the unfolded protein response, which would explain the activation of GCN4 in the MSUM strains (Patil et al., 2004).

Loss of tRNA modification could cause degradation of mRNA with extreme elongation stalls

Some extreme elongation pauses, including those caused by rare codons, result in nucleolytic cleavage and degradation of the mRNA by the no-go decay pathway (Doma and Parker, 2006). A recent study in mice showed that ribosome pausing resulting from a defective tRNA was only manifest upon loss of a ribosome recycling factor (Ishimura et al., 2014), suggesting that prolonged pauses lead to premature ribosome release. Both of these mechanisms could substantially reduce protein output from specific mRNAs, but would be difficult to detect with Ribo-seq. To further determine the effect of MSUM on translation elongation rate, the footprint profiling experiments need to be repeated in double-mutant strains lacking these ribosome rescue pathways as well as MSUM genes.

A signaling response to hypomodified tRNA

There is precedent for GCN2-independent activation of GCN4 translation by defects in tRNA processing, but the mechanism of this activation is unknown (Qiu et al., 2000). It is possible that this activation is downstream of defective translation of a particular protein, or that lack of MSUM leads to perturbed interactions with tRNA synthetases, or other proteins that interact with tRNAs. Recent studies have shown that the repertoire of yeast tRNA modifications changes in response to cellular stress (Chan et al., 2010; Patil et al., 2012), and that s^2U can be dethiolated during oxidative conditions (Sochacka et al., 2013). This raises the intriguing possibility that the genetic

ablation of MSUM mimics a natural stress condition that would be alleviated by GCN4 activation. Additional studies are needed to determine the molecular basis for sensing of unmodified tRNA and subsequent activation of GCN4.

Regulation of translation initiation by factor competition

Translation factor levels are reduced in many cellular stress conditions and differentiation responses, including the nutrient-deprived conditions where wild yeast spend most of their life (Berset et al., 1998; Gasch et al., 2000; Vaidyanathan et al., 2014). These reduced factor levels will likely increase competition between mRNAs for these factors, and this competition could privilege translation of proteins required for survival in these conditions, while reducing translation of others.

Translation rates are likely to be determined by redundant interactions with multiple initiation factors

Due to the cooperative and disordered nature of 43S complex recruitment, it is likely that no single interaction is rate-limiting for this process and that different steps are limiting for different messages. Thus the rate of 43S recruitment for an mRNA will be a complex function of the affinity of multiple factors for each other and the mRNA. The overall initiation rate will additionally depend on the rate of scanning through the TL, as well as the efficiency of start codon recognition. Several other translation initiation factors including eIF4A, eIF4B, and several subunits of eIF3 have RNA binding activity. Determining the RNA binding specificity of these proteins will be necessary for

quantitatively determining the contribution of individual RNA sequences to 43S recruitment.

eIF4G paralogs with different RNA binding properties could contribute to gene-specific translational efficiencies.

The yeast genome encodes two paralogs of eIF4G, eIF4G1 and eIF4G2, (Goyer et al., 1993) that could play different roles in regulating translation. The greatest amount of sequence divergence between these two proteins is at the amino terminus, which includes the first RNA binding domain (Berset, 2003; Clarkson et al., 2010), raising the possibility that they have different RNA binding specificities. This would allow the cell to modulate translation of different sets of genes by changing the relative levels of eIF4G1 and eIF4G2. In mammalian cells, depletion of eIF4G1 or eIF4G2 inhibits translation of different groups of genes, with implications for viral infection, nutritional signaling and cancer progression (Badura et al., 2012; Castelló et al., 2006; Silvera et al., 2009; Thoreen et al., 2012). It will be of great interest to compare the RNA binding preferences of eIF4G2 to those of eIF4G1, and to see if different RNA binding specificities of eIF4G paralogs can mediate differential translation efficiency in yeast and mammalian cells.

References

Badura, M., Braunstein, S., Zavadil, J., and Schneider, R.J. (2012). DNA damage and eIF4G1 in breast cancer cells reprogram translation for survival and DNA repair mRNAs. Proc. Natl. Acad. Sci. U.S.A. *109*, 18767–18772.

Berset, C., Trachsel, H., and Altmann, M. (1998). The TOR (target of rapamycin) signal

transduction pathway regulates the stability of translation initiation factor eIF4G in the yeast Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U.S.A. *95*, 4264–4269.

Berset, C., Zurbriggen, A., Djafarzadeh, S., Altmann, M., and Trachsel, H. (2003). RNA-binding activity of translation initiation factor eIF4G1 from Saccharomyces cerevisiae. RNA *9*, 871–880.

Castelló, A., Álvarez, E., and Carrasco, L. (2006). Differential Cleavage of eIF4GI and eIF4GII in Mammalian Cells EFFECTS ON TRANSLATION. J. Biol. Chem. *281*, 33206–33216.

Chan, C.T.Y., Dyavaiah, M., DeMott, M.S., Taghizadeh, K., Dedon, P.C., and Begley, T.J. (2010). A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. PLoS Genet *6*, e1001247.

Clarkson, B.K., Gilbert, W.V., and Doudna, J.A. (2010). Functional overlap between eIF4G isoforms in Saccharomyces cerevisiae. PLoS ONE *5*, e9114.

Doma, M.K., and Parker, R. (2006). Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. Nature *440*, 561–564.

Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., and Brown, P.O. (2000). Genomic expression programs in the response of yeast cells to environmental changes. Mol. Biol. Cell *11*, 4241–4257.

Goyer, C., Altmann, M., Lee, H.S., Blanc, A., Deshmukh, M., Woolford, J.L., Trachsel, H., and Sonenberg, N. (1993). TIF4631 and TIF4632: two yeast genes encoding the high-molecular-weight subunits of the cap-binding protein complex (eukaryotic initiation factor 4F) contain an RNA recognition motif-like sequence and carry out an essential function. Mol Cell Biol *13*, 4860–4874.

Ishimura, R., Nagy, G., Dotu, I., Zhou, H., Yang, X.L., Schimmel, P., Senju, S., Nishimura, Y., Chuang, J.H., and Ackerman, S.L. (2014). Ribosome stalling induced by mutation of a CNS-specific tRNA causes neurodegeneration. Science *345*, 455–459.

Patil, A., Dyavaiah, M., Joseph, F., Rooney, J.P., Chan, C.T.Y., Dedon, P.C., and Begley, T.J. (2012). Increased tRNA modification and gene-specific codon usage regulate cell cycle progression during the DNA damage response. Cell Cycle *11*, 3656–3665.

Patil, C.K., Li, H., and Walter, P. (2004). Gcn4p and Novel Upstream Activating Sequences Regulate Targets of the Unfolded Protein Response. PLoS Biol *2*, e246.

Pechmann, S., and Frydman, J. (2012). Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. Nature Structural & Molecular Biology *20*, 237–243.

