Investigating Transcription Factor Networks That Drive Biological Clocks and Oscillators

by

Christina Marie Kelliher

University Program in Genetics and Genomics
Duke University

Date: _______________________

Approved:

___________________________
Steven B. Haase, Supervisor

___________________________
Paul M. Magwene, Chair

___________________________
Beth A. Sullivan

___________________________
Gregory A. Wray

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the University Program in Genetics and Genomics in the Graduate School of Duke University

2017
ABSTRACT

Investigating Transcription Factor Networks That Drive Biological Clocks and Oscillators

by

Christina Marie Kelliher

University Program in Genetics and Genomics
Duke University

Date: _______________________

Approved:

___________________________
Steven B. Haase, Supervisor

___________________________
Paul M. Magwene, Chair

___________________________
Beth A. Sullivan

___________________________
Gregory A. Wray

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the University Program in Genetics and Genomics in the Graduate School of Duke University

2017
Abstract

Biological systems are highly dynamic, yet our temporal resolution of such dynamical processes is often limited or difficult to test in the laboratory. The 24-hour circadian rhythm and the approximately 75-minute cell cycle of a budding yeast cell are both examples of dynamical processes that contain precisely ordered events, repeating over each cycle. Organisms utilize such biological clock processes to time a particular function. Dynamic cellular events are ordered, in part, by coordinated programs of periodic gene expression. Up to 40% of all mouse genes are periodically expressed with respect to the circadian cycle, and almost 20% of all yeast genes are periodic during the cell cycle. Furthermore, more than half of the most frequently prescribed drugs in human patients target an effector whose expression is under circadian control. Given the large proportion of genes that are periodically expressed across different biological processes, it is critically important to understand mechanisms that regulate dynamics in biology.

In this dissertation, I focus on two biological processes that are dynamic and are not yet fully understood: the eukaryotic cell cycle and malaria parasite development. Large programs of periodic genes emerge when these biological clock processes are synchronized and profiled over time. Gene regulatory networks composed of transcription factors, kinases, and other transcriptional regulators play a critical role in generating periodicity in gene expression programs, ordering clock events, and maintaining oscillations in subsequent cycles.

Many previous studies have profiled gene expression during the cell cycle in the budding yeast Saccharomyces cerevisiae. I have added to this detailed body of work by demonstrating that regulatory motifs involving negative feedback are required to
maintain normal gene expression levels. Additionally, I showed that many periodic mRNAs are also periodically abundant at the protein level during the cell cycle. Both projects provide evidence for the hypothesis that cell-cycle dynamics are driven by a network of transcription factors with complex protein dynamics and with negative feedback motifs. Using this ground truth cell-cycle network in *S. cerevisiae*, I next performed a comparative transcriptomics study on cell-cycle genes in the less studied, but more human health relevant fungal pathogen, *Cryptococcus neoformans*. This work not only begins to identify a cell-cycle network in *C. neoformans* but also has implications for future antifungal drug development, as some genes that are important for fungal virulence were found to be expressed periodically during the cell cycle.

During infection, the human malaria parasite *Plasmodium falciparum* cyclically develops and re-infects red blood cells. Many groups have shown that a very large program of gene expression occurs during this red blood cell developmental cycle. In this dissertation, I deploy the experimental and analysis tools that I used to characterize the fungal cell cycle to ask if a network of transcription factors can explain developmental gene expression dynamics and cycle period control in malaria.

Biological systems are highly dynamic to respond to environmental signals, grow, and survive. As the application of genetics and genomics has moved toward characterizing complex diseases, host-pathogen interactions, or even the cell cycle of a single yeast cell, it has become increasingly clear that networks of interacting genes are required to explain biological mechanisms. Results from this dissertation where I investigate dynamic gene regulatory networks are broadly applicable to our understanding of both basic molecular biology and of human infectious diseases.
# Contents

Abstract ........................................................................................................................................ iv

List of Tables .................................................................................................................................. x

List of Figures ................................................................................................................................... xi

Acknowledgements .......................................................................................................................... xiii

1. Introduction: from the Central Dogma of Molecular Biology to Gene Regulatory Networks ............................................................................................................................................ 1

   1.1 The Circadian Rhythm Gene Regulatory Network ............................................................... 5

   1.2 The Eukaryotic Cell-Cycle is a Model for Building Gene Regulatory Networks .............. 8

   1.3 Transcription Factor Network (TFN) Model in *S. cerevisiae* ........................................ 11

   1.4 Significance: Cancer, Heart Disease, and Infectious Diseases ....................................... 14

   1.5 Innovation: Practicing Systems Biology ............................................................................ 15

   1.6 Dissertation Outline ........................................................................................................... 16

2. The Role of Negative Feedback in the *Saccharomyces cerevisiae* Cell-Cycle Network ............................................................................................................................................... 20

   2.1 Author Contributions ......................................................................................................... 20

   2.2 Introduction ....................................................................................................................... 21

   2.3 Results ............................................................................................................................... 24

   2.4 Discussion ........................................................................................................................ 35

   2.5 Future Work ..................................................................................................................... 40

   2.6 Materials and Methods .................................................................................................... 41

      2.6.1 Yeast strains, cultures, and synchronization ............................................................. 41

      2.6.2 RNA Isolation, Gene Expression Analysis, and Normalization ............................... 43

3. Layers of Regulation on Cell-Cycle Gene Expression in *Saccharomyces cerevisiae* 44
6.6.2 RNA isolation, sequencing, and analysis ................................................................. 139
6.6.3 RNA-Sequencing data analysis pipeline ................................................................. 141
7. Conclusions: Building, Testing, and Refining Gene Regulatory Networks .............. 142
References ...................................................................................................................... 151
Biography ...................................................................................................................... 183
List of Tables

Table 2.1: S. cerevisiae yeast strains used in Chapter 2................................................41

Table 3.2: JTK correlation scores for the 44 S. cerevisiae peptides compared to their respective mRNA transcripts from Figure 3.10. .............................................................56

Table 3.3: Previously annotated and predicted protein degradation mechanisms for cell-cycle proteins of interest. ...............................................................................................59

Table 3.4: S. cerevisiae yeast strains used in Chapter 3................................................73

Table 4.5: RNA-Seq values derived from two libraries prepared with the same RNA sample are more variable than values derived from the same library sequenced in two different lanes. ...............................................................................................................87

Table 4.6: RNA quality metrics for samples from this Chapter and our previous work showed that lower RNA integrity values (RIN) correlate with striped samples......................91

Table 4.7: Illumina barcodes and their reverse complement sequences map to the yeast genome with varying frequencies, and one particular barcode is associated with 2 striped samples. .............................................................................................................94

Table 5.8: TF network components in S. cerevisiae and sequence orthologs in C. neoformans have generally diverged in expression timing..................................................115

Table 6.9: 92 putative sequence-specific transcription factors and cyclin/CDK genes in Plasmodium falciparum. ..............................................................................................133

Table 7.10: 87 putative sequence-specific transcription factors and cyclin/CDK genes in Plasmodium falciparum have orthologous genes in P. vivax and P. knowlesi. ............148
List of Figures

Figure 1.1: Periodic gene expression programs are an emergent property of biological clock networks. ................................................................. 4

Figure 1.2: Topologies of the yeast cell-cycle TFN and the mammalian circadian cycle network.................................................................13

Figure 2.3: Simplified topology of the S. cerevisiae TFN at cell-cycle Start in G1/S phase. ...................................................................................23

Figure 2.4: Perturbation of negative feedback TFs alters re-budding oscillations........27

Figure 2.5: Perturbation of negative feedback TFs alters dynamics in the program of periodic transcription in two mutant backgrounds lacking B-type cyclin/CDK activity. ....30

Figure 2.6: Perturbation of negative feedback TFs alters expression levels of core genes in the S. cerevisiae TFN. .......................................................31

Figure 2.7: Perturbation of negative feedback TFs significantly increases the expression levels of SBF and MBF targets in three mutant backgrounds. ............................................34

Figure 2.8: Detailed topology of the S. cerevisiae TFN from G1 to S phase. .............36

Figure 2.9: Conflicting network models for the control of cell-cycle transcription........39

Figure 3.10: Proteome dynamics of cell-cycle regulators follow the transcriptome with delay during the cell cycle in S. cerevisiae.................................52

Figure 3.11: Method validation and supplementation of targeted mass spectrometry by comparing to time series immunoblots........................................54

Figure 3.12: Gene expression for many RNA-protein pairs is positively correlated in wild-type cells, including some proteins targeted for degradation, while mutant gene expression pairs are less correlated. ..................................................58

Figure 3.13: E3 ubiquitin ligase targets have landmark curve features in cycling wild-type cells, some of which are altered in clb1-6 mutant cells where much of the targeted destruction machinery is no longer active or periodic.................................62

Figure 3.14: Gene expression dynamics of G1/S TFs in wild-type and clb1-6 mutant cells reveal that repressor proteins are more periodically expressed than activators. ..........67

Figure 3.15: An integrated cell-cycle network includes windows of targeted E3 ubiquitin ligase activity. ...............................................................69
Figure 4.16: Expression results from technical replicates had significant variability pre- and post-normalization for five yeast time points. ..........................................................85

Figure 4.17: The periodic cell-cycle genes HSP150 and CIS3 are expressed at different times..............................................................................................................................88

Figure 4.18: Technical replicates show variability in RNA-Seq read coverage across two periodic genes. ..............................................................................................................89

Figure 5.19: Population synchrony for S. cerevisiae and C. neoformans over > 2 cell cycles. .........................................................................................................................105

Figure 5.20: About 20% of all S. cerevisiae and C. neoformans genes are periodically expressed during the cell cycle..............................................................................................107

Figure 5.21: Periodic, orthologous genes between S. cerevisiae and C. neoformans are differentially ordered during the cell cycle. ........................................................................109

Figure 5.22: DNA replication, spindle assembly, and mitosis genes are highly conserved in temporal ordering during the fungal cell cycles, while budding orthologs vary in their temporal expression pattern in C. neoformans. ........................................................................112

Figure 5.23: Novel periodic TFs in C. neoformans could regulate periodic gene expression..................................................................................................................116

Figure 5.24: Evidence for conservation of the TF network topology at G1/S in C. neoformans. ................................................................................................................118

Figure 6.25: A large program of periodic gene expression in the Plasmodium falciparum laboratory strains 3D7 (short, 39-hour period) and FVO-NIH (approximately wild-type, 45-hour period). .................................................................131

Figure 7.26: Periodic gene expression programs are an emergent property of biological clock networks, which are driven by networks of transcription factors and other regulators. ........................................................................143
Acknowledgements

Thank you to my advisor, Steve Haase. Thank you for your support, mentorship, teaching advice, and, most importantly, sports analogies! I truly appreciate how you taught me to think critically about everything from reviewing papers to the very first slide of weekly lab meetings. I would also like to thank my committee members Paul Magwene, Beth Sullivan, and Greg Wray for your guidance and mentorship throughout graduate school.

My research projects in the Haase Laboratory were highly collaborative, and I have many people to thank who were instrumental for the success of this work. Thank you to John Harer (Duke), Kevin McGoff (UNCC), Xin Guo (HKPU), Anastasia Deckard (Mimetics), and Francis Motta (Duke)—for your patience in explaining math to someone who only made it through Calculus 2! For technical assistance, thank you to the Duke Microarray Facility (especially Laura-Leigh Rowlette), the Duke Sequencing Facility, and the Duke Proteomics Core. A major project in my dissertation research would not have been possible without the time, experimental efforts, and patience of Matt Foster, Erik Soderblom, and Arthur Moseley in the Proteomics Facility. Thank you to Stacey Smalls Everett for your help in cleaning glassware after all of my time course experiments! On the organizational and logistical side, I would also like to thank Liz Labriola, Carolyn Weinbaum, Leslie Mavengere, Anne Lacey, Blythe Boquist, and Carla Neville. A special thank you to Jo Bernhardt for putting up with my constant emails to reserve conference rooms for lab meetings and collaborator visits!

Thank you to all of the members of the Haase Laboratory who helped me along the way. You have all inspired me, challenged me, and made me a stronger scientist
and person: Adam Leman, Chun-Yi Cho, Sara Bristow, Michael Ortiz, Crystal Sierra, Leslie Peteya, Ariana Eily, Michael Mayhew, and Mark Chee. Thank you to former lab members Dave Orlando, Charles Lin, and Laura Simmons Kovacs. Scientifically, I would also like to thank Dr. Steve James for inspiring me to pursue research beyond my undergraduate work and for the famous “fun-guys” and “fun-gal” puns!

To all of my graduate school and Duke Basketball friends, I am certain that I would not have (sanely) made it through my time in graduate school without you. I am a smarter and better member of our society thanks to you wonderful people, especially Abbe, Tiffany, Sarah, and Jessie. And, most importantly, NCAA National Champs 2015!! Thank you to my friends before graduate school, for sticking with me and putting up with the science jargon: Kate, Dennis, RJ, Elena, and Erin.

Finally, I thank my family. You are all such an inspiration and full of support for me. Mom, Dad, TJ, Marylou, Steve, Kayla, PG, Audrey, and Jack—thank you all for your visits, encouraging notes, and smiling faces. To Susan, Steve, and the rest of the St. Thomas family, thank you for welcoming me into your family with such love and enthusiasm. Brian, I feel so lucky to have had such a caring, intelligent, and wonderful person help me through graduate school each and every day.

Thank you to all of my family, friends, and academic community for helping me get where I am today and for giving me the confidence to achieve my goals for tomorrow.
1. Introduction: from the Central Dogma of Molecular Biology to Gene Regulatory Networks

DNA is transcribed into RNA, which is then translated into protein—this linear transfer of functional information in all living cells is known as the central dogma of molecular biology (Crick 1970). We now appreciate that cellular information transfer is more complex. For example, RNA can be reverse transcribed into DNA (Baltimore 1970; Temin and Mizutani 1970), and RNA can participate in catalytic reactions in addition to enzymatic proteins (Guerrier-Takada et al. 1983; Kruger et al. 1982).

A similar trend of parsimonious to complex models has occurred for our understanding of gene expression in eukaryotes. Initially, each genetic locus was thought to encode a unique enzyme protein (Beadle and Ephrussi 1936; Beadle and Tatum 1941), which fit with early descriptions of human metabolic diseases caused by single-gene mutations (Garrod 1902). Further work on quantitative traits in humans such as height (Lango Allen et al. 2010), IQ (Deary et al. 2010), and diseases with non-Mendelian inheritance (Voineagu et al. 2011) have demonstrated that variation in many genes generates variation in humans and other eukaryotic species.

DNA sequencing of the human genome revealed fewer protein-coding genes than expected for the complexity of the human species (Lander et al. 2001). Instead, it was proposed that messenger RNAs in humans are alternatively spliced and regulated in complex and cell-type specific ways. Current cutting-edge technologies are being deployed in an attempt to characterize all regulatory regions (Encode Consortium 2012) and transcription patterns (Djebali et al. 2012) in human cells. Mounting evidence suggests that regulation of gene expression levels, in addition to genome sequence
variation, accounts for physiological differences between cell types and between human individuals.

Biological systems are dynamic in time as they respond to environmental signals, grow, reproduce, and survive. At the cellular level, signal transduction pathways and transcription factors (TFs) lead to changes in gene expression as cells respond to a particular environmental stimulus. For many years, geneticists have applied forward genetic screens to identify the gene(s) responsible for a given phenotype and/or environmental response. Genetic screens in model organisms have been extremely successful in documenting the contributions of single genes to cellular phenotypes and, in some cases, to the causative mutations in human diseases with Mendelian inheritance. However, mutations in single genes are often not sufficient to explain complex diseases or quantitative traits in biological systems. It has become increasingly clear that networks of interacting genes are required to mechanistically understand complex diseases such as cancer (reviewed in: Hanahan and Weinberg 2011) or host-pathogen interactions. Furthermore, many biological processes are dynamic across time—from the 24-hour sleep/wake cycle of humans to the 1.5-hour period for a single yeast cell to divide.

In this dissertation, I argue that, in addition to regulating gene expression in complex ways, cells build complexity through networks of interacting proteins. In three species of eukaryotic organisms, I will investigate a “central dogma” of network biology, where a transcription factor (TF) protein binds to the promoter of a second TF and affects the rate of its target TF’s RNA production, forming a network of TFs. Specifically, I will discuss building gene regulatory networks of TFs that participate in dynamic biological processes over time. The human genome and ENCODE projects have been
transformative in understanding the basic structure and variation in the steady state human genome. However, biological systems are dynamic across time from the cellular to whole organism scales.

Here, I investigate networks of interacting TFs and regulatory proteins that generate dynamics in cellular processes like the cell cycle, the circadian rhythm, and the developmental cycle of a human pathogen. Building networks of genes that control a given process is critical for a complete mechanistic understanding of a biological system (reviewed in: Bar-Joseph et al. 2012). Understanding dynamics and network topology also allows us to predict how gene networks will respond to perturbation, such as drug treatment. Drug resistance is a major problem in human disease from cancer treatment to infectious diseases.

Organisms utilize biological clock networks to time a particular process or function, and many phenomena are common to such clocks. For example, large programs of periodic gene expression emerge when biological clocks are synchronized and profiled over time (Figure 1.1). The dynamics of these gene expression programs are qualitatively similar despite evolutionary divergence between eukaryotic species and processes.
Figure 1.1: Periodic gene expression programs are an emergent property of biological clock networks.

Heatmaps are shown for biological rhythms across many different time scales from 75 minutes to ~40 hours. During the budding yeast cell cycle, periodic gene expression is shown for 845 genes (12.9% of the genome) from synchronized *Saccharomyces cerevisiae* cells, data from: (Kelliher et al. 2016) (A). Periodic transcription is also observed in cycling *Cryptococcus neoformans* cells, a fungal pathogen (981 genes, 14.4% of the genome, data from: (Kelliher et al. 2016)) (B). During the human cell cycle, periodic gene expression also occurs (1495 genes, 3.9% of the genome, K562 data: unpublished) (C). Circadian rhythms from mouse liver cells display a periodic gene expression program (1079 genes, 4.6% of probe IDs, data from: (Hughes et al. 2009)) (D). Humans infected with malaria are plagued with periodic fever cycles, which correspond to the malaria parasites’ developmental cycle in red blood cells. Almost the entire transcriptome of *Plasmodium falciparum* is periodic over this developmental cycle in vitro (4617 genes, 85.4% of the genome, data: unpublished) (E). For each heatmap (A-E), genes are displayed relative to their respective mean level.
To demonstrate this similarity quantitatively, I calculated average peak widths (difference between the 50% peak time and the 50% trough time) for each list of periodic genes shown for each dataset. I observed that periodic gene programs are expressed with similar duration peak widths as a percentage of total cycle time (Figure 1.1, A: 32.1±15.7%; B: 29.0±14.2%; C: 28.9±15.6%; D: 28.6±12.6%; E: 25.8±12.2%). Commonalities in gene expression programs suggest that gene regulatory networks (composed of TFs, kinases, protein degradation machinery, and other transcriptional regulators) have evolved to buffer dynamical signals against noise and may share structural network features. We posit that biological clock networks have the following unifying characteristics: 1) composed of transcription factors, 2) contain feedback motifs and redundant nodes (to increase robustness of the clock network), and 3) respond to environmental stimuli.

In this Introduction Chapter, I provide background knowledge for the dynamic biological processes of interest for my dissertation projects. The Introduction concludes with an outline of completed experiments and results aimed at characterizing gene regulatory networks in dynamic processes of three different eukaryotes.

1.1 The Circadian Rhythm Gene Regulatory Network

One of the best-characterized gene regulatory networks is the circadian rhythm, which is present in almost all organisms to anticipate environmental light-dark cycles (reviewed in: E. E. Zhang and Kay 2010). In the mouse, nearly 40% of all expressed genes across different tissues have expression signatures of circadian regulation (R. Zhang et al. 2014). Thus, the circadian gene regulatory network drives the expression of
many genes and maintains the approximately 24-hour cycle period length (Bünning 1960).

The simplest circadian clock network is found in bacteria and is composed of three major components that form a negative feedback loop (reviewed in: Cohen and Golden 2015). The cyanobacterial KaiA, KaiB, KaiC oscillator can function autonomously in a test tube with the addition of ATP (Nakajima et al. 2005). Unlike eukaryotic molecular clocks, the cyanobacterial oscillator is truly biochemical and does not require new transcription to produce circadian dynamics and period length (Tomita et al. 2005). Bacterial and eukaryotic rhythms do share a core negative feedback loop and intriguing connections with metabolic processes (reviewed in: Reddy and Rey 2014).

*Neurospora crassa* has been widely used as a fungal model system for the study of the eukaryotic circadian clock. Mutations in the frequency (FRQ) locus were identified in screens for arrhythmic fungi (Loros and Feldman 1986; McClung et al. 1989). The FRQ protein participates in a core negative feedback loop that regulates the circadian period length (reviewed in: Hurley et al. 2016). The screens that identified FRQ in *N. crassa* and its orthologs in other eukaryotes represent the utility of genetic approaches when one gene has a large effect on the biological phenotype of interest (analogous to mapping human disease genes with Mendelian inheritance). The positive regulators of the *N. crassa* circadian network, white collar-1 (WC-1) and white collar-2 (WC-2), were characterized more than a decade later than FRQ due to complex and partially redundant roles in activating light-responsive genes in the network (Crosthwaite et al. 1997).

A circadian clock network with related structure to fungi has been characterized in the insect model *Drosophila melanogaster* and in the mammalian model *Mus*
musculus. The activating transcription factors in the mouse network form a dimer to activate target genes, and mutations in the activator TFs were found to alter the circadian period length (Gekakis et al. 1998; King et al. 1997; Vitaterna et al. 1994). Two families of transcriptional repressors form the murine negative feedback loop: Cryptochrome proteins (Kume et al. 1999) and Period proteins (Okamura et al. 1999). Homologous core clock genes were identified in flies (Allada et al. 1998; Darlington et al. 1998; Konopka and Benzer 1971), and it was further discovered that the positive arm of the network directly activates the repressive arm, forming an approximately 12-hour negative feedback loop.

Circadian rhythms in plants are essential for photosynthesis and for protection from pathogens, as plants are not mobile like many other eukaryotic species (reviewed in: Harmer 2009). The circadian network in plants was recently found to directly coordinate processes involved in metabolism and immunity with time of day (Zhou et al. 2015).

The human circadian clock network contains homologous genes to the mouse network (reviewed in: Takahashi 2017). Human mutations in circadian genes have been found to affect sleep duration (He et al. 2009) and other aspects of human physiology. Shifting an individual’s molecular rhythm out-of-phase with daily light/dark cycles, as observed in jet lag, shift workers, and military personnel, leads to a higher risk of cancer and metabolic disease. Fascinatingly, circadian perturbations to the mammalian microbiome community have also been linked to obesity and inflammation (Thaiss et al. 2014).

The circadian clock gene regulatory network has many interesting dynamical features. For example, the network is temperature-compensated such that the
approximately 24-hour period length is maintained at a range of temperatures (reviewed in: Hogenesch and Ueda 2011). The network is also designed to be robust to perturbation of individual genes. "Paralog compensation" has been demonstrated in the mouse network, where the strength of interacting TFs changes in response to genetic perturbation of a network component (Baggs et al. 2009). Thus, network robustness can come from multiple genes with partially overlapping functions (e.g. Per1, 2, 3 and Cry1, 2 paralogs) or from the network structure itself. Genes that exist in an intricate and robust gene regulatory network can foil classical genetic approaches of single or double mutant screens to characterize a phenotype of interest. Network inference on dynamical systems will become invaluable for learning these complex mechanisms.

1.2 The Eukaryotic Cell-Cycle is a Model for Building Gene Regulatory Networks

My lab has posited that the network controlling cell-cycle dynamics, like the circadian rhythm, is ancient in origin, is closely linked to metabolism, and contains highly redundant regulatory pathways (Simmons Kovacs et al. 2008). The cell cycle, metabolic networks, and circadian rhythms are important for organism fitness and are likely maintained by selection throughout evolutionary time.

The cell division cycle is a fundamental biological process underlying growth and reproduction in all organisms. The eukaryotic cell cycle is divided into four phases: Gap 1 (G1), Synthesis (S), Gap 2 (G2), and Mitosis (M). The central purpose of the cell cycle is for cells to accurately duplicate their genomic content, followed by faithful segregation of DNA and other cellular contents into two new cells (Morgan 2007). Cyclin-dependent kinases (CDKs) and their cyclin binding partners regulate the initiation and proper ordering of cell-cycle events, such as DNA replication and spindle formation (reviewed

The budding yeast *Saccharomyces cerevisiae* has one cell-cycle CDK homolog, Cdc28, which binds a variety of periodically transcribed cyclins for activation. During G1 phase, Cdc28 binds Cln1, 2, and 3 to initiate budding and other Start events, which is analogous to the restriction point in animal cells (Pardee 1974; Richardson et al. 1989). During the S/G2/M phases, Cdc28 binds six B-type cyclins, Clb1-6, to initiate DNA replication, spindle elongation, and other mitotic events (Fitch et al. 1992; Schwob and Nasmyth 1993).

Many genes that encode regulators of cell-cycle events are also periodically transcribed during the cell cycle (e.g. cyclin genes). Programs of periodic gene expression have been identified in many eukaryotes including fungi, plants, mice fibroblasts, and human cell lines (Bar-Joseph et al. 2008; Grant et al. 2013; Ishida et al. 2001; Kelliher et al. 2016; Menges et al. 2005; Oliva et al. 2005; Peng et al. 2005; Rustici et al. 2004). Many cell-cycle genes peak in mRNA expression level before their protein products are used in cell-cycle events. One canonical example is DNA replication origin firing, where replication origin proteins are assembled into a complex before S phase, activated, and then degraded or changed in localization to prevent re-replication (reviewed in: Bell and Dutta 2002). This gene expression phenomenon has been called “just-in-time transcription” (reviewed in: Haase and Wittenberg 2014). Physical cellular events occur during the cell cycle along with active gene expression in the nucleus, and recently a mechanism has been described in yeast where newly replicated regions of
DNA are actively modified to prevent copy-number issues for transcription (Voichek et al. 2016). Just-in-time transcription has also been investigated in the context of co-regulated metabolic genes, and expressing genes periodically was found to conserve cellular energy (G. Z. Wang et al. 2015). Taken together, previous studies demonstrate that many genes play a role in cell-cycle regulation and that many of them are precisely regulated and periodically transcribed.

Periodic cell-cycle genes have been described extensively in the budding yeast *Saccharomyces cerevisiae* (Cho et al. 1998; de Lichtenberg et al. 2005; P. Eser et al. 2014; Granovskaia et al. 2010; Hereford et al. 1981; Orlando et al. 2008; Pramila et al. 2006; Spellman et al. 1998), which makes *S. cerevisiae* an ideal model system to construct and test gene regulatory networks. Many previous studies have identified yeast transcription factors and their binding sites in the genome (Harbison et al. 2004; Horak et al. 2002; Hu et al. 2007; T. I. Lee et al. 2002; Maclsaac et al. 2006; Simon et al. 2001; Venters et al. 2011; Workman et al. 2006). High-throughput genetic experiments and TF binding datasets are curated in open access yeast databases (Cherry et al. 2012; Teixeira et al. 2014). Quantitative models have utilized these datasets to predict an interconnected network of periodically expressed transcription factors that is capable of driving cell-cycle transcription (Hillenbrand et al. 2016; F. Li et al. 2004; McGoff et al. 2016; Orlando et al. 2008; Sevim et al. 2010; Simmons Kovacs et al. 2012). In *S. cerevisiae*, many of the putative cell-cycle network transcription factors are regulated at the protein level by cyclin/CDKs and by ubiquitin ligases such as the Anaphase-Promoting Complex/Cyclosome (APC/C) (Christiano et al. 2014; Holt et al. 2009; Landry et al. 2014; Ostapenko and Solomon 2011; Ubersax et al. 2003).
Specific to cell-cycle regulation, checkpoint pathways are mechanisms to transiently halt cell-cycle progression upon cellular damage (Hartwell and Weinert 1989). Checkpoint effectors are thought to interact directly with cell-cycle transcription factors during arrest, but more work is needed to understand these network connections (Bristow et al. 2014; Edenberg et al. 2014; Travesa et al. 2012). My lab has proposed that feedback and cross talk between cyclin/CDKs, checkpoints, and transcription factors are crucial for faithful cell-cycle progression in budding yeast.

Molecular players in the fungal cell cycle are not as well understood outside of the model species Saccharomyces cerevisiae, Schizosaccharomyces pombe, Neurospora crassa, Ashbya gossypii, and Aspergillus nidulans. Pathogenic fungi such as Aspergillus fumigatus, Candida albicans, and Cryptococcus neoformans are of particular interest because they can cause human disease (Brown et al. 2012). By combining information from timing of gene expression during the cell cycle, evolutionarily related fungal genes, and previous laboratory work in a “comparative functional genomics” approach, future work will build mechanistic, testable hypotheses about cell-cycle gene function in pathogenic fungi.

1.3 Transcription Factor Network (TFN) Model in S. cerevisiae

As described in the previous section, much experimental effort went into understanding cell cycle regulators in the budding yeast Saccharomyces cerevisiae. Our group and others have proposed that an interconnected network of transcription factors, cyclin/CDKs, protein degradation machinery, and checkpoints drive ordered cell-cycle events and periodic transcription (Bristow et al. 2014; Chen et al. 2004; Hillenbrand et al. 2016; F. Li et al. 2004; Orlando et al. 2008; Simmons Kovacs et al. 2012; Simon et al. 2001). Transcription factors drive periodic transcription of cell-cycle genes, including
regulatory subunits such as cyclin genes. In turn, cyclin/CDKs and E3 ubiquitin ligase complexes regulate the activity, abundance, and localization of many network TFs.

Using algorithms that rank periodic expression over time, approximately 70 out of hundreds of annotated yeast transcription factor genes were classified as periodic during the cell cycle (Deckard et al. 2013; P. Eser et al. 2014; Orlando et al. 2008). Periodic TFs were then filtered by regulation data in order to build an interconnected network, leaving less than 20 TFs that have binding evidence to the promoter of another periodic TF gene(s) and have evidence of regulation by an upstream periodic TF (Harbison et al. 2004; Orlando et al. 2008; Teixeira et al. 2014). The proposed yeast TFN is interconnected such that periodic TFs sequentially regulate the synthesis of downstream TFs. Smaller sub-networks of this TFN are capable of producing oscillations in a Boolean modeling framework (Orlando et al. 2008; Sevim et al. 2010; Simmons Kovacs et al. 2012). Thus, the TFN in budding yeast had a solid research foundation prior to work presented in this dissertation. More experiments were needed to understand the interconnected network of TF proteins and to investigate the level of redundancy required for generating yeast TFN dynamics.

Transcription factor negative feedback motifs are critical for generating oscillatory behavior in the circadian clock network and in a variety of synthetic biological contexts (Elowitz and Leibler 2000; Tigges et al. 2009) (reviewed in: Ferrell et al. 2011; Novák and Tyson 2008). Using the Xenopus extract system, a negative feedback loop between B-type cyclin/CDK activity and B-type cyclin protein destruction was shown to drive rapid embryonic cell cycles (Pomerening et al. 2003). In one project in this dissertation, I asked if negative feedback on TF activators at G1/S phase is important for yeast cell-cycle progression. Gene regulatory network diagrams for the yeast cell-cycle TFN and
for the circadian clock network illustrate key negative feedback loops at G1/S phase and at the light/dark transition, respectively (Figure 1.2).

**Figure 1.2: Topologies of the yeast cell-cycle TFN and the mammalian circadian cycle network.**

Yeast regulators are shown on a cell-cycle timeline approximately by peak time of mRNA expression (A). Periodic TFs (activators in green, repressors in red), cyclin/CDKs (blue), and an E3 ubiquitin ligase (blue, Anaphase-Promphase Complex or APC/C) form an interconnected network that orders cell-cycle events and drives periodic gene expression. Pointed arrows indicate gene expression activation or increase in protein activity; blunted arrows indicate gene expression repression or decrease in protein activity. Mammalian circadian clock regulators are shown on a 24-hour timeline by approximate peak expression (B). TFs (activators in green, repressors in red) and F-box degradation machinery (blue) form interlocked feedback loops that output circadian gene expression dynamics and an approximately 24-hour molecular period length. Rhythmic changes in protein localization and degradation also contribute to the dynamics of the delayed negative feedback loop between Clock/Bmal heterodimers and their TF repressors (reviewed in: Takahashi 2017).

To date, many quantitative models of the cell cycle have relied on fitting mRNA expression data (Hillenbrand et al. 2016; Orlando et al. 2008; Sevim et al. 2010; Simmons Kovacs et al. 2012). These models have made the assumption that periodic mRNA expression is a reliable proxy for TF protein activity and/or that TF protein...
expression can be directly inferred from mRNA dynamics using simple mathematical
transformations. In other words, the protein dynamics of yeast network TFs must closely
match their periodic mRNA expression profiles for these past modeling assumptions to
be valid. In another project in this dissertation, I directly tested this assumption by
quantifying protein expression levels of core network components.

A long-term goal of this work is to fully characterize the gene regulatory networks
that control periodic gene expression and order cell-cycle events in many biological
systems and then to compare the network structures (Figure 1.2). In the last two major
projects of this dissertation, I explore dynamics data in two less well-understood models:
the cell cycle in Cryptococcus neoformans and the developmental cycle of Plasmodium
falciparum, the deadliest species of human malaria. Identification of transcription factors
and of their binding sites in the genome will be essential information to continue these
network inference projects in future work. A transcription factor (TF) deletion collection
has been generated and carefully phenotyped in C. neoformans (Jung et al. 2015). Much less genetic and genomic data are available for P. falciparum compared to genetic
model organisms.

