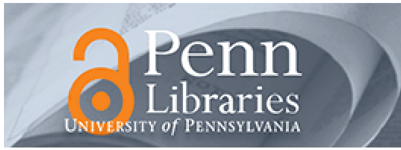

University of Pennsylvania
ScholarlyCommons

Publicly Accessible Penn Dissertations

1-1-2013

Effects of Temperature on Global Gene Expression in Natural Strains of Budding Yeast

Hoa Giang
University of Pennsylvania, hoagiang@sas.upenn.edu

Follow this and additional works at: http://repository.upenn.edu/edissertations

Part of the Biology Commons

Recommended Citation
Giang, Hoa, "Effects of Temperature on Global Gene Expression in Natural Strains of Budding Yeast" (2013). *Publicly Accessible Penn Dissertations*. 866.
http://repository.upenn.edu/edissertations/866

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/edissertations/866
For more information, please contact libraryrepository@pobox.upenn.edu.

Effects of Temperature on Global Gene Expression in Natural Strains of Budding Yeast

Abstract
Natural biological systems are resilient from the simplest form of unicellular organisms to the most complex form of multi-organ organisms. This resilience of a system manifests itself in two ways: "returns to its current attractor or moves to a new attractor that maintains the system's functions" (Hiroaki Kitano, 2004). That is, a system can work to maintain its current state or change to a new state that allows it to properly function under perturbations. One such complex system is the regulation of gene expression in biological organisms in which recruitment of transcriptional machinery to gene regulatory regions activates and controls transcription of target genes. Systemic responses of gene expression to perturbations result in alteration or stability of gene expression in individual genes as well as the state of cellular functions.

The objective of this work is to investigate the consequences of temperature perturbation on genome-wide gene expression with respect to cellular growth in two contrasting attributes: variation and robustness. We first characterize variation of genome-wide gene expression across five temperature conditions in three Saccharomyces cerevisiae strains--two natural strains and one laboratory strain--and investigate potential regulatory mechanisms of this expression variation. We show that as many as half of the number of genes in the genome exhibit expression variation but this gene expression variation is mostly specific to each strain. However, the global transcriptome displays a simple linear response to the temperature gradient manifested as a one-dimensional subspace, suggesting a global coordination of transcription against temperature perturbation.

Next, we characterize the robustness of genome-wide gene expression against temperature perturbation and compare it against the genetic differences in gene expression among these three strains. We provide evidence to support a hypothesis that selective forces potentially drive congruent evolution of genetic and temperature robustness of genome-wide gene expression. We present results to support the hypothesis that greater selection for gene expression robustness against temperature perturbation occurs in the natural strains compared to the laboratory strain, and that the evolution of gene expression robustness likely involves trans-factors.

In summary, we propose that a global regulatory coordination of transcription via trans-factors likely modulates genome-wide gene expression in relation to growth-permissive perturbations and drives congruent evolution of genetic robustness in the unicellular eukaryote, Saccharomyces cerevisiae.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Biology

This dissertation is available at ScholarlyCommons: http://repository.upenn.edu/edissertations/866

First Advisor
Junhyong Kim

Keywords
budding yeast, gene expression

Subject Categories
Biology

This dissertation is available at ScholarlyCommons: http://repository.upenn.edu/edissertations/866

EFFECTS OF TEMPERATURE ON GLOBAL GENE EXPRESSION IN NATURAL STRAINS OF BUDDING YEAST

Hoa Giang

A DISSERTATION

in

Biology

Presented to the Faculties of the University of Pennsylvania

in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2013

Supervisor of Dissertation

*Signature*_______________

Junhyong Kim, Ph.D., Professor of Biology

Graduate Group Chairperson

*Signature*_______________

Doris Wagner, Ph.D., Professor of Biology

Dissertation Committee

Paul Sniegowski, Ph.D., Professor of Biology (Dissertation Committee Chair)

Mark Goulian, Ph.D., Professor of Biology

Michael Lampson, Ph.D., Associate Professor of Biology

Arjun Raj, Ph.D., Assistant Professor of Bioengineering

EFFECTS OF TEMPERATURE ON GLOBAL GENE EXPRESSION

IN NATURAL STRAINS OF BUDDING YEAST

COPYRIGHT

2013

Hoa Giang

ACKNOWLEDGEMENTS

I would first like to sincerely thank my advisor, Dr. Junhyong Kim. My words are not enough to deliver what Junhyong is capable of as an exceptional mentor. Junhyong has been the constant source of guidance and inspiration to nurture me throughout this enduring marathon.

I would also like to express my gratitude to my thesis committee: Dr. Paul Sniegowski, Dr. Mark Goulian, Dr. Michael Lampson, Dr. Arjun Raj and Dr. Vijay Kumar. Without their encouragement and challenges, I would not be able to learn so many things and develop my research throughout the years.

When I joined the lab, I had the fortune to work with Dan Simola, who has guided me through the interface of high-throughput studies: how to design an experiment, how to carry out statistical tests and how to interpret large-scale datasets. I am deeply indebted to Dan for his inspirational spirit.

I want to thank many members in the lab who has shared their interesting and motivating discussions: Stephen, Miler, Sheng, Shreedhar, Chantal, Mugdha, Hannah, Sarah. You all make the lab a wonderful environment to work and explore sciences. Specifically, Stephen has supplied everything we need to do science, including yoga classes at your studio to help us getting well physically.

Last but not least, I would like to thank my wife and my parents for endless support and encouragement.

ABSTRACT

EFFECTS OF TEMPERATURE ON GLOBAL GENE EXPRESSION

IN NATURAL STRAINS OF BUDDING YEAST

Hoa Giang

Junhyong Kim

Natural biological systems are resilient from the simplest form of unicellular organisms to the most complex form of multi-organ organisms. This resilience of a system manifests itself in two ways: "returns to its current attractor or moves to a new attractor that maintains the system's functions" (Hiroaki Kitano, 2004). That is, a system can work to maintain its current state or change to a new state that allows it to properly function under perturbations. One such complex system is the regulation of gene expression in biological organisms in which recruitment of transcriptional machinery to gene regulatory regions activates and controls transcription of target genes. Systemic responses of gene expression to perturbations result in alteration or stability of gene expression in individual genes as well as the state of cellular functions.

The objective of this work is to investigate the consequences of temperature perturbation on genome-wide gene expression with respect to cellular growth in two contrasting attributes: variation and robustness. We first characterize variation of genome-wide gene expression across five temperature conditions in three *Saccharomyces cerevisiae* strains—two natural strains and one laboratory strain—and investigate potential regulatory mechanisms of this expression variation. We show that as many as half of the number of genes in the genome exhibit expression variation but this gene expression variation is mostly specific to each strain. However, the global transcriptome displays a simple linear response to the temperature gradient manifested as a one-dimensional subspace, suggesting a global coordination of transcription against temperature perturbation.

Next, we characterize the robustness of genome-wide gene expression against temperature perturbation and compare it against the genetic differences in gene expression among these three strains. We provide evidence to support a hypothesis that selective forces potentially drive congruent evolution of genetic and temperature robustness of genome-wide gene expression. We present results to support the hypothesis that greater selection for gene expression robustness against temperature perturbation occurs in the natural strains compared to the laboratory strain, and that the evolution of gene expression robustness likely involves *trans*-factors.

In summary, we propose that a global regulatory coordination of transcription via *trans*-factors likely modulates genome-wide gene expression in relation to growth-permissive perturbations and drives congruent evolution of genetic robustness in the unicellular eukaryote, *Saccharomyces cerevisiae*.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

ABSTRACT iv

TABLE OF CONTENTS vi

LIST OF TABLES viii

LIST OF FIGURES x

Chapter 1: INTRODUCTION 1

1.1 Regulation of genome-wide gene expression under temperature perturbation 2

1.2 Overview of phenotypic robustness...5

1.3 Evolution of phenotypic robustness...7

1.4 Robustness of gene expression in *Saccharomyces cerevisiae* 9

1.5 Dissertation structure 11

References..13

Chapter 2: VARIATION OF GENOME-WIDE GENE EXPRESSION IN *SACCHAROMYCES CEREVISIAE* REVEALS GLOBAL SCALING PATTERN UNDER TEMPERATURE PERTURBATION 17

2.1 Introduction 17

2.2 Growth rate shows a linear response to temperature change but not relative timing of the G1/S transition 19

2.3 Genome-wide gene expression exhibits distinct steady-state responses to temperature perturbation...20

2.3.1 Temperature perturbation induces steady-state expression responses in a large number of genes, including antisense expression 21

2.3.2 Steady-state expression responses to temperature perturbation are unrelated to transient expression responses to heat shock 25

2.4 Strain-specific expression responses to temperature reflect the adaptation of natural strains to their microenvironment 26

2.5 Steady-state expression responses show strain-specific differences in temperature response curves 31

2.6 Global pattern of steady-state gene expression displays major one-dimensional linear subspace of temperature responses....................35
2.7 Discussion 36
2.8 Materials and methods 41
References....................64

Chapter 3: ROBUSTNESS OF GENOME-WIDE GENE EXPRESSION IN NATURAL STRAINS OF THE BUDDING YEAST 67

3.1 Introduction 67
3.2 Statistical association between temperature and genetic variance of genome-wide gene expression supports the congruent evolution of these two variance components 70
3.3 Reduced expression variation in the natural strains with respect to the laboratory strain is consistent with selection for robustness of gene expression 77
3.3.1 Natural strains exhibit greater temperature robustness compared to the laboratory strain 78
3.3.2 *F*-statistics reveal reduced expression variation in the natural strains with respect to the laboratory strain 79
3.4 Molecular evolution analysis indicates the potential role of *trans* factors in reduced expression variation 81
3.5 Discussion 85
3.6 Materials and methods 87
References....................112

Chapter 4: CONCLUSIONS 115

References 120

LIST OF TABLES

Table 2.1 Gene Ontology (GO) terms overrepresented in genes with significant temperature effects in natural strains 54

Table 2.2 Gene Ontology (GO) terms overrepresented in genes with significant temperature effects at the G1/S checkpoint 55

Table 2.3 Gene Ontology (GO) terms overrepresented in genes with significant temperature effects at the M/G1 checkpoint 56

Table 2.4 Gene Ontology (GO) terms overrepresented in genes with significant temperature effects at 18°C and the G1/S checkpoint 57

Table 2.5 Gene Ontology (GO) terms overrepresented in genes with significant temperature effects at 34°C and the G1/S checkpoint 58

Table 2.6 Gene Ontology (GO) terms overrepresented in genes with significant temperature effects at 34°C and the M/G1 checkpoint 59

Table 2.7 Genes showing significant strain-specific temperature effects in the three yeast strains 60

Table 2.8 Evolutionary rates in the 5' *cis*-element, CDS, and 3' *cis*-element regions in strain-specific responsive genes. Strain-specific responsive genes are genes showing significant temperature effects in specific strains. Common genes are strain-specific responsive genes shared in all three strains (Figure 2.3) 61

Table 2.9 Evolutionary rates in the 5' *cis*-element, CDS, and 3' *cis*-element regions in the transcription factors of two classes of genes (the common responsive genes—TF-common, and the strain-specific responsive genes—TF-specific), the general transcription factors (TF-general) and non-transcription factors (Non-TFs)........ 62

Table 2.10 Number of genes showing linear and quadratic responses in the three yeast strains. 63

Table 3.1 Mean temperature and genetic variance in different classes of genes........ 93

Table 3.2 Number of genes showing decreased and increased temperature variation in the natural strains with respect to the laboratory strain. 94

Table 3.3 Means and confidence intervals of evolutionary rates in the coding sequence (dN/dS), 5' *cis*-element (d5/dS), and 3' *cis*-element (d3/dS) regions in the genes showing robust expression against temperature perturbations in both natural strains at both checkpoints 95

Table 3.4 Means and confidence intervals of evolutionary rates in the coding sequence (dN/dS), 5' *cis*-element (d5/dS), and 3' *cis*-element (d3/dS) regions in the genes showing robust expression against temperature perturbations solely in YPS2055 .. 96

Table 3.5 Means and confidence intervals of evolutionary rates in the coding sequence (dN/dS), 5' *cis*-element (d5/dS), and 3' *cis*-element (d3/dS) regions in the genes showing robust expression against temperature perturbations solely in YPS2073 .. 97

Table 3.6 List of genes expressing robustness in YPS2055 at the G1/S checkpoint........ 98

Table 3.7 List of genes expressing robustness in YPS2055 at the M/G1 checkpoint 100

Table 3.8 List of genes expressing robustness in YPS2073 at the G1/S checkpoint...... 103

Table 3.9 List of genes expressing robustness in YPS2073 at the M/G1 checkpoint 106

Table 3.10 Gene Ontology (GO) terms overrepresented in genes with robust expression in natural strains at both checkpoints.. 109

Table 3.11 Gene Ontology (GO) terms overrepresented in genes with robust expression in natural strains at the G1/S checkpoint 110

Table 3.12 Gene Ontology (GO) terms overrepresented in genes with robust expression in natural strains at the M/G1 checkpoint 111

LIST OF FIGURES

Figure 2.1 Phenotypic measures of yeast cell-division cycle as a function of strain and temperature 49

Figure 2.2 Venn diagram showing the number of genes shared by three subsets: G1/S - genes with significant temperature effects at G1/S checkpoint, M/G1 - genes with significant temperature effects at M/G1 checkpoint, ESR - genes in the Environmental Stress Responses program.. 50

Figure 2.3 Number of genes showing significant strain-specific temperature effects in three yeast strains: YPS183 - a derivative of the laboratory strain S288c, YPS2055 - a natural strain collected in Pennsylvania, YPS2073 - a natural strain collected in New Jersey... 51

Figure 2.4 Steady-state expression responses to temperature in multi-dimensional transcriptome state space 52

Figure 2.5 Distribution of the loading values of genes onto the two first principal components in Figure 2.4... 53

Figure 3.1 Association of genetic and temperature variances in the transcriptome of the budding yeast 92

CHAPTER 1

INTRODUCTION

Cells are instructed by their genotypes and modulated by stimuli in the environment to direct developmental programs that lead to progressive changes in phenotypes and eventually dictate cell fates. When perturbations threaten the growth and survival of cells, one important part of the adaptive response consists of a massive reorganization of the gene expression program. Two basic strategies for transcriptional regulation are observed in evolved biological systems: expression variation and expression robustness [1]. Expression variation represents the changeability of gene expression in different environments, whereas expression robustness represents the invariance of gene expression in different environments. These two features are seemingly contradictory, yet both coexist ubiquitously in evolved biological systems [2].

Unicellular eukaryotes are subject to various kinds of environmental stress, such as natural fluctuations in temperature, sunlight, salinity, and nutrient abundances. To cope with these specific changes, cells rely on both generic and stress-specific adaptive responses at different levels: transcriptional, post-transcriptional and translational [3,4]. Studies in budding yeast have reported gene expression changes as a major component of environmental stress responses, along with alterations in metabolism and cellular physiology [5,6]. On the other hand, cells have to maintain homeostatic functions to sustain their normal growth and developmental program. Previous studies already identified the robust transcriptional regulation underlying the progression of cell-division

cycle [7-9]. While much is understood about each subject individually, how cells regulate the transcriptional response programs to environmental changes and simultaneously coordinate the cell-division cycle is still a major question for current research [10].

The main theme studied in this dissertation is the regulation of gene expression associated with the cell-division cycle under growth-permissive temperature perturbation in a unicellular eukaryote, *Saccharomyces cerevisiae*. In particular, we dissect two fundamental features of the regulated transcriptome in *S. cerevisiae*: expression variation and expression robustness. Our first objective is to elucidate gene expression variation under temperature perturbation at two checkpoints of the cell-division cycle and molecular evolution of gene expression variation. The second objective is to explore robustness of gene expression in natural populations of *S. cerevisiae* and test a specific hypothesis about the evolution of expression robustness in biological systems.

1.1 REGULATION OF GENOME-WIDE GENE EXPRESSION UNDER TEMPERATURE PERTURBATION

Among physiological changes, temperature has been one of the fundamental influences in the history of living organisms. Even a small change in temperature will alter the growth of different microbial species [11,12]. Physiological effects of temperature involve cell cycle progression, metabolic activity, cell wall and membrane dynamics, protein aggregation and sequestration [3]. For example, in the budding yeast, temperature increase may lead to transient arrest in G1 phase, reduced metabolic activity, and protein misfolding. In addition to physiological effects, heat shock evokes a signaling cascade that activates a transcriptional stress response program [10].

Temperature change disrupts the protein folding process and causes the accumulation of denatured proteins, which activates two independent stress response programs. First, the heat shock transcription factor Hsf1 induces transcription of downstream heat shock proteins to block expression of cyclins and result in transient arrest. Hsf1 target genes contain multiple copies of a 5-bp sequence of nGAAn called "heat shock elements", or HSEs [11]. Simultaneously, the general stress response Msn2/4 mediates the transcription of genes with "stress responsive elements" (STREs), 5-nucleotide sequences of CCCCT or AGGGG [10,11]. The Msn2/4 complex has two functional proteins, Msn2 and Msn4, in which Msn4 expression is stress induced. Msn2/4 is a more general stress response program that responds to a variety of stresses in addition to temperature increase [2,3,12].

Nevertheless, findings in Gasch *et al* also revealed that the transient expression responses disappeared when the stress was extended [3]. Moreover, the steady-state expression at the heat shock temperature resembled the steady-state expression at the normal temperature rather than the transient expression at the heat shock temperature. Whereas the transient programmed responses represent the necessary protective mechanisms, the steady-state programmed responses allow cells to rearrange their regulatory systems and continue their growth. Therefore, understanding homeostatic programmed responses is necessary to investigating how biological systems are modulated by environmental signals to promote progressive changes in developmental programs such as the cell-division cycle.

In eukaryotes, regulation of transcription is modulated by combinatorial effects of the structural state of DNA (i.e. chromatin formation) and regulatory proteins called

transcription factors. DNA structural state regulates transcription by limiting the accessibility of transcriptional machinery to the promoter regions, representing global transcriptional regulation. Transcription factors bind to specific DNA sequences in the promoter regions and activate transcription of target genes, representing specific transcriptional regulation [13,14]. Combined together, global and specific transcriptional regulation constitute a coordinated regulatory network that dynamically governs cellular functions. In unicellular eukaryotes such as the budding yeast, the cell-division cycle process involves a network of ~800 genes; only small numbers of key regulators are responsible for the control of this complex process [15].

While the transient response to heat shock stress involves two specific regulatory pathways (Hsf1 and Msn2/4), the steady-state response likely relies on global transcriptional regulation to drive cellular functions under temperature perturbations. Theoretical studies in the yeast cell-division cycle network claimed that such networks are stable and robust against perturbations to biological systems [18]. Two recent experimental-computational studies using network modeling of gene expression have revealed that global transcriptional machinery is likely the main coordinator of gene expression during cell growth transitions [19,20]. Experimental studies in fission yeast also indicated the presence of a coordinated global regulation of transcription that controls cell growth [21]. Therefore, preliminary evidence appears to support a key role of global transcriptional regulation to guide developmental programs, with specific transcriptional regulation playing a complementary cast as protective mechanisms against environmental perturbation.

1.2 OVERVIEW OF PHENOTYPIC ROBUSTNESS

Robustness is a property of a biological system that allows itself to maintain its functions against perturbations. Specifically, phenotypic robustness denotes the consistency of expressed phenotype under such changes. Phenotypic robustness is found at all levels of organization order: molecular structures, metabolic flux, gene expression, and developmental processes [1,22-24]. Robustness is different from homeostasis because robustness is concerned with maintaining system functions rather than maintaining system states [25,26].