Qiu, H., Hu, C., Anderson, J., Bjork, G.R., Sarkar, S., Hopper, A.K., and Hinnebusch, A.G. (2000). Defects in tRNA processing and nuclear export induce GCN4 translation independently of phosphorylation of the alpha subunit of eukaryotic translation initiation factor 2. Mol Cell Biol *20*, 2505–2516.

Silvera, D., Arju, R., Darvishian, F., Levine, P.H., Zolfaghari, L., Goldberg, J., Hochman, T., Formenti, S.C., and Schneider, R.J. (2009). Essential role for eIF4GI overexpression in the pathogenesis of inflammatory breast cancer. Nature Cell Biology *11*, 903–908.

Sochacka, E., Bartos, P., Kraszewska, K., and Nawrot, B. (2013). Bioorganic & Medicinal Chemistry Letters. Bioorganic & Medicinal Chemistry Letters *23*, 5803–5805.

Thoreen, C.C., Chantranupong, L., Keys, H.R., Wang, T., Gray, N.S., and Sabatini, D.M. (2012). A unifying model for mTORC1-mediated regulation of mRNA translation. Nature *486*, 109–113.

Vaidyanathan, P.P., Zinshteyn, B., Thompson, M.K., and Gilbert, W.V. (2014). Protein kinase A regulates gene-specific translational adaptation in differentiating yeast. RNA *20*, 912–922.

Zhang, G., Hubalewska, M., and Ignatova, Z. (2009). Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. Nature Structural & Molecular Biology *16*, 274–280.

Appendix I: Condition-specific Perturbation of Ribosome Footprints by Cycloheximide Treatment

Abstract

Ribosome footprint profiling (Ribo-seq) has allowed the interrogation of translational processes in living cells with unprecedented detail and resolution. In order to properly interpret data produced from this novel technique, it is important to understand possible sources of bias. Most previous studies have used the translational inhibitor cycloheximide (CHX) to prepare ribosome footprints, and although the mechanism of action of this drug is well understood, it's effect on the complicated *in vivo* distribution of ribosomes is unclear. We examined the effect of CHX on ribosome footprint distributions from yeast undergoing glucose starvation, and found very large and complex condition-specific effects on ribosome distribution. These effects may have confounded previous studies reporting widespread regulated translation in transcript leaders.

Introduction

The ribosome footprint profiling (Ribo-seq) technique (Ingolia et al., 2009) allows the determination of ribsome positions genome-wide with nucleotide resolution. The technique consists of isolating mRNA-bound ribosomes, digesting them with an RNase, and isolating and sequencing the ribosome-protected mRNA fragments. In addition to allowing the inference of initiation rates and protein levels from the ribosome density on an mRNA, the positional information from this technique has allowed the probing of

detailed translational events, such as translation of upstream open reading frames (Brar et al., 2012; Ingolia et al., 2009), translational pausing (Guydosh and Green, 2014; Han et al., 2014; Ishimura et al., 2014; Shalgi et al., 2012), production of alternative protein isoforms (Lee et al., 2012), and codon-level modulation of decoding speed (Qian et al., 2012; Stadler and Fire, 2013; Zinshteyn and Gilbert, 2013). In order to reliably isolate mRNA-bound ribosomes, cells are usually incubated with translational inhibitors such as CHX, which binds the E-site of the ribosome and prevents its translocation (Garreau de Loubresse et al., 2014; Schneider-Poetsch et al., 2010). It is generally assumed that this treatment faithfully captures the *in-vivo* positions of ribosomes.

In this appendix, we set out to study gene-specific translational effects in yeast adjusting to glucose starvation (Vaidyanathan et al., 2014), a condition in which a dramatic reduction in protein synthesis has occurred (Ashe et al., 2000). We noticed large condition-specific changes in ribosome positions, which we initially interpreted as a novel mode of translational regulation. However, these effects were entirely dependent on pretreatment of yeast with CHX. These findings indicate that great care must be taken when interpreting positional information from Ribo-seq data, particularly when translational inhibitors have been used.

Results

CHX treatment leads to accumulation of ribosomes at start codons and in transcript leaders

In addition to quantifying effects on gene-specific translational efficiency, Ribo-seq provides information about the positions of ribosomes along the message. We noticed prominent accumulation of Ribo-seq reads over the initiation codons of many genes following 3 hours of glucose withdrawal (e.g. *YDR224C*, Figure I.1A). To determine how general this effect was, read density as a function of position was determined for each gene. Averaging positional densities for all genes revealed a dramatic increase in footprint reads at AUG initiation codons as well as a substantial increase in reads upstream (Figure I.1B). Most genes showed an increased fraction of start codon reads with a median increase of ~4-fold (Figure I.1C).