An improved, basic biological understanding of the TF network—a novel cell-
cycle control module—will diversify our ability to treat diseases of aberrant proliferation.
Cell-cycle control modules (cyclin/CDKs, checkpoints, and transcriptional motifs) are
highly conserved across eukaryotes.

1.4 Significance: Cancer, Heart Disease, and Infectious Diseases

Failure to sustain checkpoint arrest or to control once-and-only-once duplication
events during the cell cycle can lead to hallmarks of cancer, such as genome instability
and aneuploidy. By dissecting all of the modules that control cell proliferation, we will be better equipped to treat cancers and infectious diseases at a cell biological level.

Many chemotherapeutic drugs have been designed to activate cell-cycle checkpoints, and drugs targeting CDK activity have had limited success in the clinic (reviewed in: Vermeulen et al. 2003). Cell proliferation is also improperly regulated in heart disease (Sriram and Patterson 2001). A complex and interconnected network of cyclin/CDKs, checkpoint effectors, and transcription factors contributes to cell-cycle dynamics in budding yeast cells. Cell-cycle findings in yeast are likely to generate hypotheses that are directly testable in mammalian systems due to high levels of functional conservation. For example, in human cells, retinoblastoma protein (RB) inhibits the E2F family of TFs prior to the restriction point, acting as a switch to couple cell-cycle transcription with environmental growth signals. This inhibitory interaction is functionally equivalent to Whi5 inhibiting the TF complex SBF in S. cerevisiae (reviewed in: Bertoli et al. 2013) (Figure 1.2 A).

Based on experiments from multiple human cell lines, 5-10% of protein-coding genes in the human genome have evidence for periodic expression during the cell cycle. Approximately 3000 periodic genes have been identified from synchronized human fibroblasts (Cho et al. 2001), HeLa cells (Whitfield et al. 2002), U2OS cells (Grant et al. 2013), and K562 cells (Figure 1.1 C, unpublished data). Periodic genes are implicated in functions that are relevant to cancer, such as maintenance of genome integrity, response to extracellular growth signals, and invasive behavior.

1.5 Innovation: Practicing Systems Biology

The objectives and gaps in knowledge addressed in this dissertation incorporate tools from genetics to quantitative biology. To answer questions about dynamics in
biology, I first developed many of the required experimental and computational tools while investigating the cell-cycle TFN in the budding yeast *Saccharomyces cerevisiae*. The *S. cerevisiae* genome is arguably the best-annotated eukaryotic genome (Cherry et al. 2012), and budding yeast cells are extremely genetically tractable.

Over the course of my dissertation projects, our quantitative group met weekly to discuss biological questions, build models, design optimal experiments to test in the lab, and iterate (Kitano 2002). High throughput experimental approaches have been used to study cell-cycle transcription for more than a decade (Spellman et al. 1998), but systems-level modeling of the eukaryotic cell cycle remains incomplete. Many previous mathematical models of the cell cycle have focused on cyclin/CDK activity, over-simplified transcriptional synthesis parameters, and/or omitted some critical TF interactions during the cell cycle (Chen et al. 2004; Rahi et al. 2016). Our quantitative group has approached modeling dynamical systems with a comparative approach in describing algorithms to rank periodic gene expression (Deckard et al. 2013) and in designing a network inference method (McGoff et al. 2016). A long-term goal of our quantitative group is to develop a generalizable analysis pipeline to infer gene regulatory networks from time series data. The datasets described in this dissertation directly augment these quantitative modeling goals.

### 1.6 Dissertation Outline

This dissertation furthers scientific knowledge in the cell cycle and systems biology fields and elucidates gene regulatory networks controlling dynamics in biology. I investigate cell-cycle dynamics in both *Saccharomyces cerevisiae* and *Cryptococcus neoformans* (Figure 1.1 A and B). I then apply these approaches to the developmental cycle of the malaria species *Plasmodium falciparum* (Figure 1.1 E). These projects are
unified under the theme of gene expression dynamics driven by transcription factors (Figure 1.1). Each project was conducted in a highly collaborative environment, and contributions to experiments and publications are described specifically for each Chapter.

In Chapters 2 and 3, I discuss our current understanding of the interconnected network of transcription factors (TFs) and other regulators that drive cell-cycle gene expression in the budding yeast *Saccharomyces cerevisiae* (Figure 1.1 A, Figure 1.2 A). In these sections, I address two research questions: 1) do TF repressors and negative feedback motifs regulate periodic gene expression, and 2) do protein expression dynamics of cell-cycle regulators track with their periodic mRNA levels?

Experimental approaches in Chapter 2 are modeled after the genetic perturbation approaches utilized in the circadian rhythm field to identify network components that control the 24-hour period length. I asked whether negative feedback motifs were important for sustaining oscillations in cell-cycle transcription in *S. cerevisiae*. This work is featured in a second-author publication, currently in preparation. Briefly, we found that negative feedback is required for maintaining normal gene expression levels and plays a role in maintaining cell-cycle period length.

In Chapter 3, I directly test the assumption that periodic mRNA expression accurately predicts dynamics in protein abundance, which is of interest to cell cycle and systems biologists alike. This proteomic study is in preparation as a first author manuscript. We showed that expression levels of many proteins are dynamic during a wild-type cell cycle, mirroring mRNA expression. We also profiled mutant cells and demonstrate that layers of regulation, such as targeted degradation and cyclin/CDK phosphorylation, fine-tune the protein expression levels of cell-cycle regulators.
Through *S. cerevisiae* cell-cycle projects, I developed laboratory and data analysis skills that could then be applied to other dynamical biological systems. All relevant data analysis pipelines and quantitative algorithms are described for each Chapter. I will also organize coding projects and visualization tools into a github or other repository for future laboratory members. Chapter 4 is technical in scope and discusses unpublished work on the noise and potential biases associated with RNA-sequencing technology. RNA-sequencing was applied for gene expression quantification in many of the projects described in this dissertation.

In Chapter 5, I discuss cell-cycle transcription in a budding yeast species that is diverged from *S. cerevisiae* by over 500 million years of evolution, *Cryptococcus neoformans* (Figure 1.1 B). *C. neoformans* is an opportunistic human fungal pathogen. Understanding its cell cycle is extremely important for human health and for developing novel antifungal drug targets. In this section, I address two specific research questions: 1) does periodic transcription occur during the *C. neoformans* cell cycle, and 2) how conserved are the programs of cell-cycle-regulated genes between *S. cerevisiae* and *C. neoformans* (Figure 1.1 A and B)? To address these questions, I describe work from a first-author publication, where we showed that genes are indeed periodically expressed during the *C. neoformans* cell cycle and found evidence for partial conservation of cell-cycle regulation with *S. cerevisiae*.

Chapter 6 is the final data chapter and focuses on our current, somewhat limited understanding of the red blood cycle developmental cycle in a human malaria species *Plasmodium falciparum* (Figure 1.1 E). Malaria infections are characterized by periodic fevers in human patients, which correspond to parasites bursting from host red blood cells to infect new cells. Across many different malaria hosts, parasite red blood cell
developmental cycles almost always occur in multiples of the host’s 24-hour circadian rhythm period. We and others have observed that some laboratory strains of *P. falciparum* have sped up their cycle period length, likely due to in vitro culturing conditions away from the typical infectious life cycle. In this Chapter, we investigate the genetics underlying cycle period length in malaria development.

I conclude with a discussion about the state of the systems biology field on studying gene networks and understanding network evolution. In the Chapters described above, we focused on two comparisons to ask about relationships between gene regulatory networks—cell cycle regulators in the budding yeasts *C. neoformans* and *S. cerevisiae* (separated by 500 million years of evolution) and the red blood cell developmental cycle in different laboratory strains of *P. falciparum* (separated by ~50 years of laboratory culturing). Evidence is emerging from our work and others that evolution can operate on networks of genes and/or motifs in networks. In some cases, it appears that convergent evolution has independently arrived on specific network motifs, and in other cases that selective pressure has maintained the ancestral network structure. I propose that network conservation should enable comparison of dynamical networks between species and allow us to identify putative network components in novel and uncharacterized biological systems.
2. The Role of Negative Feedback in the *Saccharomyces cerevisiae* Cell-Cycle Network


### 2.1 Author Contributions

This Chapter further investigates a proposed gene regulatory network that functions during the *S. cerevisiae* cell cycle using genetic approaches. I deleted three genes that encode repressor TFs and profiled gene expression over time in perturbation and control cells. Some experiments described in this Chapter appear in the first manuscript cited above, and our findings are relevant to the second citation.

In the first manuscript, my laboratory colleague Chun-Yi Cho wrote the manuscript, generated yeast strains and performed all transcriptome-profiling experiments for 4 of the main text figures. I contributed a mutant yeast strain and two datasets, which will be described in the Results section of this Chapter.

The second manuscript is conceptually more philosophical and describes different models in the cell-cycle field about how periodic genes are regulated. We revisit data from a recent publication from the Cross laboratory (Rahi et al. 2016) and offer an alternative interpretation. I analyzed the raw RNA-Sequencing data from the recent
publication and performed periodicity-ranking analyses. My colleague Chun-Yi Cho assembled the figures and wrote the manuscript along with Steve Haase. Our collaborators in the Math Department, Francis Motta and Anastasia Deckard, designed and implemented a quantitative model that provided key evidence for our re-interpretation of data from the recent paper (Rahi et al. 2016). Conclusions from this manuscript will be incorporated in the Discussion section of this Chapter and motivate our current thinking about cell-cycle dynamics in budding yeast.

2.2 Introduction

After the discovery that cyclin/CDKs play a conserved role in ordering eukaryotic cell-cycle events, the cell-cycle field next asked if cyclin/CDKs were also responsible for driving periodic gene expression (reviewed in: Haase and Wittenberg 2014; Simmons Kovacs et al. 2008; Wittenberg and Reed 2005). My lab used the genetic model organism *Saccharomyces cerevisiae* to directly ask if periodic cyclin/CDK activity is required for periodic gene expression. Surprisingly, the wild-type program of periodic gene expression partially persisted in a B-type cyclin/CDK mutant background (*clb1-6*) (Haase and Reed 1999; Orlando et al. 2008) and in cells with a temperature-sensitive CDK allele (*cdc28-4*) (Simmons Kovacs et al. 2012). In both cyclin/CDK mutant experiments, a population of cells was synchronized in early G1 phase, released, and then physically arrested at the G1/S border, as DNA replication cannot be initiated in the absence of B-type cyclin/CDK activity. Taken together, these results showed that periodic cyclin/CDK activity and physical cell-cycle progression are not required to drive much of the periodic transcription program.

In *S. cerevisiae*, a transcription factor network (TFN) was proposed to explain gene expression dynamics in the absence of oscillatory cyclin/CDK activity (see
Introduction Section 1.3). Using quantitative models, our group and others have proposed that a network of serially activated TFs is capable of driving oscillations in periodic transcription (Hillenbrand et al. 2016; Orlando et al. 2008; Sevim et al. 2010; Simmons Kovacs et al. 2012).

Functional redundancy and stabilizing feedback loops are both common in oscillatory biological networks (reviewed in: Ferrell et al. 2011). Biological systems are highly robust to deal with fluctuations in environmental inputs, random noise, and genetic mutations (Kitano 2002). Negative feedback is a core element of transcription oscillations in synthetic (Elowitz and Leibler 2000) and natural (U. Alon 2007; Denby et al. 2012; E. E. Zhang and Kay 2010) biological systems. In this Chapter, I implemented classical genetic techniques to delete negative feedback genes individually and in combination and used modern genomics techniques to quantify changes in transcription dynamics, which was modeled after previous functional genomics work (Kemmeren et al. 2014). Genetic redundancy has been demonstrated in the circadian network (Baggs et al. 2009) as well as other biological pathways such as galactose utilization in yeast (Acar et al. 2010). S. cerevisiae is a particularly good model to study network redundancy because the species underwent an ancestral whole genome duplication, and extant paralogs are thought to serve partially redundant functions (Kafri et al. 2005; Kaganovich and Snyder 2012). The G1/S transition of the budding yeast cell-cycle TFN appeared to have many functionally redundant TFs (Figure 2.3).
Figure 2.3: Simplified topology of the *S. cerevisiae* TFN at cell-cycle Start in G1/S phase.

Periodic TFs (activators in green, repressors in red) are placed on the cell-cycle timeline approximately by peak mRNA expression. Edges between TFs represent evidence for regulation: ChIP-chip data for TF binding and/or genetic evidence for regulation type, compiled in: (McGoff et al. 2016; Orlando et al. 2008). Pointed arrows indicate activation, and blunted arrows mark repression. TF complexes: MBF = Mbp1 and Swi6; SBF = Swi4 and Swi6.

At cell-cycle Start, approximately 200 periodic genes are activated by the heterodimeric TFs SBF and MBF (Swi4/6 cell cycle box (SCB) Binding Factor; Mlu1 cell cycle box (MCB) Binding Factor) (U. Eser et al. 2011; V. R. Iyer et al. 2001). SBF and MBF are functionally analogous to the E2F family of TFs in humans (reviewed in: Bertoli et al. 2013). Genetic evidence shows that SBF and MBF make up partially redundant paths for activating G1/S transcription, and a double *swi4 mbp1* knockout is lethal (Bean et al. 2005; Koch et al. 1993; Nasmyth and Dirick 1991). The SBF and MBF paths have distinct sets of regulators that form negative feedback loop motifs. The repressors of SBF, Yhp1 and Yox1 (Yeast Homeo-Protein and Yeast Homeobox), repress *SWI4* gene
expression levels by blocking ECB (Early Cell Cycle Box) sites in its promoter (MacKay et al. 2001). \textit{YOX1} and \textit{YHP1} are genetically required for the periodic expression of some cell-cycle genes (Pramila et al. 2002) and are paralogs from the \textit{S. cerevisiae} whole genome duplication (Kellis et al. 2004). The repressor of MBF, Nrm1 (Negative regulator of MBF targets), represses MBF activity by binding to the TF complex and acting as a co-repressor (de Bruin et al. 2006). Nrm1 is also actively regulated during checkpoint arrest, which leads to high expression of MBF target genes (de Bruin et al. 2006; Travesa et al. 2012). I hypothesized that Nrm1, Yhp1, and Yox1 make up redundant negative feedback loops early in the cell cycle, which may be required for the generation of one transcriptional wave per cycle and for proper period control.

\subsection*{2.3 Results}

In early experiments to ask if the TFN is directly involved in cell-cycle control, previous lab members assayed cell-cycle period alterations in TF deletion or overexpression mutants (Simmons Kovacs et al. 2012). Yeast cells lacking B-type cyclin/CDK activity are physically arrested at G1/S phase but continue to periodically drive new bud emergence, which first led our lab to propose a cyclin/CDK independent oscillator (Haase and Reed 1999). Periodic re-budding events in \textit{clb1-6} cells were later found to correlate with peaks of periodic G1/S genes (Orlando et al. 2008), and thus, periodic re-budding events could be used as a readout for TFN activity. Measuring re-budding as a proxy for TFN activity was simple to perform by microscopy and much more cost-efficient than transcriptional profiling for every TFN mutant. TF perturbation and control strains were synchronized in late G1 phase with alpha-factor mating pheromone and released. Budding index was monitored over time to measure re-budding cycle period length. Using an algorithmic model called CLOCCS (Characterize
Loss of Cell Cycle Synchrony), the cell-cycle period length parameter lambda (λ) was compared between paired perturbation and control experiments (Orlando et al. 2007, 2009).

A distribution of period changes (long, no change, short) was observed after individually knocking out six TF repressors or overexpressing nine different TF activators under the GAL1-10 promoter, including perturbations that significantly decreased period length (genetic perturbations that increase re-budding period length could be due to stress, metabolic changes, or other confounding variables). However, TF perturbation strains differed from the control period length by 30% or less (Figure 4 from:) (Simmons Kovacs et al. 2012). Perturbing one or two TF genes never led to an arrest in re-budding cycles. Furthermore, TF genetic perturbations only affected cell-cycle period length in cells lacking B-type cyclin/CDK activity and did not have a significant effect on period length in a wild-type background. These results suggested that the TF network is buffered against genetic perturbation to individual TFs.

I hypothesized that the TFN contains functionally redundant negative feedback motifs in G1/S phase (Figure 2.3). To test this hypothesis, I made combinatorial deletions of NRM1, YHP1, and YOX1 in both clb1-6 and GAL-SIC1Δ3P backgrounds. GAL-SIC1Δ3P cells express hyperstable Sic1, an inhibitor of B-type cyclins, and phenocopy the re-budding cycles of clb1-6 cells (Schwob et al. 1994; Verma et al. 1997). B-type cyclins also participate in negative feedback loops on the SBF/MBF TF activators. Clb2/CDK inhibits SBF protein activity (Amon et al. 1993), B-type cyclin/CDKs appear to have inhibiting effects on MBF targets in the absence of Nrm1 (de Bruin et al. 2006), and Swi6 nuclear localization is thought to be controlled by CDK phosphorylation (Sidorova et al. 1995).
Mutant cells were synchronized by alpha-factor mating pheromone (arrest at the G1/S border) or by centrifugal elutriation (mechanically select small G1 daughter cells from a size gradient) (Leman et al. 2014). I sampled synchronous populations of cells over time to count re-budding cycles. Using the CLOCCS algorithm, I calculated the period of re-budding cycles in each negative feedback mutant and paired control strains. Loss of individual repressors in the TF network did not significantly alter cell-cycle oscillations (Simmons Kovacs et al. 2012). Mutant cells lacking all negative feedback motifs had a significantly lengthened re-budding period compared to single and double TF perturbations (Figure 2.4). The G1/S negative feedback loops appeared to be partially redundant because period length increased additively as each TF repressor was knocked out (Figure 2.4 D). A sub-population of the triple mutant cells budded only once when a majority of control cells completed two re-budding cycles, indicating a partial arrest in the re-budding cycle (Figure 2.4 C). I hypothesized that the sub-population of \textit{GAL-SIC1Δ3P nmt1Δ yhp1Δ yox1Δ} mutant cells that halt re-budding cycles may fail to re-initiate waves of periodic transcription in the absence of negative feedback.
Figure 2.4: Perturbation of negative feedback TFs alters re-budding oscillations.
Mutant cells that lack B-type cyclin/CDK activity re-bud with a similar period length to a wild-type cell cycle (Haase and Reed 1999; Orlando et al. 2008) (A). GAL-SIC1Δ3P control and TF repressor mutant cells were synchronized in G1 phase using either alpha-factor or centrifugal elutriation and released into YEP 2% galactose for induction of the stabilized Sic1 construct. Re-budding events were monitored by counting at least 200 cells per time point. The period length (λ) was estimated using the CLOCCS algorithm (Orlando et al. 2007, 2009). As shown previously, when two TF repressors are deleted, re-budding period is significantly lengthened (Simmons Kovacs et al. 2012). Representative budding curves from GAL-SIC1Δ3P yhp1Δ yox1Δ mutant cells (red) compared to the GAL-SIC1Δ3P control (black) show a 23% increase in re-budding cycle length (230.9 and 187.7 minutes, respectively) (B). When three TF repressors were deleted, re-budding period length was further lengthened. Representative budding curves from GAL-SIC1Δ3P nrm1Δ yhp1Δ yox1Δ mutant cells (red) compared to GAL-SIC1Δ3P controls (black) show a 42% increase in re-budding cycle length (246.9 and 173.5 minutes, respectively) (C). Serial deletion of TF repressors demonstrated that the triple mutant had the most significant effect on re-budding period length relative to internal control cells (D). The three TF repressors were also deleted in the clb1-6 strain background, and re-budding period length was compared to internal controls (D, purple bar). Average percent change in re-budding period length is shown for each mutant relative to paired controls, as in (Simmons Kovacs et al. 2012). Bars indicate standard deviations, and asterisks (**) indicate p-values < 0.01 (paired t-test).

Partial arrest in re-budding cycles indicated that G1/S transcript levels may be perturbed in the GAL-SIC1Δ3P nrm1Δ yhp1Δ yox1Δ background. Interestingly, the triple TF mutant in the clb1-6 background had a longer re-budding cycle period length compared to controls, but the period extension was not as dramatic as in the GAL-SIC1Δ3P background (Figure 2.4 D, purple bar). Therefore, we profiled transcriptome dynamics of triple mutant cells in both backgrounds to ask if negative feedback plays a role in TFN function using microarray (GAL-SIC1Δ3P experiments) or RNA-Seq (clb1-6 experiments) (Materials and Methods).

Perturbation of B-type cyclins and three TF network repressors (NRM1, YHP1, and YOX1) affected transcription dynamics compared to cycling wild-type cells and to control cells lacking B-type cyclin/CDK activity only (Figure 2.5). The transcriptome dynamics of these mutants were complex, and so we also visualized the expression levels of core TFN genes by line graph (Figure 2.6). Two to three cell-cycle oscillations
can be seen for core TFN genes in wild-type experiments, with expected expression-level damping over time due to asymmetric cell division in budding yeast. Mutant cells lacking B-type cyclin/CDK activity (blue/green lines) showed at least one wave of gene expression over time (Figure 2.5, Figure 2.6), as shown previously (Orlando et al. 2008). Mutant GAL-SIC1Δ3P nrm1Δ yhp1Δ yox1Δ cells exhibited constitutively high gene expression levels for TFN components (Figure 2.6) and for many early cell-cycle genes in the gene expression program (Figure 2.5). The clb1-6 nrm1Δ yhp1Δ yox1Δ mutant had a more modest effect on increasing TFN gene expression levels (Figure 2.6).
Figure 2.5: Perturbation of negative feedback TFs alters dynamics in the program of periodic transcription in two mutant backgrounds lacking B-type cyclin/CDK activity.

A program of ~880 genes was shown to maintain periodic expression in the absence of B-cyclin/CDKs (Orlando et al. 2008), which is illustrated here by heatmap (A-B). Wild-type data (A, period = 77.1 minutes) were obtained from the previous study, where cells were synchronized by elutriation and released into 2% YEPD plus 1M sorbitol. GAL-SIC1Δ3P (A, re-budding period = 150.9 mins) and GAL-SIC1Δ3P nrm1Δ yhp1Δ yox1Δ (A, re-budding period = 233.8 mins) cells were synchronized by elutriation and released into 2% YEPG (A). Wild-type (B, period length = 68.8 mins), clb1-6Δ (B, re-budding period = 108.9 mins), and clb1-6Δ nrm1Δ yhp1Δ yox1Δ (B, re-budding period = 97.5 mins) cells were synchronized by alpha-factor mating pheromone and released into 2% YEPD (B). Wild-type data were obtained from a previous study (Kelliher et al. 2016). Genes along the y-axis are in the same order for all heatmaps (Figure 4: (Orlando et al. 2008)). Transcript levels are depicted as a z-score change relative to mean expression.
Figure 2.6: Perturbation of negative feedback TFs alters expression levels of core genes in the S. cerevisiae TFN.

A TF network model was proposed to explain transcription dynamics of wild-type and B-type cyclin mutant cells (Figure 4: (Orlando et al. 2008)). The core chain of activators consists of SBF/MBF (SWI4 gene shown), Hcm1 (HCM1 shown), SFF (NDD1 shown), and Swi5/Ace2 (SWI5 shown). Line graphs depict the absolute transcript expression levels (arbitrary units, AU) in wild-type and mutant datasets from microarray (top 2 rows) and RNA-Seq profiling experiments (bottom 2 rows). TFN genes were activated in order in wild-type cells (biological replicates in red, orange lines). TFN genes maintained ordering in B-type cyclin mutant cells (biological replicates in green, blue lines) in both GAL-SIC1Δ3P (row 2) and the clb1-6 background (row 4). In the absence of TF repressors (purple lines, rows 2 and 4), TFN activators were more highly expressed.
Perturbation of both B-type cyclin/CDK activity and three TF repressors led to higher gene expression levels and significantly longer period length. Both phenotypes were more pronounced in the GAL-SIC1Δ3P background than in clb1-6 cells. Since Nrm1, Yhp1, and Yox1 repressors act directly on MBF and SBF, I next focused on putative SBF/MBF target genes (Bristow et al. 2014; U. Eser et al. 2011; Ferrezuelo et al. 2010). B-type cyclin/CDKs can inhibit SBF/MBF activity in addition to the TF repressors (Amon et al. 1993; de Bruin et al. 2006; Sidorova et al. 1995). For example, more recent work has shown that SBF target genes do not oscillate in cells arrested with high Clb2/CDK activity (Bristow et al. 2014). Some G1/S genes did continue to be periodically expressed in these cdc20Δ mutant cells despite high Clb2/CDK expression. It was proposed that these G1/S genes can still oscillate in high B-type cyclin/CDK conditions through the MBF/Nrm1 negative feedback loop (Bristow et al. 2014). Therefore, I deleted NRM1 in the cdc20Δ background in order to test this hypothesis.

Taken together, these experiments were designed to assay the specific contribution of TF negative feedback on G1/S gene expression in both high Clb/CDK activity (cdc20Δ nrm1Δ) and low Clb/CDK activity conditions (GAL-SIC1Δ3P nrm1Δ yhp1Δ yox1Δ and clb1-6 nrm1Δ yhp1Δ yox1Δ).

SBF/MBF target genes in G1/S phase have altered expression patterns in the absence of TF negative feedback (Figure 2.7). Triple TF mutant cells in the GAL-SIC1Δ3P background have significantly high, constitutive expression of G1/S genes compared to GAL-SIC1Δ3P alone (Figure 2.7 A). Perturbing negative feedback in the clb1-6 background had a more modest effect on G1/S gene expression dynamics, but SBF/MBF target genes were expressed at significantly higher average levels in triple
mutant cells compared to controls (Figure 2.7 B). Unexpectedly, the program of
SBF/MBF target genes in clb1-6 triple mutants continued to be sharply inactivated at
about 50 minutes into the experiment compared to triple mutant cells in the GAL-
SIC1Δ3P background. Furthermore, perturbation of NRM1 in the cdc20Δ background did
not halt oscillations for a subset of G1/S genes (Figure 2.7 C). In summary, these data
show that TF negative feedback is required to maintain normal gene expression levels,
and perturbation of negative feedback leads to higher and more constitutive G1/S gene
expression in all mutants. However, perturbation of negative feedback TFs was not
sufficient to halt G1/S gene expression oscillations in two of three mutant conditions.
Figure 2.7: Perturbation of negative feedback TFs significantly increases the expression levels of SBF and MBF targets in three mutant backgrounds.
A list of 160 SBF and MBF target genes was obtained from a previous study (Bristow et al. 2014), filtered by the program of ~880 periodic genes (Figure 2.5), and the remaining 133 genes are shown in heatmaps (A-C). GAL-SIC1Δ3P (A, top) and GAL-SIC1Δ3P nrm1Δ yhp1Δ yox1Δ (A, bottom) cells were synchronized by elutriation and released into YEP 2% galactose. clb1-6 (B, top) and clb1-6 nrm1Δ yhp1Δ yox1Δ (B, bottom) cells were synchronized by alpha-factor and released into YEP 2% dextrose. cdc20Δ (C, top) and cdc20Δ nrm1Δ (C, bottom) cells were synchronized by elutriation and released into YEP 2% dextrose. cdc20Δ data (C, top) were obtained from a previous study, where cells released into YEP 2% dextrose with 1M sorbitol media (Bristow et al. 2014). Genes along the y-axis are in the same order for all heatmaps, and this order was taken from a previous study (Figure 4 from: (Orlando et al. 2008)). Transcript levels are depicted as a z-score change relative to mean expression for each gene in each individual dataset, where values represent the number of standard deviations away from the mean (A-C). Mean expression levels for the 133 periodic SBF/MBF target genes are plotted as boxplots in repressor mutants (purple) compared to controls (green). Bars indicate standard deviations, and asterisks (**) indicate p-values < 0.01 (paired t-test).

2.4 Discussion

Perturbation of Nrm1, Yhp1, and Yox1 negative feedback TFs caused higher gene expression levels of cell-cycle genes, especially for G1/S phase genes, and lengthened cell-cycle period in B-type cyclin/CDK mutant backgrounds. These findings are summarized in the context of the current TFN model (Figure 2.8). However, variability in the effect size of negative feedback perturbation was observed between the three cell-cycle mutants tested.
Figure 2.8: Detailed topology of the S. cerevisiae TFN from G1 to S phase.

Periodic TFs (activators in green, repressors in red) are placed on the cell-cycle timeline approximately by peak mRNA expression. In blue, G1 (CLN) and B-type (CLB) cyclin/CDKs, the Anaphase-Promoting Complex (APC/C), and Sic1 regulate each other and TFs in the network (A). Edges between TFs represent evidence for regulation: ChIP-chip data for TF binding and/or genetic evidence for regulation type, compiled in: (McGoff et al. 2016; Orlando et al. 2008). Edges between regulatory proteins and TFs represent protein-level modifications (e.g. phosphorylation or ubiquitination). Pointed arrows indicate activation, and blunted arrows mark repression. TF complexes: MBF = Mbp1 and Swi6; SBF = Swi4 and Swi6. In the absence of negative feedback (gray) (B), cell-cycle genes accumulated at higher expression levels.

The GAL-SIC1Δ3P and clb1-6 mutant backgrounds were genetically engineered to inhibit or remove B-type cyclin/CDK activity, respectively. These cells exhibit similar re-budding cycle phenotypes and were thought to be phenocopies of the same mutant cell-cycle state. However, GAL-SIC1Δ3P and clb1-6 cells showed variable results in the absence of negative feedback (Figure 2.5). Specifically, G1/S transcripts still appear to be inactivated in clb1-6 triple TF mutant cells compared to GAL-SIC1Δ3P (Figure 2.7). I hypothesize that differences in growth media may explain the variability between GAL-SIC1Δ3P and clb1-6 dynamics. After synchronization, clb1-6 cells were released into
rich dextrose sugar media, while \textit{GAL-SIC1Δ3P} cells were released into galactose sugar media. Galactose treatment is known to activate different carbon source genes and lead to some canonical stress response signatures in yeast (Gasch et al. 2000). Cells sense both carbon source and stress signals in G1 phase and integrate these signals before activating SBF/MBF to commit to the cell cycle (Ewald et al. 2016; Zhao et al. 2016). Recent work has even proposed that some cell-cycle genes are under direct control by metabolic TFs (Papagiannakis et al. 2017).

Given this difference in media conditions, there are two models to explain different dynamics of SBF/MBF target genes in the absence of Nrm1, Yhp1, and Yox1 (Figure 2.7): an uncharacterized co-repressor of SBF/MBF is present in rich media (YEPD) and not in poor media (YEPG), or an unknown co-activator of SBF/MBF is present in poor media (YEPG) only. Molecularly, these models could also be fit with previously identified co-factors of SBF/MBF complexes. For example, the Swi6 co-activator protein could be differentially localized in different media conditions, while Swi4 and Mbp1 remain bound to DNA constitutively. The Msa1 and Msa2 co-factors, which have been previously implicated in a G0 quiescence-like state in yeast, could have differential activity in the two media conditions (S. Miles and Breeden 2016). Alternatively, an unknown co-factor could be responsible for the observed difference in SBF/MBF activity.

\textit{GAL-SIC1Δ3P nrm1Δ yhp1Δ yox1Δ} mutant cells had the most dramatic effect on gene expression dynamics through the TFN (Figure 2.5, Figure 2.6). Interestingly, SBF/MBF target genes did not continuously increase in expression levels over the time series in the absence of negative feedback (Figure 2.6, Figure 2.7), which could suggest that the SBF/MBF complexes are not fully active over time. I posit that the TFN is highly
interconnected, and these results may reflect network compensation to the removal of negative feedback. Previous genetic data also indicate that there are redundant paths to activate the cell-cycle transcription program outside of SBF/MBF. Although $swi4\Delta$ $mbp1\Delta$ double mutant cells are inviable (Koch et al. 1993), this mutant condition can be rescued in a $P_{CLN3-CLN2} swi4\Delta$ $mbp1\Delta$ background (Bean et al. 2005). Other groups have proposed that S phase TFs may have overlapping target genes with SBF and MBF (Pramila et al. 2006). Perhaps strong activation is not a dynamical feature of SBF/MBF at cell-cycle Start, and instead, the TFN has evolved to robustly and redundantly activate the program of gene expression. It would be interesting to test the model that SBF/MBF do not act as constitutive activators using a genetic overexpression “fast-on” system for the $SWI4$ or $MBP1$ genes (McIsaac et al. 2013, 2014). A $SWI4$ fast-on strain has been obtained from the Botstein laboratory.

We have proposed that the cell-cycle gene regulatory network in yeast is highly complex and interconnected with machinery like cyclin/CDKs and ubiquitin ligase complexes (Cho et al. 2017, manuscript submitted). Other groups in the field have argued that biochemical periodicity in cyclin/CDK activity and destruction drives the cell cycle and that cell-cycle genes are controlled by an unconnected group of TF regulons (SBF/MBF in G1/S phase; Hcm1/Tos4/Plm2 in S phase; SFF in G2/M phase) (Rahi et al. 2016) (Figure 2.9).
Figure 2.9: Conflicting network models for the control of cell-cycle transcription.