Biologist C.H. Waddington first used the term "canalization", or robustness as presently used, to describe a biological system's ability to produce the same end-result regardless of variations in conditions [27]. The hypothesis was supported by a common observation in nature, that is, the wild type of organisms usually displays less variable traits as compared to the mutant. A simple reason for canalization to evolve was the fact that because of the deleterious effects of mutations on an adaptive trait, any modifier to ensure the optimal production of this trait would also be selected for [27]. Waddington demonstrated that canalization occurred during artificial selection of cross-veinless wings in *Drosophila melanogaster* under temperature shock [28]. The fact that the phenotype continued to appear after the selection stopped suggested that the developmental process was canalized to produce the adaptive phenotype. In fact, canalization is usually observed in developmental processes. In recent decades, the basis for this canalization process has been discovered to be a buffering mechanism to heat shock by the molecular chaperone Hsp90 [29].

Two main sources of perturbation distinguish two general types of robustness in biological systems. Genetic robustness prescribes the consistency of phenotype against heritable changes such as mutations. A natural case of genetic robustness is gene duplication, in which deleterious mutations occurring on either of the duplicate copies may not affect the phenotype [30]. Analogously, environmental robustness refers to the consistency of phenotype against variation in the environment, such as temperature fluctuation, sunlight, or chemical stimuli.

Another type of robustness in biological systems is stochastic robustness, that is, the consistency of phenotype against random noise inside systems. An example of stochastic robustness would be the variance of phenotype measured in multiple single cells of identical genotype [31,32]. In several cases, stochastic robustness is considered as developmental robustness because the developmental process produces similar final phenotypes despite the random noise during this process [2,33].

The impact of robustness on evolution of biological systems is another interesting question. It would seem that robustness reduces the evolvability (that is, a system's ability to produce heritable variation) and the adaptability (that is, a system's ability to adapt to new environments) of biological systems. If robustness is favored, the rigidity of the phenotype will be increased. As a result, biological systems may lose the capacity to cope with a new environment. However, recent studies have shown that robustness need not necessarily reduce evolvability and adaptability but may actually facilitate both evolvability and adaptability [34-36].

In the case of genetic robustness, when mutations occur and the phenotype is robust, these mutations will become hidden variation in the population. Therefore, a larger genotype space arises in the population and subsequently can facilitate adaptation when selection occurs [37-39]. For example, McBride and colleagues have demonstrated that robust clones of the RNA virus ϕ6 have higher evolvability of thermotolerance as compared to less robust clones [40]. In the case of environmental robustness, many theoretical studies have shown that phenotypic robustness can increase phenotypic variability and thus facilitate adaptation [41-43].

1.3 EVOLUTION OF PHENOTYPIC ROBUSTNESS

Despite the differences between the two sources of perturbation, the mechanisms underlying the evolution of genetic and environmental robustness may be related. The evolution of environmental robustness can be viewed simply as the consequence of adaptation during multiple recurrences of environmental changes. How genetic robustness evolves still under debate. Currently, three hypothetical scenarios may lead to evolution of genetic robustness: intrinsic evolution, adaptive evolution, and congruent evolution of genetic robustness [22,23,44,45].

Under intrinsic evolution, genetic robustness arises as a coupling component of the phenotype itself. In other words, robustness is acquired during the selection for the phenotype rather than in a direct selection for robustness itself. For example, genetic robustness may evolve during selection of global gene network properties. In their paper, Siegal and Bergmann demonstrated that increased genetic robustness may be due to selection for increased developmental stability of the network [46]. A different example

is the microRNA precursors in *Drosophila* that are highly robust against mutations. Simulation study has cautiously claimed that the robustness of these microRNA precursors is likely to have evolved as an intrinsic consequence of selection for secondary structures in these microRNA precursors [47]. A rather controversial example is the evolution of dominance, that is, a heterozygous locus is robust against all the mutations occurring in the recessive allele. Even though R.A. Fisher first argued that dominance evolves under direct selection as a modifier to the phenotype [48], it was later accepted that dominance may also evolve as an inevitable property of metabolic pathways due to the constraints of biochemical reactions [49].

The second hypothesis argues that natural selection can act directly on genetic robustness, so-called adaptive evolution. A simple scenario is that selection for the phenotype occurs first and then subsequently selection for robustness of this phenotype in the face of mutation. However, the conditions in which genetic robustness may be selected for are very limited [50,51]. For example, one of the conditions is the mutation rate. For genetic robustness to be selective, a high mutation rate in the population may be necessary [52]. Except for viruses, most species have low mutation rates due to selection against high mutation rate. Therefore, it is extremely rare to find satisfactory populations for adaptive evolution of genetic robustness.

The third hypothesis posits that genetic robustness may evolve as a by-product of environmental robustness because they are processing through the same biological systems. Theoretical work from Ancel and Fontana has shown that RNA structures that were robust against thermodynamic perturbation were also robust against mutations [44].

Several heat shock capacitors such as Hsp90 or GroEL not only can protect cells from environmental changes, but are also able to buffer against mutations [29,53]. In this case, environmental robustness and genetic robustness may be correlated in biological systems. Recent studies on gene expression of the whole genome have revealed correlations between genetic and environmental robustness in bacteria and budding yeast [45,54,55].

For the first two scenarios, genetic robustness and environmental robustness are independently evolved via different mechanisms. In contrast, the congruence hypothesis proposes that evolution of environmental and genetic robustness occurs in the same event. This leads to different consequences. In the congruence hypothesis, phenotypes that are robust to environmental changes will also exhibit genetic robustness. As a result, phenotypes with high robustness potentially evolve under stronger selection. In the other scenarios, phenotypes that exhibit genetic robustness may not display environmental robustness. Therefore, the different consequences help distinguish the congruence hypothesis of genetic robustness from other hypotheses.

1.4 ROBUSTNESS OF GENE EXPRESSION IN *SACCHAROMYCES CEREVISIAE*

The early literatures focused mainly on robustness of single genes and molecular phenotypes due to technical issues in obtaining large-scale experimental data. Recent high-throughput technologies have expanded the ability to collect numerous traits of interest from a single experiment. For example, microarray and next-generation sequencing allow researchers to study thousands of genes in an inexpensive and simple

procedure, which subsequently provides data for the systemic study of biological organisms.

The increasing amount of genomic data has launched more efforts to investigate the evolution hypothesis of genetic robustness in a systematic scale. Landry and colleagues used measured gene expression from four mutation accumulation (MA) lines in *Saccharomyces cerevisiae* to investigate the congruence of genetic robustness in these MA lines with environmental and stochastic robustness from available datasets [54]. However, these four MA lines were experimentally diverged for only 4000 generations which might be insufficient to observe the effects of mutations on gene expression variation.

More recently, Proulx and colleagues used gene expression data in *Saccharomyces cerevisiae* to test the evolution of robustness [45]. The authors estimated robustness from two independent datasets: genetic robustness from gene-deletion experiments and environmental robustness from stress-induced experiments. They found moderate correlation between genetic and environmental robustness in addition to evidence supporting adaptive evolution of genetic robustness. Their findings supported both adaptive and congruence hypotheses.

In a similar study to Proulx *et al*, Ben Lehner compared correlation of genetic and environmental robustness from multiple independent datasets [55]. In this paper, the author relied on the phenotypes of each single gene deletion in an array of stresses to determine environmental robustness. In two separate datasets, double-mutant effects were used to score the genetic robustness. Statistical associations between genetic and

environmental robustness were observed among these datasets despite the discrepancy of measures in different methods. The findings, therefore, appeared to support the congruence hypothesis of genetic and environmental robustness.

However, in Proulx *et al*, the measures of gene expression in these two datasets represented transient responses and perhaps did not reflect the expression responses in the context of life history traits such as cell cycle checkpoints. Therefore, a study of steady-state gene expression under perturbation will reveal robustness of expression in the constraint of developmental process. In Lehner's study, counting the conditions in which phenotypes were preserved would depend on the number of conditions investigated. Because genetic and environmental conditions are high-dimensional, this might contribute to the moderate correlation among independent datasets.

1.5 DISSERTATION STRUCTURE

This dissertation aims to dissect two aspects of regulation of gene expression in biological systems: expression variation and robustness. Chapter 2 presents an analysis of the gene expression variation associated with cell-division cycle progression in three strains of *Saccharomyces cerevisiae* with respect to temperature perturbation. Chapter 3 presents an analysis of robustness of gene expression with respect to temperature perturbation.

In Chapter 2, we characterize steady-state expression variation in two natural strains and a derivative of a laboratory strain under five temperature conditions. Because of the diverse geographic and ecological niches between natural and laboratory strains, we

hypothesize that evolution of expression variation is different among strains. We then explore two simple models of expression as a function of temperature gradients, which represents the effects of physiological kinetics and programmed regulation of expression. Lastly, we examine the global transcriptional regulation of steady-state gene expression associated with the cell-division cycle under temperature perturbation by analyzing the multi-dimensional structure of the transcriptome. We hypothesize that despite the gene expression variation in each strain, the global steady-state transcriptome may serve as a linear combination of individual genes to assist the cell-division cycle progression in *Saccharomyces cerevisiae*.

In Chapter 3, we characterize the robustness of steady-state gene expression and test the congruence hypothesis between genetic and temperature robustness of gene expression. We hypothesize that because of their specific ecological niches, the natural strains would evolve expression robustness to adapt to temperature fluctuation as compared to the laboratory strain. Because of the global transcriptional regulation, we predict systemic congruence between genetic and temperature robustness of gene expression using the variance component estimates from the generalized linear model of strain and temperature. We predict that the degree of congruence may be dependent on mechanisms of transcription regulation.

REFERENCES

[1] Kitano, H. (2004) Biological robustness. *Nat Rev Genet* ***5***, 826–837.

[2] Kaneko, K. (2012) Evolution of robustness and plasticity under environmental fluctuation: formulation in terms of phenotypic variances. *J Stat Phys* ***148***, 687–705.

[3] Halbeisen, R. E. and Gerber, A. P. (2009) Stress-dependent coordination of transcriptome and translatome in yeast. *PLoS Biol.* *7*, e1000105.

[4] de Nadal, E., Ammerer, G. and Posas, F. (2011) Controlling gene expression in response to stress. *Nat Rev Genet* ***12***, 833–845.

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

[6] Causton, H. C., Ren, B., Koh, S. S., Harbison, C. T., Kanin, E., Jennings, E. G., Lee, T. I., True, H. L., Lander, E. S. and Young, R. A. (2001) Remodeling of yeast genome expression in response to environmental changes. *Molecular Biology of the Cell* ***12***, 323–337.

[7] Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D. and Futcher, B. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Molecular Biology of the Cell* ***9***, 3273–3297.

[8] Futcher, B., Zhu, G., Spellman, P. T., Volpe, T., Brown, P. O., Botstein, D. and Davis, T. N. (2000) Two yeast forkhead genes regulate the cell cycle and pseudohyphal growth. *Nature* ***406***, 90–94.

[9] Bähler, J. (2013) Cell-cycle control of gene expression in budding and fission yeast. *Annu. Rev. Genet.* ***39***, 69–94.

[10] Struhl, K. (1995) Yeast transcriptional regulatory mechanisms. *Annu. Rev. Genet.* ***29***, 651–674.

[11] Sweeney, J. Y., Kuehne, H. A. and Sniegowski, P. D. (2004) Sympatric natural and populations have different thermal growth profiles. *FEMS Yeast Research* ***4***, 521–525.

[12] Vanoni, M., Vai, M. and Frascotti, G. (1984) Effects of temperature on the yeast cell cycle analyzed by flow cytometry. *Cytometry* ***5***, 530–533.

[13] Verghese, J., Abrams, J., Wang, Y. and Morano, K. A. (2012) Biology of the heat shock response and protein chaperones: budding yeast (*Saccharomyces cerevisiae*) as a model system. *Microbiology and Molecular Biology Reviews* ***76***, 115–158.

[14] Sorger, P. K. and Pelham, H. R. B. (1987) Purification and characterization of a

heat-shock element binding protein from yeast. *The EMBO Journal* ***6***, 3035.

[15] Sorger, P. K. and Pelham, H. R. B. (1988) Yeast heat shock factor is an essential DNA-binding protein that exhibits temperature-dependent phosphorylation. *Cell* ***54***, 855–864.

[16] Schmitt, A. P. and McEntee, K. (1996) Msn2p, a zinc finger DNA-binding protein, is the transcriptional activator of the multistress response in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences* ***93***, 5777–5782.

[17] Martínez-Pastor, M. T., Marchler, G., Schüller, C., Marchler-Bauer, A., Ruis, H. and Estruch, F. (1996) The *Saccharomyces cerevisiae* zinc finger proteins Msn2p and Msn4p are required for transcriptional induction through the stress response element (STRE). *The EMBO Journal* ***15***, 2227–2235.

[18] Li, F., Long, T., Lu, Y., Ouyang, Q. and Tang, C. (2004) The yeast cell-cycle network is robustly designed. *Proc. Natl. Acad. Sci. U.S.A.* ***101***, 4781–4786.

[19] Berthoumieux, S., de Jong, H., Baptist, G., Pinel, C., Ranquet, C., Ropers, D. and Geiselmann, J. (2013) Shared control of gene expression in bacteria by transcription factors and global physiology of the cell. *Mol Syst Biol* ***9***.

[20] Gerosa, L., Kochanowski, K., Heinemann, M. and Sauer, U. (2013) Dissecting specific and global transcriptional regulation of bacterial gene expression. *Mol Syst Biol* ***9***.

[21] Zhurinsky, J., Leonhard, K., Watt, S., Marguerat, S., Bähler, J. and Nurse, P. (2010) A coordinated global control over cellular transcription. *Current Biology* ***20***, 2010–2015.

[22] Visser, J. A. G. M., Hermisson, J., Wagner, G. P., Meyers, L. A., Bagheri-Chaichian, H., Blanchard, J. L., Chao, L., Cheverud, J. M., Elena, S. F., Fontana, W., et al. (2003) Perspective: evolution and detection of genetic robustness. *Evolution* ***57***, 1959–1972.

[23] Wagner, A. (2007) Robustness and evolvability in living systems.

[24] Hartling, J. and Kim, J. (2008) Mutational robustness and geometrical form in protein structures. *J. Exp. Zool. B Mol. Dev. Evol.* ***310***, 216–226.

[25] Waddington, C. H. (1953) The 'Baldwin effect,' 'genetic assimilation' and "homeostasis." *Evolution* ***7***, 386–387.

[26] Kitano, H. (2007) Towards a theory of biological robustness. *Mol Syst Biol* ***3***.

[27] Waddington, C. H. (1942) Canalization of development and the inheritance of acquired characters. *Nature* ***150***, 563–565.

[28] Waddington, C. H. (1953) Genetic assimilation of an acquired character. *Evolution* ***7***, 118–126.

[29] Rutherford, S. L. and Lindquist, S. (1998) Hsp90 as a capacitor for morphological evolution. *Nature* ***396***, 336–342.

[30] Gu, Z., Steinmetz, L. M., Gu, X., Scharfe, C., Davis, R. W. and Li, W.-H. (2003) Role of duplicate genes in genetic robustness against null mutations. *Nature* ***421***, 63–66.

[31] Blake, W. J., KÆrn, M., Cantor, C. R. and Collins, J. J. (2003) Noise in eukaryotic gene expression. *Nature* ***422***, 633–637.

[32] Raser, J. M. and O'Shea, E. K. (2005) Noise in gene expression: origins, consequences, and control. *Science* ***309***, 2010–2013.

[33] Kaneko, K. (2007) Evolution of Robustness to Noise and Mutation in Gene Expression Dynamics. *PLoS ONE* ***2***, e434.

[34] Wagner, A. (2008) Robustness and evolvability: a paradox resolved. *Proceedings of the Royal Society B: Biological Sciences* ***275***, 91–100.

[35] Masel, J. and Siegal, M. L. (2009) Robustness: mechanisms and consequences. *Trends in Genetics* ***25***, 395–403.

[36] Masel, J. and Trotter, M. V. (2010) Robustness and evolvability. *Trends in Genetics* ***26***, 406–414.

[37] Hermisson, J. (2004) The population genetic theory of hidden variation and genetic robustness. *Genetics* ***168***, 2271–2284.

[38] Draghi, J. A., Parsons, T. L., Wagner, G. P. and Plotkin, J. B. (2010) Mutational robustness can facilitate adaptation. *Nature* ***463***, 353–355.

[39] Lauring, A. S., Frydman, J. and Andino, R. (2013) The role of mutational robustness in RNA virus evolution. *Nat Rev Micro* ***11***, 327–336.

[40] McBride, R. C., Ogbunugafor, C. B. and Turner, P. E. (2008) Robustness promotes evolvability of thermotolerance in an RNA virus. *BMC Evol Biol* ***8***, 231.

[41] Parter, M., Kashtan, N. and Alon, U. (2008) Facilitated variation: how evolution learns from past environments to generalize to new environments. *PLoS Comput. Biol.* ***4***, e1000206.

[42] Espinosa-Soto, C., Martin, O. C. and Wagner, A. (2011) Phenotypic robustness can increase phenotypic variability after nongenetic perturbations in gene regulatory circuits. *Journal of Evolutionary Biology* ***24***, 1284–1297.

[43] Stewart, A. J., Parsons, T. L. and Plotkin, J. B. (2012) Environmental robustness and the adaptability of populations. *Evolution* ***66***, 1598–1612.

[44] Ancel, L. W. and Fontana, W. (2000) Plasticity, evolvability, and modularity in RNA. *J. Exp. Zool.* ***288***, 242–283.

[45] Proulx, S. R., Nuzhdin, S. and Promislow, D. E. L. (2007) Direct selection on genetic robustness revealed in the yeast transcriptome. *PLoS ONE* ***2***, e911.

[46] Siegal, M. L. and Bergman, A. Waddington's canalization revisited: Developmental stability and evolution. *pnas.org*.

[47] Price, M. N., Deutschbauer, A. M., Skerker, J. M., Wetmore, K. M., Ruths, T., Mar, J. S., Kuehl, J. V., Shao, W. and Arkin, A. P. (2013) Indirect and suboptimal control of gene expression is widespread in bacteria. *Mol Syst Biol* ***9***.

[48] Fisher, R. A. (1928) The possible modification of the response of the wild type to recurrent mutations. *The American Naturalist* ***62***, 115–126.

[49] Kacser, H. and Burns, J. A. (1981) The molecular basis of dominance. *Genetics* ***97***, 639–666.

[50] Meiklejohn, C. D. and Hartl, D. L. (2002) A single mode of canalization. *Trends in Ecology & Evolution* ***17***, 468–473.

[51] Wagner, G. P., Booth, G. and Bagheri-Chaichian, H. (1997) A population genetic theory of canalization. *Evolution* ***51***, 329–347.

[52] Montville, R., Froissart, R., Remold, S. K., Tenaillon, O. and Turner, P. E. (2005) Evolution of mutational robustness in an RNA virus. *PLoS Biol.* ***3***, e381.

[53] Fares, M. A., Ruiz-González, M. X., Moya, A., Elena, S. F. and Barrio, E. (2002) Endosymbiotic bacteria: GroEL buffers against deleterious mutations. *Nature* ***417***, 398–398.

[54] Landry, C. R., Lemos, B., Rifkin, S. A., Dickinson, W. J. and Hartl, D. L. (2007) Genetic properties influencing the evolvability of gene expression. *Science* ***317***, 118–121.

[55] Ben Lehner. (2010) Genes confer similar robustness to environmental, stochastic, and genetic perturbations in yeast. *PLoS ONE* ***5***, e9035.

CHAPTER 2

VARIATION OF GENOME-WIDE GENE EXPRESSION IN *SACCHAROMYCES CEREVISIAE* REVEALS GLOBAL SCALING PATTERN UNDER TEMPERATURE PERTURBATION

2.1 INTRODUCTION

Physiological stresses can threaten cellular and organismal functions in yeast, in part by altering patterns of gene regulation that guide developmental programs and determine cell fates. Examples of such physiological stresses include osmotic shock, high salinity, high temperature, and exposure to radiation and toxic agents. The effect of these and other stresses on genome-wide gene expression have been examined in budding yeast [1]. While most stresses elicit only moderate responses in gene expression, heat shock in particular has been shown to induce a massive and rapid alteration of genome-wide gene expression levels [2].