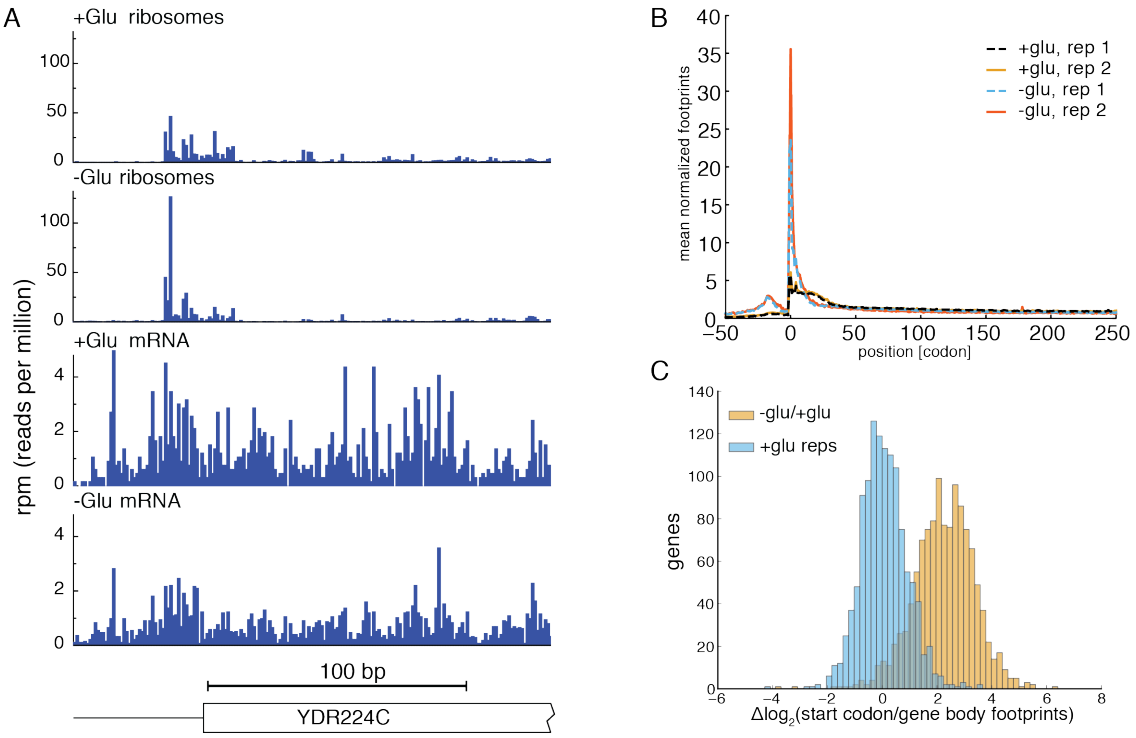

Figure I.1 : Ribosomes footprints accumulate on start codons in Ribo-seq data from glucose starved yeast.
(A) Distribution of Ribo-seq and RNA-seq read 5' ends across a typical gene (*YDR224C*). The peaks of ribo-seq reads 12nt upstream of the start codon are from ribosomes with the start codon in their P site.
(B) Glucose starvation increased Ribo-seq read density at start codons and upstream. Read density as a function of position was determined for each gene and then all genes were averaged.
(C) Most genes showed increased read density at start codons in glucose-starved cells. The distributions of changes in the $\log_2$ ratio of (start codon reads/gene body reads) are plotted for +glu replicates in blue, and for –glu vs +glu in yellow.

Increased start codon footprint density suggested the intriguing possibility that glucose starvation induced a novel form of translational regulation – AUG pausing. However, these libraries were made from cells incubated with cycloheximide (CHX) *in vivo*, a treatment that increased start codon reads in unstarved control samples (Figure I.2A) consistent with previous observations (Ingolia et al., 2009). This effect of CHX incubation can be explained by the fact that CHX does not prevent loading of ribosomes

but traps any newly loaded ribosomes at the first translocation step in elongation (Schneider-Poetsch et al., 2010). To determine whether the observed ~4-fold increase in ribosome density at initiation codons represented a biological effect of glucose starvation, we performed footprint profiling on starved cells without CHX incubation. To minimize ribosome run-off during handling, the cells were harvested by rapid filtration and fast freezing in liquid nitrogen. Fast freezing preserved Ribo-seq read density in the 5' regions of ORFs compared to normal processing without CHX, as expected (Figure I.2B). In the absence of CHX incubation, no increase in start codon footprints was observed following glucose withdrawal. In fact, starved cells showed reduced ribosome density at initiation codons compared to unstarved cells processed in parallel (Figure I.2C). Glucose starvation also caused a pronounced increase in 5' transcript leader (TL) reads that was entirely dependent on CHX incubation (Figure I.2D). Thus, we conclude that glucose starvation does not induce AUG pausing.

Ribosome footprint accumulations at start codons are caused by ribosome run-on during CHX pre-treatment

In light of these observations, we considered an alternative explanation for CHX-dependent increased AUG read density in glucose-starved cells. Starved cells showed globally reduced polysome size (Vaidyanathan et al., 2014), consistent with reduced ribosome density in the body of many ORFs. If additional ribosomes were captured at initiation codons during the in vivo CHX incubation in both starved and unstarved conditions, these AUG-bound ribosomes would make up a larger fraction of total

ribosome footprints in the starved sample. This model predicts that ORFs with relatively high ribosome density in starved cells should show smaller AUG peaks. Consistent with this hypothesis, the starvation-induced increase in start codon read density was anti-correlated with the change in TE (Figure I.2E), and this relationship was not observed without CHX incubation (Figure I.2F).

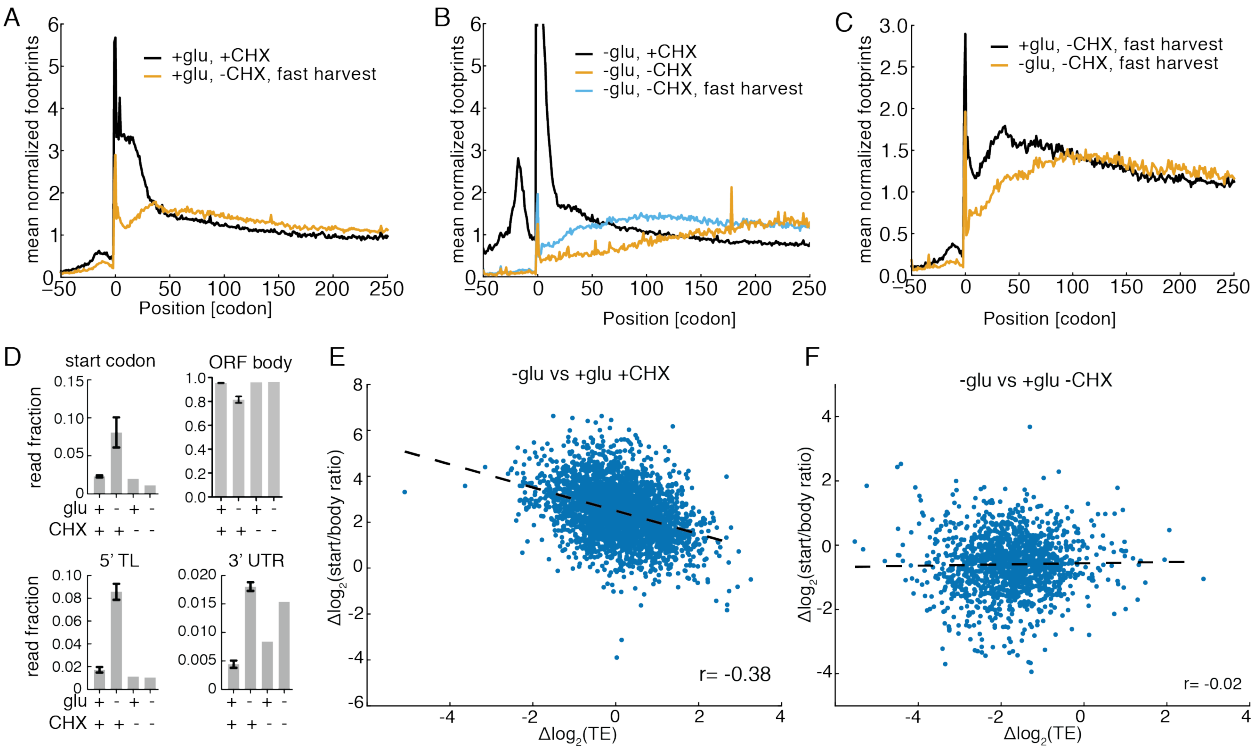

Figure I.2: Ribosome footprint accumulations at start codons are artifacts of ribosome run-on during CHX pre-treatment.
(A) Incubation with CHX increased read density at start codons and 5'regions in glucose media. Standard Ribo-seq with in vivo incubation with CHX for 2 min was compared with Ribo-seq of cells harvested by rapid filtration and processed without CHX.
(B) Incubation with CHX was required to see increased start codon and 5'TL read density in glucose starved cells.
(C) Direct comparison of distributions of ±glucose footprints prepared by fast freezing without CHX.
(D) Quantitation of the effects of glucose and CHX on footprint distribution. Error bars represent S.D. of two biological replicates except for the –CHX libraries for which there is a single replicate.