Periodic TFs (activators in green, repressors in red) are placed on the cell-cycle timeline approximately by peak time of mRNA expression. In blue, G1 (CLN) and B-type (CLB) cyclin/CDKs, APC/C, Cdc14, and Sic1 regulate each other and TFs in the network. Edges from TFs represent evidence for transcriptional regulation: ChIP-chip data for TF binding and/or genetic evidence for regulation type (compiled in: (McGoff et al. 2016; Orlando et al. 2008)). Edges between regulatory proteins and TFs represent protein-level modifications (e.g. phosphorylation or ubiquitination). Pointed arrows indicate gene expression activation or increase in protein activity; blunted arrows indicate repression of gene expression or decrease in protein activity. In a recent paper, the Cross lab argued that cell-cycle transcription is primarily controlled by a biochemical oscillator of CDK and APC/C activity (blue nodes, solid edges) (Figure S1 from: (Rahi et al. 2016)). The Haase lab posits that cell-cycle transcription is driven by an integrated network of cyclin/CDKs, transcription factors (TFs), and E3 ubiquitin ligase activity (all nodes, all edges). Dashed edges are included in the Haase lab network model but not in the Cross lab model.

In this Chapter, I have shown that TF negative feedback is required for maintaining normal gene expression levels and for cell-cycle period length in B-type cyclin/CDK mutants. These results demonstrate that perturbing regulatory edges between TFs lead to changes in cell-cycle gene expression levels. These functional genomics experiments and others (Cho et al. 2017, manuscript submitted) argue that
TFs play an important and interconnected role in driving cell-cycle gene expression dynamics. Therefore, we suggest that the cell-cycle gene regulatory network in *S. cerevisiae* has both transcriptional and biochemical components (Figure 2.9).

### 2.5 Future Work

By genetically perturbing TFN components in G1/S phase, I have shown that negative feedback contributes to cell-cycle gene expression dynamics. Our results also suggest that the SBF and MBF TF complexes may not produce constitutively high activation of target genes in the absence of negative feedback. As suggested in the Discussion section above, it would be interesting to test this hypothesis in future work by overexpressing *SWI4* in a "fast-on" yeast strain and assaying expression levels of its target genes over time (McIsaac et al. 2013, 2014).

The finding that G1/S negative feedback is important to maintain gene expression levels and cell-cycle period length could be further tested in a wild-type background. Mutant *nrm1Δ yhp1Δ yox1Δ* cells could be co-cultured with wild-type cells in a survival assay. I expect that triple mutant cells would be less viable than wild type due to prolonged expression of G1/S transcripts. Furthermore, gene expression dynamics could be profiled in triple *nrm1Δ yhp1Δ yox1Δ* mutants to ask if negative feedback on SBF/MBF targets is delayed until the predicted Clb/CDK negative feedback.

Future work could also compare dynamics signatures in serial perturbations of TF repressors to learn about the contributions of each negative feedback motif. To ask how periodic curve shapes change in mutant conditions, our group has developed algorithms to quantify curve shape correlations (see Chapter 3). This functional genomics approach would further clarify both the redundancy and unique contributions of negative feedback motifs to dynamics.
Negative feedback motifs were a logical first step to investigate TFN oscillator control mechanisms. Other transcriptional motifs likely play a role in TFN function and dynamics. The systematic perturbation-and-dynamics experiments described in this Chapter can be directly applied to study other motifs. Future work examining other transcriptional motifs and connections between the TFN, cyclin/CDKs, E3 ubiquitin ligases, and checkpoint effectors will continue to define cell-cycle control modules at the molecular level.

2.6 Materials and Methods

2.6.1 Yeast strains, cultures, and synchronization

All yeast strains are derivatives of *S. cerevisiae* BF264-15D MATa (Reed et al. 1985). Gene deletions were performed using standard yeast methods (Gietz and Schiestl 2007). Yeast strain genotypes from this Chapter are summarized in the Table below.

Table 2.1: *S. cerevisiae* yeast strains used in Chapter 2.

<table>
<thead>
<tr>
<th>Strain ID</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBY661</td>
<td>MATa arg4 clb1::URA3 clb2::LEU2 clb3::TRP1 clb4::HIS2 clb5 clb6::ADE1 P_gal1-CLB1::LEU2</td>
<td>(Haase and Reed 1999)</td>
</tr>
<tr>
<td>SBY934</td>
<td>MATa bar1::hpMX4 arg4 clb1::URA3 clb2::LEU2 clb3::TRP1 clb4::HIS2 clb5::ARG4 clb6::ADE1 P_gal1-CLB1::LEU2</td>
<td>(Chee and Haase 2010)</td>
</tr>
<tr>
<td>SBY1399</td>
<td>MATa bar1 GAL-SIC1Δ3P::URA3</td>
<td>(Simmons Kovacs et al. 2012)</td>
</tr>
<tr>
<td>SBY1394</td>
<td>MATa bar1 GAL-SIC1Δ3P::URA3 yox1Δ::kanMX6</td>
<td></td>
</tr>
<tr>
<td>SBY1552</td>
<td>MATa bar1 GAL-SIC1Δ3P::URA3 yhp1Δ::hpMX4 yox1Δ::kanMX6</td>
<td></td>
</tr>
<tr>
<td>SBY1554</td>
<td>MATa bar1 GAL-SIC1Δ3P::URA3 yhp1Δ::hpMX4</td>
<td></td>
</tr>
<tr>
<td>SBY2312</td>
<td>MATa bar1 GAL-SIC1Δ3P::URA3 nrm1Δ::natMX4</td>
<td>This Chapter</td>
</tr>
<tr>
<td>SBY2320</td>
<td>MATa bar1 GAL-SIC1Δ3P::URA3 nrm1Δ::natMX4 yhp1Δ::hpMX4 yox1Δ::kanMX6</td>
<td>(Cho et al. 2017, unpub.)</td>
</tr>
<tr>
<td>SBY2377</td>
<td>MATa bar1 GAL-SIC1Δ3P::URA3 nrm1Δ::natMX4 yox1Δ::kanMX6</td>
<td>This Chapter</td>
</tr>
</tbody>
</table>
Yeast cultures were grown in standard YEP medium (1% yeast extract, 2% peptone, 0.012% adenine, 0.006% uracil supplemented with 2% sugar). For centrifugal elutriation of \textit{GAL-SIC1AΔ3P} strains, cultures were grown to mid-log phase in YEP sucrose (YEPS) media at 30°C. Elutriated early G1 cells were then resuspended in pre-warmed YEP galactose (YPEG) media at 30°C to induce expression of the stabilized Sic1AΔ3P construct. Aliquots were sampled every 10 minutes and subsequently assayed by microarray. \textit{GAL-CLB1 clb1-6} strains were cultured in YEP galactose (YPEG) media at 30°C and grown to mid-log phase. For synchronization, alpha-factor mating pheromone was added at a concentration of 5 μg/ml for \textit{BAR1} strains or 30 ng/ml for \textit{bar1} strains (Table 2.1). Synchronous G1 cells were then resuspended in pre-warmed YEP dextrose (YPED) media at 30°C. Aliquots were sampled every 10 minutes and subsequently assayed by RNA-Sequencing. For centrifugal elutriation of \textit{GALL-CDC20 cdc20Δ} strains, cultures were grown to mid-log phase in YEP galactose (YPEG) media at 30°C. Elutriated early G1 cells were then resuspended in pre-warmed YEP dextrose (YPED) media at 30°C to remove Cdc20 expression and cause cell-cycle arrest at the metaphase-to-anaphase transition (Bristow et al. 2014). Aliquots of \textit{cdc20Δ nrm1Δ} cultures were sampled every 10 minutes and subsequently assayed by microarray.
2.6.2 RNA Isolation, Gene Expression Analysis, and Normalization

Total RNA was isolated by acid phenol extraction as described previously (Leman et al. 2014). RNA samples were cleaned using RNA Clean and Concentrator kits (Zymo Research) if required to improve quality control measurements (Nanodrop).

Microarray samples were submitted to Duke Microarray Facility for labeling, hybridization, and imaging. mRNA was amplified and labeled by Ambion MessageAmp Premier kit (Ambion Biosystems) and hybridized to the Yeast Genome 2.0 Array (Affymetrix). Raw data CEL files were normalized together using the dChip method from the Affy package in Bioconductor, as described (Bristow et al. 2014). Microarray gene expression datasets from this Chapter are publically available at the NCBI Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) under accession numbers GSE8799 (Orlando et al. 2008), GSE49650 (Bristow et al. 2014), and GSE75694 (Cho 2017, manuscript in preparation). The \( \text{cdc20}^{\Delta} \text{nrm1}^{\Delta} \) data are unpublished.

RNA-Sequencing samples were submitted to the Duke Sequencing Facility for library preparation and sequencing. Our RNA-Seq data analysis pipeline is described in detail for datasets in Chapters 3-5 below. The \( \text{clb1-6} \) RNA-Seq gene expression datasets described in this Chapter are currently unpublished (see Chapter 3 for pending \( \text{clb1-6} \) data publication).
3. Layers of Regulation on Cell-Cycle Gene Expression in \textit{Saccharomyces cerevisiae}

\textbf{Citation:} Kelliher CM, Foster MW, Motta FC, Deckard A, Soderblom EJ, Moseley MA, Haase SB. (2017). Layers of regulation on cell-cycle gene expression in the budding yeast \textit{Saccharomyces cerevisiae}. \textit{Manuscript in Preparation}.

\subsection*{3.1 Author Contributions}

This Chapter describes the well-characterized gene regulatory network that functions during the \textit{S. cerevisiae} cell cycle. We measured protein expression levels of key cell-cycle regulators. These experiments will appear in the manuscript cited above.

In this manuscript, I performed the wild-type and \textit{clb1-6} mutant \textit{S. cerevisiae} time series transcriptome and proteome profiling experiments and all relevant lead-up experiments. Matt Foster developed the targeted proteomics method and ran the assay. I analyzed the raw RNA-sequencing data and raw mass spectrometry data from both experiments (with assistance from Matt Foster and Erik Soderblom) and produced the figures and tables herein. Our collaborators in the Math Department, Francis Motta and Anastasia Deckard, wrote code and implemented algorithms to compare gene expression curve shapes and to quantify noise levels in the datasets. Matt Foster and I wrote the manuscript with guidance from Steve Haase.

\subsection*{3.2 Introduction}

In the budding yeast \textit{Saccharomyces cerevisiae}, transcription factors (TFs) regulate the periodic expression of many genes, including gene products required for progression through cell-cycle events. Quantitative models have shown that a network of
interconnected transcription factors is capable of producing the cell-cycle gene expression oscillations observed in vivo. These quantitative models were built using periodic mRNA expression as a proxy for TF activity. To determine whether transcript dynamics match protein expression, we applied a multiplexed targeted mass spectrometry approach (parallel reaction monitoring) to quantify TF protein dynamics during the *S. cerevisiae* cell cycle. Protein expression dynamics of many TFs and cell-cycle regulators closely followed their respective transcript dynamics in cycling wild-type cells, but discordant mRNA/protein expression dynamics were also observed, for example, in proteins targeted for degradation by E3 ubiquitin ligase complexes such as SCF (Skp1/Cul1/F-box) and APC/C (anaphase-promoting complex/cyclosome). We further profiled mutant cells lacking B-type cyclin/CDK activity (*clb1-6*), where oscillations in ubiquitin ligase activity, cyclin/CDKs, and cell-cycle progression are halted. We found that a number of proteins were no longer periodically degraded in *clb1-6* mutants compared to wild type. Surprisingly, we found that activating TFs at G1/S phase are more constitutively expressed than their periodic mRNA expression pattern in both wild-type and *clb1-6* cells. We posit that the abundance of the negative regulators of G1/S TFs plays a major role in the periodic expression of G1/S genes in both wild-type and mutant cells. Our findings show that multiple layers of regulation (transcription, protein stability, and proteasome targeting) affect protein expression dynamics during the cell cycle.

The eukaryotic cell cycle is a complex biological process, where many regulatory proteins have been characterized (Morgan 2007). In particular, the cell cycle is very well understood in the budding yeast model system *Saccharomyces cerevisiae* with nearly 50 years of experimentation (Elliott and McLaughlin 1978; Gullov and Friis 1984;
Hartwell et al. 1974; Lord and Wheals 1981). In *S. cerevisiae*, cell cycle events are ordered and regulated by cyclins, cyclin-dependent kinases (CDKs), protein degradation machinery, and transcription factors (Bai et al. 1996; J. Bloom and Cross 2007; Haase and Wittenberg 2014; Visintin et al. 1997). Core TFs control the expression of cyclin gene products and the expression of many other genes that are periodically expressed during the cell cycle (Bristow et al. 2014; Cho et al. 1998; de Lichtenberg et al. 2005; P. Eser et al. 2014; Granovskaia et al. 2010; T. I. Lee et al. 2002; Orlando et al. 2008; Pramila et al. 2006; Simmons Kovacs et al. 2012; Simon et al. 2001; Spellman et al. 1998). Cell-cycle machinery is highly conserved across eukaryotes (Elledge and Spottswood 1991; M. G. Lee and Nurse 1987; Ninomiya-Tsuji et al. 1991), and periodic gene expression occurs in many species (Bar-Joseph et al. 2008; Cho et al. 2001; Grant et al. 2013; Ishida et al. 2001; Kelliher et al. 2016; Menges et al. 2005; Oliva et al. 2005; Peng et al. 2005; Rustici et al. 2004; Whitfield et al. 2002). We have proposed that cell-cycle genes are expressed “just-in-time” for their respective events, a phenomenon which may have evolved in eukaryotes to aid in the proper ordering of cell-cycle events (Haase and Wittenberg 2014; Kelliher and Haase 2017; Simmons Kovacs et al. 2008).

Despite the observation that many regulatory components of the eukaryotic cell cycle are periodically transcribed, it is largely unknown whether changes in mRNA alone are sufficient to produce dynamic protein expression. Post-transcriptional modifications such as phosphorylation and ubiquitination are known to affect protein activity and rate of turnover, respectively, for cell-cycle regulators (Archambault et al. 2004; Breitkreutz et al. 2010; Swaney et al. 2013; Ubersax et al. 2003). Indeed, a large body of work has shown that there can be a poor correlation between mRNA and protein expression (de Godoy et al. 2008; Futcher et al. 1999; Gygi et al. 1999; Jensen et al. 2006; Lackner et
(reviewed in: (Vogel and Marcotte 2012)). However, the majority of studies have focused on the analysis of relative mRNA and protein expression between experimental groups, without regard to the timing or kinetics of mRNA and protein turnover. A few previous studies have profiled proteome dynamics for *S. cerevisiae*, *Schizosaccharomyces pombe*, and human cells at time points corresponding to major cell-cycle phases (Carpy et al. 2014; Flory et al. 2006; Ly et al. 2014). With more dense sampling over the cell cycle, our group and others have previously demonstrated in *S. cerevisiae* that cell-cycle proteins can have tightly-controlled periodic expression reflecting their mRNA expression levels, although these analyses have only been performed on a small number of proteins (Ball et al. 2011; Serikawa et al. 2003).

Here, we quantified the relationship between mRNA and protein expression over a dense sampling (20 or more time points) and across of a large number (~50) of *S. cerevisiae* TFs and other cell-cycle proteins. While RNA-sequencing technologies now enable complete quantitation of the *S. cerevisiae* transcriptome, a complementary methodology is still lacking for proteome-wide quantitation of protein expression. ORF-tagging strategies have enabled single-cell fluorescence or immunoblotting analysis of protein expression (Chong et al. 2015; Ghaemmaghami et al. 2003; Huh et al. 2003; Kulak et al. 2014), but available tools do not cover the entirety of the *S. cerevisiae* proteome and/or are not amenable to multiplexing. Likewise, quantitative proteomic approaches using liquid chromatography tandem mass spectrometry (LC-MS/MS) can achieve nearly complete coverage of the *S. cerevisiae* proteome (~5000 proteins) but require extensive fractionation and are currently limited to comparisons of 10 or fewer samples (Paulo et al. 2015). Targeted proteomic assays, on the other hand, offer high
sensitivity and quantitative precision and are capable of multiplexing proteins across unlimited numbers of samples (Costenoble et al. 2011; Picotti et al. 2009). We sought to apply a targeted proteomic approach, for the first time, to quantify cell-cycle-dependent protein expression across a large set of *S. cerevisiae* proteins.

In wild-type *S. cerevisiae* cells, we found that many cell-cycle regulator proteins match their respective mRNA dynamics. Interesting exceptions included proteins targeted for degradation by E3 ubiquitin ligase complexes, such as SCF and APC/C, and the G1/S transcriptional activator complexes SBF and MBF. We then profiled mutant cells and demonstrated that periodic cyclin/CDK and periodic ubiquitin ligase activity are not required for the dynamic signatures of some cell-cycle proteins of interest.

### 3.3 Results

#### 3.3.1 Targeted mass spectrometry method development

Using data from our laboratory and others, we selected 45 cell-cycle proteins of interest, including low-abundance periodic TFs, and 4 constitutively-expressed controls for cytoplasmic (Rim11, Vps9) and nuclear (Cic1, Taf12) localization (Supplementary Table 1). We developed a targeted proteomic assay for multiplexed quantitation of these proteins. To select peptides for quantitation, we used available discovery-based and targeted proteomic data from PeptideAtlas and other sources, along with previously published guidelines for optimal peptide selection (Materials and Methods) (Desiere et al. 2006; Mirzaei et al. 2013). In the absence of prior experimental data, we chose peptide sequences from those that were selected by Aebersold and colleagues to quantify the entire yeast proteome (Picotti et al. 2013). In total, 129 stable isotope-labeled (SIL) internal standards were synthesized to these 49 proteins, and we were
able to identify 97 (75.2%) of these peptides by data-dependent LC-MS/MS analysis in the time series experiments (Supplementary Table 1).

To test the assay, SIL peptides were spiked into whole-cell tryptic digests of asynchronous wild-type S. cerevisiae cells, and 1 µg of digests were analyzed by LC-MS/MS using parallel reaction monitoring (PRM), a highly sensitive targeted proteomic approach. Native yeast peptides were identified based on the retention time and MS/MS spectra of the SIL peptide standards. After removing targets that had poor reproducibility across triplicate analyses or were undetectable above noise, we were able to quantify 38 peptides belonging to 22 proteins (only 45% of the proteins of interest). Normalization to SIL provided an approximate measure of protein quantity. Compared to other approaches for estimating protein abundance in S. cerevisiae, our quantitative analysis best fit to copy number determined by label-free proteomics versus analyses of epitope- or GFP-tagged yeast strains (Supplementary Table 1). Nonetheless, many of our targets of interest were undetectable by PRM under these conditions. Since many cell-cycle regulators are transiently expressed during the cell cycle, we hypothesized that undetectable proteins in asynchronous yeast samples were diluted below the limits of detection.

3.3.2 Many cell-cycle regulators exhibit dynamic protein expression during a wild-type cell cycle

We recently profiled transcription dynamics from wild-type S. cerevisiae cells across 3 cell cycles using RNA-sequencing (time points collected every 5 minutes) (Kelliher et al. 2016). To compare cell-cycle dependent protein expression under identical conditions, S. cerevisiae cells were cultured in YEPD media, arrested in G1 phase using alpha-factor mating pheromone, and released into rich media at 30°C. Cells
were collected over time to monitor the budding index (a marker for cell cycle entry) and to extract protein (Materials and Methods). Protein expression was quantified by PRM across two replicate time series (20 time points per replicate, collected every 7 minutes). For the two biological replicate experiments, QC samples were designed by pooling equal amounts of each of the 20 time points for internal technical replicates. To maximize duty cycle and sensitivity, 2 peptides were selected per targeted protein of interest based on scouting of the pooled QC samples. For each biological replicate, the time series samples were analyzed in a random order and were interspersed with five analyses of the QC pools.

Using the SIL peptides as internal references, we selected interference-free product ions for quantitation of each of the peptides (Supplementary Table 1) and normalized the expression of native proteins to the SIL peptide standards (Materials and Methods). In total, we quantified about 80 peptides corresponding to 48 proteins across each of the replicates. Our ability to quantify peptides from time series samples was greater than from asynchronous cells, presumably due to the high abundance of target proteins at one or more time points across the cell cycle (Supplementary File 1). We further filtered the data to include only peptides that were reproducibly quantified across QC pools and/or between the two biological replicate experiments. In addition, we analyzed expression level noise between peptides using the QC pools and individual samples over the two biological replicates (Supplementary Table 2). Based on these multiple factors, we assigned the highest confidence to the quantitation of 44 peptides belonging to 31 unique proteins (excluding the nuclear and cytoplasmic controls, which are not dynamically expressed during the cell cycle).
The replicate wild-type time series experiments were aligned to a common cell-cycle timeline using the CLOCCS algorithm (Orlando et al. 2007, 2009) (Supplementary File 1), and the relative expression of mRNA and protein was visualized across the 31 cell-cycle genes. Qualitatively, we found that many proteins follow their respective RNA dynamics with delay (Figure 3.10). Our results support previous immunoblotting experiments for 14 individual cell-cycle proteins (Supplementary Table 1) and demonstrate cell-cycle dynamics for 17 proteins for the first time. Interestingly, a cluster of proteins whose mRNA transcripts peak near S phase did not appear to match their periodic mRNA expression pattern (Fhl1, Fkh1, Fkh2, Msn2, Swi6, Mbp1, and Ixr1).
Figure 3.10: Proteome dynamics of cell-cycle regulators follow the transcriptome with delay during the cell cycle in *S. cerevisiae*.

Wild-type budding yeast cells were grown in 2% YEPD rich media, synchronized by alpha-factor mating pheromone, released into YEPD, and monitored over about 2 cell cycles. Samples were collected every 5 minutes for RNA-sequencing (Kelliher et al. 2016) (A) or every 7 minutes for total protein extraction (B-C). Population synchrony was monitored by counting at least 200 cells per time point for the presence or absence of a bud (bottom). 31 high confidence proteins (44 total peptides) are shown, with multiple high confidence peptides per protein for: Cdc28, Swi4, Cln2, Tos4, Hcm1, Plm2, Nrm1, Ndd1, Fhl1, Swi5, Ace2, and Sic1. Genes and proteins were ordered on the y-axis by peak time of mRNA expression (A). Transcript and protein levels are depicted as a z-score change relative to expression mean in the respective dataset, where values represent the number of standard deviations away from the mean. Each column represents a lifeline point on a common cell-cycle timeline determined by the CLOCCS algorithm.
At cell-cycle Start in *S. cerevisiae*, approximately 200 periodic genes are activated by the heterodimeric TFs SBF and MBF (U. Eser et al. 2011; V. R. Iyer et al. 2001). Intriguingly, the components of these TF activator complexes (SBF: Swi4 and Swi6; MBF: Mbp1 and Swi6) did not appear to be periodically expressed at the protein level despite periodic mRNA expression (Figure 3.10). SBF and MBF are thought to be inactivated during S phase by repressor TFs (Yhp1, Yox1, and Nrm1) and later in time by Clb2/CDK phosphorylation (Amon et al. 1993; de Bruin et al. 2006; Pramila et al. 2002). Two key negative regulators of SBF and MBF, Clb2 and Yhp1, were not detected with high confidence in the wild-type PRM assay, and we performed time series immunoblotting experiments in triplicate to examine their dynamics (Figure 3.11). Additionally, we compared the protein expression dynamics from PRM to immunoblotting results for three TFs: Swi4, Nrm1, and Yox1 (Figure 3.11). Taken together, it appeared that repressor proteins were more dynamically expressed than activator TFs at G1/S phase (Figure 3.10, Figure 3.11); however, we sought a method to more quantitatively compare RNA and protein curve shapes over the yeast cell cycle.
Wild-type cells expressing NRM1-HA3 (A), SWI4-13MYC (B), YOX1-13MYC (C), CLB2-HA (D), or YHP1-13MYC (E) were grown in 2% YEPD media, synchronized by alpha-factor mating pheromone, released into YEPD, and monitored over about 2 cell cycles. Samples were collected every 7 minutes for total protein extraction. Protein immunoblots were normalized to Cdc28/Pho85 (PSTAIR; constitutive levels over the cell cycle) with ImageJ. One representative Western Blot is shown for each triplicate set of experiments.

To assess reproducibility between PRM and immunoblotting, three replicates of Western Blot data were compared to targeted mass spectrometry peptide data for NRM1_1 (A), SWI4_1 (B), and YOX1_1 (C). All expression datasets were aligned on a common cell-cycle timeline using CLOCCS. mRNA expression, peptide heavy : light ratios, and Western Blot data were scaled to maximum expression for each gene or protein ([0, 100] linear scale). Line plots for mRNA expression (black, dashed), replicate peptide expression (replicate 1 in green, replicate 2 in blue), and three replicate Western Blots (replicates 1-3 in gold) are shown.

To determine a quantitative relationship between mRNA and protein expression curves, we repurposed the JTK_CYCLE periodicity-ranking algorithm. JTK_CYCLE (JTK) was developed in the circadian rhythm field for identifying periodic genes from
whole-transcriptome time series data (Hughes et al. 2010). The existing JTK algorithm uses a cosine curve as a reference signal and compares it to the time series data of interest to score periodic expression (Materials and Methods). We modified JTK to take two time series and compare them to each other. Since we were interested in scoring different time delays between mRNA and protein expression, we further modified the algorithm to compare only the two times series profiles without any phase shifting. Lastly, the original algorithm scores both highly positive and negative correlations to a cosine shape as periodic (using an absolute function on the correlation measure when computing the JTK statistic). In order to preserve the difference between positive and negative correlations, we removed the absolute function from JTK. The modified JTK algorithm gives the smallest scores to shapes that are the most similar (positive correlation between RNA and protein), middle scores to shapes that are not similar, and the largest scores to shapes that are reflections on the x-axis (negative or anti-correlation between RNA and protein).

Using this modified JTK method, we found that 23 proteins (31 peptides, or 70.5%) were positively correlated with their respective mRNA expression during the cell cycle (Table 3.2; Figure 3.12 A, red and green dots). Some proteins—Cdc28, Fhl1, Fkh2, Msn2, Swi4, Swi6, and Mbp1—had a poor correlation score (Figure 3.12 A, black dots) because stabilized proteins did not correlate well with periodic mRNA expression, as we observed above (Figure 3.10). We hypothesized that the remaining poorly correlated RNA-protein pairs could indicate post-translational modifications acting specifically on protein expression. Therefore, we mined the literature for previously reported and predicted protein degradation mechanisms of cell-cycle regulators of interest (Table 3.3). We observed that proteins targeted for destruction at specific times
during the cell cycle, such as Cin8, Clb5, and Sic1, had poor RNA-protein correlation scores in cycling wild-type cells (Figure 3.12 A).

**Table 3.2: JTK correlation scores for the 44 S. cerevisiae peptides compared to their respective mRNA transcripts from Figure 3.10.**

In order to place transcriptome and proteome data on an identical time scale, experiments were aligned to a cell-cycle timeline with the CLOCCS algorithm. Then, data points were interpolated to 30 samples along the cell-cycle timeline interval from 50-250 lifeline points (i.e. all time series were sampled in silico approximately every 6.9 points). Interpolation calculations were done in the R Statistical Programming Environment using the function approxfun (arguments: method="linear", rule=2) (R Core Development Team 2017). Any negative values resulting from the interpolation were set to zero. Interpolated datasets were then run through the modified JTK correlation algorithm. Peptide-RNA correlations with p-values less than 0.01 are shown in red font. Peptides that are not significantly correlated with mRNA include the more constitutively expressed cell-cycle proteins (e.g. Cdc28, Fhl1, Fkh2, Gat1, Ixr1, Mbp1, Msn2, Swi4, and Swi6) and canonical proteasome targets (e.g. Cin8, Clb5, Pds1, and Sic1). Clb2 and Yhp1 measurements from Western Blot experiments are also included from Figure 3.11.

<table>
<thead>
<tr>
<th>RNA ID</th>
<th>peptide ID</th>
<th>wild-type rep1 (adj. p-value)</th>
<th>wild-type rep2 (adj. p-value)</th>
<th>clb1-6 rep1 (adj. p-value)</th>
<th>clb1-6 rep2 (adj. p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE2</td>
<td>ACE2_1</td>
<td>7.549E-04</td>
<td>8.125E-04</td>
<td>0.0126</td>
<td>1.9985</td>
</tr>
<tr>
<td>ACE2</td>
<td>ACE2_2</td>
<td>0.2008</td>
<td>4.875E-04</td>
<td>1.888E-07</td>
<td>0.0099</td>
</tr>
<tr>
<td>ASH1</td>
<td>ASH1_2</td>
<td>2.184E-10</td>
<td>3.271E-06</td>
<td>0.4446</td>
<td>0.0720</td>
</tr>
<tr>
<td>CDC20</td>
<td>CDC20_1</td>
<td>1.649E-05</td>
<td>6.528E-05</td>
<td>0.1014</td>
<td>0.0658</td>
</tr>
<tr>
<td>CDC28</td>
<td>CDC28_1</td>
<td>0.2272</td>
<td>0.3384</td>
<td>4.345E-07</td>
<td>1.191E-08</td>
</tr>
<tr>
<td>CDC28</td>
<td>CDC28_2</td>
<td>0.1884</td>
<td>0.1170</td>
<td>8.461E-06</td>
<td>2.283E-08</td>
</tr>
<tr>
<td>CIN8</td>
<td>CIN8_1</td>
<td>0.3755</td>
<td>0.2272</td>
<td>1.9823</td>
<td>1.6118</td>
</tr>
<tr>
<td>CLB2</td>
<td>CLB2_1_wb</td>
<td>3.352E-07</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>CLB2</td>
<td>CLB2_2_wb</td>
<td>4.379E-10</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>CLB2</td>
<td>CLB2_3_wb</td>
<td>9.209E-05</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>CLB5</td>
<td>CLB5_1</td>
<td>0.0534</td>
<td>0.7502</td>
<td>0.2463</td>
<td>0.6385</td>
</tr>
<tr>
<td>CLN2</td>
<td>CLN2_1</td>
<td>3.352E-07</td>
<td>3.486E-05</td>
<td>1.7020</td>
<td>0.2010</td>
</tr>
<tr>
<td>CLN2</td>
<td>CLN2_2</td>
<td>3.352E-07</td>
<td>0.0431</td>
<td>0.0786</td>
<td>1.7616</td>
</tr>
<tr>
<td>FHL1</td>
<td>FHL1_1</td>
<td>0.1825</td>
<td>0.0631</td>
<td>1.4300</td>
<td>0.0158</td>
</tr>
<tr>
<td>FHL1</td>
<td>FHL1_2</td>
<td>0.2272</td>
<td>0.4899</td>
<td>1.4300</td>
<td>0.7246</td>
</tr>
<tr>
<td>FKH1</td>
<td>FKH1_1</td>
<td>0.0450</td>
<td>0.0025</td>
<td>1.3190</td>
<td>0.4928</td>
</tr>
<tr>
<td>FKH2</td>
<td>FKH2_1</td>
<td>0.3384</td>
<td>0.0685</td>
<td>0.9532</td>
<td>1.2008</td>
</tr>
<tr>
<td>GAT1</td>
<td>GAT1_1</td>
<td>0.0491</td>
<td>0.6203</td>
<td>0.7100</td>
<td>0.3564</td>
</tr>
<tr>
<td>HCM1</td>
<td>HCM1_1</td>
<td>4.268E-08</td>
<td>1.118E-05</td>
<td>0.2800</td>
<td>1.0468</td>
</tr>
<tr>
<td>HCM1</td>
<td>HCM1_2</td>
<td>1.590E-07</td>
<td>2.203E-05</td>
<td>0.2890</td>
<td>1.3190</td>
</tr>
<tr>
<td>IXR1</td>
<td>IXR1_2</td>
<td>0.3567</td>
<td>0.0450</td>
<td>0.1100</td>
<td>0.1874</td>
</tr>
<tr>
<td>MBP1</td>
<td>MBP1_1</td>
<td>1.9740</td>
<td>0.0534</td>
<td>1.4030</td>
<td>1.9280</td>
</tr>
<tr>
<td>Gene</td>
<td>Sample 1</td>
<td>Sample 2</td>
<td>Sample 3</td>
<td>Sample 4</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>MCM1</td>
<td>0.0033</td>
<td>0.0010</td>
<td>0.1874</td>
<td>0.0372</td>
<td></td>
</tr>
<tr>
<td>MCM1</td>
<td>1.7288</td>
<td>1.5430</td>
<td>1.9977</td>
<td>1.9999</td>
<td></td>
</tr>
<tr>
<td>MCM1</td>
<td>0.0116</td>
<td>5.961E-05</td>
<td>0.0099</td>
<td>8.340E-05</td>
<td></td>
</tr>
<tr>
<td>NDD1</td>
<td>0.0053</td>
<td>0.0123</td>
<td>0.3564</td>
<td>0.0374</td>
<td></td>
</tr>
<tr>
<td>NDD1</td>
<td>0.0021</td>
<td>0.0015</td>
<td>0.0336</td>
<td>0.5437</td>
<td></td>
</tr>
<tr>
<td>NRM1</td>
<td>4.996E-05</td>
<td>1.642E-04</td>
<td>0.0658</td>
<td>0.0336</td>
<td></td>
</tr>
<tr>
<td>PCL1</td>
<td>1.7288</td>
<td>1.5430</td>
<td>1.9977</td>
<td>1.9999</td>
<td></td>
</tr>
<tr>
<td>PCL2</td>
<td>8.702E-04</td>
<td>8.428E-05</td>
<td>1.888E-07</td>
<td>2.508E-04</td>
<td></td>
</tr>
<tr>
<td>PDS1</td>
<td>0.0314</td>
<td>0.1213</td>
<td>1.9888</td>
<td>1.9842</td>
<td></td>
</tr>
<tr>
<td>PHD1</td>
<td>9.989E-05</td>
<td>0.0020</td>
<td>0.0168</td>
<td>0.0973</td>
<td></td>
</tr>
<tr>
<td>PLM1</td>
<td>1.421E-09</td>
<td>3.451E-11</td>
<td>1.6436</td>
<td>0.8911</td>
<td></td>
</tr>
<tr>
<td>PLM2</td>
<td>1.590E-07</td>
<td>1.649E-05</td>
<td>1.6638</td>
<td>1.1089</td>
<td></td>
</tr>
<tr>
<td>SFG1</td>
<td>1.576E-11</td>
<td>6.149E-10</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>SIC1</td>
<td>0.2137</td>
<td>0.2272</td>
<td>2.0000</td>
<td>2.0000</td>
<td></td>
</tr>
<tr>
<td>SIC1</td>
<td>1.1680</td>
<td>0.3036</td>
<td>1.8808</td>
<td>2.0000</td>
<td></td>
</tr>
<tr>
<td>SWI4</td>
<td>0.1010</td>
<td>1.5949</td>
<td>1.5785</td>
<td>1.5785</td>
<td></td>
</tr>
<tr>
<td>SWI4</td>
<td>0.3384</td>
<td>1.6433</td>
<td>0.9221</td>
<td>1.4030</td>
<td></td>
</tr>
<tr>
<td>SWI4</td>
<td>0.2560</td>
<td>1.6433</td>
<td>0.5179</td>
<td>0.1621</td>
<td></td>
</tr>
<tr>
<td>SWI5</td>
<td>1.231E-07</td>
<td>1.226E-06</td>
<td>0.0158</td>
<td>0.0973</td>
<td></td>
</tr>
<tr>
<td>SWI5</td>
<td>0.0014</td>
<td>1.394E-04</td>
<td>0.4103</td>
<td>0.1014</td>
<td></td>
</tr>
<tr>
<td>SWI6</td>
<td>0.2342</td>
<td>0.1257</td>
<td>0.0857</td>
<td>0.2463</td>
<td></td>
</tr>
<tr>
<td>TOS4</td>
<td>4.268E-08</td>
<td>7.501E-06</td>
<td>0.0933</td>
<td>0.8602</td>
<td></td>
</tr>
<tr>
<td>TOS4</td>
<td>4.379E-10</td>
<td>5.550E-06</td>
<td>0.0112</td>
<td>0.3774</td>
<td></td>
</tr>
<tr>
<td>YHP1</td>
<td>2.334E-07</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>YHP1</td>
<td>2.184E-10</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>YHP1</td>
<td>0.0215</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>YOX1</td>
<td>1.705E-06</td>
<td>2.263E-04</td>
<td>0.2463</td>
<td>0.9844</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.12: Gene expression for many RNA-protein pairs is positively correlated in wild-type cells, including some proteins targeted for degradation, while mutant gene expression pairs are less correlated.