The impact of temperature on gene expression can be manifested in two ways. First, as a global physiological parameter, temperature can affect the kinetics of biochemical reactions and thermodynamic equilibrium. Therefore, changes in gene expression due to temperature perturbation may be caused in part by altered biochemical kinetics. Second, temperature change may trigger programmed alterations in gene regulation. There are two main classes of programmed regulatory responses: immediate and transient responses and stable, long-term responses. Immediate programmed responses are typically transient and function as protective mechanisms, such as the expression of heat shock and cold shock response genes [3,4]. On the other hand, long-term programmed responses may be

steady-state and related to homeostatic functions or life-history modulation functions, such as the timing of G1/S checkpoint in yeast.

Importantly, immediate programmed responses have been well-studied in yeast, but homeostatic programmed responses remain poorly understood. In this study, we investigated the steady-state expression responses to temperature changes in three yeast strains: two natural strains and a derivative of a laboratory strain. We first dissected the steady-state expression responses in each individual gene. We found that approximately half the transcriptome responded to temperature perturbation. Interestingly however, only one-fifth of these genes also exhibit transient responses to heat shock. This suggests that steady-state responses involve distinct gene regulatory mechanisms from transient responses. Second, we evaluated the mechanism of steady-state responses, either due to kinetic changes or programmed responses, for each gene. Surprisingly, few genes displayed similar types of responses across strains, suggesting that temperature-dependent steady-state expression might evolve in a strain-specific manner (that is, may be dependent on historical evolutionary trajectories). Lastly, analysis of global patterns of genome-wide expression in response to temperature perturbation revealed a one-dimensional linear subspace parameterized by a temperature gradient. This linear subspace is uniform in all three yeast strains and at different checkpoints. Therefore, we propose that the temperature-dependent steady-state transcriptome may be regulated as a linear combination of genome-wide gene expression to sustain life-history traits such as the cell-cycle checkpoints.

2.2 GROWTH RATE SHOWS A LINEAR RESPONSE TO TEMPERATURE CHANGE BUT NOT RELATIVE TIMING OF THE G1/S TRANSITION

The G1/S or "Start" transition is the mitotic cell-cycle checkpoint at which a budding yeast cell devotes resources to either to completing the cell-division cycle by initiating DNA replication (S phase) or to prolonging the phase of cell growth (G1 phase). A class of proteins called G1 cyclins regulates this transition by binding and activating cyclin-dependent kinases, which in turn initiate the G1/S transcriptional program by phosphorylating Whi5 [5]. Thus, G1 cyclins orchestrate the changes in gene expression dynamics necessary for cell-cycle progression.

Temperature has been previously characterized as a major physiological parameter that significantly impacts cell-cycle dynamics, specifically the duration of the cell-division cycle, the timing of checkpoint transitions [6,7], and the size and molecular complement of a cell at division, which can be viewed as major life-history parameters [14]. However, the specific impact that temperature variation has on genome-wide gene expression levels remains unknown. To examine this, we first performed phenotypic measurements of the yeast cell-division cycle as a function of strain and temperature. Figure 2.1 shows the length of cell cycle and the relative timing of the G1/S during the cell cycle as a function of strain and temperature. In general, the length of the cell-division cycle appears to be a linear function of temperature, declining at a constant rate with increase in temperature (Figure 2.1A). The relative rate of decrease with temperature is similar for all strains. In contrast, the timing of the "Start" transition in the cell cycle reveals a more complicated relationship with temperature (Figure 1B). If all the checkpoints scale uniformly with

temperature, we would expect a constant ratio of the G1/S transition. But as can be seen in the figure, from 18°C to 26°C, there is a relative acceleration of the G1/S transition in all strains, especially in strain YPS2073. There is a general reversal of this trend between 26°C and 34°C. One possible interpretation is that the yeast strains prefer cell division over cell growth at optimal temperatures while delaying the cell-division decision at less optimal temperatures.

To study the impact of temperature on steady-state gene expression involving life history functions such as the G1/S checkpoint, we sequenced the yeast transcriptome at the G1/S checkpoint in five different temperature conditions in comparison to the M/G1checkpoint.

2.3 GENOME-WIDE GENE EXPRESSION EXHIBITS DISTINCT STEADY-STATE RESPONSES TO TEMPERATURE PERTURBATION

To characterize the molecular changes related to cell-division cycle progression following temperature perturbation, we sequenced the transcriptome of three yeast strains sampled at two cell-division cycle checkpoints (G1/S and M/G1) following incubation at five temperature conditions (18-34°C with 4°C interval). mRNA samples extracted from each condition were prepared using serial analysis of gene expression (SAGE) and sequenced by next-generation sequencing; this yielded an average of 5.2 million 22-nucleotide reads per sample. Reads were mapped to the reference genome at an average rate 72% of total reads, in which 42.5% mapped uniquely to mRNAs and 3.2% mapped uniquely to the antisense strands of mRNAs. The number of reads mapping uniquely to mRNAs was normalized to generate the transcriptome profiling in each sample.

2.3.1 TEMPERATURE PERTURBATION INDUCES STEADY-STATE EXPRESSION RESPONSES IN A LARGE NUMBER OF GENES, INCLUDING ANTISENSE EXPRESSION

A total of 4344 protein-coding genes showed detectable expression in all samples (that is, at least a uniquely mapped read was found in each sample). A generalized linear model with strain and temperature as the independent variables and normalized gene expression level as the dependent outcome was applied [8,9]. Genes have significant temperature responses if their expression changes significantly in one or more temperature conditions. The linear model suggested 2405 genes with significant temperature responses at the G1/S checkpoint and 1879 genes with significant temperature responses at the M/G1 checkpoint (Benjamini-Hochberg correction, p-value<0.05). Overall, 1330 genes showed significant temperature responses at both stages.

To analyze the functional categories of genes showing significant temperature responses, we used the Gene Ontology annotation from the *Saccharomyces* Genome Database (SGD; http://yeastgenome.org; January 2013). Genes with significant temperature responses are overrepresented in metabolism and translation activities. The enriched gene ontology terms for 1330 genes significant at both stages involved structural constituents of the ribosome and components of the cytoplasm, mitochondria and ribosomes (Table 2.1). Genes significant at the G1/S checkpoint were enriched in the translation process, structural constituents of the ribosome, and many molecular functions related to cytoplasm and metabolism (Table 2.2). Genes significant at the M/G1 checkpoint were enriched in the large ribosomal subunit (Table 2.3). The gene ontology enrichment

suggested that regulation of translation activities is critical to assist proper life history traits such as cell-division cycle checkpoints.

Findings in Schade *et al* suggested that immediate transcriptional response of the budding yeast to cold shock differed to the immediate response to heat shock [4]. The growth-permissive temperature range in our study stretches from a moderately cold extreme, to the typically optimal range, and lastly to a moderately hot extreme. To analyze the significant responses to a particular temperature, we tested three comparisons for the genes with significant temperature responses: 18°C against 22-30°C, representing the low temperature extreme against normal range; 34°C against 22-30°C, representing the high temperature extreme against normal range; and 26°C against 22 and 30°C, representing the optimal temperature against immediate sub-optimal temperatures.

Compared to the normal range of growth temperatures, the low temperature extreme yielded 1152 and 291 significantly differentially expressed genes at G1/S and M/G1 stages, respectively (Benjamini-Hochberg correction among genes, p-value<0.05 with Bonferroni correction for the number of comparisons per gene). At G1/S, genes were enriched in cytoplasm and mitochondrion components and the proteasome complexes (Table 2.4). The enrichment of mitochondrion components suggests that energy (i.e. ATP) metabolism is particularly important in cold conditions, whereas enrichment of the proteasome complexes suggests the disruptive effects of low temperature to protein folding.

At the other end, the high temperature extreme yielded 687 and 348 significantly expressed genes at G1/S and M/G1 stages as compared to normal range, respectively.

The G1/S-significant genes were enriched in cytoplasmic translation processes involving structural constituents of the ribosome and heat shock protein binding (Table 2.5). The M/G1-significant genes were also enriched in components of ribosomes and unfolded protein binding functions, a regular component of the heat shock response (Table 2.6). This suggested that steady-state expression in relation to life-history traits such as the G1/S checkpoint may require cautiously protective mechanisms of the heat shock program, even at a growth-permissive heat condition. In contrast with our findings in synchronized cell populations, Gasch *et al* showed that unsynchronized steady-state expression at 33°C was similar to unsynchronized steady-state expression at 29°C [3]. A potential explanation is that the measurement of gene expression in unsynchronized cultures averages true variation over the distribution of cells at different stages of the cell-division cycle.

As noted in figure 1B, at 26°C all three strains had the shortest transition between M/G1 to G1/S in relation to the total cell cycle length. At this temperature, we found 25 and 21 genes significantly differentially expressed at G1/S and M/G1 compared to sub-optimal temperatures of 22 and 30°C. These genes represent functional categories of metabolism, translation, oxidation, ribosomal activities. This suggests that optimizing a complex trait such as cell growth may involve broadly optimizing specific components in multiple pathways, but not necessarily all the components of the pathways.

In regulation of transcription, a previous study has reported that genes showing antisense expression had larger sense expression variability compared to genes without antisense expression [10]. The model proposed that antisense transcription inhibits sense

expression when sense expression is low, and this inhibition is relaxed if sense expression is high. A potential function of antisense expression is to rapidly and efficiently regulate mRNA levels post-transcriptionally by forming double-strand bonds with *cis*-complementary mRNAs and thereby obstructing the translation machinery. Therefore, we explored the possibility of antisense expression regulating steady-state sense expression under temperature stress.

In our study, an average 3% of the reads mapped to the antisense strand, which is comparable with previous reports of antisense transcription [11]. Among 1225 genes expressing antisense transcription in all samples, 343 and 305 genes showed significant temperature effect at G1/S and M/G1 stages, respectively. When compared to the sense expression, 206 (out of 343) and 141 (out of 305) genes also showed significant temperature effects for sense expression at G1/S and M/G1 checkpoints. These accounted for 60% and 47% of the total number of genes showing significant temperature effect in antisense expression, at the respective checkpoints. Sense and antisense expression in each sample were weakly correlated overall (average Pearson's correlation coefficient r=0.14 in all samples), which was consistent with previous studies [10,11]. Assuming that sense and anti-sense expression are nearly independent of each other, the expected fraction of genes with significant antisense temperature effect that overlaps with significant sense effect is 187 and 124, respectively (Fisher's exact test, p-value=0.007 for G1/S checkpoint, p-value=0.015 for M/G1 checkpoint). The high fractions of genes showing significant temperature effects in both sense and antisense expression raised the possibility of temperature-dependent antisense expression and the possibility that the

antisense expression has a functional role in overall transcriptional regulation. In fact, we found that genes showing similar significant temperature effects in both sense and antisense expression (defined as the foreground set) exhibited significantly higher sense expression variability as compared to other genes having antisense expression (defined as the background set; mean temperature variance of 3.777 versus 1.655 at G1/S checkpoint, Wilcoxon rank-sum test, p-value=1.4×10^{-6}; mean temperature variance of 2.532 versus 0.966 at M/G1 checkpoint, Wilcoxon rank-sum test, p-value=2.2×10^{-16}). In contrast, temperature variance of the genes in the background set was similar to that of genes without antisense expression (1.655 versus 1.609). Therefore, our result suggests that enhanced variability of sense expression under temperature perturbation may be partly mediated via regulation of antisense expression.

2.3.2 STEADY-STATE EXPRESSION RESPONSES TO TEMPERATURE PERTURBATION ARE UNRELATED TO TRANSIENT EXPRESSION RESPONSES TO HEAT SHOCK

Based on multiple studies, a large set of ~900 genes has emerged that display transient changes in expression following various environmental stresses [2,3,12]. That is, expression levels of these transient genes returned to normal levels as the stress conditions continued. Out of these 868 transient stress-response genes [3], we found 406 and 278 genes with significant temperature responses at G1/S and M/G1, respectively. Therefore, 47% and 32% of transient stress-response genes showed differential steady-state expression at the temperature ranges assayed in this study. However, these genes only accounted for 17% and 15% of temperature-responsive genes in our study (Figure

2.2). This suggests that the steady-state expression response program is different from the transient expression responses to temperature changes.

2.4 STRAIN-SPECIFIC EXPRESSION RESPONSES TO TEMPERATURE REFLECT THE ADAPTATION OF NATURAL STRAINS TO THEIR MICROENVIRONMENT

While the laboratory strain of yeast is generally cultured in a near-optimal temperature range at ~25°C, the natural strains used in this study have experienced routine daily and seasonal temperature fluctuations for thousands of years. This contrast allows us to examine whether different strains have evolved specific gene expression responses to temperature. We ran analysis of variance separately for each strain. The laboratory strain, YPS183, had the largest number of genes showing a significant temperature response at any checkpoints among all three strains (1475 and 377 genes at G1/S and M/G1 checkpoints as compared to 676 and 205 in YPS2055, 244 and 0 in YPS2073; Figure 2.3 and Table 2.7). We hypothesize that the smaller number of significant genes in the natural strains as compared to the laboratory strain reflects the homeostatic adaptation of the natural strains to routine climate fluctuations, which is likely to have been lost in the laboratory strain.

Between the two natural strains, YPS2055 had more genes showing a significant temperature response than YPS2073 (676 compared to 244 at G1/S, and 205 compared to none at M/G1 checkpoint). The differences between the natural strains were surprising since the strains were collected from nearby regions (100 miles apart) with similar range of temperature fluctuation year-round (YPS2055 from Tyler Arboretum in Pennsylvania,

YPS2073 from Mettler's Woods in New Jersey) [13,14]. We postulated that the similar responses would involve temperature changes, and the different responses might be results of other microenvironment effects. In that case, we would expect that the common genes (that is, genes that show significant temperature effects in both natural strains) would be involved in the temperature-dependent regulation pathway, but the strain-specific genes (that is, genes that show significant temperature effects in only one of natural strains) would not be related to temperature adaptation. However, the common genes were not enriched in any gene ontology category. Neither were the strain-specific genes.

Regulation of transcription in eukaryotes is a result of transcription factor assembly binding to upstream non-coding sequences of coding genes [15]. Therefore, one hypothesis is that evolutionary changes in *cis*-element sequences have altered transcription factor binding patterns resulting in different responses to temperature change. To understand the evolutionary forces responsible for the expression divergence among these three strains, we examined changes in noncoding *cis*-regulatory elements for genes displaying strain-specific temperature effects. We used novel genome assemblies of the two natural strains and the laboratory strain (see details in Materials and Methods) to estimate the molecular divergence between these strains and an outgroup species *Saccharomyces paradoxus*.

For each gene, we defined a 5' *cis*-element region that extended from the start codon to 500 base pairs upstream of the target gene. If another gene fell within this region, we removed any overlapping sequence. Consequently, some 5' *cis*-element regions may be

shorter than 500 base-pairs. Similarly, we defined a 3' *cis*-element region that extends from a gene's stop codon up to 500 base pairs downstream of the gene, excluding overlapped segments. The 3' *cis*-element of a gene serves a negative control for the identification of the 5' *cis*-element since the majority of known *cis*-elements reside in the 5' region [16].

We aligned the 5' *cis*-element region, 3' *cis*-element region, and the coding sequences (CDS) for each gene using gene annotations corresponding to the *S288c* reference genome (SGD release 64/UCSC sacCer3). We then computed evolutionary rates of DNA sequence change for each region. For coding sequences, we estimated synonymous and non-synonymous sites using the method of Li-Pamilo-Bianchi [17,18]. For the 5' *cis*-element and 3' *cis*-element, we used the Kimura two-parameter model [19-21]. An assembly of the *Saccharomyces paradoxus* genome was used as the outgroup species for all distance estimates [22]. The ratios of the evolutionary rate of the non-synonymous sites (dN), the 5' *cis*-element (d5), and the 3' *cis*-element (d3) to the rate of the synonymous sites (dS) was used to infer the mode of evolution in each gene.

We tested the hypothesis that the current *S. cerevisiae* strains have evolved temperature-dependent expression adaptations due to natural selection. We hypothesized that if expression responses were selectively advantageous for each gene and strain independently, we would observe higher rates of relative divergence in the 5' *cis*-element of genes that show a strain-specific expression response compared to those that have a strain-independent response to temperature change. To examine this, we defined a set of genes as common genes that show a significant temperature effect on expression in all

three strains, and another set of specific genes that show significant strain-specific temperature effects based on our general linear model analysis (see above and in Materials and Methods). We then compared the rates of molecular evolution in the specific genes for each strain with the rates of molecular evolution in the common genes.

Surprisingly, we found significant differences between the specific genes and the common genes in the coding regions but not in the 5' and 3' *cis*-element regions (Table 2.8). In all three strains, the coding sequence of strain-specific genes had a significantly higher dN/dS ratio than the common genes. This suggested that strain-specific genes might have been under slightly positive selection, which might be linked to functional differences under temperature-dependent expression. That is, if a gene is important for temperature homeostasis during the cell cycle and its expression changes with temperature, then it may be subject to stronger directional selection. However, the similarity of the evolutionary rates of 5' *cis*-elements between common genes and specific genes suggested that the *cis*-elements did not contribute to the temperature-dependent expression divergence among these strains. The range of d5/dS and d3/dS values was in agreement with recent reports of the evolutionary rates of 5'-flanking regions in *Drosophila* genes [20]. Thus, no genes showed significant evidence of elevated rates of molecular evolution for sequences that may modulate gene expression by *cis* effects.

To test the evolution of *trans* factors, we compared the evolutionary rates among regulators of the common genes (named "TF-common"), regulators of the strain-specific genes (named "TF-specific"), and the general transcription factors that do not regulate the

temperature-dependent genes (named "TF-general"). We observed different trends of molecular evolution in the 5' *cis*-element, the coding sequence, and 3' *cis*-element (Table 2.9). The regulators of the strain-specific genes had higher rates in the coding sequences and lower rates in the 5' *cis*-elements as compared to the general transcription factors, whereas the rates in the 3' *cis*-elements were inconsistent across three strains. Even though it was not significant, the higher rates in coding sequences and lower rates in 5' *cis*-elements were consistent across the three strains. For the regulators of the common temperature-dependent genes, the coding sequences showed similar rates to the general transcription factors, but the 5' *cis*-elements showed lower rates similar to those of the strain-specific genes.

The higher evolutionary rates in the coding sequence suggested that there were weak positive selection forces acting on the protein function of the regulators of the strain-specific temperature-dependent genes. As shown in Table 2.9, we found lower evolutionary rates in 5' *cis*-element regions of the regulators of the temperature-dependent genes suggesting negative selection forces acting on *cis* effects compared to the general transcription factor genes. In fact, lower expression variability across different genetic backgrounds in these specific regulators as compared to other subsets of regulators provided further support for negative selection. Altogether, these findings suggest that modifications in the protein function of the *trans* regulators might contribute to the strain-specific temperature-dependent expression rather than the *cis*-element of the regulated genes.

Overall, we observed strain-specific temperature-dependent expression in different subsets of genes in three yeast strains. We showed that the coding sequences of these strain-specific genes might have experienced greater positive selection relative to temperature responsive genes common to all strains. We examined the *cis*-elements of these strain-specific genes for possible molecular changes related to temperature-dependent expression differences across strains but did not find any differences in rates. We next examined genes that regulate the expression of these temperature-dependent genes to see if changes in their protein function may be related to strain-specific differences in temperature-dependent gene expression. We found weak selective forces in the coding sequence evolutionary rates as compared to general transcription factors. However, we found patterns of molecular evolution consistent with enhanced negative selection in the 5' *cis*-element regions of the genes that regulate the expression of the temperature-dependent genes. Together, these results suggest that while the expression of *trans*-regulators may be under stabilizing selection, their amino acid sequences are evolutionarily variable, which could cause broad downstream regulatory effects due to changes in DNA binding affinity, target specificity, or protein interaction.

