Determinants of variation in mRNA degradation rates

by

Benjamin Samuel Neymotin

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Biology

New York University

January, 2015

_____

David Gresham, PhD

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

DEDICATION

To my family and friends, who allow me to always keep my priorities straight, and remind me to always keep perspective of the truly important aspects of life.

ACKNOWLEDGEMENTS

As I sat in a courtroom for jury duty, a discussion arose about the proper way to raise a child. I vividly remember when one mother stood up and announced, “it takes a village to raise a child”. This phrase resonated deeply with me. It suggested that only through community effort could a child become a respectable and responsible adult. In reflecting on my years at NYU, I can’t help but think that as I complete my PhD, the appropriate phrase is “it takes a department to raise a scientist”.

Over the course of more than five years, I have interacted with a tremendous number of individuals, each of whom contributed to my development as a scientist. Although impossible to list each one, there are some who need special acknowledgement.

First and foremost, I wish to thank the current and previous Chairs of the department, Steve Small and Gloria Coruzzi. Under their leadership, the department has grown tremendously, with each new faculty member bringing great science, fresh ideas, and spectacular enthusiasm. More importantly, Steve and Gloria were extremely kind to me, offering help and advice on several personal matters over the past few years. In addition, Gloria allowed to me rotate in her laboratory, which exposed me to the incredible field of genomics and systems biology. I am truly grateful for their guidance.

Edo Kussell introduced me to NYU in the summer of 2007. He allowed me to work in his laboratory, and sparked my interest in metagenomics. Were

it not for the great experience I had that summer, I'm not sure I would have known to apply to NYU two years later. I am eternally thankful.

Dave Fitch, Rich Bonneau, Patrick Eichenberger, and Jon Warner all served on my thesis committee throughout the duration of my PhD. With each committee meeting they offered me encouragement and ideas of how to improve my research project. Jon introduced me to approach to equilibrium kinetics, which eventually became the basis for RATE-seq. Dave always kept me focused on the big picture and offered great suggestions, Rich provided critical insight, and Patrick brought a wealth of knowledge about different aspects of microbiology and biology in general. In addition, Patrick allowed me to rotate in his lab. To my committee members, I am grateful.

To all of the people I worked with directly on wetlab and computational projects, and those who offered sagely advice, I am extremely thankful. This group includes Gabriel Krouk, Peter McKenney, John Burns, and Rebecca Davidson. I am especially thankful to Rebecca, who helped me get started in programming, and continues to be extremely helpful, not only for her knowledge, but more importantly for her friendship.

In the Gresham Lab I was privileged to work with many great scientists including Niki Athanasiadou, Nathan Brandt, Jungeui Hong , Naomi Ziv, Darach Miller, and Stephanie Lauer. Niki was extremely helpful in offering advice both in experimental biology and bioinformatics. She often serves as a guiding light at the end of a dark tunnel. Nathan single handedly maintained the lab's

homeostasis. It is hard to imagine successful completion of my PhD without his innumerable contributions to the lab, and specifically my projects. Jungeui was the computational ninja of the lab, solving any problems that we faced. Naomi was the critical analyst. She was always able to pinpoint the strengths and weaknesses of every presentation and paper, and offer extremely helpful suggestions for enhancement. Darach was always willing to provide perfectly timed puns and jokes. He has extremely exciting ideas for projects, and will no doubt produce great things in the near future. Steph is the newest member of the lab, but her presence already permeates throughout. Her warm nature and exuberance make everyone feel a little bit happier. Victoria Ettorre and Christina Nunez were undergrads who allowed me to be their mentor and gain experience serving in a leadership role. To all of them, I am extremely grateful.

No PhD student can succeed without strong mentorship. My mentor, David Gresham, was an outstanding source of knowledge, guidance, confidence, and inspiration. David's passion for understanding complex issues in science, as well as his broad interests in biology, results in many novel ideas and really great science. More importantly, he cares about the success of his PhD students. Numerous times I was downtrodden by the difficulties of my project, and each time he was able to suggest a means of overcoming seemingly insurmountable obstacles. I was very lucky to have such a mentor, and I know he will continue to invest much time and energy into current and

future PhD candidates. No doubt, the lab is destined for great success and achievement.

Outside of lab, I am fortunate to have a family who is supportive and caring. All of my in-laws are continuously concerned about my health and well-being. My father-in-law has helped me out many times in difficult situations over the past few years. My parents always help put things in perspective and offer invaluable life experience and wisdom. My sisters give me practical perspective on the many important aspects of life. My Grandfather, who unfortunately passed away in the middle of my PhD studies, always emphasized the importance of hard work, and the fact that nothing in life comes without great effort and drive. To all of them I am continually grateful.

Finally, and most importantly, I wish to thank my wife Leah, and my daughter Leora. They are the ones who have had to deal with my anxiety and stress, and they are the ones who have kept me going. No matter the day, no matter the situation, they always welcome me home with a hug, kiss and unconditional love and affection. This dissertation is no less their achievement than it is mine. It proves that together, we can achieve anything with the support and care of loved ones.

ABSTRACT

The expression level of a gene in the form of RNA is determined by its rates of synthesis and degradation. Together, these two processes determine the overall abundance of a transcript and through these processes, a cell can respond to changes in environmental conditions. While many investigations have focused on how particular changes to environment affect overall RNA levels, few have determined the underlying mechanism governing those changes. In the absence of kinetic analyses of mRNA degradation, it is impossible to fully understand gene regulation at its most basic level. Therefore, it is imperative to uncouple the processes of synthesis and degradation in any analysis in order to realize the complex working of cellular responses.

In this dissertation, I have investigated how a cell regulates gene expression post-transcriptionally. In chapter 1, I describe the importance of post-transcriptional gene regulation, and review the current understanding of degradation pathways and cellular processes that may affect RNA degradation. In chapter 2, I describe the development and application of a method, RATE-seq, which enables simultaneous measurement of RNA synthesis and degradation rates genome-wide. Using these data, as well as other genome-wide datasets of gene specific parameters, I explore the potential factors responsible for that variation in Chapter 3. In Chapter 4, I analyze the role of

cellular growth rate in rates of RNA synthesis and degradation. Finally, in Chapter 5, I give a perspective on future studies of interest in post-transcriptional gene regulation and specifically mRNA degradation. Together, these studies describe determinants of degradation rates for different mRNA transcripts and they suggest areas of interest for future studies of mRNA kinetics.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABREVIATIONS

DOX: Doxycycline

ERCC: External RNA Controls Consortium

GFP: Green fluorescent protein

GO: Gene Ontology

Biotin-HPDP: N- [6-(Biotinamido)hexyl]-3′-(2′-pyridyldithio)-propionamide

M-MLVRT: Moloney Murine Leukemia Virus Reverse Transcriptase

qRT-PCR: quantitative reverse transcription polymerase chain reaction

RATE-seq: RNA Approach to Equilibrium Sequencing

RNABP: RNA binding protein

UDP: Uridine diphosphate

UMP: Uridine monophosphate

UPRT: Uracil Phosphoribosyl Transferase

UTR: Untranslated region

4sU: 4-thiouridine

4tU: 4-thiouracil

CHAPTER 1: INTRODUCTION

1.1: Post-transcriptional control of gene expression

In response to extracellular signals, a cell must regulate the expression of thousands of genes. The DNA of a gene is first transcribed into RNA, and the RNA is then processed and translated into protein. Each of these steps is subject to extensive regulation and there is extensive complexity at each of these different stages. Although transcription is the first step in gene expression, it is a relatively rare and slow event (Pelechano et al. 2010), suggesting that rapid responses to environmental signals require additional levels of regulation. Indeed, some responses occur so rapidly that they can only be explained by post-transcriptional regulatory mechanisms (Elkon et al. 2010). Consequently, an understanding of gene expression regulation requires an understanding of how mRNA is post-transcriptionally regulated.

Once an mRNA is processed into its mature state, it enters the cytoplasm where it is translated, stored in a translationally repressed state, or degraded (**Figure 1.1**). If the mRNA is to be translated, it is first bound by different translation initiation factors which bring the 5' and 3' ends of the message in close proximity. The 40S ribosomal subunit finds the mRNA cap and scans the transcript for a start codon (**Figure 1.2**). Once found, the large 60S subunit binds and the fully formed ribosome proceeds to translate the message (Lodish et al. 2008). Because the two ends of the message are physically connected when in such conformation, ribosomes that have

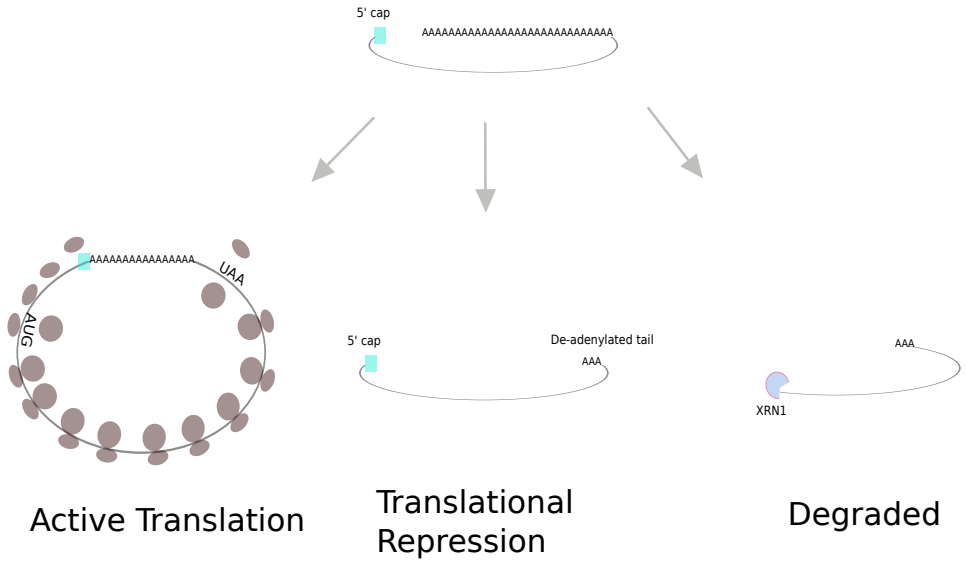


Figure 1.1: The state of an mRNA when in the cytoplasm
Once a mature mRNA leaves the nucleus, it can be translated by ribosomes (left), repressed in subcellular locations (middle), or degraded (right).

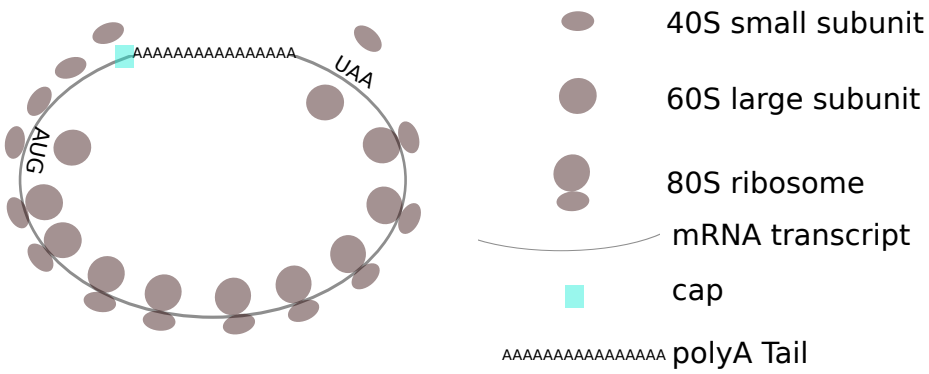


Figure 1.2: Translation of mRNA
The mRNA transcript is translated once the fully formed ribosome is assembled at the first start codon. For simplicity, the translation initiation factors and the resulting polypeptides are omitted.

completed translation can rapidly find the cap and initiate translation anew. When not in such conformation the mRNA can either be stored in a translationally repressed manner or degraded.

Translational repression and storage of mRNA is a key step in many biological processes. In the early developing embryo, maternally loaded transcripts are initially in oocyte processing bodies where they lack a polyA tail and they are translationally repressed (Wilt 1973). Following fertilization, the mRNA is polyadenylated, ribosomes bind, and the transcripts engage in active translation. The conversion of these maternal transcripts to a translationally active state is imperative since the early zygote is transcriptionally silent (Stitzel and Seydoux 2007) and all initial protein products are a result of these transcripts (Schier 2007). In addition, the morphogen gradient that establishes the polarity and body structure of the organism depends on translational repression of certain mRNAs in space and time (Lodish et al. 2008). In the absence of translational repression, the developing embryo is compromised. Translational repression is common in many organisms, often with mRNA stored in cytoplasmic granules referred to as processing bodies (p-bodies) (Brengues et al. 2005). mRNAs stored in p-bodies can re-enter the translational pool for at least a subset of the transcriptome (Arribere et al. 2011). P-bodies are dynamic structures, that appear or disappear, depending on a variety of extracellular conditions (Teixeira et al. 2005). They are also the location of many components of the

degradation machinery (Sheth and Parker 2003), consistent with the observation that the first steps in mRNA decay involve translational repression (Coller and Parker 2005; 2004). Although translational repression may lead to an eventual return to translation, in most cases it is thought to be the first step in degradation of the transcript.

Ultimately, whether an mRNA is translationally active or repressed, it is degraded. Once degraded, the transcript can only be replaced through new transcriptional events, which are rare and costly. The cell must therefore carefully control initiation of the irreversible step of degrading a particular mRNA. How the cell makes such a critical decision for different species of mRNA is the focus of this thesis.

1.2: The factors controlling mRNA degradation

The degradation pathways of mRNA have been established based on individual transcript studies primarily in *Saccharomyces cerevisiae* (**Figure 1.3**). Most eukaryotic mRNAs contain both a cap and polyA tail that assist in preventing mRNA degradation (Furuichi et al. 1977; Coller et al. 1998). When targeted for degradation, an mRNA is first deadenylated (Muhlrad and Parker 1992) by the CCR4 deadenylase (Tucker et al. 2001), decapped by the DCP1/DCP2 decapping enzymes (Stevens 1980a; Beelman et al. 1996), and then degraded in the 5'→3' direction by the XRN1 exonuclease (Larimer et al. 1992) to 5' mononucleotides (Stevens 1980b). In a second degradation

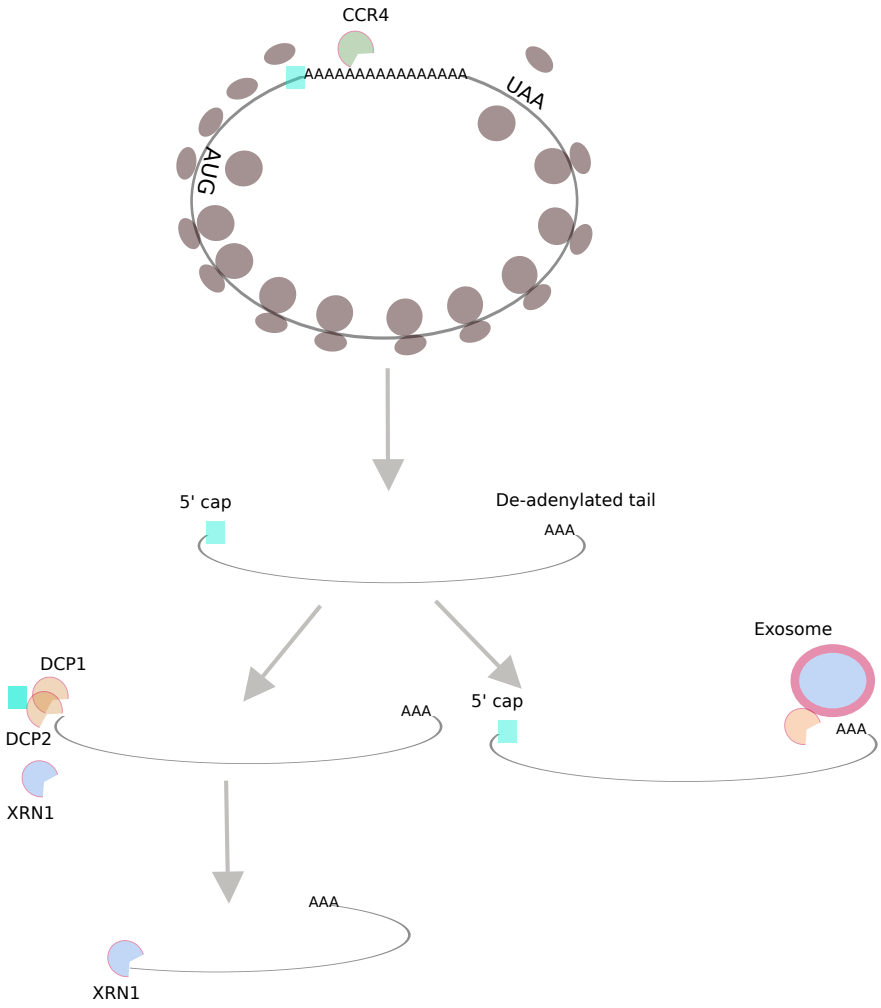


Figure 1.3: The degradation pathways of normal mRNA
Following deadenylation by CCR4, the transcript is either decapped by DCP1/DCP2 and then degraded by XRN1 in the 5'→3' direction, or it is degraded in the 3'→5' direction by the exosome.

pathway, the exosomal complex degrades mRNA in the 3'→5' direction (Mitchell et al. 1997; Anderson and Parker 1998; Schmid and Jensen 2008). Interestingly, 5'→3' degradation machinery is localized in the same concentrated p-bodies where mRNAs are translationally repressed. Because p-body size and number vary as a function of environmental conditions (Teixeira et al. 2005), this suggests the overall translational status and degradation status of the pool of mRNA changes as well. Degradation pathways do not appear to be redundant, and there are differences in specificity of mRNA transcripts targeted (Parker 2012; Sun et al. 2012; 2013; Houalla et al. 2006; He et al. 2003). The pathways and players in mRNA degradation are now well characterized and described in exquisite detail (Parker 2012).

Studies of individual transcripts have shown that rates of mRNA degradation differ between transcripts (Harpold et al. 1981; Herrick et al. 1990), but the determinants of this variation were initially unclear. Early studies of RNA decay suggested that translation plays a role in the rate at which transcripts degrade (Singer and Penman 1972; Graves et al. 1987; Wisdom and Lee 1991; Stimac et al. 1984). Subsequent work has shown that the primary sequence of the mRNA transcript plays a major role in its degradation kinetics. These sequence determinants can be found in the coding region (Caponigro et al. 1993), as well as the 3' and 5' untranslated regions (UTR) (Shaw and Kamen 1986; Muhlrad and Parker 1992; Gupta et al. 2014).

More recently, promoter elements have been shown to effect mRNA decay rates (Bregman et al. 2011; Trcek et al. 2011). In addition, RNA binding proteins (Olivas and Parker 2000; Grigull et al. 2004; Chen et al. 2001; Coller et al. 1998) as well as signaling pathways involved in response to extracellular conditions (Albig and Decker 2001), appear to impact the degradation kinetics of different transcripts. These observations suggest mRNA decay rates are determined by a combination of many different factors. Although informative, most of these studies were conducted for individual transcripts. To understand transcriptome-wide effects of the multiple factors affecting mRNA degradation, mRNA decay analyses must be performed on a genome-wide scale.

1.3: Global methods for measurement of mRNA degradation

A number of methods exist for measuring mRNA degradation on a global scale, but they all rely on the same general principle. Since mRNA abundance is determined by synthesis and decay, then if synthesis is inhibited, degradation can be measured by tracking the decrease in transcript abundance with time. By inhibiting RNA polymersase II (RNA Pol II), mRNA synthesis can be halted. Most early analyses of RNA degradation used chemical inhibitors of different steps of transcription, including thiolutin (Tipper 1973), which inhibits transcription initiation and α-amanitin, and Actinomycin (Schindler and Davies 1975) which inhibit transcription elongation. A major advance in transcriptional inhibition was introduced with

temperature sensitive alleles of RNA Pol II (Nonet et al. 1987). The inhibition of transcription using these different methods enabled the direct measurement of mRNA decay for individual transcripts (Herrick et al. 1990) and when combined with microarray technology, on the genome-wide scale as well (Wang et al. 2002; Grigull et al. 2004; Shalem et al. 2008).

While these analyses were extremely informative, one major issue is the relevance of the data to physiological conditions. When transcription is inhibited, cells rapidly die (Nonet et al. 1987). Therefore, the results may not accurately reflect what is actually occurring in living cells. One of the methods, thiolutin, has actually been shown to effects the degradation machinery itself, systematically overestimating the stability of transcripts (Pelechano and Pérez-Ortín 2008). Temperature sensitive alleles require a shift to a higher temperature, which inevitably results in changes to mRNA folding, which is thought to affect degradation (Muhlrad and Parker 1992). Interestingly, one study found that each method of transcriptional inhibition affects subsets of genes differently (Grigull et al. 2004). This suggests that the method used for transcriptional inhibition plays a major role in the resulting estimates of mRNA decay rates. Clearly a different method, which does not affect cellular physiology, is much desired.

An alternative method to transcriptional inhibition is the discrimination of newly synthesized transcripts through use of labeled RNA precursors, usually uracil or uridine. In the case of uracil, certain cell types

can salvage extracellular uracil from the environment and rapidly convert it into uridine monophosphate in a single step (**Figure 1.4**). Such enzymes exist in protozoa (Cleary et al. 2005), fungi (Kern et al. 1990), and plants (Mainguet et al. 2009). In higher level organisms, similar conversions are feasible using uridine, which is transported into the cell and then phosphorylated to uridine monophosphate. Early kinetic studies used radiolabled uracil and uridine for analysis of decay rates of mRNA pools and specific transcripts (Greenberg 1972; Wiegers et al. 1975; Harpold et al. 1981; Kim and Warner 1983). By estimating the specific activity of the RNA, it is possible to determine how the abundance of labeled transcripts changes with time. While useful for studying individual transcripts, radiolabeling of transcripts cannot be used on a genome wide scale.

Analysis of transcriptome-wide mRNA degradation rates requires an alternative to radiolabeled uracil that enables fractionation of new from existing transcripts. The uracil analogue 4-thiouracil (4tU) offers such an approach. Following addition of 4tU, the number of labeled transcripts increases with time as with radiolabeled uracil. The thiolated transcripts can subsequently be conjugated to a molecule that will form a disulfide bond with the thiolated nucleotide. Inclusion of a biotin in this molecule allows for separation with streptavidin magnetic beads. Such an approach with either 4tU or 4-thiouridine (4sU) has been utilized in a variety of organisms and cell

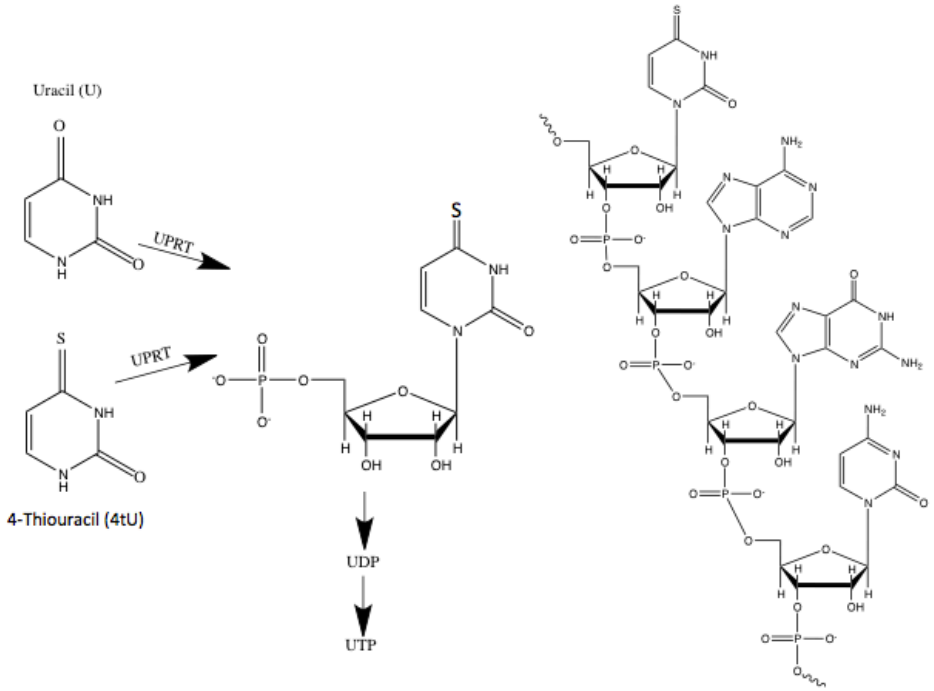


Figure 1.4: The use of labeled nucleotides for metabolic labeling of mRNA
When provided uracil in the media, yeast cells will convert the uracil nucleobase into uridine monophosphate (UMP)(left) via the Uracil phosphoribosyl transferase (UPRT) enzyme. The UMP will then be converted into Uridine diphosphate (UDP) and uridine triphosphate (UTP). If a labeled nucleotide such as thiouracil is provided in the media, then the labeled nucleotide will be incorporated into newly synthesized mRNA transcripts (right).

types (Cleary et al. 2005; Dölken et al. 2008; Rabani et al. 2011; Munchel et al. 2011). Use of metabolic labeling of RNA transcripts enables uncoupling of synthesis from degradation. When carefully controlled, the effect on cellular physiology is minimal. In combination with high throughput sequencing, RNA degradation can be quantitatively measured, enabling a global picture of mRNA degradation rates.

1.4: The effect of cellular growth rate on transcriptome dynamics

The abundance of an mRNA species varies with environmental conditions (Gasch et al. 2000). However, changes in environmental conditions frequently lead to concomitant changes in cellular growth rate. A critical question regarding mRNA kinetics is the role of cellular growth rate in transcriptome dynamics.

To investigate the role of growth rate on transcriptome dynamics, the rate at which cells grow needs to be modulated. Growth rate control in batch cultures is performed by use of different molecular forms of essential nutrients. For example, glucose and galactose as carbon sources support growth rates that differ by a factor of two (Tyson et al. 1979). Although growth rate has been controlled, the secondary effects of the different nutrients on metabolism and physiology confound interpretation of results. Therefore, a different method for the control of growth rate is desirable.

The chemostat is a system in which growth rate is controlled in defined environments (Hoskisson and Hobbs 2005; Ziv et al. 2013a). The chemostat

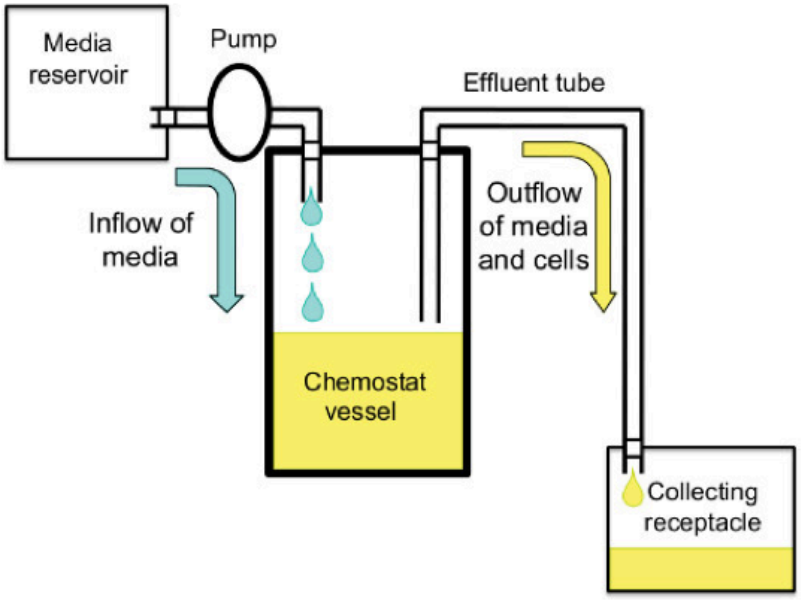


Figure 1.5: The use of chemostat cultures to control growth rate
In the chemostat, fresh media is diluted into growing cultures of cells. The growth rate is then controlled via the dilution rate. When the culture is diluted at a high rate, then the cells maintain a fast growth rate. When the culture is diluted at a slow rate, then the cells maintain a slow growth rate. Figure is taken from (Ziv et al. 2013a).

provides a constant supply of new media, while expelling old media and cells (**Figure 1.5**). By controlling the rate at which the new media is supplied, different steady state growth rates are attained using the same media composition. Previous studies using the chemostat have shown that a large fraction of the transcriptome is regulated in a growth rate dependent manner (Regenberg et al. 2006; Castrillo et al. 2007; Brauer et al. 2008). In fact, irrespective of the nutrient limitation used to control cell growth, more than a quarter of the yeast transcriptome is up or downregulated as a simple linear function of growth rate (Brauer et al. 2008).

Despite the importance of regulated cell growth and its effect on gene expression programs, little is known about how variation in RNA synthesis and degradation are modulated as a function of growth rate. Studies in bacteria of several individual transcripts showed that mRNA stability changes when doubling time is modulated using different sources of carbon (Nilsson et al. 1984). However, because transcription was inhibited in this study by chemical inhibition, cells are not actually growing when degradation rates were measured. In addition, growth rate regulated using different molecular forms of carbon cannot discriminate effects of nutrient versus effects of growth rate. Control of growth rates using chemostats provides the ideal system for addressing the critical question of how mRNA kinetics change as a function of growth rate.

1.5: Determinants of variation in mRNA degradation rates

The expression level of a gene in the form of RNA is determined by its rates of synthesis and degradation. Together, these two processes determine the overall abundance of a transcript and through these processes, a cell can respond to changes in environmental conditions. While many investigations have focused on how particular changes to environment affect overall RNA levels, few have determined the underlying mechanism governing those changes. In the absence of kinetic analyses of mRNA degradation, it is impossible to fully understand gene regulation at its most basic level. Therefore, it is imperative to uncouple the processes of synthesis and degradation in any analysis in order to realize the complex working of cellular responses.

In this dissertation, I have investigated how a cell regulates gene expression post-transcriptionally. In chapter 1, I described the importance of post-transcriptional gene regulation, and reviewed the current understanding of degradation pathways and cellular processes that may affect RNA degradation. In chapter 2, I describe the development and application of a method, RATE-seq, which enables simultaneous measurement of RNA synthesis and degradation rates genome-wide. Using these data, as well as other genome-wide datasets of gene specific parameters, I explore the potential factors responsible for that variation in Chapter 3. In Chapter 4, I analyze the role of growth rate in rates of RNA synthesis and degradation. Finally, in Chapter 5, I give a perspective on future studies of interest in post-

transcriptional gene regulation and specifically mRNA degradation. Together, these studies describe determinants of degradation rates for different mRNA transcripts and they suggest areas of interest for future studies of mRNA kinetics.

CHAPTER 2: DETERMINATION OF IN-VIVO RNA KINETICS USING RATE-SEQ

This chapter was published in the journal *RNA* as:

Neymotin B, Athanasiadou R, Gresham D. 2014. Determination of in vivo RNA kinetics using RATE-seq. *RNA* **20**: 1645–1652.

Supplemental tables of genome-wide measurements are available on the *RNA* website

2.1: Abstract

The abundance of a transcript is determined by its rate of synthesis and its rate of degradation; however, global methods for quantifying RNA abundance cannot distinguish variation in these two processes. Here, we introduce RNA approach to equilibrium sequencing (RATE-seq), which uses in vivo metabolic labeling of RNA and approach to equilibrium kinetics, to determine absolute RNA degradation and synthesis rates. RATE-seq does not disturb cellular physiology, uses straightforward normalization with exogenous spike-ins, and can be readily adapted for studies in most organisms. We demonstrate the use of RATE-seq to estimate genome-wide kinetic parameters for coding and noncoding transcripts in *Saccharomyces cerevisiae*.

2.2: Introduction

Remodeling of gene expression is critical for a broad range of biological processes from the cell division cycle and embryo development (Schier 2007) to cellular responses to extracellular signals (Gasch et al. 2000). Regulation of

transcript abundance is controlled by the combined action of transcript synthesis and transcript degradation. Although the regulation of transcript synthesis has historically been the primary focus of investigation, there is accumulating evidence that RNA degradation plays an important role in dynamic biological processes (Elkon et al. 2010). A comprehensive understanding of the regulation of gene expression programs, and the development of mathematical models that explain the dynamics of gene expression, requires the accurate estimation of absolute rates of both RNA synthesis and RNA degradation in vivo.

A variety of high-throughput methods have been introduced with the goal of estimating in vivo rates of either RNA synthesis or degradation. Genomic run on assays (García-Martínez et al. 2004) provide a means of estimating mRNA synthesis rates; however, these methods require isolation of nuclei or permeabilization of cells, which likely compromises the physiology of cells. Until recently, mRNA decay rates have been estimated using transcriptional inhibition (Wang et al. 2002; Grigull et al. 2004; Shalem et al. 2008) using either temperature-sensitive alleles of RNA polymerase II or chemical inhibition of transcript production. While these methods succeed in inhibiting transcript synthesis, they typically result in a stress response or cellular death (Nonet et al. 1987) resulting in the estimation of mRNA decay rates that may have little physiological relevance.

Recently, methods using in vivo metabolic labeling of mRNAs (Cleary et al. 2005; Dölken et al. 2008) have been introduced using either the nucleobase 4-thiouracil (4tU) or nucleoside 4-thiouridine (4sU), which introduce a reactive thiol group into RNAs. Following RNA purification, the presence of a thiol group in RNAs enables conjugation to N-[6-(Biotinamido)hexyl]-3′-(2′-pyridyldithio)-propionamide (biotin-HPDP) and subsequent fractionation using streptavidin-coated magnetic beads. Genome-wide estimation of in vivo kinetic parameters using metabolic labeling of RNA with 4tU has been reported using different experimental designs. Pulse-chase labeling with 4tU (Munchel et al. 2011) represents a promising approach to estimating mRNA degradation rates. However, internal recycling of labeled nucleotides (Puckett et al. 1975; Nikolov and Dabeva 1985) may result in an incomplete chase thereby confounding the estimation of mRNA degradation rates. Alternatively, comparative Dynamic Transcriptome Analysis (cDTA) (Sun et al. 2012)(an updated version of Dynamic Transcriptome Analysis [DTA]) (Miller et al. 2011)estimates rates of mRNA degradation by determining the ratio of labeled to total RNA using hybridization to a DNA microarray at a single time point following addition of 4tU. However, cDTA requires the manufacture of customized dual species DNA microarrays to normalize hybridization signals, and relies on a single time-point after labeling, which may not accurately capture kinetic parameters. Indeed, the use of different individual time points has a significant effect on the estimated degradation rates for at least a subset

of transcripts (Dölken et al. 2008), which is likely the case for similar approaches using RNA-seq (e.g., (Schwanhäusser et al. 2011)).

Here, we report a general method for accurate measurement of absolute RNA kinetic parameters in vivo. We use approach to equilibrium labeling (Greenberg 1972), which minimizes exposure of cells to 4tU and is not affected by nucleotide recycling. We undertook a series of rigorous controls to optimize each step of the RATE-seq method. By using strand-specific sequencing (Parkhomchuk et al. 2009) in combination with ribosomal depletion, we measured rates of decay for a variety of different types of RNA, including noncoding RNA and snRNA. We developed a normalization method using multiple spike-in RNAs that also enables identification and correction for technical artifacts. To account for the nature of the data (i.e., overdispersed count data in which the variance is greater than the mean) in model fitting we used a weighted nonlinear regression to estimate parameters. We used RATE-seq to define the regulatory landscape of steady-state transcript levels, defined as a function of the underlying kinetic parameters genome-wide, and find that many transcripts in budding yeast have similar steady-state levels but differ greatly in their rates of production and degradation. RATE-seq can be readily implemented in any organism, making it a generally applicable method for characterizing the steady-state in vivo kinetics of RNA with unprecedented resolution.

2.3: Results

2.3.1: Thiouracil labeling follows approach to equilibrium kinetics

The rate of change in RNA abundance (d[RNA]/dt) can be modeled as a function of a constant rate of synthesis (k) and a degradation rate proportional to RNA abundance (α[RNA]) using the relationship d[RNA]/dt = k-α[RNA]. If a labeled nucleotide is added to the culture the concentration of labeled transcript will increase with time to an equilibrium value at a rate solely determined by the transcript's degradation rate constant (α_{RNA}) and the cells' division rate constant (α_{growth}) (i.e., $\alpha = \alpha_{RNA} + \alpha_{growth}$). Approach to equilibrium labeling, using radiolabeling, was developed over 40 yr ago to estimate the rate of total mRNA turnover (Greenberg 1972) and was subsequently used to study individual transcripts using transcript-specific probes (Harpold et al. 1981; Kim and Warner 1983). To apply approach to equilibrium labeling on a genome-wide scale we developed a method using 4tU-labeling and RNA-seq (**Figure 2.1A**). Our method relies on the presence of an endogenous copy of uracil phosphoribosyltransferase (UPRT) in *Saccharomyces cerevisiae* (encoded by FUR1), which converts 4tU into 4-thiouridine monophosphate allowing its incorporation in RNA. For the purpose of normalizing RNA-seq libraries from different time points following labeling, we added a constant quantity of three different in vitro-transcribed thiolated transcripts (Supplemental Table 2.S1) to isolated RNA prior to fractionation. As deadenylated transcripts can persist in the cytoplasm or be readenylated in some species (Wilt 1973), we used rRNA depletion rather than poly(A)

fractionation. The lack of poly(A) selection step also enables the analysis of both coding and noncoding transcripts.