In order to place transcriptome (31 genes, from Figure 3.10) and proteome (44 peptides) data on an identical time scale, time series experiments were aligned to a cell-cycle timeline using the CLOCCS algorithm and interpolated as described in the Table 3.2 legend. Interpolated datasets were then run through the modified JTK correlation algorithm. The log-distribution of RNA-peptide correlation scores is shown between the two biological replicates for cycling wild-type cells (A). RNA-protein pairs with positive correlation scores are shown as red or green dots (red = less than 0.01 in both replicates; dark red = less than 0.01 in only one biological replicate; green = less than
RNA-protein pairs with poor to anti-correlation scores (greater than 0.05 in both biological replicates) are shown as black dots. A subset of E3 ubiquitin ligase targets had poor RNA-protein correlation scores in wild-type cells. The degradation kinetics of select E3 ubiquitin ligase targets are plotted after peak mRNA expression during the cell cycle (B). Three verified SCF targets (Cln2, Hcm1, and Yox1), three APC/C targets (Cin8, Nrm1, and Pds1), and two proteins targeted by both SCF and APC/C (Ndd1 and Tos4) were shown from Table 3.3 (B). The maximum RNA expression value of each gene was found, and normalized protein expression values were plotted for each protein after peak mRNA expression (SCF in green: CLN2_1, HCM1_2, YOX1_1; APC/C in purple: CIN8_1, NRM1_1, PDS1_2; Both in blue: NDD1_1, TOS4_1). The log-distribution of peptide-RNA correlation scores from the modified JTK correlation algorithm is shown for biological replicates of mutant clb1-6 cells (C). Three SCF targets (Cln2, Hcm1, and Yox1), three APC/C targets (Cin8, Nrm1, and Pds1), and two proteins targeted by both SCF and APC/C (Ndd1 and Tos4) were plotted from clb1-6 data as described above for wild-type cells (D).

**Table 3.3: Previously annotated and predicted protein degradation mechanisms for cell-cycle proteins of interest.**

Degradation pathways targeting specific proteins of interest are shown with literature references. The amino acid sequences of all proteins were also tested for various protein instability metrics: N-end rule (Bachmair et al. 1986), PEST sequence predictor (Rogers et al. 1986), F-box predictor from the SMART database (Letunic et al. 2015), and APC/C sequence predictor from the GPS-ARM tool (Z. Liu et al. 2012). The four proteins used as controls for constitutive expression during the cell cycle (Rim11, Taf12, Vps9, and Cic1) are not shown, as they are not dynamically expressed. Proteins annotated as “unknown” did not match any bioinformatics sequence predictors and did not have experimental evidence for targeted protein destruction.

<table>
<thead>
<tr>
<th>Targeted protein</th>
<th>Putative Degradation Mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE2</td>
<td>N-end rule (D)</td>
</tr>
<tr>
<td>ASE1</td>
<td>APC target (2 KEN-box) (Juang et al. 1997)</td>
</tr>
<tr>
<td>ASH1</td>
<td>SCF-CDC4 target (Q. Liu et al. 2011)</td>
</tr>
<tr>
<td>BCK2</td>
<td>unknown</td>
</tr>
<tr>
<td>CDC20</td>
<td>APC target (2 D, 1 KEN-box) (Foe et al. 2011)</td>
</tr>
<tr>
<td>CDC28</td>
<td>unknown</td>
</tr>
<tr>
<td>CDH1</td>
<td>unknown</td>
</tr>
<tr>
<td>CIN8</td>
<td>APC target (1 KEN-box) (Chee and Haase 2010; Hildebrandt and Hoyt 2001)</td>
</tr>
<tr>
<td>CLB2</td>
<td>APC target (1 D, 1 KEN-box) (Lim et al. 1998; Seufert et al. 1995; Wäsch and Cross 2002)</td>
</tr>
<tr>
<td>CLB5</td>
<td>APC target (1 D-box) (Seufert et al. 1995; Shirayama et al. 1999)</td>
</tr>
<tr>
<td>CLB6</td>
<td>SCF-CDC4 target (Jackson et al. 2006)</td>
</tr>
<tr>
<td>CLN2</td>
<td>SCF-GRR1 target; PEST sequence (1) (Barral et al. 1995; Lanker et al. 1996)</td>
</tr>
<tr>
<td>CLN3</td>
<td>PEST sequence (2 potential) (Tyers et al. 1992; Yaglom et al. 1996)</td>
</tr>
</tbody>
</table>
To further investigate targeted protein destruction mechanisms, we selected verified SCF targets (Cln2, Hcm1, and Yox1), APC/C targets (Cin8, Nrm1, and Pds1), and both SCF and APC/C targets (Ndd1 and Tos4) from Table 3.3. We found the peak mRNA expression value for each gene and then plotted normalized protein expression kinetics after the mRNA peak for SCF (green) versus APC/C (purple) targets (Figure 60).
3.12 B). These results suggested that SCF and APC/C targets have different degradation kinetics during the cell cycle. For the two proteins with previous literature supporting both SCF and APC/C targeting, Tos4 protein dynamics more closely matched verified SCF targets, while Ndd1 dynamics mirrored APC/C targets.

We visualized SCF and APC/C proteasome targets over the cell cycle to determine if E3 ubiquitin ligase targets shared protein expression shape characteristics (Figure 3.13). Many SCF target proteins appeared to be highly unstable, where proteins were not expressed in the absence of the respective mRNA precursor (Figure 3.13 A-B). Peak expression of the APC/C co-factor Cdc20 (Figure 3.13 E, dark gray lines) timed the destruction of many canonical APC/C targets in G2/M phase (Figure 3.13 G-H). Unexpectedly, the accumulation of some APC/C target proteins also appeared to be regulated on the “front side” of the protein curves (Figure 3.13 F-H, light gray lines). Based on previous work, we predict that APC/C-Cdh1 activity can inhibit the accumulation of some APC/C targets until S phase (Huang et al. 2001; Yeong et al. 2001).
Figure 3.13: E3 ubiquitin ligase targets have landmark curve features in cycling wild-type cells, some of which are altered in clb1-6 mutant cells where much of the targeted destruction machinery is no longer active or periodic.

RNA expression and peptide heavy : light ratios were scaled to the maximum value for each gene or protein ([0, 100] linear scale). Line plots for mRNA expression (black, dashed) and biological replicates of wild-type peptide expression (replicate 1 in green, replicate 2 in blue) are shown for canonical SCF targets: CLN2_1 (A), HCM1_1 (B), SIC1_1 (C), and SWI5_3 (D) or APC/C targets: CDC20_1 (E), CLB5_1 (F), PDS1_2 (G), and NDD1_1 (H). When multiple peptides per protein were detected (Figure 3.10), the peptide with lower noise levels was selected (Supplementary Table 2). Each RNA-
protein line graph (A-H) contains a light gray line marking the peak protein expression level for Cln2p (A) as a proxy for APC/C-Cdh1 activity shut-off and a dark gray line at the peak protein expression level for Cdc20p as a proxy for peak APC/C-Cdc20 activity (E). Line plots for mRNA expression (black, dashed) and biological replicates of clb1-6 mutant peptide expression (replicate 1 in red, replicate 2 in purple) are shown for canonical SCF targets: HCM1_2 (I), SIC1_1 (J) or APC/C targets: PDS1_2 (K), and NDD1_1 (L). When multiple peptides per protein were detected, the peptide with lower noise levels was selected (Supplementary Table 2). In the cell-cycle timeline for clb1-6 cells, S and G2/M phases are grayed to indicate that B-cyclin mutant cells are physically arrested at the G1/S border.

To supplement our qualitative inspection of protein destruction (Figure 3.12 B, Figure 3.13), we also calculated delays between RNA and protein expression dynamics for all peptides shown in Figure 3.10 (Supplementary Table 3). We investigated three timing metrics between mRNA and protein curves: time delay to 50% peak expression (“front side” of curves), peak time delay (“peak” of curves), and time delay to 50% trough expression (“back side” of curves). For the positively correlated RNA-protein pairs (Figure 3.12, red dots), the average of all three mRNA-to-protein delay time metrics was 14.4 +/- 16.6 cell-cycle timeline points (approximately 10.1 +/- 11.7 minutes for a 70.4 minute cell cycle). We observed high variability in delay times for different RNA-protein pairs, which is reflected in the large standard deviation among the most positively correlated RNA-protein pairs. When grouped specifically by putative proteasome regulation, SCF targets have a shorter average mRNA-to-protein delay time of 15.3 +/- 15.1 cell-cycle timeline points (approximately 10.8 +/- 10.7 minutes; SIC1_1 and SIC1_2 were removed from this average calculation because protein peaked before mRNA) compared to 18.9 +/- 11.6 cell-cycle points (approximately 13.3 +/- 8.1 minutes) for APC/C targets (Supplementary Table 3).
3.3.3 In the absence of periodic cyclin/CDK and ubiquitin ligase activities, some regulatory proteins continue to be dynamically expressed

To query the importance of periodic protein destruction for regulating cell-cycle protein expression, we profiled clb1-6 cells where cell-cycle progression is physically halted, but many cell-cycle genes continue to be periodically expressed (Haase and Reed 1999; Orlando et al. 2008). GAL-CLB1 clb1-6 cells were cultured in YEPG media, arrested in G1 phase using alpha-factor mating pheromone, supplemented with dextrose to inhibit CLB1 expression, and then released into YEPD media at 30°C. Cells were collected over time to monitor the re-budding index, isolate mRNA, or extract protein (Materials and Methods). Replicate time series experiments were aligned to a common cell-cycle timeline using the CLOCCS algorithm (Orlando et al. 2007, 2009) (Supplementary File 1).

Cells lacking B-type cyclin activity are physically arrested at the G1/S border and cannot initiate DNA replication (Haase and Reed 1999). Cell-cycle proteins that are phosphorylated by B-cyclin/CDKs and targeted for degradation by SCF should therefore be stabilized. Additionally, APC/C-Cdh1 should be active only at the beginning of the time series (as we observed in wild-type cells), followed by inactivation by Cln/CDKs and other kinases (Hall et al. 2004; Zachariae et al. 1998). Later in time, APC/C-Cdc20 should not become activated because functional Cdc20 requires B-cyclin/CDK phosphorylation (Rahal and Amon 2008; Rudner and Murray 2000). Therefore, clb1-6 mutant protein expression dynamics should be largely dependent on mRNA dynamics and protein half life.

In general, cell-cycle regulatory proteins were less correlated with their respective mRNA precursors in clb1-6 mutant cells compared to wild type, with only 15
positively correlated RNA-peptide pairs representing 12 unique proteins (Figure 3.12 A, C). This result was likely not due to noise in protein expression data, as the magnitude of noise values was similar between the wild-type and mutant experiments (Supplementary Table 2). A subset of core cell-cycle TFs, including Tos4, Nrm1, Mcm1, Ndd1, Ace2, and Swi5, maintained a positive correlation score between mRNA and protein expression during the mutant cycle (Figure 3.12 C, red and green dots). We hypothesized that lack of periodic protein destruction could explain the decreased correlation between some RNA-protein pairs in mutant cells compared to wild type.

Both APC/C and SCF targets had variable degradation kinetics after peak mRNA expression in mutant cells compared to wild type (Figure 3.12 B, D). We hypothesized that SCF targets requiring Clb/CDK phosphorylation would be stabilized and that SCF targets degraded after Cln/CDK phosphorylation would be highly unstable in mutant cells (Figure 3.12 D, green lines). We found that Sic1 showed one, early peak of protein expression in clb1-6 cells, which suggests persistent Cln/CDK phosphorylation and SCF degradation later in the time series (Figure 3.13 J). On the other hand, Hcm1 was not turned over to low levels in clb1-6 cells (Figure 3.13 I), which agrees with previous data that Hcm1 is targeted by SCF after B-type cyclin/CDK phosphorylation (Landry et al. 2014). Many canonical APC/C targets matched the predicted dynamics, where proteins accumulated after APC/C-Cdh1 removal and were not destroyed in later in time because Cdc20 is inactive in clb1-6 mutants (Figure 3.12 purple lines, Figure 3.13 K-L). In the absence of oscillatory ubiquitin ligase activity, we also wanted to ask if proteins changed in absolute expression level in clb1-6 compared to wild type. In our PRM assay, normalized values for each peptide were calculated relative to a constant amount of heavy labeled SIL peptides (Materials and Methods). Therefore, peptide ratios are
directly comparable between wild-type and mutant experiments. Two APC/C-Cdc20 targets, Ndd1 and Pds1, showed more than 2-fold increase specifically in protein accumulation in clb1-6 cells. The average expression levels of the NDD1_1 and PDS1_2 peptides in wild-type replicates were 0.00682 and 0.00653, respectively, which increased to 0.0374 and 0.0306 in mutant cells (Supplementary Tables 4 and 5).

3.3.4 Repressor TFs are more periodically expressed at the protein level than G1/S activator TFs in both wild-type and mutant cells

After commitment to the cell cycle, about 200 genes are periodically activated at G1/S phase by the TF complexes SBF and MBF in S. cerevisiae (U. Eser et al. 2011; V. R. Iyer et al. 2001). These TFs are functionally analogous to the E2F family of TFs in mammalian cells (reviewed in: (Bertoli et al. 2013)). The subunits of SBF and MBF (Swi4, Swi6, and Mbp1) are periodically transcribed during the cell cycle (Spellman et al. 1998). The program of G1/S genes is periodically activated in both wild-type and clb1-6 mutant cells, suggesting that the activity of SBF and MBF is periodic in both conditions (Orlando et al. 2008).

We compared the mRNA and protein expression dynamics of G1/S activator (Swi4) and repressor TFs (Yox1, Yhp1, Nrm1) in wild-type and clb1-6 mutant cells (Figure 3.14). As we observed previously for wild-type cells (Figure 3.10, Figure 3.11), the repressor TFs were more dynamically expressed at the protein level in clb1-6 mutant cells (Figure 3.14 E-H). We posit that the abundance of inhibiting co-factors of SBF and MBF control periodicity in their activator activity, rather than periodic abundance of the activators alone. The subunits of SBF and MBF are clearly periodic at the mRNA level (Figure 3.10). We predict that nascent pools of unmodified SBF and MBF proteins are required each cell cycle from periodic mRNA synthesis, while the total protein
abundance of SBF and MBF remains constitutive over the cell cycle. Alternatively, the localization of Swi4, Swi6, and/or Mbp1 could be altered without changing the total protein abundance.

Figure 3.14: Gene expression dynamics of G1/S TFs in wild-type and clb1-6 mutant cells reveal that repressor proteins are more periodically expressed than activators.
RNA and protein expression datasets were aligned on a common cell-cycle timeline using CLOCCS. Line plots for mRNA expression (dashed) and replicates of peptide expression (wild-type replicate 1 in green, replicate 2 in blue; clb1-6 replicate 1 in red, replicate 2 in purple) are shown for Swi4 (A, E), Yox1 (B, F), Yhp1 (C, G), and Nrm1 (D, H). mRNA expression values (fpkm units) on the y-axes were normalized together for the two datasets, and wild-type expression dynamics are shown in blue and clb1-6 expression in purple (A-D). Peptide expression values (heavy : light ratios) on the y-axes are comparable between the time series datasets because a constant amount of SIL peptides was used in the experiments (E-H). Wild-type protein expression levels of Yhp1-13MYC were taken from Figure 3.11. Two replicates of immunoblotting data were scaled relative to the maximum value for each experiment ([0, 100] linear scale) followed by scaling the maximum value to match the YHP1_5 PRM data from clb1-6 cells ([0, 0.02] linear scale, arbitrary units) (G).

3.4 Discussion

During the cell cycle, an interconnected network of transcription factors, cyclin/CDK complexes, and E3 ubiquitin ligase machinery regulates the order of events and periodic genes (Figure 3.15). Here, we have provided further evidence for the interconnected regulatory relationships between periodic E3 ubiquitin ligase activity and cell-cycle regulatory proteins (Figure 3.12, Figure 3.13). Multiple studies have proposed that a TF network regulates periodic gene expression during the S. cerevisiae cell cycle (Hillenbrand et al. 2016; Orlando et al. 2008; Sevim et al. 2010; Simmons Kovacs et al. 2012). Quantitative modeling evidence for the function of this TF network relies on the assumption that mRNA expression is a reliable proxy for cell-cycle protein expression. Here we applied a targeted mass spectrometry approach to demonstrate that, generally, dynamic mRNAs produced dynamic protein expression during the cell cycle (Figure 3.10). We also profiled RNA and protein expression from clb1-6 mutant cells and demonstrated that some cell-cycle regulators continued to correlate with mRNA expression independently of B-type cyclin/CDK activity (Figure 3.12), allowing for some periodic transcription to persist in clb1-6 cells (Haase and Reed 1999; Orlando et al. 2008). Other cell-cycle regulator proteins depended on periodic post-transcriptional
regulatory mechanisms and were not expressed consistently with their mRNA expression pattern in mutant cells (Figure 3.13). Taken together, our results show that redundant layers of regulation control the expression of cell-cycle proteins during the yeast cell cycle.

Figure 3.15: An integrated cell-cycle network includes windows of targeted E3 ubiquitin ligase activity.

We posit that cell-cycle ordering is maintained by cyclin/CDK activity, an interconnected network of transcription factors (TFN), and E3 ubiquitin ligase activity. Approximate windows of peak degradation machinery are shown based on wild-type data (Figure 3.13). Periodic TFs (activators in green, repressors in red) are placed on the cell-cycle timeline approximately by peak mRNA expression. In blue, G1 (CLN) and B-type (CLB) cyclin/CDKs, APC/C, Cdc14, and Sic1 regulate each other and TFs in the network. Edges from TFs represent evidence for transcriptional regulation: ChIP-chip data for TF binding and/or genetic evidence for regulation type (compiled in: (McGoff et al. 2016; Orlando et al. 2008)). Edges between regulatory proteins and TFs represent protein-level modifications (e.g. phosphorylation or ubiquitination). Pointed arrows indicate activation, and blunted arrows mark repression or protein degradation.
We demonstrated that there is not one single delay time for RNA-to-protein synthesis for the cell-cycle regulators of interest. Along with previous work, this finding has implications for quantitative models where biochemical rate parameters are assumed to be equal across network components (Ball et al. 2013). We estimated that the transcription-translation delay time for the average dynamic yeast protein is about 10 minutes with a large standard deviation. Proteins targeted for degradation by the SCF complex had short durations of protein expression compared to mRNA (Supplementary Table 3), while the protein accumulation of canonical APC/C targets was gated by both Cdh1 and Cdc20 activity (Figure 3.13). Some cell-cycle regulators are turned over by multiple mechanisms (Table 3.3), which indicates redundancy in cell-cycle control. We posit that redundant destruction mechanisms buffer the cell-cycle regulatory network against ectopic expression of regulators in the incorrect cell-cycle phase. This redundancy model could be tested by synchronizing yeast cells, constitutively expressing a cell-cycle TF(s), and querying the degree to which downstream target gene expression is altered.

Curiously, we observed that some proteins did not damp in expression level as much as the respective RNA in the second cell cycle (Figure 3.13). After synchronization, populations of budding yeast cells are expected to damp in gene expression levels over time due to asymmetric cell division between mother and daughter cells (Guo et al. 2013). This protein expression “anti-damping” phenomenon is likely not due to data normalization, as immunoblotting results (where proteins of interest were normalized to a constitutive band) also showed evidence for lack of protein expression damping (Figure 3.11). This anti-damping observation could be explained by
previous findings that translation may be more efficient in the second cell cycle after release from alpha-factor synchronization (Goranov et al. 2009; Serikawa et al. 2003).

Currently, DNA or RNA genomic technologies are used more frequently than mass spectrometry technologies to assay gene expression. Therefore, we do not have a complete understanding of the noise models and/or all sources of variance in proteomic datasets. We took advantage of both time series sampling and biological replicate experiments to quantify noise in our PRM datasets (Supplementary Table 2). We dealt with the technical noise in the mass spectrometry data by attempting to target multiple peptides per protein and by quantifying on multiple transitions per peptide (Supplementary Table 1). Future experiments will further our understanding of thresholds, sources of bias, and limits of detection in targeted mass spectrometry.

The current model of the budding yeast cell-cycle TFN is that unstable TF proteins sequentially activate a large program of gene expression each cell cycle. The heavy peptide SIL standards used for PRM targeted mass spectrometry in this Chapter can be re-used for many future experiments. Future work could query proteasome mutants (SCF, APC/C, proteasome inhibitor treatment, etc.) to ask about stabilization of cell-cycle proteins and the resulting gene expression dynamics. The abundance of cell-cycle regulators is only part of the complete mechanism—localization, TF complex formation, and affinity for DNA binding sites require further exploration to fully characterize the yeast cell-cycle network.

3.5 Future Work

Our current model of the budding yeast cell-cycle TFN is that unstable TF proteins sequentially activate a large program of gene expression each cycle. Remaining questions include: Are cell-cycle proteins particularly unstable compared to
housekeeping proteins and/or proteins involved in other dynamic processes? Does ordering matter—if TFs are ectopically expressed at the wrong time during the cycle, can the periodic gene expression program be improperly re-started?

The heavy peptide standards applied in this Chapter for targeted mass spectrometry can be re-used for many future experiments, which could profile E3 ubiquitin ligase mutants (SCF, APC/C) or proteasome inhibitor treatment to query the effects of stabilizing cell-cycle proteins and the resulting gene expression dynamics. In addition to sampling for total protein over time, future work could enrich for nuclear extracts to add the dimension of TF localization over the cell cycle to our analyses (Mirzaei et al. 2013). Some TF network components are already known to move in and out of the nucleus at specific cell-cycle phases (Huh et al. 2003). Finally, expression and localization profiling could be combined with phospho-enrichment mass spectrometry to characterize precise post-transcriptional modifications of TFs in the network.

3.6 Materials and Methods

3.6.1 Yeast strains, cultures, and synchronization

Saccharomyces cerevisiae strains are derivatives of BF264-15D MATa bar1. Strains were constructed using standard yeast methods. Yeast cultures were grown in standard YEP media (1% yeast extract, 2% peptone, 0.012% adenine, 0.006% uracil, and 2% dextrose or galactose sugar). Strain genotypes from this Chapter are summarized in the Table below.
Table 3.4: *S. cerevisiae* yeast strains used in Chapter 3.

<table>
<thead>
<tr>
<th>Strain ID</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF264-15Dau</td>
<td>MATa ade1 his2 leu2-3,112 trp1-1a ura3Δns</td>
<td>(Reed et al. 1985)</td>
</tr>
<tr>
<td>SBY151</td>
<td>MATa bar1</td>
<td>(Cole and Reed 1991)</td>
</tr>
<tr>
<td>SBY661</td>
<td>MATa arg4 clb1::URA3 clb2::LEU2 clb3::TRP1 clb4::HIS2 clb5 clb6::ADE1 P&lt;sub&gt;GAL1&lt;/sub&gt;-CLB1::LEU2</td>
<td>(Haase and Reed 1999)</td>
</tr>
<tr>
<td>SBY934</td>
<td>MATa bar1::hphMX4 arg4 clb1::URA3 clb2::LEU2 clb3::TRP1 clb4::HIS2 clb5::ARG4 clb6::ADE1 P&lt;sub&gt;GAL1&lt;/sub&gt;-CLB1::LEU2</td>
<td>(Chee and Haase 2010)</td>
</tr>
<tr>
<td>SBY1205</td>
<td>MATa bar1 YOX1-13MYC-T&lt;sub&gt;ADH1&lt;/sub&gt;::kanMX6</td>
<td>This Chapter</td>
</tr>
<tr>
<td>SBY1258</td>
<td>MATa bar1 CLB2-HA::kanMX6</td>
<td>(Bristow et al. 2014)</td>
</tr>
<tr>
<td>SBY1328</td>
<td>MATa bar1 SWI4-13MYC-T&lt;sub&gt;ADH1&lt;/sub&gt;::kanMX6</td>
<td>This Chapter</td>
</tr>
<tr>
<td>SBY1340</td>
<td>MATa bar1 YHP1-13MYC-T&lt;sub&gt;ADH1&lt;/sub&gt;::kanMX6</td>
<td>This Chapter</td>
</tr>
<tr>
<td>SBY2371</td>
<td>MATa bar1 NRM1-HA3::kanMX6</td>
<td>This Chapter</td>
</tr>
</tbody>
</table>

For alpha-factor synchronization experiments, yeast cells were cultured for 2 days, grown overnight at 30°C to mid-log phase, and on the morning of day 3, arrested using 30 ng/ml alpha-factor for approximately one cell-cycle duration (wild-type cells: 110-115 minutes, *clb1-6* mutant cells: 150-160 minutes). Wild-type cultures were maintained in YEP dextrose media throughout arrest-release. Mutant *P<sub>GAL1</sub>*-CLB1 *clb1-6Δ* cells were cultured in YEP galactose media. With 40-45 minutes remaining in alpha-factor arrest, *clb1-6* cells were treated with 20% dextrose to a final concentration of 2% to inhibit Clb1 expression. Synchronized wild-type and *clb1-6* cultures were washed and resuspended in fresh, pre-warmed YEPD media at 30°C at a concentration of about
1x10^7 cells/ml. Aliquots were taken at each time point and assayed for budding index counts, protein extraction, or RNA extraction.

### 3.6.2 RNA isolation and RNA-Sequencing analyses

Wild-type RNA-Sequencing time series data collection and analysis were described previously (Kelliher et al. 2016). Raw RNA-Sequencing data from wild-type cells can be found at the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE80474. Wild-type data were re-normalized to mutant clb1-6 data for this Chapter.

For clb1-6 cells, total RNA was isolated by acid phenol extraction as described previously (Leman et al. 2014). Samples were sent to the Duke University Sequencing Facility for stranded library preparation. mRNA was amplified and barcoded using KAPA Stranded mRNA-Seq Library Preparation Kits, and reads were sequenced in accordance with standard Illumina HiSeq protocols. Libraries of 50 base-pair single-end reads were prepared, and 12 samples were multiplexed for sequencing together in each lane. A detailed description of the RNA-Sequencing data analysis pipeline has been described previously (Kelliher et al. 2016). RNA-Seq mapping statistics for this study are presented (Supplementary File 1). Normalized RNA-Seq output (fpkm units) was used in the analyses. Raw RNA-Sequencing data for clb1-6 cells described in this Chapter will be available at GEO upon publication.

### 3.6.3 JTK_CYCLE algorithm details

JTK_CYCLE compares a reference signal (by default, a cosine curve) to the time series data of interest and scores periodic shapes. For each time series profile, a pairwise comparison is made for all the values to determine if each value is greater than or less than each other value, which captures the shape of time series while ignoring its
absolute amplitudes. The increasing/decreasing pattern of the time series is then compared to the increasing/decreasing pattern of the reference time series (cosine curve) at phase shifts for each user-specified period length. To determine the statistical significance of the correlation, the Jonckheere-Terpstra test and Kendall’s tau (which counts of discordant pairs between the increasing/decreasing patterns for the two time series) are used. The original JTK-CYCLE implementation in R was provided by the authors (Hughes et al. 2010).

3.6.4 Protein isolation and Western Blotting

Cell samples were collected on filters (Millipore) at each time point and flash frozen in liquid nitrogen. For each set of time series samples, protein extraction was performed in a cold room at 4°C to inhibit yeast proteases and other enzymatic activity. Cell filters were thawed on ice and then washed in 200 μl of cold 1X PBS with 0.01% sodium azide (NaN3, which further inhibited yeast cellular processes). Cells were pelleted, resuspended in 1 ml of cold 10% trichloroacetic acid (TCA), and incubated on ice for 5 minutes. Cells were then resuspended in 100 μl of cold 10% TCA, and acid-washed glass beads (Sigma-Aldrich) were added to each tube. Samples were vortexed for 10 minutes at 4°C. Cell lysates were then transferred to a fresh tube using a gel-loading pipette tip. Glass beads were washed two times with 100 μl of cold 10% TCA, and cell lysates were transferred to the respective tube. Lysates were cleared by centrifugation at maximum speed for 10 minutes at 4°C. After aspirating the supernatant, protein pellets were washed gently with 1 ml of cold 100% acetone. Aspirated pellets were flash frozen in liquid nitrogen and stored at -80°C. Mass spectrometry samples were transferred to the Duke Proteomics Facility (https://genome.duke.edu/cores-and-services/proteomics-and-metabolomics) for further processing described below.
Protein pellets for Western Blot experiments were resuspended in 100 μl of Thorner Buffer (8 M Urea, 5% SDS, 40 mM Tris-HCl pH 6.8, 0.1 mM EDTA, 0.4 mg/ml Bromophenol Blue, 1% β-Mercaptoethanol) plus 3 μl of 2 M Tris Base (unbuffered), boiled for 5 minutes, and stored at -20°C. SDS-PAGE gel electrophoresis was performed using the Laemmli method. Samples on the gel were transferred to an Immobilon-P PVDF membrane (Millipore) for about 2 hours using a Semi-Dry Transfer System.

Primary antibody solutions were composed of 1X TBS, 0.1% Tween-20, 5% dry milk (w/v), 0.02% NaN₃, and antibody. The primary antibodies used for tagged protein detection from this Chapter were c-Myc (mouse 9E10, Santa Cruz Biotechnology, used at 1:1000) and HA (mouse anti-HA.11, Covance, used at 1:1000). Cdc28 and Pho85 served as cell-cycle loading controls and were detected with mouse anti-PSTAIR (mouse, Abcam, used at 1:10,000 or 1:16,667). Secondary antibodies were horseradish peroxidase-conjugated (horse anti-mouse IgG, Cell Signaling Technology, used at 1:3000). Blots were visualized with SuperSignal™ West Pico Chemiluminescent Substrate (Thermo Fisher Scientific). Immunoblot quantification and normalization to the loading control were performed in ImageJ.

3.6.5 Protein sample preparation for mass spectrometry

Stable isotope-labeled (SIL) peptides (SpikeTide TQL) were ordered from JPT Peptide Technologies (Berlin, DE) and were digested with trypsin as previously described (Foster et al. 2014). Three peptides per protein of interest were selected based on available data, preferencing: 1) peptides utilized in a prior targeted study of yeast TFs (Mirzaei et al. 2013); and 2) peptides with optimal characteristics for targeted proteomics that had been identified in discovery-based proteomics studies (http://www.peptideatlas.org/) (Desiere et al. 2006). In the absence of these data, we
selected proteotypic peptides as previously designed and analyzed by Aebersold and coworkers (Picotti et al. 2013).

Yeast protein samples were isolated as described above. TCA pellets were resuspended in 100 µl of 0.25% (w/v) acid labile surfactant (ALS-1) in 50 mM ammonium bicarbonate, pH 8 (AmBic) followed by probe sonication for 4 x 3 seconds and shaking at 50°C for 10 minutes. After centrifugation at 15,000 xg for 20 minutes, protein concentrations of supernatants were measured by Bradford assay. Twenty µgs of protein per sample were reduced and denatured by heating at 80°C for 10 minutes in buffer containing 0.2 % ALS-1 and 10 mM DTT. After cooling, samples were alkylated with 25 mM iodoacetamide for 30 minutes and digested overnight with 1:50 (w/w) of Sequencing Grade Modified Trypsin (Promega) at 37°C. Following digestion, 2% (v/v) acetonitrile and 1% (v/v) TFA were added and the digests were incubated at 60°C for 2 hours. After centrifugation, SIL peptides (Supplementary Table 1) were added to a final concentration of 5 fmol of each peptide per µg of peptide digests. Finally, peptides were transferred to Maximum Recovery LC Vials (Waters).