(E-F) Reduced translation was correlated with elevated start codon reads only in CHX-treated cells.

CHX treatment altered the codon-level distributions of ribosome footprints

We examined whether CHX incubation affected other codon-level distributions of ribosome footprints. Bulk codon occupancy for each sense codon was determined for A, P and E sites with and without CHX treatment (Zinshteyn and Gilbert, 2013). Codon occupancy was reproducible between replicates (Figure I.3A,C) and globally perturbed by incubation with CHX (Figure I.3B, D). A recent report (Lareau et al., 2014) showed that there are actually two distinct populations of yeast ribosome footprints, likely representing different conformations of the ribosome. The longer footprints are stabilized by CHX and captured in our assay, and the distribution of these two states depends on the amino acid encoded by the codon in the ribosomal A-site. It is thus possible that CHX has different activities depending on the identity of the codon in the A site. We do not assume that the fast-frozen samples without CHX represent a true snapshot of in vivo ribosome distributions as some run-off likely occurred despite rapid collection. Importantly, calculations of gene-specific TEs and the effects of glucose starvation on TE were insensitive to CHX and to inclusion or omission of start codon reads (Figure I.4). This was understandable given that start codon reads made up <10% of any library. The apparent prominence of the AUG peak in metagene analyses was due to the concentration of 7-10% of reads at ~0.2% of possible codon positions (given a median ORF length of 409 codons). Thus, while some metrics exhibited strong CHX

dependence (TL reads, start/body ratios, codon occupancy), others (FP RPKM, TE) were robust to the method of Ribo-seq library preparation.

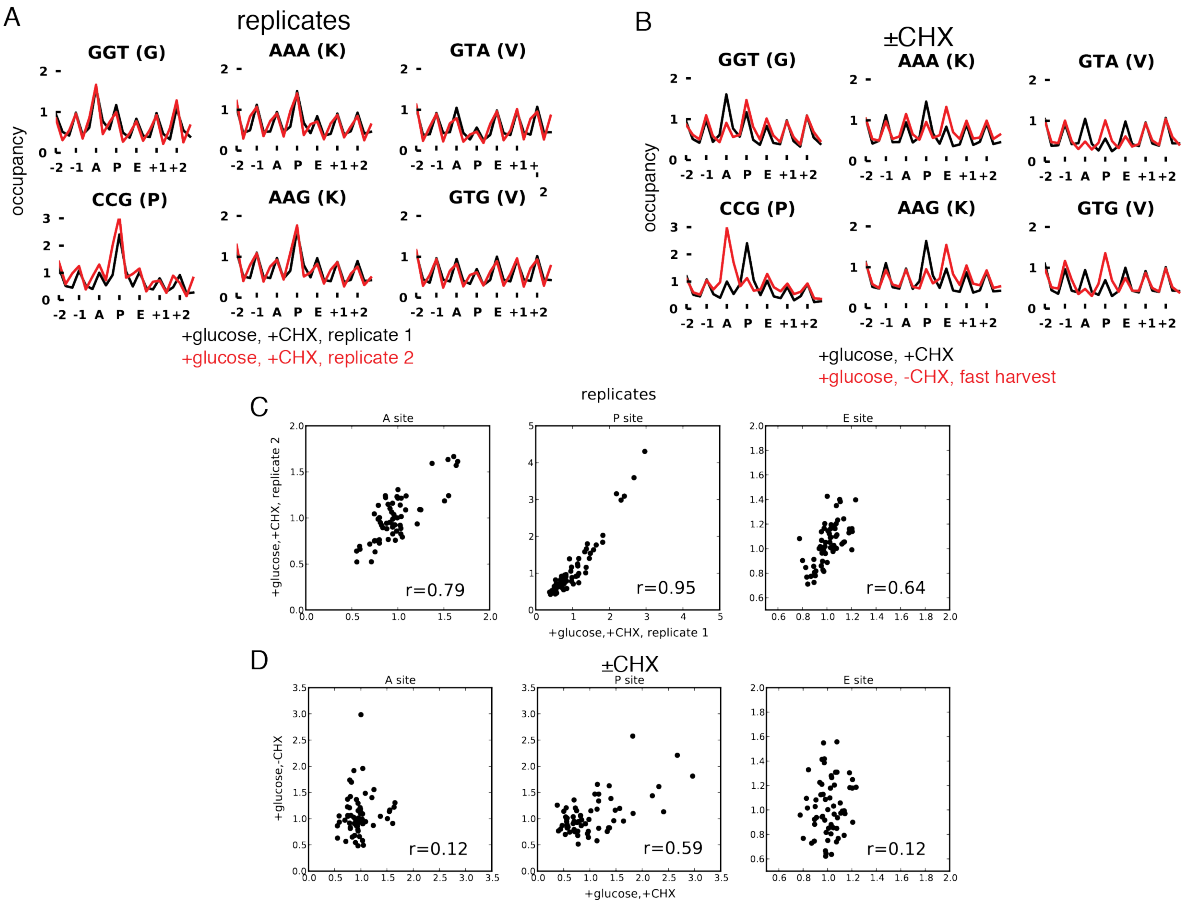

Figure I.3: CHX Altered the Codon-level Distribution of Ribosome Footprints
(A, B) Example codons that were strongly perturbed by in vivo incubation with CHX.
(C) Codon occupancy was reproducible between replicates. Codon occupancy values are the peak heights for each codon in a given site, as computed in panel A.
(D) CHX treatment caused global alterations in A, P and E site codon occupancy.