We first tested the efficiency of 4tU incorporation in *S. cerevisiae* and its physiological consequences. Consistent with previous reports (Munchel et al. 2011) we find that yeast cells take up 4tU provided in the growth medium and incorporate it into RNA (Supplemental **Fig. 2.S1**). However, we find that cells lacking a functional uridine monophosphate biosynthetic pathway (i.e., ura3$^-$) cannot grow when supplemented with 4tU alone (Supplemental **Fig. 2.S2**), suggesting that highly thiolated transcripts are not tolerated by the cell. As we found comparable 4tU incorporation in a ura3$^-$ strain and prototrophic strain (Supplemental **Fig. 2.S1**) we performed all subsequent experiments in a prototrophic strain. Over the timescale and concentrations of 4tU used for RATE-seq we detect no effect on cell growth (Supplemental **Fig. 2.S3**), although prolonged exposure and higher concentrations appear to have slight effects (Supplemental **Fig. 2.S4**). We confirmed that the concentration of 4tU used for RATE-seq does not affect global gene expression (Supplemental **Fig. 2.S5**). Using a dot blot and colorimetric assay (Materials and Methods), we find that the pool of newly synthesized mRNA approaches equilibrium consistent with a model of constant synthesis and exponential degradation (**Figure 2.1B**). Consistent with expectation, the equilibrium value of labeled RNA differs with different concentrations of 4tU, but the kinetics of the

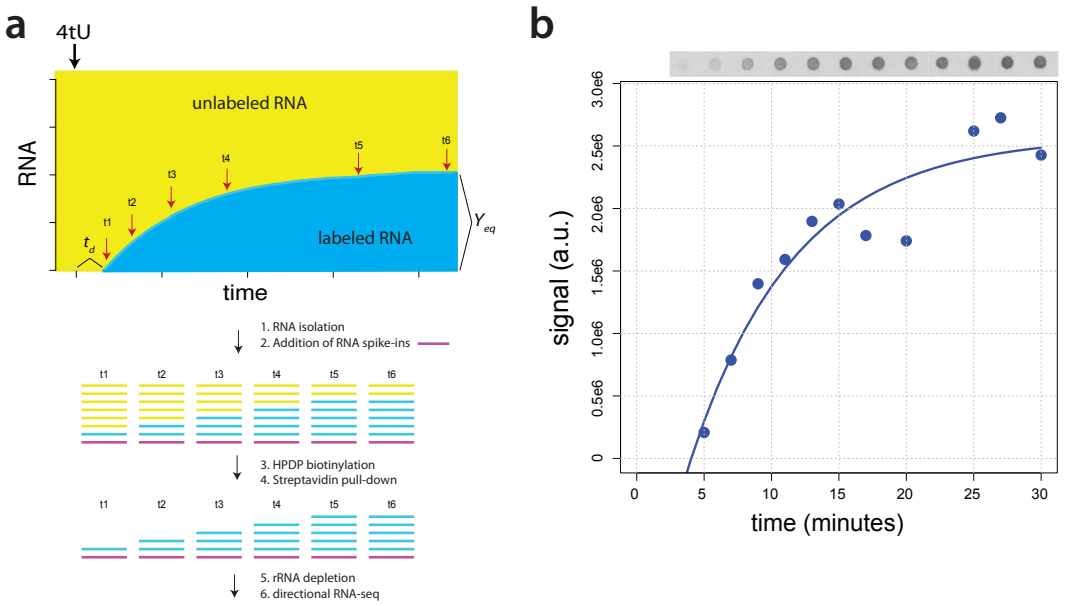


Figure 2.1: RATE-seq enables in vivo measurement of RNA kinetics
(**a**) Overview of approach to equilibrium labeling and analysis using RATE-seq. The increase in labeled transcript with time *Y(t)* modeled using the relationship $Y(t) = Y_{eq}(1 - e^{-(\alpha_{RNA} + \alpha_{growth})(t - t_d)})$ where Y_{eq} is the abundance of labeled transcript at steady state, α_{RNA} is the transcript's degradation rate constant, α_{growth} is the growth rate constant of the culture, t is the time after addition of label, and t_d is a time delay between the addition of label and the time at which labeled transcripts can be detected. Red arrows indicate points at which the RNA samples are recovered following addition of 4tU. (**b**) Incorporation of 4tU conforms to approach to equilbrium kinetics. An equivalent quantity of biotinylated polyadenylated RNA from timepoints following addition of 4tU was bound to a membrane and visualized using streptavidin alkaline phosphatase and chemifluorescence. Values are shown along with the model fit.

approach to equilibrium is unaffected (Supplemental **Fig. 2.S6**). As with radiolabeling experiments in mammalian cells (Greenberg 1972), the mRNA fraction approaches equilibrium faster than total RNA (Supplemental **Fig. 2.S7**), which reflects the relative stability of rRNA compared with mRNA.

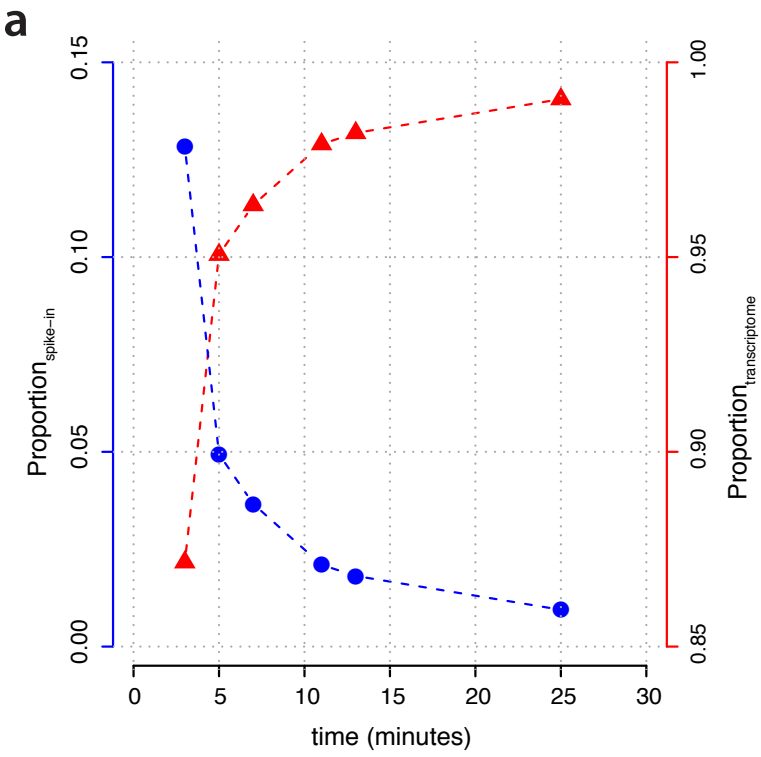
2.3.2: Measurement of RNA degradation rates transcriptome-wide

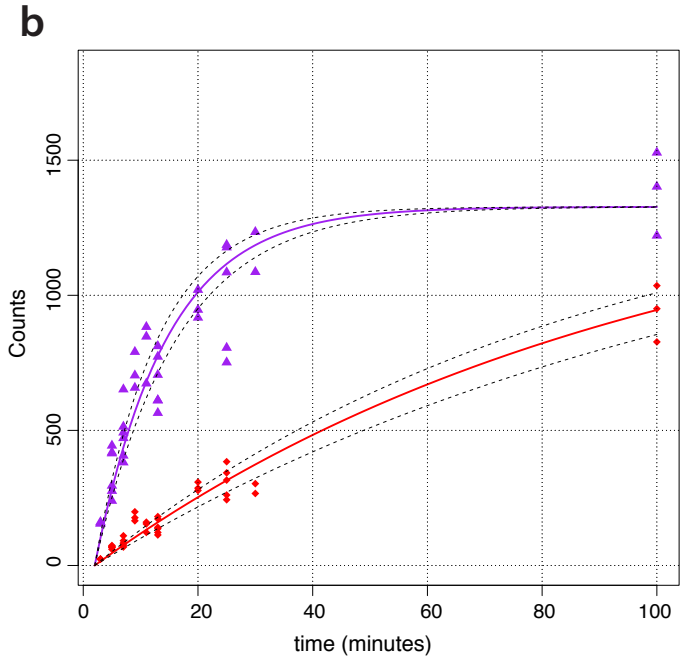
We performed RATE-seq using replicate yeast populations growing in a defined rich medium during log phase. Following RNA-seq analysis, the relative counts (Supplemental Tables 2.S2, 2.S3) of spike-ins are observed to decrease with time and concomitantly, the proportion of counts mapping to the transcriptome increases (**Figure 2.2A**). We found that the use of multiple spike-ins facilitated identification of technical biases potentially introduced during library preparation (Supplemental **Fig. 2.S8**). The correlation of per transcript counts between replicates at the same time point is high (Spearman $\rho = 0.98$; Supplemental **Fig. 2.S9**). To normalize transcript counts (Supplemental Table 2.S4) we first determined the ratio of counts for each transcript to each spike-in, scaled each ratio, and then multiplied by the mean count of all spike-ins across all experiments to preserve the scale of the data (Materials and Methods). We studied the mean-variance relationship at each time point and found that the data are overdispersed (Supplemental **Fig. 2.S10**). Therefore, to estimate the degradation rate constant for individual transcripts we performed a nonlinear weighted regression using normalized counts from the combined data set (**Figure 2.2B**) (Materials and Methods). We

determined confidence intervals for the estimated decay constant for each transcript using bootstrapped values from each time-point (Materials and Methods).

Using RATE-seq we determined degradation rate constants, and corresponding half-lives, with 95% confidence intervals (CI) for 5308 mRNAs (Supplemental Table 2.S5). Most transcripts are rapidly degraded, with a mean and median half-life of 15 and 10 min, respectively (**Figure 2.2C**). Thus, RATE-seq analysis estimates RNA half-lives that are shorter than most previous global estimates (Wang et al. 2002; Grigull et al. 2004; Miller et al. 2011; Munchel et al. 2011). Using bootstrapped values we find that for the majority of transcripts the estimated degradation rates have confidence intervals of ±20% (Supplemental Table 2.S5). A previous study (Wang et al. 2002)showed that transcripts encoding functionally related gene products have similar decay rates. We find that genes within the same Gene Ontology (GO) terms also have similar decay rates (Supplemental Table 2.S6) although the agreement between the estimated rates from the two studies is poor. Functional categories representing the most rapidly degraded transcripts include “Helicase activity” and “Regulation of cell cycle” whereas categories representing the most stable transcripts include “Cytoplasmic translation” and “Ribosome” (**Figure 2.3A**).

In addition to variation in mRNA degradation rates we find evidence for variation in rates of noncoding transcripts including small nuclear RNA


a
0.15
0.10
0.05
0.00
Proportion spike−in
1.00
0.95
0.90
0.85
Proportion transcriptome
0
5
10
15
20
25
30
time (minutes)


b
1500
1000
500
0
Counts
0
20
40
60
80
100
time (minutes)

C

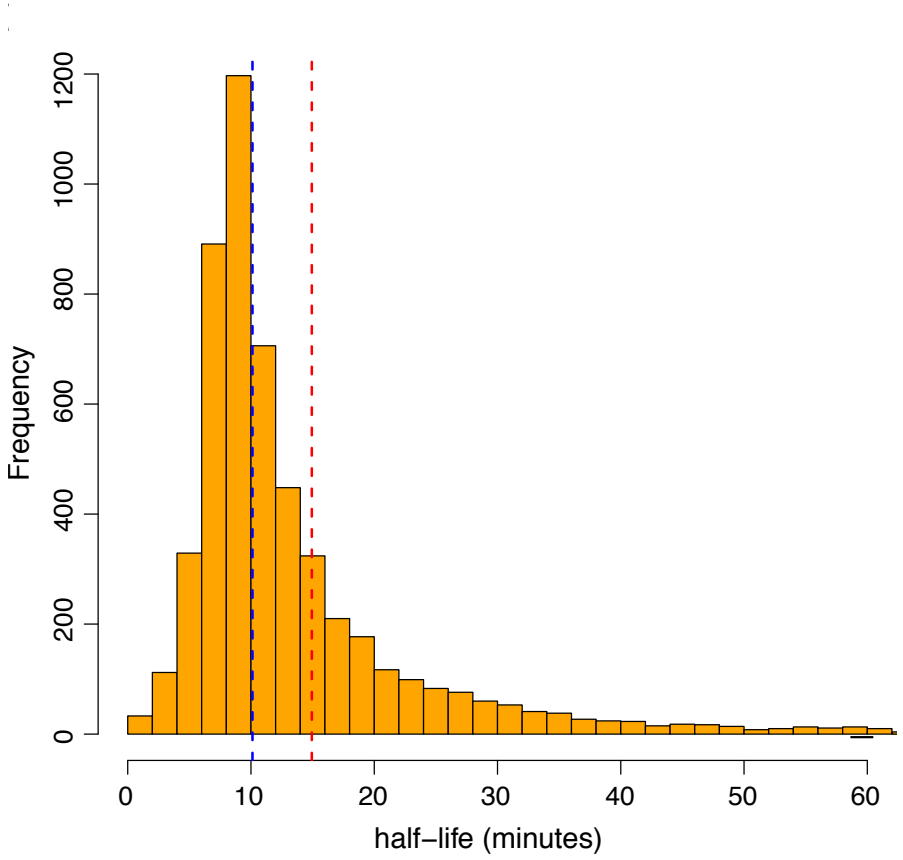


Figure 2.2: Global RNA kinetics determined using RATE-seq

(**a**) The relative fraction of reads mapping to the transcriptome increases with time whereas the fraction of reads mapping to each spike in decreases. (**b**) Representative example of RATE-seq data for a rapidly degraded gene *(CTK1*, purple) and a slowly degraded gene (*GIM4*, red). 95% CI for the estimated degradation rate constant are indicated by dashed lines. (**c**) The distribution of half-lives for all coding transcripts with the mean (red line) and median (blue line) half-life indicated.

(snRNA) and long noncoding RNAs (lncRNA) (**Figure 2.3B**). The population snRNAs appear to be more stable than coding transcripts (Figure 2.3B) and have similar half-lives, suggesting that the post-synthesis fate of snRNAs is coordinately regulated.

We compared mRNA half-lives estimated using RATE-seq to previously reported estimates in *Saccharomyces cerevisiae* (Supplemental **Fig. 2.S11**). As noted in previous reports (Miller et al. 2011; Munchel et al. 2011), the agreement among mRNA half-lives using different methods is poor. Surprisingly, RATE-seq estimates correlate poorly with those reported using pulse-chase labeling with 4tU (Munchel et al. 2011). Our method has a number of differences that may account for this including the absence of poly(A) selection, the use of multiple spike-ins for normalization, and the use of untransformed data for nonlinear model fitting, which avoids errors introduced by linear transformation of data. In addition, mathematical modeling suggests that nucleotide recycling may slow the observed chase resulting in a systematic underestimation of mRNA decay rates (Supplemental **Fig. 2.S12**). Our estimates are most similar to results obtained using DTA (Miller et al. 2011), which may reflect the fact that both methods isolate newly synthesized transcripts following label addition. Importantly, consistent with both existing in vivo labeling methods in budding yeast (Miller et al. 2011; Munchel et al. 2011), we find that the half-lives for ribosomal protein-coding genes is greater than the median half-life of all mRNAs (Supplemental **Fig.**

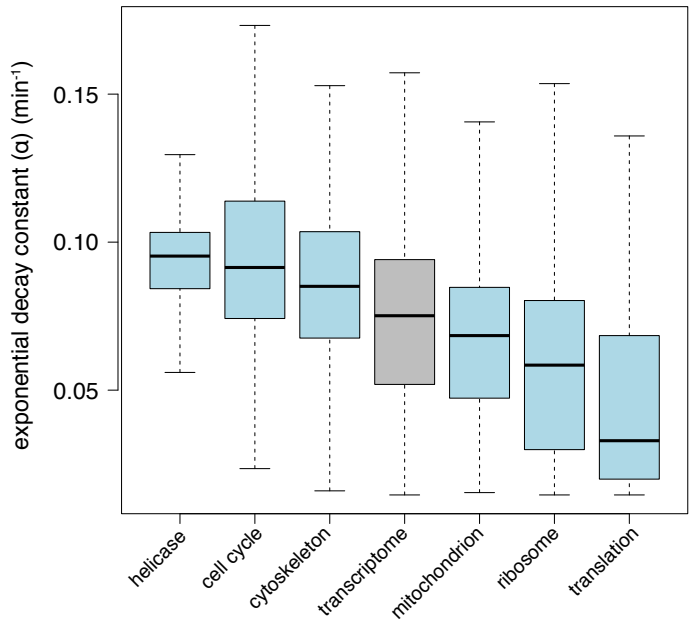
2.S13). In contrast, in all studies using transcriptional inhibition ribosomal protein-coding transcripts are found to degrade as rapidly as the transcriptome average, which may reflect a stress response to the profound impact on cell physiology caused by these methods.

2.3.3: The landscape of regulated transcript abundance

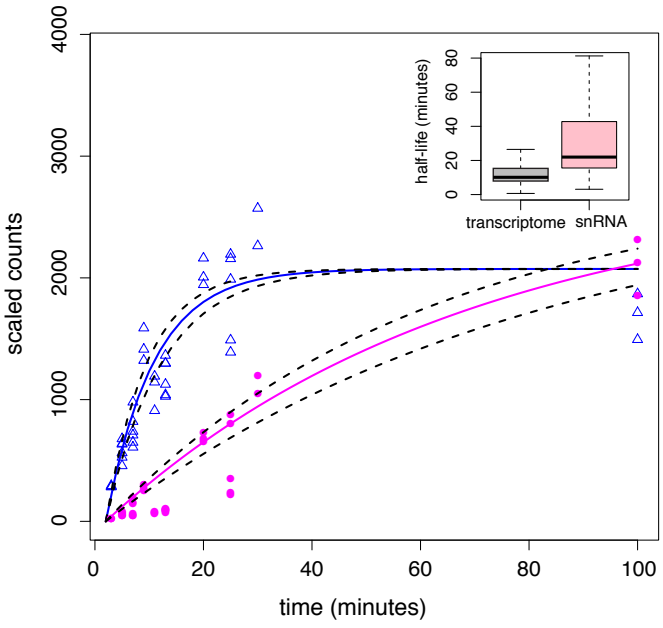
At steady state, transcript abundance levels are constant (i.e., d[RNA]/dt = 0) and transcript synthesis and degradation are related by the expression $k = \alpha[RNA]$. Therefore, the rate of transcript production can be estimated using the degradation rate constant and the steady-state abundance of the transcript. As only a fraction of transcripts are labeled with 4tU (**Figure 2.1A**), RATE-seq does not quantify RNA abundance. Therefore, we used published estimates of absolute transcript abundance from quantitative sequencing data (Lipson et al. 2009) to estimate rates of mRNA synthesis in steady-state conditions (Supplemental Table 2.S5). Our estimates of mRNA synthesis rates are in good agreement with previous estimates using Genomic-Run-On (GRO) assays (Supplemental **Fig. 2.S14**;(Pelechano et al. 2010)) with a linear correlation coefficient of $r = 0.5$. A source of discrepancy between the two data sets may be that our study estimates the rate of production of mature transcripts, whereas GRO estimates nascent transcription rates.

The combinatorial effect of variation in synthesis and degradation rates defines the landscape of regulated transcript abundance (**Figure 2.3C**). Within

a


exponential decay constant (α) (min-1)
0.15
0.10
0.05
helicase
cell cycle
cytoskeleton
transcriptome
mitochondrion
ribosome
translation

b


scaled counts
0
1000
2000
3000
4000
0
20
40
60
80
100
time (minutes)
half-life (minutes)
0
20
40
60
80
transcriptome
snRNA

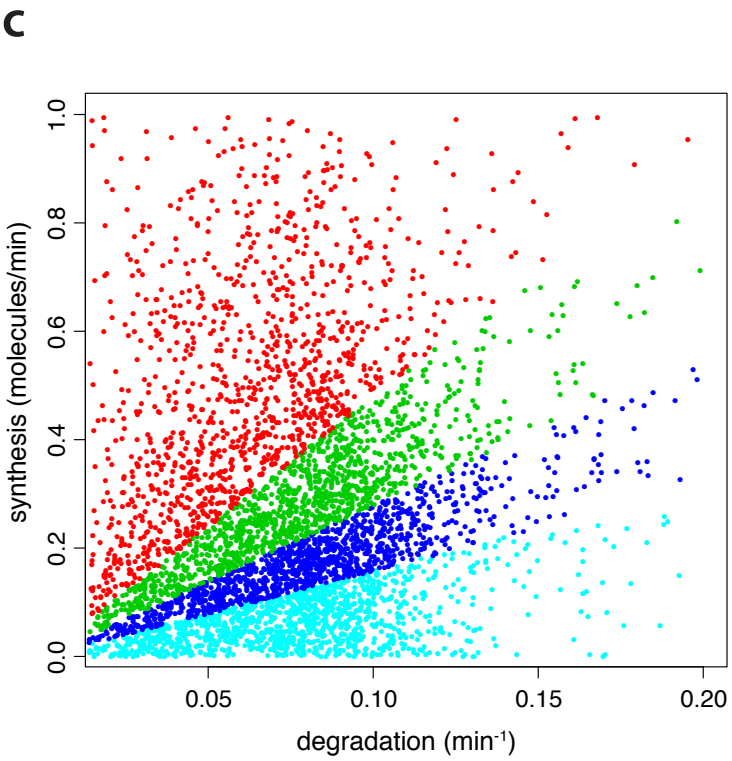


Figure 2.3: Coordinated regulation of mRNA abundance levels

(**a**) Decay rates for sets of functionally related genes defined by GO term categories are non-randomly distributed. (**b**) RATE-seq can be used to estimate the kinetics of snRNAs (e.g. snR48 (pink circles)) and lncRNAs (e.g. ICR1 (blue triangles)). The boxplot (inset) displays the half-lives for snRNA and all other transcripts. (**c**) The global relationship between transcript synthesis and degradation in *S. cerevisiae*. The abundance of each mRNA is categorized into quartiles (first quartile: 0.001-1.58 mRNAs/cell (cyan), second quartile: 1.6-2.8 mRNAs/cell (blue), third quartile: 2.8-4.8 mRNAs/cell (green), fourth quartile: 4.8-66 mRNAs/cell (red)).

this landscape it is clear that classes of transcripts defined by rapid synthesis and degradation have equivalent steady-state levels to classes of transcripts that are comparatively slowly synthesized and degraded. Understanding the sources and consequences of these different kinetics is central to understanding gene expression regulation.

2.4: Discussion

The abundance of a transcript is determined by both its rate of synthesis and its rate of degradation. To fully characterize the regulation of mRNA levels these rates must be uncoupled. Moreover, studying transcripts under their native control is critical as transcript stability may depend on cis-acting factors that associate with promoter regions (Bregman et al. 2011; Trcek et al. 2011).

RATE-seq is an efficient and general means of estimating transcriptome-wide absolute rates of RNA synthesis and degradation in steady-state conditions. In contrast to existing methods, it does not interfere with the cell's physiology, provides enhanced accuracy, obviates the potential impact of nucleotide recycling, and can be applied to a variety of types of transcripts on a genome-wide basis. In principle, incorporation of 4tU is feasible in all organisms using either endogenous or heterologous expression of UPRT (Cleary et al. 2005). Alternatively, 4sU can be used in organisms without endogenous nucleotide salvage pathways (Dölken et al. 2008). Therefore, we expect that RATE-seq will be of great utility for investigating the

relationship between RNA synthesis and degradation in a variety of genotypes and organisms.

2.5: Materials and methods

2.5.1: Strains and growth conditions

Experiments were performed using either FY4 (MATa) or FY3 (MATa ura3-52), which are isogenic to S288C. All RATE-seq analyses were performed using the prototrophic strain FY4 in which a single colony was inoculated into an overnight culture in synthetic complete medium containing 500 μM uracil. The saturated overnight culture was back-diluted 1:50 into fresh medium of the same composition. Log phase cells were treated with 4tU to a final concentration of 500 μM. Cells were collected at multiple time points over the course of 100 min by vacuum filtration onto nitrocellulose filters and immediately frozen in liquid nitrogen.

2.5.2: Synthesis of polyadenylated thiolated spike-in RNAs

To generate three RNA spike-ins with similar GC content to *S. cerevisiae* mRNAs, we cloned three different regions of the *Bacillus subtilis* genome. The three spike-ins (spike-in$_{700}$, spike-in$_{900}$, spike-in$_{1200}$) have a GC content of 0.42 and lengths of 700, 900, and 1200 bases, respectively. Three regions of the *B. subtilis* genome were PCR amplified and cloned into the pSP64 poly(A) in vitro transcription vector (Promega). Plasmids were linearized using EcoRI restriction and run off transcription performed as recommended by the

manufacturer with the addition of thiolated UTP:UTP at a ratio of 2:1 in the reaction. Spike-in RNA was subsequently treated with DNAse and purified.

2.5.3: RNA extraction

RNA was purified from cells using a hot acid phenol/chloroform extraction. Briefly, 750 µL of lysis buffer (10 mM EDTA, 10 mM Tris, 0.5%SDS) was added to each sample and vortexed. An equal volume of acid phenol was then added to the sample and vortexed. Samples were incubated for 1 h at 65°C with occasional vortexing. Filters were removed and samples were placed on ice for 10 min. After centrifugation, the aqueous phase was transferred to Phase Lock Gel (PLG) tubes and an equal volume of chloroform added. The aqueous phase was collected and RNA was precipitated using two volumes of 95% ethanol and 0.1 volume of 3 M Sodium Acetate. RNA pellets were washed with 70% ethanol twice and dried at room temperature for half an hour and resuspended in RNAse free water.

2.5.4: RNA biotinylation and streptavidin pull down

For biotinylation reactions 100 µg of total RNA was added to a solution of 10 mM Tris-HCl (pH 7.4), and 1 mM EDTA. Biotin-HPDP (1 mg/mL) was added to a final concentration of 2 µg for each 1 µg of RNA (Supplemental **Fig. 2.S15A**). In addition, the three spike-in RNAs were pooled and 12 ng of the mixture added to the reaction mixture containing 100 µg of RNA sample. The reaction was allowed to proceed for 3 h in the dark, after which reactants were removed using chloroform extraction. RNA pellets were precipitated with 1

volume of isopropanol and 1/10 volume of 5 M NaCl. RNA pellets were washed once with 75% ethanol and resuspended in RNase-free water.

The biotinylated RNA was fractionated from unlabeled RNA using streptavidin magnetic beads (NEB) (Supplemental **Fig. 2.S15B**). Pull downs were performed essentially as previously described (Zeiner et al. 2008). Beads were washed four times and then transcripts were cleaved from magnetic beads using β-mercaptoethanol (5%). RNA was precipitated with 1 volume isopropanol, 1/10 volume NaCl, and 3 μg of glycogen (Supplemental **Fig. 2.S15C**).

2.5.5: Dot blot analysis

For isolation of poly(A) RNA from total RNA, Oligo d(T)25 magnetic beads (New England Biosciences) were used in combination with a 12-tube magnetic rack. Beads were washed once in a binding buffer/wash buffer (20 mM Tris-HCl at pH 7.5, 500 mM LiCl, 1 mM EDTA) similar to manufacturer recommendations except that DTT was left out of the buffer, as this would cleave the RNA conjugated to biotin-HPDP. At least 40 μg of total RNA was added to 200 μL of beads. Samples were washed in 1x binding buffer, then 1x low-salt buffer, and eluted from beads in TE buffer following incubation for 3 min at 50°C.

For each sample, 200 ng of mRNA was blotted onto a Zeta-Probe nylon membrane (BioRad) using a BioRad DotBlot. The RNA was cross-linked using a UV cross-linker. The blot was blocked using blocking buffer (PBS, 10% SDS, 1

mM EDTA) for 20 min. Samples were then probed with Streptavidin Alkaline phosphatase in blocking solution (1:1000). The membrane was washed in PBS at decreasing concentrations of SDS (10%, 1%, 0.1%) for 10 min each. Spots were visualized using ECF substrate (GE Healthcare), visualized on a Typhoon FLA 9500, and analyzed using ImageQuant software.

2.5.6: Depletion of ribosomal RNA

Following fractionation of thiolated transcripts, 100 ng was depleted of 18S and 25S ribosomal transcripts. Two rounds of ribosomal depletion were performed using LNA probes provided in the Ribominus kit (Invitrogen). RNA was then precipitated using 2 volumes ethanol, 1/10 volume 3 M sodium acetate, and glycogen. Pellets were resuspended in 6 µL of RNAse free water.

2.5.7: Library preparation for Illumina sequencing

First strand synthesis of rRNA-depleted RNA was performed using the Super Script III kit (Life Technologies) and random priming using random hexamers. Second-strand synthesis was performed with dUTP in place of dTTP to enable strand-specific sequencing (Parkhomchuk et al. 2009). Samples were end repaired, A-tailed, and ligated to NEXTflex DNA Barcodes for multiplex sequencing. Adapter dimers were removed using AMPure beads (Agencourt). Samples were then treated with UNG and amplified using 10 cycles of PCR prior to sequencing. Samples were sequenced using an Illumina 2000 single-end 50-bp run.

2.5.8: Sequence alignment

Ilumina sequencing reads were first filtered for rRNA sequences by aligning to the ribosomal DNA of the yeast genome using Bowtie with default settings. All remaining reads were then aligned to the rest of the yeast genome and the three spike-in sequences using Bowtie2 (Langmead et al. 2009). After converting SAM files to BAM files, reads were filtered based on quality scores of 20 or higher. The resulting BAM files were then used to calculate total counts per transcript using the featureCounts function of the Subread package in R, using the argument for strand specific counting. Each library had between 5 and 13 million reads mapping to non-rRNA transcripts.

2.5.9: Data normalization

We performed RATE-seq using two biological replicates, with time points k = 3, 5, 7, 11, 13 and 25 minutes following label addition for replicate 1, and time points k = 5, 7, 9, 13, 20, 25, 30 and 100 minutes following label addition for replicate 2. To normalize data within each time series we employed the following normalization scheme:

1) We first computed a ratio, A, between the read count M, for each gene i in replicate j at time point k and the read count S for each spike-in n in replicate j at time point k:

$A_{injk} = M_{ijk} / S_{njk}$ where n=1,2,3 for each of the three spike-ins and j=1 or 2 depending on the replicate

2) We then computed a scaling factor, β, for each spike-in by calculating the average ratio between each spike-in and a reference spike-in across all K time points within a replicate j:

$$\beta_{nj} = (\sum_{k=1}^{k=K_j} S_{njk} / S_{n=1\,jk}) / K_j$$

3) To normalize the data within a replicate, j we multiplied the ratio, A for each gene by the scaling factor:

$$C_{injk} = A_{injk} \cdot \beta_{nj}$$

4) To return the data to the original scale we then multiplied the normalized ratio for each gene by the average spike-in count across all K time points from both replicates:

$$N_{injk} = C_{injk} \cdot (\sum_{j=1}^{j=2} S_{njk} / \sum_{j=1}^{j=2} K_j)$$

We excluded all data from a time-point if a spike-in was deviant in its expected behavior. Thus, each time point has between 2-6 values depending on whether the time point was replicated and whether any data were removed.

2.5.10: Model fitting

The approach to equilibrium method assumes transcript decay follows first order kinetics and transcript synthesis follows zeroth order kinetics. This leads to the following two equations:

1. $Y_{unlabeled} = Y_{ss} e^{-\alpha t}$
2. $Y_{ss} = Y_{unlabeled} + Y_{labeled}$,

where $Y_{unlabeled}$ is the abundance of unlabeled transcript, $Y_{labeled}$ is the abundance of labeled transcript, Y_{ss} is the total abundance of a transcript at time t, and α is the transcript's degradation rate constant. The approach to equilibrium equation is then obtained by substitution of equation 1 for the value of $Y_{unlabeled}$ in equation 2 and solving for $Y_{labeled}$ leading to $Y_{labeled} = Y_{ss}(1 - e^{-\alpha t})$

Based on this equation, we modeled the abundance of labeled transcript for each mRNA as $Y(t) = Y_{eq}(1 - e^{-(\alpha_{RNA} + \alpha_{growth})(t - t_d)})$, where *Y(t)* is the amount of the labeled transcript at time *t*, Y_{eq} is the abundance of labeled transcript at steady state, α_{RNA} is the transcript's degradation rate constant, α_{growth} is the growth rate constant of the culture, *t* is the time after addition of label, and t_d is a time delay between the addition of label and the time at which labeled transcripts can be detected.

To calculate the degradation rate constant for each transcript, we performed non-linear regression, estimating both α (the summation of α_{RNA} and α_{growth}) and Y_{eq}. To account for biological variation, we combined the data from both replicates to generate a single parameter estimate for α_{RNA} and Y_{eq}. Because the variance of the RATE-seq data increases with increasing time we used weighted least squares regression with weights of 1/Y, which avoids undue influence of later time points on the model fit, using the *gnls* function in the *nlme* package in R. To minimize the parameters that we needed to estimate we set the time delay parameter, t_d equal to 2 minutes since labeled transcripts were pulled down as early as 3 minutes after addition of 4tU. As the doubling time of the culture is 100 minutes, α_{growth} is equal to 0.0069. Transcript half-lives were calculated as ln(2)/(α-0.0069*)*.

We calculated 95% confidence intervals of the estimated degradation rate constant by randomly sampling with replacement the equivalent number of points from normalized data for each mRNA 1000 times. When resampling fails to sample timepoints towards the latter part of the curve we found that the non-linear regression frequently failed to converge. Therefore, we used bootstrapped values to estimate only α keeping Y_{eq} the same for all iterations.

The data and model fit for each gene can be visualized using the available R script *rateSeqFit.R* using the function curve.generator().

2.5.11: Assessment of labeling efficiency and bias

We estimated the amount of 4tU labeling using a colorimetric Dot Blot analysis of labeled RNA and a synthesized oligonucleotide containing a 5′ biotin label. A standard curve was generated by diluting known quantities of the labeled oligonucleotide and used to estimate the number of labels in an RNA sample of known mass. Assuming an average transcript length of 1200 nucleotides we estimate that approximately one out of 500 uracil is labeled after 35 min of labeling under our conditions.

To test whether 4tU is preferentially incorporated, we performed a DNA microarray analysis of 4tU labeled RNA compared with the unfractionated sample (Supplemental **Fig. 2.S16**). Consistent with previous observations (Miller et al. 2009), there is a slight dependency of label incorporation on length. Thus, RNA-seq analysis of 4tU labeled transcripts is expected to result in increased counts for longer transcripts as a result of both

increased labeling efficiency and the larger target size of longer transcripts (Supplemental **Fig. 2.S17**). This labeling bias affects the steady-state equilibrium value for each transcript (Supplemental **Fig. 2.S18**). However, this bias does not affect the estimate of the decay rate as there is no relationship between the counts of labeled transcripts at any time point and our estimate of the decay rate constant (Supplemental **Fig. 2.S19**).

2.5.12: Estimation of mRNA synthesis rates

To estimate the synthesis rate for each transcript we assumed that the rate of change in mRNA (dRNA/dt) at steady-state is equal to zero and therefore used the relationship $k = \alpha[mRNA]$. We also assumed 60,000 mRNA/cell (Zenklusen et al. 2008). Confidence intervals for mRNA synthesis rates were calculated using the 95% CI values determined for α_{RNA}.

2.5.13: Gene enrichment analysis

Gene enrichment analysis was performed as in (Gresham et al. 2010). Non-random distribution of decay rates for each GO SLIM category as compared with the genes not in the category were identified using the Wilcoxon-Mann-Whitney test in R.

2.5.14: R functions and packages

All analyses were performed using R (Team) and several open source packages. The functions featureCounts of the Subread package and gnls of the nlme package were used for data analysis of nonlinear regression. The following functions and packages, in addition to base functions in R and

custom written functions were used for presentation of figures: subplot of the TeachingDemos package, axis.break of Plotrix, and heatscatter of LSD.

2.5.15: Accession codes

Sequencing data are available through the Sequence Read Archive under BioProject ID PRJNA236614.

2.5.16: Acknowledgements

We thank Christine Vogel, Daniel Tranchina and members of the Gresham lab for helpful comments. This work was supported by the National Institute of Health (GM107466 to D.G.); the National Science Foundation (MCB-1244219 to D.G.) and a Dupont Young Professor award (to D.G).