3.6.6 Quantitative LC-MS/MS

Targeted quantitation of yeast proteins was performed using parallel reaction monitoring (PRM) (Gallien et al. 2012, 2014). Briefly 1 µg of peptide digests per sample were analyzed using a nanoACQUITY UPLC system (Waters) coupled to a QExactive Plus high resolution accurate mass tandem mass spectrometer (Thermo) via a nanoelectrospray ionization source. Peptides were trapped on a Symmetry C18 180 µm × 20 mm trapping column (5 µl/min at 99.9/0.1 v/v H₂O/MeCN) followed by an analytical separation using a 1.7 µm ACQUITY HSS T3 C18 75 µm × 250 mm column (Waters) with a 90 minute gradient of 5 to 40% MeCN with 0.1% formic acid at a flow rate of 400
nl/min and column temperature of 55°C. Data collection on the QExactive Plus MS was performed in targeted MS/MS mode 17,500 resolution (m/z 200) with a target AGC value of 5x10⁴ ions, an isolation width of 1.0 m/z and an ion fill time of 240 ms. Targeted MS/MS was triggered by an inclusion list, with 2 minute retention time windows for each precursor.

### 3.6.7 PRM Data Analysis

PRM data was analyzed using Skyline (MacLean et al. 2010). To generate a spectral library, 50 fmol of neat SpikeTides were analyzed by data-dependent LC-MS/MS using nanoACQUITY and QExactive Plus MS followed by database searching using Mascot v2.5. Raw MS/MS data was imported into Skyline. Data was imported using the following transition setting tabs: filter, y-ions, >precursor m/z to last ion; and Full-scan, targeted acquisition using Oribtrap analyzer with 17,500 resolution @ 200 m/z. Default peptide settings were used except for the structural modification, carbamidomethyl-Cys; and the isotope modifications, label: \(^{13}\text{C}_6^{15}\text{N}_4\)-Arg and label: \(^{13}\text{C}_6^{15}\text{N}_2\)-Lys isotope. Raw data was deposited in the Chorus database (https://chorusproject.org/) under the project “Yeast PRM Time-course”. Curated Skyline files were uploaded to the Panorama Targeted Proteomics data repository (panoramaweb.org) (Sharma et al. 2014). Links to download Skyline data from analyses of wild-type yeast and of clb1-6 knockout cells can be found in the publication. Normalized PRM data from this study are available in Supplementary Tables 4 and 5.
4. Sources of variability in RNA-Sequencing datasets from technical replicates in *Saccharomyces cerevisiae*

**Citation:** Kelliher CM, Roth CJ, Haase SB. (2017). RNA-Sequencing library preparation generates more variability than sequencing lane usage in technical replicates. *Manuscript in Preparation.*

### 4.1 Author Contributions

This Chapter discusses technical noise in RNA-Sequencing data. We utilized RNA-Seq to profile gene expression dynamics in Chapters 2, 3, 5, and 6. Thus, it was important to explore potential biases and noise in RNA-Seq measurements. These analyses will appear in the manuscript cited above.

I mentored Cullen Roth during his rotation in the Haase Laboratory in the 2015 Spring Semester. Cullen quantified the reproducibility of RNA-Seq data in sets of technical replicates described in this Chapter. For the manuscript, I analyzed the raw RNA-Sequencing data and produced the figures and tables based on results from Cullen’s rotation project. I wrote the manuscript and compiled edits with assistance from Steve Haase.

### 4.2 Introduction

RNA-Sequencing is a modern tool that is frequently applied in the biological sciences to quantify RNA expression levels. Microarrays are an alternative technology for measuring RNA expression and have been in use since the 1990s. Thus, sources of variation in microarray-based experiments are better characterized compared to RNA-
Here, we explore technical noise in RNA-Seq results from five replicated sets of RNA samples taken from the budding yeast *Saccharomyces cerevisiae*. Importantly, all of the RNA samples presented in this Chapter passed standard quality control metrics before sequencing, yet we still observed variation in the mRNA expression results. Overall, we demonstrate that noise between library preparations is significantly higher than lane-to-lane sequencing variability. We specifically follow up on the noisiest set of technical replicates and identify some “red flags” for resulting RNA-Seq data quality, including a potential species-specific effect from Illumina barcode selection.

Microarray technology was the first high-throughput method to quantify RNA expression from a biological sample. Microarray chips are designed to contain short DNA oligonucleotide probes that are complementary to many sequences of interest (A. C. Pease et al. 1994; Schena et al. 1995; Shalon et al. 1996). Thus, microarray experiments require prior knowledge of the desired transcripts to measure. Additionally, there are physical space limits on each chip to detect all features in complex transcriptomes. Gene expression values are quantified by hybridization and fluorescence between microarray probes and prepared biological samples. Normalization methods for microarrays are well developed and computationally accessible (G Gentleman et al. 2004; C. Li and Wong 2001).

RNA-Sequencing is also widely applied for high-throughput quantification of RNA expression. In general, RNA-Seq protocols consist of extracting RNA from a biological sample, preparing RNA for sequencing (known as library preparation), and next-generation sequencing. Typically, raw sequencing reads are then aligned to a reference genome of interest, and the abundance of RNA transcripts is quantified and normalized for each sample. A priori knowledge of transcripts is not needed for RNA-Seq.
experiments because most reads in a library preparation sample are sequenced at sufficient depth. RNA-Seq can also elucidate transcriptome details such as mRNA splicing variants (Nagalakshmi et al. 2008; Trapnell et al. 2010), non-coding RNAs (Wilhelm et al. 2008; Yassour et al. 2010), anti-sense transcription (Levin et al. 2010), and others (reviewed in: Z. Wang et al. 2009). To date, sequencing costs are decreasing, protocols and experimental kits are well described (Heyer et al. 2015; L. Wang et al. 2011), data storage is improving, and analysis pipelines are more streamlined and straightforward (Conesa et al. 2016).

Given the informational advantages of RNA-Sequencing approaches, many groups have switched from microarray to RNA-Seq to assay transcription levels. Next-generation sequencing technologies have been applied for only about a decade (Shendure et al. 2005). Therefore, we have less knowledge about biases in RNA-Seq sample preparation and next-generation sequencing errors. Several groups have subjected the same biological samples or conditions to both microarray and RNA-Seq profiling and observed good correlation between the results (J. S. Bloom et al. 2009; Fu et al. 2009; Nookaew et al. 2012; Su et al. 2014; Yassour et al. 2009) (reviewed in: Kratz and Carninci 2014). In both technologies, additional controls are required to estimate absolute mRNA quantities per cell. Sequencing is slightly more sensitive than microarrays for detecting lowly expressed genes because microarrays are subject to background hybridization noise. Additionally, microarray probes can become saturated when detecting very highly expressed genes, while RNA-Seq is not limited by hybridization or fluorescence.

Previous studies have investigated potential biases in RNA-Sequencing data acquisition. Both microarray and RNA-Seq assays require reverse transcription from
RNA to cDNA and sample labeling. Both of these procedures are subject to biases introduced by random primers and to PCR amplification biases. Microarrays are uniquely subject to noise in the chemistry of complementary DNA hybridization (discussed above) and in the optics of fluorescence detection. For RNA-Seq, a variety of library preparation techniques are available, but they are known to generate slight differences in the resulting read coverage across expressed genes (Lahens et al. 2014). One step in many RNA-Seq library preparation methods includes the ligation of 6-nucleotide barcode identifiers onto samples from each unique library (Craig et al. 2008). This barcoding process allows researchers to sequence multiple samples in a single sequencing lane and to link each sequencing read with the biological sample from which it was derived (known as multiplexing). For small RNA libraries, select barcodes have been shown to bias the gene expression results (S. Alon et al. 2011; van Nieuwerburgh et al. 2011).

Our group is particularly interested in transcription dynamics during the eukaryotic cell cycle. To study the cell cycle, we synchronize populations of cells in a certain phase of the cycle, release cells into fresh media, and take samples over time to profile RNA levels using a separate microarray or RNA-Seq sample for each time point. Many groups have conducted high-throughput expression profiling experiments to better understand cell-cycle transcription (Spellman et al. 1998) (reviewed in: Haase and Wittenberg 2014). We recently profiled the transcriptome of the budding yeast *Saccharomyces cerevisiae* using RNA-Seq, sampling cells every 5 minutes over about 3 cell cycles (Kelliher et al. 2016). Here, we describe five different time point samples from *S. cerevisiae* for which we replicated RNA-Seq library preparation to study two aspects of potential technical noise in the resulting mRNA expression measurements.
4.3 Results

Our cell-cycle transcriptome study involved multiplexing 10 time point samples in 6 sequencing lanes (60 total samples) (Kelliher et al. 2016). To better understand technical noise, we investigated whether RNA-Seq data is noisier between different sequencing lanes or between different library preparations of the same biological sample. We selected five time points (0, 45, 90, 135, and 180 minutes post release from synchronization) to perform RNA-Seq in technical replicates (N=3). We compared the same total RNA sample prepared with two different libraries followed by sequencing in the same lane, and we compared the same library preparation split and sequenced in two different lanes (Materials and Methods).

Raw RNA-Seq data from time points 0, 45, 90, 135, and 180 minutes were aligned to the S288C reference genome (Dobin et al. 2013). RNA-Seq reads mapping uniquely to annotated genes in *S. cerevisiae* were counted using HTSeq (Anders et al. 2015). These alignment-derived counts represent gene count data before normalization with other samples in the time series. We reasoned that, since replicated samples were derived from the same total RNA sample, counts might be similar between replicates even before standard RNA-Seq normalization methods were applied.

Our group is particularly interested in quantifying dynamically expressed genes during the cell cycle. Thus, we examined the gene expression levels of a list of 1246 genes that are periodically transcribed during the cell cycle (Fig 2A from:) (Kelliher et al. 2016). Furthermore, it is difficult to identify a set of constitutively expressed housekeeping genes across the cell cycle, as many commonly-used yeast loading controls display different expression levels across different environmental conditions (Teste et al. 2009) and over the cell cycle. We extracted the 1246 genes of interest from
each technical replicate before normalization and examined log-normalized distributions of read counts (Figure 4.16 A-E).

We found that the three technical replicates had a significant effect on the variation in raw gene counts across all 5 time points (ANOVA, Figure 4.16 legend). We posit that the significant variability of raw count results between technical replicates is due to uneven sequencing of multiplexed libraries. During the next-generation sequencing process, different libraries in a single lane can be slightly over- or under-sequenced, which will then produce raw data that is systematically over- or under-counted for each replicate sample prior to standard data normalization techniques.
Figure 4.16: Expression results from technical replicates had significant variability pre- and post-normalization for five yeast time points.

We considered 1246 periodic genes for the analysis of technical RNA-Seq replicates from five time points. We first looked at the distribution of gene expression values before data normalization using read counts from the HTSeq algorithm. Counts data were log-normalized and plotted as histograms for three technical replicates from time point 0 (A), time 45 (B), time 90 (C), time 135 (D), and time 180 (E). To ask if the technical replicates had significantly different gene count values from each other (and controlling for any individual gene effects), we ran ANOVAs in R (using aov with setup: log(expr_value) ~ tech_rep + Error(1246_gene_IDs / tech_rep)). For all five (A-E), we found that technical replicates had a significant effect on read counts (ANOVA, time point 0: p<2x10^{-16}; time 45: p<2x10^{-16}; time 90: p<2x10^{-16}; time 135: p<2x10^{-16}; time 180: p<2x10^{-16}). After normalization with standard methods, we examined the distribution of gene expression values (fpkm units). Expression data were log-normalized for three technical replicates and shown as histograms from time point 0 (F), time 45 (G), time 90 (H), time 135 (I), and time 180 (J). We ran the same ANOVA setup to ask if the technical replicates had significantly different normalized gene expression values. All five technical replicates (F-J) were significantly variable (time point 0: p= 0.0014; time 45: p= 7.8x10^{-6}; time 90: p= 4.5x10^{-13}; time 135: p= 0.0445; time 180: p= 3.0x10^{-14}).
Time series RNA-Seq samples were then normalized as described previously (Kelliher et al. 2016). In brief, data are first normalized by the number of reads derived from each sample, or library complexity (discussed above). Data are also normalized by average RNA-Seq read coverage across annotated genes (geometric mean by default in Cufflinks) and by gene length (e.g. if gene A is two times longer in genomic sequence than gene B, mRNA A should generate about twice as many RNA-Seq reads as mRNA B). After normalization, we examined the log-normalized distributions of mRNA expression values (fragments per kilobase of transcript per million mapped reads, or fpkm units) for the 1246 periodic cell-cycle genes of interest from each technical replicate (Figure 4.16 F-J). We found that all technical replicates continued to have a significant effect on the variation in normalized gene expression values (ANOVA; Figure 4.16 legend). To ask which technical replicates (library preparation or sequencing lane) generated the most variation in gene expression values, we performed pairwise t-tests on the log-normalized 1246 gene expression values from each technical replicate. For all five time points, we found that replicates with different library preparation had a more significant effect on the average gene expression values than replicates in different sequencing lanes (Table 4.5). This finding indicates that the library preparation step has more of an effect on the resulting gene expression value than sequencing lane used.
Table 4.5: RNA-Seq values derived from two libraries prepared with the same RNA sample are more variable than values derived from the same library sequenced in two different lanes.

Log-normalized RNA expression values for 1246 periodic genes (fpkm units) were obtained from technical replicates of five different time points (0, 45, 90, 135, 180 minutes). Pairwise t-tests between replicates were computed in the R statistical computing environment (using the function t.test with argument: paired=T). The 180-minute technical replicates showed the most significant variation between different library preparations.

<table>
<thead>
<tr>
<th>Technical Replicate</th>
<th>Sequencing Lane test (Lib. 1 &amp; Lane I versus Lib. 1 &amp; Lane II)</th>
<th>Library Preparation test (Lib. 1 &amp; Lane I versus Lib. 2 &amp; Lane I)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 minutes</td>
<td>0.0458</td>
<td>0.00103</td>
</tr>
<tr>
<td>45 minutes</td>
<td>0.515</td>
<td>0.000116</td>
</tr>
<tr>
<td>90 minutes</td>
<td>0.431</td>
<td>7.34x10^-8</td>
</tr>
<tr>
<td>135 minutes</td>
<td>0.329</td>
<td>0.131</td>
</tr>
<tr>
<td>180 minutes</td>
<td>0.163</td>
<td>8.67x10^-9</td>
</tr>
</tbody>
</table>

The 180-minute technical replicates showed the most significant variability even after normalization with standard methods (Table 4.5, Figure 4.16 J). Variability in RNA-Seq read coverage across genes has been previously observed for library preparation protocols that utilize a poly-A tail enrichment step (Lahens et al. 2014). Since our library preparation protocol included this step, we wanted to zoom in on a region in the yeast genome containing periodic gene(s) in our list. We identified a region where two highly periodic genes, HSP150 and CIS3, are arranged in a tail-to-tail orientation in the yeast genome (Cherry et al. 2012). These two genes are also periodically activated at different times during the cell cycle (Figure 4.17). Thus, this region of the genome allowed us to examine RNA-Seq read mapping patterns for low, intermediate, and high expression levels of two periodic genes. We visualized the RNA-Seq read distributions across the HSP150 and CIS3 gene loci for three significantly variable technical replicates (Figure 4.18). As expected, we observed patterns of slightly uneven coverage across both
genes in all technical replicates because our samples were prepared with a poly-A enrichment step (Lahens et al. 2014). However, RNA-Seq reads from the 180-minute samples were particularly unevenly distributed across the HSP150 and CIS3 genes (Figure 4.18 F).

**Figure 4.17: The periodic cell-cycle genes HSP150 and CIS3 are expressed at different times.**

Line plots depict RNA expression levels (fpkm units) over time for HSP150 (green) and CIS3 (blue) from a synchronous population of wild-type S. cerevisiae cells over about 3 cell cycles. Normalized time series data were taken from a previous study (Kelliher et al. 2016). Time points with technical replicate samples (0, 45, 90, 135, and 180 minutes) are depicted with red dots. Each set of technical replicates covers high, low, and intermediate gene expression values for both genes.
Figure 4.18: Technical replicates show variability in RNA-Seq read coverage across two periodic genes.

The periodically expressed \textit{HSP150} and \textit{CIS3} genes are arranged in a tail-to-tail orientation in the yeast genome. We quantified the number of reads at each coordinate in the \textit{HSP150} and \textit{CIS3} region using the depth feature in samtools (H. Li et al. 2009). Coverage of reads mapping to gene coordinates is plotted as a pair of histograms for the technical replicates from time point 0 (A, D), time 45 (B, E), and time 180 (C, F). Sequencing lane comparisons are shown in the top row (A, B, C). Library preparation comparisons are shown in the bottom row (D, E, F). The 180-minute sample has increased variance compared to other internal control samples.
Due to the significant variation observed in the 180-minute technical replicates, we asked if the 180-minute sample showed an unusual visual pattern in the context of the original cell-cycle time series experiment. We noticed that the time point samples corresponding to 180, 145, 175, and 190 minutes appeared “striped” in the time series heatmap (Fig 2A from:) (Kelliher et al. 2016). In other words, the striped time points visually stood out compared to adjacent time point samples.

We hypothesized that there may be differences in RNA quality metrics from striped samples compared to the rest of the time series. All of the RNA samples sequenced in our study passed standard quality control (QC) metrics; however, we reasoned that striped samples might have been slightly lower quality. We compiled RNA quality data for all 60 samples described in this Chapter and previously (Table 4.6). Three RNA quality metrics—concentration, 260/280, and 260/230 ratios—were not significantly different for the striped versus normal samples (t-test: p=0.516, p=0.782, and p=0.160, respectively). On average, the RNA Integrity (RIN) value was significantly lower for striped compared to normal samples (t-test: p=0.000983; striped sample mean: 8.03 versus non-stripe mean: 8.62). Low RIN numbers indicate some level of RNA degradation, and it has been previously shown that degraded RNA affects the coverage and quality of RNA-Seq results (Adiconis et al. 2013).
Table 4.6: RNA quality metrics for samples from this Chapter and our previous work showed that lower RNA integrity values (RIN) correlate with striped samples.

Samples are labeled with the corresponding time point as well as striped versus non-stripe designation (Fig 2A from:) (Kelliher et al. 2016). Yeast RNA samples were extracted, and three quality metrics were measured on a Nanodrop—concentration, 260/280, and 260/230 ratio. The 60 samples all passed quality control cutoffs of concentration > 20 ng/μl, 260/280 > 1.8, and 260/230 > 1.0, respectively (Duke Sequencing Facility). RNA Integrity (RIN) number was calculated using a Bioanalyzer on the “Plant” setting, which works best for yeast RNA. All samples passed the quality control cutoff of RIN > 7; however, we show that lower relative RIN values correlates with lower quality RNA-Seq data. Time points in red font denote striped samples.

<table>
<thead>
<tr>
<th>Time Point (minutes)</th>
<th>Sample Description</th>
<th>Designated as Striped?</th>
<th>Concentration (ng/μl)</th>
<th>260/280 Ratio</th>
<th>260/230 Ratio</th>
<th>RIN Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>(Kelliher et al. 2016); This Chapter</td>
<td>No</td>
<td>213.1</td>
<td>2.11</td>
<td>2.69</td>
<td>8.8</td>
</tr>
<tr>
<td>0</td>
<td>This Chapter</td>
<td>No</td>
<td>213.1</td>
<td>2.11</td>
<td>2.69</td>
<td>8.8</td>
</tr>
<tr>
<td>0</td>
<td>This Chapter</td>
<td>No</td>
<td>213.1</td>
<td>2.11</td>
<td>2.69</td>
<td>8.8</td>
</tr>
<tr>
<td>5</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>208.7</td>
<td>2.08</td>
<td>2.68</td>
<td>8.9</td>
</tr>
<tr>
<td>10</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>229.7</td>
<td>2.08</td>
<td>2.65</td>
<td>8.9</td>
</tr>
<tr>
<td>15</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>221.8</td>
<td>2.06</td>
<td>2.67</td>
<td>9.1</td>
</tr>
<tr>
<td>20</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>204.8</td>
<td>2.08</td>
<td>2.63</td>
<td>9.2</td>
</tr>
<tr>
<td>25</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>202</td>
<td>2.12</td>
<td>2.59</td>
<td>9</td>
</tr>
<tr>
<td>30</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>189.7</td>
<td>2.06</td>
<td>2.7</td>
<td>9.3</td>
</tr>
<tr>
<td>35</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>179.6</td>
<td>2.06</td>
<td>2.67</td>
<td>9.2</td>
</tr>
<tr>
<td>40</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>164</td>
<td>2.05</td>
<td>2.75</td>
<td>9.5</td>
</tr>
<tr>
<td>45</td>
<td>(Kelliher et al. 2016); This Chapter</td>
<td>No</td>
<td>176.3</td>
<td>2.05</td>
<td>2.66</td>
<td>9.5</td>
</tr>
<tr>
<td>45</td>
<td>This Chapter</td>
<td>No</td>
<td>176.3</td>
<td>2.05</td>
<td>2.66</td>
<td>9.5</td>
</tr>
<tr>
<td>45</td>
<td>This Chapter</td>
<td>No</td>
<td>176.3</td>
<td>2.05</td>
<td>2.66</td>
<td>9.5</td>
</tr>
<tr>
<td>50</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>187.5</td>
<td>2.06</td>
<td>2.7</td>
<td>9.6</td>
</tr>
<tr>
<td>55</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>236.2</td>
<td>2.09</td>
<td>2.66</td>
<td>9.6</td>
</tr>
<tr>
<td>60</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>187.8</td>
<td>2.06</td>
<td>2.66</td>
<td>8.7</td>
</tr>
<tr>
<td>65</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>205.7</td>
<td>2.07</td>
<td>2.67</td>
<td>8.8</td>
</tr>
<tr>
<td>70</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>182.1</td>
<td>2.07</td>
<td>2.7</td>
<td>9.1</td>
</tr>
<tr>
<td>75</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>185.1</td>
<td>2.03</td>
<td>2.72</td>
<td>9</td>
</tr>
<tr>
<td>80</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>193.6</td>
<td>2.09</td>
<td>2.71</td>
<td>9.1</td>
</tr>
<tr>
<td>85</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>186.9</td>
<td>2.08</td>
<td>2.67</td>
<td>8.9</td>
</tr>
<tr>
<td>90</td>
<td>(Kelliher et al. 2016); This Chapter</td>
<td>No</td>
<td>183.3</td>
<td>2.01</td>
<td>2.67</td>
<td>9.2</td>
</tr>
<tr>
<td>90</td>
<td>This Chapter</td>
<td>No</td>
<td>183.3</td>
<td>2.01</td>
<td>2.67</td>
<td>9.2</td>
</tr>
<tr>
<td>Page</td>
<td>This Chapter</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>183.3</td>
<td>2.01</td>
<td>2.67</td>
</tr>
<tr>
<td>------</td>
<td>-------------</td>
<td>------------------------</td>
<td>----</td>
<td>--------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>90</td>
<td>This Chapter</td>
<td>No</td>
<td>115</td>
<td>2.02</td>
<td>2.74</td>
<td>9.1</td>
</tr>
<tr>
<td>95</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>138.7</td>
<td>2.03</td>
<td>2.68</td>
<td>9.1</td>
</tr>
<tr>
<td>100</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>148.5</td>
<td>2.04</td>
<td>2.68</td>
<td>9.1</td>
</tr>
<tr>
<td>105</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>122.7</td>
<td>2.04</td>
<td>2.8</td>
<td>9.1</td>
</tr>
<tr>
<td>110</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>149.9</td>
<td>1.99</td>
<td>2.77</td>
<td>9.1</td>
</tr>
<tr>
<td>115</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>153.4</td>
<td>2.05</td>
<td>2.78</td>
<td>9.2</td>
</tr>
<tr>
<td>120</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>165.3</td>
<td>2.03</td>
<td>2.76</td>
<td>9.3</td>
</tr>
<tr>
<td>125</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>144.1</td>
<td>2.01</td>
<td>2.76</td>
<td>9.4</td>
</tr>
<tr>
<td>130</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>184.8</td>
<td>1.98</td>
<td>2.71</td>
<td>8</td>
</tr>
<tr>
<td>135</td>
<td>(Kelliher et al. 2016); This Chapter</td>
<td>No</td>
<td>184.8</td>
<td>1.98</td>
<td>2.71</td>
<td>8</td>
</tr>
<tr>
<td>140</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>197.1</td>
<td>2.07</td>
<td>2.61</td>
<td>8</td>
</tr>
<tr>
<td>145</td>
<td>(Kelliher et al. 2016)</td>
<td>Yes</td>
<td>164</td>
<td>2.04</td>
<td>2.64</td>
<td>7.6</td>
</tr>
<tr>
<td>150</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>180.4</td>
<td>2.07</td>
<td>2.69</td>
<td>7.9</td>
</tr>
<tr>
<td>155</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>175</td>
<td>2.03</td>
<td>2.7</td>
<td>7.8</td>
</tr>
<tr>
<td>160</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>222.4</td>
<td>2.08</td>
<td>2.67</td>
<td>7.9</td>
</tr>
<tr>
<td>165</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>205.3</td>
<td>2.08</td>
<td>2.66</td>
<td>7.8</td>
</tr>
<tr>
<td>170</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>185.3</td>
<td>2.03</td>
<td>2.64</td>
<td>7.8</td>
</tr>
<tr>
<td>175</td>
<td>(Kelliher et al. 2016)</td>
<td>Yes</td>
<td>150.5</td>
<td>2.05</td>
<td>2.7</td>
<td>8.2</td>
</tr>
<tr>
<td>180</td>
<td>(Kelliher et al. 2016); This Chapter</td>
<td>Yes</td>
<td>203.2</td>
<td>2.05</td>
<td>2.65</td>
<td>8</td>
</tr>
<tr>
<td>185</td>
<td>(Kelliher et al. 2016)</td>
<td>Yes</td>
<td>203.2</td>
<td>2.05</td>
<td>2.65</td>
<td>8</td>
</tr>
<tr>
<td>190</td>
<td>(Kelliher et al. 2016)</td>
<td>Yes</td>
<td>164.6</td>
<td>2.06</td>
<td>2.7</td>
<td>8.4</td>
</tr>
<tr>
<td>195</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>169.4</td>
<td>2.02</td>
<td>2.76</td>
<td>7.6</td>
</tr>
<tr>
<td>200</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>214.2</td>
<td>2.09</td>
<td>2.7</td>
<td>7.6</td>
</tr>
<tr>
<td>205</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>222.8</td>
<td>2.05</td>
<td>2.68</td>
<td>7.6</td>
</tr>
<tr>
<td>210</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>289.7</td>
<td>2.09</td>
<td>2.61</td>
<td>7.6</td>
</tr>
<tr>
<td>215</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>209.2</td>
<td>2.06</td>
<td>2.67</td>
<td>7.5</td>
</tr>
<tr>
<td>220</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>195.5</td>
<td>2.01</td>
<td>2.65</td>
<td>7.6</td>
</tr>
<tr>
<td>225</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>230.1</td>
<td>2.04</td>
<td>2.66</td>
<td>7.5</td>
</tr>
<tr>
<td>230</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>223.3</td>
<td>2.05</td>
<td>2.69</td>
<td>7.7</td>
</tr>
<tr>
<td>235</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>226.9</td>
<td>2.07</td>
<td>2.63</td>
<td>8.1</td>
</tr>
<tr>
<td>240</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>191.2</td>
<td>2.08</td>
<td>2.7</td>
<td>7.7</td>
</tr>
<tr>
<td>245</td>
<td>(Kelliher et al. 2016)</td>
<td>No</td>
<td>129.6</td>
<td>2.03</td>
<td>2.66</td>
<td>8.1</td>
</tr>
</tbody>
</table>
Importantly, the slightly lower RIN numbers for striped samples were all above the minimum QC value recommended by the sequencing facility, and so we continued to investigate variables specific to RNA-Seq protocols that could explain the striped pattern. We first asked if the amount of sequencing reads generated from a striped sample (with potentially higher RNA degradation than normal samples) was decreased. Three RNA-Seq alignment metrics—total reads, total reads aligned uniquely to the yeast genome, and total reads mapped to annotated genes—were not significantly different for the striped versus normal samples (t-test: p=0.306, p=0.381, and p=0.236, respectively). Based on findings from previous work, we also wanted to investigate if the 6-nucleotide barcode sequence could be a source of variability for striped samples. Here, we reasoned that the addition of 6 extra nucleotides during library preparation could catalyze the formation of atypical cDNA secondary structures.

We mapped the occurrence of each 6-nucleotide sequence and its reverse complement to the S288C reference genome (which includes the mitochondrial genome sequence) (Table 4.7). We found that one particular barcode, 5' ACTTGA 3', mapped frequently to the *S. cerevisiae* yeast genome and occurred in 2 of the striped samples (145, 180 minutes). However, the frequency of barcode mapping to the yeast genome was not significantly different between striped versus non-stripe samples (t-test: p=0.303). Taken together, we predict that a combination of input RNA quality and barcode selection can affect resulting RNA-Seq data quality.
Table 4.7: Illumina barcodes and their reverse complement sequences map to the yeast genome with varying frequencies, and one particular barcode is associated with 2 striped samples.

6-nucleotide barcode and reverse complement sequences were counted in the S288C yeast genome. Barcodes are ordered by the most to least frequent in the yeast genome. Time points in red font denote striped samples.

<table>
<thead>
<tr>
<th>Barcode Sequence (5' to 3')</th>
<th>Reverse Complement Sequence (5' to 3')</th>
<th>Total Mappings to S288C Genome</th>
<th>Time Point Samples with Barcode (minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATTCCT</td>
<td>AGGAAT</td>
<td>9069</td>
<td>0</td>
</tr>
<tr>
<td>ACTTGA</td>
<td>TCAAGT</td>
<td>8732</td>
<td>75, 115, 145, 180</td>
</tr>
<tr>
<td>ATGTCA</td>
<td>TGACAT</td>
<td>7777</td>
<td>135, 215</td>
</tr>
<tr>
<td>CTTGTA</td>
<td>TACAAG</td>
<td>7526</td>
<td>90, 165, 200</td>
</tr>
<tr>
<td>AGTCAC</td>
<td>TTGACT</td>
<td>7331</td>
<td>170, 205</td>
</tr>
<tr>
<td>GCCAAT</td>
<td>ATGGGC</td>
<td>6883</td>
<td>65, 105, 135</td>
</tr>
<tr>
<td>ACTGAT</td>
<td>ATCAGT</td>
<td>6879</td>
<td>40</td>
</tr>
<tr>
<td>TAGCTT</td>
<td>AAGCTA</td>
<td>6602</td>
<td>85, 125, 155, 190, 245</td>
</tr>
<tr>
<td>TGACCC</td>
<td>TGGTCA</td>
<td>6270</td>
<td>15, 55, 95, 225</td>
</tr>
<tr>
<td>AGTCAC</td>
<td>GGAAACT</td>
<td>5313</td>
<td>175, 210</td>
</tr>
<tr>
<td>CAGATC</td>
<td>GATCTG</td>
<td>4718</td>
<td>70, 110, 140, 235</td>
</tr>
<tr>
<td>ACAGTG</td>
<td>CACTGT</td>
<td>4717</td>
<td>20, 60, 100, 230</td>
</tr>
<tr>
<td>GTTCGA</td>
<td>CGAAAC</td>
<td>4320</td>
<td>25</td>
</tr>
<tr>
<td>GATCAG</td>
<td>CTGATC</td>
<td>4216</td>
<td>80, 120, 150, 185, 240</td>
</tr>
<tr>
<td>CGATGT</td>
<td>ACATCG</td>
<td>4061</td>
<td>5, 45</td>
</tr>
<tr>
<td>ATCACG</td>
<td>CGTGTG</td>
<td>3329</td>
<td>0</td>
</tr>
<tr>
<td>GAGTGAC</td>
<td>CCACCTC</td>
<td>3126</td>
<td>35</td>
</tr>
<tr>
<td>TTAGCC</td>
<td>GCTACTA</td>
<td>2738</td>
<td>10, 50, 90</td>
</tr>
<tr>
<td>GGCTAC</td>
<td>GTAGCCT</td>
<td>2638</td>
<td>45, 130, 160, 195</td>
</tr>
<tr>
<td>CCGTGC</td>
<td>GGACGG</td>
<td>1621</td>
<td>220</td>
</tr>
<tr>
<td>GTCCGC</td>
<td>GCGGCC</td>
<td>1331</td>
<td>180</td>
</tr>
<tr>
<td>CGTACG</td>
<td>CGTACG</td>
<td>841</td>
<td>30</td>
</tr>
</tbody>
</table>

4.4 Discussion

RNA-sequencing is becoming the most prevalent method for measuring gene expression levels. Here, we utilized technical replicates from the budding yeast *S. cerevisiae* to investigate potential sources of noise in RNA-Seq data. We observed
greater noise between two library preparations on the same RNA sample than between
the same sample applied to two different sequencing lanes (Figure 4.16, Table 4.5). We
also identified some red flags for inconsistent mRNA expression results between
technical replicates, including lower RNA integrity number and a potential effect from
sequencing barcode selection (Table 4.6, Table 4.7). We demonstrate that lower quality
RNA-Seq data appeared to have more aberrant read coverage patterns across genes
compared to high quality samples (Figure 4.18).