2.5 STEADY-STATE EXPRESSION RESPONSES SHOW STRAIN-SPECIFIC DIFFERENCES IN TEMPERATURE RESPONSE CURVES

Our results have shown that some genes have significant gene expression changes following steady-state temperature perturbations. Moreover, some of these genes also have significant strain-specific effects. However, the relationship between a gene's expression level and temperature is complex and gene-specific. Here we investigated two

possible patterns of temperature-dependent expression change using a regression model: (1) monotonic linear response to temperature, which may be driven by temperature dependent rates of chemical reaction; and (2) quadratic response to temperature, suggesting regulated expression levels responding to some optimal temperature range.

For each gene and each strain, we applied a linear regression model and a quadratic regression model to the measured expression responses as a function of temperature gradients. We tested for significant linear and quadratic regression independently, and used an F-statistic to test the increase in coefficient of determination (R-squared) from linear to quadratic [8]. Genes exhibit a quadratic response if they satisfy either of the following conditions: (i) the quadratic regression is significant and the linear regression is not; (ii) two kinds of regression are significant and the quadratic model is significantly better than the linear model. For genes failing both conditions, they have linear responses if the linear regression was significant; otherwise, we assumed that genes failing both types of response were too irregular or too constant to fit these two models.

Table 2.10 lists the number of genes with different types of responses in the three strains. First, the laboratory strain generally had more genes that fitted both the linear and quadratic model than either of the natural strains, suggesting that the natural strains have more buffering of expression regulation than the laboratory strain. Second, there were more genes that fitted these two models at the G1/S checkpoint than the M/G1 checkpoints, especially for the quadratic model. The G1/S checkpoint regulates the decision to devote cellular resources for DNA synthesis, which is the most energetically costly step in the budding yeast cell cycle [23]. There is a possibility that more genes are

involved in optimizing temperature-dependent decision to commit to DNA synthesis than optimizing mitosis. The number of genes showing a quadratic response at the M/G1 checkpoint was especially low for the natural strains at less than 10% of the genes showing linear response at the same checkpoint. Therefore, we hypothesized that in natural strains, mitosis may not be optimized for a particular temperature as much as the decision to proceed to DNA synthesis.

Because the linear response represents a potentially non-regulated response, we expected that many genes would display similar linear responses in all strains. Surprisingly, few genes showed the same types of response in all three strains. Only 51 and 21 genes had linear responses in all strains at the G1/S and M/G1 checkpoints, respectively. Of these, only three genes - *MRPL38*, *KNH1*, *KTI11* - had linear responses at both checkpoints. A subset of 17 genes had quadratic responses in all strains at the G1/S checkpoint, but none at the M/G1 checkpoint. These genes with common temperature-dependent response patterns were enriched in the protein unfolding process, suggesting that buffering protein folding from temperature effects is the most basic component of temperature dependent gene expression. When compared to the transient stress responses, few genes with linear responses (7 out of 51 and 3 out of 21) were overlapped. However, there were 7 out of 17 genes with quadratic responses were found in the transient stress responses.

The number of genes sharing the same temperature response profile was greater in the natural strains than in all three strains. We found 116 and 89 genes with shared linear responses at G1/S and M/G1 checkpoints, respectively. These linear genes were significantly enriched in metabolic processes. When we compared strain YPS2055 to the

laboratory strain, we found 142 and 47 genes with shared linear responses at G1/S and M/G1 checkpoints. In the case of strain YPS2073, there were 75 and 41 genes with shared quadratic responses at G1/S and M/G1 checkpoints.

We found 49 and 3 genes with shared quadratic responses at G1/S and M/G1 checkpoints in the natural strains. These quadratic genes were not enriched in any functional category. If temperature fluctuation dictated the same adaptation occurring in the natural strains, we would expect more common regulated responses between the natural strains. When we compared strain YPS2055 to the laboratory strain, we found 61 and 2 genes with shared quadratic responses at G1/S and M/G1 checkpoints. In the case of strain YPS2073, there were 45 and 5 genes with shared quadratic responses at G1/S and M/G1 checkpoints. It is surprising to see that the number of genes with regulated responses in pairs of strains were similar. At the beginning, we hypothesized that temperature fluctuation guides the evolution of regulated expression, perhaps leading to similar regulated expression against temperature changes. However, our results suggested that the evolution may occur independently to achieve same goals and thus, few genes shared similar temperature response curves between the natural strains.

Altogether, the analysis of temperature response curves illustrates two elementary models of temperature-dependent expression in three yeast strains: monotonic response representing first-order chemical reactions, and quadratic response representing regulated expression with an optimal temperature point. Consistent with the previous findings that the natural strains have adapted to temperature fluctuation by buffering expression variation due to temperature change, they have few genes having regulated responses to

temperature as compared to the laboratory strain. However, in all three strains, optimizing gene expression at the time of progression to DNA synthesis involves more genes than at mitosis. Furthermore, many temperature-dependent regulated expression curves were strain-specific, which suggested that cells might employ different regulatory responses to guide developmental programs under environmental changes.

2.6 GLOBAL PATTERN OF STEADY-STATE GENE EXPRESSION DISPLAYS MAJOR ONE-DIMENSIONAL LINEAR SUBSPACE OF TEMPERATURE RESPONSES

In the previous sections, we have characterized the steady-state expression responses, differentiated steady-state responses from transient responses, and modeled the expression response as a function of temperature. In general, temperature impacted the steady-state expression of half the transcriptome, a much larger effect than seen in the transient responses. Analyses suggested strain-specific temperature-dependent responses during the cell-division cycle. However, these specific responses likely resulted from specific transcriptional regulation of gene expression (that is, modular regulation of transcription in a subset of genes involving a specific transcription factor) in joint control with the global transcriptional regulation (that is, generic transcription machinery involving most genes in the genome) in a cell. Two recent studies used a model-based metabolic network approach to evaluate the contribution of specific and global transcriptional regulation during cellular growth in bacteria [24,25]. Both studies have shown that global transcriptional regulation plays a dominant role during growth transitions. Therefore, we hypothesize that global regulatory coordination of genome-

wide expression is the main mechanism that regulates the temperature-dependent responses despite the observed strain-specific transcriptional responses.

To assess the global structure of the genome-wide expression, we examined the major axes of variation using principal component analysis (PCA). Each principal component in PCA analysis represents a weighted linear combination of all the expressed genes, in which the ordering of the principal components (PC) reveals decreasing levels of variation. Figure 2.4 shows a projection of all measured transcriptomes onto the first two PCA axes. In this figure, PC1 generally seems to separate the natural strains from the laboratory strain while PC2 seems to mainly show responses to temperature. While not shown in Figure 2.4, we also found the PC3 axis separates the cell cycle checkpoints. While the previous gene-centric analysis suggested considerable variation in which genes show temperature effects in which strains and checkpoints, the global multi-dimensional picture suggests a surprisingly simple one-dimensional temperature-dependent gradient of expression pattern. The individual gene effects are projections of this linear gradient to each gene axis, which may show non-linear response patterns such as quadratic responses, dependent on the parameterization by temperature. In addition, which genes show large changes may differ dependent on strain and stage, but the global pattern with respect to temperature dependent variation shows a simple linear subspace of expression response. We hypothesize that even though environmental conditions may shape the expression landscape of many individual genes, the global transcriptome variation follows a simple linear scaling to promote homeostatic functions or life-history modulation functions.

2.7 DISCUSSION

Regulation of gene expression is instructed by genotypes and modulated by environmental signals to guide the developmental program in unicellular eukaryotes. Using next-generation sequencing, we characterized the regulation of genome-wide expression in three yeast strains under five different temperature conditions. Gene expression was quantified at two key checkpoints of the cell-division cycle, representing the steady-state genome-wide transcriptome under a growth-permissive temperature gradient. We analyzed individual gene expression to investigate regulatory pathways that simultaneously respond to temperature and coordinate cell cycle, and to systematically model the expression of individual genes as a function of temperature gradients. We observed that most steady-state gene expression responses to temperature gradient were strain-specific. However, we hypothesized that despite the strain-specific expression variation we observed in half of the transcriptome, the global structure of the transcriptome with regard to growth transitions reflects whole-transcriptome response of nearly all the genes, which was surprisingly consistent with a multi-dimensional linear response model of temperature gradients.

We detected large-scale changes in the steady-state transcriptome responses to temperature changes: 50% of genes showing significant temperature responses at one of the two checkpoints, of which 1330 genes displayed responses in both checkpoints. However, when we evaluated gene expression in strain-specific contexts, fewer genes displayed significant temperature responses, especially in the natural strains. This suggested that strain-specific transcriptional regulation has differentially influenced the gene-by-gene expression variation in response to temperature conditions, perhaps in the

strain-specific microenvironment. A potential explanation from our observation of fewer significant temperature responses in natural strains is that because of their frequent experience with temperature fluctuation, they have acquired robustness against temperature changes and therefore display modest expression responses as compared to the laboratory strain, as observed in several occasions in wild strains of budding yeasts [26,27].

Expression plasticity describes the adaptive expression variation in response to distinct environmental conditions that may be selected for [28,29]. For example, in our study, several ribosomal genes have higher expression at optimal temperature but much lower expression at unfavorable temperatures. Genes displaying expression plasticity might be advantageous because they facilitate adaptation to specific temperature conditions. We actually found five genes with enhanced expression plasticity as genes displaying significantly increased expression variation in the natural strains compared to the laboratory strain (*LEU2*, *DAL5*, *DAL2*, *VBA1*, *AAD16*). Interestingly, two of these genes (*DAL5*, *DAL2*) are in the allantoin degradation pathway that allows yeast to use allantoin as a nitrogen source [30]. Nitrogen starvation has been linked to detrimental effects during the cell-division cycle in both budding yeasts and fission yeasts [31,32]. Therefore, we speculate that because of the reduced metabolic activities at lower temperatures, natural strains might have recruited the allantoin pathway as a secondary supplier for nitrogen, which is required for proper cell division. In fact, we observed that the expression of the two allantoin genes in the natural strains were highest at 18°C and gradually declined as temperature increased.

In recent years, regulatory non-coding RNAs have emerged as a key mechanism to control gene expression with respect to stress responses, including antisense RNAs, microRNAs (miRNAs), and long non-coding RNAs (lncRNAs) (reviewed in [33]). One potential function of antisense transcripts is to rapidly inhibit translation of sense transcripts by complementary binding, therefore serving as a protective mechanism for energy-costly metabolisms activities in stress conditions. However, little is known about the context in which antisense expression plays a primary role in transcriptional regulation. Our finding of genes with significant temperature effects on both sense and antisense expression suggests that under temperature fluctuation antisense expression may confer higher variability of sense expression, despite the weak correlation between sense and antisense expression. A potential mechanism of antisense regulation of gene expression involves chromatin remodeling in which natural antisense transcripts cause methylation/demethylation at the promoter regions of sense transcripts and subsequently silence the expression of sense transcripts [34-36]. A more recent study reported that the antisense transcript of a gene PHO84 affects transcription in the sense strand by blocking the promoter region rather than binding to the sense transcript [37]. These two examples may represent potential generic mechanisms of antisense regulation of transcription in *cis*-complementary strands; but further large-scale studies may be required to address antisense regulation of sense transcription.

A challenge in biological systems is unraveling the roles of specific and global transcriptional regulation to promote growth transitions in unicellular organisms. Despite the large amount of temperature-dependent expression variation in all three strains, we

could not infer specific regulatory pathways that might drive the observed expression variation. This led us ask whether there is a mostly global transcriptional regulation of steady-state gene expression in response to alterations of temperature—possibly a global physiological parameter. When we dissected the global structure of the transcriptome, we found a major one-dimensional subspace that is concordant with the temperature gradient. Surprisingly, such a simple response surface was seen for each of the three strains as well as for the different cell cycle checkpoints (Figure 2.2). The strains and the checkpoints differed from each other in terms of the specific genes with temperature-dependent variation but nevertheless followed the simple global response pattern shown in Figure 2.2. Hence, our result supports a regulation mechanism via global transcriptional machinery to assist growth transitions in a unicellular eukaryote, which is similar to recent findings in bacteria comparing the contribution of specific and global transcriptional regulation to gene expression at growth transitions [24,25]. Furthermore, if specific transcriptional pathways regulate the temperature-dependent expression variation, they are likely mediated via *cis*-regulatory elements in the upstream region of a gene [38]. Our molecular analyses suggested that *trans*-acting factors rather than *cis*-elements may be responsible for gene expression variation against temperature changes, providing further support for the involvement of global transcriptional regulation to steady-state genome-wide expression during temperature perturbation. We speculate that the mechanism of regulation is likely via transcription mediators that affect nucleosome architecture, as discussed in a paper by Zaugg and Luscombe [39].

Overall, we characterized and analyzed the steady-state genome-wide expression of *Saccharomyces cerevisiae* under temperature perturbation. Our results showed that the steady-state expression response to a temperature gradient was strain-specific, perhaps due to independent lineages of evolution. Despite their independent evolution, the global genome-wide expression as a response to temperature gradient is uniformly consistent in all three strains. We hypothesize that while cell lineages may evolved expression variation in individual genes, the evolution of temperature-dependent expression response involves a system-level "tuning", perhaps involving the global architecture of the yeast transcriptome, rather than individual gene effects or pathways. A potential mechanism for this global tuning of transcriptome is via *trans*-acting factors, such as transcription factors and global epigenetic changes.

2.8 MATERIALS AND METHODS

Yeast strains

Two natural strains—YPS2055 and YPS2073—are heterothallic haploid MATa derivatives from homothallic diploids *Saccharomyces cerevisiae* strains, collected in Pennsylvania and New Jersey, USA [13,14]. The laboratory strain YPS183 is a derivative of BY4741 (genotype: *HO*Δ::*kanMX*, *leu*2Δ). In all three strains, the mating-type loci were replaced by a Kanamycin resistance cassette to prevent yeast cells switching from a-mating type to α-mating type.

Calibration of the length of cell-division cycle and Start checkpoint

The length of the cell-division cycle (CDC) and the timing of G1/S checkpoint are strain-specific and temperature-dependent. CDC length was estimated as the time needed to double the number of cells in liquid cultures. Synchronized yeast cultures were released from G1 arrest and inoculated in synthetic dextrose (SD) media at controlled temperature conditions (18, 22, 26, 30, and 34°C). Samples of 200ul culture were collected every 30-60 minutes for at least one cell cycle. The number of yeast cells was counted in triple replicates using a cell counter (Beckman Coulter, Z2 series). We then applied a linear regression to determine the doubling time for particular strain and temperature conditions.

The G1/S checkpoints were determined from the budding index profiles. For each combination of strains and temperatures, yeast cell cultures were collected every five minutes in one cell cycle. Yeast cells were fixed and stained with DAPI (4',6-diamidino-2-phenylindole; Sigma Part D9542). Microscopy determined the timing of cultures having 25% budded cells after releasing from G1 arrest. At least 200 cells were counted to estimate the fraction of budded cells. The length of CDC and the relative timing of G1/S checkpoint are summarized in Figure 2.1.

Collection of synchronized yeast cultures

Frozen stocks were plated on a rich medium (YPD) and incubated overnight. Freshly grown colonies were inoculated in minimal SD media at 30°C, 225rpm. The next day cultures were diluted into new SD media and grown to an optical density of OD600=0.3. Alpha-factor pheromone (Zymo Research Part Y1001) was added to the media at concentration of 4μM. Yeast cultures were incubated for an hour and checked under microscopy to confirm G1 cell synchronization before washing twice with designated-

temperature SD media. Synchronized cultures were released into new SD media and incubated at the designated temperature, 225rpm. Cultures were collected at the G1/S and M/G1 checkpoints.

Overall, we sampled three yeast strains in five temperature conditions (18, 22, 26, 30, and 34°C) at two checkpoints (G1/S and M/G1), each condition in two biological replicates. Yeast cells were harvested and stored in 1mL TRIzol Reagent (Invitrogen) at -80°C.

RNA sequencing pipeline

We combined serial analysis of gene expression with strain-specific RNA sequencing to characterize the genome-wide expression for three yeast strains in five temperature conditions at two checkpoints. Total RNAs were obtained from synchronized yeast cell cultures following the TRIzol protocol (Invitrogen Part 15596026). RNA quality and quantity was controlled and measured by NanoDrop instrument (Thermo Scientific). The SOLiD SAGE with Barcoding Adaptor kit (Life Technologies Part 4452811) was used to convert 1 μg of total RNAs into sequencing libraries. In short, messenger RNAs were first bound to oligo-dT beads and primed for cDNA synthesis. A restriction enzyme *Nla*I processed cDNAs at the specific sites of "CATG" sequence. Adapters were attached to the 5-prime and 3-prime end of the tags through a series of cut-and-ligation steps. The final product was a library of tags, in which each tag contained a 22-27 nucleotide sequence of the messenger RNA inserted between two adapter sequences. The libraries were barcoded and sequenced using SOLiD 4 systems (Applied Biosystems, Life Technologies) at the Penn Genome Frontiers Institute (PGFI, University of Pennsylvania) with assistance from the manufacturer's instructions.

The sequencing run returned reads with high-quality base calling. A total of sixty libraries had an average of 5.3 million reads per sample, ranging from 2.5 million to 10 million reads. Reads were 35 base pairs long, which suggested that they might carry a portion of the adapters in the opposite end during sequencing. To assure that the reads captured only the mRNAs, we decided to trim the reads to stringent 22 base-pair sequences.

Alignment of sequencing reads to the Saccharomyces cerevisiae Reference Genome

Sequencing reads aligned to the reference genome S288c downloaded from the *Saccharomyces* Genome Database (SGD, http://www.yeastgenome.org, version R64). All alignments were performed using the Bowtie aligner with no more than two mismatches allowed per sequencing read [40].

The reads were initially aligned to a list of 5548 protein-coding genes of the yeast genome, which only included open-reading frames (ORFs) with evident proteins (SGD, Jan 2012). The annotated sequence of each gene consisted of the coding sequences inserted between the 5' and 3' untranslated regions determined in yeast by Xu *et al* [41]. This step was to compute the number of reads uniquely mapped to the mRNAs.

The strand-specific sequencing procedure allowed us to determine whether the reads came from mRNAs or its antisense. An average of 42.5% of reads mapped to the sense strand of the genes and ~3% of reads mapped to the antisense strand of the genes. These numbers were consistent with previous reports [10,11].

To examine the unmapped reads, we aligned them to the whole genome. There were approximately 10% of reads that mapped uniquely to the genome, mostly to the intergenic regions. In addition, an average of 16% of reads mapped to multiple loci in the genome. Overall, we were able to align ~72% of reads to the yeast reference genome.

Genome-wide gene expression profiling

In the yeast transcriptome, we found 4344 genes with at least one read mapping uniquely in all libraries. Using the number of reads mapping uniquely to the transcriptome, we computed the gene expression profiling in each sample (i.e. the read count per gene for all the genes). We filtered and normalized the read counts to adjust for the differences in mean and variation across the libraries. The normalization followed a method using negative-binomial distribution to correct for over-dispersion in count data [42]. The normalized read count for each gene represented the expression of that particular gene in the samples. Gene expression profiling was performed in three strains across five different temperatures (18-34°C with 4°C interval) at two checkpoints of the cell-division cycle (G1/S and M/G1). Each combination of strain, temperature, and checkpoint had two biological replicates. However, in our assessment, we decided to leave out one sample (YPS2073 at 34°C at M/G1 checkpoints, biological replicate A) due to inconsistent clustering with the other samples. In total the data included 59 gene expression profiles, each containing expression levels for 4344 genes.

Generalized Linear Model of gene expression

The dataset was analyzed as a two-way random model in which three strains and five temperatures were included. The model was performed at each checkpoint separately. All ANOVA analyses were carried out using PROC GLM in SAS (SAS Institute, Cary, NC, USA) [8,9].