2.5.17: Author contributions

B.N. developed and optimized the RATE-seq method and performed the experiments. N.A. developed the RNA-seq library preparation protocol. B.N. performed computational analyses with help from N.A. B.N. and D.G. conceived of the method and wrote the paper with support from N.A.

2.6 Supplemental Figures:

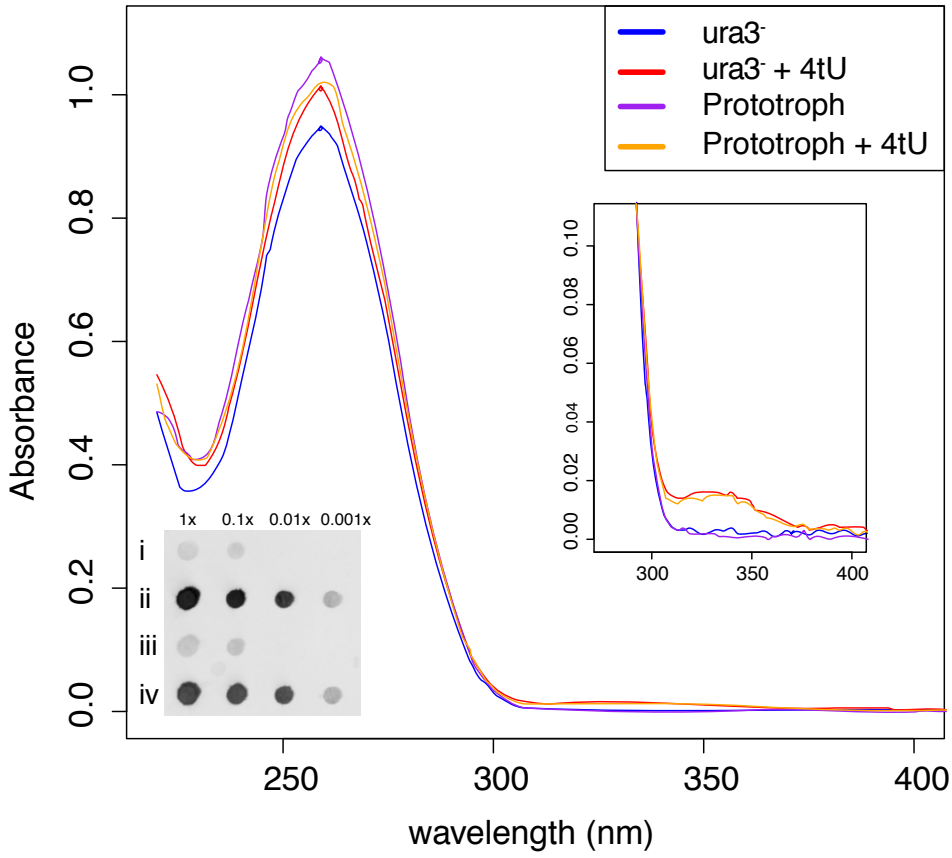


Figure 2.S1: Thiouracil is taken up by prototrophic and auxotrophic cells with equal efficiencies

The absorbance spectra of equivalent amounts of RNA purified from a *ura3-* and prototrophic strain grown in the presence of uracil and either the presence or absence of 4tU. 4tU has maximal absorption at 327nm resulting in a unique peak of absorbance in RNA labeled with 4tU (top inset). 2µg of the same total RNA sample from an auxotrophic cell grown in the (i) absence and (ii) presence of 4tU was compared to the same amount of RNA isolated from a prototrophic strain grown in the (iii) absence and (iv) presence of 4tU.

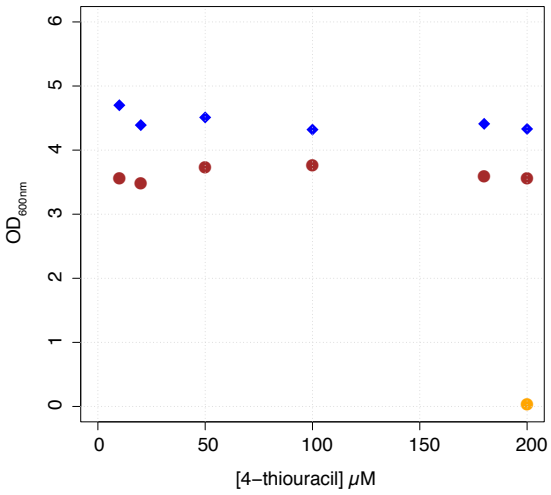


Figure 2.S2: Thiouracil cannot supplement a uridine monophosphate auxotroph

Cell densities were determined following 24 hours growth in the presence of 100μM uracil and different concentrations of 4tU. The presence of 4tU has no visible effect on final culture density for a prototrophic strain (blue diamonds) or an auxotrophic strain (brown circles). A *ura3*$^{-}$ strain does not grow when supplemented with 4tU in the absence of uracil (orange circle).

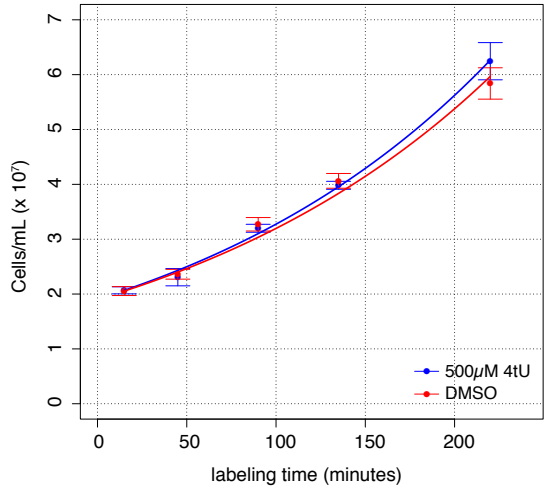


Figure 2.S3: Thiouracil has no detectable effect on cell growth for the duration of labeling

Cells in log phase growth supplemented with 500μM uracil were treated with 500μM 4tU (blue curve) or an equivalent volume of DMSO (red curve). Error bars correspond to 2 standard errors of the mean (95% confidence interval) of two biological replicates and at least two technical measurements of each.

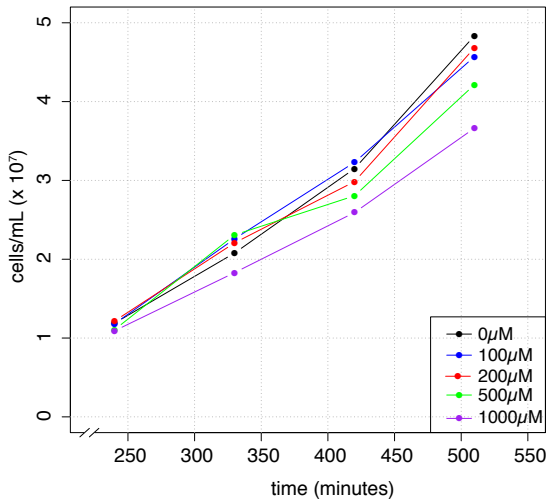


Figure 2.S4: Thiouracil minimaly affects cell growth at higher concentrations and longer durations of labeling
Cells were grown in 500μM uracil and varying concentrations of 4tU. Measurement of changes in cell density with time was initiated after at least two population doublings. The presence of 4tU seems to have a slight effect on rate of growth with increasing exposure times and higher concentrations.

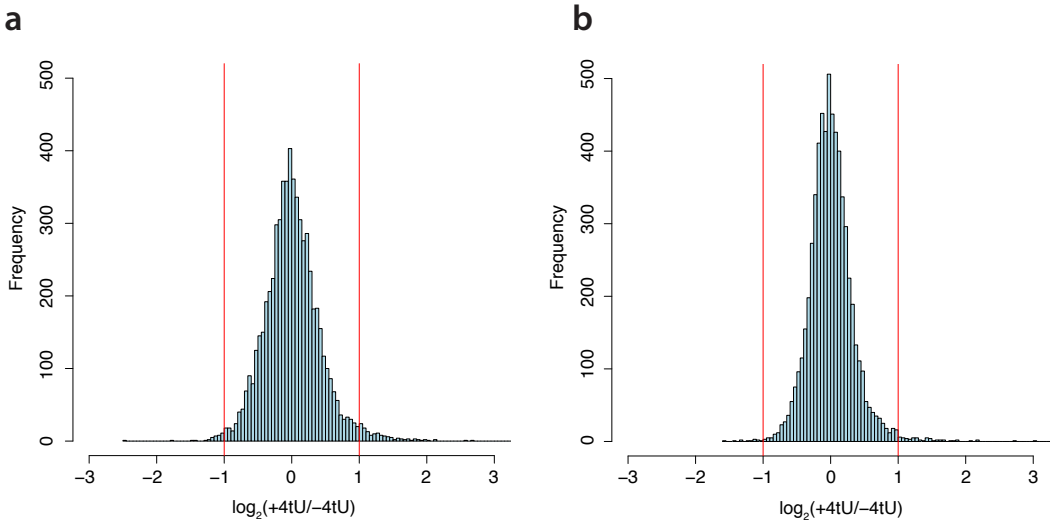


Figure 2.S5: Effect of 4tU exposure on steady-state transcript levels
RNA from cells treated with 4tU for an extended period of time was compared to RNA from cells never exposed to 4tU. Microarray analysis shows that in both (**a**) a strain defective in UMP biosynthesis (*ura3*$^{-}$) and (**b**) a prototrophic strain, the differences in steady state RNA levels are minimal, with 99% of transcripts having a $\log_2$ fold change from-1 to 1 (**Table S7**).

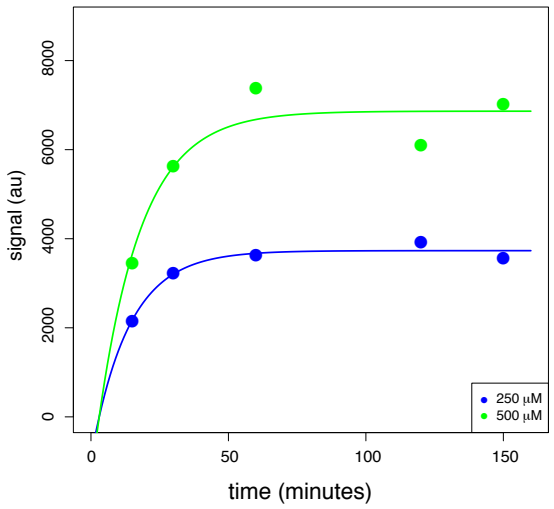


Figure 2.S6: The extent of 4tU labeling is concentration and time dependent
The kinetics of the approach to equilibrium is essentially the same for total RNA grown in the presence of 250µM or 500µM 4tU. The expected difference in the final steady state level of labeled RNA is the only observed difference.

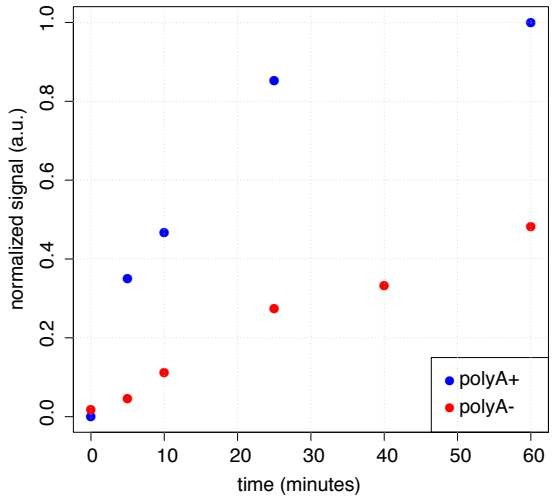


Figure 2.S7: Total RNA and mRNA have distinct labeling kinetics
Equivalent masses of biotinylated polyadenylated RNA and non-polyadenylated RNA were bound to a nitrocellulose membrane using a dot blot and visualized with streptavidin alkaline phosphatase and chemifluorescence. The signal for each RNA sample was normalized to the maximal signal following 24 hours of growth. Non-polyadenylated RNA, which is primarily rRNA, approaches its equilibrium value more slowly than polyadenylated mRNA.

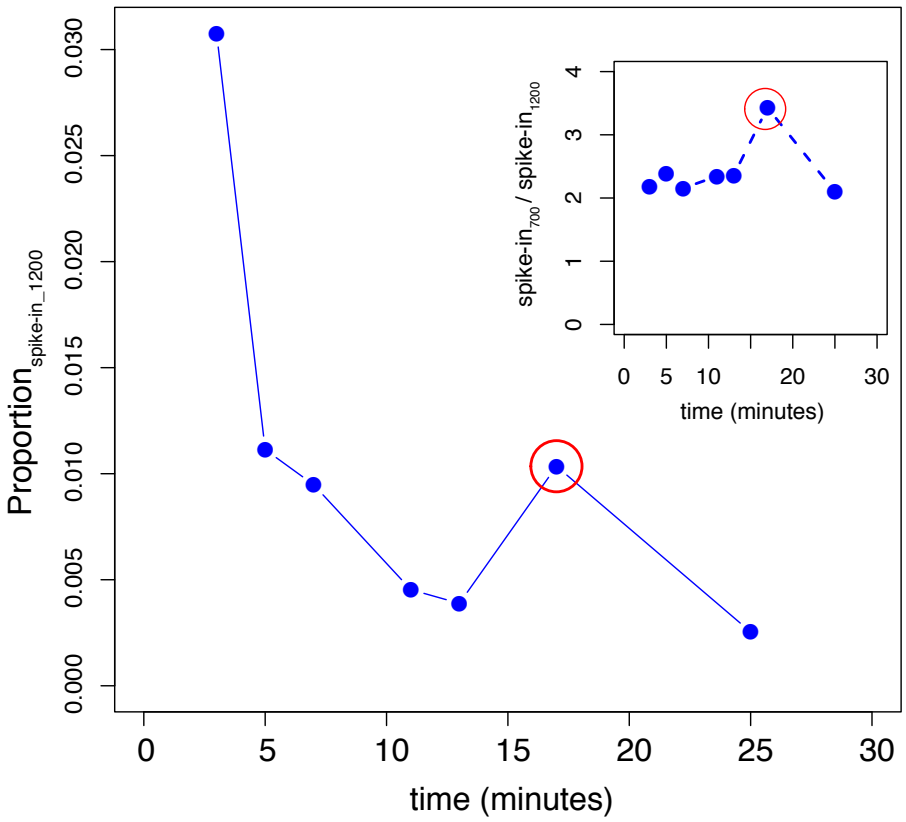


Figure 2.S8: Identification of technical biases in library preparation using multiple spike-ins

In our experiments, we expect the relative abundance of each spike-in to decrease with time. We also expect that the ratio between spike-ins, which have been added from the same pool, should be relatively constant across time. We found that at timepoint t=17 minutes spike-in$_{1200}$ is deviant from its expected behavior (shown) as are spike-in$_{700}$ and spike-in$_{900}$ (not shown). Simultaneously, the ratio between spike-in$_{700}$ and spike-in$_{1200}$ (inset) is relatively uniform across time, with the exception of t=17 minutes. These two metrics suggest that technical biases in library preparation have been introduced in this sample, and the timepoint was excuded from further analysis.

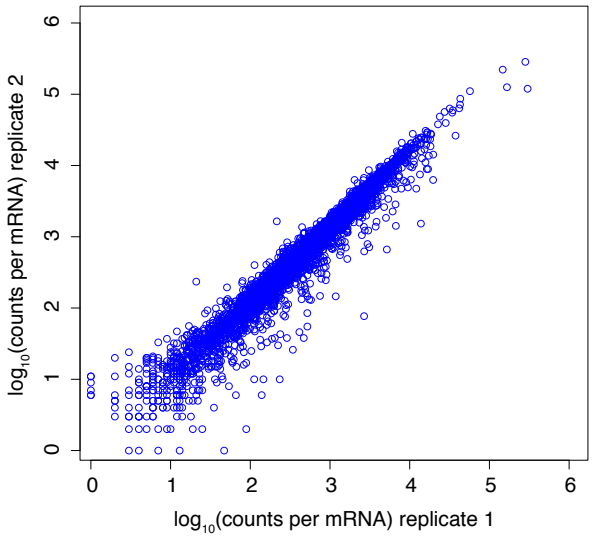


Figure 2.S9: Total counts per transcript are highly correlated between replicates
Total counts per mRNA were compared for replicate experiments at the same time point (5 minutes) following label addition. The Spearman correlation between replicates is ρ=0.98.

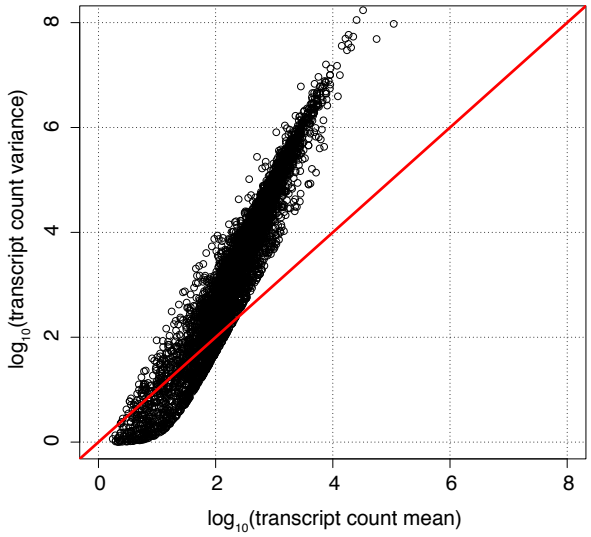


Figure 2.S10: Normalized transcript count data are over-dispersed
The mean-variance relationship for normalized data. The expectation from a Poisson distribution of equality between transcript mean and variance is indicated by the red line.

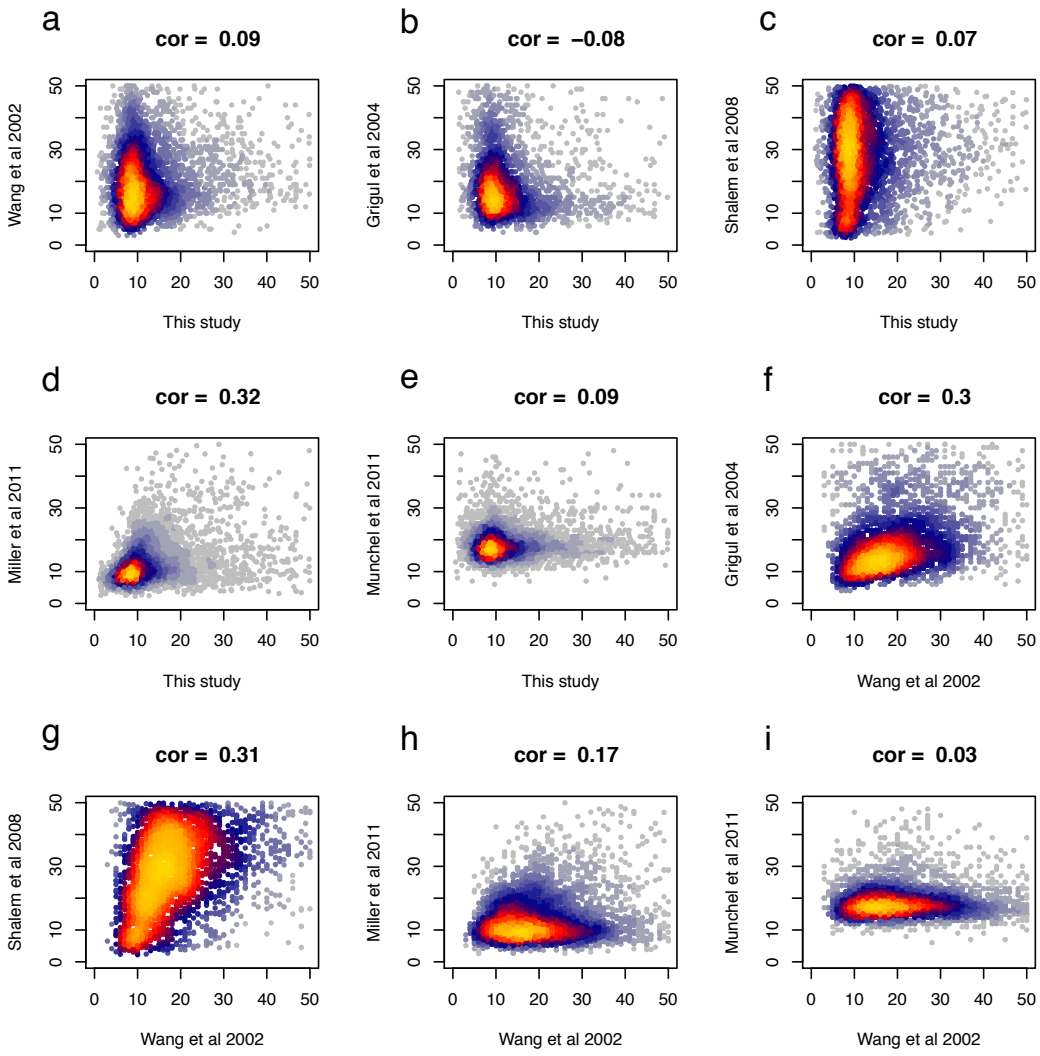


Figure 2.S11: Comparison of estimated RNA half-lives between our studies and previous publications

The Pearson correlation between our study and (**a**) (Wang et al. 2002), (**b**) (Grigull et al. 2004), (**c**) (Shalem et al. 2008), (**d**) (Miller et al. 2011), and (**e**) (Munchel et al. 2011) is indicated. The Pearson correlation between (Wang et al. 2002) and (**f**) (Grigull et al. 2004), (**g**) (Shalem et al. 2008), (**h**) (Miller et al. 2011), and (**i**) (Munchel et al. 2011) is indicated.

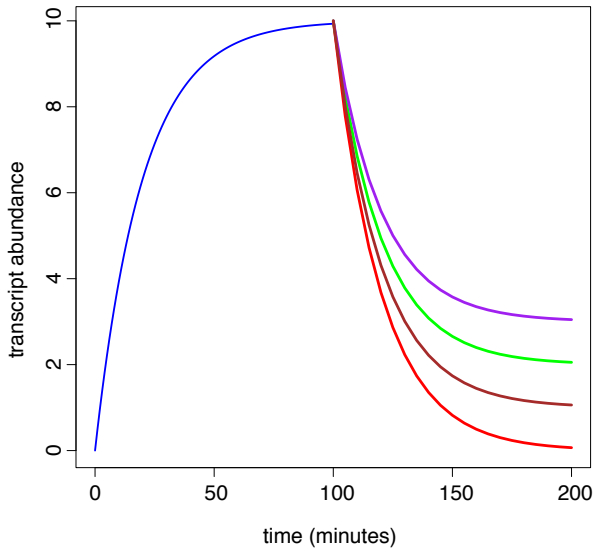


Figure 2.S12: The effect of nucleotide recycling on pulse-chase labeling
The population of labeled transcripts can be modeled by the equation $dY/dt = x\beta - \alpha Y$ where α is the degradation rate constant, Y is the number of transcripts, β is the rate of synthesis, and x is a coefficient ranging from 0-1 denoting the fraction of newly synthesized transcripts that are labeled. If x=1 all new transcripts are labeled following the chase whereas when x=0 no newly synthesized transcripts incorporate the label following the chase. In the case of a pulse-chase experiment one starts with a labeled population of transcripts. In the scenario of a perfect chase, x=0, and the observed rate of mRNA degradation (red) is the same as that estimated using approach to equilibrium (blue curve), which does not depend on synthesis kinetics. However, if the label is recycled then x>0 and the label will be incorporated into newly synthesized transcripts following the chase. For example, if x = 0.1, 0.2, or 0.3 the degradation rate is under-estimated by 29% (brown curve), 45% (green curve), and 57% (purple curve) respectively.

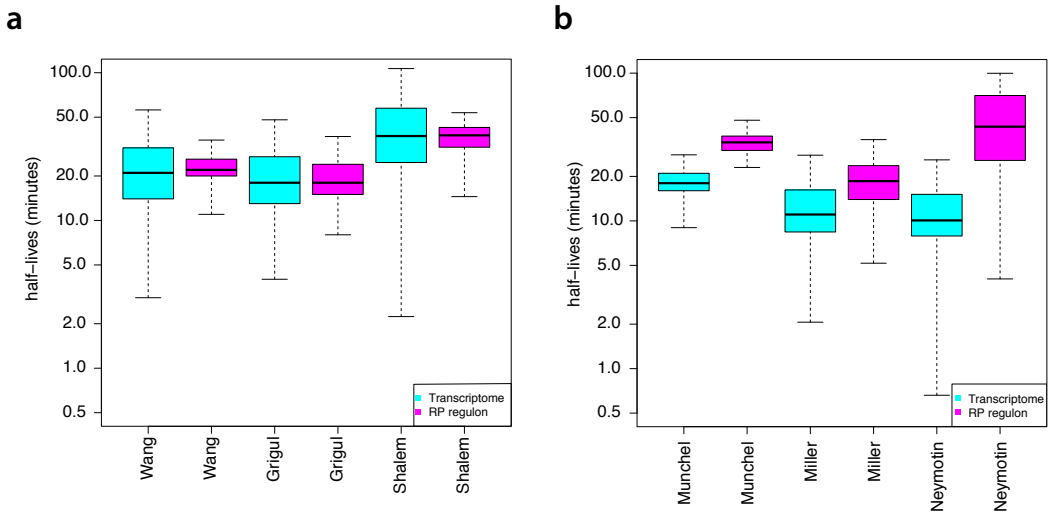


Figure 2.S13: Genome-wide degradation rate estimates for the ribosomal protein (RP) regulon are most similar based on method
We compared methods using transcriptional shut-off (Wang et al. 2002; Grigull et al. 2004; Shalem et al. 2008) with methods using 4tU labeling (Munchel et al. 2011; Miller et al. 2011) including the current study. The study is indicated by the first author's name. (**a**) Using transcriptional shut off, transcripts belonging to RP regulon (pink) are found to have half-lives similar to the entire transcriptome (blue) whereas (**b**) all estimates using *in vivo* metabolic labeling find that the RP transcripts are long-lived relative to the rest of the transcriptome.

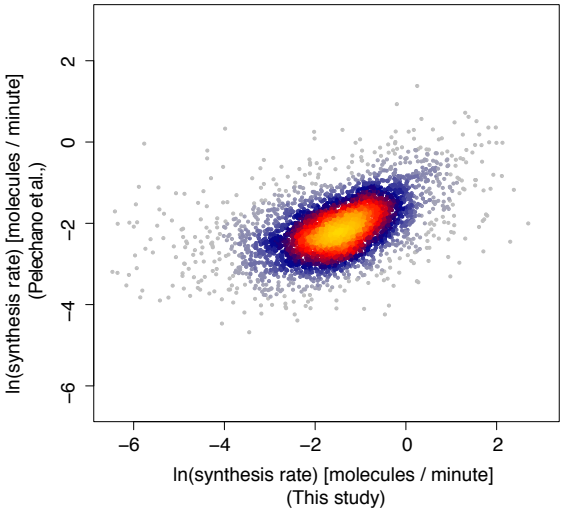


Figure 2.S14: Comparison of mRNA synthesis rates estimates using RATE-seq and GRO
We compared rates of synthesis using RATE-seq with those estimated using Genomic Run On (Pelechano et al. 2010). The estimates are positively correlated.

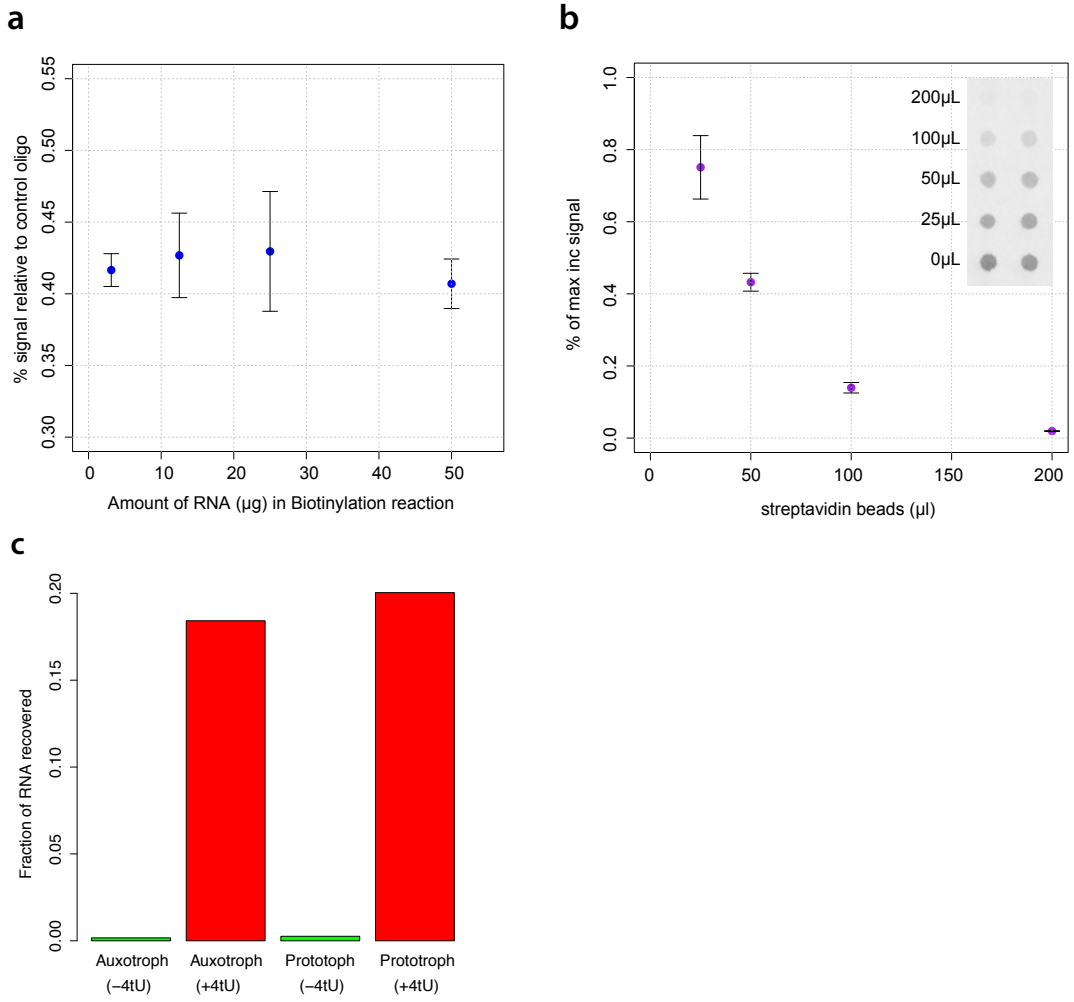


Figure 2.S15: Optimization of 4tU labeling protocol

Each step of the 4tU labeling procedure was optimized as follows. (**a**) We determined that the quantity of biotin-HPDP used for conjugating to 4tU-labeled RNA was sufficient by using a fixed amount of biotin-HPDP with differing quantities of 4tU -labeled RNA. There is no decrease in signal with increasing quantities of RNA using a colorimetric dot blot analysis indicating that biotin-HPDP is in excess for the range of RNA concentrations used in RATE-seq. (**b**) We optimized the quantity of streptavidin-labeled beads to ensure that the entire labeled fraction is captured. RNA that was not captured by the streptavidin-labeled beads was assayed for 4tU content using a colorimetric dot blot assay. We determined that 200µl of beads is sufficient to capture all labeled RNA. (**c**) We determined the pull down efficiency of samples treated with 4tU as compared with untreated cells by quantifying the total amount of recovered RNA following biotin-HPDP conjugation and streptavidin beads of a fixed mass of RNA. For an equal quantity of RNA, the yield of RNA is 100-fold higher for the labeled samples compared with unlabeled sample in both a *ura3*$^{-}$ auxotroph and a prototrophic strain.

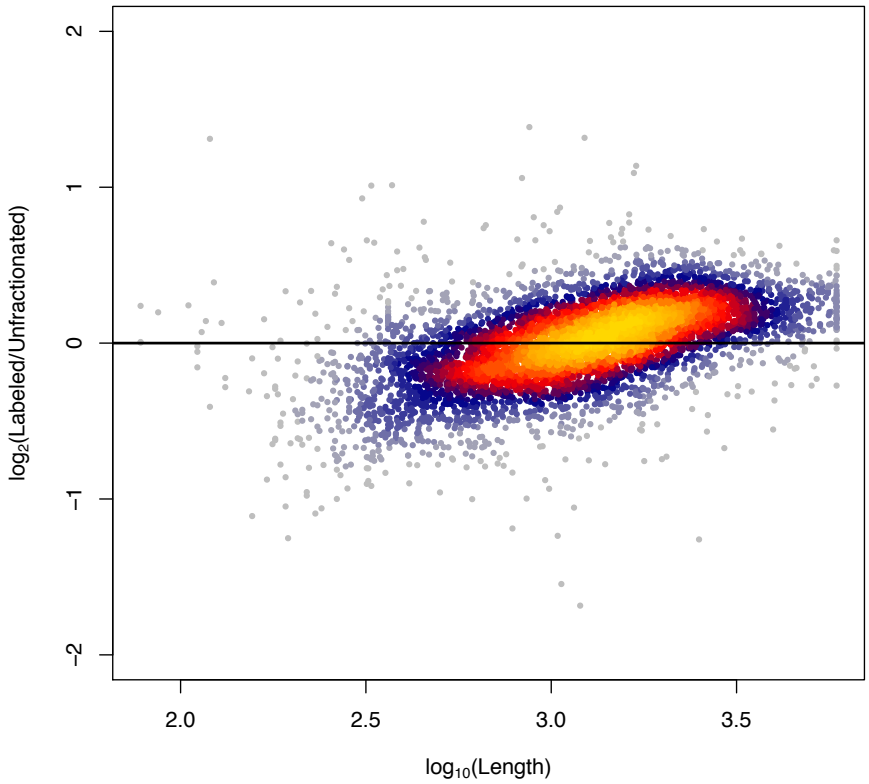


Figure 2.S16: Labeling bias in 4tU treatment
Comparison of coding sequence length with the relative microarray signal (log_2) (**Table S7**) from fractionated thiolated RNA compared to the unfractionated sample indicates that longer transcripts are preferentially labeled and/or recovered.

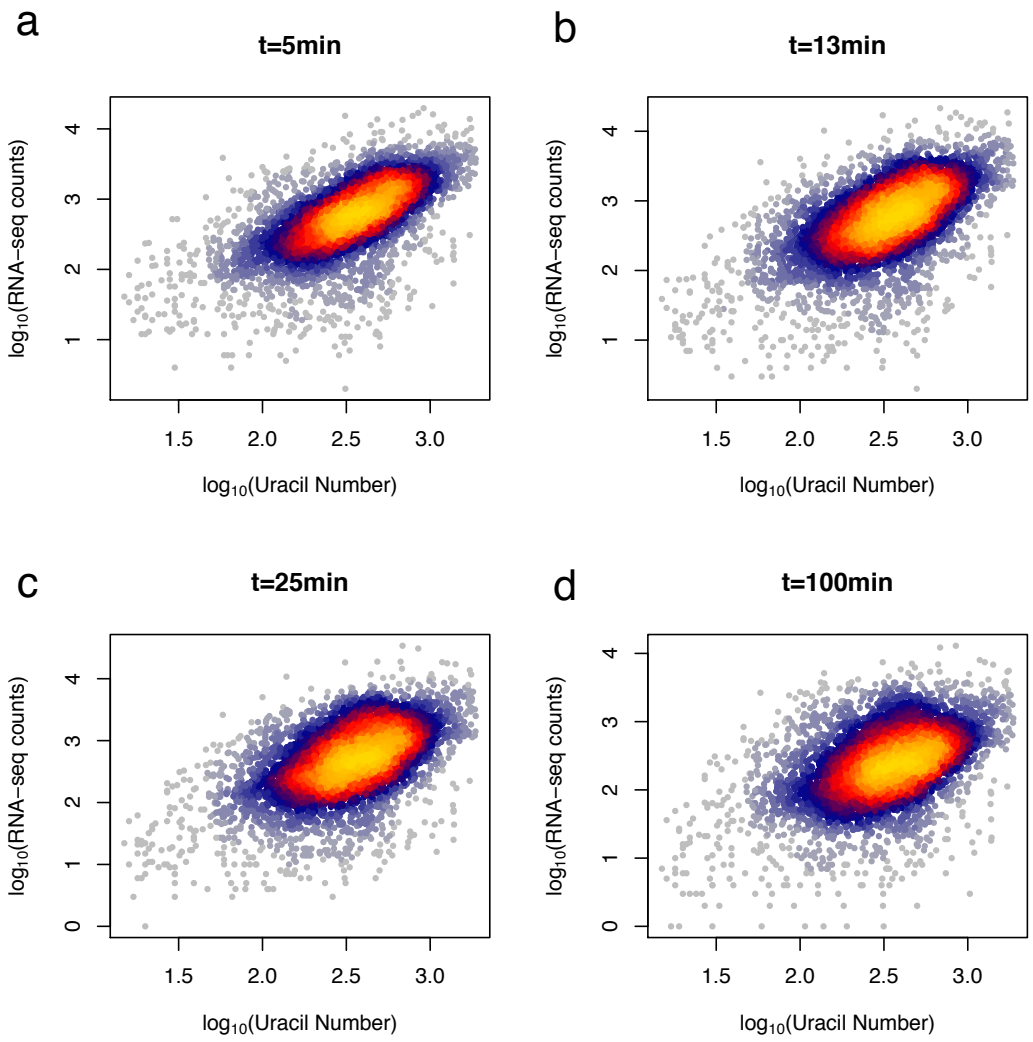


Figure 2.S17: Sequencing read counts are positively correlated with the amount of uracil in a transcript

The total number of uracils in a transcript is a function of its length. A longer transcript typically has more counts at each time point during the labeling as a result of both increased label incorporation and increased transcript length. The relationship between uracil number in a transcript and RNA-seq counts is apparent at (**a**) 5, (**b**) 13, (**c**) 25, and (**d**) 100 minutes after labeling.