In two cases, we identified an association between the Illumina TruSeq barcode
5’ ACTTGA 3’ and poorer quality RNA-Seq results in the context of the yeast time series
experiment. This barcode and its reverse complement sequence appear at high
frequency in the *S. cerevisiae* genome (Table 4.7). We predict that the introduction of
this barcode may cause atypical cDNA secondary structures to form, which could then
bias sequencing results. However, barcode selection alone does not predict poor quality
results because two other non-stripe time point samples were also prepared in libraries
with the 5’ ACTTGA 3’ barcode (Table 4.7). In a future set of experiments with increased
sample size for statistical tests, it will be important to deconvolve the effects of individual
barcodes and different mixtures of barcodes during multiplexing on RNA-Seq results
(Auer and Doerge 2010). This potential barcode bias is an important direction for future
work because different genomes subjected to RNA-Seq will have a different set of 6-
nucleotide frequencies from the yeast samples described in this Chapter.

We also identified an RNA quality metric, low RIN number, which correlated with
poorer quality RNA-Seq results (Table 4.6). It is not surprising that RNA degradation
would affect the ability to sequence and quantify RNA expression levels. Some RNA-
Seq library preparation protocols involve an enrichment step of RNAs with poly-A tails
(including data from this Chapter), and 5’ RNA degradation could then lead to biased increases in read coverage on the 3’ ends of transcripts. Future work could test this hypothesis by subjecting a slightly degraded sample to both RNA-Seq and microarray profiling. Since microarray probes can be designed across the entire length of a gene, it would be interesting to compare these results to an RNA-Seq library with a poly-A enrichment step, where we would predict over-representation of 3’ read coverage.

Reproducibility is a key aspect of scientific progress, and the relatively low level of noise between sequencing lanes is encouraging for experiments being performed in different sequencing facilities. However, the observed noise in library preparation is worth considering for reproducibility in RNA-Seq technology. An additional improvement for precise and accurate quantification of mRNA is to standardize the use of spike-in controls (Jiang et al. 2011; Risso et al. 2014) (reviewed in: Lovén et al. 2012). Future sequencing technologies with less involved library preparation steps such as amplification free sequencing (Lipson et al. 2009) and nanopore sequencing are promising for reducing the biases observed here.

The most common experimental application of RNA-Seq technology is to identify differential gene expression, or DGE, in triplicate from “case” and control conditions (e.g. tumor versus tissue-matched normal cells). There are a variety of tools available to analyze such DGE datasets (Soneson and Delorenzi 2013); however, there is some disagreement in the field about proper quantification, normalization, and statistical methods to generate a DGE list of up- and down-regulated genes. Triplicate samples are leveraged to average out biological noise in these case/control datasets, but systematic noise in a given sample could then systematically bias a DGE list (S. Li et al. 2014). We propose that such case/control studies should check for low quality data by visualizing
the coverage of reads across genes of interest before averaging with other biological replicates.

RNA-sequencing and microarray approaches reveal steady-state levels of RNA from a given sample. These expression values represent the sum of new transcription (nascent RNAs, processed transcripts) and RNA degradation across a population of cells. Steady-state levels of RNA can be helpful for identifying DGE gene lists in the case/control experimental design described above; however, sampling a single time point can mask many dynamic cellular processes and/or different cell types in a sample (e.g. patient tumor biopsies). Here, we leveraged data from five different time points taken from synchronized populations of yeast cells and examined genes that are known to be dynamically expressed. We demonstrate that much of the technical noise in RNA-Seq experiments is explained by variability in library preparation. We identify some red flags for RNA-Seq data quality that can also be applied to steady-state differential gene expression experiments—low RNA integrity value and Illumina barcode selection, which manifest as aberrant read mapping patterns across expressed genes.

4.5 Materials and Methods

4.5.1 Yeast strains, cultures, and synchronization

The wild-type *Saccharomyces cerevisiae* strain is a derivative of BF264-15D MATα bar1 (Cole and Reed 1991; Reed et al. 1985). Yeast cultures were grown in standard YEPD medium (1% yeast extract, 2% peptone, 0.012% adenine, 0.006% uracil supplemented with 2% dextrose sugar). Alpha-factor arrest/release experiments were performed and sequenced as described previously (Kelliher et al. 2016).

Total RNA was extracted from yeast samples taken at time points 0 minutes, 45 minutes, 90 minutes, 135 minutes, and 180 minutes after release from alpha-factor
arrest. Libraries were prepared for sequencing in triplicate from the five time point samples (50 base pair, single end reads using Illumina TruSeq Stranded mRNA kits). Each sample was prepared in two separate libraries and sequenced in the same sequencing lane (multiplexed with 10 total samples per lane). In addition, one of the library preparations was split into two aliquots and sequenced in two separate lanes.


4.5.2 RNA-Sequencing analysis and normalization

Samples were prepared and sequenced at the Duke Sequencing Facility as described previously (Kelliher et al. 2016). The *S. cerevisiae* S288C genome (Ensembl build R64-1-1) was downloaded from Illumina iGenomes in March 2016. Raw FASTQ files were aligned to the reference genome using STAR (Dobin et al. 2013). Raw FASTQ files were examined for quality using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Raw reads mapping uniquely to annotated genes were counted using HTSeq (Anders et al. 2015). Aligned reads were assembled into transcripts, quantified, and normalized using Cufflinks (Trapnell et al. 2013). The 15 replicated samples described above were normalized together using the CuffNorm feature. The normalized output fpkm gene expression levels were used in the analyses presented. To avoid fractional and zero values, 1 was added to every fpkm value using the R statistical programming environment (R Core
Development Team 2017). Raw RNA-sequencing data from the 15 technical replicate samples described in this Chapter will be available at the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) upon publication.
5. Cell-Cycle Transcription in the Fungal Pathogen Cryptococcus neoformans


5.1 Author Contributions

In this Chapter, we investigate the cell-cycle transcription program of a previously uncharacterized species of fungi, Cryptococcus neoformans, and compare findings to the distantly related yeast Saccharomyces cerevisiae.

For the manuscript, I performed the wild-type S. cerevisiae time series transcriptome profiling experiment. Adam Leman and Crystal Sierra performed the C. neoformans time series profiling experiment, all of the relevant lead-up experiments, and Biosafety Level 2 laboratory protocols. I analyzed the raw RNA-Sequencing data from both yeast experiments and produced the figures and tables herein. I wrote the manuscript and compiled co-author edits with guidance from Steve Haase throughout.

5.2 Introduction

About 500 million years of evolution separate the fungal phyla Ascomycota and Basidiomycota (Stajich et al. 2009; Taylor et al. 1999). The cell cycle is an essential biological process driving cell division of these distantly related yeasts, and therefore may be under strong selective pressure for conservation. Both Saccharomyces cerevisiae (Ascomycota) and Cryptococcus neoformans (Basidiomycota) can grow and
divide asymmetrically in a budding yeast form. *C. neoformans* is a causative agent of deadly fungal meningitis, primarily in immune-compromised patients (Brown et al. 2012; Park et al. 2009). Many groups studying *C. neoformans* focus on virulence factors for human infection, such as the yeast’s polysaccharide capsule, melanin production, Titan cell formation, and others (Buchanan and Murphy 1998; Cairns et al. 2016; Karkowska-Kuleta et al. 2009; Kozubowski and Heitman 2012; O. W. Liu et al. 2008). We propose that the function of cell-cycle regulators, which are essential for proliferation in the host, merit further investigation as virulence factors. Furthermore, there is evidence that virulence pathways are perturbed when cell-cycle progression is slowed, which suggests direct connections between cell-cycle regulators and virulence pathways (García-Rodas et al. 2014, 2015).

The cell cycle is the process by which a cell duplicates its contents and faithfully divides into two genetically identical cells. In eukaryotes, a biochemical oscillator drives sequential cell-cycle events, where the cyclin-dependent kinase (CDK) and its variety of cyclin binding partners initiate events by phosphorylation, followed by destruction of kinase activity in mitosis by the anaphase-promoting complex (APC). Another common feature of the eukaryotic cell cycle is a temporally regulated program of transcription, which has been demonstrated in *S. cerevisiae*, *Schizosaccharomyces pombe*, *Arabidopsis thaliana*, mouse fibroblasts, and human tissue culture cells (Bar-Joseph et al. 2008; Cho et al. 1998; Grant et al. 2013; Ishida et al. 2001; Menges et al. 2005; Oliva et al. 2005; Orlando et al. 2008; Peng et al. 2005; Pramila et al. 2006; Rustici et al. 2004; Spellman et al. 1998). These programs of periodic genes include cyclin mRNAs, DNA replication factors, APC activators, and other cellular components that are utilized at specific times during the cell cycle. Our group and others have proposed that this “just-
The in-time transcription mechanism is an important aspect of energy-efficient and faithful cell divisions (Haase and Wittenberg 2014; G. Z. Wang et al. 2015). In *S. cerevisiae*, an interconnected network of periodic transcription factors (TFs) is capable of driving the periodic program of cell-cycle gene expression (F. Li et al. 2004; Orlando et al. 2008; Simmons Kovacs et al. 2012; Simon et al. 2001). Aspects of this yeast TF network are conserved in human cells; for example, G2/M genes are activated by a periodic forkhead domain-containing TF in both eukaryotes (Grant et al. 2013; Reynolds et al. 2003). The topology of cell-cycle entry is also functionally conserved, where a repressor (*S. c.* WHI5, *H. s.* RB1) is removed by G1 cyclin/CDK phosphorylation to activate a G1/S transcription factor complex (*S. c.* SBF/MBF, *H. s.* E2F-TFDP1) (Bertoli et al. 2013). However, the genes involved in cell-cycle entry are not conserved at the sequence level between fungi and mammals (Medina et al. 2016), suggesting that the fungal pathway could be targeted with drugs without affecting mammalian host cells.

Sequence-specific DNA-binding TFs have been identified in *C. neoformans* and phenotypically profiled by single gene knockouts (Jung et al. 2015; O. W. Liu et al. 2008; Maier et al. 2015). This TF deletion collection was profiled over many virulence factor-inducing conditions to discover pathways that regulate disease and drug response genes (Jung et al. 2015). Serial activation of TFs during capsule production has also been studied to elucidate the order in which TFs control virulence gene products (Maier et al. 2015). However, the cell cycle has not been investigated in synchronous populations of cells to date. Although the phenotypes of some single mutant cell-cycle TFs have been examined from asynchronous populations, these studies offer limited understanding of temporal aspects of gene expression during the cell cycle.
Here we investigate transcriptional dynamics of the pathogenic yeast *C. neoformans* using cells synchronized in the cell cycle. We compare our findings to the cell-cycle transcriptional program in *S. cerevisiae*. We find that a similar percentage of all genes (~20%) are periodically transcribed during the cell cycle, and we present a comprehensive periodicity analysis for all expressed genes in both yeasts. We show that S-phase gene orthologs are highly conserved and temporally precede M-phase gene orthologs in both yeasts. Additionally, we find that many TFs in the cell-cycle entry pathway are conserved in sequence homology, periodicity, and timing of expression in *C. neoformans*, while others, notably genes involved in budding, are not. We also identify 40 virulence genes that appear to be cell-cycle-regulated, along with nearly 100 orthologous fungal genes that are periodic in the same cell-cycle phase. Taken together, these cell-cycle genes represent candidates for further study and for novel antifungal drug development.

### 5.3 Results

#### 5.3.1 Cell-cycle synchronization and determination of periodic gene expression

Identifying approaches for synchronizing populations of *C. neoformans* has been challenging. We succeeded in synchronizing by centrifugal elutriation, a method that has been very successful for *S. cerevisiae* cells (Bristow et al. 2014; Orlando et al. 2008; Simmons Kovacs et al. 2012). For *C. neoformans*, we isolated early G1 daughter cells by centrifugal elutriation and released the population into rich media (YEPD) at 30°C to monitor cell-cycle progression, as described previously (Leman et al. 2014). This size-gradient synchrony procedure is conceptually similar to the *C. neoformans* synchrony procedure presented by Raclavsky and colleagues (Raclavsky et al. 2002). For *S.
*cerevisiae*, we isolated G1 cells by alpha-factor mating pheromone treatment (Breeden 1997). We utilized this synchrony technique to isolate larger *S. cerevisiae* cells and to offset some loss of synchrony over time due to asymmetric cell divisions. A functional mating pheromone peptide for *C. neoformans* has been described but is difficult to synthesize in suitable quantities (Davidson et al. 2000). After release from synchronization, bud formation and population doubling were counted for at least 200 cells over time (Figure 5.19). The period of bud emergence was about 75 minutes in both budding yeasts grown in rich media, although the synchrony of bud emergence after the first bud in *C. neoformans* appeared to be less robust (Figure 5.19 A-B). Each yeast population completed more than two population doublings over the course of the experiments.
Figure 5.19: Population synchrony for *S. cerevisiae* and *C. neoformans* over > 2 cell cycles.

*S. cerevisiae* cells were grown in 2% YEPD media, synchronized by alpha-factor mating pheromone, and released into YEPD (A). *C. neoformans* cells were grown in 2% YEPD rich media; small daughter cells were isolated by centrifugal elutriation and released into YEPD (B). Population synchrony was estimated by counting at least 200 cells per time point for the presence or absence of a bud, and doubling time was also monitored (C-D). Orange arrows indicate the time points where each population passed a complete doubling in cell concentration from the previous cycle (gray lines).

Total RNA was extracted from yeast cells at each time point (every 5 minutes for *S. cerevisiae*, or every 10 minutes for *C. neoformans*) and multiplexed for stranded RNA-Sequencing. Between 87-92% of reads mapped uniquely to the respective yeast genomes (S1 File). To identify periodic genes, we applied periodicity algorithms to the time series gene expression datasets. Four algorithms were used to determine periodicity rankings for all genes in each yeast: de Lichtenberg, JTK-CYCLE, Lomb-Scargle, and persistent homology (Cohen-Steiner et al. 2010; de Lichtenberg et al. 2005;
Deckard et al. 2013; Hughes et al. 2010; Lomb 1976; Scargle 1982). Since each algorithm favors slightly different periodic curve shapes (Deckard et al. 2013), we summed the periodicity rankings from each algorithm and ranked all yeast genes by cumulative scores for *S. cerevisiae* and for *C. neoformans* (S1 Table and S2 Table, respectively). By visual inspection, the top 1600 ranked genes in both yeasts appeared periodically transcribed during the cell cycle (S1 Fig). There was no clear “threshold” between periodic and non-periodic genes during the cell cycle—rather, we observed a distribution of gene expression shapes and signatures over time (S1 Fig). Previous work on the *S. cerevisiae* cell cycle has reported lists ranging from 400-1200 periodic genes.

To validate our RNA-sequencing time series dataset for the *S. cerevisiae* cell cycle, we compared the top-ranked 1600 periodic genes to previously published cell-cycle gene lists and found a 57-89\% range of overlap with previous periodic gene lists (S2 Fig) (Bristow et al. 2014; Cho et al. 1998; de Lichtenberg et al. 2005; P. Eser et al. 2014; Granovskaia et al. 2010; Orlando et al. 2008; Pramila et al. 2006; Spellman et al. 1998).

Three filters were applied to each budding yeast dataset to estimate and compare the number of periodic genes (S1 File). First, we pruned noisy, low-expression genes from each dataset, leaving 5913 expressed genes in *S. cerevisiae* (S1 Table) and 6182 expressed genes in *C. neoformans* (S2 Table). Next, we took the top 1600 expressed genes from the cumulative ranking of the four periodicity algorithms described above. Finally, we applied a score cutoff to each list of top 1600 genes using the Lomb-Scargle algorithm (see S1 file) (Deckard et al. 2013; Lomb 1976; Scargle 1982). We estimated that there are 1246 periodic genes in *S. cerevisiae* (~21\% expressed genes) and 1134 periodic genes in *C. neoformans* (~18\% expressed genes) (Figure 5.20). We
also provided multiple criteria for evaluating the cell-cycle expression patterns of individual genes in each yeast (S1 Table, S2 Table, S1 Fig).

![Graph](image)

**Figure 5.20**: About 20% of all *S. cerevisiae* and *C. neoformans* genes are periodically expressed during the cell cycle.

Four periodicity-ranking algorithms were run on the time series gene expression datasets at a period of 75 minutes (see S1 File). The top-ranked periodic genes (1-1600) were then filtered by the Lomb-Scargle algorithm to identify (A) 1246 periodic genes in *S. cerevisiae* and (B) 1134 periodic genes in *C. neoformans*. Genes in each periodic gene list were ordered along the y-axis by peak time of expression in the respective yeast dataset. As expected, the second and third cell cycles showed expression level damping due to asymmetric cell divisions in both budding yeasts. Transcript levels are depicted as a z-score change relative to mean expression for each gene, where values represent the number of standard deviations away from the mean. Each row represents transcript levels of a unique gene across the time series. Each column represents a time point in minutes.

Cellular processes that contribute to virulence are a major focus of work in the *C. neoformans* field. We took advantage of the partial *C. neoformans* deletion collection and genetic screens for virulence factors (O. W. Liu et al. 2008) and searched for periodic virulence genes. We found that 40 genes (about 16% of the virulence genes characterized by the Madhani group and many previous studies) were periodically
expressed in *C. neoformans* during the cell cycle (S3 Table). These virulence genes are periodic during normal cycles in rich media, which suggests that some virulence processes are directly cell-cycle-regulated. For example, budding and cell wall synthesis are coupled to cell-cycle progression in *S. cerevisiae*. A subset of 14 periodic virulence genes in *C. neoformans* had capsule and/or cell wall phenotypes reported in previous studies (S3 Table). We then asked if the 40 periodic virulence genes might be co-regulated during the *C. neoformans* cell cycle (S3 Fig). Over half of the periodic virulence genes clustered together and peaked in a similar cell-cycle phase (20-30 minutes into cycle 1). 11 of the 14 capsule / cell wall genes were contained in this cluster (S3 Fig, S3 Table).

Next, we wanted to ask if periodicity and temporal ordering of orthologous genes is evolutionarily conserved between the two budding yeasts. We compiled the largest list to date of putative sequence orthologs between *C. neoformans* and *S. cerevisiae* from the literature, databases, and additional BLAST searches (S1 File, S4 Table) (Janbon et al. 2014; Johnson et al. 2008; Jung et al. 2015; Stajich et al. 2012). About half of the periodic genes from each yeast (Figure 5.20) had at least one sequence ortholog in the other species. However, there were only about 230 pairs of orthologous genes that were labeled periodic in both yeasts. Those pairs of periodic orthologs have diverged in temporal ordering between *C. neoformans* and *S. cerevisiae* (Figure 5.21, S5 Table). These results indicated that the programs of periodic gene expression, and possibly the regulatory pathway, have diverged to some degree between the two budding yeasts. This altered temporal ordering between *S. cerevisiae* and *C. neoformans* periodic orthologous genes was likely not due to the experimental synchrony procedure. We obtained transcriptome data from two previous studies on *S. cerevisiae* cell-cycle-
regulated transcription (which applied a different cell-cycle synchrony procedure, used different lab strains of *S. cerevisiae*, and/or measured gene expression on different platforms), and our list of periodic *S. cerevisiae* genes maintained temporal ordering during the cell cycle in all three datasets (S4 Fig).

![Diagram](image)

**Figure 5.21:** Periodic, orthologous genes between *S. cerevisiae* and *C. neoformans* are differentially ordered during the cell cycle.

In *S. cerevisiae*, 753 genes out of the 1246 periodic genes had at least one ortholog in *C. neoformans* (60.4%). In *C. neoformans*, 593 genes out of the 1134 periodic genes had at least one ortholog in *S. cerevisiae* (52.3%). The intersection of these two gene lists contained 237 unique *S. cerevisiae* (**A**) and 225 unique *C. neoformans* (**B**) gene orthologs that were periodic in both budding yeasts. *C. neoformans* orthologs were plotted in the same relative order as their ortholog in *S. cerevisiae* (**B**), and we observed that many periodic genes have diverged in temporal ordering between the two yeasts. Transcript levels are depicted as a z-score change relative to mean expression for each gene, where values represent the number of standard deviations away from the mean. Orthologous periodic gene pairs are in the same relative order for (**A-B**) (for exact ordering of gene pairs and multiple-mapping orthologs, see S5 Table). Each column represents a time point in minutes.

Cell-cycle regulated gene expression has also been investigated in a species of pathogenic Ascomycota, *Candida albicans* (Côte et al. 2009). To ask about common
periodic gene expression in an evolutionarily intermediate budding yeast species, we further identified putative periodic orthologous genes shared between S. cerevisiae, C. neoformans, and C. albicans. A core set of almost 100 orthologs appeared to have both conserved periodicity and temporal ordering between all three budding yeasts (S5 Fig, S5 Table). This fungal gene set was enriched for functions in mitotic cell cycle and cell-cycle processes, which suggested that core cell-cycle regulators are under strong selection for conservation at the sequence level and by timing of periodic gene expression.

5.3.2 Conservation of known cell-cycle regulators

We reasoned that some cell-cycle events must be invariable in temporal ordering between fungi (S5 Fig). DNA replication (S-phase) should be highly conserved across organisms because duplication of genetic material is essential for successful division. Segregation of genomic content during mitosis (M-phase) is also essential for division, and duplication must precede division. Using annotations for S. cerevisiae (Cherry et al. 2012) we identified lists of genes known to be involved in regulating events in various cell-cycle phases including bud formation and growth (Bi and Park 2012; Howell and Lew 2012), DNA replication (Fragkos et al. 2015; Masai et al. 2010), and spindle formation, mitosis, and mitotic exit (Biggins 2013; Weiss 2012; Winey and Bloom 2012). We filtered the resulting gene lists by periodicity in S. cerevisiae (Figure 5.20 A, S6 Table). We then identified orthologous genes in C. neoformans without enforcing a periodicity filter.

We have previously shown that expression timing of canonical cell-cycle orthologs in S. cerevisiae and S. pombe can vary—some gene pairs shared expression patterns while others diverged (Orlando et al. 2007). To temporally align orthologous
gene plots between *S. cerevisiae* and *C. neoformans*, we used the algorithmic approach described previously with *S. cerevisiae* and *S. pombe* time series transcriptome data (Orlando et al. 2007). The first, most synchronous cycle of budding data from each yeast was fit using the CLOCCS algorithm (Figure 5.19, S6 Fig) (Orlando et al. 2007, 2009). Time points in minutes were then transformed into cell-cycle lifeline points to visualize the data (see S1 File).

As observed previously, *S. cerevisiae* genes that regulate budding, S-phase, and mitosis were largely transcribed periodically in the proper phases (Figure 5.22 A, D, and G) (Cho et al. 1998; Orlando et al. 2008; Pramila et al. 2006; Spellman et al. 1998). Cell-cycle gene expression peak time patterns were examined to quantitatively compare cell-cycle phases (S7 Fig). Bud assembly and growth genes peaked throughout the cell-cycle transcription program, and the temporal ordering of these genes repeated across cell cycles (Figure 5.22 A, S7A-B Fig). Similarly, spindle assembly and mitosis genes peaked in the mid-to-late phases of the transcription program (Figure 5.22 G). DNA replication genes peaked in a defined window in the middle phase of the transcription program (Figure 5.22 D). We observed analogous expression patterns for *C. neoformans* orthologs associated with S-phase and mitosis (Figure 5.22 E and H), but orthologs associated with budding appeared to be expressed with less restriction to a discrete cell-cycle phase or strict temporal order (S7 Fig). This budding gene pattern can be observed qualitatively where the unrestricted expression timing creates a more “speckled” appearance in the *C. neoformans* heatmap (Figure 5.22 B) and differentially timed gene expression peaks (Figure 5.22 C).
Figure 5.22: DNA replication, spindle assembly, and mitosis genes are highly conserved in temporal ordering during the fungal cell cycles, while budding orthologs vary in their temporal expression pattern in *C. neoformans*.

*S. cerevisiae* genes annotated as bud assembly and growth genes were identified and filtered by periodicity (77 genes) (A). Many budding genes had an ortholog in *C.
*C. neoformans* (61 genes, 79.2%), and some orthologs were labeled periodic (20 genes, 32.8%) (B). Genes annotated as DNA replication genes were identified and filtered by periodicity (61 genes) (D). Almost all DNA replication genes had an ortholog in *C. neoformans* (53 genes, 86.9%), and over half of the orthologs were labeled periodic (28 genes, 52.8%) (E). Genes annotated as mitotic and spindle assembly genes were also identified and filtered by periodicity (143 genes) (G). Over half of the M-phase genes had an ortholog in *C. neoformans* (87 genes, 60.8%), and many orthologs were called periodic (53 genes, 60.9%) (H). Transcript levels are depicted as a z-score change relative to mean expression for each gene, where values represent the number of standard deviations away from the mean. Orthologous periodic gene pairs are in the same order for (A-B, D-E, or G-H) (for exact ordering of gene pairs and multiple-mapping orthologs, see S6 Table). Each column represents a time point in minutes. Canonical budding (C), DNA replication (F), and mitotic (I) gene orthologs are plotted to compare transcript dynamics between *S. cerevisiae* (blue) and *C. neoformans* (green). Global alignment E-values for ortholog pairs can be found in S4 Table. Line plots for orthologs are shown on a mean-normalized scale (same linear scaling method as heatmaps) (C, F, and I). This mean-normalization was used because *C. neoformans* genes have higher fold-change expression levels than *S. cerevisiae* genes (S1 Fig). Orthologous genes are plotted on a common cell-cycle timeline in CLOCCS lifeline points as described (see S1 File). In both yeasts, S-phase genes generally precede M-phase genes in temporal order (D-F, G-I).

We hypothesize that bud emergence and bud growth are not as tightly coordinated with cell-cycle progression in *C. neoformans* cells. Unlike *S. cerevisiae* where bud emergence occurs primarily at the G1/S transition, *C. neoformans* bud emergence can occur in a broad interval from G1 to G2 phases (Ohkusu et al. 2001, 2004). The difference in budding transcript behaviors between *S. cerevisiae* and *C. neoformans* orthologs could therefore reflect the difference in the cell biology of bud emergence and growth (Figure 5.22 A-B). Only about 33% of the orthologous budding gene pairs were periodically expressed in *C. neoformans*, compared to 53% DNA replication and 61% mitosis orthologs (Figure 5.22 B, E, H). Furthermore, budding orthologs that were periodic in both *C. neoformans* and *S. cerevisiae* showed some divergence in expression timing (Figure 5.22 C). We also observed that bud emergence of *C. neoformans* cells during the time series appeared less synchronous in second and third cycles than *S. cerevisiae* cells (Figure 5.19 A-B). Bud emergence in *C. neoformans*
could be controlled by both stress pathways and TF inputs because the first budding cycle is highly synchronous after elutriation synchrony, which causes a transient stress response in released cells (Figure 5.19 B). However, our data do not rule out a model where some budding genes in *C. neoformans* are controlled post-transcriptionally by localization, phosphorylation, or other periodic mechanisms. It is also possible that budding orthologs are more difficult to identify than other cell-cycle genes due to sequence divergence or that novel budding genes have evolved in the *C. neoformans* lineage.

5.3.3 Partial conservation of the transcription factor (TF) network control module

We have previously shown that a network of periodically expressed TFs is capable of driving the program of periodic genes during the *S. cerevisiae* cell cycle (Orlando et al. 2008; Simmons Kovacs et al. 2012). We hypothesized that a network of periodic TFs could also function in *C. neoformans* to drive a similar fraction of cell-cycle genes. Thus, the temporal re-ordering of part of the *C. neoformans* gene expression program (Figure 5.21) could be explained by two models: evolutionary re-wiring of shared network TFs with *S. cerevisiae* or novel TF network components arising in *C. neoformans* to drive cell-cycle genes. First, we asked if network TFs were conserved from *S. cerevisiae* to *C. neoformans*. Indeed, a majority of network TFs and key cell-cycle regulators have putative orthology between the two yeasts (Table 5.8) (Medina et al. 2016). As observed for other cell-cycle genes (Figure 5.22), orthologs of some network TFs were expressed in the same phase in both yeasts, while others were expressed at different times (Table 5.8).
Table 5.8: TF network components in *S. cerevisiae* and sequence orthologs in *C. neoformans* have generally diverged in expression timing.

Putative orthologous gene pairs were identified, if any (S4 Table) (Medina et al. 2016). The peak time (minutes) and time to half-peak expression (minutes) was identified for the first cell cycle in each yeast. Peak times were similar for some pairs (e.g. SWI4, CNAG_07464), but many pairs have diverged in ordering (e.g. FHL1, CNAG_05934, and CNAG_05535). The protein global alignment score is also shown for each putative ortholog pair (S4 Table). Some reported ortholog pairs did not have a significant global alignment score (i.e. E-value > 10), which was likely due to similar local sequence matches (e.g. homologous protein domains) and divergent regions elsewhere in the proteins (see S1 File).

<table>
<thead>
<tr>
<th><em>S. cerevisiae</em> gene ID</th>
<th>Peak time (minutes)</th>
<th>Half peak time (min)</th>
<th><em>C. neoformans</em> gene ID</th>
<th>Peak time (minutes)</th>
<th>Half peak time (min)</th>
<th>Global E-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDH1</td>
<td>5</td>
<td>0</td>
<td>CNAG_03191</td>
<td>30</td>
<td>20</td>
<td>9.80E-101</td>
</tr>
<tr>
<td>SWI4</td>
<td>20</td>
<td>15</td>
<td>CNAG_07464</td>
<td>10</td>
<td>0</td>
<td>&gt;10</td>
</tr>
<tr>
<td>YOX1</td>
<td>20</td>
<td>15</td>
<td>CNAG_03229</td>
<td>70</td>
<td>10</td>
<td>&gt;10</td>
</tr>
<tr>
<td>CLN1/2</td>
<td>25,25,5</td>
<td>15,15,0</td>
<td>CNAG_06092</td>
<td>10</td>
<td>0</td>
<td>0.78; 2.90E-02</td>
</tr>
<tr>
<td>HCM1</td>
<td>25</td>
<td>15</td>
<td>CNAG_03116</td>
<td>30</td>
<td>20</td>
<td>&gt;10</td>
</tr>
<tr>
<td>STB1</td>
<td>25</td>
<td>15</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDC28</td>
<td>30</td>
<td>20</td>
<td>CNAG_01664</td>
<td>20</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>CLB3/4</td>
<td>35,45</td>
<td>25,30</td>
<td>CNAG_02095</td>
<td>30</td>
<td>20</td>
<td>1.2E-88; 4.40E-79</td>
</tr>
<tr>
<td>NDD1</td>
<td>40</td>
<td>25</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FHL1</td>
<td>45</td>
<td>25</td>
<td>CNAG_05934</td>
<td>0</td>
<td>0</td>
<td>&gt;10</td>
</tr>
<tr>
<td>FKH1</td>
<td>45</td>
<td>30</td>
<td>CNAG_05861</td>
<td>0</td>
<td>0</td>
<td>0.25</td>
</tr>
<tr>
<td>SWI6</td>
<td>45</td>
<td>25</td>
<td>CNAG_01438</td>
<td>10</td>
<td>0</td>
<td>1.50E-14</td>
</tr>
<tr>
<td>MBP1</td>
<td>45</td>
<td>30</td>
<td>CNAG_07464</td>
<td>10</td>
<td>0</td>
<td>6.80E-41</td>
</tr>
<tr>
<td>FHL1</td>
<td>45</td>
<td>25</td>
<td>CNAG_05535</td>
<td>50</td>
<td>40</td>
<td>&gt;10</td>
</tr>
<tr>
<td>YHP1</td>
<td>45</td>
<td>15</td>
<td>CNAG_03229</td>
<td>70</td>
<td>10</td>
<td>&gt;10</td>
</tr>
<tr>
<td>NRM1</td>
<td>45</td>
<td>20</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLB1/2/5/6</td>
<td>50,65,25,20</td>
<td>40,40,15,15</td>
<td>CNAG_04575</td>
<td>80</td>
<td>60</td>
<td>5.8E-72; 9.30E-91; &gt;10</td>
</tr>
<tr>
<td>WHI5</td>
<td>50</td>
<td>25</td>
<td>CNAG_05591</td>
<td>10</td>
<td>0</td>
<td>&gt;10</td>
</tr>
<tr>
<td>FKH2</td>
<td>55</td>
<td>30</td>
<td>CNAG_02566</td>
<td>10</td>
<td>0</td>
<td>2.90E-04</td>
</tr>
<tr>
<td>ACE2</td>
<td>60</td>
<td>40</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SWI5</td>
<td>60</td>
<td>45</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDC20</td>
<td>65</td>
<td>50</td>
<td>CNAG_07756</td>
<td>30</td>
<td>20</td>
<td>2.40E-90</td>
</tr>
<tr>
<td>MCM1</td>
<td>75</td>
<td>60</td>
<td>CNAG_07924</td>
<td>30</td>
<td>20</td>
<td>&gt;10</td>
</tr>
</tbody>
</table>
Second, we asked if there were any novel periodic TFs in *C. neoformans* (i.e. TFs with no predicted ortholog in *S. cerevisiae*, or TFs with an ortholog in *S. cerevisiae* that is not known to function in the TF network). We constructed a list of periodic *C. neoformans* TFs by filtering a previously annotated transcription factor list (Jung et al. 2015) with our list of periodic genes (Figure 5.23, S7 Table). Indeed, 30 novel TF genes were periodic during the *C. neoformans* cell cycle (Figure 5.23 A). Taken together, results from Table 5.8 and Figure 5.23 suggested that both network TF re-wiring and novel periodic TFs in *C. neoformans* could explain the differential ordering of periodic genes during the cell cycle (Figure 5.21).