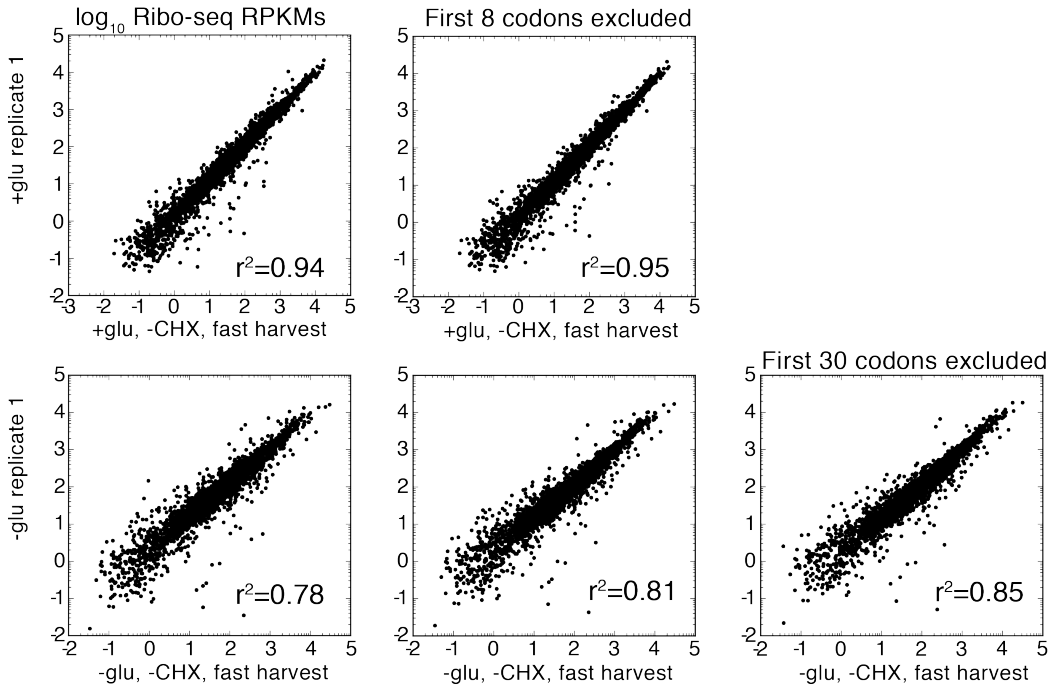

Figure I.4: Gene-level measures of translational activity are robust to CHX treatment.
RPKMS from +CHX and –CHX Ribo-seq libraries in both +glucose and –glucose conditions are well correlated.

CHX-dependent accumulation of ribosomes at uORFs

The robust CHX-dependent accumulation of ribosome footprints in TLs was of particular interest, as several studies have claimed that increased translation of upstream open reading frames (uORFs) occurs during stress responses or developmental phases (Brar et al., 2012; Gerashchenko et al., 2012; Ingolia et al., 2009; 2011; Lee et al., 2012). Moreover, it has been claimed that most of these translation events initiate at non-AUG codons, an altogether surprising finding, considering that the vast majority of previously-described initiation events require an AUG codon (Kozak, 1980), and the existence of a dedicated methionine tRNA that is required for canonical initiation (Lomakin et al., 2006). We found that in glucose-starved yeast, ribosome

footprints were increased over both AUG and near-AUG uORFs (Figure I.5, left panels), but this effect was not present in libraries prepared without CHX (Figure I.5, right panels). Some instances of non-AUG initiation found by Ribo-seq are likely to be real, as they can still be detected in the absence of elongation inhibitors (Ingolia et al., 2011), and many have ribosome density in our –CHX dataset. However, it appears that the change in uORF translation observed between conditions is largely an artifact of CHX treatment.

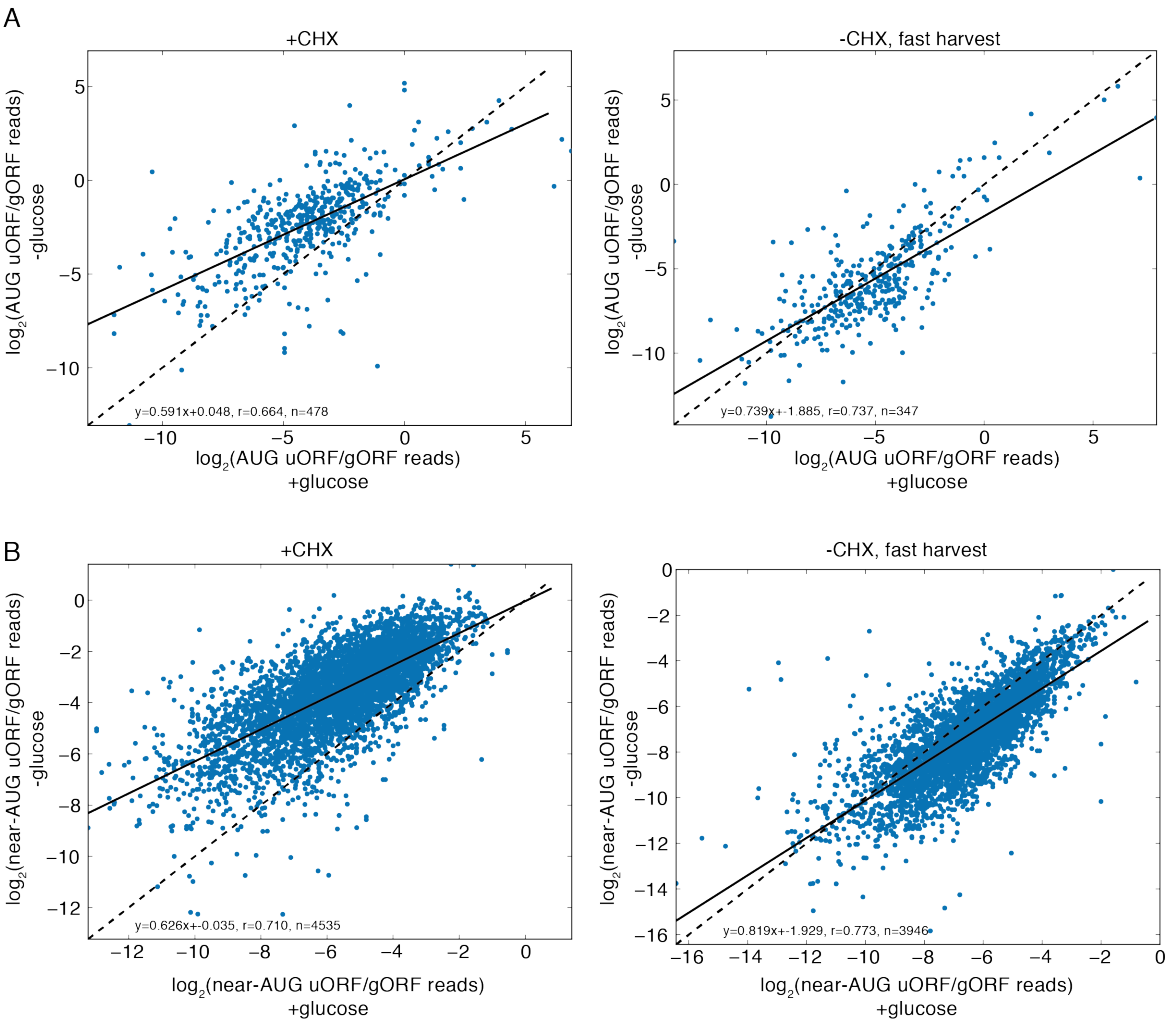

Figure I.5: CHX-dependent accumulation of ribosome footprints on uORFs in glucose-starved yeast.