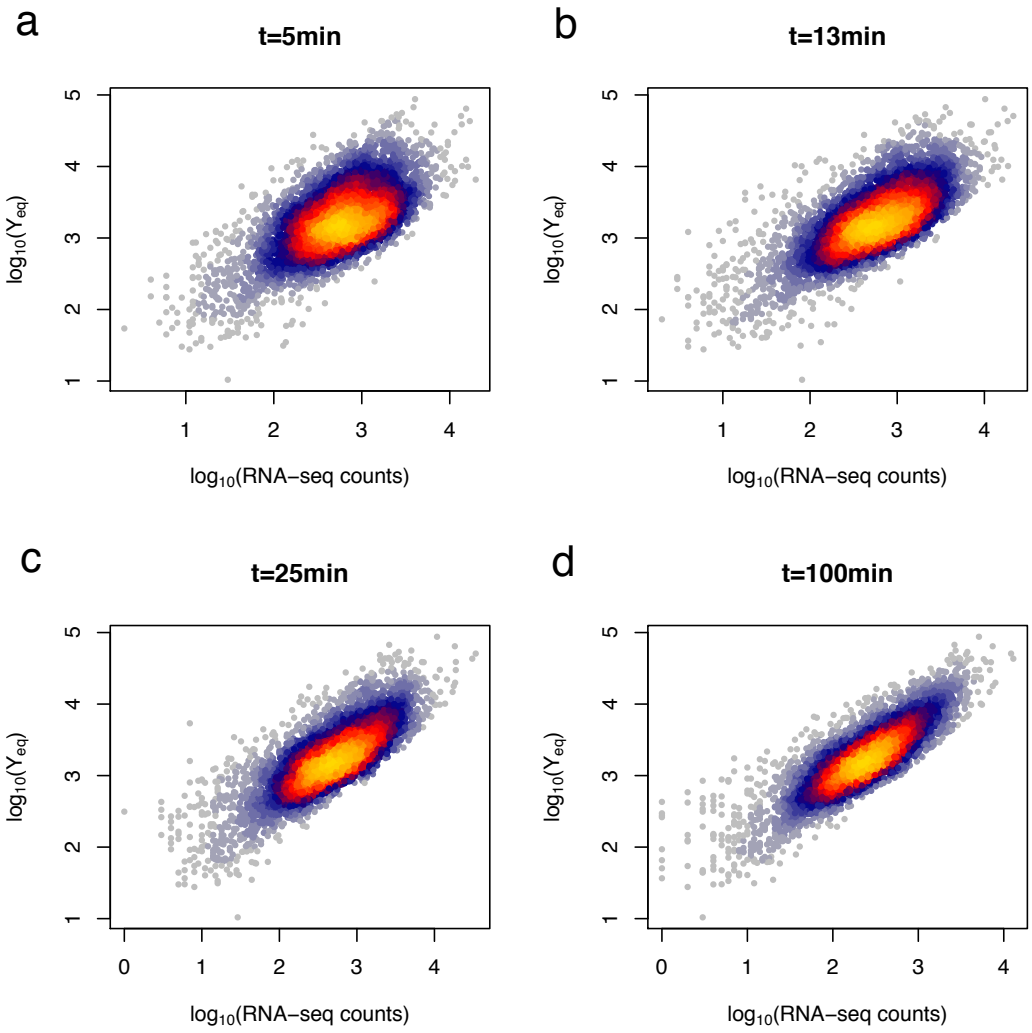


Figure 2.S18: Labeling bias affects the equilibrium value for each gene
The steady state equilibrium value (Y_{eq}) estimated by our non-linear modeling is positively correlated with the total number of sequence counts per transcript at each time point. The relationship between RNA-seq counts and Y_{eq} is apparent at (**a**) 5, (**b**) 13, (**c**) 25, and (**d**) 100 minutes after labeling.

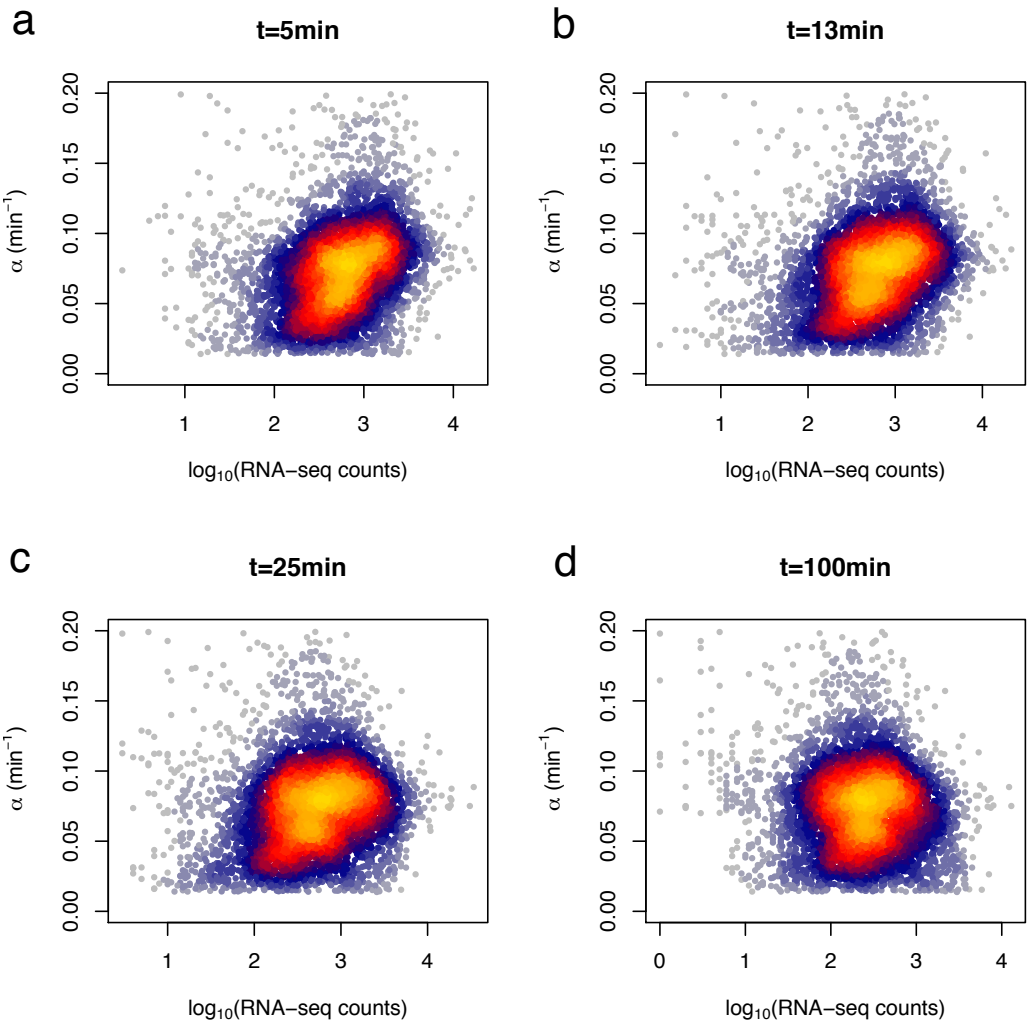


Figure 2.S19: Labeling bias does not influence the estimated degradation rate constant

The estimated degradation rate constant (α) is not correlated with the number of counts for each transcript at any time point during the approach to equilibrium. There is no relationship between RNA-seq counts and α detectable after (**a**) 5, (**b**) 13, (**c**) 25, and (**d**) 100 minutes of labeling.

CHAPTER 3: GC CONTENT AND TRANSLATIONAL EFFICIENCY CONTRIBUTE TO VARIATION IN MRNA DEGRADATION RATES

3.1: Abstract

The abundance of a transcript is determined by its rates of synthesis and degradation. Studies of individual mRNAs have shown that *cis* and *trans* elements affect mRNA decay rates. However, transcriptome-wide studies have failed to show a strong relationship between mRNA decay rates and any individual property of mRNA. We investigated the contributions of *cis* and *trans* factors on transcriptome wide decay rate variation in the yeast *Saccharomyces cerevisiae* using multiple regression analysis. We find that many inherent properties of transcripts have a relationship with decay rate and that multiple regression increases the explanatory power of the variation. We find that irrespective of dataset or method of measurement, several predictors, including transcript abundance, ribosome density, and GC content of the wobble position in a codon are significant predictors of mRNA decay rate across most datasets. Using conditionally regulated promoters of individual transcripts, we find that as predicted by the multiple regression model, decreasing ribosome density by changing the translational start site destabilizes mRNA for some, but not all transcripts. We postulate that this depends on the function of the encoded protein. We also experimentally validate that GC content of the third codon position affects mRNA steady state

levels as well as decay kinetics for individual transcripts. Our results suggest that translation efficiency is a primary determinant of degradation rates. We confirm that mRNA decay is a regulated process governed by multiple elements in a transcript specific manner.

3.2: Introduction

Remodeling of gene expression programs requires modulation of synthesis rate, decay rate, or both. Both mRNA synthesis and decay are critical for cell survival, as termination of either results in rapid cellular death (Nonet et al. 1987; Anderson and Parker 1998). As such, a thorough understanding of each process is imperative to fully interpret the mechanism of mRNA transcript level regulation. Although extensive evaluations on the elements affecting mRNA synthesis have been conducted and much is known about transcriptional regulation (reviewed in (Hager et al. 2009)), far less is known about the *cis* and *trans* factors affecting mRNA decay, and the degree to which those factors affect degradation. A complete understanding of gene regulation requires identification of the determinants of mRNA degradation expression.

The degradation pathways themselves and components of those pathways are understood (reviewed in (Parker 2012)) as a result of analyses of select individual transcripts (Decker and Parker 1993; Muhlrad et al. 1995; Beelman et al. 1996). Individual transcript studies also proved that mRNA degrades partially as a function of both factors in *cis* to the transcript, including sequence elements in the coding (Parker and Jacobson 1990;

Wisdom and Lee 1991) and untranslated (Shaw and Kamen 1986; Muhlrad and Parker 1992) regions, as well as factors in *trans*, including regulation by RNA binding proteins (Olivas and Parker 2000; Chen et al. 2001). More recently, promoter regions have also been implicated in determination of transcript stability (Bregman et al. 2011; Trcek et al. 2011). However, these observations are for a select few genes, in particular environments. To obtain a global picture of determinants of mRNA degradation, one needs study the trends available in genome wide studies.

Genome wide mRNA decay rates are available across all levels of life including bacteria (Selinger et al. 2003), plants (Narsai et al. 2007), flies (Thomsen et al. 2010), mouse (Rabani et al. 2011) and different human cell lines (Duan et al. 2013). However, no organism is more extensively studied than the yeast *S. cerevisiae*. In yeast, genome-wide RNA decay rates have been measured multiple times using different methods including transcriptional inhibition (Wang et al. 2002; Grigull et al. 2004; Shalem et al. 2008), genomic-run-on (García-Martínez et al. 2004), and metabolic labeling (Miller et al. 2011; Munchel et al. 2011; Neymotin et al. 2014). In addition, most of the knowledge in mRNA decay is from individual transcript studies in yeast. Finally, there is a rich array of datasets in yeast measuring gene specific parameters for the entire genome including protein levels (Ghaemmaghami et al. 2003), protein half-life (Belle et al. 2006), RNA abundance (Lipson et al. 2009), transcription rates (Pelechano et al. 2010), UTR lengths (Nagalakshmi

et al. 2008), and ribosome density (Ingolia et al. 2009). All of these datasets were generated in yeast under rich conditions of exponential growth. Therefore, yeast is the optimal organism for study of determinants of mRNA decay.

Recently, we introduced RNA Approach to Equilibrium Sequencing (RATE-seq), a new metabolic labeling method for determination of *in-vivo* mRNA synthesis and degradation rates genome-wide (Neymotin et al. 2014). Unlike previous methods, RATE-seq does not require transcriptional shut-off and has no effect on cellular physiology. In addition, because our method tracks the appearance of newly labeled transcripts with time, it does not suffer the problem associated with a slow chase, as in methods of pulse-chase labeling. We determine decay rates using non-linear regression of multiple time points following labeling and therefore are not limited in any problems associated with use of a single time point to determine kinetic rates. As a result RATE-seq offers the most accurate representation of physiological mRNA synthesis and degradation rates.

Collectively, the results of the mRNA decay studies in yeast have shown three important features within transcriptome wide decay rates. First, there is variation in the rates at which different transcripts are degraded, some by as much as an order of magnitude. Second, transcripts for genes involved in coordinately regulated processes are regulated at the level of mRNA decay (Wang et al. 2002; Neymotin et al. 2014). Third, no single transcript feature

has been reported in those publications that can explain a large amount of the observed variation (Wang et al. 2002; Munchel et al. 2011; Miller et al. 2011). This suggests that either the responsible factors are masked in genome wide studies, or a combination of different parameters affect rates of decay each contributing in degree in a transcript specific manner. By comparing genome wide rates of decay with any single transcript feature, one is actually comparing all of the differences between transcripts, including sequence elements, transcript length, GC content, transcript abundance, and folding properties. Strong relationships between transcript features and mRNA decay rates may not be evident simply because every mRNA transcript is inherently different. To fully address the question of determinants of variation, all available data about a gene need be utilized simultaneously.

We have analyzed mRNA degradation rates in *S. cerevisiae* through multiple regression analysis (Eck and Stephan 2008; Duan et al. 2013) and the many datasets describing gene specific parameters. To determine the underlying factors responsible for variation in rates of decay, we model mRNA degradation rate as a dependent variable on multiple predictors. In so doing, we address the possibility of a multifactor dependence for mRNA decay. We find that some predictors are significant within methods of measurement, but others are significant across methods, suggesting it is not a technical artifact of the method. We then biologically tested the predictor for individual transcripts, attempting to keep all other transcript features constant. In so

doing, we have addressed the issue of global studies masking underlying determinants of decay for individual species of mRNA.

Our study shows that irrespective of method of measurement, ribosome density, the GC content of the wobble position in a codon, and steady state levels are significant predictors in multiple regression analysis. For individual transcripts, we show that changing ribosome density affects both degradation rate and steady state levels in a transcript specific manner. We also show that altering the GC content of the wobble position affects decay kinetics, but only to a defined extent. We suggest that translation efficiency also contributes to mRNA decay kinetics. Taken together, our results indicate that mRNA decay is determined by multiple factors, many of which are intimately linked to aspects of translation.

3.3: Results

3.3.1: Multiple regression analysis reveals common predictors across datasets

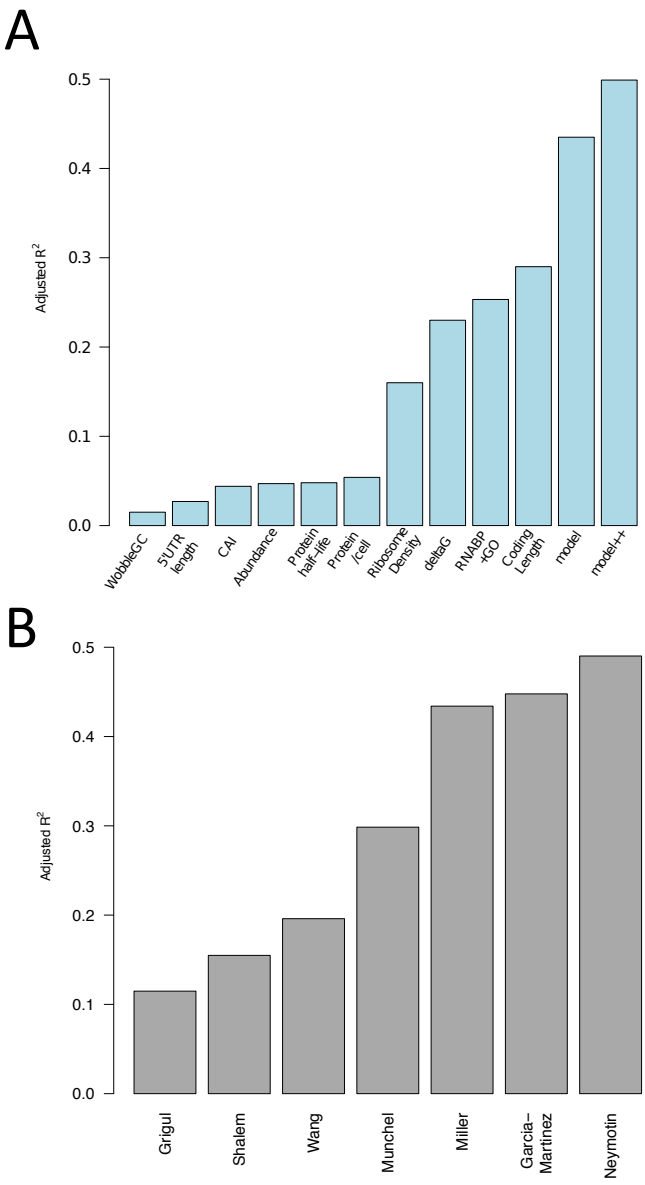
We performed multiple regression analysis in attempt to explain variation in mRNA decay rates genome wide. The predictor datasets for multiple regression were chosen based on factors implicated in regulating rates of mRNA decay for individual transcripts and are described in **Table 3.1** (**Table 3.S1**). A potential problem with multiple regression is that the significance of a predictor is calculated assuming all other predictors in the model are held constant. What may happen, then, is that a predictor is deemed significant or insignificant only because of other predictors in the

Table 3.1:

Dataset / variable	number	original units	Reference
Coding Length	5850	nucleotides	SGD
3'UTR Length	4950	nucleotides	SGD/Nagalakshmi et al, 2008
5'UTR Length	4345	nucleotides	SGD/Nagalakshmi et al, 2008
5'UTR GC content	4345	percent	SGD/Nagalakshmi et al, 2008
3'UTR GC content	4911	percent	SGD/Nagalakshmi et al, 2008
Abundance	5488	transcript/million	Lipson et al, 2009
Protein per cell	3818	protein/cell	Ghaemmaghami et al, 2003
Ribosome Density	5269	rpkM	Ingolia et al, 2009
Transcription Rate	4346	molecules/min	Pelechano et al, 2010
codGC1	5850	percent	SGD
codGC2	5850	percent	SGD
WobbleGC	5850	percent	SGD
Protein half-life	3164	min	Belle et al, 2006
deltaG	5850	kcal/mol	SGD/GeneRFold
Codon Adaptation Index (CAI)	5850	relative scale	CAI function in R
Protein per mRNA	3583	protein/cell/transcript	Ghaemmaghami et al, 2003 and Lipson et al, 2009
Munchel	5311	min	Munchel et al, 2011
Miller	4407	min	Miller et al, 2011
Wang	3967	min	Wang et al, 2002
Grigul	2702	min	Grigul et al, 2002
Garcia-Martinez	4624	min	Garcia-Martinez et al, 2004
Shalem	4639	min	Shalem et al, 2008
Neymotin	4802	min	Neymotin et al, 2014

model. Therefore, prior to modeling, we first observed the linear relationship of each predictor with mRNA decay. We found that the p-value for many of the predictors—an indication of whether the regression coefficient for that predictor differs from zero—is significant (**Table 3.S2**). This suggests a linear relationship between many of the predictors and decay rate, and also that the predictors, although not having a strong relationship with decay rate individually, might play a role in combination with other predictors.

We built multiple regression models to explain variation in mRNA decay rates. Analysis of the RATE-seq dataset indicates that many aspects of translation are often predictors of mRNA decay rates (**Figure 3.1A**). We also find that the explained variation, as measured by adjusted r-squared is greater when multiple predictors are included, suggesting decay rates are determined by a combination of transcript features ("model"). In addition, the amount of variation explained increases further when functional annotation is added in the form of Gene Ontology and RNA binding proteins ("model++"). This is consistent with the observation that transcripts encoding proteins in similar functional categories degrade with similar kinetics (Wang et al. 2002; Neymotin et al. 2014). We confirmed similar trends in other decay datasets as well (**Figure 3.S1**). Interestingly, we find that methods for measuring RNA decay that utilize transcriptional inhibition tend to explain far less variation than less disruptive methods (**Figure 3.1B,** compare first three datasets to last four), with RATE-seq explaining the most variation overall. By regressing the


A
Adjusted R2
0.0
0.1
0.2
0.3
0.4
0.5
WobbleGC
5'UTR length
CAI
Abundance
Protein half-life
Protein /cell
Ribosome Density
deltaG
RNABP +GO
Coding Length
model
model++
B
Adjusted R2
0.0
0.1
0.2
0.3
0.4
0.5
Grigul
Shalem
Wang
Munchel
Miller
Garcia-Martinez
Neymotin

C

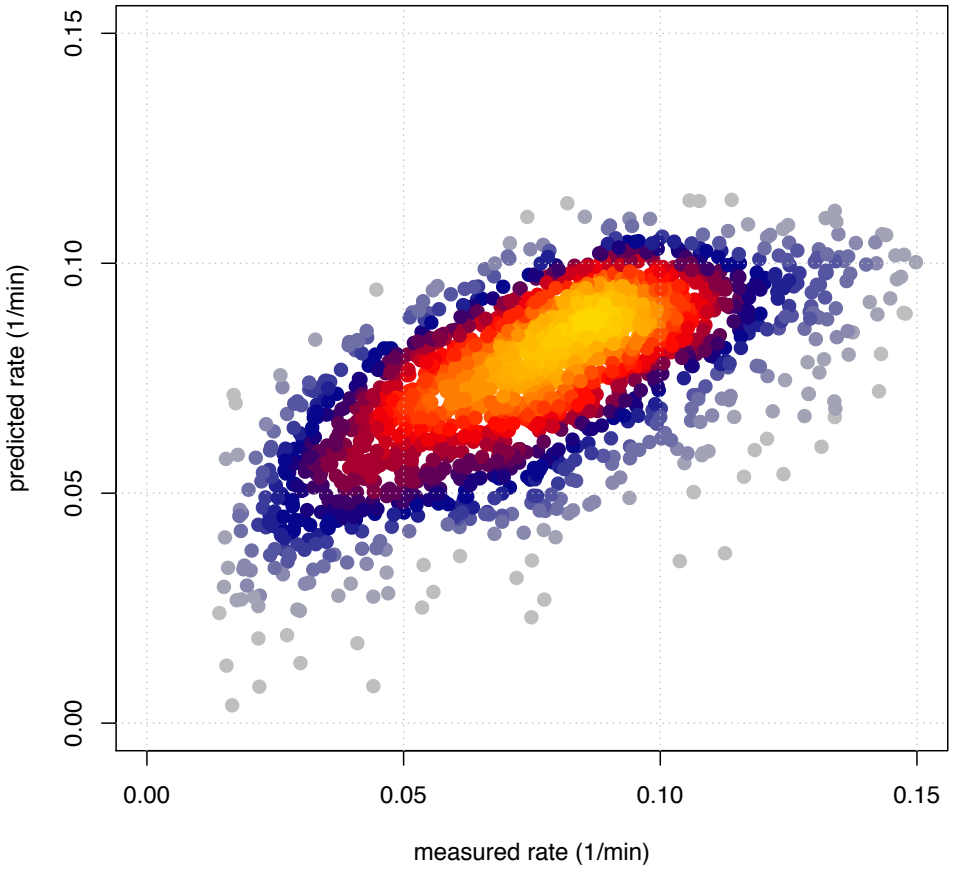


Figure 3.1: Multiple predictors best describe variation in mRNA decay
(a) Predictors of variation in mRNA decay in the RATE-seq dataset involve aspects of translation. (b) We determined the adjusted R^2 for each dataset when incorporated into a multiple regression model that includes functional annotation, RNA binding protein affinity (Hogan et al. 2008), and significant continuous variable predictors for each dataset. (c) By comparing the predicted decay rates vs. the measured decay rates, one can readily see the explanation of variation in decay rates. Data shown is from (Neymotin et al. 2014).

RATE-seq dataset we explain as much as 50% of the observed variation in decay rates (**Figure 3.1C**). Interestingly, irrespective of method of measurement, several predictors are significant, including steady state mRNA level, ribosome density, and WobbleGC content (**Figure 3.S1**). This fact strongly suggests that these predictors are not artifacts of any particular method, and they have an effect on rates of mRNA decay.

3.3.2: Ribosome density affects mRNA decay rate in a transcript specific manner

To determine the effect of ribosome density on mRNA degradation, we mutated the start codon of two different transcripts from ATG to GTG. This should disable the small ribosomal subunit from locating an ATG and recruiting the large subunit for fully formed ribosomes. Such an approach should reduce the number of ribosomes bound, but not eliminate ribosome binding in its entirety. In the absence of the normal start codon, downstream ATG may serve as start codons and translation initiation sites where the 80s ribosome forms (**Figure 3.2A**). To control transcription, these genes were placed under Doxycycline repressible promoters (Garí et al. 1997), which have little to no effect on cellular physiology, and have no effect on global gene expression (Wishart et al. 2005). By comparing to a constitutively expressed housekeeping gene, we are able to affect a single transcript without a global physiological response to measure degradation rates.

We tested both a GFP transcript and the GAP1 transcript for defects in degradation as a function of start codon removal. The GFP transcript encodes a fluorescent protein with no physiological role in the cell. In contrast, GAP1 encodes an amino acid permease that is necessary for efficient survival in nitrogen poor conditions. In the absence of a start codon, the steady state level of GFP is slightly different between the ATG and GTG strains (**Figure 3.2B**), although this difference is not statistically significant (p=0.0696, two sample t-test). The kinetics of mRNA decay is also unaffected **(Figure 3.2C).** In contrast, we find that mutation of the *GAP1* start codon significantly decreases steady state abundance levels (**Figure 3.2D,** p= 0.0024, two sample t-test), and the transcript degrades more rapidly (**Figure 3.2E**), although both reach the same final abundance levels. Taken together, our results suggest that a decrease in ribosome density strongly affects the decay kinetics for the GAP1 transcript, but has minimal effect on the GFP transcript. Possible explanations for these conflicting results are described below.

3.3.3: Changing WobbleGC content affects mRNA level and decay kinetics

Our regression model predicts that factors involved in mRNA translation contribute to observed mRNA decay rates (**Figure 3.1A**). One factor is the GC content of the wobble position in a codon (WobbleGC). WobbleGC affects mRNA levels, and mRNA degradation patterns in *E. coli* (Kudla et al. 2009). In addition, WobbleGC affects mRNA levels in mammalian

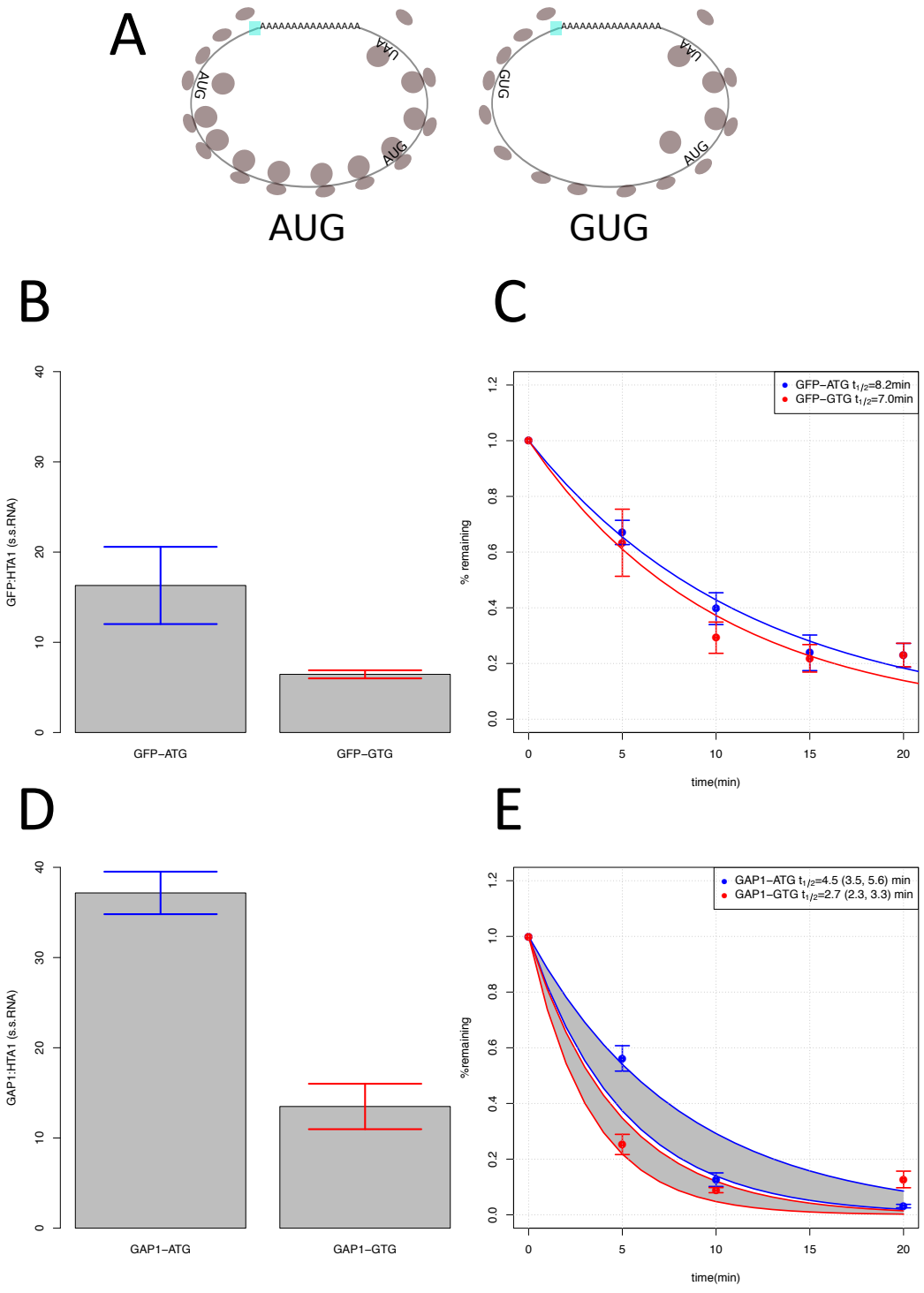
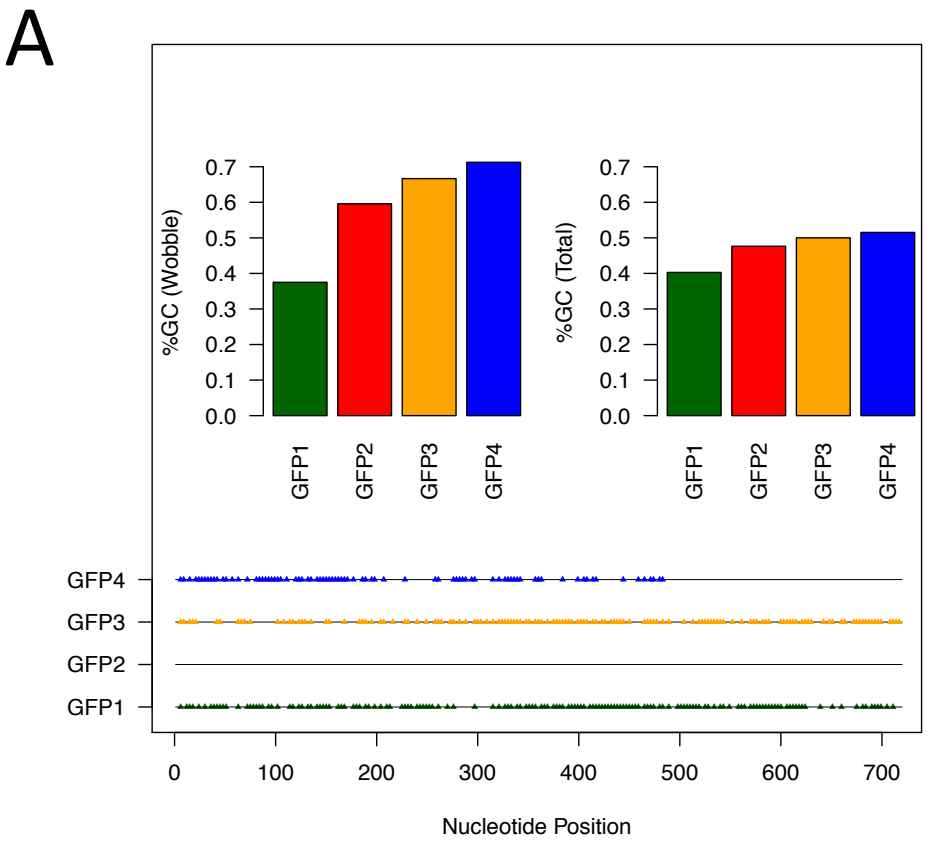


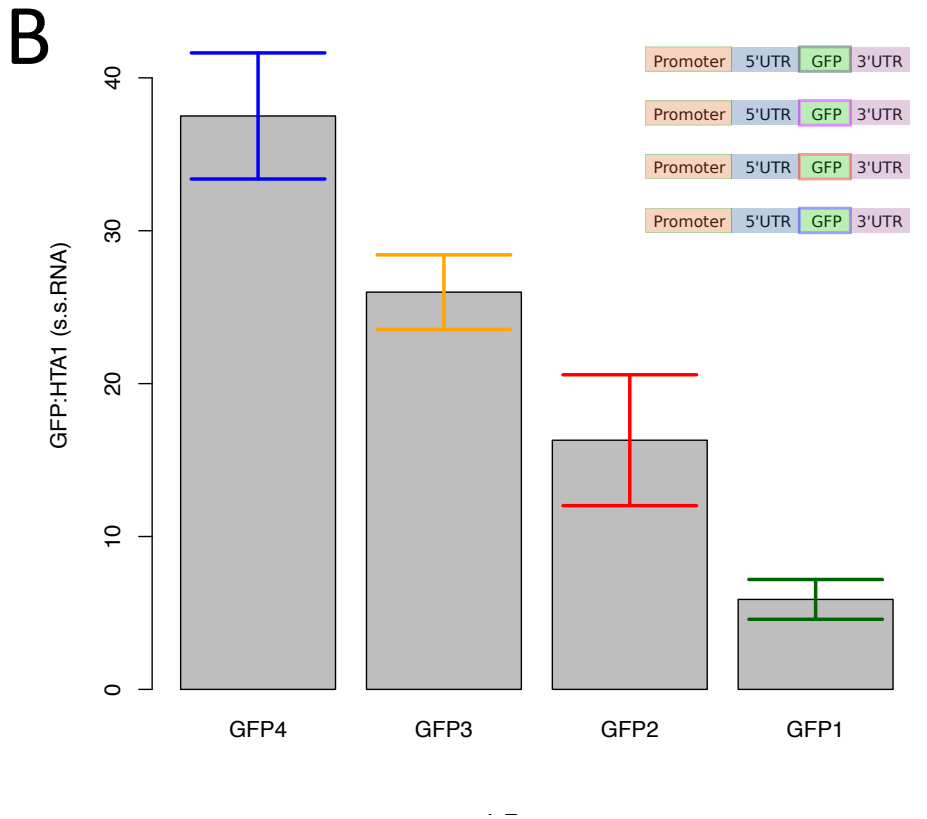
Figure 3. 2: The effect of ribosome density on mRNA decay rates
(A) In the presence of the normal start codon, the 40S ribosomal subunit scans the transcript for a start codon and when it is found, the 60S large subunit is

recruited to produce a fully formed ribosome (Left). In the absence of the normal start codon, the 40S ribosomal subunit will pass the original site of translation initiation and scan until the next start codon. At that point downstream, the fully formed ribosome may form (Right). Mutation of the start codon in GFP from ATG to GTG has little effect on the steady state level (B) and decay kinetics (C) of the transcript. Mutation of the start codon in GAP1 from ATG to GTG affects the steady state level (D) which is at least in part due to a change in the initial decay kinetics (E) of the transcript. In (E) the gray shading indicates bootstrapped 95% confidence intervals.

cells (Kudla et al. 2006), but not decay rates, suggesting that mRNA synthesis or processing is the underlying contributor to differences in mRNA levels.