A gene list of *C. neoformans* TFs was obtained (180 genes) (Jung et al. 2015) and filtered by periodicity (36 genes, 20.0%). 6/36 periodic TFs were putative orthologs to previously known TFs in the *S. cerevisiae* cell-cycle network (Table 5.8). The remaining 30 novel periodic TFs are shown (A). Transcript levels are depicted as a z-score change relative to mean expression for each gene, where values represent the number of standard deviations away from the mean. Each column represents a time point in minutes. 18/30 periodic TFs have no documented ortholog in *S. cerevisiae*. 12/30 periodic TFs do have a putative ortholog in *S. cerevisiae*, but that gene is not currently known to participate in the *S. cerevisiae* cell-cycle network (S7 Table). Three examples of these ortholog pairs are shown between periodic *C. neoformans* TFs and their
putative *S. cerevisiae* ortholog (B). Line plots for orthologs are shown on a mean-normalized scale (z-score of fpkm units, same linear scaling method as heatmaps) (B). This mean-normalization was used because *C. neoformans* genes have higher fold-change expression levels than *S. cerevisiae* genes (S1 Fig). Orthologous genes are plotted on a common cell-cycle timeline in CLOCCS lifeline points as described (see S1 File).

Putative S-phase regulators in *C. neoformans* exhibited transcript behaviors that were very similar in periodicity and in ordering to their *S. cerevisiae* orthologs (Figure 5.22 D-F). Thus, we predicted that the network motifs and TFs controlling the transcription of periodic S-phase genes could be conserved. Orthologous genes in the G1/S topology were largely conserved in periodic expression dynamics at cell-cycle entry (Figure 5.24). The expression timing of some genes had shifted earlier in the *C. neoformans* cell cycle (Figure 5.24 C, Figure 5.24 F, Table 5.8), but this result does not refute the hypothesis that these genes are activated and functional at G1/S phase. Therefore, the network topology of cell-cycle entry appeared largely conserved in *C. neoformans* both by sequence and by gene expression dynamics. The prediction of this model is that a common G1/S transcriptional network drives a common set of S-phase periodic genes. To test this model, we examined promoter sequences from TF network genes in *S. cerevisiae* and *C. neoformans*, as well as the promoters of 38 periodic DNA replication ortholog pairs, and did an unbiased search for enriched TF binding sequences. The core motif “ACGCGT” for SBF/MBF transcription factors (V. R. Iyer et al. 2001; Lowndes et al. 1991; McIntosh et al. 1991) was identified in both *S. cerevisiae* and *C. neoformans* promoters. The motif was not enriched in randomly selected periodic gene promoters, suggesting that SBF/MBF is functionally conserved in *C. neoformans* to drive TF network oscillations and DNA replication gene expression (S8 Fig).
Figure 5.24: Evidence for conservation of the TF network topology at G1/S in C. neoformans.

At cell-cycle entry in S. cerevisiae, the repressors Whi5 and Stb1 are removed from the SBF/MBF complexes by G1 cyclin/CDK phosphorylation. The heterodimeric TF complexes SBF (Swi4, Swi6) and MBF (Mbp1, Swi6) can then activate ~200 periodic genes at the G1/S border. SBF/MBF activate the downstream transcriptional activator Hcm1 to continue the temporal activation of S-phase genes. The transcriptional repressors Yox1, Yhp1, and Nrm1 then repress SBF/MBF (A). Ortholog pairs are shown for SBF/MBF (CNAG_07464 or MBS1) (B), SWI6 (CNAG_01438 or MBS2) (C), G1 cyclins (CNAG_06092) (D), HCM1 (CNAG_03116) (E), and WHI5 (CNAG_05591) (F). Line plots for orthologs are shown on a mean-normalized scale (z-score of fpkm units, same linear scaling method as heatmaps) (B-F). This mean-normalization was used because C. neoformans genes have higher fold-change expression levels than S. cerevisiae genes (S1 Fig). Orthologous genes are plotted on a common cell-cycle timeline in CLOCCS lifeline points as described (see S1 File).

5.4 Discussion

Here, we present the first RNA-Sequencing dataset of transcription dynamics during the cell cycle of C. neoformans. Despite evolutionary distance between Basidiomycota and Ascomycota, S. cerevisiae and its extensive genome annotation provided an excellent analytical benchmark to compare to cell-cycle transcription in C. neoformans. RNA-Sequencing has been shown to be more quantitative than microarray technology for lowly- and highly-expressed genes using asynchronous S. cerevisiae
cells due to microarray background fluorescence and saturation of fluorescence, respectively (Yassour et al. 2009).

We demonstrate that 20% or more of all genes in the budding yeast genomes are periodically transcribed during the cell cycle. A ranking of periodicity for transcript dynamics in *C. neoformans* is provided (S2 Table). For the sake of comparison, we have presented gene sets of 1100-1200 periodic genes with the highest relative periodicity scores as “cell-cycle-regulated”; however, there is a continuum of periodic gene expression dynamics during the cell cycle in both yeasts (S1 Fig). The four periodicity algorithms applied here yielded a range of periodicity scores with no clear distinction between “periodic” and “non-periodic” gene sets (S1-S2 Tables). These results suggest that yeast mRNAs fluctuate in expression with various degrees of cell-cycle periodicity.

We propose that the top 20% periodic genes presented in this study are directly regulated by periodic cell-cycle TFs in *C. neoformans* and in *S. cerevisiae*. We also posit that some of the remaining 80% genes are weakly cell-cycle regulated. For example, some genes could be subject to complex regulation with one regulatory input from a cell-cycle periodic TF and another input from a constitutively expressed TF.

We raise two important questions about the yeast periodic gene expression programs: is periodic expression of a core set(s) of genes required for the fungal cell cycle, and how are periodic gene dynamics controlled in each yeast?

In both yeasts, periodic transcription is a high dimensional cell-cycle phenotype because transcriptional state reflects the phase-specific biology of the cell cycle over repeated cycles (Figure 5.20, Figure 5.22). In other words, G1-, S-, and M-phase genes follow a defined temporal ordering pattern. *S. cerevisiae* cells synchronized by different methods and/or grown in different conditions display similar ordering of periodic cell-
cycle genes, despite different cell-cycle period lengths (S4 Fig). Here, we examined the transcriptome of cycling C. neoformans cells at 30°C. Other groups have shown that C. neoformans cells spend more time in G1 phase at 24°C (Yamaguchi et al. 2007). We predict that future studies examining cell-cycle transcription of C. neoformans cells grown in different conditions (i.e. non-rich media or 37°C infection temperature) would continue to display a similar temporal ordering of cell-cycle genes. These findings provide more evidence that “just-in-time transcription” is a conserved feature of eukaryotic cell cycles (Haase and Wittenberg 2014).

We show that some orthologous periodic genes have diverged in temporal ordering during the cell cycles of S. cerevisiae and C. neoformans over evolutionary time (Figure 5.21). We specifically investigated genes that play a role in bud emergence and bud growth, and we find that many budding gene orthologs are not controlled in a defined temporal order during the C. neoformans cell cycle (Figure 5.19 A-B and Figure 5.22 A-B). On the other hand, DNA replication and mitosis genes do appear to be conserved by sequence homology, periodic expression, and temporal ordering (Figure 5.22 D-I). Lastly, we find that a set of about 100 orthologous genes is both periodic and expressed in proper cell-cycle phase in the budding yeasts S. cerevisiae, C. neoformans, and C. albicans (S5 Fig) (Côte et al. 2009). These findings suggest that there may be a conserved set of fungal cell-cycle-control genes, which represent novel therapeutic targets for fungal infections.

We posit that a network of periodic transcription factors (TFs) could control the periodic gene expression program in C. neoformans, which has been shown in S. cerevisiae and suggested in human cells (Grant et al. 2013; Orlando et al. 2008; Simmons Kovacs et al. 2012; Simon et al. 2001). Many orthologous genes to S.
*cerevisiae* TF network components have diverged in expression timing in *C. neoformans* cells (Table 5.8). However, we show that the G1/S network topology is likely conserved between *S. cerevisiae* and *C. neoformans* because orthologous genes display similar expression dynamics (Figure 5.24). Furthermore, we find that the promoters of G1/S TF network orthologs and promoters of periodic DNA replication orthologs are enriched for an “ACGCGT” sequence motif, which matches the SBF/MBF binding site consensus in *S. cerevisiae* (S8 Fig) (V. R. Iyer et al. 2001; Lowndes et al. 1991; McIntosh et al. 1991). Therefore, we propose that the G1/S transcriptional motif—where a co-repressor is removed by G1 cyclin/CDK phosphorylation and a TF activator complex is derepressed—is also conserved in *C. neoformans* (Figure 5.24 B-D, G) (Bertoli et al. 2013; Medina et al. 2016). Downstream of the G1/S activator complex, the *C. neoformans* TF network may also contain a common forkhead domain S-phase activator and homeobox domain G1/S repressor (Figure 5.24 E, Table 5.8) (de Bruin et al. 2006; Pramila et al. 2002, 2006). This partially conserved TF network model in *C. neoformans* explains the common G1/S topology, on-time DNA replication gene transcription, as well as differential expression of budding and other cell-cycle genes by divergent parts of the TF network.

The regulation of periodic transcription and the function of a putative TF network warrant further investigation as virulence factors of fungal meningitis caused by *C. neoformans*. It has been previously shown that fluconazole drug treatment can affect cell ploidy in *C. neoformans* (Sionov et al. 2010). More recently, polyploid Titan cells were shown to produce haploid and aneuploid daughter cells during *C. neoformans* infection (Gerstein et al. 2015). Therefore, future work on proper regulation of DNA replication and the contribution of periodic gene products could greatly benefit our understanding of
genome stability in *C. neoformans*. The *C. neoformans* TF deletion collection was recently phenotyped, and the potential of targeted TF therapies was discussed (Bahn 2015; Jung et al. 2015). We have added to the *C. neoformans* genotype/phenotype map by documenting the functional outputs of cell-cycle TFs over synchronized cell cycles. We also propose that a conserved G1/S topology of cell-cycle TFs may initiate the cell-cycle transcription network in *C. neoformans*. It is possible that a multi-drug combination targeting cell-cycle regulators and previously characterized virulence pathways could yield more successful antifungal therapies (Bahn 2015). For example, a combination therapy could target TFs at the conserved G1/S topology to slow cell-cycle entry and also target fungal cell wall or capsule growth. In the circadian rhythm field, it has been shown that drugs targeting Clock Controlled Genes are most potent when administered at the time of the target gene’s peak expression (R. Zhang et al. 2014).

Interestingly, deletion of the known SBF/MBF ortholog, Mbs1 (CNAG_07464), is viable in *C. neoformans* (Jung et al. 2015; Song et al. 2012). These genetic results do not match *S. cerevisiae*, where *swi4 mbp1* double mutants are inviable (Koch et al. 1993). In fact, deletion of the single known G1 cyclin ortholog, CNAG_06092, is also viable in *C. neoformans* (García-Rodas et al. 2014). Mbs1 and the G1 cyclin are likely important for cell-cycle progression in *C. neoformans* because mutant phenotypes are highly defective in capsule formation in G1 phase, melanin production, and response to Hydroxyurea treatment during S phase (García-Rodas et al. 2014, 2015; Jung et al. 2015; Song et al. 2012). However, the genetics are inconsistent with findings in *S. cerevisiae* and warrant further investigation to characterize the G1/S TF network topology of *C. neoformans*. It is possible that uncharacterized, redundant genes exist in the *C. neoformans* G1/S network motif.
We find that 40 candidate virulence genes are periodically expressed during the *C. neoformans* cell cycle (S3 Table, S3 Fig). An important direction for future work is to identify the mechanistic links between cell-cycle regulators and virulence pathways. 14 periodic virulence genes have annotated phenotypes in capsule formation and/or cell wall secretion. Fungal cells must secrete new cell wall and capsule during growth, and the direct links between cell cycle and these virulence factors in *C. neoformans* warrants further study because the cell wall and capsule are not present in host cells. The ultimate goal of this work is to identify the regulatory mechanism of periodic gene expression in *C. neoformans* and to find optimal drug targets and combination therapies for disrupting the fungal cell cycle.

**5.5 Future Work**

We were pleased to be invited to write a short review about the significance of this manuscript in *Current Genetics*, entitled “Connecting virulence pathways to cell-cycle progression in the fungal pathogen *Cryptococcus neoformans*” (Kelliher and Haase 2017).

Future work on this project will be focused on identifying a complete transcription factor network (TFN) that regulates cell-cycle genes in *C. neoformans*. This collaborative work is currently underway by the groups of John Harer at Duke, Tomas Gedeon at Montana State University, and Konstantin Mischaikow at Rutgers University. Our quantitative collaborators are building a network inference pipeline tool for general use by systems biologists. The *S. cerevisiae* and *C. neoformans* yeast cell-cycle datasets will be used to test network-building tools and will be featured in this future publication.

Another useful direction for future work on understanding gene regulatory networks in *C. neoformans* will be to profile cell-cycle synchronized cells in non-rich
media and/or at high temperature, as it is already known that steady-state gene expression levels change in response to poor media conditions (Janbon et al. 2014).

After cell-cycle transcription factors and network motifs are identified in C. neoformans, future research will investigate TFN function by perturbing candidate nodes and measuring cell-cycle effects. We predict that perturbation of critical C. neoformans cell-cycle TFs will alter the rate of cell-cycle progression (Simmons Kovacs et al. 2012). The Heitman lab has kindly offered access to the C. neoformans TF single deletion collection (Jung et al. 2015). With nearly 500 million years of evolution separating the C. neoformans (Basidiomycota) and S. cerevisiae (Ascomycota) budding yeasts, this project offers fascinating insights into the question: can evolution operate on networks of interacting TFs? For example, we can compare cell-cycle TFN topologies between the two yeasts to ask if convergent evolution is arriving on specific network motifs, and/or if there is there selective pressure to maintain the cell-cycle network structure of the budding yeasts' common ancestor to produce strikingly similar cell-cycle gene expression dynamics.

5.6 Materials and Methods

5.6.1 Yeast strains, cultures, and synchronization

The wild-type Saccharomyces cerevisiae strain is a derivative of BF264-15D MATa bar1 (Cole and Reed 1991; Lew et al. 1992). The wild-type Cryptococcus neoformans var. grubii serotype A strain is a derivative of H99F (Janbon et al. 2014). Yeast cultures were grown in standard YEP medium (1% yeast extract, 2% peptone, 0.012% adenine, 0.006% uracil supplemented with 2% dextrose sugar). For centrifugal elutriation, cultures were grown in YEP-dextrose (YE PD) medium at 30°C overnight. Elutriated early G1 cells were then resuspended in fresh YEPD medium at 30°C for time
series experiments. For α-factor arrest, cultures were grown in YEPD medium at 30°C and incubated with 30 ng/ml α-factor for about 110 minutes. Synchronized cultures were then resuspended in fresh YEPD medium at 30°C. Aliquots were taken at each time point and subsequently assayed by RNA-Sequencing.

5.6.2 RNA isolation and RNA-Sequencing analyses

Total RNA was isolated by acid phenol extraction as described previously (Leman et al. 2014). Samples were submitted to the Duke Sequencing Facility (https://www.genome.duke.edu/cores-and-services/sequencing-and-genomic-technologies) for stranded library preparation and sequencing. mRNA was amplified and barcoded (Illumina TruSeq Stranded mRNA Library Preparation Kit for S. cerevisiae and KAPA Stranded mRNA-Seq Library Preparation Kit for C. neoformans) and reads were sequenced in accordance with standard Illumina HiSeq protocols. For S. cerevisiae, libraries of 50 base-pair single-end reads were prepared, and 10 samples were multiplexed and sequenced together in each single lane. For C. neoformans, libraries of 125 base-pair paired-end reads were prepared (due to larger and more complex yeast transcriptome with introns), and 12 samples were multiplexed and sequenced together in each single lane. Raw FASTQ files were aligned to the respective yeast genomes using STAR (Dobin et al. 2013). Aligned reads were assembled into transcripts, quantified, and normalized using Cufflinks2 (Trapnell et al. 2013). Samples from each yeast time series were normalized together using the CuffNorm feature. The normalized output FPKM gene expression levels were used in the analyses presented. A detailed description of each analysis pipeline is presented in the S1 File.
6. Developmental Cycle Transcription and Period Control in Laboratory Strains of the Malaria Parasite *Plasmodium falciparum*

6.1 Author Contributions

In this Chapter, we investigate the developmental cycle transcription program of *Plasmodium falciparum*, the deadliest species that causes malaria in humans. Unlike the *S. cerevisiae* and *C. neoformans* fungi discussed in previous Chapters, much less is known about the complete genome structure or about transcriptional regulation in *Plasmodium* species. Thus, the data described in this Chapter represent an unknown case with which we can test gene regulatory network inference pipelines developed on yeast cell-cycle datasets in future work.

This malaria project is highly collaborative and spans multiple institutions. Steve Haase oversees and manages collaborations and project directions. Garima Chopra and LTC Norman Waters at Walter Reed Army Institute of Research (WRAIR) performed the time series experiments and malaria developmental staging described here. Adam Leman extracted RNA from the WRAIR samples and coordinated Biosafety Level 2 laboratory protocols. I analyzed the raw RNA-Sequencing data from all experiments and produced the figures herein. Hayden Walcott, an undergraduate student in Duke Biology, and I compiled a list of putative transcription factor genes in *P. falciparum*. Our quantitative collaborators John Harer, Anastasia Deckard, Francis Motta (Duke Mathematics), Tomas Gedeon, Bree Cummins (Montana State Mathematics), and Konstantin Mischaikow (Rutgers Mathematics) are currently building gene regulatory networks from the data described here. Finally, our genomics collaborators Greg Wray,
Shauna Morrow, Ben Redelings, and Madison Rogers (Duke Biology) are currently working on PacBio sequencing and annotating the *P. falciparum* laboratory strains described in this Chapter.

### 6.2 Introduction

Malaria is a significant global health concern with about 200 million new cases each year, which lead to approximately 500,000 deaths. A majority of malaria fatalities occur in Africa and are caused by the malaria species *Plasmodium falciparum* (World Health Organization, Malaria Report 2016). Preventative measures like mosquito netting and pesticides have reduced the number of malaria cases over the past decade (Bhatt et al. 2015). However, malaria eradication cannot be achieved with pesticides and netting only, vaccines to date have not been effective, and drug resistance is increasingly problematic in the clinic.

Most malaria cases in humans are caused by three species: *Plasmodium falciparum*, *P. vivax*, and *P. malariae*. After infection, clinical presentation of malaria includes a periodic fever that occurs every 48 hours (for *P. falciparum* and *P. vivax* infection) or 72 hours (*P. malariae* infection). During the malaria developmental cycle, parasites invade red blood cells, proliferate asexually, and burst to infect new red blood cells, which triggers the periodic host fevers. The major stages of this developmental cycle are ring stage, trophozoite stage, schizont stage, and merozoites to infect new red blood cells. Different malaria species infect a variety of mammalian hosts, yet the developmental cycle period length almost always exists as a ratio to the host circadian rhythm of 24 hours (reviewed in: Mideo et al. 2013).

Preliminary evidence in mouse and in human patients suggests that malaria parasites actively synchronize with host circadian rhythms and that malaria parasites
proliferate more effectively when synchronized. Two populations of mice were entrained to 12-hour periods of light and dark conditions at opposite times (12:12 standard and 12:12 reverse light conditions) and were then infected with a rodent species of malaria, *Plasmodium chabaudi* (24-hour developmental cycle). After parasites synchronized to the host mice in both populations, new populations of circadian-matched and circadian-reverse mice were cross-infected with synchronous *P. chabaudi*. Parasites that were out-of-phase with the host rhythms performed significantly worse during the course of the infection (O’Donnell et al. 2011). This study demonstrated that mammalian host rhythms are important for malaria parasites to monitor for efficient infection. One case study with 4 human patients also directly tested the model that *P. malariae* parasites synchronize their developmental cycle with host circadian rhythm (Young et al. 1940). In all four individuals infected with *P. malariae*, fevers were observed after parasites appearing in the segmenter stage at about 9 AM in the morning. After synchronous fevers were observed, the sleep cycle of all four patients was reversed, and two of four patients were kept in constant light (free-running circadian) conditions. Within two to three *P. malariae* developmental cycles (6-9 days), parasites shifted about 12 hours and re-synchronized to the new host rhythm. Although this case study is limited to only four individuals, the data indicate that *P. malariae* parasites actively synchronize to human circadian rhythms during infection.

The red blood cell developmental period length in malaria species correlates with the 24-hour circadian rhythm, and this ratio relationship appears to be important for an effective malaria infection (reviewed in: Mideo et al. 2013). To better understand the biology of the asexual cycle, researchers have been able to culture *Plasmodium falciparum* in red blood cells in the laboratory for more than four decades (Trager and
Jensen 1976). The developmental cycle period length of some contemporary *P. falciparum* laboratory strains is altered when grown in vitro compared to the 48-hour cycle period seen in the clinic. Given the importance of maintaining period length and synchronizing with the 24-hour host circadian rhythm, we sought to understand the genetics behind altered period length in vitro for *P. falciparum* laboratory strains.

The *P. falciparum* sequencing project revealed that this species has an extremely AT-rich genome (Gardner et al. 2002). It has also been challenging to identify gene homologs compared to other sequenced eukaryotic genomes. Some repetitive and AT-rich regions of the *P. falciparum* genome were recently resolved with long-read PacBio DNA sequencing efforts (Vembar et al. 2016). The mutation rate in *P. falciparum* has been estimated from natural isolates (Manske et al. 2012; Volkman et al. 2007) and from laboratory cultures (Bopp et al. 2013; A. Miles et al. 2016; Mu et al. 2010). Therefore, genomic resources are readily available for *P. falciparum* (http://plasmodb.org/plasmo/; https://www.malariagen.net/projects/pf3k), but orthology to well-understood genetic model organisms and transcriptome annotation in malaria is an ongoing challenge for the field.

Previous transcriptome profiling work in *P. falciparum* has revealed a large program of periodic gene expression during the red blood cell developmental cycle. At least 60% of all measured genes were expressed dynamically in the HB3 laboratory strain with very dense sampling over time; however, early generation microarrays led to many missing data points from this study (Bozdech et al. 2003). The developmental transcriptomes of three *P. falciparum* laboratory strains (HB3, Dd2, and 3D7) were comparatively analyzed in previous work, but these datasets were also generated with early custom microarrays (Llinás et al. 2006). More recent studies have used RNA-
Sequencing to profile the 3D7 laboratory strain over time, but these studies were sampled less densely than original profiling studies (Broadbent et al. 2015; Otto et al. 2010). Taken together, these studies agree that many genes are periodically expressed during the asexual cycle, suggesting that regulation of transcription plays an important role in malaria development.

Here, we compare the transcriptome dynamics of two laboratory strains of \textit{P. falciparum} grown in vitro that cycle at short period (39 hours) and a more typical infection period length (45 hours). Our time series studies utilized a dense sampling scheme (at least 10 time points per cycle) in order to measure dynamics of the developmental cycle, along with quantitative microscopy staging data for each time point (Materials and Methods).

\textbf{6.3 Results}

\textit{P. falciparum} 3D7 and FVO-NIH strains were synchronized, released, and sampled every three hours over about 1.5 developmental cycles (Materials and Methods). Progression through ring, trophozoite, and schizont stages was monitored by microscopy over time. We profiled the time series samples by stranded RNA-Sequencing. To identify periodic genes, normalized time series datasets were run through periodicity-ranking algorithms as described previously (Deckard et al. 2013). There are 5713 annotated ORFs in the \textit{P. falciparum} 3D7 reference genome, and 4911 genes (86\%) in our 3D7 dataset scored better than a Lomb-Scargle algorithm cutoff of 0.5 (this cutoff was previously applied to identify periodic genes in yeast cell-cycle data: (Bristow et al. 2014)). Our findings support previous work that a majority of genes in the \textit{P. falciparum} genome are periodically activated during the red blood cell developmental cycle (Figure 6.25).
Figure 6.25: A large program of periodic gene expression in the *Plasmodium falciparum* laboratory strains 3D7 (short, 39-hour period) and FVO-NIH (approximately wild-type, 45-hour period).

Periodicity-ranking algorithms were run on the 3D7 dataset (with the first 2 time points removed due to presumed stress response after synchronization) at a period of 40 hours, as described previously (Deckard et al. 2013). The Lomb-Scargle algorithm was used to filter top periodic genes in 3D7 data (score ≤ 0.5), which produced a list of 4911 annotated genes. In both of the 3D7 (A) and FVO-NIH (B) heatmaps, the 4911 periodic genes were ordered along the y-axis in the same order by peak time of expression in the 3D7 dataset. Transcript levels are depicted as a z-score change relative to mean expression for each gene, where values represent the number of standard deviations away from the mean. Each row represents transcript levels of a unique gene across the time series. Each column represents a time point in hours. Parasite developmental staging (ring, trophozoite, or schizont) was counted by microscopy for each laboratory strain over time.
The *P. falciparum* 3D7 strain has a shortened ring stage compared to the FVO-NIH strain by both microscopic staging data and by gene expression signatures (Figure 6.25). We posit that, during an in vivo infection, the ring stage may be an important developmental window where parasites are sensing and synchronizing with a host circadian signal(s). In rodent models, peak levels of melatonin (a circadian-regulated hormone) correspond with late schizont stage in vivo, and parasites mount a cellular response to melatonin in vitro (Hotta et al. 2000) (reviewed in: Lima et al. 2013). Therefore, we propose that parasite cells may sense host circadian signals other than melatonin during ring stage.

In both 3D7 and FVO-NIH laboratory strains, a large program of periodic genes was observed with very similar peak time ordering over the developmental cycle in red blood cells (Figure 6.25). This finding agrees with previous comparative transcriptomics on the HB3, Dd2, and 3D7 laboratory strains (Llinás et al. 2006). We hypothesized that a common set of regulators drives this common program of gene expression in *P. falciparum*. Thus, we mined the literature and bioinformatics databases for putative genes that encode *P. falciparum* transcription factors (Table 6.9).
Table 6.9: 92 putative sequence-specific transcription factors and cyclin/CDK genes in *Plasmodium falciparum*.

Putative TF genes (columns 1-2) were found using bioinformatics and literature searches (column 3). Putative cyclin and cyclin dependent kinase (CDK) genes are also listed. The malaria developmental cycle in red blood cells involves multiple mitotic divisions to grow and produce more parasites. Therefore, we reasoned that cyclin/CDK genes would be directly involved in the developmental cycle as well as transcription factors that control the many periodic genes observed in each stage (*Figure 6.25*).

<table>
<thead>
<tr>
<th>Standard ID</th>
<th>Other IDs</th>
<th>Literature Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF3D7_0109600</td>
<td>PFA0470c</td>
<td></td>
</tr>
<tr>
<td>PF3D7_0308100</td>
<td>PFC0345w</td>
<td>(Bischoff and Vaquero 2010)</td>
</tr>
<tr>
<td>PF3D7_0407600</td>
<td>PFD0375w</td>
<td></td>
</tr>
<tr>
<td>PF3D7_0409800</td>
<td>PFD0485w</td>
<td></td>
</tr>
<tr>
<td>PF3D7_0517900</td>
<td>PFE0895c</td>
<td></td>
</tr>
<tr>
<td>PF3D7_0522900</td>
<td>PFE1145w</td>
<td></td>
</tr>
<tr>
<td>PF3D7_0525000</td>
<td>PFE1245w</td>
<td></td>
</tr>
<tr>
<td>PF3D7_0602000</td>
<td>PFF0095c</td>
<td></td>
</tr>
<tr>
<td>PF3D7_0602200</td>
<td>PFF0105w</td>
<td></td>
</tr>
<tr>
<td>PF3D7_0607100</td>
<td>PFF0350w</td>
<td></td>
</tr>
<tr>
<td>PF3D7_0729700</td>
<td>PF07 0124</td>
<td></td>
</tr>
<tr>
<td>PF3D7_0818100</td>
<td>MAL8P1.70</td>
<td></td>
</tr>
<tr>
<td>PF3D7_0906600</td>
<td>PFI0325c</td>
<td></td>
</tr>
<tr>
<td>PF3D7_0909700</td>
<td>PFI0470w</td>
<td></td>
</tr>
<tr>
<td>PF3D7_0927200</td>
<td>PFI1335w</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1008600</td>
<td>PF10 0083</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1009400</td>
<td>PF10 0091</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1019300</td>
<td>PF10 0186</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1119300</td>
<td>PF11 0200</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1134600</td>
<td>PF11 0357</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1205500</td>
<td>PFL0275w</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1209100</td>
<td>PFL0455c</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1210200</td>
<td>PFL0510c</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1216900</td>
<td>PFL0815w</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1243100</td>
<td>PFL075c</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1307900</td>
<td>MAL13P1.37</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1308100</td>
<td>PF13 0043</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1309800</td>
<td>PF13 0054</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1358600</td>
<td>PF13 0314</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1425600</td>
<td>PF14 0236</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1439500</td>
<td>PF14 0374</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1440400</td>
<td>PF14 0383</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1443800</td>
<td>PF14 0416</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1450400</td>
<td>PF14 0479</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1458600</td>
<td>PF14 0559</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1464200</td>
<td>PF14 0610</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1464400</td>
<td>PF14 0612</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1468400</td>
<td>PF14 0652</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1468900</td>
<td>PF14 0657</td>
<td></td>
</tr>
<tr>
<td>Gene ID</td>
<td>Description</td>
<td>References</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------</td>
<td>----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>PF3D7_1473900</td>
<td>PF14_0707</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1146600</td>
<td>PfNF-YB; PF11_0477</td>
<td>(Bischoff and Vaquero 2010; Lima et al. 2013)</td>
</tr>
<tr>
<td>PF3D7_0611200</td>
<td>PFF0550w</td>
<td></td>
</tr>
<tr>
<td>PF3D7_0622900</td>
<td>PFF1100c</td>
<td></td>
</tr>
<tr>
<td>PF3D7_0934400</td>
<td>PF1665w</td>
<td>(Bischoff and Vaquero 2010; Campbell et al. 2010)</td>
</tr>
<tr>
<td>PF3D7_1305200</td>
<td>PF13_0026</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1350900</td>
<td>PF13_0267</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1456000</td>
<td>PF14_0533</td>
<td></td>
</tr>
<tr>
<td>PF3D7_0317200</td>
<td>Pfcrk-4</td>
<td></td>
</tr>
<tr>
<td>PF3D7_0415300</td>
<td>Pfcrk-3</td>
<td></td>
</tr>
<tr>
<td>PF3D7_0417800</td>
<td>Pfcrk-1</td>
<td></td>
</tr>
<tr>
<td>PF3D7_0518400</td>
<td>Pf cyc-3</td>
<td>cyclin/CDKs</td>
</tr>
<tr>
<td>PF3D7_0615500</td>
<td>Pfcrk-5</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1014400</td>
<td>Pfmrk</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1227500</td>
<td>Pfyc-c2</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1337100</td>
<td>PIPK6</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1356900</td>
<td>PIPK5</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1304700</td>
<td>Pfyc-4; PF13_0022</td>
<td>cyclin/CDKs; (Bischoff and Vaquero 2010)</td>
</tr>
<tr>
<td>PF3D7_1463700</td>
<td>Pfyc-1; PF14_0605</td>
<td>cyclin/CDKs</td>
</tr>
<tr>
<td>PF3D7_0604100</td>
<td>PFF0200c</td>
<td>(Bischoff and Vaquero 2010; Campbell et al. 2010; De Silva et al. 2008)</td>
</tr>
<tr>
<td>PF3D7_1466400</td>
<td>PF14_0633</td>
<td>(Bischoff and Vaquero 2010; Campbell et al. 2010; De Silva et al. 2008; L. M. Iyer et al. 2008)</td>
</tr>
<tr>
<td>PF3D7_1315800</td>
<td>PIMyb1; PF13_0088</td>
<td>(Bischoff and Vaquero 2010; Gissot et al. 2005)</td>
</tr>
<tr>
<td>PF3D7_0315800</td>
<td>PFC0690c; MAL3P5.22</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1033600</td>
<td>PF10_0327</td>
<td>(Bischoff and Vaquero 2010; L. M. Iyer et al. 2008)</td>
</tr>
<tr>
<td>PF3D7_1123000</td>
<td>PF11_0241</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1205800</td>
<td>PFL0290w</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1225200</td>
<td>PFL1215c</td>
<td></td>
</tr>
<tr>
<td>PF3D7_0404100</td>
<td>PFD0200c</td>
<td></td>
</tr>
<tr>
<td>PF3D7_0516800</td>
<td>PFE0840c</td>
<td></td>
</tr>
<tr>
<td>PF3D7_0613800</td>
<td>PFF0670w</td>
<td></td>
</tr>
<tr>
<td>PF3D7_0802100</td>
<td>MAL8P1.153</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1007700</td>
<td>PF10_0075</td>
<td>(Bischoff and Vaquero 2010; Campbell et al. 2010; L. M. Iyer et al. 2008)</td>
</tr>
<tr>
<td>PF3D7_1107800</td>
<td>PF11_0091</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1115500</td>
<td>PF11_0163</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1143100</td>
<td>PF11_0442</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1222400</td>
<td>PFL1075w</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1239200</td>
<td>PFL1900w</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1317200</td>
<td>PF13_0097</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1408200</td>
<td>PF14_0079</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1429200</td>
<td>PF14_0271</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1449500</td>
<td>PF14_0471</td>
<td></td>
</tr>
<tr>
<td>PF3D7_0420300</td>
<td>PFD0985w</td>
<td>(Bischoff and Vaquero 2010; Campbell et al. 2010; Chaubey et al. 2014; L. M. Iyer et al. 2008)</td>
</tr>
<tr>
<td>PF3D7_0730300</td>
<td>PF07_0126</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1139300</td>
<td>PF11_0404</td>
<td></td>
</tr>
<tr>
<td>PF3D7_1342900</td>
<td>PF13_0235</td>
<td></td>
</tr>
</tbody>
</table>
PF3D7_1222600   PFL1085w; PfAP2-G  (Bischoff and Vaquero 2010; Campbell et al. 2010; L. M. Iyer et al. 2008; Kafsack et al. 2014)

PF3D7_1209300   PFL0465c  (Bischoff and Vaquero 2010; L. M. Iyer et al. 2008; Jurgeleinaite et al. 2009; Reilly Ayala et al. 2010)

PF3D7_1014600   PF10_0143  (Bischoff and Vaquero 2010; L. M. Iyer et al. 2008; Jurgelenaite et al. 2009)

PF3D7_1033700   BDP1  (Josling et al. 2015)

PF3D7_1212900   BDP2  (Josling et al. 2015)

PF3D7_1011800   NA  (Komaki-Yasuda et al. 2013)

PF3D7_0506800   NA  NA

PF3D7_0530600   NA  NA

There is some disagreement in the malaria field about the significance of general transcription mechanisms (chromatin regulators, epigenetic marks, RNA binding proteins, etc.) versus sequence-specific TFs in controlling gene expression during the asexual cycle. Sequence-specific TFs in *Plasmodium* species appear to be diverged and/or evolutionarily lost compared to characterized TFs in other annotated eukaryotic genomes (e.g. only about 50 TFs were originally identified by standard comparative bioinformatics (Bischoff and Vaquero 2010)). Gene expression variation in clonal *P. falciparum* parasites has been attributed to epigenetic modifications (Rovira-Graells et al. 2012). Therefore, it is likely that sequence-specific TFs and more general mechanisms both affect transcription dynamics during the red blood cell stage.