(A) Comparison of the ribosome footprints mapping to each AUG uORF from (Brar et al., 2012), normalized to the footprints mapping to the corresponding genic ORF, between + and – glucose. The datasets for the left plot were produced with CHX preincubation, and the ones on the right were made without CHX. Solid lines are lines of best fit, and dashed lines are y=x.

(B) Same as (A), but for near-AUG uORFs.

We next considered the possibility that the some of the detected near-AUG uORFs are an artifact of the detection method. The large number of CHX-dependent ribosome footprints in the TL suggests that there may be a decrease in initiation fidelity during incubations with CHX. These sets of uORFs were identified by the presence of a bolus of ribosome footprints at their start codon (Brar et al., 2012), precisely the sort of artifactual accumulation that appears to be cause by CHX incubation, so we evaluated their translation potential via another analytical method. The ribosome release score (Guttman et al., 2013) relies on the fact that, unlike non-productive assemblies of 80S ribosomes and other artifacts, real translation events will show a sharp decline after the first in-frame stop codon. By comparing the ribosome density over the ORF to the density in its 3' UTR, the metric more reliably separates bona-fide noncoding RNAs from coding ones in metazoan cells (Chew et al., 2013; Guttman et al., 2013). We applied this metric to AUG and near-AUG uORFs. As positive controls, we used the annotated yeast genic ORFs, which are known to be translated with high confidence and should have very high ribosome release scores. As negative control, we generated a set of uORFs by translating *in silico* from random positions in TLs with no other known uORFs. In all datasets examined, genic ORFs had the highest ribosome release scores, followed by AUG uORFs, near-AUG uORFs, and then random uORFs (Figure I.6). This result suggests that there is some amount of translation of both AUG and near-AUG uORFs. However, the majority of near-AUG uORFs had ribosome release scores less than 1, and in the absence of CHX, their ribosome release scores approached those of the random uORFs (Figure I.6) arguing that most are not translated. The effect of CHX

on ribosome release scores likely comes from large CHX-dependent accumulations of ribosomes on the start codons of uORFs, which are a large fraction of the reads for these short ORFs. These results are consistent with previous findings that yeast AUG uORFs, but not near-AUG uORFs, are conserved, reduce translational efficiency and elicit nonsense-mediated mRNA decay (Arribere and Gilbert, 2013).

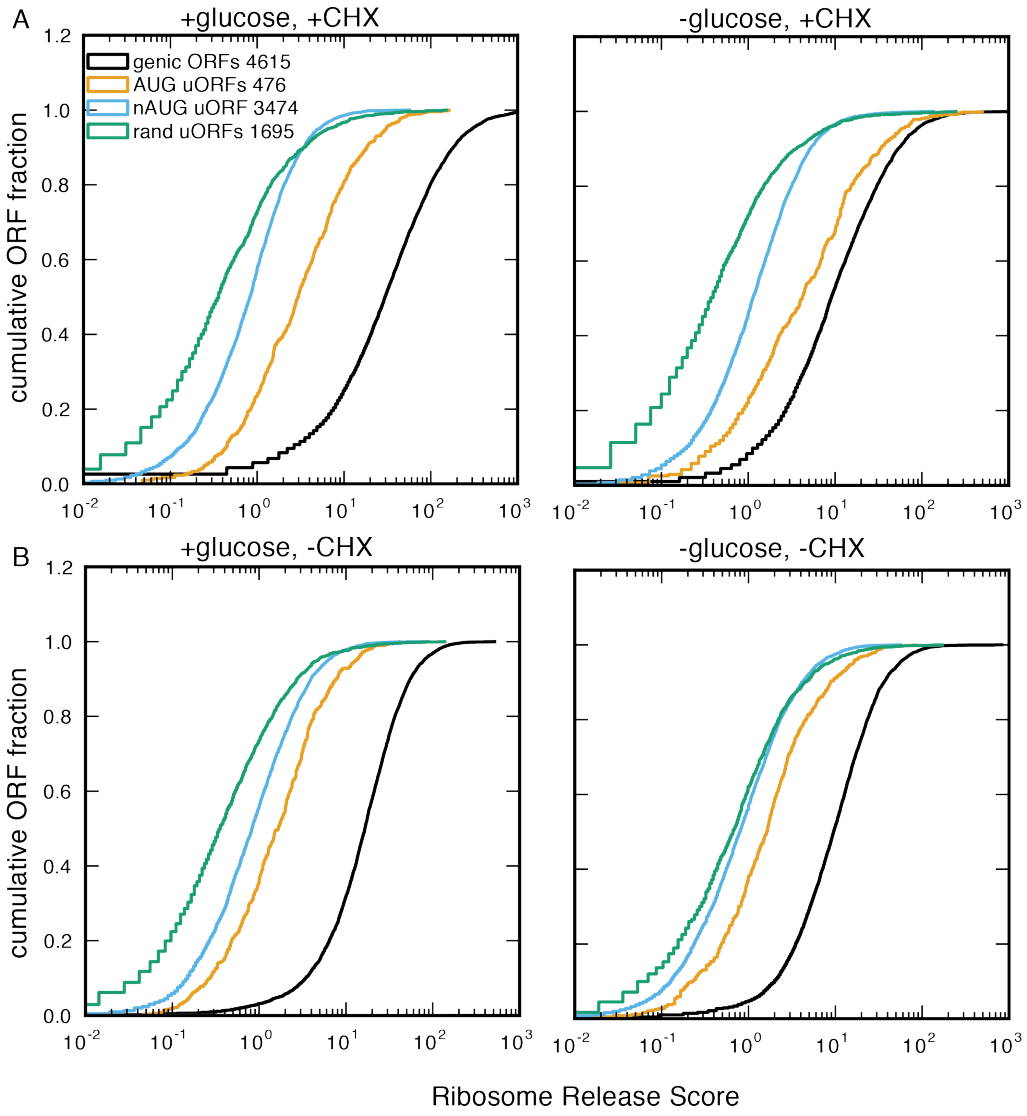

Figure I.6: Ribosome release scores suggest that a large fraction of putative near-AUG uORFs are not real translation events.
(A) Cumulative distributions of ribosome release scores for AUG uORFs, near aUG (nAUG) uORFs, genic ORFs, and a set of uORFs with randomly selected start codons (rand uORFs).
(B)Same as in (A), but for datasets lacking CHX.