To examine the contribution of WobbleGC to mRNA degradation, we utilized four different GFP constructs. The four GFP are very different in WobbleGC, but are similar in overall GC content (**Figure 3.3A**). They are all under the exact same context of regulatory regions including the same Doxycycline regulated promoter and the same UTR (**Figure 3.3B**), and are therefore expected to have the same rate of transcription. Any differences in steady state abundance must then result from differences in decay kinetics or processing. We find that the mRNA levels are directly related to WobbleGC (**Figure 3.3B**), consistent with similar observations in mammalian cells(Kudla et al. 2006). This immediately suggested that decay rates are affected. Following addition of Doxycycline to repress transcription, we find that indeed, three of our strains degrade differentially in a WobbleGC dependent manner (**Figure 3.3C).** Consequently, differences in steady state levels can at least partially be explained by differences in decay rates, without the need for concluding that synthesis rates or processing are responsible (**Figure 3.3D**). However, GFP1, which has the lowest WobbleGC, has significant overlap with GFP2, which has the next lowest WobbleGC (**Figure 3.4A**). This suggests that WobbleGC affects the decay rate within a specific range, beyond which other factors are contributing to observed mRNA degradation.


A
%GC (Wobble)
%GC (Total)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
GFP1
GFP2
GFP3
GFP4
GFP4
GFP3
GFP2
GFP1
0
100
200
300
400
500
600
700
Nucleotide Position


B
GFP:HTA1 (s.s.RNA)
0
10
20
30
40
GFP4
GFP3
GFP2
GFP1
Promoter
5'UTR
GFP
3'UTR

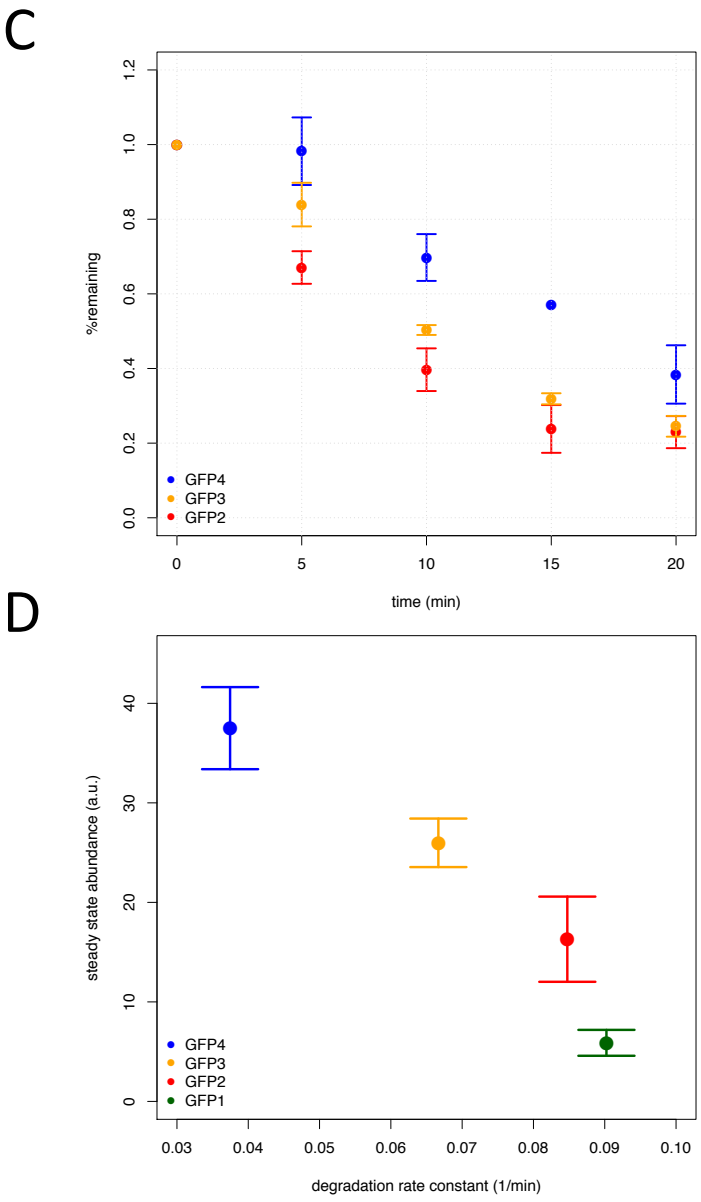


Figure 3.3: WobbleGC content affects decay kinetics and steady state levels

(a) The four GFP transcripts are similar in overall GC content (right), but are very different in WobbleGC (left). At the bottom is the alignment of each GFP relative to GFP2. Positions of similarity in sequence are depicted by gray line and differences are in triangles. (b) All GFP transcripts are under control of the same regulatory elements (inset). Differences in WobbleGC affect the transcript steady state levels. (c) GFP2-4 degrade with different kinetics in a WobbleGC dependent manner. (d) Degradation rate is at least partially responsible for observed changes in mRNA abundance levels.

We next investigated what other factors could explain the similarity in decay rate of GFP1 and GFP2. Our multiple regression analysis suggested that a variety of factors involved in translation contribute to variation in mRNA decay kinetics. Therefore, we looked to see if differences in translation could explain the similarities and differences in mRNA decay kinetics. As shown in **Figure 3.4B,** there is no clear relationship between the amount of protein produced and mRNA decay rate. However, when we determined translational efficiency, as measured by the amount of protein produced per total mRNA levels, we see that the rapidly degrading GFP1 and GFP2 have very similar efficiency (**Figure 3.4C**). In contrast, GFP3 and GFP4, which are more slowly degrading transcripts, are much lower in overall translational efficiency. These patterns are also true when we calculate the codon adaptation index (CAI) for each transcript, which is a separate measure of translatability (**Figure 3.S2**). From these data we conclude that within a certain range, an increase in the GC content of the wobble position decreases degradation rate. Simultaneously, decreased translational efficiency is associated with decreased degradation rate. Possible explanations and mechanism for this observation are discussed below.

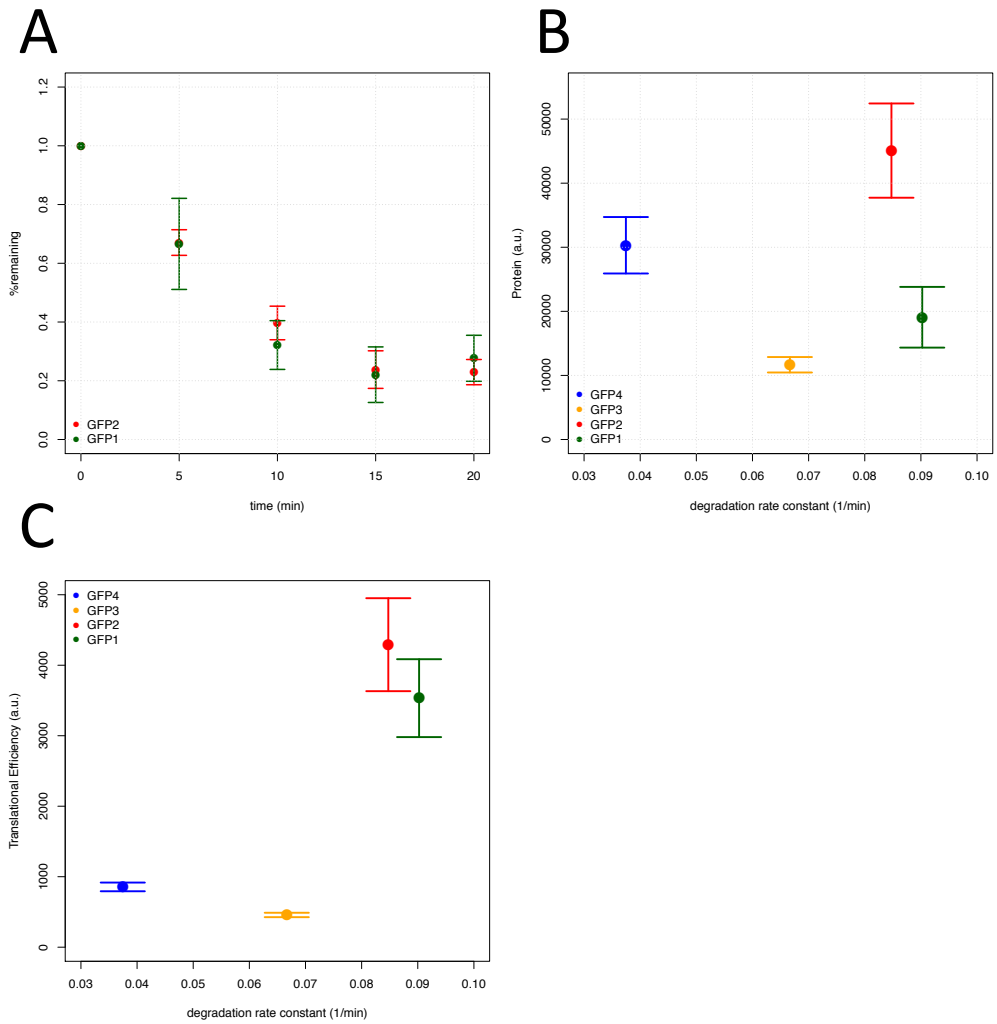


Figure 3.4: Translational efficiency contributes to degradation rates
(a) GFP1 and GFP2 are similar in their degradation kinetics. We calculated a decay rate constant for each strain using non-linear regression and compared it to (b) protein level of each GFP strain and (c) translational efficiency based on ratio of protein expression level with steady state mRNA levels. The transcripts with higher translational efficiency (GFP1 and GFP2) have faster degradation rate constants.

3.4: Discussion

The expression level of every gene is determined both by the rate at which it is synthesized and the rate at which it is degraded. Studies of several individual transcripts determined *cis* and *trans* features that had effect on observed mRNA decay rate. However, the variation in genome-wide mRNA half-life studies cannot be explained to a great degree based on any single transcript feature. In this paper we tried to address the reason for this observation, as well as investigate other determinants of decay rates through analysis of transcriptome wide datasets.

Through multiple regression analysis of each, we determined significant predictors of mRNA decay. We found that the RATE-seq dataset explained more variation than any of the other analyzed datasets. In addition many of the significant predictors of variation involved different aspects of translation.

Previous studies regarding the role of translation on mRNA degradation have shown differing results. In (Yun and Sherman 1996) *CYC1* was mutagenized such that all start codons were removed from its coding region with the intention of inhibiting ribosome binding and translation. They found that following transcription inhibition with thiolutin, the translationally impaired transcript degraded no differently than the translationally intact transcript. Similarly, in (Beelman and Parker 1994) translation of the MFA2 transcript was inhibited, but with strong secondary structures in the 5' region of the transcript. There too, following a transcriptional pulse-chase with the

gal inducible system, the MFA2 transcript degrades with similar kinetics. In contrast, the PGK1 transcript, when inhibited translationally with strong secondary structures as in MFA2, degrades more rapidly than in the absence of secondary structures (Muhlrad et al. 1995). A more recent analysis of transcriptome-wide decay datasets suggests that on average, increasing ribosome density increases mRNA stability (Edri and Tuller 2014), an observation we confirm in our decay dataset **(Figure 3.S3)**. A potential resolution of these observations, then, is that translation affects mRNA degradation to different degrees, depending on the transcript under investigation.

We attempted to determine the role of ribosome density for two separate transcripts, keeping all other features about the transcript and protein produced constant. Our approach was to mutate the start codon from ATG to GTG, to prevent ribosome formation at the start codon. By mutating the start codon of GFP, we generated a yeast strain that does not produce a fluorescent protein (**Figure 3.S4**), and should therefore at least be reduced in ribosome density. One scenario in which this would not be the case is if the GTG mutation can serve as a non-canonical start site of translation (Ingolia et al. 2009). However, in such a scenario, we would still expect a fluorescent protein to be produced. Since no fluorescence is observed, we must assume that the GTG is not a start codon for ribosome formation in this case. Therefore, if ribosomes form on the transcript, they do so further downstream

of this codon, potentially at nucleotide 234, which is the next start codon. For the GAP1 transcript, we were unable to show that mutation of the start codon results in loss of production of the GAP1 protein (**Figure 3.S5**). However, assuming that the GTG cannot serve as a start codon, the next start codon is 288 nucleotides (96 amino acids) downstream. Encoded within this putative non-translated region of the transcript are amino acid residues known to affect the extent of functionality and localization of the encoded permease (Merhi et al. 2011). This decrease in protein function may explain why the transcript is potentially destabilized and why we see a large drop in the steady state levels of GAP1 as well as an increase in the degradation rate.

One possible explanation for the difference between the GFP and GAP1 transcripts is the functional role of the encoded protein. The GFP transcript encodes a protein that has no role in the cell. In contrast, GAP1 encodes an amino acid permease that is necessary for efficient survival in nitrogen poor conditions. One possibility is that the function of the protein encoded by a given transcript affects the decay kinetics of the transcript itself. This is consistent with our previous observation that transcripts encoding genes in similar GO term categories have similar degradation rate constants (Neymotin et al. 2014). The relationship we see between ribosome density and mRNA decay rate on the transcriptome-wide level is then either a result of correlation with some unidentified factor affecting decay rate, or most likely, is

an indication of ribosome density affecting the decay rate of some, but not all transcripts.

Another predictor of variation that we investigated was the GC content of the wobble position in a transcript. This has previously been suggested to affect decay in human lymphoblastoid cells (Duan et al. 2013). This question was directly addressed in mammalian cells when looking at two different forms of GFP. In that study, where transcription was globally inhibited, they found an effect on mRNA levels, but not mRNA decay rates. In our studies we find that WobbleGC content determines both mRNA levels and decay rates. The effect is true within a certain range. Outside of that range, other factors contribute to observed decay rates, as is clear from two of our transcripts, which are very different in WobbleGC, but degrade with similar kinetics.

We searched for similarities and differences between the rapidly degrading and slowly degrading transcripts. Despite the large differences in decay kinetics between GFP2 and GFP4, their sequence is identical in the latter portion of their transcripts (**Figure 3.3A**). This suggested that any sequence factors affecting decay kinetics were in the first portion of the transcripts. Given that translation and mRNA decay are intimately linked, we checked how codon adaptation index (CAI), a measure of the translatability of the transcript, differed between the four strains in the initial third of the transcript, which is the source of most of the sequence divergence between GFP2 and GFP4. We found that the CAI of the two rapidly degrading strains is almost identical

(GFP1 and GFP2), whereas those of the more stable transcripts (GFP3 and GFP4) are lower (**Figure 3.S2**), suggesting similarities in translation efficiency. This observation is consistent with that found in (Carlini 2005), that more stable transcripts tend to have more rare codons, and should therefore have lower CAI. We then measured translation efficiency directly as the ratio of proteins produced per mRNA transcript, and also found GFP1 and GFP2 quite similar, with GFP3 and GFP4 yielding much lower levels. These data are potentially consistent with the idea that mRNA decay is co-translational, and that the ribosomes continue elongation in the 5'→3' direction, as they are being degraded by the exonuclease (Hu et al. 2009). The slower degrading transcripts, which are higher in WobbleGC, lower in CAI and lower in translatability would therefore have ribosomes proceeding more slowly along the transcript. However, this explanation does not explain why GFP3, which has lower translatability than GFP4, degrades more rapidly. Most likely the translational efficiency in addition to the effects of a change in GC content together affect the observed decay rate.

It is clear from our analyses that variation in mRNA decay rates is best explained by a combination of different transcript features as suggested more than two decades ago (Caponigro et al. 1993). At the same time, controlling for all factors while changing one feature at a time is difficult, as seen with WobbleGC content in our present study. We intended to change WobbleGC content only, but that resulted in changes to protein levels, mRNA steady state

abundance, and translatability. Interestingly, from our multiple regression analysis we find that methods for measuring RNA decay that utilize transcriptional inhibition tend to explain far less variation than less disruptive methods. RATE-seq, which does not perturb cellular physiology, seems to explain the largest amount of variation in decay rates as a function of natural gene features. Future studies of mRNA decay and the factors underlying variation need consider the physiological effects of the methods employed. Metabolic labeling methods, which minimally perturb the cell, rather than transcriptional inhibition, which causes rapid cellular death, should be employed so as to obtain physiologically relevant data. Although all the analyzed datasets of mRNA decay and predictor variables were generated under similar environments, ideally, a single study would measure decay, transcription, and protein levels in the same strain and environmental conditions. Our analysis of the available datasets give a glimpse into the complexities of mRNA decay and suggests WobbleGC and translation efficiency are features of interest for future decay studies.

3.5: Materials and Methods

3.5.1: Plasmid construction

Plasmid pCM188 (Garí et al. 1997) was the backbone for all plasmids. This CEN4 plasmid harbors the *URA3* gene, the tetracycline transactivator constitutively expressed, and a multiple cloning site with a *CYC1* TATA region upstream, all under control of two copies of the tetracycline operator.

Transcription of the gene of interest is repressed in the presence of tetracycline or its derivatives, which in our case is Doxycycline. Plasmids DGP147, DGP148, DGP149, and DGP231 are pCM188 with degenerate forms of GFP (Kudla et al. 2009) ranging in GC content in the third position of each codon as 0.71, 0.38, 0.6, and 0.67 respectively (GFP4, GFP1, GFP2, GFP3). The coding sequence of each GFP was cloned into the BamHI and NotI sites. Plasmid DGP198 is the same as DGP149, except the start codon of GFP has been mutated to GTG. Plasmid DGP217 is pCM188 with the *GAP1* gene + 3'UTR cloned into the BamH I and Not I sites. Plasmid DGP218 is the same as DGP217, except the start codon of *GAP1* has been mutated to GTG.

3.5.2: Strains and Growth Conditions

All yeast strains are FY3 (*MATa*, isogenic strain to the S288C derivative FY4 with a *ura3-52* mutation) with one of the above-indicated plasmids. DGY696, 697, 698, 898, and 1281 have plasmids DGP147, 148, 149, 198, and 231 respectively. DGY1193 and 1194 harbor plasmids DGP217 and DGP218 respectively, with the KANMX4 cassette inserted in place of the endogenous *GAP1* locus.

A single colony for each strain was inoculated for an overnight culture in synthetic complete media without uracil, to maintain selection of the plasmids. Studies of the GAP1 transcript were performed in nitrogen limiting media with proline as the limiting nitrogen source, as described in (Hong and Gresham 2014). Saturated cultures were back-diluted 1:50 into media of the

same composition as overnight. Cells were allowed to grow for 5.5 hrs (~2.5 doublings) before transcription was inhibited with Doxycycline at a final concentration of 10μg/ml. Cells were collected by filtration on nitrocellulose membranes and snap frozen in liquid nitrogen.

3.5.3: RNA extraction, RT, qRT-PCR analysis

RNA was extracted as previously described (Neymotin et al. 2014), using the hot phenol-chloroform method. Purified RNA was then treated with RQ1 DNAse according to manufacturer recommendations. Reverse transcription was performed using random hexamers and MMLVRT enzyme. Quantitative Reverse Transcription PCR (qRT-PCR) was performed using the Syber Green system and a Roche Light Cycler. RNA levels were quantified in comparison to the HTA1 housekeeping gene, which is unaffected by Doxycycline addition. Ratios were calculated using the formula: $Y=2^{(-(HTA1\_ct-Gene\_ct))}$ with ct being the calculated cycle threshold. RNA levels from each time point were then normalized to t=0, which was set to 1. All analyses with error bars are the mean +/- the standard error for 3-6 biological replicates. Values without error bars are the average of two replicates. Primer sequences had amplification efficiencies of at least 95% on RT products. The amplified product for all GFP strains begins between position 463 and 586 of the transcript. The amplified products are all between 80 and 120 base pairs long. The sequences used for qRT-PCR analysis are as follows:

HTA1:

Forward: 5'- GCTGGTAATGCTGCTAGGGATA-3'

Reverse: 5'- TTACCCAATAGCTTGTTCAATT-3'

GFP2, GFP4, GFP-ATG, GFP-GTG:

Forward: 5'-TTGCCGGATAACCACTACCT-3'

Reverse: 5'-CCTGCTGCAGTCACAAACTC-3'

GFP1:

Forward: 5'-GCCGATAAGCAGAAGAATGG-3'

Reverse: 5'-TGTTGATAATGGTCCGCAAG-3'

GFP3:

Forward: 5'- CGACCATTACCAGCAGAACA-3'

Reverse: 5'- GGGTCCTTTGACAGAGCAGA-3'

GAP1-ATG, GAP1-GTG

Forward: 5'-TTTGTTCTGTCTTCGTCAC-3'

Reverse: 5'-CTCTACGGATTCACTGGCAGCA-3'

3.5.4: Fluorescence measurements and images

Four biological replicates each of DGY696, DGY697, DGY698, and DGY1281 were grown to log phase and fluorescence was quantified using a Becton Dickinson Accuri cell counter. For each sample, the fluorescence of 1000 events was measured and the mean of those events was used as a single

value for the biological replicate. Fluorescent images were acquired using a Nikon Eclipse phase contrast microscope.

3.5.5: Multiple Regression Analysis

To normalize the data for our multiple regression analysis, we utilized decay rates rather than half-lives, which are typically log-normally distributed, and removed any outliers based on interquartile range. Decay rates for most datasets were calculated as $\ln(2)/t_{half}$, except in (Munchel et al. 2011) and (Neymotin et al. 2014) where the effects of dilution as a result of cellular growth was also considered. Counts were taken from (Lipson et al. 2009) and normalized based on an assumption of ~60,000 mRNA/cell (Zenklusen et al. 2008). Protein per mRNA was calculated as the values from (Ghaemmaghami et al. 2003) divided by the values for mRNA transcript/cell. Codon adaptation index (Sharp and Li 1987) was calculated for each transcript based on the codon frequency tables in the seqinr package in R (R core team). To normalize the data we log transformed the predictor variables ($\log_{10}$(Variable) or $\log_{10}$(Variable +1)), except for GC content of each codon position, which is approximately normal in distribution, and deltaG, which is negative in value.

In a linear multiple regression model (Crawley 2012), a parameter of interest is modeled as being dependent on two or more predictors. We use the degradation rate constant as the parameter of interest, and the other measurements as predictors. To build the model, we followed two separate approaches. In approach 1, we first determined the p-value of the pair wise

correlation of each predictor to degradation rate. This indicates whether the regression coefficient is significantly different from zero, and whether or not the predictor has any effect on decay rate. Next we included all predictors that have a p-value less than 0.05 into the multiple regression models. We then performed stepwise deletion of terms by removing sequentially the predictors with the highest p-value. The final model is then the reduced model where only significant terms remained. In approach 2, we performed the regression in a slightly different manner. We first calculated the significance of each predictor when it is the only one in the model, as above. We then started adding to the model based on the predictors with the lowest p-value. With each additional term we checked to see that all of the terms in the model were significant. If a new term was added and it was not significant, we removed it from the model. If a new term was added and a different predictor lost its significance, we tested a model with either the new predictor or the one that lost significance, and retained the one that explained more variation. We did not add terms that were insignificant in the pair wise correlation with decay rate. Both models gave similar results regarding which predictors were significant

We must note that several predictors are highly correlated with one another including the number of transcripts with ribosome density and protein per cell, and coding length with free energy (**Figure 3.S6**). In effect, this means that if either of these pairs is in a model together, the variation that

could be explained by one of these predictors is distributed (although not necessarily equally) among the correlated variables. Nevertheless, the fact that a combination of multiple predictors explains more variation than any one of the factors individually, strongly suggests that the combinatorial effects of multiple predictors determine mRNA decay kinetics.

3.5.6: R functions and packages

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

3.6: Supplemental Figures

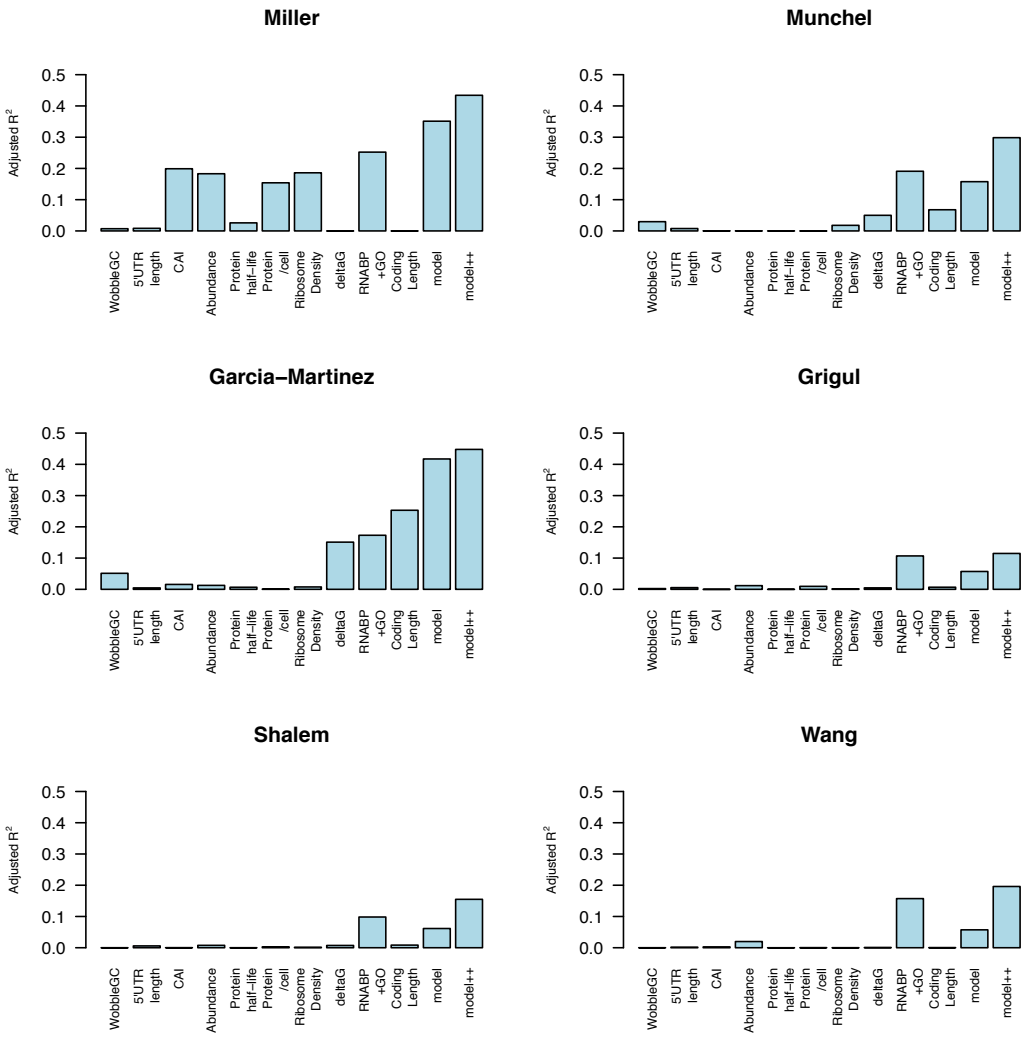


Figure 3.S1: Multiple predictors best explain variation in mRNA degradation in all studied degradation datasets
We show the adjusted R^2 for individual predictors in each model, as well as for the model with multiple terms ("model"), GO terms with RNA binding Proteins ("RNABP+GO"), and the final model which includes both continuous variables as well as the categorical GO and RNABP ("model++). First author indicates each dataset.

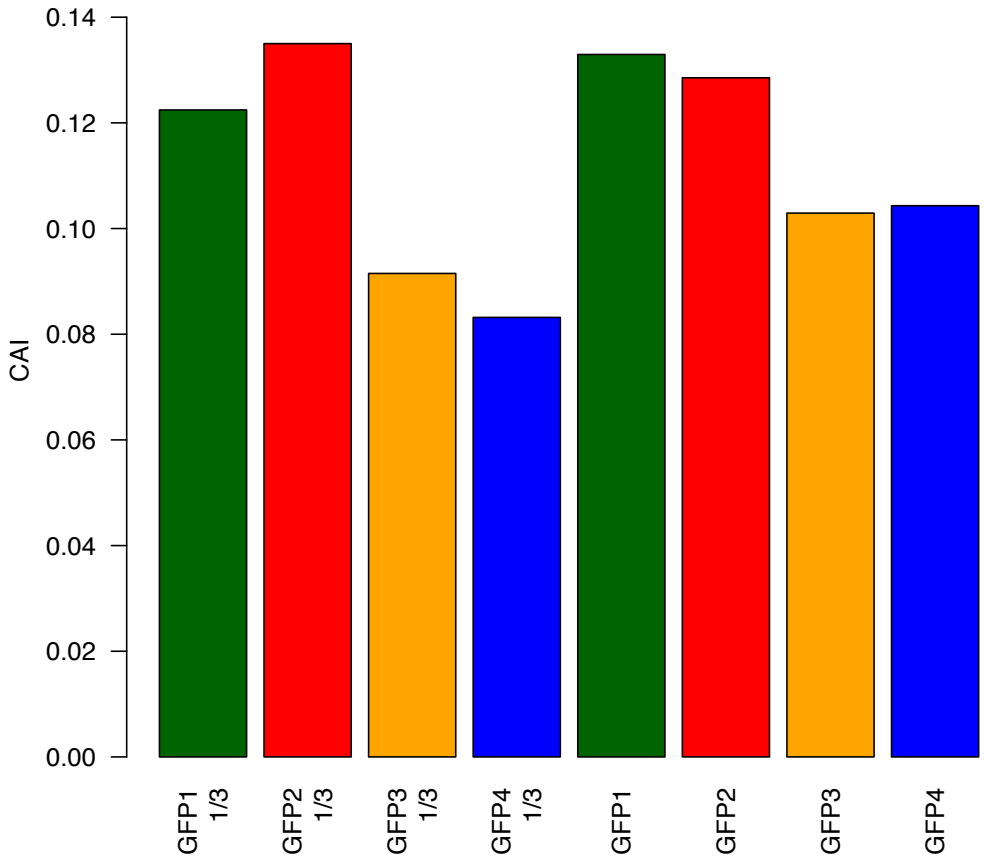


Figure 3.S2: CAI confirms translational effects on mRNA degradation
We calculated the codon adaptation index (CAI) for all four GFP transcripts. The "GFP 1/3" indicates the CAI of the first 240 nucleotides, which is the primary source of differences between transcripts. The "GFP" indicates the CAI for the entire transcript.

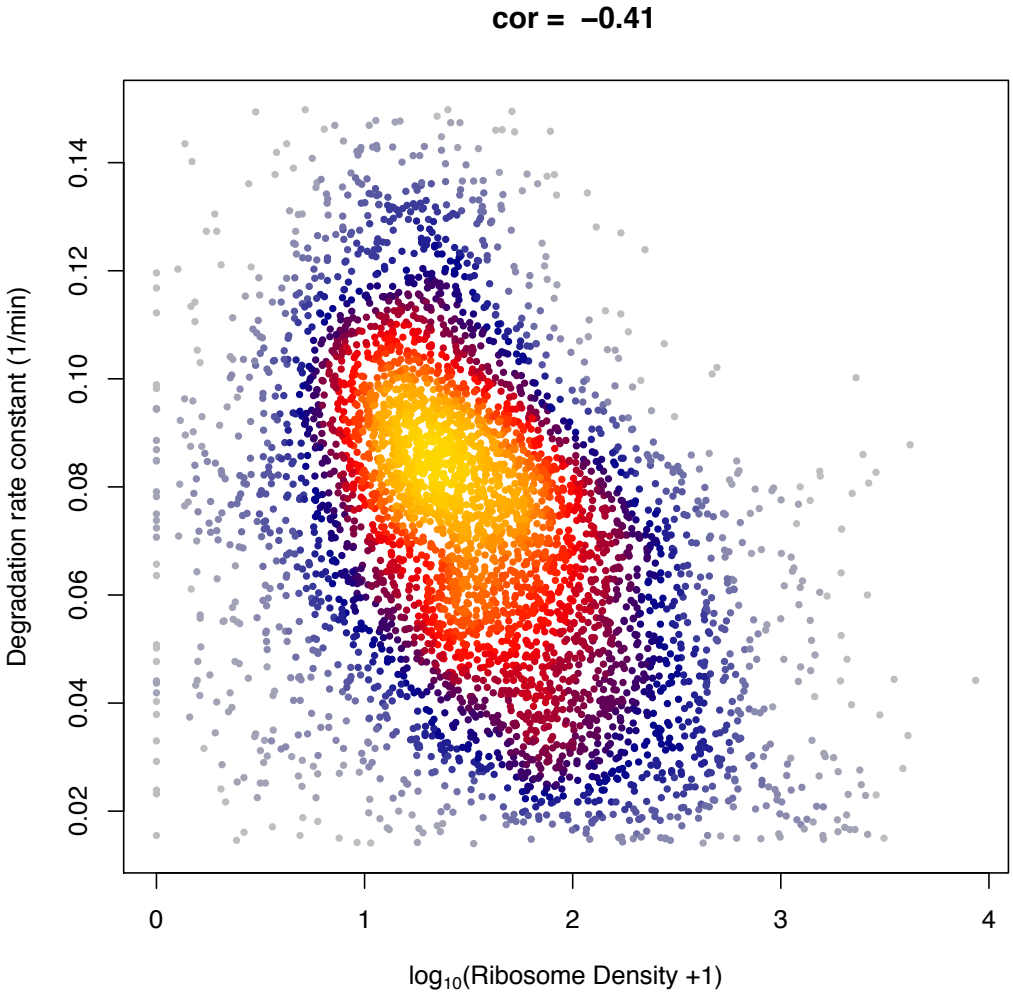


Figure 3.S3: Degradation rate decreases with increased ribosome density
The ribosome density of a transcript is inversely correlated with degradation rate, showing that the more ribosomes present, the more stable the transcript. Data are from (Neymotin et al. 2014).

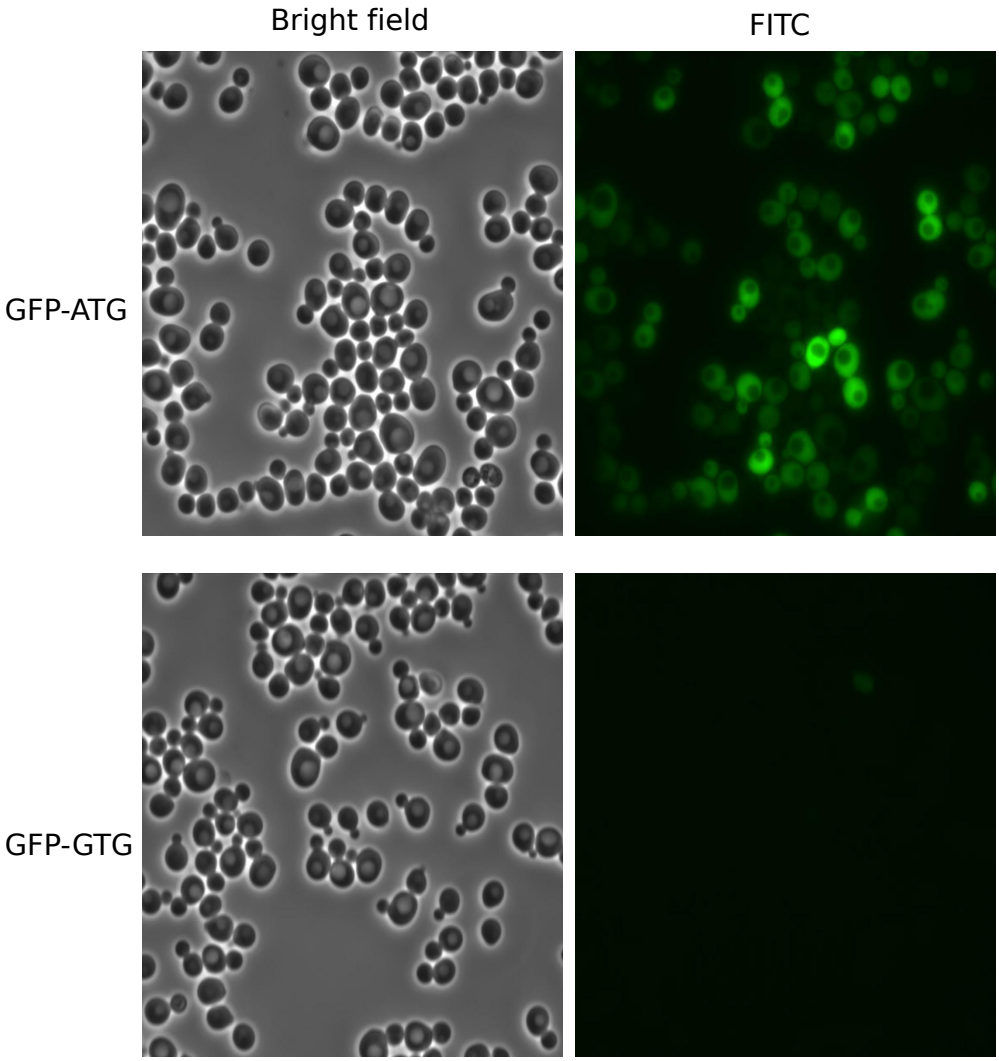


Figure 3.S4: Mutation of ATG disrupts fluorescent protein production
Using phase contrast microscopy, we visualized the GFP transcript that has either the wild type ATG start codon, or the GTG in place of the start codon.