For each gene, we performed two-way Analysis of Variance (ANOVA) to determine the effect of temperature and strain on the expression of the gene. We classified genes as showing temperature responses if the effect of temperature on expression was statistically significant after adjusting for the number of genes with a False Discovery Rate (FDR) at 0.05. In a subsequent analysis, we tested three comparisons for significant temperature responses at particular temperatures: 18°C against optimal temperatures (22, 26, 30°C), 34°C against optimal temperatures (22, 26, 30°C), and 26°C against near-optimal temperatures (22 and 30°C). We also performed one-way ANOVA to test the strain-specific significant temperature responses in each strain.

To test the temperature response curve of gene expression, we applied a linear regression and a quadratic regression of gene expression as a function of temperature gradient. If the linear regression and the quadratic regression of gene expression on temperature were significant, we used an F-statistic to assess the increase in coefficient of determination from linear to quadratic regression. We set up rules to define the temperature response curve of each gene: (i) If both regressions were significant and the F-statistic was significant, genes would have quadratic responses; (ii) If both regressions were significant but the F-statistic was not, genes would have linear responses; (iii) If quadratic regression was significant and linear regression was not, genes would have quadratic

responses; (iv) If quadratic regression was not significant and linear regression was significant, genes would have linear responses. Most analyses were performed using PROC REG in SAS.

Gene ontology analysis

Gene enrichment analyses were performed using the *Saccharomyces* Genome Database (SGD, www.yeastgenome.org) with the list of 4344 genes as the background set. All significant tests were adjusted for multiple testing with FDR 0.05.

Transcription factor analysis

To test the evolution of *trans* effects, we obtained transcriptional regulators of a gene from a curated repository (YEASTRACT, http://www.yeastract.com), which compiled a list of transcription factors and their target genes from thousands of references. We only considered the association between the transcription factors and their target genes if there was evidence of direct interaction.

Estimating evolutionary rates in the coding sequence (CDS), the 5' cis-element and the 3' cis-element region.

To investigate the molecular evolution of the three yeast strains, we used data from a genome sequencing project in our lab to assemble the draft genomes. The genome project sequenced many yeast strains (including the three yeast strains in this study) in Illumina paired-end reads of 100 base-pairs. Genome assembly was performed using the SPAdes package [43]. Only large high-quality contigs (length larger than 500 base pairs, average coverage higher than 20 reads per site) were retained for subsequent analyses. Whole-

genome alignments were performed using Lastz [44] and were used to estimate evolutionary rates against *Saccharomyces paradoxus*, a sister species of *S. cerevisiae*.

For each gene, we computed the evolutionary rates in the 5' *cis*-element, coding sequence, and 3' *cis*-element regions. We defined the 5' and 3' *cis*-elements as the 500 base-pairs upstream and downstream of the coding sequence. If the *cis*-elements overlapped with the coding sequence of the neighbor genes, we removed the overlapping segment from the *cis*-elements. The Li-Pamilo-Bianchi method was used to compute the synonymous and non-synonymous rates in the coding sequence [17,18]. The Kimura 2-parameter model was used for the 5' and 3' *cis*-element regions [19-21].

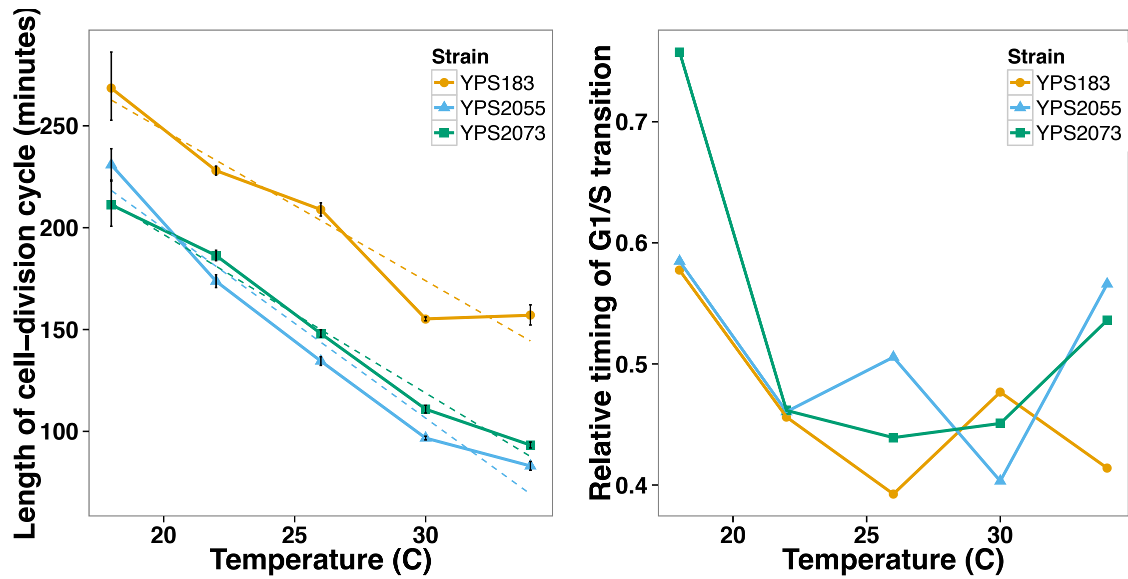

Figure 2.1 Phenotypic measures of yeast cell-division cycle as a function of strain and temperature. (A) Length of cell-division cycle is a linear function of temperature; dashed lines represent the fitted linear regression. The bars represent the standard errors estimated at each checkpoint. (B) Relative timing of the G1/S transition to the cell-division cycle as a function of temperature. The relative timing of the G1/S transition is determined by budding index.

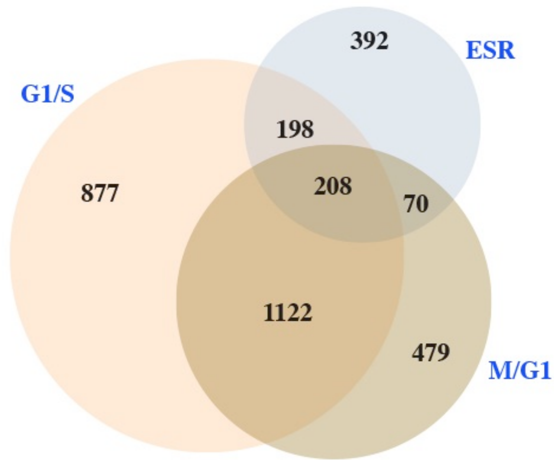

Figure 2.2 Venn diagram showing the number of genes shared by three subsets: G1/S - genes with significant temperature effects at G1/S checkpoint, M/G1 - genes with significant temperature effects at M/G1 checkpoint, ESR - genes in the Environmental Stress Responses program

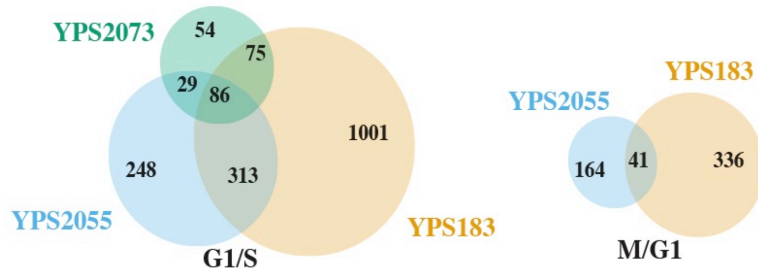

Figure 2.3 Number of genes showing significant strain-specific temperature effects in three yeast strains: YPS183 - a derivative of the laboratory strain S288c, YPS2055 - a natural strain collected in Pennsylvania, YPS2073 - a natural strain collected in New Jersey. Left diagram: at the G1/S checkpoint. Right diagram: at the M/G1 checkpoint.

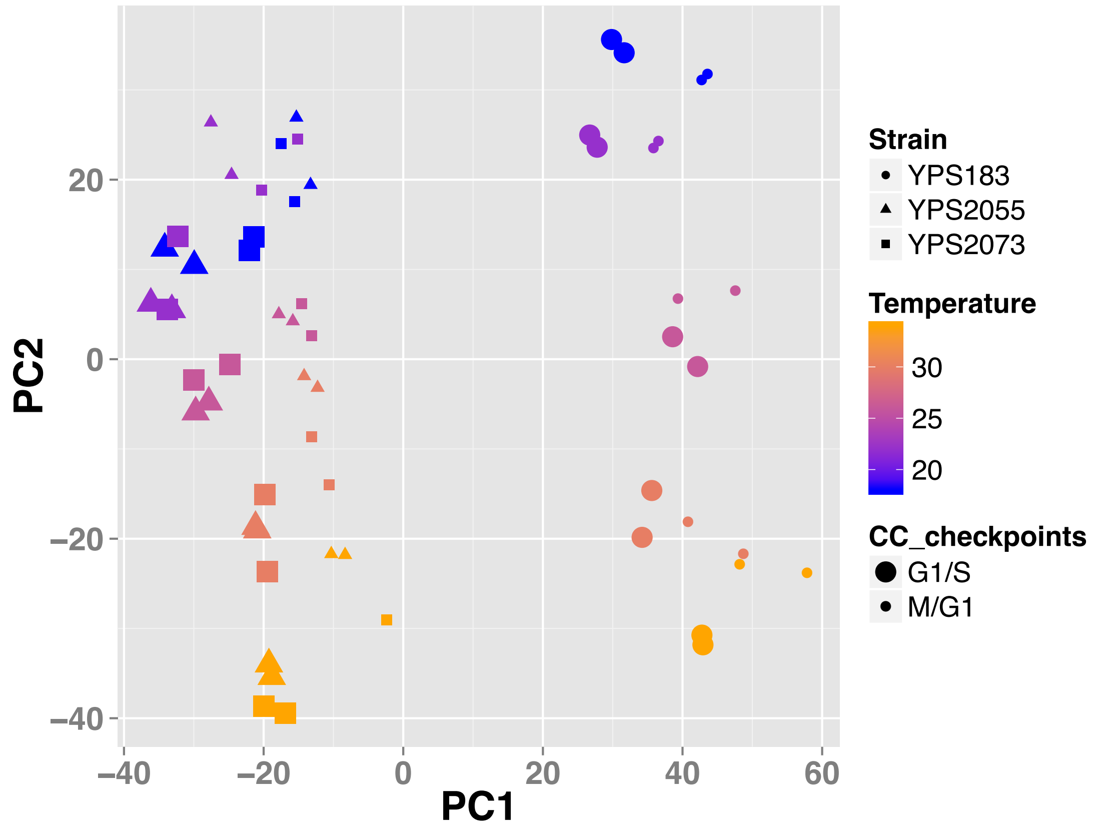

Figure 2.4 Steady-state expression responses to temperature in multi-dimensional transcriptome state space. The two dimensions showing greatest variation among samples are consistent with the direction of genetic background and decreasing temperature. The x- and y-axis are two linear combinations of all the expressed genes in the genome. Each point is a sample with a description of strain by shape of the point, temperature by color, and checkpoint by size.

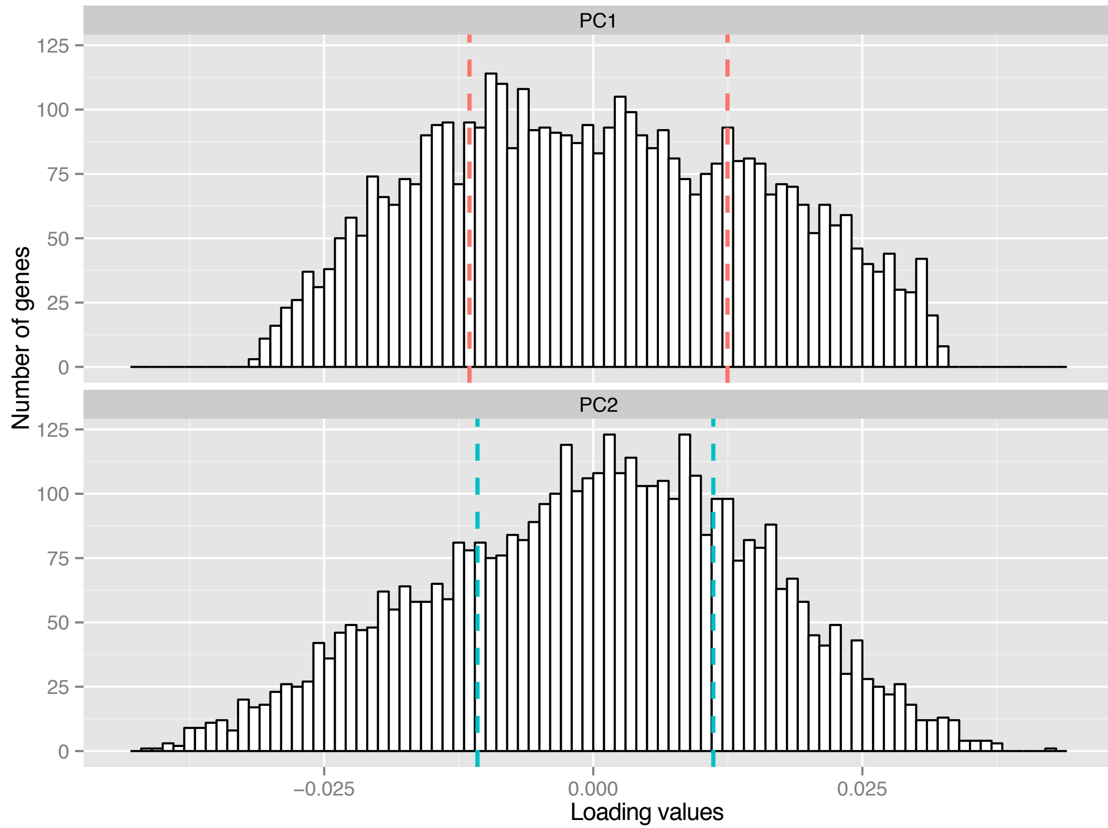

Figure 2.5 Distribution of the loading values of genes onto the two first principal components in Figure 2.4. The dashed lines represent the 25% and 75% percentile in each histogram.

Table 2.1 Gene Ontology (GO) terms overrepresented in genes with significant temperature effects in natural strains. Orange shading indicates GO terms involving molecular functions. Blue shading indicates GO terms involving biological processes; green shading indicates GO terms involving cellular components.

GOID	GO terms	*p*-value
3735	structural constituent of ribosome	0.044620001
313	organellar ribosome	0.00269666
5761	mitochondrial ribosome	0.00269666
5759	mitochondrial matrix	0.003183219
44444	cytoplasmic part	0.004534992
44391	ribosomal subunit	0.005923138
315	organellar large ribosomal subunit	0.017029186
5762	mitochondrial large ribosomal subunit	0.017029186
5739	mitochondrion	0.01787797
5737	cytoplasm	0.018187716
44429	mitochondrial part	0.019186734
5840	ribosome	0.038137759

Table 2.2 Gene Ontology (GO) terms overrepresented in genes with significant temperature effects at the G1/S checkpoint. Orange shading indicates GO terms involving molecular functions; blue shading indicates GO terms involving biological processes; green shading indicates GO terms involving cellular components.

GOID	GO terms	*p*-value
6412	translation	0.000121228
44444	cytoplasmic part	8.40E-07
44445	cytosolic part	4.76137E-06
44391	ribosomal subunit	8.41776E-06
5840	ribosome	1.16469E-05
44429	mitochondrial part	3.20958E-05
5759	mitochondrial matrix	8.0313E-05
5737	cytoplasm	0.000139023
5739	mitochondrion	0.000224559
15935	small ribosomal subunit	0.001817048
502	proteasome complex	0.002284645
22626	cytosolic ribosome	0.002327249
313	organellar ribosome	0.002763192
5761	mitochondrial ribosome	0.002763192
31597	cytosolic proteasome complex	0.003683997
34515	proteasome storage granule	0.003683997
44425	membrane part	0.003943549
314	organellar small ribosomal subunit	0.003944338
5763	mitochondrial small ribosomal subunit	0.003944338
5829	cytosol	0.003951878
30529	ribonucleoprotein complex	0.00400091
42175	nuclear outer membrane-endoplasmic reticulum membrane network	0.004946883
12505	endomembrane system	0.006746925
31090	organelle membrane	0.007691033
15934	large ribosomal subunit	0.010790339
43229	intracellular organelle	0.014245068
16021	integral to membrane	0.015129658
5789	endoplasmic reticulum membrane	0.016935047
43226	organelle	0.017392837
31224	intrinsic to membrane	0.017877282
44432	endoplasmic reticulum part	0.025977008
31967	organelle envelope	0.038862507
31975	envelope	0.038862507
44424	intracellular part	0.041665451
5783	endoplasmic reticulum	0.044121624
3735	structural constituent of ribosome	0.000160826
5198	structural molecule activity	0.023150009

Table 2.3 Gene Ontology (GO) terms overrepresented in genes with significant temperature effects at the M/G1 checkpoint. Green shading indicates GO terms involving cellular components.

GOID	GO terms	*p*-value
315	organellar large ribosomal subunit	0.027263189
5762	mitochondrial large ribosomal subunit	0.027263189

Table 2.4 Gene Ontology (GO) terms overrepresented in genes with significant temperature effects at 18°C and the G1/S checkpoint. Green shading indicates GO terms involving cellular components.

GOID	GO terms	*p*-value
44444	cytoplasmic part	0.000582246
5737	cytoplasm	0.002886197
44429	mitochondrial part	0.00552509
44445	cytosolic part	0.006625891
502	proteasome complex	0.006805286
5739	mitochondrion	0.011339454
31597	cytosolic proteasome complex	0.012298702
34515	proteasome storage granule	0.012298702

Table 2.5 Gene Ontology (GO) terms overrepresented in genes with significant temperature effects at 34°C and the G1/S checkpoint. Orange shading indicates GO terms involving molecular functions; blue shading indicates GO terms involving biological processes; green shading indicates GO terms involving cellular components.

GOID	GO terms	*p*-value
2181	cytoplasmic translation	0.000165008
6412	translation	0.000216079
3735	structural constituent of ribosome	1.01145E-06
5198	structural molecule activity	0.00328341
31072	heat shock protein binding	0.010572543
44391	ribosomal subunit	2.31E-07
5840	ribosome	3.51E-07
22626	cytosolic ribosome	1.066E-06
30529	ribonucleoprotein complex	4.55843E-05
44445	cytosolic part	7.28726E-05
15935	small ribosomal subunit	0.000119686
5829	cytosol	0.000153273
22627	cytosolic small ribosomal subunit	0.008202371
15934	large ribosomal subunit	0.010190619
22625	cytosolic large ribosomal subunit	0.015731078
5759	mitochondrial matrix	0.043088045

Table 2.6 Gene Ontology (GO) terms overrepresented in genes with significant temperature effects at 34°C and the M/G1 checkpoint. Orange shading indicates GO terms involving molecular functions; green shading indicates GO terms involving cellular components.

GOID	**GO terms**	***p*-value**
51082	unfolded protein binding	0.000680325
44445	cytosolic part	0.001327186
22626	cytosolic ribosome	0.02268602
5829	cytosol	0.027909885
5840	ribosome	0.038855057

Table 2.7 Genes showing significant strain-specific temperature effects in the three yeast strains (Benjamini-Hochberg correction, *p*-value<0.05)

	YPS183	**YPS2055**	**YPS2073**
G1/S	1475	676	244
M/G1	377	205	0

Table 2.8 Evolutionary rates in the 5’ *cis*-element, CDS, and 3’ *cis*-element regions in strain-specific responsive genes. Strain-specific responsive genes are genes showing significant temperature effects in specific strains. Common genes are strain-specific responsive genes shared in all three strains (Figure 2.3). The mean and two standard errors are shown in the table. Molecular evolution rates are estimated using *S. paradoxus* as the reference outgroup. Significance levels of Wilcoxon rank-sum test: (*): p-value<0.05; (**):p-value<0.01; (***):p-value<0.001.