CHX omission leads to ribosome accumulation at stop codons

Even though the use of CHX in the preparation of samples for Ribo-seq leads to some artifactual perturbations in ribosome position, it does not mean that datasets prepared without the drug accurately reflect the *in* vivo distribution of ribosomes. In fact, even with fast harvesting protocols taking less than a minute, there is sufficient time for yeast cells to inhibit translation initiation, leading to depletion of ribosomes from the 5' end of messages (Guydosh and Green, 2014; Ingolia et al., 2009; Lareau et al., 2014). To see if this runoff could affect the ribosome distribution at stop codons, which was previously shown to be perturbed by different translation inhibitors (Ingolia et al., 2011), we examined the ribosome footprints in a region around the stop codon. In samples prepared without CHX treatment, there is a strong accumulation of ribosomes at stop codons, especially in glucose-starved cells (Figure I.7). In addition, strong secondary and possibly even tertiary ribosome queuing events are visible upstream of the stop codon pause, at intervals of 28 nucleotides, where upstream ribosomes would be expected to stop if sterically blocked by a downstream ribosome (Wolin and Walter, 1988). The increased size of these pausing and queuing events in glucose-starved cells suggests that translation termination or ribosome recycling is slowed in this condition.

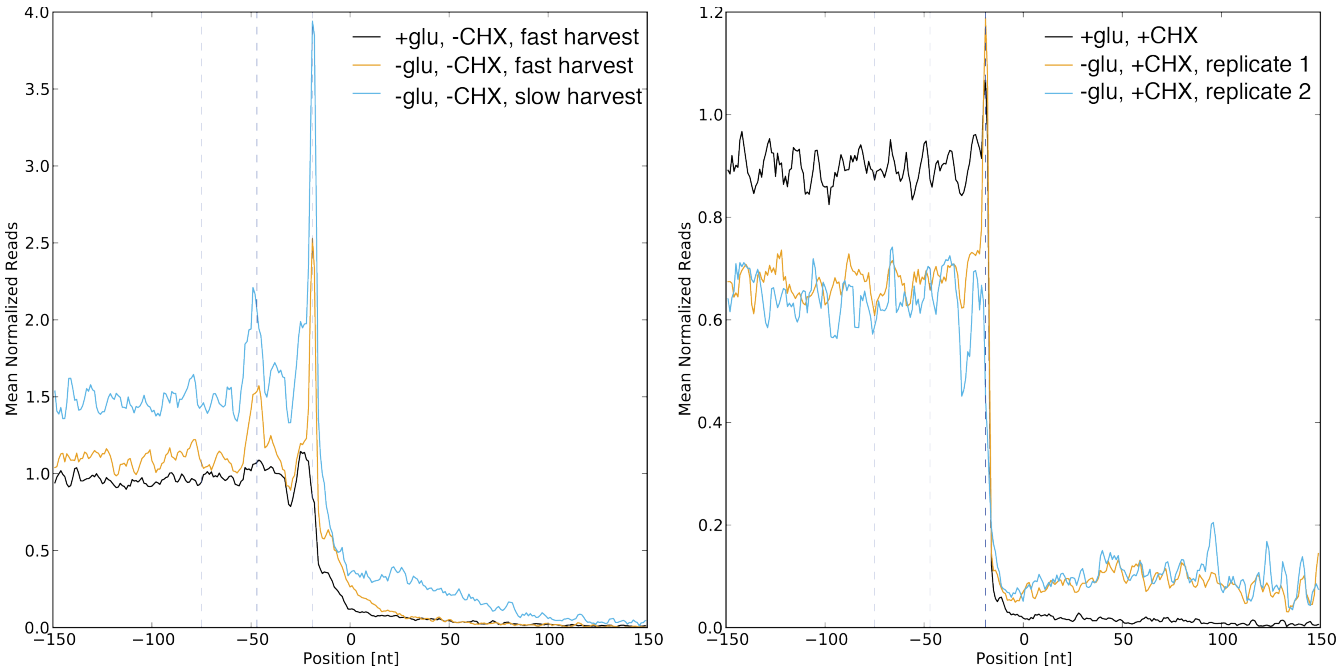

Figure I.7: CHX omission causes ribosomes to run into and queue upstream of stop codons.
Average read density for all genes, centered on stop codons, for libraries prepared with and without CHX. Vertical lines indicate the expected positions for reads corresponding to ribosomes with stop codons in their A site, and primary and secondary queuing events. For clarity, the 3nt periodicity of ribosome footprints was smoothed with a 3nt sliding window.

Discussion

In the present work we have shown that treatment of living cells with CHX causes non-physiological and condition-specific accumulations of ribosomes at start codons and in transcript leaders. This result has serious implications for any study comparing ribosome distributions or gene expression levels between conditions with varying levels of translational activity. As global modulation of translation rate is quite common in different stress conditions and environmental stages, great care must be taken to rule out inhibitor-specific effects when comparing two conditions, and controls without inhibitors should be performed in all experimental conditions or cell types, as the relative

magnitude of the effect is inversely proportional to the ribosome density on each mRNA. Similar effects of elongation inhibitors on ribosome distribution have now been reported in mammalian cells (Ingolia et al., 2011), although the effect on TL ribosomes appears to have been ignored. In light of these findings, several claims of widespread regulated translational events in TLs will need to be re-evaluated.

What is the cause of the spurious ribosome accumulations in TLs? After addition of CHX, there is nothing stopping continued initiation event, during the relatively long time in which the start codon is blocked by a CHX-bound ribosome. The initiation complex will thus spend an unusually long time in a relatively constrained region of the TL. Although the intrinsically high fidelity of the initiation machinery is presumably unaffected, it will sample a small set of sub-optimal positions many times, eventually initiation at a non-physiological position. This model may also have implications for messages that are elongation-limited, and the start codon is naturally blocked by a slow-moving ribosome.

We also found a perturbation in codon-specific ribosome distributions upon treatment with CHX, but it is not clear from this data which distributions best represent the *in vivo* state. In our datasets untreated with cycloheximide, we found substantial accumulations of ribosomes at stop codons, indicating that substantial ribosome movement is occurring during sample harvesting. This also indicates that, due to the rapid translational response of yeast to filtering or centrifugation, there is a need for technical improvement in the methods used to harvest yeast cultures.

Materials and Methods

Yeast Strains and Culture Conditions

Strains were in the Sigma 1278b background (MATa *ura3 leu2 trp1 his3)*. Yeast cultures were grown in YPAD (1% Yeast extract, 2% Peptone, 0.01% Adenine hemisulfate, 2% Dextrose) at 30˚C in baffled flasks with vigorous shaking. Glucose-starved cultures were prepared from YPAD cultures grown to OD600 = 1.0–1.1, harvested by centrifugation for 5 min at 12,000 x g, resuspended in prewarmed YPA medium lacking glucose and returned to shaking at 30°C for various times.

Ribo-seq and RNA-seq

Ribo-seq and RNA-seq were performed essentially as described (Brar et al., 2012; Ingolia et al., 2009). Except for the dataset labeled “slow harvest”, libraries without CHX were prepared by the fast freezing method in which yeast cultures were directly harvested on to a 0.45μm filter (Millipore) and immediately scraped into a tube containing liquid nitrogen to minimize polysome runoff. The time from culture to freezing was ~80-90 sec. 1.5ml of 1x PLB (20 mM HEPES-KOH pH 7.4, 2 mM $Mg(OAc)_2$, 100 mM KOAC, 1% Triton-X 100, 3 mM DTT) per gram of yeast was dripped into liquid nitrogen, and added to the cryomill (Retsch) with the frozen yeast before lysis. For the “slow harvest” sample, the yeast culture was collected by centrifugation. Subsequent library preparation steps were the same for samples with or without CHX, except for the omission of CHX in all buffers for -CHX libraries.