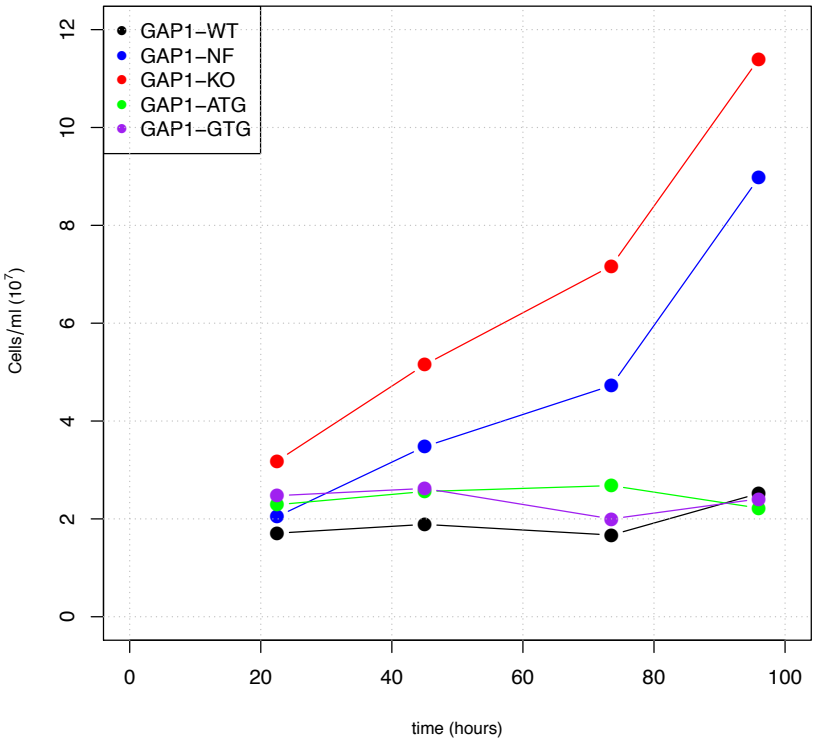


Figure 3.S5: Effect of mutated start codon on GFP protein
We grew yeast in minimal media containing D-histidine (D-his). D-his is toxic to cells and leads to cell growth inhibition and death. In the absence of a functional GAP1 protein, cells do not have the permease and do not take up D-his. We show five strains: GAP1-WT=wild type strain, GAP1-KO=strain with complete knockout of coding sequence of GAP1, GAP1-NF=Derived strain with non-functional GAP1, GAP1-ATG=plasmid GAP1 with normal start codon, and GAP1-GTG=plasmid GAP1 with GTG in place of start codon.

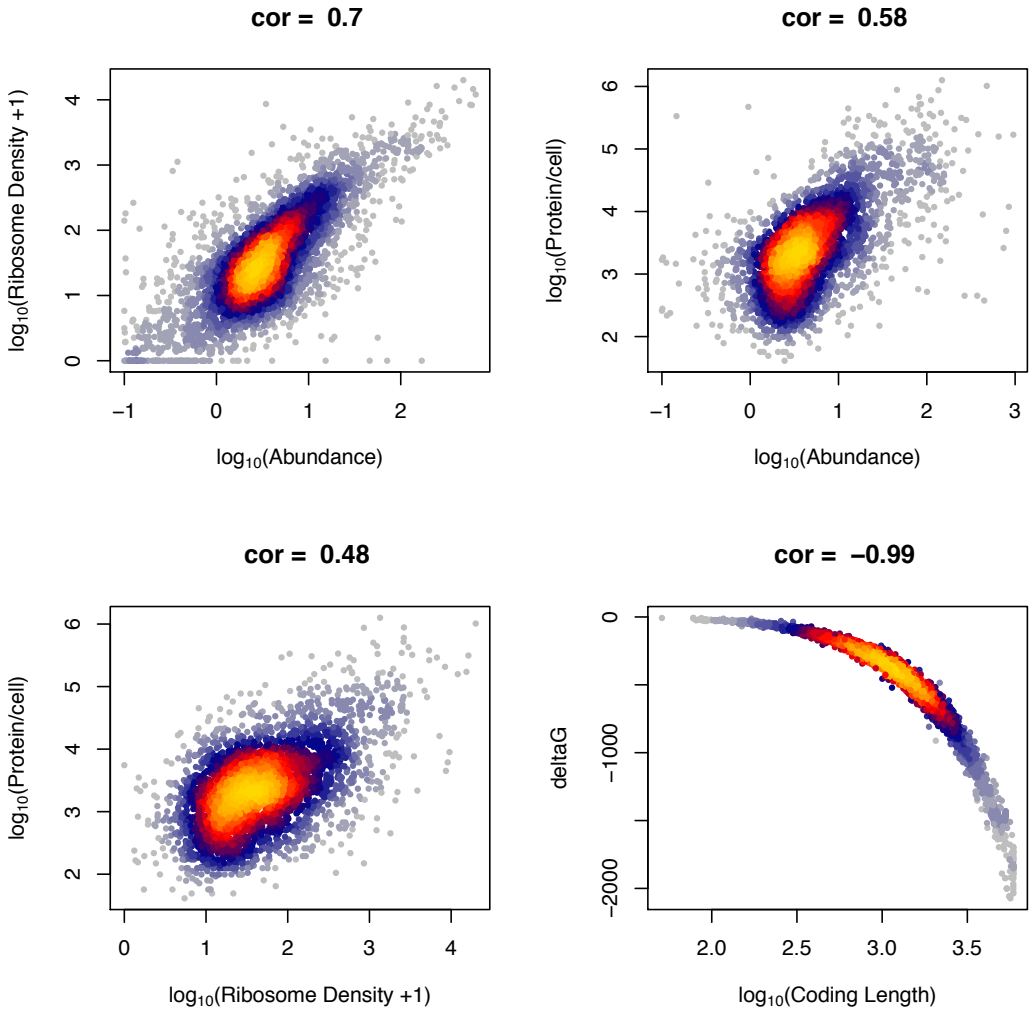


Figure 3.S6: Predictors of mRNA decay variation are highly correlated
We show that a number of predictors are highly correlated. The spearman rank correlation is indicated for each pair wise comparison.

Tables:
Table 1: Source for each dataset with number of observations.
Supplemental Tables:
Table S1: Matrix with all values used for regression analysis.
Table S2: Matrix with p-value, adjusted R^2, and correlation for each predictor and a given decay rate.
Table S3: Values for all qPCR experiments.

CHAPTER 4: GLOBAL TUNING OF GENE EXPRESSION AS A FUNCTION OF CELLULAR GROWTH RATE

4.1: Abstract

Gene expression programs are regulated in response to extracellular signals. At the same time, these signals can affect the rate of cell growth. Previous studies using microarrays have shown an effect of growth rate in relative mRNA abundance. However, these data are potentially confounded by the observation that the total RNA amount per cell is also a function of growth rate. In this study we determined the absolute abundance of all transcripts on a per cell basis as a function of growth rate, utilizing externally provided RNA spike-ins. We find that all classes of RNA increase as a function of increasing growth rate, but to different extents. This suggests that assumptions of constant relationships between different classes of RNA are incorrect, but that the use of external spike-ins can address this issue. To determine the extent to which mRNA synthesis and decay affect observed changes in mRNA levels, we applied RATE-seq to chemostat cultures of three different growth rates. We find that both mRNA synthesis and degradation increase in response to cellular growth rate. However, the response to growth rate of mRNA synthesis exceeds that of degradation thereby resulting in increased steady-state abundance of virtually all mRNAS. This study provides the first complete picture of transcriptome dynamics as a function of cellular growth rate.

4.2: Introduction

Regulated cell growth is critical for the normal development and viability of all organisms. In the early developing embryo, the initial fertilized egg undergoes rapid nuclear replication and cellular division. All initial protein products are generated from maternally loaded mRNA and protein, which must be degraded in proper space and time (Schier 2007). Subsequent cellular growth depends on the differentiation of the cell type as well as environmental cues (Lodish et al. 2008). Similarly in single celled organisms, growth rate is tightly regulated as a function of nutrient availability (Ziv et al. 2013b), as well as other environmental factors. In the absence of proper regulation of growth rate and cell cycle, the development and well being of an organism may be compromised. In addition to the growth rate itself, gene expression levels must be regulated to ensure proper cellular survival. Both growth rate and gene expression levels determine cellular homeostasis, but the extent to which one determines the other is unclear.

Genome-wide studies of mRNA levels provide the most comprehensive view of gene regulation (Gasch et al. 2000). In a single microarray or RNA-sequencing run, it is possible to observe the gene expression profile of a given organism (DeRisi et al. 1997; Nagalakshmi et al. 2008). However, analysis of such large data necessitates extensive normalization for removal of technical artifacts introduced during sample processing. A common strategy involves the assumption of total RNA levels not changing across samples and a mean normalization to some arbitrary amount. Often times such assumptions are

invalid. Different cell types are known to have different total mRNA amounts and the same cell type may have different amounts of total RNA as a function of growth rate or nutrient availability (Waldron and Lacroute 1975). Recent studies have shown that such mean normalization in microarray technology can mask underlying biological phenomena, simply as a result of incorrect assumptions of constant RNA levels (Lovén et al. 2012). In addition, it is not known if the relative abundance of different types of RNA, for example rRNA or mRNA, within a cell change in response to environmental conditions. Therefore, comparisons that rely on assumptions of equivalent total RNA levels may be misleading analyses.

The rate of cell growth has a profound effect on gene expression. In yeast, more than 25% of the yeast transcriptome changes in abundance level as a function of growth rate (Brauer et al. 2008). Such changes in mRNA levels can be achieved through modulation of synthesis rate, degradation rate, or both (Hargrove et al. 1991). Analyses of a GFP library measuring promoter activity indicates that the entire transcriptional activity of the genome is scaled as a function of growth rate (Keren et al. 2013). In addition, individual transcript studies in *E. coli* indicate that certain transcripts are regulated at the mRNA degradation level (Nilsson et al. 1984). These observations indicate that RNA levels are controlled with growth rate, and that each process of synthesis and degradation may play a role in this regulation.

In this study, we quantified the effect of cellular growth rate on changes in mRNA levels. Using externally provided RNA spike-ins (External RNA Controls Consortium 2005; Baker et al. 2005), we determined the absolute number of RNA transcripts on a per cell basis. Results using RNA spike-in are highly reproducible and obviate the need for normalization based on assumptions of unchanging RNA levels. We show that RNA levels are directly affected by changes in cellular growth rate. Using our recently described method of RNA Approach to Equilibrium Sequencing (RATE-seq) (Neymotin et al. 2014), we explored the underlying causes of changes in RNA levels by measuring synthesis and degradation rates genome-wide. RATE-seq provides a means for determination of kinetic parameters without compromising cellular physiology. We find that both synthesis and degradation rates are regulated in a growth rate dependent manner, but synthesis is the primary determinant of increases in mRNA levels associated with increased growth rates. Our study shows that much of the transcriptome is globally tuned in a growth rate dependent manner.

4.3: Materials and methods

4.3.1: Strains, media, growth conditions, and sample collection

We performed all experiments using FY4, a MATa haploid derivative of S288C. Cells were grown for two days on YPD plates. A single colony for each experiment was inoculated overnight in minimal media containing a limiting

nutrient concentration of either glucose or ammonium sulfate. Glucose limiting media contained: 1x vitamins, 1x metals, 100uM uracil, 1x carbon limited salts, and 0.08% glucose. Nitrogen limiting media contained: 1x vitamins, 1x metals, 100uM uracil, 1x nitrogen limited salts, 2% glucose, and 400uM ammonium sulfate. Chemostat vessels were established as in (Ziv et al. 2013a) at a volume of 450ml. We monitored the dilution rate over several hours to confirm that the culture was being diluted at the desired rate. For each nutrient limitation, we performed our analysis at dilution rates of 0.12/hr (slow), 0.2/hr (medium), and 0.3/hr (fast) in biological triplicate. Once the desired dilution rate was attained, cultures were grown for ~8 generations (generation time = ln(2)/Dilution rate). We determined a steady-state was attained by measuring cell density using a Coulter particle counter.

To sample a defined number of cells we collected approximately 12mL of culture from the chemostat vessel by lowering the drop tube. Exactly 10mL of culture was then pipetted and cells were collected on nitrocellulose filters. These samples, designated as steady state samples, were further processed as below.

We next performed RATE-seq (Neymotin et al. 2014) on the remaining culture. We added 4-thiouracil (4tU) in a single bolus to the chemostat vessel to instantaneously shift the concentration to 500uM 4tU. Simultaneously, we switched the feed media lines to a media carboy containing the identical same medium plus 500uM 4tU already present (**Figure 4.1A**). This ensures that the

label is not diluted out as the experiment continues. We determined the appropriate time points for sample collection for each dilution rate based on the labeling kinetics (**Figure 4.S1**). For each dilution rate, we collected 8 samples of ~25ml each. The dilution rate (D) is determined by the flow (F) in mL/hr and volume (V) in mL of the culture as D=F/V. With each collected sample we change the volume of the culture. To maintain a consistent dilution rate, we adjusted the flow to account for the change in volume. After each sample, we changed the pump settings accordingly. At several points throughout the experiment we measured cell density to confirm that our adjustments were compensating for the change in volume. Following filtration of the samples we immediately placed them in liquid nitrogen.

4.3.2: Quantitative RNA preparations

We determined the cell density of the steady-state chemostat cultures by taking cell measurements with a Coulter counter (Beckman Coulter) in technical triplicate and calculating the average of these measurements. A defined volume of each steady-state culture was harvested for RNA extraction in eight technical replicates. Extractions were performed using the acid phenol method. Purified RNA was resuspended in a pre-defined volume of molecular grade water and the sample concentration was measured.

4.3.3: Processing of steady state samples and library construction

Before RNA extraction of steady state cultures, 0.2µl of the ERCC spike-in mastermix (mix 1, Life Technologies) was added to 10 million cells together

with the Lysis Buffer. RNA was extracted using a standard hot acid phenol protocol and depleted of rRNA (Ribo-Zero Magnetic Gold Kit [Yeast], Epicentre). Sequencing libraries for both steady state and RATE-seq samples were prepared using the directional RNA-seq protocol (Parkhomchuk et al. 2009). We used random hexamers without polydT for reverse-transcription to avoid bias against transcripts that are not polyadenylated. Samples were converted to cDNA and ligated to custom adapters for multiplex sequencing on an Illumina Hiseq 2500.

4.3.4: RATE-seq sample preparation

We performed the biotinylation reaction as in (Neymotin et al. 2014). We biotinylated ~45-95ug of total RNA for each sample. To normalize libraries we pooled four different in-vitro transcribed thiolated transcripts. Three of the spike-ins are described in (Neymotin et al. 2014) and a fourth is based on the sequence of B0041.8.1 in *C. elegans*. For every 100ug, we added 4 ng of each thiolated spike-in. The same pool of spike-ins was used for all biotinylation reactions. Following biotinylation, samples were spun and the supernatant transferred to a new tube as in (Zeiner et al. 2008). RNA was precipitated with one volume isopropanol and 1/10 volume 5M NaCl. Pellets were washed with 70% EtOH and resuspended in nuclease free water such that the final concentration was ~500ng/ul. Pulldown of labeled RNA was as in (Neymotin et al. 2014), except we washed the beads four times, one of which was with 65C wash buffer, and 1ul of 20mg/ml GlycoBlue was used for

precipitation of the RNA pellet. The RNA pellet was resuspended in 10ul of nuclease free water. To remove unwanted ribosomal transcripts, we utilized the RiboZero kit (Epicentre), following a scaled down version of the manufacturer's protocol. We then proceeded to make libraries for sequencing.

4.4: Results

4.4.1: Chemostat enables precise control of growth rate

To study the effects of cellular growth rate on RNA metabolism, we utilized chemostat cultures (**Figure 4.1A**) (Hoskisson and Hobbs 2005; Ziv et al. 2013a). In the chemostat, fresh media limited in a particular nutrient is supplied to growing cells at a particular rate. Simultaneously, old media and cellular debris are expelled from the culture. By modulating the rate at which the culture is diluted it is possible to establish a scenario of steady state growth, where the growth rate is directly determined by the dilution rate of the culture. Unlike experiments in batch culture where growth rate can be controlled by use of different molecular forms of a key nutrient, the chemostat can control growth rate without secondary effects of different nutrients. Therefore, the chemostat is the ideal system for determining effects of growth rate on gene expression programs and the underlying kinetic processes.

We first analyzed the contribution of cellular growth rate on different physical properties of the cell. We grew 18 different chemostat cultures corresponding to three different dilution rates, in two different nutrient limitations, in biological triplicate. As expected, the culture density changes in

a growth rate dependent manner, such that cells growing more slowly have higher culture density (**Figure 4.1B**)(Brauer et al. 2008). In addition the bud index, which is a measure of the proportion of cells progressing through the cell cycle, increases with growth rate (**Figure 4.1C**) as expected (Johnston et al. 1980; Brauer et al. 2008). Faster growing cells tend to be larger (Tyson et al. 1979), which we find to be the case in our analysis as well (**Figure 4.1D**). The culture density, bud index, and cell size are all determined in a growth rate dependent manner.

Using quantitative RNA preparations, we found that the total RNA content increases with growth rate (**Figure 4.2**), consistent with previous observations (Waldron and Lacroute 1975). Interestingly, as with the physical properties of cell size, culture density, and bud index the degree of difference between RNA content as a function of growth rate is dependent on the nutrient limitation: cells that are growing at different rates in environments limited for nitrogen show a stronger response to growth rate variation than cells growing in carbon limitation.

Because with this method we measure total RNA content, it is impossible to know if only certain RNA classes (such as rRNAs, tRNAs, mRNAs,

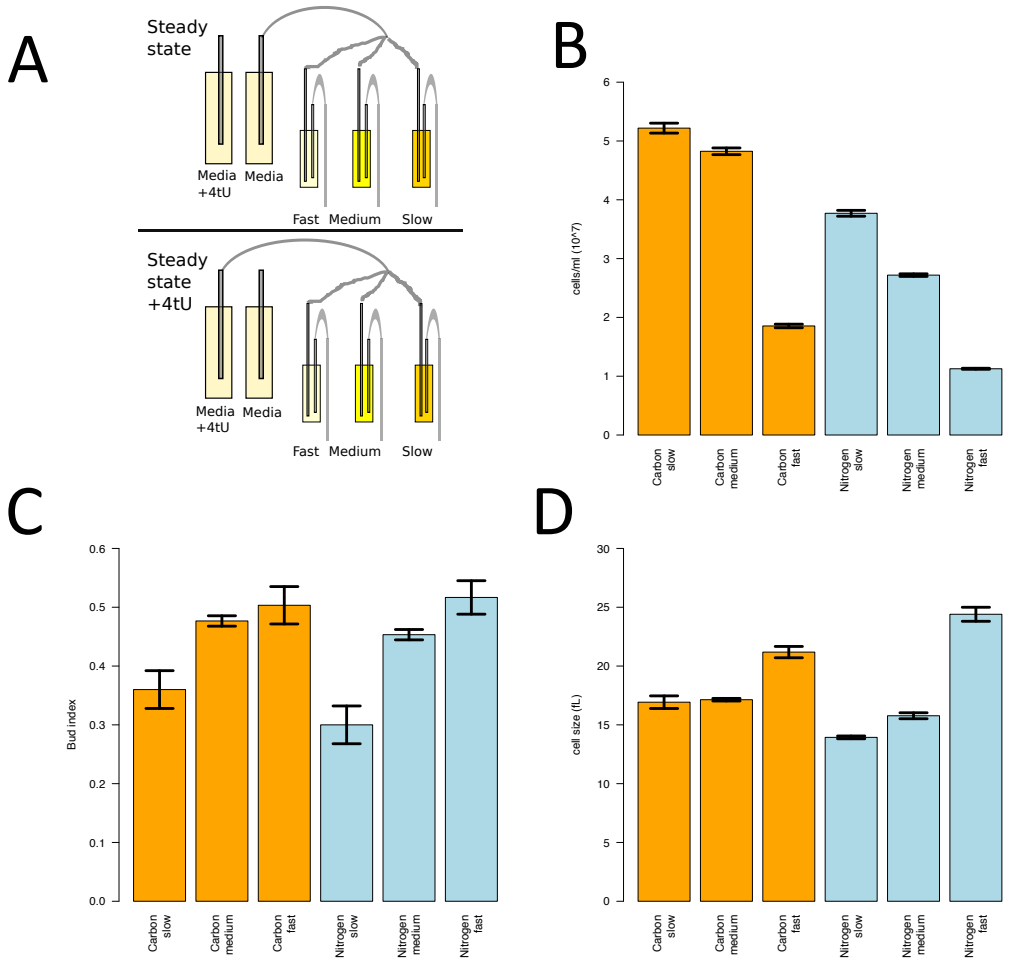


Figure 4.1: Chemostat maintains growth rate at different steady state levels

We grew yeast cultures at three different growth rates by supplying a continuous flow of new nutrients. The culture density of each vessel, as indicated by the darkness of color, increases with increasing growth rate. For the RATE-seq experiments, we added a bolus of 4-thiouracil as well as switched the media lines to a media carboy that contains the exact same media, but with 4-thiouracil in it as well. Once steady state was attained, we measured the B) cell density, C) bud index, and D) cell size.

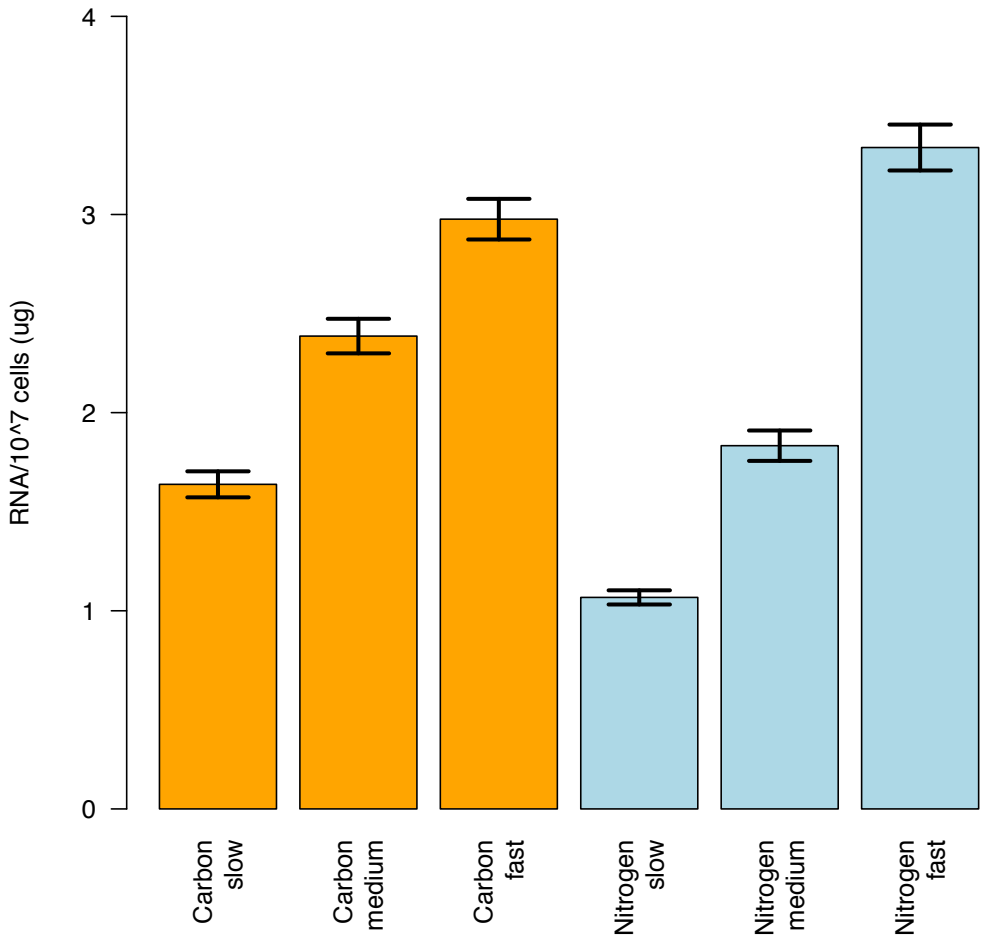


Figure 4.2: Quantitative RNA content analysis
We performed RNA extraction for samples from each growth rate from a defined number of cells.

etc) are disproportionately contributing to variation in RNA content or it reflects a globally coordinated increase in transcription with increasing growth rates. To address this question we performed directional RNA-seq analysis of all growth rates in both limiting conditions.

4.4.2: External RNA spike-ins are accurate and reproducible calibrators in RNA-seq experiments

Normalization of genome-wide data is imperative for proper interpretation of the results. Common normalization methods assume that differences in absolute intensities between samples (*e.g.* total microarray signal or library sequencing depth) are due to technical variation (*e.g.* pippeting errors, variation in reaction efficiencies, errors in quantification etc). This assumption is necessary as this variation is introduced routinely and needs to be eliminated to ensure reproducibility in high-throughput experiments (Weis and Consortium 2005). However, recent studies have revealed scenarios where the total intensity of the sample measured is not an experimental artifact, but reflects a true total transcriptome amplification (Lovén et al. 2012). In this case, the use of external RNA spike-ins as calibrators enables accurate measurement of changes in gene expression that would otherwise be masked with the standard normalization procedures.

To evaluate the use of spike-ins for normalization purposes, we added external RNA spike-ins to a defined number of cells rather than to a defined

amount of RNA. We aliquoted the same biological sample (C-limited chemostat, growth rate=0.35) into six replicate tubes, each containing the same number of cells. Prior to RNA extraction, the same amount of spike-in mix was carefully added to three of the cell aliquots, while the other set of three tubes received 4.4-fold more spike-in per cell in each tube. The results in **Figure 4.3** support the conclusion that this method is highly reproducible and yields accurate quantitative measurements.

4.4.3: Calculation of growth rate effects on gene expression

We next applied external RNA spike-ins to our growth rate samples. We added the same amount of spike-in to 10 million cells of each of the 18 different chemostat samples. Sequencing data show that the correlation between biological replicates is strong (Figure S2) and that the strength of the correlation decreases across growth rate. Consistent with observations above, the variation is higher between growth rates in the nitrogen limiting conditions than in the carbon limiting conditions.

We next tested how differential expression using spike-ins compares with a more commonly used normalization method, namely EdgeR (BioconductoR). Specifically, we determined the effect of EdgeR normalization on genes we found highly (50-20 fold) or lowly (1.2-2 fold) differentially expressed between growth rates based on spike-in normalization. We find that EdgeR normalization underestimated the extent of differential expression for the highly differentially expressed genes, and reversed the direction of

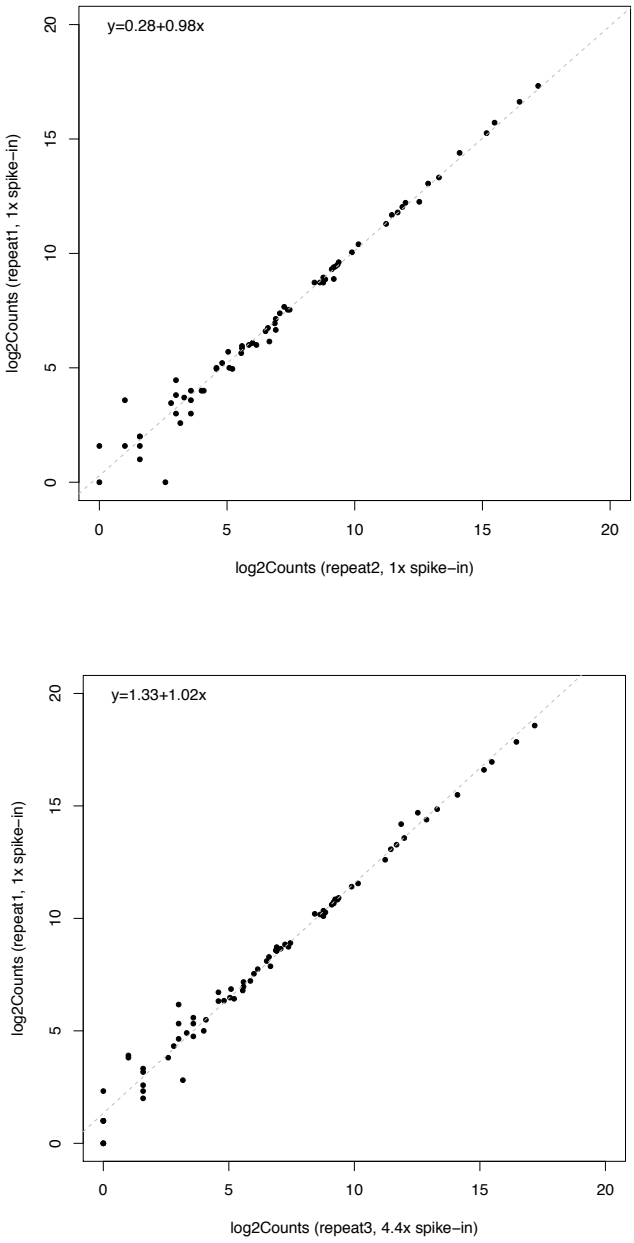


Figure 4.3: Use of RNA spike-in enables accurate and reproducible quantification of RNA transcript number.
We added the same (top) or different (bottom) amounts of RNA spike-in to two separate containers containing an equal number of cells from the same biological sample. We then plotted the length normalized transcript counts for RNA spike-in.

differential expression for the lowly differentially expressed genes (**Figure 4.S3C**). This observation agrees with the observations of Loven *et al.* (2012) and our simulations (**Figure 4.S3A and 4.S3B**).

Given the similarities between our experimental design and a previous study (Brauer et al. 2008), we can compare our results using spike-ins to those using microarray normalization. For consistency, we performed the comparisons on similar growth rates between studies (between 0.1 and 0.3 h^{-1}) and in samples derived from the same limiting conditions (glucose- and ammonium-limiting) as present in the previous study. We also log-transformed our data prior to calculating the gene-wise slopes—which are a measure of the expression response of each gene to the growth rate—for compatibility with the M values used previously.

Comparison of each gene's slope response to growth rate between the two methods (**Figure 4.S4**), confirmed that our gene-wise measurements correlate significantly with the previous dataset. However, there were big differences between the two distributions of the estimated slopes (**Figure 4.4**). In the analysis of (Brauer et al. 2008) the authors concluded that certain Gene Ontology classes of transcripts are affected positively and others negatively by the growth rate (positive and negative slopes). We hypothesized that our measurements, calculated in an RNA content-independent manner, would reveal a greater number of transcripts responding positively to the growth rate. The results in **Figure 4.4** show exactly that; not only is the

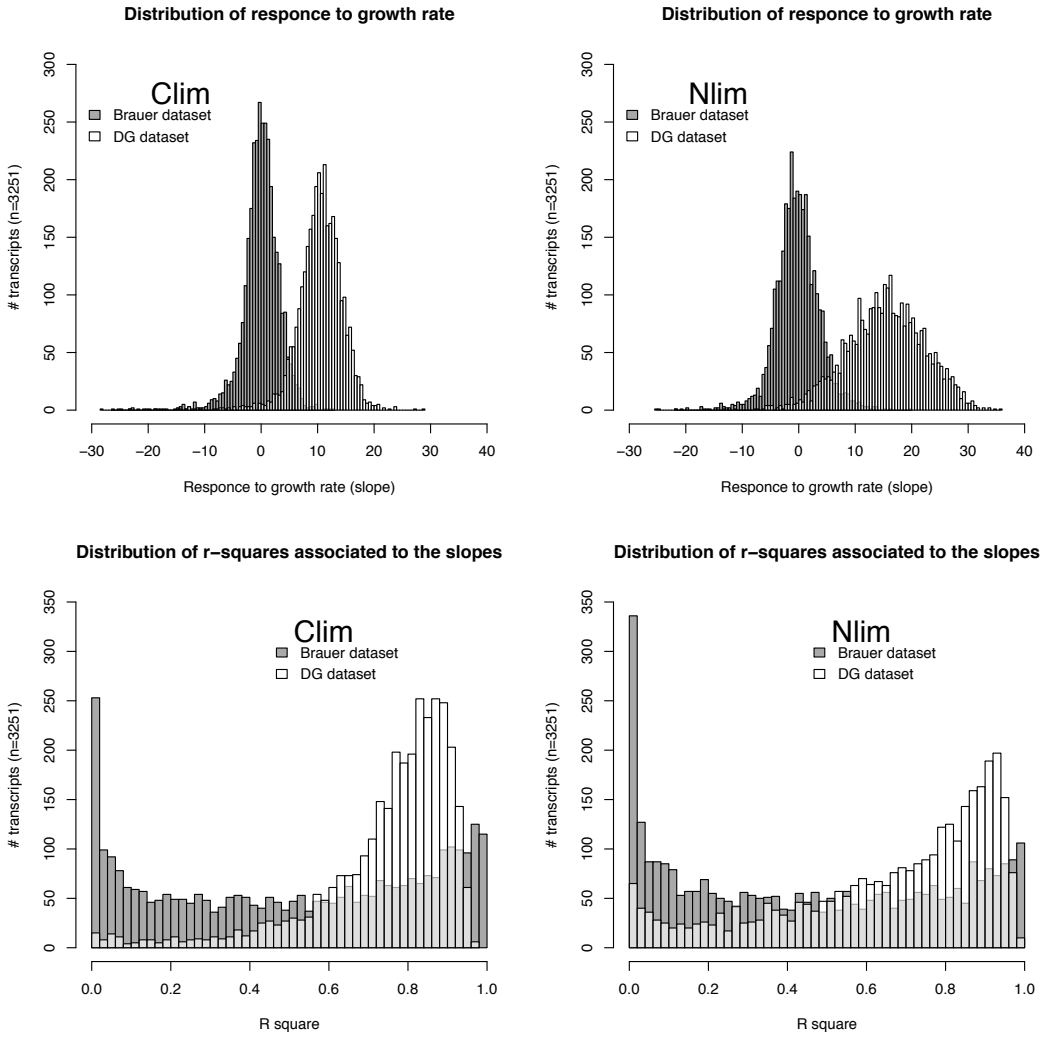


Figure 4.4: Accurate quantification using RNA spike-ins
(A),(B) Distributions of the gene-wise intensities of response to growth rate (slopes) as measured in this (light) and previous (dark) studies (Brauer, 2008). (C),(D) r squares of the goodness-of-fit when calculating the slopes. (A),(C) carbon limited samples (B),(D) nitrogen limited samples.

distribution of the slopes calculated in this study shifted to higher values, the vast majority of the transcripts we measured have a positive slope. This means that the observed increase in RNA content with growth rate (**Figure 2.4**) is due, in part, to a global increase in mRNA transcript numbers for almost every gene. Moreover, using external RNA spike-ins for sample calibration seemed to result in a better model fit to the data (r-squares of the correlations calculated for each estimated linear model fit of the data, **Figure 4.4**).

Using abundance values calibrated by external RNA spike-in standards seemed to outperform traditional normalization methods. However, the comparison is potentially confounded by the different nature of the measurements we were comparing. In (Brauer et al. 2008) the measurements used for the calculation of slopes were M values, the log signal ratio of a comparative high-density microarray hybridization experiment. In our study we used log-transformed estimates of molecules per cell estimates derived from directional RNA-seq.

To exclude the possibility that the different nature of the two measurement units was acting as a scaling factor, we set the mean value of the lowest growth rate in each dataset to one, and then normalized all other values to this dataset-specific constant. This way the two studies become comparable, albeit in arbitrary units. After this numerical transformation, the differences between the two datasets was less pronounced but the overall trend remains the same: our calculated slopes were consistently larger and

primarily positive (**Figure 4.S5**).