	YPS183			YPS2055			YPS2073		
	dN/dS	**d5/dS**	**d3/dS**	**dN/dS**	**d5/dS**	**d3/dS**	**dN/dS**	**d5/dS**	**d3/dS**
Specific	0.12±0.01*	0.76±0.07	0.71±0.05	0.13±0.02*	0.65±0.05	0.65±0.04	0.13±0.02***	0.69±0.10	0.65±0.07
Common	0.09±0.02	0.68±0.17	0.67±0.18	0.09±0.02	0.67±0.16	0.67±0.11	0.09±0.02	0.68±0.16	0.67±0.18

Table 2.9 Evolutionary rates in the 5' *cis*-element, CDS, and 3' *cis*-element regions in the transcription factors of two classes of genes (the common responsive genes—TF-common, and the strain-specific responsive genes—TF-specific), the general transcription factors (TF-general) and non-transcription factors (Non-TFs). The mean and two standard errors are showed in the table. Molecular evolution rates are estimated using *S. paradoxus* as the reference outgroup. Significance levels of Wilcoxon rank-sum test: (*): p-value<0.05; (**):p-value<0.01; (***):p-value<0.001.

	YPS183			YPS2055			YPS2073		
	dN/dS	**d5/dS**	**d3/dS**	**dN/dS**	**d5/dS**	**d3/dS**	**dN/dS**	**d5/dS**	**d3/dS**
Non-TFs	0.146±0.004	0.700±0.024***	0.667±0.017	0.146±0.004	0.692±0.021***	0.661±0.018	0.145±0.004	0.696±0.024***	0.663±0.019
TF-general	0.141±0.027	0.557±0.064	0.588±0.070	0.143±0.028	0.549±0.065	0.597±0.069	0.145±0.026	0.541±0.063	0.615±0.068
TF-common	0.154±0.015	0.450±0.030	0.552±0.047	0.154±0.016	0.447±0.031	0.548±0.046	0.155±0.016	0.446±0.031	0.548±0.046
TF-specific	0.192±0.043	0.478±0.074	0.641±0.106	0.174±0.034	0.475±0.075	0.602±0.108	0.166±0.039	0.479±0.072	0.568±0.116

Table 2.10 Number of genes showing linear and quadratic responses in the three yeast strains.

	fitting model	YPS183	YPS2073	YPS2055	all strains
G1/S	quadratic	569	219	207	17
	linear	568	236	502	51
M/G1	quadratic	198	21	16	0
	linear	244	260	276	21

REFERENCES

[1] de Nadal, E., Ammerer, G. and Posas, F. (2011) Controlling gene expression in response to stress. *Nat Rev Genet* ***12***, 833–845.

[2] Verghese, J., Abrams, J., Wang, Y. and Morano, K. A. (2012) Biology of the heat shock response and protein chaperones: budding yeast (*Saccharomyces cerevisiae*) as a model system. *Microbiology and Molecular Biology Reviews* ***76***, 115–158.

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

[4] Schade, B. (2004) Cold adaptation in budding yeast. *Molecular Biology of the Cell* ***15***, 5492–5502.

[5] Bertoli, C., Skotheim, J. M. and de Bruin, R. A. M. (2013) Control of cell cycle transcription during G1 and S phases. *Nat Rev Mol Cell Biol* ***14***, 518–528.

[6] Sweeney, J. Y., Kuehne, H. A. and Sniegowski, P. D. (2004) Sympatric natural and populations have different thermal growth profiles. *FEMS Yeast Research* ***4***, 521–525.

[7] Vanoni, M., Vai, M. and Frascotti, G. (1984) Effects of temperature on the yeast cell cycle analyzed by flow cytometry. *Cytometry* ***5***, 530–533.

[8] Sokal, R. R. and Rohlf, F. J. (1995) Biometry.

[9] Saxton, A. M.SAS Institute. (2004) Genetic analysis of complex traits using SAS.

[10] Xu, Z., Wei, W., Gagneur, J., Clauder-Münster, S., Smolik, M., Huber, W. and Steinmetz, L. M. (2011) Antisense expression increases gene expression variability and locus interdependency. *Mol Syst Biol 7*.

[11] Goodman, A. J., Daugharthy, E. R. and Kim, J. (2013) Pervasive antisense transcription is evolutionarily conserved in budding yeast. *Mol. Biol. Evol.* ***30***, 409–421.

[12] Causton, H. C., Ren, B., Koh, S. S., Harbison, C. T., Kanin, E., Jennings, E. G., Lee, T. I., True, H. L., Lander, E. S. and Young, R. A. (2001) Remodeling of yeast genome expression in response to environmental changes. *Molecular Biology of the Cell* ***12***, 323–337.

[13] Sniegowski, P. D., Dombrowski, P. G. and Fingerman, E. (2002) *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* coexist in a natural woodland site in North America and display different levels of reproductive isolation from European conspecifics. *FEMS Yeast Research* ***1***, 299–306.

[14] Simola, D. F., Francis, C., Sniegowski, P. D. and Kim, J. (2010) Heterochronic evolution reveals modular timing changes in budding yeast transcriptomes. *Genome Biol.* ***11***, R105.

[15] Reményi, A., Schöler, H. R. and Wilmanns, M. (2004) Combinatorial control of gene expression. *Nat Struct Mol Biol* ***11***, 812–815.

[16] Hahn, S. and Young, E. T. (2011) Transcriptional regulation in *Saccharomyces cerevisiae*: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics* ***189***, 705–736.

[17] Pamilo, P. and Bianchi, N. O. (1993) Evolution of the Zfx and Zfy genes: rates and

interdependence between the genes. *Mol. Biol. Evol.* ***10***, 271–281.

[18] Li, W.-H. (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J Mol Evol* ***36***, 96–99.

[19] Kimura, M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* ***16***, 111–120.

[20] Kohn, M. H. (2003) Inference of positive and negative selection on the 5' regulatory regions of *Drosophila* genes. *Mol. Biol. Evol.* ***21***, 374–383.

[21] Osada, N. (2005) Substitution rate and structural divergence of 5'UTR evolution: comparative analysis between human and cynomolgus monkey cdnas. *Mol. Biol. Evol.* ***22***, 1976–1982.

[22] Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. A., Davey, R. P., Roberts, I. N., Burt, A., Koufopanou, V., et al. (2009) Population genomics of domestic and wild yeasts. *Nature* ***458***, 337–341.

[23] Daga, R. R., Bolaños, P. and Moreno, S. (2003) Regulated mRNA stability of the Cdk inhibitor Rum1 links nutrient status to cell cycle progression. *Current Biology* ***13***, 2015–2024.

[24] Berthoumieux, S., de Jong, H., Baptist, G., Pinel, C., Ranquet, C., Ropers, D. and Geiselmann, J. (2013) Shared control of gene expression in bacteria by transcription factors and global physiology of the cell. *Mol Syst Biol* ***9***.

[25] Gerosa, L., Kochanowski, K., Heinemann, M. and Sauer, U. (2013) Dissecting specific and global transcriptional regulation of bacterial gene expression. *Mol Syst Biol* ***9***.

[26] Guan, Q., Haroon, S., Bravo, D. G., Will, J. L. and Gasch, A. P. (2012) Cellular memory of acquired stress resistance in *Saccharomyces cerevisiae*. *Genetics* ***192***, 495–505.

[27] Lewis, J. A., Elkon, I. M., McGee, M. A., Higbee, A. J. and Gasch, A. P. (2010) Exploiting natural variation in *Saccharomyces cerevisiae* to identify genes for increased ethanol resistance. *Genetics* ***186***, 1197–1205.

[28] Bateson, P., Bateson, P. P. G. and Gluckman, P. (2011) Plasticity, Robustness, Development and Evolution.

[29] Kaneko, K. (2012) Phenotypic plasticity and robustness: evolutionary stability theory, gene expression dynamics model, and laboratory experiments. *Adv. Exp. Med. Biol.* ***751***, 249–278.

[30] Cooper, T. G., Gorski, M. and Turoscy, V. (1979) A cluster of three genes responsible for allantoin degradation in Saccharomyces cerevisiae. *Genetics* ***92***, 383–396.

[31] Nurse, P. and Thuriaux, P. (1977) Controls over the timing of DNA replication during the cell cycle of fission yeast. *Experimental Cell Research* ***107***, 365–375.

[32] Gimeno, C. J., Ljungdahl, P. O., Styles, C. A. and Fink, G. R. (1992) Unipolar cell divisions in the yeast S. cerevisiae lead to filamentous growth: Regulation by starvation and RAS. *Cell* ***68***, 1077–1090.

[33] Amaral, P. P., Dinger, M. E. and Mattick, J. S. (2013) Non-coding RNAs in homeostasis, disease and stress responses: an evolutionary perspective. *Briefings in Functional Genomics* ***12***, 254–278.

[34] Tufarelli, C., Stanley, J. A. S., Garrick, D., Sharpe, J. A., Ayyub, H., Wood, W. G. and Higgs, D. R. (2003) Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nat. Genet.* ***34***, 157–165.

[35] Imamura, T., Yamamoto, S., Ohgane, J., Hattori, N., Tanaka, S. and Shiota, K. (2004) Non-coding RNA directed DNA demethylation of Sphk1 CpG island. *Biochemical and Biophysical Research Communications* ***322***, 593–600.

[36] Faghihi, M. A. and Wahlestedt, C. (2009) Regulatory roles of natural antisense transcripts. *Nat Rev Mol Cell Biol* ***10***, 637–643.

[37] Castelnuovo, M., Rahman, S., Guffanti, E., Infantino, V., Stutz, F. and Zenklusen, D. (2013) Bimodal expression of PHO84 is modulated by early termination of antisense transcription. *Nat Struct Mol Biol* ***20***, 851–858.

[38] Carroll, S. B. (2008) Evo-Devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* ***134***, 25–36.

[39] Judith B Zaugg, N. M. L. (2012) A genomic model of condition-specific nucleosome behavior explains transcriptional activity in yeast. *Genome Res.* ***22***, 84.

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

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

[42] Anders, S. and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol.* ***11***, R106.

[43] Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., et al. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* ***19***, 455–477.

[44] Harris, R. S. Improved pairwise alignment of genomic DNA. *Ph.D. Thesis, The Pennsylvania State University*.

CHAPTER 3

ROBUSTNESS OF GENOME-WIDE GENE EXPRESSION IN NATURAL STRAINS OF THE BUDDING YEAST

3.1 INTRODUCTION

Organisms may be viewed as complex systems; they can be highly sensitive to specific environmental signals yet are also resilient to both severe and broad environmental disturbances as well as to the constant barrage of mutations that affect their functional configuration. For example, there are many genes which are not only non-essential for viability but which also have little effect on the phenotype when inactivated [1,2]. This property of resilience of a biological system to unwanted perturbations is called robustness, defined here as the persistence of a standard, or "wild type", phenotype against genetic, environmental, and stochastic perturbations. One key feature of robustness is its presence at all levels of biological organization: molecular structures, gene expression, developmental programs, and ultimately the fitness of organisms [3]. Despite the ubiquity of robustness and its fundamental importance in the maintenance of individual and collective life, the broad principles and specific molecular mechanisms conferring robustness upon individual organisms are still poorly understood.

Robustness is the consistency of a trait in the face of perturbation, relative to some mean value for a reproducing population in a specific environment. It is usually classified by the source of perturbation: either genetic, environmental, or stochastic. Genetic robustness concerns the reduced variation of trait values in the face of heritable changes in DNA sequence, specifically mutations. Environmental robustness refers to the reduced

variation in the face of non-heritable changes external to the organism, such as temperature, sunlight, osmolarity, or chemical stimuli. The typical measure of robustness is the inverse of the variance of the trait when perturbed.

The evolution of environmental robustness can be viewed simply as the consequence of adaptation during multiple recurrences of environmental changes [4]. The evolution of genetic robustness is harder to formalize. Currently, three main models have been proposed: intrinsic evolution, adaptive evolution, and congruent evolution of genetic robustness [3-6]. Under intrinsic evolution, genetic robustness arises as the coupling component of the phenotype itself: it is required for the selection of the phenotype. For example, the selection for a dominant trait would also lead to mutational robustness in the recessive alleles, since mutations in the recessive alleles would not affect the outcome of the traits. The adaptive hypothesis argues that natural selection can act directly on genetic robustness as a heritable trait. A simple scenario of the latter model is that selection for the phenotype occurs first and then subsequently selection for robustness of this phenotype in the face of mutation. For these two scenarios, genetic robustness and environmental robustness are independently evolved in different mechanisms. In contrast, the congruence hypothesis posits that genetic robustness evolves as a by-product of evolution for environmental robustness. In this scenario, one would expect that traits robust to environmental changes may also be robust to mutations. Despite their differences, the three scenarios are not mutually exclusive.

In general, robustness can be examined indirectly using two approaches that differ according to the evolutionary time scale under consideration. One approach considers the

phenotypic effects of short-term evolutionary divergence of well-defined populations under controlled conditions. This method provides measures of traits in the starting populations as a reference to determine evolved robustness. However, monitoring controlled evolution experiments over hundreds to thousands of generations is typically intractable except for certain bacterial species and viruses [7,8]. Therefore, this approach is typically unable to detect evolutionary changes in the robustness of a trait. Here, the measure of robustness is the insensitivity or invariance of trait values despite accumulation of mutations. That is, mutational robustness. The second approach involves using comparative analyses of wild populations to examine the long-term evolution of robustness in the natural environment. The source of perturbation in this approach is the difference of genetic backgrounds in natural strains, which may have accumulated multiple mutations under a full spectrum of evolutionary modes. However, the lack of knowledge of ancestral states hinders the ability to make inferences about robustness of traits in present states.

In the present study, we used two natural strains and one laboratory strain of the budding yeast *Saccharomyces cerevisiae* to investigate the evolution of robustness with respect to genetic perturbations and temperature fluctuations. The laboratory strain was obtained from a natural isolate of budding yeast (EM93) and first cultivated in laboratory condition 70 years ago [9,10]. The natural strains in our study were collected from oak trees in northeast America and estimated to diverge around 3 to 4 thousand years ago [11,12]. While the natural strains have experienced continual natural selection pressure, the laboratory strain has evolved in comparatively relaxed conditions. Therefore, the

laboratory strain can serve as an excellent reference point for the evolution of robustness in natural populations. Using the levels of gene expression as the traits of interest, we are able to study a large number of traits simultaneously with over six thousand genes in the yeast transcriptome. A further review of previous studies on the evolution of genetic and temperature robustness of gene expression in budding yeasts was discussed extensively in Chapter 1.

3.2 STATISTICAL ASSOCIATION BETWEEN TEMPERATURE AND GENETIC VARIANCE OF GENOME-WIDE GENE EXPRESSION SUPPORTS THE CONGRUENT EVOLUTION OF THESE TWO VARIANCE COMPONENTS

Coupling among genetic, environmental, and stochastic robustness has been studied using multiple datasets [6,13]. However, because the genetic and environmental robustness were usually computed from different datasets or approaches, only moderate correlations were found. Here we investigated the correlation between genetic and temperature robustness in a nested-factor experimental design using a novel dataset of genome-wide gene expression dynamics for three yeast strains and five temperature conditions (details described in Materials and Methods). To analyze the data for each gene, we applied a generalized linear model and estimated variance components for genetic and temperature factors, including an interaction term between genetic and temperature factors if it was significant. To correct for effects of gene-to-gene variability of mean expression on variance components, we scaled the genetic and temperature variance components by the

residual variance component to give us a scaled genetic and temperature variance for each gene in the transcriptome.

We first examined the genome-wide association between the genetic and temperature variances in the dataset. Our results indicated a significant correlation between genetic variance and temperature variance (Spearman's rank correlation coefficient ρ=0.22, p-value=2.2×10^{-16}, n=6090). The magnitude of the correlation coefficient is consistent with results found in previous studies [6,14].

While the large number of genes involved in these data provided strong statistical power, we were concerned with the moderate degree of association. One possibility is that confounding factors could mask stronger association. For example, the genetic variance estimated for each individual gene can be partitioned into three components. The first component involves the between strain variation that may play a role in strain-specific function of a gene, possibly through past selective adaptation. The second component may arise as a by-product of system-level plasticity and robustness to environmental input. The last component involves random noise along with neutral divergence.

Since both temperature-related variation and strain-related variation may have multiple components, we hypothesized that different functional classes of genes may exhibit different levels of robustness for both environmental and genetic variation. To examine whether different functional classes consequently have different relationships between genetic and temperature variance, we dissected levels of genetic and temperature variances and their correlation for four classes of genes: genes classified as essential (i.e. genes required for growth in rich media), transcription factors, TATA-box dependent

genes, and genes with a high degree of protein-protein interaction. For each gene class, we computed the genetic and environmental variance components for the foreground genes (i.e., genes within each class) and compared the values against a set of background genes selected as specific control for each class. We next examined the correlation between the two variances in the foreground and background sets. All results are summarized in Table 3.1.

Essential genes are genes that are indispensable for growth in normal conditions. In budding yeast, roughly a thousand genes in the transcriptome are required for growth in rich glucose medium and have therefore been defined as essential genes [1,15]. Because of their importance, we postulated that essential genes would have overall lower expression variability compared to the nonessential genes. We compared a foreground set of essential genes with a background set of nonessential genes. While the temperature variances were similar (Table 3.1; 1.31 in the foreground versus 1.44 in the background), the genetic variance of the foreground gene set was significantly smaller than that of the background (Table 3.1; 2.87 versus 4.44; Wilcoxon rank-sum test, p-value=$3x10^{-6}$). The smaller genetic variance in essential genes is consistent with the lethal effect of these genes. Therefore, for this class of genes there is evidence for stabilizing selection maintaining particular levels of gene expression despite independent divergence across the lineages. The lack of significant difference for the environmental variance shows that these essential genes respond to temperature perturbations in a similar manner to non-essential genes. This suggests two possibilities. First, the variability in gene expression as a function of temperature may be a simple property of biochemical reaction rates, which

is unlikely to differentially affect essential versus non-essential genes. However, results described in Chapter 2 suggest that different genes have specific patterns of temperature dependent response (e.g., a quadratic form with a peak or valley at middle temperatures); therefore, it is unlikely that most of the temperature dependent expression variation is due to simple physical changes in reaction rates. Second, the variability in gene expression as a function of temperature may involve system-level adjustment of gene expression to either optimize cellular function (e.g., cell division) at particular temperatures or to maintain homeostasis of normal function. Our results suggest that there is no broad difference in the degree of expression plasticity between essential and non-essential genes as components of such system-level dynamics. Thus, essentiality of a gene in terms of allowing cell growth does not manifest as special patterns of variation under temperature related fluctuations.

Genetic regulation involves transcription factors binding to the *cis*-regulatory region of a gene to regulate its expression. Maintaining robust expression of transcription factors is important to proper cell growth and division. Therefore, changes in expression of the transcription factors would subsequently impact both the level and timing of their target gene [12,16]. Here we assessed two types of variance in the foreground set of transcription factors and the background set of genes that are not transcription factors. Table 3.1 shows that transcription factors have lower variance for both temperature and genetic components compared to non-transcription factors but the difference is not significant (mean temperature variance of 1.00 in the foreground versus 1.42 in the background, Wilcoxon rank-sum test, p-value=0.25; mean genetic variance of 3.64 in the

foreground versus 4.13 in the background, Wilcoxon rank-sum test, *p*-value=0.7). One possibility for lack of statistical significance is the large difference in the sample size of the transcription factors versus non-transcription factors. The observation of lower genetic variation, albeit not statistically significant, was consistent with weakly purifying selection forces in the 5' *cis*-element of transcription factors as we reported in Chapter 2. Thus, because of their potential for pleiotropic effects over large number of genes, we hypothesize that there may be stronger stabilizing selection operating on the transcription factors' gene expression.