Data Analysis

Sequencing reads were processed as described previously (Ingolia et al., 2009) using Python scripts developed in-house. Genes for which there were at least 128 read counts across both replicate libraries in each condition were included. Metagene profiles were constructed by normalizing the Ribo-seq profile of each gene by the average Ribo-seq read density across the body of the entire transcript, including TL and 3'UTR (as annotated by (Xu et al., 2009)).

uORF annotations

uORF annotations were taken from (Brar et al., 2012), and mapped onto the Sigma 1278b genic ORF annotations, requiring that the sequence and location of the start codon relative to the genic start codon be maintained in both genomes.

Ribosome release scores

The metric used here is modified from (Guttman et al., 2013). We define the RRS as the ratio: (footprints per nt in the coding sequence)/(footprints per nt in the 3' UTR). Reads were offset by 12nt, the distance from the 5' end of a footprint to the ribosomal P site (Ingolia et al., 2009). For genic ORFs, the UTR is defined by the annotations of (Xu et al., 2009). For uORFs, the 3'UTR is defined as the region starting immediately after the uORF stop codon, up to the start codon of the next downstream uORF or genic ORF. Random uORFs were generating by choosing random start locations in the TLs

(annotated by (Xu et al., 2009)) of genes with no uORFs, and then translating to the first in-frame stop codon.

Accession Numbers

Sequencing data used for the analyses presented here can be found in the GEO database with accession number GSE51532.

Acknowledgements

We thank David Bartel for the initial suggestion that CHX causes accumulations at start codons, and members of the Gilbert lab for helpful discussions. The +CHX, +glu sequencing libraries were prepared by Mary Kay Thompson. This project was started as a collaboration with Pavan Vaidyanathan on work that was published in RNA (Vaidyanathan *et al. 2014)*. This work was supported by the NIH R01 (GM094303).

References

Arribere, J.A., and Gilbert, W.V. (2013). Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. Genome Research *23*, 977–987.

Ashe, M.P., De Long, S.K., and Sachs, A.B. (2000). Glucose depletion rapidly inhibits translation initiation in yeast. Mol. Biol. Cell *11*, 833–848.

Brar, G.A., Yassour, M., Friedman, N., Regev, A., Ingolia, N.T., and Weissman, J.S. (2012). High-Resolution View of the Yeast Meiotic Program Revealed by Ribosome Profiling. Science *335*, 552–557.

Chew, G.L., Pauli, A., Rinn, J.L., Regev, A., Schier, A.F., and Valen, E. (2013). Ribosome profiling reveals resemblance between long non-coding RNAs and 5' leaders of coding RNAs. Development *140*, 2828–2834.

Garreau de Loubresse, N., Prokhorova, I., Holtkamp, W., Rodnina, M.V., Yusupova, G., and Yusupov, M. (2014). Structural basis for the inhibition of the eukaryotic ribosome. Nature *513*, 517–522.

Gerashchenko, M.V., Lobanov, A.V., and Gladyshev, V.N. (2012). Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. Proceedings of the National Academy of Sciences *109*, 17394–17399.

Guttman, M., Russell, P., Ingolia, N.T., Weissman, J.S., and Lander, E.S. (2013). Ribosome Profiling Provides Evidence that Large Noncoding RNAs Do Not Encode Proteins. Cell *154*, 240–251.

Guydosh, N.R., and Green, R. (2014). Dom34 Rescues Ribosomes in 3' Untranslated Regions. Cell *156*, 950–962.

Han, Y., Gao, X., Liu, B., Wan, J., Zhang, X., and Qian, S.-B. (2014). Ribosome profiling reveals sequence-independent post-initiation pausing as a signature of translation. Cell Research *24*, 842–851.

Ingolia, N.T., Ghaemmaghami, 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.

Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell *147*, 789–802.

Ishimura, R., Nagy, G., Dotu, I., Zhou, H., Yang, X.L., Schimmel, P., Senju, S., Nishimura, Y., Chuang, J.H., and Ackerman, S.L. (2014). Ribosome stalling induced by mutation of a CNS-specific tRNA causes neurodegeneration. Science *345*, 455–459.

Kozak, M. (1980). Evaluation of the “scanning model” for initiation of protein synthesis in eucaryotes. Cell *22*, 7–8.

Lareau, L.F., Hite, D.H., Hogan, G.J., and Brown, P.O. (2014). Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. Elife *3*, e01257–e01257.

Lee, S., Liu, B., Lee, S., Huang, S.-X., Shen, B., and Qian, S.-B. (2012). Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. Proceedings of the National Academy of Sciences *109*, E2424–E2432.

Lomakin, I.B., Shirokikh, N.E., Yusupov, M.M., Hellen, C.U.T., and Pestova, T.V. (2006). The fidelity of translation initiation: reciprocal activities of eIF1, IF3 and YciH. Embo J. *25*, 196–210.

Qian, W., Yang, J.-R., Pearson, N.M., Maclean, C., and Zhang, J. (2012). Balanced Codon Usage Optimizes Eukaryotic Translational Efficiency. PLoS Genet *8*, e1002603.

Schneider-Poetsch, T., Ju, J., Eyler, D.E., Dang, Y., Bhat, S., Merrick, W.C., Green, R., Shen, B., and Liu, J.O. (2010). Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. Nat Chem Biol *6*, 209–217.

Shalgi, R., Hurt, J.A., Krykbaeva, I., Taipale, M., Lindquist, S., and Burge, C.B. (2013).

Widespread regulation of translation by elongation pausing in heat shock. Molecular Cell *49*, 439–452.

Stadler, M., and Fire, A. (2013). Conserved Translatome Remodeling in Nematode Species Executing a Shared Developmental Transition. PLoS Genet *9*, e1003739.

Vaidyanathan, P.P., Zinshteyn, B., Thompson, M.K., and Gilbert, W.V. (2014). Protein kinase A regulates gene-specific translational adaptation in differentiating yeast. RNA *20*, 912–922.

Wolin, S.L., and Walter, P. (1988). Ribosome pausing and stacking during translation of a eukaryotic mRNA. Embo J. *7*, 3559–3569.

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L.M. (2009). Bidirectional promoters generate pervasive transcription in yeast. Nature *457*, 1033–1037.

Zinshteyn, B., and Gilbert, W.V. (2013). Loss of a Conserved tRNA Anticodon Modification Perturbs Cellular Signaling. PLoS Genet *9*, e1003675.