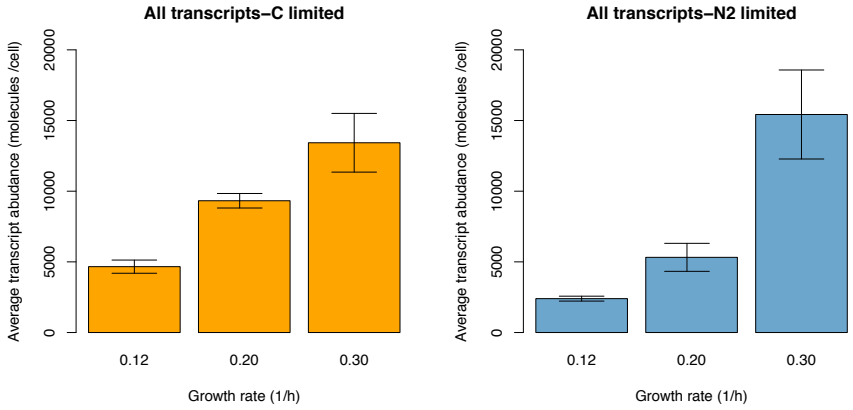
4.4.4: Responses of transcript abundance to changes in growth rate

We next investigated the effect of growth rate on all classes of RNAs within a cell. We examined the total number of RNA molecules in each cell, as measured by RNA-seq (**Figure 4.5A**). These measurements are a clear reflection of the results obtained by quantitative RNA purification (**Figure 4.2**) in that we observe an increase in the total RNA content per cell with increasing growth rates, and this effect is more pronounced in the nitrogen-limited conditions. Further exploration of each RNA class individually (**Figure 4.5B-4.5 D**) indicates that the effect of growth rate on gene expression is global, affecting all transcript classes examined; mRNA (**Figure 4.5B**), non-coding RNA (**Figure 4.5C**) and tRNA (**Figure 4.5D**) levels all appear to increase with increasing growth rate. Whether this is a result of changes in RNA synthesis or degradation is unclear from this analysis of total RNA abundance levels.

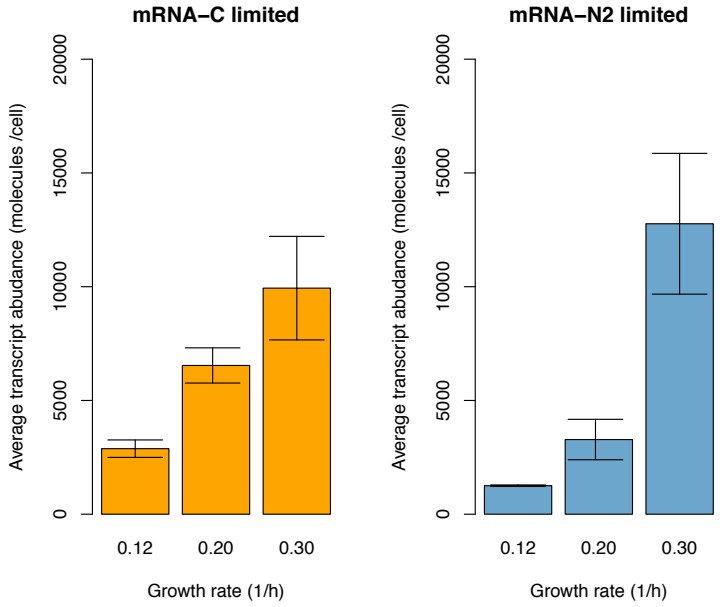
4.4.5: Degradation rates are regulated in a growth rate dependent manner

The absolute abundance of a transcript is determined by its rates of synthesis and degradation. To study the contribution of variation in degradation rate to observed differences in mRNA abundance, we performed RATE-seq on nitrogen-limited chemostat samples grown at different growth rates. Following normalization, we performed regression analysis to estimate decay rate constants for all genes at three different growth rates (**Figure 4.6A**). For the slow, medium, and fast growing cells, we determined rates for

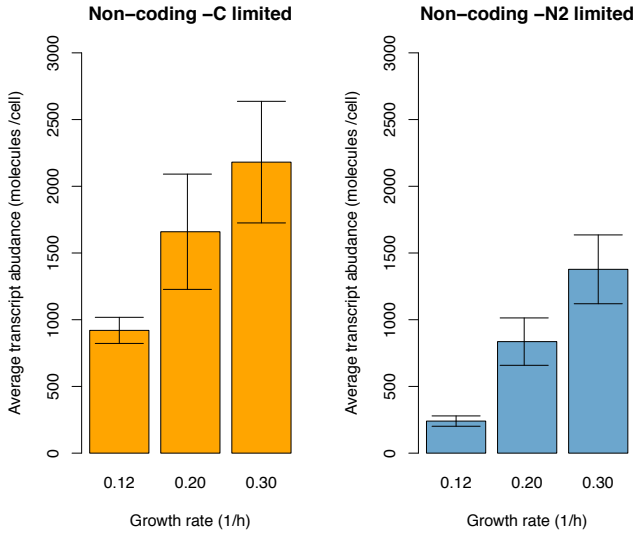
A


All transcripts–C limited
Average transcript abudance (molecules /cell)
0
5000
10000
15000
20000
0.12
0.20
0.30
Growth rate (1/h)
All transcripts–N2 limited
Average transcript abudance (molecules /cell)
0
5000
10000
15000
20000
0.12
0.20
0.30
Growth rate (1/h)

B


mRNA–C limited
Average transcript abudance (molecules /cell)
0
5000
10000
15000
20000
0.12
0.20
0.30
Growth rate (1/h)
mRNA–N2 limited
Average transcript abudance (molecules /cell)
0
5000
10000
15000
20000
0.12
0.20
0.30
Growth rate (1/h)

C



D

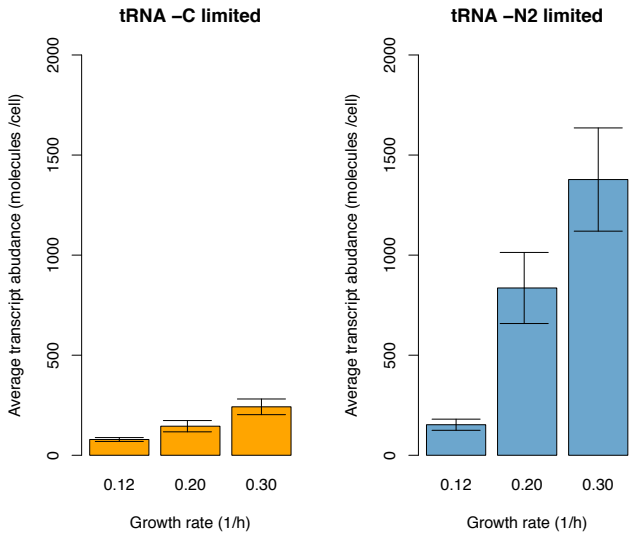
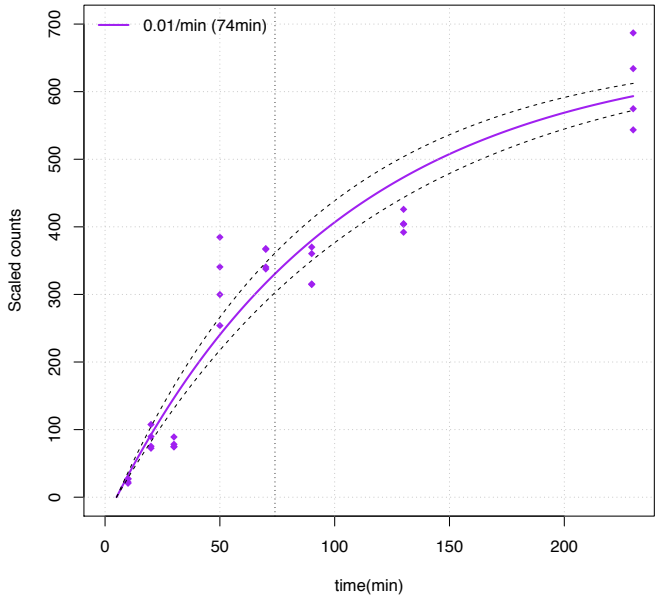


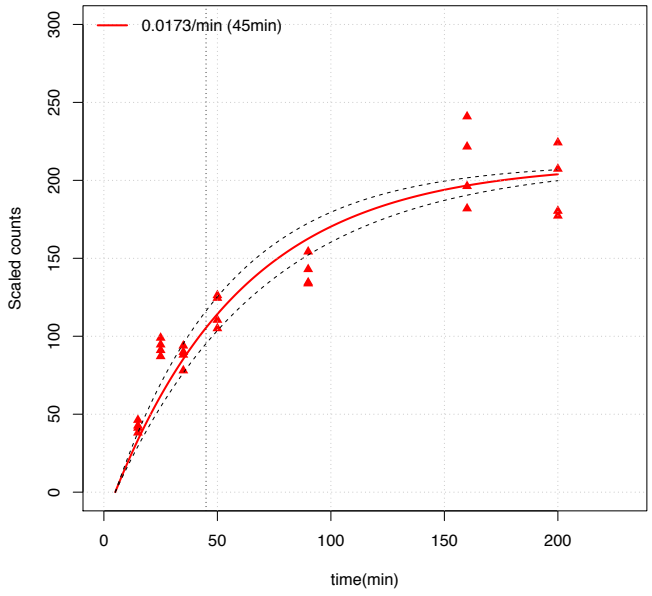
Figure 4.5: Change in transcriptome size in response to growth rate
(A) All annotated transcripts (B) mRNA, (C) non-coding RNA, and (D) tRNA.

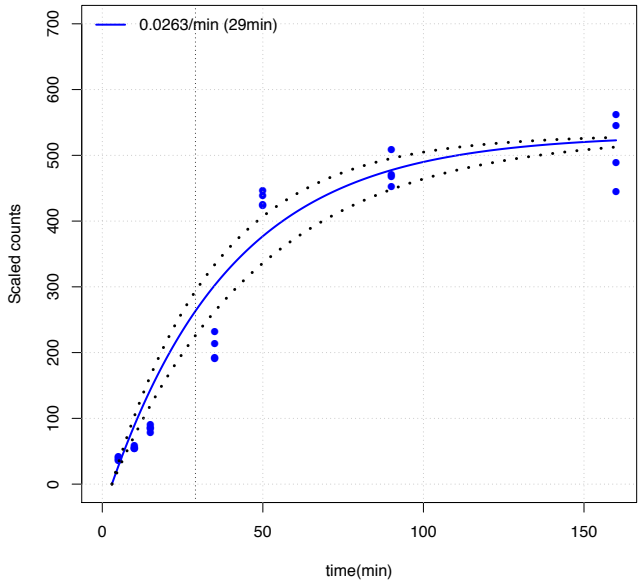
4499, 4221, and 5052 genes respectively. As shown in **Figure 4.6B**, an increase in growth rate is associated with an increase in degradation rates. However, on the global scale, there is minimal difference between the medium growth rate and fast growth rate (**Figure 4.6B**, compare "medium" and "fast"). Nevertheless these data are a clear indication that rates of mRNA degradation are regulated, at least partially, in a growth rate dependent manner.

We next investigated whether a subset of these genes can be explained in a simple relationship with growth rate. Using linear regression, we modeled degradation rates as a function of growth rate. Many genes fit well to a linear model, with 432 genes having r-squared of 0.95 or higher (**Table 4.S2**). The slopes of the regression lines differ between genes (**Figure 4.6C**), suggesting that each gene is scaled in a transcript specific manner. Using Gene Ontology enrichment analysis, we find that these 432 genes are overrepresented in genes in a number of categories, with the strongest overrepresentation in "ribosome biogenesis" and "cellular component biogenesis". In fact, we see a strong relationship between the degradation rate of the RP regulon and the cellular growth rate (**Figure 4.6D**). These observations are consistent with previous findings that ribosome levels (Ju and Warner 1994) and cellular components (Scott et al. 2010) are finely tuned with growth rates. Together, these data indicate that degradation rates are modulated on a wide scale to enable rapid responses to changes in cellular growth rate.

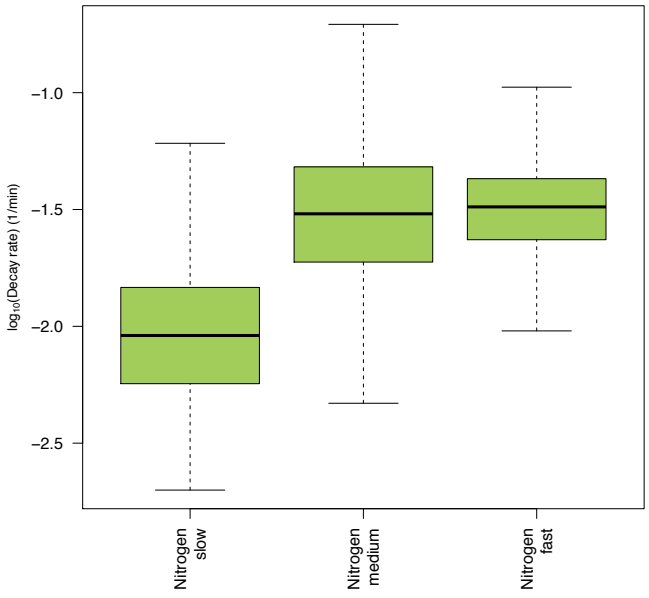
A


0.01/min (74min)
Scaled counts
time(min)
0
100
200
300
400
500
600
700
0
50
100
150
200

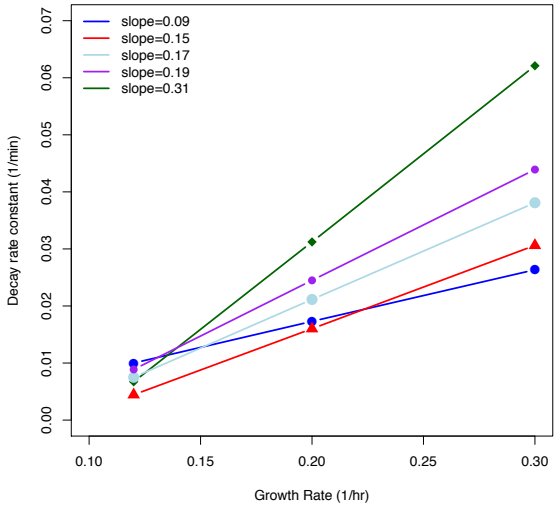

0.0173/min (45min)
Scaled counts
time(min)
0
50
100
150
200
250
300
0
50
100
150
200


0.0263/min (29min)
Scaled counts
time(min)
0
100
200
300
400
500
600
700
0
50
100
150

B


log10(Decay rate) (1/min)
−1.0
−1.5
−2.0
−2.5
Nitrogen slow
Nitrogen medium
Nitrogen fast

C



D

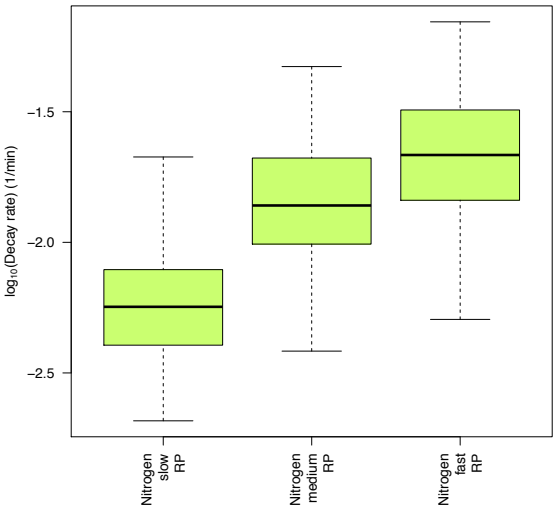


Figure 4.6: Regulation of Degradation rates as a function of growth rate
We show the non-linear regression for a single gene A) in the slow (purple), medium (red), and fast (blue) growth rates. We show the in-vivo degradation rate constant, as well as the time at which one half the equilibrium level of labeling is attained. B) Global degradation rates are determined in a growth

rate dependent manner. C) The extent of variation in degradation rate is determined in a transcript specific manner. D) The RP regulon.

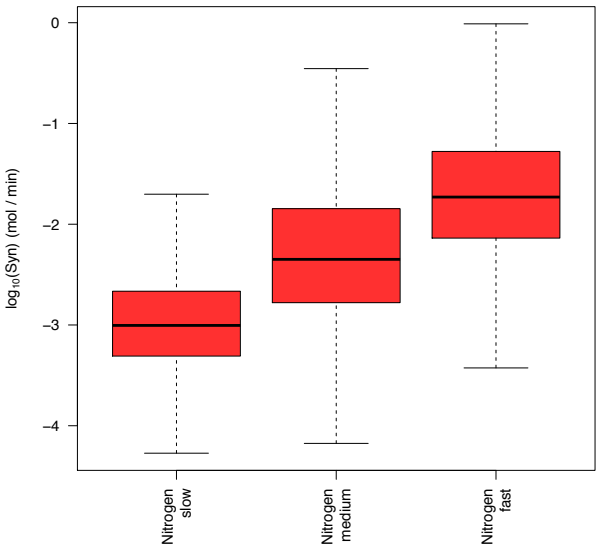
4.4.6: Synthesis rates are regulated in a growth rate dependent manner

In addition to decay measurements, we also determined rates of synthesis genome-wide. Under steady state conditions, rates of synthesis and degradation are equal, and synthesis rates can therefore be inferred with decay measurements and abundance levels. We estimated synthesis rates using our degradation rates generated through RATE-seq, and abundance levels generated from our measurements of mRNA copy number.

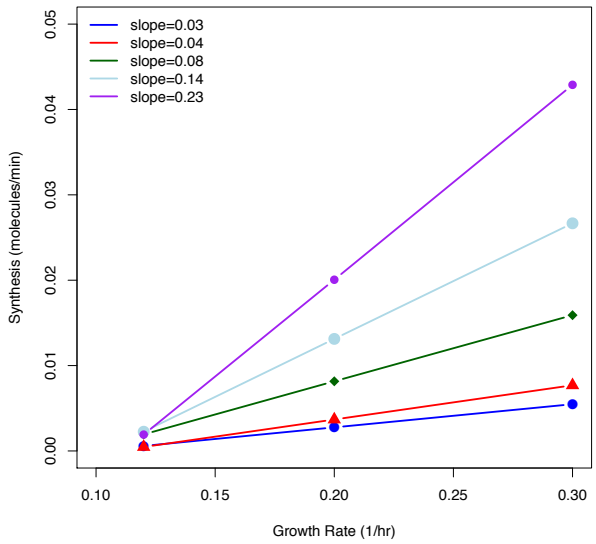
We find that similar to degradation rates, synthesis rates are determined in a growth rate dependent manner (**Figure 4.7A**). However, the relationship is much clearer with each growth rate having distinguishable distribution genome-wide. To determine the extent of this relationship for individual genes, we performed linear regression of synthesis rate as a function of growth rate. We find that the synthesis rates of more than six hundred genes can be explained in a simple linear model (**Table 4.S3**). As with degradation rates, the extent of increase in synthesis rate with increased growth rate is determined in a gene specific manner (**Figure 4.7B**). GO term analysis reveals over representation relating to factors associated with growth rate including "cell division", "cell cycle", and "biological organization". These data indicate that synthesis rates are determined as a function of growth rate.

We next compared the extent of variation of synthesis and degradation as a function of growth rate. By comparing the extent of difference between rates at the fastest and slowest growth, we can establish whether synthesis or

A


0
−1
−2
−3
−4
log10(Syn) (mol / min)
Nitrogen slow
Nitrogen medium
Nitrogen fast

B


slope=0.03
slope=0.04
slope=0.08
slope=0.14
slope=0.23
Synthesis (molecules/min)
0.00
0.01
0.02
0.03
0.04
0.05
0.10
0.15
0.20
0.25
0.30
Growth Rate (1/hr)

C

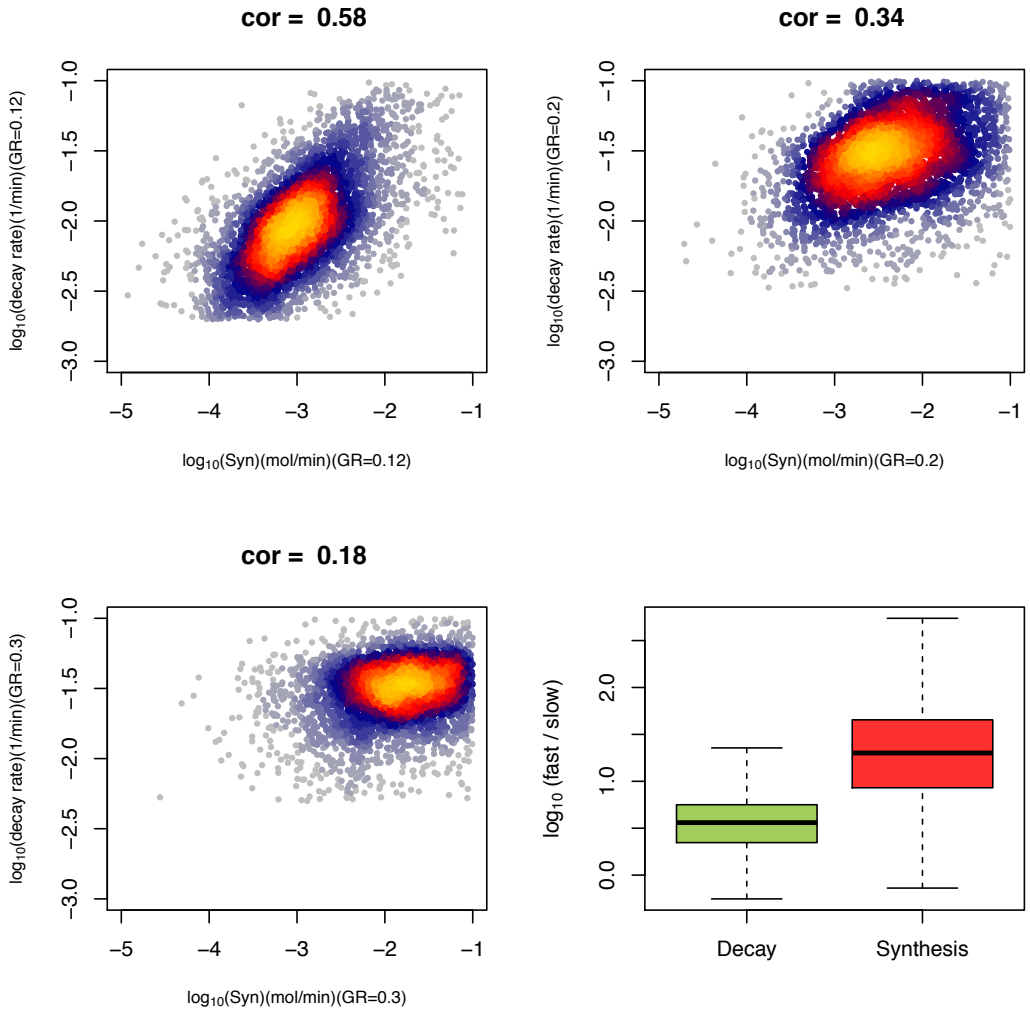


Figure 4.7: Synthesis rates are the primary determinants of changes in mRNA levels

We generated synthesis rates by combining our abundance dataset with RATE-seq estimates of degradation rate constants. A) Global synthesis rates are determined in a growth rate dependent manner. B) The extent of variation in synthesis rates is determined in a transcript specific manner. C) Synthesis rates vary to a greater extent as a function of growth rate than do degradation rates, and are more highly correlated with degradation rates at lower cellular growth rates.

degradation contributes more to modulating mRNA levels. As shown in **Figure 4.7C**, the difference between synthesis rates at the higher and lower growth rates is more pronounced than the difference between degradation rates. In addition, we find that the relationship between synthesis and degradation rates is strong at the low growth rate, and decreases with increasing growth rate (**Figure 4.7C**). Similarly, the relationship between synthesis rate and abundance levels increases with increasing growth rate (**Figure 4.S6**). These data indicate that globally, the rate at which a gene is transcribed is the primary determinant of mRNA levels whereas increases in degradation rates have a secondary effect.

4.5: Discussion

In this work, we address the effect of cellular growth rate on changes in mRNA levels. To do so, we developed a method for normalizing RNA-seq data that does not rely on assumptions of total RNA levels remaining constant across time and samples. Using ERCC spike-ins, we determined the total number of transcripts per cell in *S. cerevisiae*. We performed these analyses for technical replicates to establish effects of technical variation, as well as biological replicates to account for any biological variability. We found that the number of transcripts per cell increases as a function of growth rate. This is true for all types of RNA investigated including rRNA, tRNA, ncRNA and mRNA. These analyses show that all types of RNA are modulated as a function of growth rate.

Global normalization strategies often rely on the assumption that the proportion of different fractions of RNA does not change between samples. Our data show this to be an inaccurate assumption when growth rate varies. Because environmental perturbations or changes in extracellular condition often lead to changes in growth rate, global studies of gene expression should not assume constant RNA levels across samples. Such assumptions may lead to inaccurate interpretation of results. Our observation that a large number of genes are regulated in a growth rate dependent manner is consistent with previous observations (Brauer et al. 2008). However, we show that our response is always in the positive direction, that an increase in growth rate leads to a concomitant increase in mRNA abundance. This observation was masked in those previous observations, most likely as a result of faulty assumptions about constant RNA levels. Use of externally provided spike-ins provides the most accurate means for transcript number quantification on a per cell basis.

To determine the contributing factors to changes in mRNA levels, we utilized RATE-seq to establish rates of synthesis and degradation. We found that both mRNA synthesis and decay are modulated in a growth rate dependent manner. However, the primary determinant of mRNA levels seems to be mRNA synthesis, as the changes in mRNA levels are best explained as a function of synthesis rates. We performed linear regression of both synthesis rates and decay rates versus cellular growth rate, and found a substantial

fraction of genes for which changes in growth rate explains changes in synthesis and degradation rates respectively. Interestingly, only 3 out of 614 genes that fit to a linear model of growth rate versus synthesis rate have a negative slope. Likewise, only 7 out of 432 genes that fit a linear model of growth rate versus degradation rate have negative slope. This suggests that for the most part, increases in cellular growth rate lead to increases in degradation rates and even larger increases in synthesis rates. This results in higher overall mRNA levels. It will be interesting to see how these observations compare to degradation rates measured in carbon limiting conditions.

By utilizing chemostat cultures, ERCC spike-ins, and RATE-seq, we have provided the first complete picture of transcriptome dynamics as a function of cellular growth rate. The tools described here enable the absolute quantification of RNA levels per cell, and determination of the underlying kinetic parameters establishing those levels. We showed that there is variation in rates of mRNA synthesis and degradation as a function of growth rate, and the extent of this variation is determined in a transcript specific manner. The fact that growth rate affects mRNA kinetics is of extreme significance. This means that any study perturbing cellular growth rate will inadvertently change the synthesis and degradation rates of all transcripts. Why some transcripts increase their rates to a greater degree than other transcripts is yet to be determined. Our recent studies of yeast cells growing

in rich conditions in batch culture suggest that factors involved in translation play a role in variation of genome-wide mRNA degradation rates (Neymotin et al, in preparation). Whether this is true or not for the variation observed here is to be investigated in future studies.

4.6: Supplemental Figures

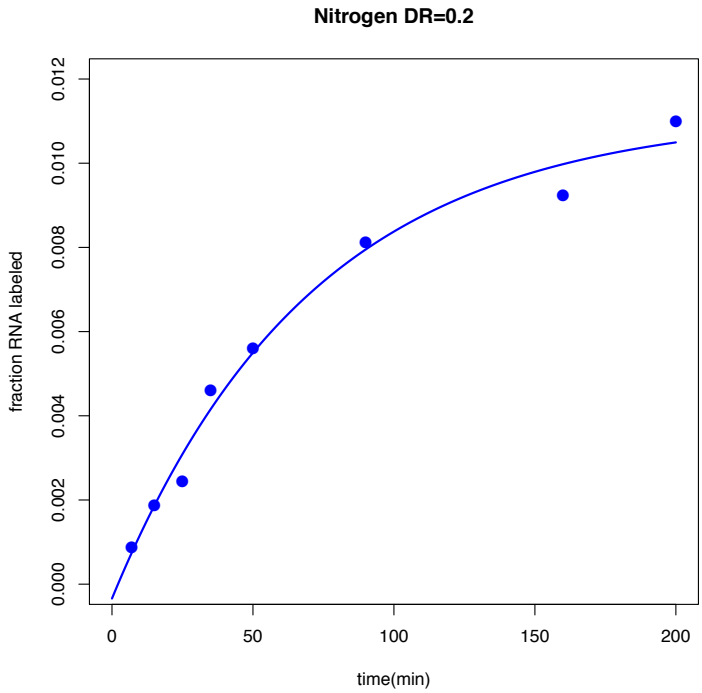


Figure 4.S1: The kinetics of labeling
We determined the appropriate time points for RATE-seq analysis based on preliminary labeling studies. Here we show an example of the amount of labeled RNA recovered for the samples grown at medium growth rate.

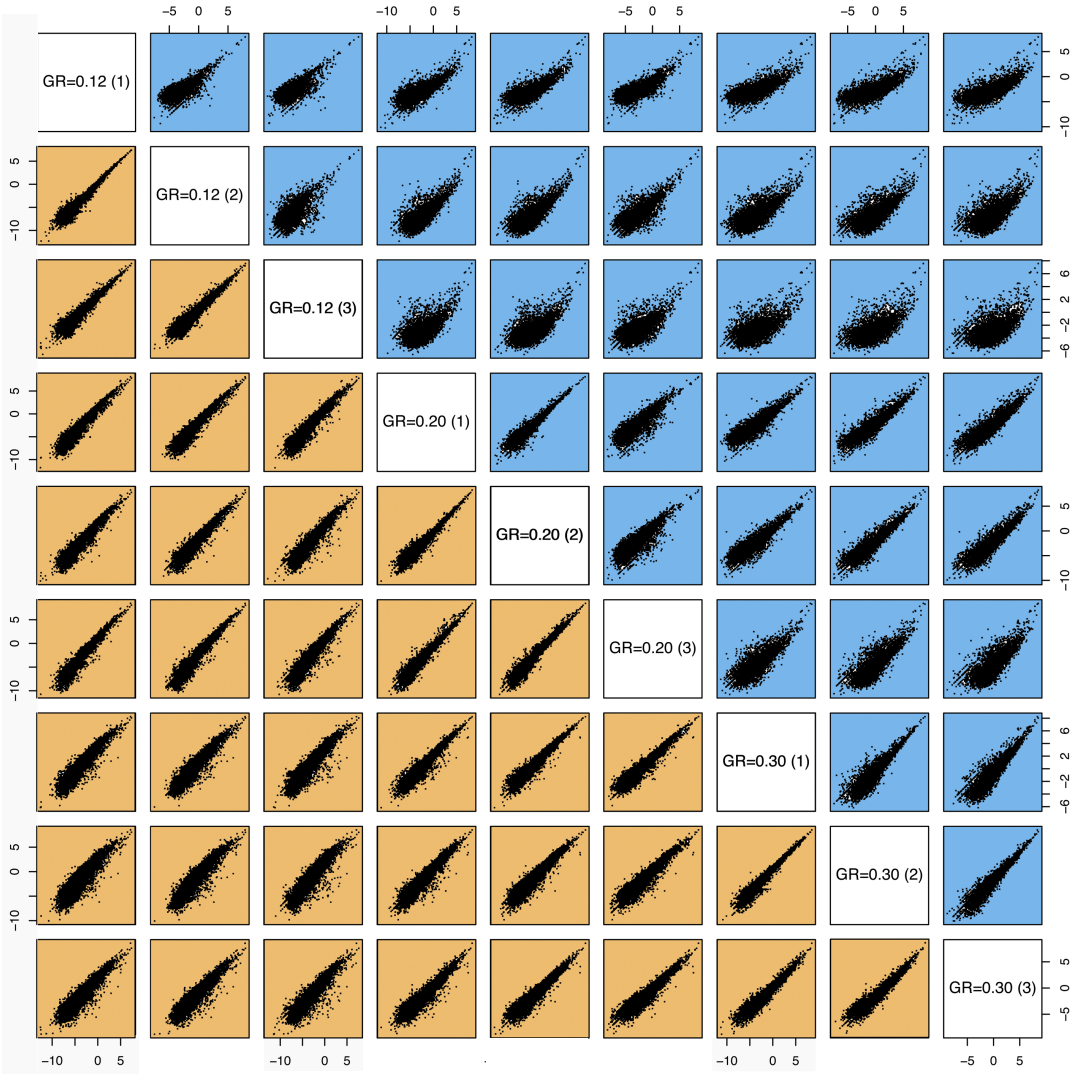
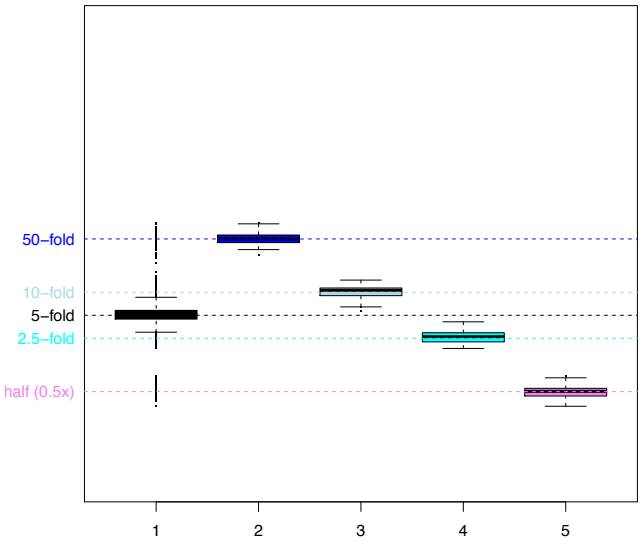
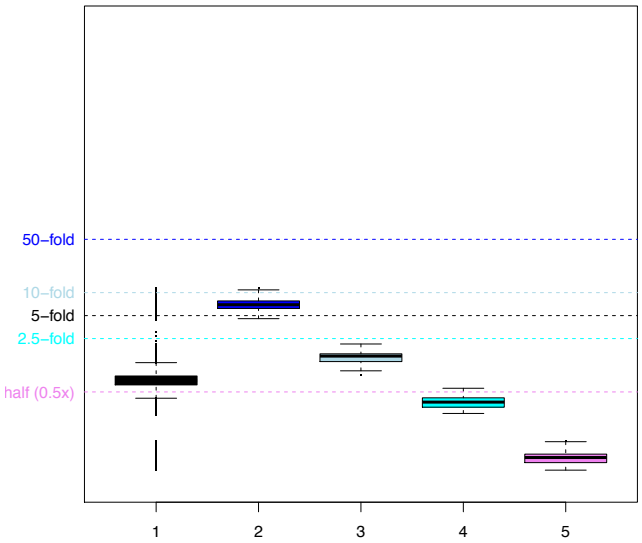


Figure 4.S2: Reproducibility for biological replicates
Scatterplot matrix of all the pair-wise correlations between the nitrogen limited (blue) and carbon limited (orange) samples. The comparisons made are indicated on the diagonal.