In addition to *trans* factors, major core *cis*-regulatory elements in gene promoters regulate the mechanism of transcription regulation, including the TATA box, the initiator (INR), and downstream promoter element (DPE). In yeast, recruitment of a general transcription factor TBP, TATA-binding protein, to gene promoters involves two distinct transcription co-activator complexes, TFIID and SAGA [17,18]. TFIID comprises of TBP and multiple conserved Tafs (i.e. TBP-associated factors), which is required for transcription activation from promoters without TATA box sequence [19,20]. SAGA is also a TBP-dependent co-activator complex of twenty subunits. Despite sharing several common Tafs with TFIID pathway, the SAGA complex mostly regulates expression of TATA-containing genes that are highly inducible [20,21]. Findings presented in Basehoar *et al* indicate that genes containing TATA box sequence in the promoter were related to stress response activities and highly regulated, whereas genes without TATA box sequence were found to be involved in housekeeping functions and less regulated [22]. Here we compared a foreground set of TATA-containing genes with a background set of

TATA-less genes. Interestingly, both the mean genetic variance and temperature variance of the TATA-containing genes were significantly larger than that of the background (Table 3.1; temperature variance of 2.26 versus 1.23, Wilcoxon rank-sum test, *p*-value=7.2×10^{-11}; genetic variance of 8.13 versus 3.26, Wilcoxon rank-sum test, *p*-value=2.2×10^{-16}). Our results are concordant with previous findings that the presence of TATA box sequence confers higher gene expression sensitivity to mutations, environmental perturbations, and stochastic noise [20,23].

In yeast, deletion of a protein in the protein-protein interaction network likely causes various degrees of deleterious effect depending on position of the protein in the interaction network [24]. Highly connected proteins interact with many different proteins and consequently have broad effects on the protein-protein interaction network when their levels change. Previous studies also reported that highly connected proteins have the capability to buffer environmental and genetic changes in *Saccharomyces cerevisiae* [6,25]. To examine how the position of proteins in an interaction network confers robustness, we classified genes by the number of direct protein-protein interactions. A gene was described as highly connected if its protein could interact with a large number of other proteins, specifically, if the number of interactions is in the top tenth percentile genome-wide. Genes with only one direct protein-protein interaction were considered lowly connected genes. Direct protein-protein interactions in *Saccharomyces cerevisiae* were collected from the BioGRID database of protein and genetic interactions (BIOGRID-ORGANISM-Saccharomyces_cerevisiae-3.1.93.tab2.txt)[26]. We compared the temperature and genetic variances in the foreground set of highly connected genes

with the background set of lowly connected genes. While the genetic variances were slightly smaller in the foreground than in the background (Table 3.1; 4.06 versus 5.03, Wilcoxon rank-sum test, *p*-value=0.057), the temperature variance was significantly higher in the foreground than in the background (Table 3.1; 1.64 versus 1.43, Wilcoxon rank-sum test, *p*-value=0.015). The magnitudes of the differences are small but the opposite trends observed in two types of variances contrast with the other classes of genes we examined. We hypothesize that highly connected genes will tend to have a potentially greater degree of stabilizing selection due to their pleiotropic effects, similar to the transcription factors, resulting in lower variation amongst the strains. We also hypothesize that when highly connected genes are involved in system-level temperature responses, they might tend towards greater degree of changes to satisfy the more complex stoichiometric relationships. Nevertheless, this last point is highly speculative and the magnitude of differences is too small to have confidence that there is a meaningful class distinction.

Having identified distinct functional classes of genes, we then examined whether the relationship between temperature and genetic variance varies for each class. Overall, the correlation of temperature and genetic variance in each of the foreground sets was consistently higher that of the background sets (Table 3.1). The combined *p*-value over the four different classes contrasting the foreground and background correlation was significantly different (Fisher's combined *p*-value test, *p*-value=0.03). However, individually only the case of TATA-containing genes versus TATA-less genes were statistically significant (Table 3.1; Spearman's correlation coefficient of 0.274 versus

0.197, Welch's t-test, *p*-value=0.02). The classes of genes we study here comprise either those that are essential for fitness or those that are likely to be strongly involved in system-wide gene function (transcription regulation or interacting with other gene products). In both cases, we might expect a stronger level of selective effects compared to the background. Therefore, the stronger correlation in the genetic and environmental variance components see in Table 3.1 for the foreground genes is consistent with the idea that selective effects on either the environmental variation or genetic variation drive the congruent evolution of the two variance components.

3.3 REDUCED EXPRESSION VARIATION IN THE NATURAL STRAINS WITH RESPECT TO THE LABORATORY STRAIN IS CONSISTENT WITH SELECTION FOR ROBUSTNESS OF GENE EXPRESSION

In the next portion of our analysis we examined variance of gene expression in natural strains of budding yeast compared to the laboratory strain. The natural strains were collected directly from natural habitats (oak trees) in the northeast regions of the United States. In this natural environment the strains experience temperature fluctuations during the entire year as well as annual seasonal changes in the average temperature. We postulated that these natural strains may have evolved an enhanced ability to buffer gene expression to better adapt to the changes of temperature (as well as to other factors). In contrast, the standard laboratory strain S288C, while originally derived from a natural strain EM93, has been maintained for over 70 years under relatively constant laboratory conditions (albeit typically experiencing freeze-thaw cycles). Therefore, we hypothesized that the laboratory strain is likely to have experienced relaxed selection for robustness of

cellular function in response to changes in the temperature. We note that robustness of cellular function under temperature fluctuations can be mediated by either robustness (i.e., low level of changes) of gene expression or by system level adjustment of gene action, which may involve a large degree of expression changes of individual genes.

3.3.1 NATURAL STRAINS EXHIBIT GREATER TEMPERATURE ROBUSTNESS COMPARED TO THE LABORATORY STRAIN

To estimate strain-specific temperature variance for each gene, we applied a single-factor linear model for each strain and cell-cycle stage separately with temperature as the independent factor (see Materials and Methods). The laboratory strain had a larger number of genes with significant temperature effects than the natural strains with 1475 and 377 genes with significant temperature effects at G1/S and M/G1 checkpoints, respectively (Benjamini-Hochberg correction, p-value<0.05) while the two natural strains YPS2055 and YPS2073 had 676, 205 and 244, 0 numbers of significant temperature effect genes at the two checkpoints, respectively (Chapter 2, Table 2.1). The natural strains show less than half the number of genes with significant temperature effects compared to the laboratory strain, suggesting that potential cell function robustness under temperature fluctuations is mediated by a mechanism that reduces temperature-dependent variation in gene expression. This mechanism may involve either direct modification of transcriptional regulation of individual genes or modification of transcription factors that *trans*-regulate these genes. We examined the expression variation in transcription factors for the natural strains versus that of the laboratory strain. At the G1/S checkpoint, the expression variation is highest in the laboratory strain, lower in YPS2055 and lowest in

YPS2073 with significant decrease from YPS183 to either natural strain (Table 3.2; mean coefficient of variation of expression among transcription factors in YPS183, YPS2055, YPS2073: 0.824, 0.742, 0.662, respectively; Wilcoxon signed rank test, YPS2073 versus YPS183: *p*-value=0.006, YPS2055 versus YPS183: *p*-value=0.03). However, the expression variation at the M/G1 checkpoint is roughly the same in all three strains (mean coefficient of variation of expression among transcription factors in YPS183, YPS2055, YPS2073: 0.706, 0.677, 0.711, respectively; Wilcoxon signed rank test, YPS2073 versus YPS183: *p*-value=0.67, YPS2055 versus YPS183: *p*-value=0.36). This suggested that decreasing temperature-dependent variation in gene expression is likely modulated via robustness of expression in *trans*-regulatory factors.

Interestingly, YPS2055 had less than half the number of genes with significant effects as the laboratory strain, but had more than double the number of genes with significant effects than YPS2073. This difference was unexpected because the two natural strains were collected within a hundred miles of each other from the same local niche of oak tree exudates. Both YPS2055 and YPS2073 show similar temperature responses in their overall cell cycle timing (Chapter 2, Figure 2.1). However, YPS2073 has significantly delayed G1/S transition relative to the total cell cycle at very low temperatures while YPS2055 shows a profile more similar to that of the laboratory strain (Chapter 2, Figure 2.1). Therefore, these strains might have experienced microenvironment differences in their temperature regimes.

3.3.2 *F*-STATISTICS REVEAL REDUCED EXPRESSION VARIATION IN THE NATURAL STRAINS WITH RESPECT TO THE LABORATORY STRAIN

Originally derived from a vineyard strain, the laboratory strains have been experiencing relaxed selection for robustness to temperature fluctuations under normal growth conditions. Therefore, we hypothesized that the temperature-dependent expression variation of each gene for the laboratory strain may serve as a null model to assess the temperature-dependent expression variation of the natural strains. We computed *F*-statistics as the ratio of the temperature variances in the natural strains to the laboratory strain, scaled by the degrees of freedom. We defined temperature-robust genes as genes having significantly decreased temperature variance compared to the laboratory strain. We identified 70 and 96 robust genes at G1/S and M/G1 checkpoints in YPS2055, and 106 and 94 robust genes in YPS2073, respectively (genes listed in Table 3.6, 3.7, 3.8, and 3.9). Overlapping subsets of 54 and 62 robust genes were common in both strains. Combined, a total of 38 genes were consistently robust in both natural strains at both checkpoints. Overrepresented among these 38 genes with robust expression are genes involved in amino acid, organic acid and oxoacid metabolic processes (Table 3.10).

Intriguingly, the gene ontology terms enriched in the robust genes at G1/S and M/G1 checkpoint were completely different. Robust genes at G1/S checkpoints were enriched in cell periphery components including mostly plasma membrane proteins, whereas robust genes at M/G1 checkpoints were enriched in metabolic processes (Table 3.11 and Table 3.12). Temperature is known to affect the physical state of the lipid bilayer and subsequently change the activities of membrane-bound proteins [27]. Changes in the lipid composition of the membrane are necessary in different environmental conditions [28,29],

but maintaining the activities of membrane proteins may be essential to cellular functions and growth, especially at a key checkpoint like G1/S.

The functional classes of 38 genes robustly expressed in two natural strains at both checkpoints are concordant with our analyses of temperature variance in different classes of genes. Only two of these 38 genes are essential for growth in rich media and none is transcription factor. However, these genes are significantly enriched in TATA-containing genes as compared to the genome-wide proportion of TATA-containing genes (the ratio of TATA-containing genes to TATA-less genes in these 38 genes is 20:18 as compared to a genome-wide ratio of 734:3610, Fisher's exact test, p-value=6.1×10^{-7}).

3.4 MOLECULAR EVOLUTION ANALYSIS INDICATES THE POTENTIAL ROLE OF *TRANS* FACTORS IN REDUCED EXPRESSION VARIATION

Transcriptional regulation involves *cis*-elements in the promoter regions and *trans* factors that recognize and bind to the *cis*-elements to initiate gene expression. Changes in either component not only affect gene expression but also impact the robustness of gene expression under perturbations. Since our analyses suggested that natural strains have evolved robustness of expression against temperature perturbations for dozens of genes, we next investigated the potential role of *cis*-elements and *trans* factors in the evolution of robust gene expression in the natural strains. Specifically, we computed the evolutionary rates in the 5’ *cis*-element, coding sequence (CDS), and 3’ *cis*-element of each gene in the three yeast strains. For these segments, using *Saccharomyces paradoxus* as the outgroup species we computed the dN/dS, d5/dS, and d3/dS ratio of molecular evolution, where dN, d5, d3 denote non-synonymous change, 5’ change, and 3’ change,

respectively, while dS refers to the synonymous rate of change for each gene. We examined three subsets of genes with significantly robust expression: (i) genes with robust expression in all conditions; (ii) genes with robust expression only in YPS2055; and (iii) genes with robust expression only in YPS2073. We tested the significance of the difference in the rates molecular evolution in these genes with robust expression against all genes with more variable expression. If *cis* effects play a role in modulating expression, we might expect the evolutionary rates in the 5' *cis*-elements to show patterns consistent with either directional or purifying selection in the natural strains but not in the laboratory strain.

For genes that showed significantly robust expression in all conditions, we found significant differences in the rate of molecular evolution within the coding sequences of these genes for all three strains. In these genes, the dN/dS ratio within the coding sequence of the genes with more robust expression was significantly lower than that of genes with more variable expression (Table 3.3; 0.101 versus 0.147 in YPS183, Wilcoxon rank sum test, *p*-value=0.015; 0.099 versus 0.148 in YPS2055, Wilcoxon rank sum test, *p*-value=0.011; 0.097 versus 0.147 in YPS2073, Wilcoxon rank sum test, *p*-value=0.002). This pattern is consistent with the idea that genes whose transcript expression seems to be under selection for low variability, as a class, are important for fitness and therefore also experience stronger purifying selection against modification of their amino-acid sequences. Somewhat unexpectedly, the relative evolutionary rates in the 5' *cis*-elements (d5/dS) were similar between the genes with robust expression versus those with more variable expression—suggesting a lack of evidence for evolutionary

changes of the 5' *cis*-sequences to modulate gene expression (Table 3.3). This may be due to three possibilities. First, the 5' region that we examined may be too broadly defined whereas gene expression may be modulated by changes in small binding motifs. Second, the small number of genes in our class of robustly expressed genes may have had their expression patterns modulated by evolutionary changes in *trans*-factors without changes in the *cis*-sequences. Finally, the pattern of robust gene expression may originate in the common ancestor to these three strains and *S. paradoxus* and therefore patterns of molecular evolution may predate the comparisons we made. Even though the strong purifying selection against changes in the coding sequences observed in all three strains supports the last point, it does not exclude the possibility of *trans*-factors involved in robust expression. In fact, our analyses in Chapter 2 concluded that *trans*-factors are likely the driving forces of gene expression variability to temperature perturbations. Interestingly, for all three genomes, the 3' *cis*-sequences show greater relative molecular evolution for the robustly expressed genes than the more variable genes—albeit only two genomes (Table 3.3; 0.800 versus 0.663 in YPS183, Wilcoxon rank-sum test, *p*-value=0.02; 0.805 versus 0.657 in YPS2055, Wilcoxon rank-sum test, *p*-value=0.005) are significant. The average ratio of these 3' *cis*-rates compared to the synonymous rate is still under 0.8, which seems to suggest more relaxed purifying selection rather than directional selection. We are uncertain why genes with significantly less variation with respect to temperature fluctuations (and showing stronger purifying selection in the coding regions) would show more relaxed selection in the 3' *cis* regions. A highly speculative hypothesis might be a tradeoff between transcription regulation versus post-transcriptional regulation mediated through the 3'UTRs. For example, if the rate of

transcription was more tightly regulated there might be less selective pressure for 3'UTR-mediated RNA degradation control.

For genes with significantly robust expression solely in YPS2055, we also found significant differences in the coding sequence of all three strains, which is again consistent with the putative fitness importance of these genes in budding yeasts. Intriguingly, the evolutionary rates in 5' *cis*-elements were significantly smaller in the genomes of both natural strains (Table 3.4, 0.574 versus 0.686 in YPS2055, Wilcoxon rank-sum test, p-value=0.05; 0.575 versus 0.691 in YPS2073, Wilcoxon rank-sum test, p-value=0.0002), as well as the laboratory strain (albeit not significant). This is in contrast to the common robustly expressed genes shown in Table 3.3 where we found no significant difference. Given that these genes had significantly robust expression only in YPS2055, we cannot interpret the signature of greater purifying selection in these genes as related to temperature-dependent gene expression. A potential alternative was that these genes were conserved in all three strains due to selection of other traits rather than robustness. Unlike the results in Table 3.3 for the 3' *cis* sequence of the common robustly expressed genes, the genomic sequences of the genes in Table 3.4 did not show any difference in the relative rates of evolution of the 3' *cis* sequences.

For genes whose expression is significantly robust solely in YPS2073, we also found significant differences in the coding sequence of all three strains, in which the robust genes appeared to be under greater purifying selection as compared to the non-robustly expressed genes (Table 3.5). However, no significant differences were found in either of the 5' or 3' *cis*-element regions.

Because evolutionary rates were estimated using a closely related natural yeast species—*Saccharomyces paradoxus*—observation of signatures for greater purifying selection in the coding sequence of common robustly expressed genes (Table 3.3) suggested that these genes were under similar selective forces since the common ancestor of the three *S. cerevisiae* strains, which given the origin of the three strains is likely to be close to the divergence of the *S. cerevisiae and S. paradoxus*. As mentioned, these results are consistent with the idea that these genes are important for fitness and therefore experience stabilizing selection for their expression levels and purifying selection for their protein sequences. The results on the 5' *cis* and 3' *cis* sequences are more variable and hard to interpret, especially in terms of possible influence on temperature-dependent gene expression response. We hypothesize that any molecular evolution related to evolution of gene expression robustness in these strains might involve *trans* factors more than *cis* factors as we discussed in Chapter 2.

3.5 DISCUSSION

In this chapter, we examined the evolution of genetic and temperature variance and characterized the state of robustness of gene expression against temperature perturbation in two natural strains of yeast using a laboratory strain as a reference. We used generalized linear models to partition the variance into genetic and temperature components and found statistical association between the two components at a genome-wide scale. This association is consistent with the model of evolution of genetic robustness through congruent evolution. However, the magnitude of the association between genetic and temperature robustness was moderate. We hypothesized that

confounding factors involving gene-specific functions might have masked the strength of congruent evolution between genetic and temperature variance. Therefore, we analyzed the genetic and temperature variation association in four classes of genes and found significantly elevated association between the two sources of variances. Next, we applied an F-statistic to characterize genes with significantly reduced expression variation in the natural strains compared to the laboratory strain as a reference point. Surprisingly, we found that these genes are dominated by TATA-containing genes, but not transcription factors, essential genes, or "hub" genes (i.e. highly connected genes). We also found that these genes with less variation in natural strains are functionally enriched in metabolic processes.

The complex regulatory network of transcription evolves robustness to cope with stochastic, environmental, and genetic perturbation. However, given the complexity of the transcriptional network, it would be difficult to predict which genes would evolve robustness of gene expression. Previous studies have shown that genes with a TATA-box sequence in the promoter regions are related to stress responses and are highly regulated [20,22]. We reported that these TATA-containing genes have significantly higher genetic and temperature variances as compared to TATA-less genes. In our analyses, we found that these TATA-containing genes are also significantly overrepresented in genes with greater robustness of gene expression against temperature perturbation in natural strains as compared to a laboratory strain. This may suggest that both variation in gene expression and the relative reduction in variation are evolved programmed responses which are manifested more strongly in the TATA-containing genes. Surprisingly, we did

not find any enrichment of essential genes, transcription factors, or highly connected genes among these genes with robust expression. This suggests that genes with higher variation of expression are likely subject to the evolution of robustness regardless of their functional identity.

In general, the genome-wide correlation between genetic and temperature variance was consistent with previous findings, which showed congruence of genetic and environmental variability in yeast [6,13,23]. In our study, the small number of yeast strains might have limited this overall correlation. However, the fact that all four specific classes of genes displayed significantly higher correlation also suggests that specific functions of genes affect both types of robustness and the degree of congruence is modulated by the role of the genes within a system level regulation of gene expression. For example, the essential genes and the highly connected genes might both be involved in a highly complex central function in the gene regulatory network, and the higher correlation genetic and temperature robustness might be the result of changes in the connected regulatory input to the genes, either by mutational changes in the cis- and trans factors or by changes in the biochemical affinities as a function of temperature. The degree of congruent evolution of genetic and environmental robustness may not be consistent between different classes of genes because of their particular mode of gene regulation and subsequently the average of all the effects may have led to the weak correlation between two types of robustness over the whole transcriptome.