Laboratory strains of *P. falciparum* complete the red blood cell cycle with slightly different period lengths in vitro (Figure 6.25). We hypothesize that laboratory culturing conditions are permissive for mutations to arise that shorten period length. In future work, we will identify mutations in laboratory strains of *P. falciparum* and further investigate the genetic control of cycle period length. Genetic screens for mutations that alter period length have identified critical transcription factor regulators in the circadian clock network (reviewed in: E. E. Zhang and Kay 2010) and in the *Saccharomyces*
cerevisiae cell-cycle network (Simmons Kovacs et al. 2012). In the malaria field, one previous study identified laboratory-specific mutations in a gene that encodes a transcription factor (Claessens et al. 2017). A QTL study was performed on two parental laboratory strains of *P. falciparum* with different cycle period lengths, and regions of association were identified (Reilly Ayala et al. 2010). A third study of clinical *P. vivax* mutations showed that one particular transcription factor gene underwent strong selective pressure in the recent past (Parobek et al. 2016). We are in the process of DNA sequencing the *P. falciparum* 3D7 and FVO-NIH strains from this study using long-read PacBio technology as in (Vembar et al. 2016). With these genomic datasets, we will search for putative causative mutations (with a focus on transcription factor genes, Table 6.9) that associate with cycle period length in each strain.

### 6.4 Discussion

We compared periodic genes between of two laboratory strains of *P. falciparum*—the 3D7 strain with a short period length (39 hours) and the FVO-NIH strain with a period length closer to in vivo cycles (45 hours). We demonstrated that the ring stage is shortened in 3D7 compared to the FVO-NIH strain (Figure 6.25). Two additional *P. falciparum* laboratory strains were also synchronized and profiled every three hours over about 1.5 developmental cycles. The transcriptome data for SA-250 (~48 hours) and D6 (~35 hours) will be available in the near future in order to ask if the ring stage is significantly shorter in the short period strains (3D7 and D6) compared to strains with a period length closer to wild type (FVO-NIH and SA-250).

Using our curated list of sequence-specific TFs (Table 6.9), we are currently working with quantitative collaborators to build networks of TFs that can explain the periodic gene expression dynamics during red blood cell development, building on
previous work (Cai et al. 2013). To test the validity of these predicted networks, future work will genetically perturb candidate core TFs and ask if cycle period length, gene expression dynamics, and/or parasite fitness are altered in vitro.

DNA sequence data will also be available for the four laboratory strains of \textit{P. falciparum} described in this study. We will use this genomic data to ask if mutations in putative regulatory TFs (Table 6.9) associate with short period (3D7 and D6 strains) versus long period length (FVO-NIH and SA-250), which is closer to clinical strains. As proof of principle, one previous QTL study, which compared the \textit{P. falciparum} laboratory strains Dd2 (short period), HB3 (long period), and 34 progeny from a genetic cross, identified a significantly-associated genomic region containing a TF gene (Reilly Ayala et al. 2010). A genetic screen also identified many mutations associated with lengthened developmental cycle periods in \textit{P. falciparum} (Balu et al. 2010).

Future work can leverage many high-throughput datasets to further characterize transcriptome regulation and organization in \textit{Plasmodium} species. Transcription start sites have been mapped (Adjalley et al. 2016), RNA polymerase II was localized (Rai et al. 2014), nucleosome occupancy was mapped (Kensche et al. 2015; Ponts et al. 2010), chromatin marks were profiled (Bártfai et al. 2010; Gupta et al. 2013), chromosome organization was mapped by chromosome conformation capture (Ay et al. 2014), RNA-binding protein-associated transcripts were identified (Bunnik et al. 2016), polysomes were profiled (Bunnik et al. 2013; Caro et al. 2014), and anti-sense transcription was characterized (Broadbent et al. 2015; Siegel et al. 2014) in \textit{Plasmodium falciparum}. Protein abundance, localization, and/or modification (Florens et al. 2002; Foth et al. 2011; Le Roch et al. 2004; Oehring et al. 2012; B. N. Pease et al. 2013) should also be
considered when predicting transcriptome organization mechanisms and gene regulatory networks.

Malaria drug and vaccine therapies are often developed using laboratory strains of *Plasmodium falciparum*, which have had variable success (Neafsey et al. 2015). Clinical isolates of malaria are known to adapt and accumulate mutations in response to laboratory culturing conditions, which could affect the ability to make vaccines from *P. falciparum* laboratory strains (Claessens et al. 2014; Yeda et al. 2016). Vaccination of genetically modified parasites has shown more success in promoting host immunity and a protective response to malaria infection (Kublin et al. 2017). Our unique approach to malaria therapy is to first understand the basic biology of developmental cycle period control. If we can identify and then disrupt parasite synchrony mechanisms to host circadian rhythms, we hope to improve malaria treatment regiments and increase the efficacy of host immune response to out-of-phase parasites.

### 6.5 Future Work

The data described in this Chapter will be part of a future manuscript featuring the four laboratory strains of *Plasmodium falciparum* and the range of period lengths observed in vitro (short: 3D7 and D6; approximately wild type: FVO-NIH and SA-250). We plan to incorporate both transcriptomic evidence (shorter ring stage) and genomic evidence (mutations in putative TF genes) to explain the difference in cycle period lengths. Projects of similar scope and impact have been published in the following journals, which we will likely target: PNAS, Cell Host & Microbe, Genome Research, or PLOS Pathogens.
Our quantitative collaborators are currently working on a continuation project from the work described in this Chapter: inferring putative gene regulatory networks that explain transcriptome dynamics in \textit{P. falciparum}.

\section*{6.6 Materials and Methods}

\subsection*{6.6.1 \textit{Plasmodium falciparum} strains, cultures, and synchronization}

\textit{P. falciparum} parasites were synchronized by alanine treatment and temperature cycling as described (Haynes and Moch 2002). For the 3D7 strain experiment, we started sampling (0 hour time point) at the first invasion after synchronization (early ring stage). For the FVO-NIH strain experiment, we began sampling (0 hour time point) earlier at the schizont stage after synchronization in order to capture the transition to rings at the beginning of the experiment (i.e. our “time 0” sampling protocol was modified after observing variability in invasion times between \textit{P. falciparum} laboratory strains). The FVO-NIH time series experiment was split into two pools of cultures after synchronization to avoid reaching stressful, high parasitemia levels over time: time points 0-33 hours were sampled from culture ‘A’, which progressed from \(~0.1 - ~4\%\) parasitemia after time 0; time points 36-69 hours were sampled from culture ‘B’, which progressed from \(~0.01 - ~5\%\) parasitemia after time 0. Blood smears from each time point were counted and developmentally staged for both the 3D7 and FVO-NIH experiments. Stage-specific parasitemia levels over time were quantified using flow cytometry.

\subsection*{6.6.2 RNA isolation, sequencing, and analysis}

Total RNA was isolated from packed RBCs by acid phenol extraction as described (Leman et al. 2014). Total RNA samples were submitted to the Duke
Sequencing Facility for library preparation and sequencing (https://www.genome.duke.edu/cores-and-services/sequencing-and-genomic-technologies). The 21 *P. falciparum* 3D7 samples were prepared in libraries of 125 base-pair, paired-end reads (Stranded Total RNA Kapa kits with rRNA Ribo-Zero Globin depletion), and 7 samples were multiplexed together for sequencing in each of the 3 lanes using an Illumina HiSeq 2000/2500 instrument. The 24 *P. falciparum* FVO-NIH samples were prepared in libraries of 50 base-pair, single-end reads (Stranded Total RNA Illumina TruSeq kits with rRNA Ribo-Zero Globin depletion), and 12 samples were multiplexed together for sequencing in each of the 2 lanes using an Illumina HiSeq 4000 instrument. Reads were sequenced in accordance with standard Illumina HiSeq protocols.

Raw RNA-Sequencing data from this Chapter will be submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) upon publication. Raw FASTQ files were aligned to the reference genome using STAR (Dobin et al. 2013). *P. falciparum* intron size parameters for mapping were set to range from 10-3000 bps (Sorber et al. 2011). The publically available *P. falciparum* 3D7 reference genome version 3.1 was downloaded from the Wellcome Trust Sanger Institute (http://www.sanger.ac.uk/resources/downloads/protozoa/plasmodium-falciparum.html) in March 2016 from:


Aligned reads were assembled into *P. falciparum* annotated transcripts, quantified, and normalized using Cufflinks (Trapnell et al. 2013). Samples from each time series were normalized together using the CuffNorm feature. The normalized fpkm
gene expression outputs ("genes.fpkm_table") were used in the analyses presented. To avoid fractional and zero values, 1 was added to every fpkm value in each dataset using the R statistical programming environment (R Core Development Team 2017). Fractions and zeros were found to interfere with the periodicity algorithms that involved log-transformation of data points (data not shown). Alignment and transcript quantification algorithms were modified from default parameters. Below is a generalized summary of the UNIX/Linux analysis pipeline, with file paths and generic file names in italics.

6.6.3 RNA-Sequencing data analysis pipeline

Step 1. Align reads:

STAR --runThreadN 1 --runMode alignReads --genomeDir path_to_Pf_genome_build --readFilesIn sample.fastq --outFilterType BySJout --alignIntronMin 10 --alignIntronMax 3000 --outFileNamePrefix ./STAR_out/ --outFilterIntronMotifs RemoveNoncanonical

Step 2. Quantify transcripts:

cuffquant --library-type=fr-firststrand path_to_Pf_transcriptome_gtf sample.aligned.sorted.sam

Step 3. Normalize time series samples together:

cuffnorm --library-type=fr-firststrand path_to_Pf_transcriptome_gtf *.cxb
7. Conclusions: Building, Testing, and Refining Gene Regulatory Networks

Biological clock networks are used to time a particular process such as the cell cycle, circadian rhythm, and parasite development (Figure 1.1). In this dissertation, I have provided further evidence that clock networks in biology contain transcription factors, feedback motifs, and redundant genes to increase network robustness. The Saccharomyces cerevisiae cell cycle network is composed of dynamic proteins (Chapter 3) and negative feedback motifs (Chapter 2). A distantly related fungal pathogen species, Cryptococcus neoformans, expresses about 20% of all genes periodically during the cell cycle and may initiate the program of transcription from a conserved topology of TFs and cyclin/CDKs at G1/S phase (Chapter 5). The human malaria parasite Plasmodium falciparum develops cyclically in red blood cells, and over 80% of the transcriptome is activated periodically during this cycle (Chapter 6). Together with quantitative modeling, experimental work described in this dissertation has furthered our knowledge about the cell cycle and the red blood cell developmental cycle in malaria (Figure 7.26).
Figure 7.26: Periodic gene expression programs are an emergent property of biological clock networks, which are driven by networks of transcription factors and other regulators.
During the *S. cerevisiae* cell cycle, periodic gene expression is shown from synchronized cells (845 genes, 12.9% of the genome, data from: (Kelliher et al. 2016)) (A). During the *C. neoformans* cell cycle, periodic transcription is shown (981 genes, 14.4% of the genome, data from: (Kelliher et al. 2016)) (B). During the human cell cycle, periodic gene expression also occurs (1495 genes, 3.9% of the genome, K562 data: unpublished) (C). The G1/S network topology in *S. cerevisiae, C. neoformans*, and human cells is functionally conserved in structure and in dynamical signatures of genes (A-C). Line plots for *S. cerevisiae* genes (*CLN1*, *SWI4*, *YHP1*, and *NRM1*), *C. neoformans* genes (CNAG_06092, CNAG_07464, CNAG_05591, and CNAG_05176), and K562 genes (cyclin E, E2F1, p107, and E2F5) are shown from each dataset. Solid edges are supported by experimental evidence, and dashed edges are predicted but not validated (A-C). During the mouse circadian cycle, liver cells display a periodic gene expression program (1079 genes, 4.6% of probe IDs, data from: (Hughes et al. 2009)) (D). Line plots are shown for *M. musculus* core clock genes (*Bmal1/Arntl* and *Per2*). During the *P. falciparum* developmental cycle, almost the entire transcriptome is periodic in vitro (4617 genes, 85.4% of the genome, data: unpublished) (E). Line plots are shown for *P. falciparum* genes (PF3D7_1139300, PF3D7_1239200, and PF3D7_1143100). Solid edges indicate experimental evidence for TF binding sites present in the promoter of target TF genes, Figure S12 from: (Campbell et al. 2010). For each heatmap (A-E), genes are displayed relative to their respective expression mean.
Over the course of this dissertation research, our quantitative group has developed analysis tools to build gene regulatory networks from time series data. The nodes, or genes participating in the network, must be identified first. For the yeast cell cycle and eukaryotic circadian rhythm networks, node selection was completed as described (Chapter 1.1-1.3) with experiments, bioinformatics, and identifying evolutionarily related genes. Additionally, our quantitative group has approached the node selection problem by identifying TF genes that are themselves periodically expressed over the cycle of interest. We rigorously compared four algorithms for scoring periodic gene expression (Deckard et al. 2013) and developed a fifth algorithm (Perea et al. 2015). Each periodicity-ranking algorithm prefers slightly different curve shape signatures, and we took advantage of this algorithmic diversity to identify a high confidence periodic gene set in both Saccharomyces cerevisiae and Cryptococcus neoformans (Chapter 5) (Kelliher et al. 2016).

The next step in constructing gene regulatory networks is to identify edges between nodes. Edges between TFs and target promoters are typically supported by evidence from DNA binding data and/or genetic TF perturbation experiments, and both data types are readily available for the budding yeast S. cerevisiae and for the murine circadian rhythm from liver tissue (Koike et al. 2012; Teixeira et al. 2014). Our quantitative group has also developed an algorithm to predict local regulatory interactions between a given TF and a downstream target gene directly from time series data (McGoff et al. 2016). After an interconnected gene regulatory network of nodes and edges is built, a critically important step is to ask if network simulations fit the time series gene expression data used to build the network model.
Fitting gene regulatory network models to global dynamics from time series data is an ongoing problem in the field. Many tools are limited by network size (i.e. number of nodes and edges) due to the number of parameters introduced by each additional node or edge in the model. Our quantitative collaborators have developed a tool that is capable of predicting global dynamics for networks up to about 10 nodes total (Cummins et al. 2016). Future work in network inference will have two major goals: computationally streamline a pipeline that moves from time series data to plausible global network model(s) and then test gene regulatory network models by perturbing core nodes and predicting dynamical signatures. Core network nodes can be investigated using traditional genetic perturbation approaches (e.g. experiments presented in Chapter 2), designing perturbations in silico followed by modeling altered network dynamics, or as described more recently, multiplexing gene perturbations and profiling single-cell dynamics (Dixit et al. 2016).

Networks of genes can be conserved from the ancestral network (e.g. G1/S phase in budding yeasts, Figure 7.26 A-B) or convergent evolution can arrive on a similar network structure (e.g. G1/S phase in fungi and animal cells, Figure 7.26 A-C) (Medina et al. 2016). In this dissertation, I have provided evidence that the budding yeasts *C. neoformans* and *S. cerevisiae* (separated by 500 million years of evolution) execute cell-cycle Start through a similar network topology. The network structure of G1/S phase in fungal and animal cells is functionally similar, but many participating proteins are not evolutionarily related. Therefore, it appears that evolution can operate on entire networks and/or motifs to arrive at similar dynamical solutions with unrelated protein components (reviewed in: Bertoli et al. 2013).
In future work, we will elucidate the network structure of the red blood cell developmental cycle of *P. falciparum* using data from four laboratory strains grown in vitro. The goal of this work is to understand the genetics of cycle period control and to predict the underlying gene regulatory network structure in *P. falciparum*. With this *P. falciparum* network in hand, we can then predict the structure of networks from other species of *Plasmodium*, which nearly all maintain a red blood cell cycle period length that is a ratio of 24 hours (reviewed in: Mideo et al. 2013). Specifically, we are interested in two other species of malaria: *P. knowlesi* that infects rhesus macaque (*Macaca mulatta*) with a 24-hour cycle and *P. vivax* that infects humans with a 48-hour cycle time. *P. falciparum*, *P. vivax*, and *P. knowlesi* are separated by millions of years of evolution, where *P. vivax* and *P. knowlesi* are more closely related (Loy et al. 2016). Based on results from our fungal cell-cycle work, we are confident that understanding the red blood cell network in *P. falciparum* will assist us in predicting the control mechanisms for dynamics in *P. vivax* and *P. knowlesi*.

Collaborators at Emory University have investigated the dynamics of gene expression during the red blood cell stage in *P. knowlesi* and are interested in discovering the gene regulatory network (Lapp et al. 2015). We will also investigate dynamics in vivo from humans infected with *P. vivax* in collaboration with AFRIMS (Armed Forces Research Institute of Medical Sciences) in Thailand. *P. vivax* parasites express many genes periodically in vitro, with similar dynamics and similar ordering to *P. falciparum* cycles (Bozdech et al. 2008). Of the putative clock genes in *P. falciparum* (Table 6.9), 87 of 92 have orthologous genes in both *P. vivax* and *P. knowlesi*. Our preliminary network inference pipeline on *P. falciparum* (Figure 7.26 E) identified 8 AP2 TFs as nodes and multiple feasible edges with experimental evidence for TF binding.
sites (Campbell et al. 2010). A majority of those \textit{P. falciparum} TFs (7/8) are conserved in both \textit{P. vivax} and \textit{P. knowlesi} (Table 7.10).

Table 7.10: 87 putative sequence-specific transcription factors and cyclin/CDK genes in \textit{Plasmodium falciparum} have orthologous genes in \textit{P. vivax} and \textit{P. knowlesi}.

The list of 92 putative network genes was obtained from Table 6.9. Orthologous genes were identified using the PlasmoDB database (http://plasmodb.org/plasmo/). Almost all orthologous genes are also related by genomic synteny (except for the multiple orthologs identified for PF3D7_1222400, PF3D7_1317200, and PF3D7_1222600). The preliminary gene regulatory network in \textit{P. falciparum} contains eight AP2 TF genes (Figure 7.26), which are shown here in blue font.

<table>
<thead>
<tr>
<th>\textit{Plasmodium falciparum} (3D7)</th>
<th>\textit{Plasmodium vivax} (P01 strain)</th>
<th>\textit{Plasmodium knowlesi} (strain H)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF3D7_0109600</td>
<td>PVP01_0205500</td>
<td>PKNH_0204200</td>
</tr>
<tr>
<td>PF3D7_0308100</td>
<td>PVP01_0832100</td>
<td>PKNH_0834600</td>
</tr>
<tr>
<td>PF3D7_0407600</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>PF3D7_0409800</td>
<td>PVP01_0309100</td>
<td>PKNH_0307900</td>
</tr>
<tr>
<td>PF3D7_0517900</td>
<td>PVP01_1016000</td>
<td>PKNH_1015300</td>
</tr>
<tr>
<td>PF3D7_0522900</td>
<td>PVP01_1011000</td>
<td>PKNH_1010000</td>
</tr>
<tr>
<td>PF3D7_0525000</td>
<td>PVP01_1008800</td>
<td>PKNH_1007800</td>
</tr>
<tr>
<td>PF3D7_0602000</td>
<td>PVP01_1146900</td>
<td>PKNH_1148700</td>
</tr>
<tr>
<td>PF3D7_0602200</td>
<td>PVP01_1146700</td>
<td>PKNH_1148500</td>
</tr>
<tr>
<td>PF3D7_0607100</td>
<td>PVP01_1141900</td>
<td>PKNH_1143200</td>
</tr>
<tr>
<td>PF3D7_0729700</td>
<td>PVP01_0215400</td>
<td>PKNH_0214600</td>
</tr>
<tr>
<td>PF3D7_0818100</td>
<td>PVP01_0517200</td>
<td>PKNH_0501900</td>
</tr>
<tr>
<td>PF3D7_0906600</td>
<td>PVP01_0705100</td>
<td>PKNH_0704400</td>
</tr>
<tr>
<td>PF3D7_0909700</td>
<td>PVP01_0708200</td>
<td>PKNH_0707700</td>
</tr>
<tr>
<td>PF3D7_0927200</td>
<td>PVP01_0725700</td>
<td>PKNH_0725500</td>
</tr>
<tr>
<td>PF3D7_1008600</td>
<td>PVP01_0808300</td>
<td>PKNH_0807600</td>
</tr>
<tr>
<td>PF3D7_1009400</td>
<td>PVP01_0809400</td>
<td>PKNH_0809100</td>
</tr>
<tr>
<td>PF3D7_1019300</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>PF3D7_1119300</td>
<td>PVP01_0919900</td>
<td>PKNH_0917000</td>
</tr>
<tr>
<td>PF3D7_1134600</td>
<td>PVP01_0935400</td>
<td>PKNH_0932800</td>
</tr>
<tr>
<td>PF3D7_1205500</td>
<td>PVP01_1304600</td>
<td>PKNH_1305200</td>
</tr>
<tr>
<td>PF3D7_1209100</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>PF3D7_1210200</td>
<td>PVP01_1309400</td>
<td>PKNH_1310200</td>
</tr>
<tr>
<td>PF3D7_1216900</td>
<td>PVP01_1435700</td>
<td>PKNH_1436500</td>
</tr>
<tr>
<td>PF3D7_1243100</td>
<td>PVP01_1460200</td>
<td>PKNH_1462400</td>
</tr>
<tr>
<td>PF3D7_1307900</td>
<td>PVP01_1408900</td>
<td>PKNH_1408300</td>
</tr>
<tr>
<td>PF3D7_1308100</td>
<td>PVP01_1409100</td>
<td>PKNH_1408500</td>
</tr>
<tr>
<td>PF3D7_1309800</td>
<td>PVP01_1410800</td>
<td>PKNH_1410200</td>
</tr>
<tr>
<td>PF3D7_1358600</td>
<td>PVP01_1113300</td>
<td>PKNH_1112900</td>
</tr>
<tr>
<td>PF3D7_1425600</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>PF3D7_1439500</td>
<td>PVP01_1267500</td>
<td>PKNH_1242800</td>
</tr>
<tr>
<td>PF3D7_1440400</td>
<td>PVP01_1266600</td>
<td>PKNH_1241900</td>
</tr>
<tr>
<td>PF3D7</td>
<td>PVP01</td>
<td>PKNH</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>1443800</td>
<td>1263200</td>
<td>1238500</td>
</tr>
<tr>
<td>1450400</td>
<td>1256600</td>
<td>1231700</td>
</tr>
<tr>
<td>1458600</td>
<td>1248300</td>
<td>1217100</td>
</tr>
<tr>
<td>1464200</td>
<td>1238400</td>
<td>1212800</td>
</tr>
<tr>
<td>1468900</td>
<td>1237900</td>
<td>1212300</td>
</tr>
<tr>
<td>1473900</td>
<td>1233000</td>
<td>1207200</td>
</tr>
<tr>
<td>1146600</td>
<td>1138000</td>
<td>1129200</td>
</tr>
<tr>
<td>1142900</td>
<td>1126600</td>
<td>1127300</td>
</tr>
<tr>
<td>1350900</td>
<td>1203900</td>
<td>1234300</td>
</tr>
<tr>
<td>1456000</td>
<td>1250900</td>
<td>1225800</td>
</tr>
<tr>
<td>0317200</td>
<td>0822900</td>
<td>0733100</td>
</tr>
<tr>
<td>0415300</td>
<td>0521800</td>
<td>0506900</td>
</tr>
<tr>
<td>0417800</td>
<td>0527200</td>
<td>0510200</td>
</tr>
<tr>
<td>0518400</td>
<td>1015500</td>
<td>1014800</td>
</tr>
<tr>
<td>0615500</td>
<td>1133600</td>
<td>1134500</td>
</tr>
<tr>
<td>1014400</td>
<td>0814500</td>
<td>0814500</td>
</tr>
<tr>
<td>1227500</td>
<td>1445700</td>
<td>1446900</td>
</tr>
<tr>
<td>1337100</td>
<td>1217600</td>
<td>1264200</td>
</tr>
<tr>
<td>1356900</td>
<td>1115000</td>
<td>1114600</td>
</tr>
<tr>
<td>1304700</td>
<td>1405600</td>
<td>1404900</td>
</tr>
<tr>
<td>1463700</td>
<td>1243100</td>
<td>1217800</td>
</tr>
<tr>
<td>0604100</td>
<td>1144800</td>
<td>1146400</td>
</tr>
<tr>
<td>1466400</td>
<td>1240400</td>
<td>1215100</td>
</tr>
<tr>
<td>1315800</td>
<td>1416700</td>
<td>1416500</td>
</tr>
<tr>
<td>0315800</td>
<td>0824300</td>
<td>0826300</td>
</tr>
<tr>
<td>1033600</td>
<td>0618700</td>
<td>0618400</td>
</tr>
<tr>
<td>1123000</td>
<td>0923600</td>
<td>0920800</td>
</tr>
<tr>
<td>1205800</td>
<td>1304900</td>
<td>1305500</td>
</tr>
<tr>
<td>1225200</td>
<td>1443400</td>
<td>1444600</td>
</tr>
<tr>
<td>0404100</td>
<td>0303400</td>
<td>0302100</td>
</tr>
<tr>
<td>0516800</td>
<td>1017200</td>
<td>1016500</td>
</tr>
<tr>
<td>0613800</td>
<td>1135300</td>
<td>1136200</td>
</tr>
<tr>
<td>0802100</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>1007700</td>
<td>0807400</td>
<td>0806500</td>
</tr>
<tr>
<td>1107800</td>
<td>0908600</td>
<td>0905400</td>
</tr>
<tr>
<td>1115500</td>
<td>0916300</td>
<td>0913100</td>
</tr>
<tr>
<td>1143100</td>
<td>0943800</td>
<td>0941000</td>
</tr>
<tr>
<td>1222400</td>
<td>1440800; 1418100; 0724600; 0908100; 0936900; 0947400</td>
<td>1441700; 1417900; 072400; 0904900; 0934200; 0944600</td>
</tr>
<tr>
<td>1239200</td>
<td>1457300</td>
<td>1459000</td>
</tr>
<tr>
<td>1317200</td>
<td>1418100; 1440800; 0724600; 0908100; 0936900; 0947400</td>
<td>1417900; 1441700; 072400; 0904900; 0934200; 0944600</td>
</tr>
<tr>
<td>1408200</td>
<td>1340800</td>
<td>1350200</td>
</tr>
<tr>
<td>1429200</td>
<td>1320200</td>
<td>1329500</td>
</tr>
<tr>
<td>PF3D7</td>
<td>PVP01</td>
<td>PKNH</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td>--------</td>
</tr>
<tr>
<td>1449500</td>
<td>1257500</td>
<td>1232600</td>
</tr>
<tr>
<td>0420300</td>
<td>0529800</td>
<td>0513100</td>
</tr>
<tr>
<td>0730300</td>
<td>0216000</td>
<td>0215200</td>
</tr>
<tr>
<td>1139300</td>
<td>0940100</td>
<td>0937300</td>
</tr>
<tr>
<td>1342900</td>
<td>1211900</td>
<td>1258500</td>
</tr>
<tr>
<td>1222600</td>
<td>1440800; PVP01_1418100; PVP01_0908100; PVP01_0936900; PVP01_0947400</td>
<td>PKNH_1441700; PKNH_1417900; PKNH_0724000; PKNH_0904900; PKNH_0934200; PKNH_0944600</td>
</tr>
<tr>
<td>1209300</td>
<td>1308400</td>
<td>1309200</td>
</tr>
<tr>
<td>1014600</td>
<td>0814700</td>
<td>0814700</td>
</tr>
<tr>
<td>1033700</td>
<td>0618800</td>
<td>0618500</td>
</tr>
<tr>
<td>1212900</td>
<td>1431800</td>
<td>1432300</td>
</tr>
<tr>
<td>1011800</td>
<td>0811900</td>
<td>0811600</td>
</tr>
<tr>
<td>0506800</td>
<td>1027200</td>
<td>1026800</td>
</tr>
<tr>
<td>0530600</td>
<td>1003300</td>
<td>1002100</td>
</tr>
</tbody>
</table>

I propose that network conservation enables future research to compare dynamics between species and hypothesize putative network structures in uncharacterized biological systems. Fungal genetic models are among the best-understood biological systems. Our quantitative group has begun to build cell-cycle networks in the fungal pathogen *Cryptococcus neoformans* using both network inference tools and comparative approaches to *Saccharomyces cerevisiae* (Chapter 5). These tools will also be applied to developmental network inference in the understudied human pathogens *Plasmodium falciparum*, *P. vivax*, and *P. knowlesi* (Chapter 6). In this way, genetic model organisms will not only be useful for comparing putative gene functions between related species but also for comparing gene regulatory network structures to design informed future experiments in non-model organisms and uncharacterized biological systems.
References


151


Bertoli C, Skotheim JM, de Bruin RAM. (2013). Control of cell cycle transcription during G1 and S phases. *Nature Reviews Molecular Cell Biology, 14*(8), 518–528. doi:10.1038/nrm3629


Bloom J, Cross FR. (2007). Multiple levels of cyclin specificity in cell-cycle control. *Nature Reviews Molecular Cell Biology, 8*(2), 149–160. doi:10.1038/nrm2105


157


Edenberg ER, Mark KG, Toczyski DP. (2015). Ndd1 Turnover by SCFGrr1 Is Inhibited by the DNA Damage Checkpoint in Saccharomyces cerevisiae. *PLOS Genetics, 11*(4), e1005162. doi:10.1371/journal.pgen.1005162


Elliott SG, McLaughlin CS. (1978). Rate of macromolecular synthesis through the cell cycle of the yeast Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences, 75*(9), 4384–8. doi:10.1073/pnas.75.9.4384


Goranov AI, Cook M, Ricicova M, Goranov AI, Cook M, Ricicova M, et al. (2009). The
rate of cell growth is governed by cell cycle stage. *Genes & Development*, 23(12), 1408–1422. doi:10.1101/gad.1777309


161


Hildebrandt ER, Hoyt MA. (2001). Cell cycle-dependent degradation of the Saccharomyces cerevisiae spindle motor Cin8p requires APC(Cdh1) and a bipartite destruction sequence. *Molecular Biology of the Cell*, 12(11), 3402–16. doi:10.1091/mbc.12.11.3402


human height. *Nature*, 467(7317), 832–838. doi:10.1038/nature09410


Pramila T, Miles S, GuhaThakurta D, Jemiolo D, Breeden LL. (2002). Conserved homeodomain proteins interact with MADS box protein Mcm1 to restrict ECB-dependent transcription to the M/G1 phase of the cell cycle. *Genes & Development*, 16(23), 3034–3045. doi:10.1101/gad.1034302


doi:10.1038/ncomms8075


Biography

Christina (Tina) Marie Kelliher was born in Saratoga Springs, New York on June 16, 1989. She grew up in Massachusetts and attended Walpole High School. In the fall of 2007, Tina matriculated at Gettysburg College in Pennsylvania, where she went on to major in Biochemistry and Molecular Biology and conduct a thesis project in the laboratory of Dr. Steven James. In her undergraduate research, Tina investigated DNA double strand break repair in the filamentous fungus, *Aspergillus nidulans*. In May 2011, she graduated fourth overall in her class, *summa cum laude*, and with a Bachelor of Science degree.

Tina joined the University Program in Genetics and Genomics at Duke University of Durham, North Carolina in August 2011. She completed her graduate work in Dr. Steven Haase’s laboratory. She published one of her research projects in PLOS Genetics in 2016 in an article titled “Investigating Conservation of the Cell-Cycle-Regulated Transcriptional Program in the Fungal Pathogen, *Cryptococcus neoformans*.” Tina and Steve Haase also wrote an invited review for Current Genetics in 2017 entitled: “Connecting virulence pathways to cell-cycle progression in the fungal pathogen *Cryptococcus neoformans*.” After completing her Ph.D. in April 2017, Tina will continue research in the fungal genetics field as a post-doctoral associate.