A


50−fold
10−fold
5−fold
2.5−fold
half (0.5x)
1
2
3
4
5

B


50−fold
10−fold
5−fold
2.5−fold
half (0.5x)
1
2
3
4
5

C

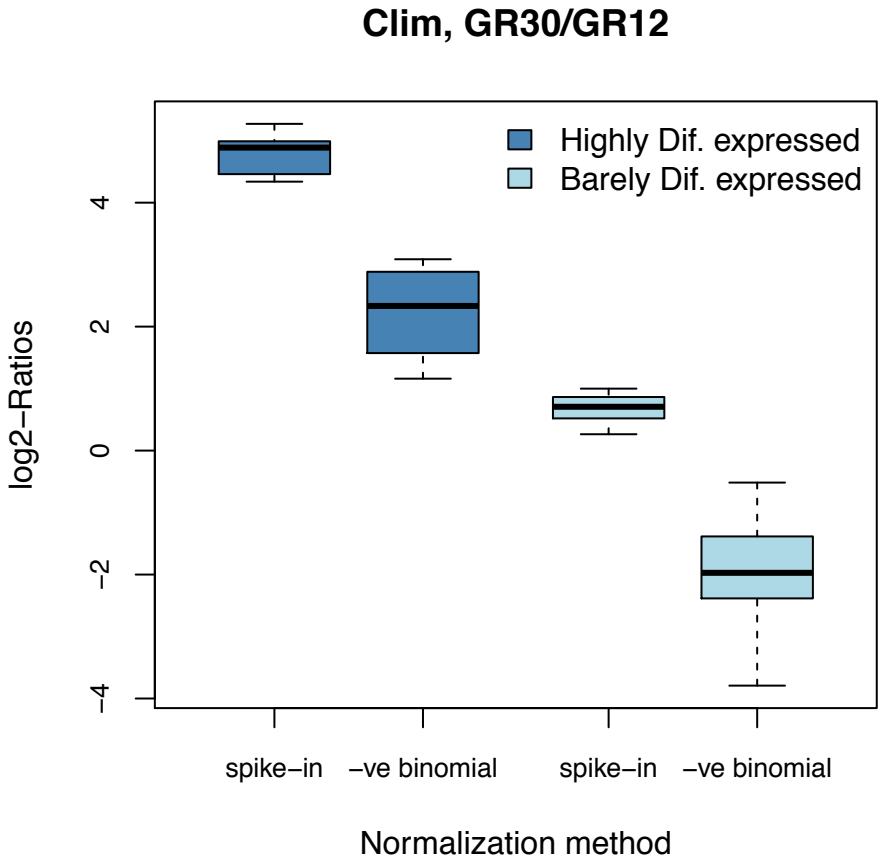
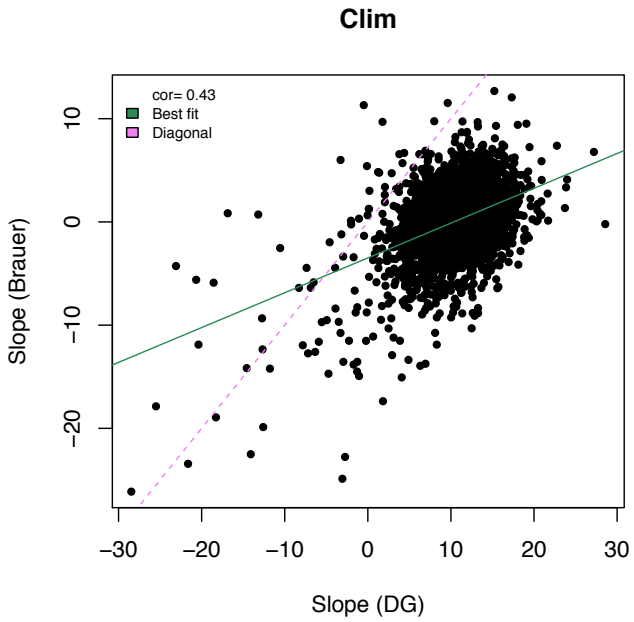


Figure 4.S3: Differences caused by normalization method
Comparison of two methods in detecting differentially expressed genes in conditions of total transcriptome amplifications. (A) Simulation showing the "real" differential expression. (B)Simulation with the data from (A), but normalized using Edge R. The gene-sets under comparison are color-coded between (A) and (B). (C) Experimental data normalized with the standard methods or RNA spike-in calibration. The "-ve binomial" refers to negative binomial.

A



B

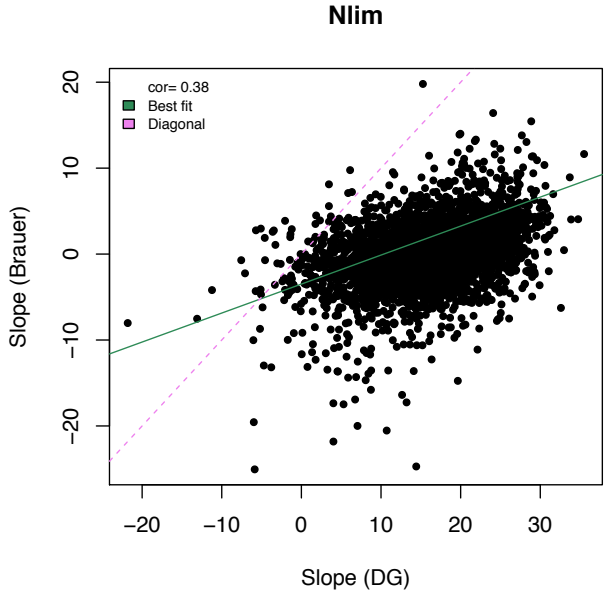


Figure 4.S4: Comparison of slopes between studies
Scatter plots of slopes from the current study (DG) and Brauer et al 2008 for (A) carbon limited and (B) nitrogen limited chemostat samples.

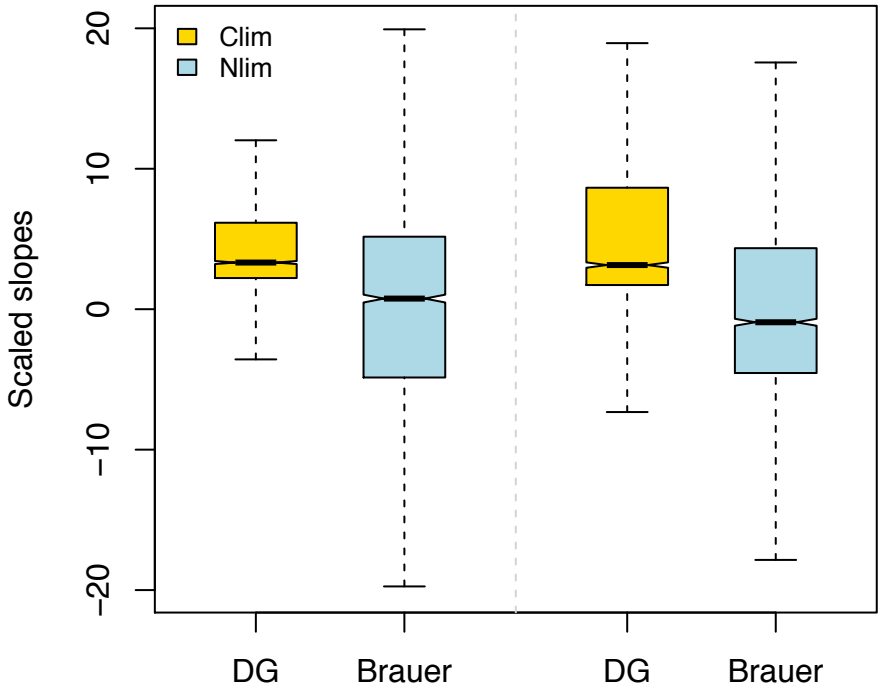


Figure 4.S5: Comparison of scaled data between studies
Box plots of slope distributions after both datasets were scaled for comparison purposes. "DG" indicates the current study.

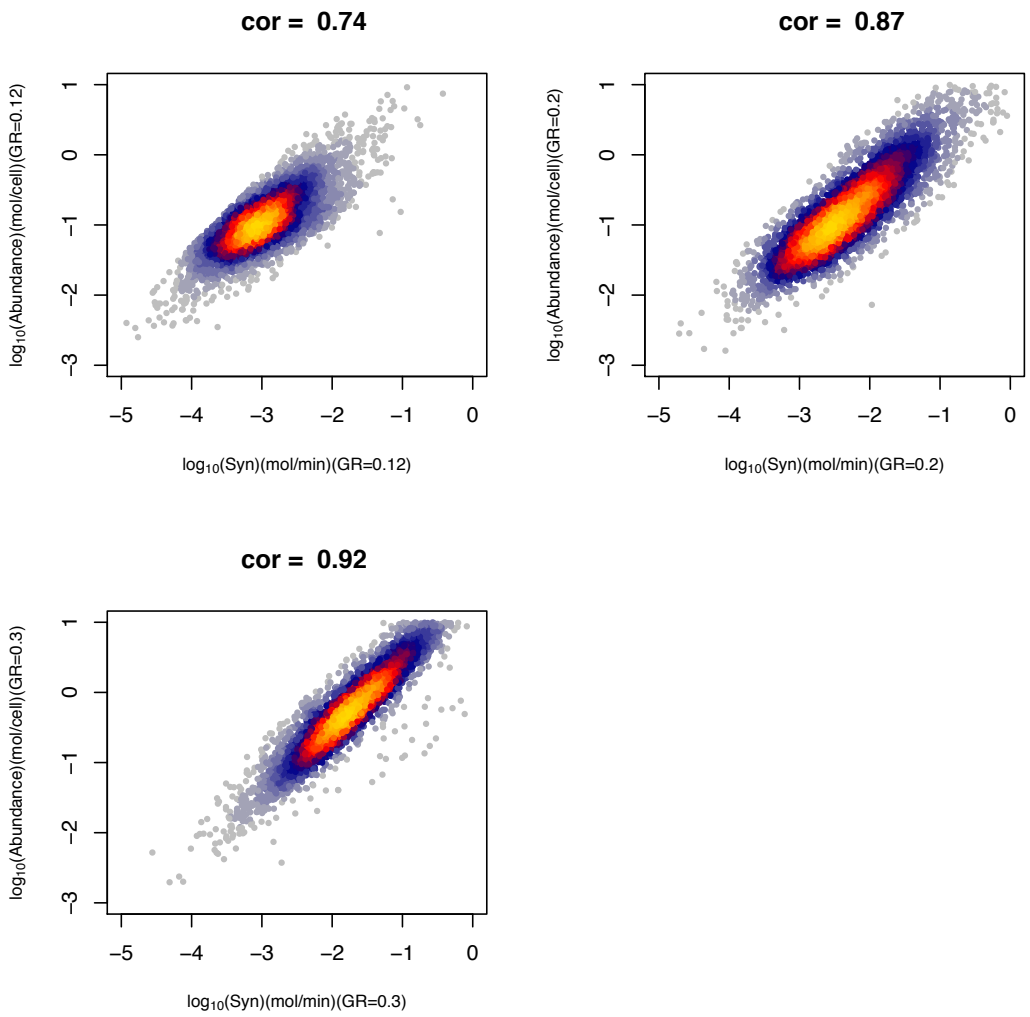


Figure 4.S6: RNA synthesis correlates well with abundance levels
We show the Spearman correlation between synthesis rate and abundance for each of the three different growth rates of slow (GR=0.12), medium (GR=0.2), and fast (GR=0.3).

Chapter 5: Conclusion

5.1: Summary of Results

For a cell to respond to the external world, it must regulate its gene expression program. Although control of transcription is often the most studied part of this regulation, it is increasingly clear that post-transcriptional regulation is a major factor in rapid transcriptome modulation. One aspect of this regulation is the rate at which different transcripts are degraded. The goal of this thesis was to characterize variation in rates of mRNA degradation transcriptome wide, identify determinants of this variation, and study how environmental conditions affect mRNA degradation rates.

5.1.1: *Measurement of mRNA decay and synthesis rates using RATE-seq*

To monitor global changes in mRNA degradation rates, we needed a method for measuring mRNA degradation rates genome wide that does not impair physiological processes (Chapter 2). RNA Approach to Equilibrium sequencing (RATE-seq) combines the uracil analogue 4-thiouracil, approach to equilibrium kinetics, and RNA sequencing, to determine decay rates. We quantified degradation rates for the entire transcriptome including coding and non-coding genes and found that transcripts are coordinately regulated based on similar biological pathways and complexes. Simultaneously we were able to quantify mRNA synthesis rates, which allowed for a complete picture of transcriptome dynamics under conditions of steady state growth in rich

media. RATE-seq is a general method with broad applications that yields accurate mRNA decay and synthesis rates.

5.1.2: *Translational efficiency and GC content effect mRNA decay rates*

To investigate the determinants of variation in mRNA degradation rates, we integrated data for RATE-seq with the vast array of available data for the features and products of all genes in the yeast *S. cerevisiae* (Chapter 3). Through multiple regression analysis, we found that variation in mRNA decay is best explained when combining multiple transcript features into a single model. Our analysis suggested that features relating to translation play a central role in variation in mRNA decay rates. To validate these observations, we designed experiments in which we modified a single transcript feature, attempting to keep all other parameters the same. We found that ribosome density affects observed degradation rates in a transcript specific manner, and we suggested that the function of the encoded protein contributes to this observation. We also determined the effect of changes in GC content in the wobble position of codons on mRNA degradation rates. Both steady state levels and degradation kinetics could be explained by changes in GC content up to a certain threshold, past which some other, undetermined factor contributed to decay. Further investigation revealed that translational efficiency, as measured by protein produced per mRNA and codon adaptation index may underlie this variation.

5.1.3: *Cellular growth rate affects mRNA synthesis and degradation rates*

We next determined the effect of cellular growth rate on rates of mRNA degradation (Chapter 4). Using the chemostat in combination with RATE-seq, we were able to determine that mRNA degradation is regulated in a growth rate dependent manner. We found that variation in mRNA degradation rates in a large subset of genes could be explained in a simple linear relationship with growth rate. Gene Onotology enrichment analysis showed an over representation for genes involved in "ribosome biogenesis" and "cellular component biogenesis". These observations are consistent with previous findings that ribosome levels (Ju and Warner 1994) and cellular components (Scott et al. 2010) are finely tuned with growth rate. By simultaneously measuring absolute abundance of mRNA transcripts per cell, we were able to infer synthesis rates. We found that at slower rates of growth, synthesis and degradation are highly correlated, but as cells increase their growth rate, the relationship between synthesis and degradation does not hold. The difference between synthesis at the fast and slow growth rates is much greater than the difference between degradation at the fast and slow growth rates. In addition, synthesis rates are highly correlated with total mRNA levels, whereas degradation is not. From these data we conclude that both synthesis and degradation are regulated in a growth rate dependent manner, but that synthesis is the primary driving force in observed differences in mRNA levels as a function of growth rate.

5.2: Perspective and Future Investigation

Our studies indicate that mRNA degradation rates are determined by multiple factors. Using a combination of bioinformatic analyses of different datasets, as well as carefully designed studies of individual transcripts, we showed that aspects of translation affect mRNA degradation. In addition, by using RATE-seq in combination with chemostats, we concluded that growth rate is also a major contributor to mRNA synthesis and degradation rates. As suggested by our multiple regression analysis, the greatest amount of explained variation in mRNA decay rates occurs when multiple predictors are integrated into a single model.

We have initiated collaborative efforts to quantify the yeast proteome as a function of cellular growth rate in the same conditions of our mRNA kinetic experiments. Previous observations indicate that protein levels fluctuate as a function of growth rate as well (Castrillo et al. 2007). In so doing, we will expand our understanding of gene regulation at the post-transcriptional level. By incorporating proteomic data in combination with mRNA abundance levels and kinetic parameters, we will enhance our understanding of the role that cell growth rate plays in post-transcriptional gene regulation.

One limitation of our studies, as well as all published studies, is the focus on model organisms and laboratory strains. It is clear that different strains of yeast from a variety of ecologies have different growth phenotypes (Ziv et al. 2013b). Based on our growth rate investigations, the suggestion

then is that the mRNA kinetics should differ between strains with different growth rates. In our lab, we have initiated studies investigating how mRNA decay rates vary based on ecological origin of different yeast strains. Using heterozygous diploids, we are investigating allele specific mRNA degradation rates. These results will inevitably be important in understanding the role of ecological adaptation in post-trancsriptional gene regulation as well as the *cis* and *trans* factors affecting variation.

RATE-seq assumes steady state growth and models mRNA degradation rates on that assumption. Although extremely powerful and informative, RATE-seq cannot be directly applied to non-steady state conditions. To address this concern, we have initiated comparable experiments to RATE-seq, but instead of tracking the increase in labeled transcript following addition of thiouracil, we chase out the label using a pulse-chase experimental design (Cleary et al. 2005; Munchel et al. 2011). Using this design, we need not assume that cells are in steady state growth, and can model degradation as the loss of labeled transcripts with time. A major caveat of pulse-chase labeling is the possibility of nucleotide recycling, as mentioned in Chapter 2. Therefore, determining the extent of recycling is necessary to accurately model the observed degradation rates. We have initiated such studies in the dynamic response of cells to preferred nitrogen sources.

Finally, our analyses showed that functionally related genes degrade with similar kinetics. It will be interesting to determine how subcellular

localization of mRNA transcripts within the cell affects mRNA decay kinetics. It is known that transcripts contain "zipcodes" in their sequences (St Johnston 2005), specifying their subcellular localization. How the spatial location of an mRNA transcript affects its degradation kinetics is of great interest. It is possible to isolate different organelles with sucrose gradient fractionation (Rieder and Emr 2001). Sequencing of the associated mRNAs in combination with RATE-seq has the potential to provide new insight into the role of mRNA localization on post-transcriptional gene expression.

REFERENCES:

Albig AR, Decker CJ. 2001. The target of rapamycin signaling pathway regulates mRNA turnover in the yeast Saccharomyces cerevisiae. *Mol Biol Cell* **12**: 3428–3438.

Anderson JS, Parker RP. 1998. The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3“ to 5” exonucleases of the exosome complex. *EMBO J* **17**: 1497–1506.

Arribere JA, Doudna JA, Gilbert WV. 2011. Reconsidering movement of eukaryotic mRNAs between polysomes and P bodies. *Mol Cell* **44**: 745–758.

Baker SC, Bauer SR, Beyer RP, Brenton JD, Bromley B, Burrill J, Causton H, Conley MP, Elespuru R, Fero M, et al. 2005. The External RNA Controls Consortium: a progress report. *Nat Methods* **2**: 731–734.

Beelman CA, Parker R. 1994. Differential effects of translational inhibition in cis and in trans on the decay of the unstable yeast MFA2 mRNA. *J Biol Chem* **269**: 9687–9692.

Beelman CA, Stevens A, Caponigro G, LaGrandeur TE, Hatfield L, Fortner DM, Parker R. 1996. An essential component of the decapping enzyme required for normal rates of mRNA turnover. *Nature* **382**: 642–646.

Belle A, Tanay A, Bitincka L, Shamir R, O'Shea EK. 2006. Quantification of protein half-lives in the budding yeast proteome. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 13004–13009.

Brauer MJ, Huttenhower C, Airoldi EM, Rosenstein R, Matese JC, Gresham D, Boer VM, Troyanskaya OG, Botstein D. 2008. Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast. *Mol Biol Cell* **19**: 352–367.

Bregman A, Avraham-Kelbert M, Barkai O, Duek L, Guterman A, Choder M. 2011. Promoter elements regulate cytoplasmic mRNA decay. *Cell* **147**: 1473–1483.

Brengues M, Teixeira D, Parker R. 2005. Movement of eukaryotic mRNAs between polysomes and cytoplasmic processing bodies. *Science* **310**: 486–489.

Caponigro G, Muhlrad D, Parker R. 1993. A small segment of the MAT alpha 1 transcript promotes mRNA decay in Saccharomyces cerevisiae: a stimulatory role for rare codons. *Mol Cell Biol* **13**: 5141–5148.

Carlini DB. 2005. Context-dependent codon bias and messenger RNA longevity in the yeast transcriptome. *Mol Biol Evol* **22**: 1403–1411.

Castrillo JI, Zeef LA, Hoyle DC, Zhang N, Hayes A, Gardner DCJ, Cornell MJ, Petty J, Hakes L, Wardleworth L, et al. 2007. Growth control of the eukaryote cell: a systems biology study in yeast. *J Biol* **6**: 4.

Chen CY, Gherzi R, Ong SE, Chan EL, Raijmakers R, Pruijn GJ, Stoecklin G, Moroni C, Mann M, Karin M. 2001. AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. *Cell* **107**: 451–464.

Cleary MD, Meiering CD, Jan E, Guymon R, Boothroyd JC. 2005. Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. *Nat Biotechnol* **23**: 232–237.

Coller J, Parker R. 2004. Eukaryotic mRNA decapping. *Annu Rev Biochem* **73**: 861–890.

Coller J, Parker R. 2005. General translational repression by activators of mRNA decapping. *Cell* **122**: 875–886.

Coller JM, Gray NK, Wickens MP. 1998. mRNA stabilization by poly(A) binding protein is independent of poly(A) and requires translation. *Genes Dev* **12**: 3226–3235.

Decker CJ, Parker R. 1993. A turnover pathway for both stable and unstable mRNAs in yeast: evidence for a requirement for deadenylation. *Genes Dev* **7**: 1632–1643.

DeRisi JL, Iyer VR, Brown PO. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**: 680–686.

Dölken L, Ruzsics Z, Rädle B, Friedel CC, Zimmer R, Mages J, Hoffmann R, Dickinson P, Forster T, Ghazal P, et al. 2008. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA* **14**: 1959–1972.

Duan J, Shi J, Ge X, Dölken L, Moy W, He D, Shi S, Sanders AR, Ross J, Gejman PV. 2013. Genome-wide survey of interindividual differences of RNA stability in human lymphoblastoid cell lines. *Sci Rep* **3**: 1318.

Eck S, Stephan W. 2008. Determining the relationship of gene expression and global mRNA stability in Drosophila melanogaster and Escherichia coli using linear models. *Gene* **424**: 102–107.

Edri S, Tuller T. 2014. Quantifying the effect of ribosomal density on mRNA stability. *PLoS ONE* **9**: e102308.

Elkon R, Zlotorynski E, Zeller KI, Agami R. 2010. Major role for mRNA stability in shaping the kinetics of gene induction. *BMC Genomics* **11**: 259.

External RNA Controls Consortium. 2005. Proposed methods for testing and selecting the ERCC external RNA controls. *BMC Genomics* **6**: 150.

Furuichi Y, LaFiandra A, Shatkin AJ. 1977. 5'-Terminal structure and mRNA stability. *Nature* **266**: 235–239.

García-Martínez J, Aranda A, Pérez-Ortín JE. 2004. Genomic run-on evaluates transcription rates for all yeast genes and identifies gene regulatory mechanisms. *Mol Cell* **15**: 303–313.

Garí E, Piedrafita L, Aldea M, Herrero E. 1997. A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in Saccharomyces cerevisiae. *Yeast* **13**: 837–848.

Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO. 2000. Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell* **11**: 4241–4257.

Ghaemmaghami S, Huh W-K, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS. 2003. Global analysis of protein expression in yeast. *Nature* **425**: 737–741.

Graves RA, Pandey NB, Chodchoy N, Marzluff WF. 1987. Translation is required for regulation of histone mRNA degradation. *Cell* **48**: 615–626.

Greenberg JR. 1972. High stability of messenger RNA in growing cultured cells. *Nature* **240**: 102–104.

Gresham D, Boer V, Caudy A, Ziv N, Brandt NJ, Storey JD, Botstein D. 2010. System-Level Analysis of Genes and Functions Affecting Survival During Nutrient Starvation in Saccharomyces cerevisiae. *Genetics*.

Grigull J, Mnaimneh S, Pootoolal J, Robinson MD, Hughes TR. 2004. Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays reveals posttranscriptional control of ribosome biogenesis

factors. *Mol Cell Biol* **24**: 5534–5547.

Gupta I, Clauder-Münster S, Klaus B, Järvelin AI, Aiyar RS, Benes V, Wilkening S, Huber W, Pelechano V, Steinmetz LM. 2014. Alternative polyadenylation diversifies post-transcriptional regulation by selective RNA-protein interactions. *Mol Syst Biol* **10**: 719.

Hager GL, McNally JG, Misteli T. 2009. Transcription Dynamics. *Mol Cell* **35**: 741–753.

Hargrove JL, Hulsey MG, Beale EG. 1991. The kinetics of mammalian gene expression. *Bioessays* **13**: 667–674.

Harpold MM, Wilson MC, Darnell JE. 1981. Chinese hamster polyadenylated messenger ribonucleic acid: relationship to non-polyadenylated sequences and relative conservation during messenger ribonucleic acid processing. *Mol Cell Biol* **1**: 188–198.

He F, Li X, Spatrick P, Casillo R, Dong S, Jacobson A. 2003. Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5“ to 3” mRNA decay pathways in yeast. *Mol Cell* **12**: 1439–1452.

Herrick D, Parker R, Jacobson A. 1990. Identification and comparison of stable and unstable mRNAs in Saccharomyces cerevisiae. *Mol Cell Biol* **10**: 2269–2284.

Hogan DJ, Riordan DP, Gerber AP, Herschlag D, Brown PO. 2008. Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *PLoS Biol* **6**: e255.

Hong J, Gresham D. 2014. Molecular specificity, convergence and constraint shape adaptive evolution in nutrient-poor environments. *PLoS Genet* **10**: e1004041.

Hoskisson PA, Hobbs G. 2005. Continuous culture--making a comeback? *Microbiology (Reading, Engl)* **151**: 3153–3159.

Houalla R, Devaux F, Fatica A, Kufel J, Barrass D, Torchet C, Tollervey D. 2006. Microarray detection of novel nuclear RNA substrates for the exosome. *Yeast* **23**: 439–454.

Hu W, Sweet TJ, Chamnongpol S, Baker KE, Coller J. 2009. Co-translational mRNA decay in Saccharomyces cerevisiae. *Nature* **461**: 225–229.

Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. 2009. Genome-wide

analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**: 218–223.

Johnston GC, Singer RA, Sharrow SO, SLATER ML. 1980. Cell Division in the Yeast Saccharomyces cerevisiae Growing at Different Rates. *Microbiology* **118**: 479–484.

Ju Q, Warner JR. 1994. Ribosome synthesis during the growth cycle of Saccharomyces cerevisiae. *Yeast* **10**: 151–157.

Keren L, Zackay O, Lotan-Pompan M, Barenholz U, Dekel E, Sasson V, Aidelberg G, Bren A, Zeevi D, Weinberger A, et al. 2013. Promoters maintain their relative activity levels under different growth conditions. *Mol Syst Biol* **9**: 701.

Kern L, de Montigny J, Jund R, Lacroute F. 1990. The FUR1 gene of Saccharomyces cerevisiae: cloning, structure and expression of wild-type and mutant alleles. *Gene* **88**: 149–157.

Kim CH, Warner JR. 1983. Messenger RNA for ribosomal proteins in yeast. *J Mol Biol* **165**: 79–89.

Kudla G, Lipinski L, Caffin F, Helwak A, Zylicz M. 2006. High guanine and cytosine content increases mRNA levels in mammalian cells. *PLoS Biol* **4**: e180.

Kudla G, Murray AW, Tollervey D, Plotkin JB. 2009. Coding-sequence determinants of gene expression in Escherichia coli. *Science* **324**: 255–258.

Langmead B, Trapnell C, Pop M, Salzberg S. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*.

Larimer FW, Hsu CL, Maupin MK, Stevens A. 1992. Characterization of the XRN1 gene encoding a 5“-->3” exoribonuclease: sequence data and analysis of disparate protein and mRNA levels of gene-disrupted yeast cells. *Gene* **120**: 51–57.

Lipson D, Raz T, Kieu A, Jones DR, Giladi E, Thayer E, Thompson JF, Letovsky S, Milos P, Causey M. 2009. Quantification of the yeast transcriptome by single-molecule sequencing. *Nat Biotechnol* **27**: 652–658.

Lodish H, Berk A, Kaiser CA, Krieger M, Scott MP, Bretscher A, Ploegh H, Matsudaira P. 2008. *Molecular Cell Biology*. sixth. Macmillan, 2008.

Lovén J, Orlando DA, Sigova AA, Lin CY, Rahl PB, Burge CB, Levens DL, Lee TI, Young RA. 2012. Revisiting global gene expression analysis. *Cell* **151**: 476–482.

Mainguet SE, Gakière B, Majira A, Pelletier S, Bringel F, Guérard F, Caboche M, Berthomé R, Renou JP. 2009. Uracil salvage is necessary for early Arabidopsis development. *Plant J* **60**: 280–291.

Merhi A, Gérard N, Lauwers E, Prévost M, André B. 2011. Systematic mutational analysis of the intracellular regions of yeast Gap1 permease. *PLoS ONE* **6**: e18457.

Miller C, Schwalb B, Maier K, Schulz D, Dümcke S, Zacher B, Mayer A, Sydow J, Marcinowski L, Dölken L, et al. 2011. Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Mol Syst Biol* **7**: 458.

Miller MR, Robinson KJ, Cleary MD, Doe CQ. 2009. TU-tagging: cell type-specific RNA isolation from intact complex tissues. *Nat Methods* **6**: 439–441.

Mitchell P, Petfalski E, Shevchenko A, Mann M, Tollervey D. 1997. The exosome: a conserved eukaryotic RNA processing complex containing multiple 3“-->5” exoribonucleases. *Cell* **91**: 457–466.

Muhlrad D, Decker CJ, Parker R. 1995. Turnover mechanisms of the stable yeast PGK1 mRNA. *Mol Cell Biol* **15**: 2145–2156.

Muhlrad D, Parker R. 1992. Mutations affecting stability and deadenylation of the yeast MFA2 transcript. *Genes Dev* **6**: 2100–2111.

Munchel SE, Shultzaberger RK, Takizawa N, Weis K. 2011. Dynamic profiling of mRNA turnover reveals gene-specific and system-wide regulation of mRNA decay. *Mol Biol Cell*.

Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**: 1344–1349.

Narsai R, Howell KA, Millar AH, O'Toole N, Small I, Whelan J. 2007. Genome-wide analysis of mRNA decay rates and their determinants in Arabidopsis thaliana. *Plant Cell* **19**: 3418–3436.

Neymotin B, Athanasiadou R, Gresham D. 2014. Determination of in vivo RNA kinetics using RATE-seq. *RNA* **20**: 1645–1652.

Nikolov EN, Dabeva MD. 1985. Re-utilization of pyrimidine nucleotides during rat liver regeneration. *Biochem J* **228**: 27–33.

Nilsson G, Belasco JG, Cohen SN, Gabain von A. 1984. Growth-rate dependent regulation of mRNA stability in Escherichia coli. *Nature* **312**: 75–77.

Nonet M, Scafe C, Sexton J, Young R. 1987. Eucaryotic RNA polymerase conditional mutant that rapidly ceases mRNA synthesis. *Mol Cell Biol* **7**: 1602–1611.

Olivas W, Parker R. 2000. The Puf3 protein is a transcript-specific regulator of mRNA degradation in yeast. *EMBO J* **19**: 6602–6611.

Parker R. 2012. RNA degradation in Saccharomyces cerevisae. *Genetics* **191**: 671–702.

Parker R, Jacobson A. 1990. Translation and a 42-nucleotide segment within the coding region of the mRNA encoded by the MAT alpha 1 gene are involved in promoting rapid mRNA decay in yeast. *Proceedings of the National Academy of Sciences of the United States of America* **87**: 2780–2784.

Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitsch S, Lehrach H, Soldatov A. 2009. Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res* **37**: e123.

Pelechano V, Chávez S, Pérez-Ortín JE. 2010. A complete set of nascent transcription rates for yeast genes. *PLoS ONE* **5**: e15442.

Pelechano V, Pérez-Ortín JE. 2008. The transcriptional inhibitor thiolutin blocks mRNA degradation in yeast. *Yeast* **25**: 85–92.

Puckett L, Chambers S, Darnell JE. 1975. Short-lived messenger RNA in HeLa cells and its impace on the kinetics of accumulation of cytoplasmic polyadenylate. *Proceedings of the National Academy of Sciences of the United States of America* **72**: 389–393.

Rabani M, Levin JZ, Fan L, Adiconis X, Raychowdhury R, Garber M, Gnirke A, Nusbaum C, Hacohen N, Friedman N, et al. 2011. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat Biotechnol* **29**: 436–442.

Regenberg B, Grotkjaer T, Winther O, Fausbøll A, Akesson M, Bro C, Hansen LK, Brunak S, Nielsen J. 2006. Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in Saccharomyces

cerevisiae. *Genome Biol* **7**: R107.

Rieder SE, Emr SD. 2001. Isolation of subcellular fractions from the yeast Saccharomyces cerevisiae. *Curr Protoc Cell Biol* **Chapter 3**: Unit 3.8.

Schier AF. 2007. The maternal-zygotic transition: death and birth of RNAs. *Science* **316**: 406–407.

Schindler D, Davies J. 1975. Inhibitors of macromolecular synthesis in yeast. *Methods Cell Biol* **12**: 17–38.

Schmid M, Jensen TH. 2008. The exosome: a multipurpose RNA-decay machine. *Trends Biochem Sci* **33**: 501–510.

Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M. 2011. Global quantification of mammalian gene expression control. *Nature* **473**: 337–342.

Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T. 2010. Interdependence of cell growth and gene expression: origins and consequences. *Science* **330**: 1099–1102.

Selinger DW, Saxena RM, Cheung KJ, Church GM, Rosenow C. 2003. Global RNA half-life analysis in Escherichia coli reveals positional patterns of transcript degradation. *Genome Res* **13**: 216–223.

Shalem O, Dahan O, Levo M, Martinez MR, Furman I, Segal E, Pilpel Y. 2008. Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. *Mol Syst Biol* **4**: 223.

Sharp PM, Li WH. 1987. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* **15**: 1281–1295.

Shaw G, Kamen R. 1986. A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell* **46**: 659–667.

Sheth U, Parker R. 2003. Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. *Science* **300**: 805–808.

Singer RH, Penman S. 1972. Stability of HeLa cell mRNA in actinomycin. *Nature* **240**: 100–102.

St Johnston D. 2005. Moving messages: the intracellular localization of mRNAs.

Nat Rev Mol Cell Biol **6**: 363–375.

Stevens A. 1980a. An mRNA decapping enzyme from ribosomes of Saccharomyces cerevisiae. *Biochem Biophys Res Commun* **96**: 1150–1155.

Stevens A. 1980b. Purification and characterization of a Saccharomyces cerevisiae exoribonuclease which yields 5'-mononucleotides by a 5‘ leads to 3’ mode of hydrolysis. *J Biol Chem* **255**: 3080–3085.

Stimac E, Groppi VE, Coffino P. 1984. Inhibition of protein synthesis stabilizes histone mRNA. *Mol Cell Biol* **4**: 2082–2090.

Stitzel ML, Seydoux G. 2007. Regulation of the oocyte-to-zygote transition. *Science* **316**: 407–408.

Sun M, Schwalb B, Pirkl N, Maier KC, Schenk A, Failmezger H, Tresch A, Cramer P. 2013. Global analysis of eukaryotic mRNA degradation reveals xrn1-dependent buffering of transcript levels. *Mol Cell* **52**: 52–62.

Sun M, Schwalb B, Schulz D, Pirkl N, Etzold S, Larivière L, Maier KC, Seizl M, Tresch A, Cramer P. 2012. Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. *Genome Res*.

Team RC. *R: A Language and Environment for Statistical Computing*.

Teixeira D, Sheth U, Valencia-Sanchez MA, Brengues M, Parker R. 2005. Processing bodies require RNA for assembly and contain nontranslating mRNAs. *RNA* **11**: 371–382.

Thomsen S, Anders S, Janga SC, Huber W, Alonso CR. 2010. Genome-wide analysis of mRNA decay patterns during early Drosophila development. *Genome Biol* **11**: R93.

Tipper DJ. 1973. Inhibition of yeast ribonucleic acid polymerases by thiolutin. *J Bacteriol* **116**: 245–256.

Trcek T, Larson DR, Moldón A, Query CC, Singer RH. 2011. Single-molecule mRNA decay measurements reveal promoter- regulated mRNA stability in yeast. *Cell* **147**: 1484–1497.

Tucker M, Valencia-Sanchez MA, Staples RR, Chen J, Denis CL, Parker R. 2001. The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in Saccharomyces cerevisiae. *Cell* **104**: 377–386.

Tyson CB, Lord PG, Wheals AE. 1979. Dependency of size of Saccharomyces cerevisiae cells on growth rate. *J Bacteriol* **138**: 92–98.

Waldron C, Lacroute F. 1975. Effect of growth rate on the amounts of ribosomal and transfer ribonucleic acids in yeast. *J Bacteriol* **122**: 855–865.

Wang Y, Liu CL, Storey JD, Tibshirani RJ, Herschlag D, Brown PO. 2002. Precision and functional specificity in mRNA decay. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 5860–5865.

Wiegers U, Kramer G, Klapproth K, Rehpenning W, Hilz H. 1975. Determination of mRNA half-life in HeLa cultures by a poly(A)-independent direct analysis of specific radioactivity of mRNA. *Eur J Biochem* **50**: 557–562.

Wilt FH. 1973. Polyadenylation of maternal RNA of sea urchin eggs after fertilization. *Proceedings of the National Academy of Sciences of the United States of America* **70**: 2345–2349.

Wisdom R, Lee W. 1991. The protein-coding region of c-myc mRNA contains a sequence that specifies rapid mRNA turnover and induction by protein synthesis inhibitors. *Genes Dev* **5**: 232–243.

Wishart JA, Hayes A, Wardleworth L, Zhang N, Oliver SG. 2005. Doxycycline, the drug used to control the tet-regulatable promoter system, has no effect on global gene expression in Saccharomyces cerevisiae. *Yeast* **22**: 565–569.

Yun D-F, Sherman F. 1996. Degradation of CYC1 mRNA in the yeast Saccharomyces cerevisiae does not require translation. *Proceedings of the National Academy of Sciences of the United States of America* **93**: 8895–8900.

Zeiner GM, Cleary MD, Fouts AE, Meiring CD, Mocarski ES, Boothroyd JC. 2008. RNA analysis by biosynthetic tagging using 4-thiouracil and uracil phosphoribosyltransferase. *Methods Mol Biol* **419**: 135–146.

Zenklusen D, Larson DR, Singer RH. 2008. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat Struct Mol Biol* **15**: 1263–1271.

Ziv N, Brandt NJ, Gresham D. 2013a. The use of chemostats in microbial systems biology. *J Vis Exp*.

Ziv N, Siegal ML, Gresham D. 2013b. Genetic and nongenetic determinants of cell growth variation assessed by high-throughput microscopy. *Mol Biol Evol* **30**: 2568–2578.