3.6 MATERIALS AND METHODS

Summary of the dataset

Genome-wide gene expression levels were measured in three haploid strains of *Saccharomyces cerevisiae*—two natural strains and the laboratory strain S288c—across five different temperatures (18-34°C with 4°C interval) at two checkpoints of the cell-division cycle (G1/S and M/G1). Briefly, cells were arrested at the G1 checkpoint using alpha-factor pheromone (Zymo Research, part number Y1001), released and collected at the G1/S and M/G1 checkpoints for each strain and temperature combination. Serial Analysis of Gene Expression (SAGE) was combined with next-generation sequencing (SOLiD 4, Applied Biosystems) to determine the number of reads registered to 5549 protein-coding genes in the yeast genome (not including hypothetical genes without evident proteins, so-called "dubious" genes). In total the data included 59 genome-wide gene expression profiles for 4344 genes. Here we only evaluated genes with detectable expression in all samples. Details on cell culture collection, cDNA sequencing preparation, and quantification procedures were described thoroughly in the Materials and Methods of Chapter 2.

Generalized Linear Model of gene expression

The dataset was analyzed as a two-way random model in which three strains and five temperatures were included. To simplify the possibility of complex interactions, the model was performed at each cell cycle checkpoint separately. All ANOVA analyses were carried out using PROC GLM in SAS (SAS Institute, Cary, NC, USA) [30,31].

For each gene, we computed the variance components using a generalized linear model (GLM). Many genes showed significant interaction between genetic and temperature factors, so we included the interaction term in the model of these genes only. Combined,

we estimated the genetic variance and temperature variance for 4344 protein-coding genes that had at least one unique read in all the samples.

The mean genetic and temperature variance was significantly larger than the mean error variation due to biological replicates. This indicated that the model was able to partition the variation due to the genetic and temperature factors with significant model effects. To show that the model was appropriate for our subsequent analyses, we randomized the gene expression among all samples within each gene and performed the same pipeline on the randomized data. The coefficient of determination (R-squared) in the randomized data was significantly smaller when compared to the experiment data.

Test of significance on the difference of two Spearman's rank correlation coefficients

To test the difference of the correlation in the foreground over the background, we utilized the one-leave-out jackknife estimate of the correlation coefficient to compare the foreground and background values, where F denotes the foreground and B denotes the background.

We applied Fisher's z-transformation to the Spearman's rank correlation coefficient ρ

$$z_F = \frac{1}{2}\ln\frac{(1+\rho_F)}{(1-\rho_F)} \qquad (3.1)$$

$$z_B = \frac{1}{2}\ln\frac{(1+\rho_B)}{(1-\rho_B)} \qquad (3.2)$$

In the foreground, we estimated the Fisher's z-transformation of the Spearman's rank correlation coefficient for one-left-out foreground datasets. For i=$\overline{1,n}$ with n being the

number of genes in the foreground, we computed ρ_{Fi} as the Spearman's correlation coefficient in the foreground when removing gene *i* from the foreground.

$$z_{Fi} = \frac{1}{2}\ln\frac{(1+\rho_{Fi})}{(1-\rho_{Fi})} \qquad (3.3)$$

$$y_{Fi} = nz_F - (n-1)z_{Fi} \qquad (3.4)$$

As a result, we have *n* estimates of y_{Fi} in the foreground

Applying a similar calculation in the background, we have *m* estimates of y_{Bj} with j=$\overline{1,m}$

$$z_{Bj} = \frac{1}{2}\ln\frac{(1+\rho_{Bj})}{(1-\rho_{Bj})} \qquad (3.3)$$

$$y_{Bj} = nz_B - (n-1)z_{Bj} \qquad (3.4)$$

A Welch t-test was used to test for a significant difference between the foreground y_{Fi} and the background y_{Bj}

Estimation of evolutionary rates in the coding sequence (CDS), the 5' cis-element and the 3' cis-element region.

As we described extensively in Chapter 2 (Materials and Methods), evolutionary rates in the 5' *cis*-element, coding sequence, and 3' *cis*-element were computed for each gene in all three strains. In brief, multiple alignment of the genome assemblies in these yeast strains was carried out using Lastz [32]. We used our custom Python scripts to calculate the synonymous (dS) and non-synonymous (dN) rates following the Li-Pamilo-Bianchi method [33,34], and the 5' and 3' *cis*-element rates (d5 and d3, respectively) following the Kimura 2-parameter model [35-37]. All evolutionary rates were estimated for each

strain against a closely related yeast species—*Saccharomyces paradoxus*—whose genome was retrieved from the Saccharomyces Genome Resequencing Project (SGRP) at Wellcome Trust Sanger Institute [38].

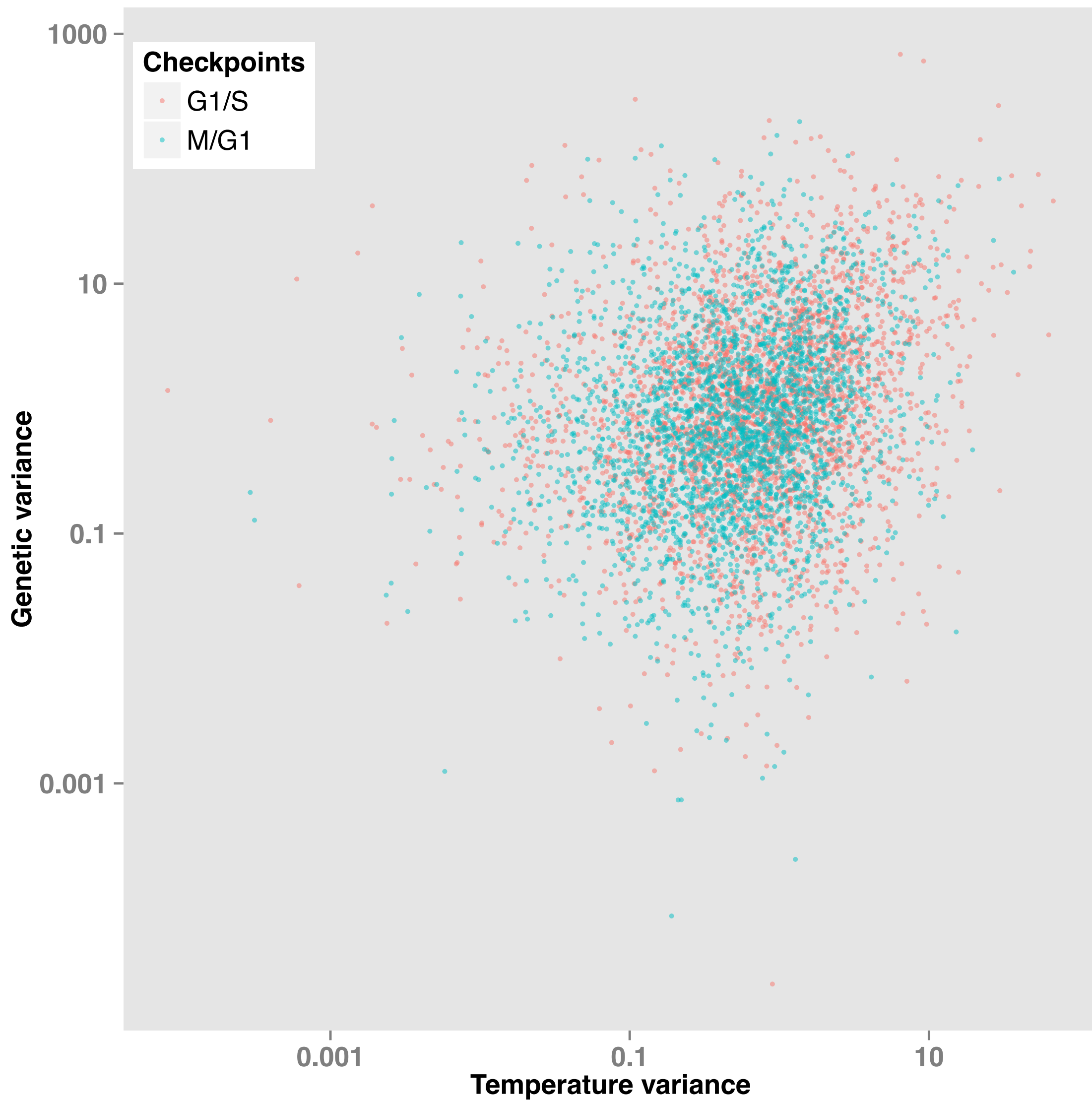

Figure 3.1 Association of genetic and temperature variances in the transcriptome of the budding yeast. Each point represents a gene with temperature and genetic variance (red-genes at G1/S checkpoint; blue-genes at M/G1 checkpoint). Spearman's rank correlation coefficient ρ=0.22, *p*-value<2.2E^{-16}, *n*=6090.

Table 3.1 Mean temperature and genetic variance in different classes of genes. Wilcoxon rank-sum test, (*): *p*-value<0.05; (**): *p*-value<0.01; (***): *p*-value<0.001.

		N	Temperature variance	Genetic variance	Spearman's correlation
Essentiality	Essential	1265	1.31±0.17	2.87±0.60***	0.257
	Non-essential	4825	1.44±0.09	4.44±0.52	0.213
TF	TFs	161	1.00±0.18	3.64±1.97	0.270
	Non-TFs	5929	1.42±0.08	4.13±0.44	0.220
TATA box	TATA-containing	1068	2.26±0.29***	8.13±2.03***	0.274*
	TATA-less	5022	1.23±0.07	3.26±0.29	0.197
Connected	Highly	555	1.64±0.31*	4.06±1.39	0.304
	Lowly	656	1.43±0.28	5.03±2.04	0.224

Table 3.2 Number of genes showing decreased and increased temperature variation in the natural strains with respect to the laboratory strain.

	Decreased temperature variation			**Increased temperature variation**		
	YPS2055	**YPS2073**	**common**	**YPS2055**	**YPS2073**	**common**
G1/S	70	106	54	11	9	**7**
M/G1	96	94	62	15	12	**10**
both stages	**43**	**54**	**38**	**8**	**6**	**5**

Table 3.3 Means and confidence intervals of evolutionary rates in the coding sequence (dN/dS), 5' *cis*-element (d5/dS), and 3' *cis*-element (d3/dS) regions in the genes showing robust expression against temperature perturbations in both natural strains at both checkpoints. Wilcoxon rank-sum test; (*): p-value<0.05; (**): p-value<0.01; (***): p-value<0.001.

		N	dN/dS	d5/dS	d3/dS
YPS183	Robust	26†	0.101±0.028*	0.720±0.142	0.800±0.187
	Non-robust	3604	0.147±0.004	0.692±0.023	0.663±0.017
YPS2055	Robust	25†	0.099±0.028*	0.671±0.106	0.805±0.154**
	Non-robust	3595	0.147±0.004	0.685±0.021	0.657±0.018
YPS2073	Robust	25†	0.097±0.028**	0.671±0.106	0.832±0.189
	Non-robust	3601	0.146±0.004	0.690±0.024	0.659±0.019

†: There were 38 genes robustly expressed in all conditions. However, we could not estimate the evolutionary rates in ~12-13 genes in each strain due to lack of genome assembly coverage or overlapping with neighbor genes.

Table 3.4 Means and confidence intervals of evolutionary rates in the coding sequence (dN/dS), 5' *cis*-element (d5/dS), and 3' *cis*-element (d3/dS) regions in the genes showing robust expression against temperature perturbations solely in YPS2055. Wilcoxon rank-sum test; (*): p-value<0.05; (**): p-value<0.01; (***): p-value<0.001.

		N	dN/dS	d5/dS	d3/dS
YPS183	Robust	64	0.111±0.021**	0.583±0.054	0.636±0.052
	Non-robust	3566	0.148±0.004	0.695±0.023	0.664±0.017
YPS2055	Robust	64	0.110±0.021**	0.574±0.054*	0.623±0.050
	Non-robust	3556	0.147±0.004	0.687±0.021	0.659±0.018
YPS2073	Robust	64	0.110±0.021***	0.575±0.054***	0.628±0.050
	Non-robust	3562	0.146±0.004	0.692±0.024	0.661±0.019

Table 3.5 Means and confidence intervals of evolutionary rates in the coding sequence (dN/dS), 5' *cis*-element (d5/dS), and 3' *cis*-element (d3/dS) regions in the genes showing robust expression against temperature perturbations solely in YPS2073. Wilcoxon rank-sum test; (*): *p*-value<0.05; (**): *p*-value<0.01; (***): *p*-value<0.001.

		N	dN/dS	d5/dS	d3/dS
YPS183	Robust	83	0.123±0.018*	0.678±0.105	0.679±0.075
	Non-robust	3547	0.148±0.004	0.693±0.023	0.664±0.017
YPS2055	Robust	83	0.123±0.018	0.696±0.156	0.688±0.110
	Non-robust	3537	0.147±0.004	0.685±0.021	0.658±0.018
YPS2073	Robust	83	0.121±0.018**	0.698±0.155	0.694±0.110
	Non-robust	3543	0.146±0.004	0.689±0.024	0.660±0.019

Table 3.6 List of genes expressing robustness in YPS2055 at the G1/S checkpoint.

Systematic name	Standard name	Gene description
YCR088W	ABP1	Actin-binding protein of the cortical actin cytoskeleton
YBR145W	ADH5	Alcohol dehydrogenase isoenzyme V
YCL025C	AGP1	Low-affinity amino acid permease with broad substrate range
YBR194W	AIM4	Protein proposed to be associated with the nuclear pore complex
YDL192W	ARF1	ADP-ribosylation factor
YDR035W	ARO3	3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase
YJR148W	BAT2	Cytosolic branched-chain amino acid (BCAA) aminotransferase
YKR068C	BET3	Subunit of the transport protein particle (TRAPP) complex
YJR025C	BNA1	3-hydroxyanthranilic acid dioxygenase
YHR122W	CIA2	Component of cytosolic iron-sulfur protein assembly (CIA) machinery
YJR109C	CPA2	Large subunit of carbamoyl phosphate synthetase
YLR216C	CPR6	Peptidyl-prolyl cis-trans isomerase (cyclophilin)
YLR286C	CTS1	Endochitinase
YER124C	DSE1	Daughter cell-specific protein
YNR067C	DSE4	Daughter cell-specific secreted protein with similarity to glucanases
YPR017C	DSS4	Guanine nucleotide dissociation stimulator for Sec4p
YHL016C	DUR3	Plasma membrane transporter for both urea and polyamines
YDR539W	FDC1	Putative phenylacrylic acid decarboxylase
YLR342W	FKS1	Catalytic subunit of 1,3-beta-D-glucan synthase
YBR047W	FMP23	Putative protein of unknown function
YDR519W	FPR2	Membrane-bound peptidyl-prolyl cis-trans isomerase (PPIase)
YMR250W	GAD1	Glutamate decarboxylase
YEL011W	GLC3	Glycogen branching enzyme, involved in glycogen accumulation
YKL152C	GPM1	Tetrameric phosphoglycerate mutase
YOL151W	GRE2	3-methylbutanal reductase and NADPH-dependent methylglyoxal reductase
YIL116W	HIS5	Histidinol-phosphate aminotransferase
YDR158W	HOM2	Aspartic beta semi-aldehyde dehydrogenase
YPR068C	HOS1	Class I histone deacetylase (HDAC) family member
YNL004W	HRB1	Poly(A+) RNA-binding protein
YLR096W	KIN2	Serine/threonine protein kinase involved in regulation of exocytosis
YDR037W	KRS1	Lysyl-tRNA synthetase
YNL239W	LAP3	Cysteine aminopeptidase with homocysteine-thiolactonase activity
YDR034C	LYS14	Transcriptional activator involved in regulating lysine biosynthesis
YLL061W	MMP1	High-affinity S-methylmethionine permease
YJL104W	PAM16	Subunit of the PAM complex and the TIM23 complex
YHR071W	PCL5	Cyclin
YBR296C	PHO89	Na+/Pi cotransporter, active in early growth phase
YER053C	PIC2	Mitochondrial phosphate carrier
YMR297W	PRC1	Vacuolar carboxypeptidase Y (proteinase C, CPY)
YER012W	PRE1	Beta 4 subunit of the 20S proteasome
YKR093W	PTR2	Integral membrane peptide transporter

YIL121W	QDR2	Plasma membrane transporter of the major facilitator superfamily
YJL217W	REE1	Cytoplasmic protein involved in the regulation of enolase (ENO1)
YIL119C	RPI1	Transcription factor, allelic differences between S288C and Sigma1278b
YDL020C	RPN4	Transcription factor that stimulates expression of proteasome genes
YDL097C	RPN6	Essential, non-ATPase regulatory subunit of the 26S proteasome lid
YER050C	RSM18	Mitochondrial ribosomal protein of the small subunit
YGL028C	SCW11	Cell wall protein with similarity to glucanases
YGL224C	SDT1	Pyrimidine nucleotidase
YEL065W	SIT1	Ferrioxamine B transporter
YMR095C	SNO1	Protein of unconfirmed function
YPL092W	SSU1	Plasma membrane sulfite pump involved in sulfite metabolism
YDR410C	STE14	Farnesyl cysteine-carboxyl methyltransferase
YJL052W	TDH1	Glyceraldehyde-3-phosphate dehydrogenase, isozyme 1
YLR178C	TFS1	Protein that interacts with and inhibits carboxypeptidase Y and Ira2p
YBR067C	TIP1	Major cell wall mannoprotein with possible lipase activity
YER175C	TMT1	Trans-aconitate methyltransferase
YER090W	TRP2	Anthranilate synthase
YJL154C	VPS35	Endosomal subunit of membrane-associated retromer complex
YDR200C	VPS64	Protein required for cytoplasm to vacuole targeting of proteins
YER072W	VTC1	Subunit of the vacuolar transporter chaperone (VTC) complex
YHR138C		Protein of unknown function
YKR018C		Protein of unknown function
YDL124W		NADPH-dependent alpha-keto amide reductase
YGL114W		Putative protein of unknown function
YGL117W		Putative protein of unknown function
YLR179C		Protein of unknown function with similarity to Tfs1p
YLR413W		Putative protein of unknown function
YML131W		Protein of unknown function
YPL257W		Putative protein of unknown function

Table 3.7 List of genes expressing robustness in YPS2055 at the M/G1 checkpoint.

Systematic name	Standard name	Gene description
YCR088W	ABP1	Actin-binding protein of the cortical actin cytoskeleton
YMR083W	ADH3	Mitochondrial alcohol dehydrogenase isozyme III
YBR145W	ADH5	Alcohol dehydrogenase isoenzyme V
YCL025C	AGP1	Low-affinity amino acid permease with broad substrate range
YBR194W	AIM4	Protein proposed to be associated with the nuclear pore complex
YER073W	ALD5	Mitochondrial aldehyde dehydrogenase
YNL065W	AQR1	Plasma membrane transporter of the major facilitator superfamily
YDL192W	ARF1	ADP-ribosylation factor
YDR035W	ARO3	3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase
YJR148W	BAT2	Cytosolic branched-chain amino acid (BCAA) aminotransferase
YJR025C	BNA1	3-hydroxyanthranilic acid dioxygenase
YML042W	CAT2	Carnitine acetyl-CoA transferase of both mitochondria and peroxisomes
YDL126C	CDC48	AAA ATPase
YGR108W	CLB1	B-type cyclin involved in cell cycle progression
YJL062W-A	COA3	Mitochondrial inner membrane protein that regulates COX1 translation
YIL111W	COX5B	Subunit Vb of cytochrome c oxidase
YJR109C	CPA2	Large subunit of carbamoyl phosphate synthetase
YLR216C	CPR6	Peptidyl-prolyl cis-trans isomerase (cyclophilin)
YPR017C	DSS4	Guanine nucleotide dissociation stimulator for Sec4p
YHL016C	DUR3	Plasma membrane transporter for both urea and polyamines
YLR342W	FKS1	Catalytic subunit of 1,3-beta-D-glucan synthase
YBR047W	FMP23	Putative protein of unknown function
YDR519W	FPR2	Membrane-bound peptidyl-prolyl cis-trans isomerase (PPIase)
YMR250W	GAD1	Glutamate decarboxylase
YLR343W	GAS2	1,3-beta-glucanosyltransferase
YKR058W	GLG1	Glycogenin glucosyltransferase
YKR067W	GPT2	Glycerol-3-phosphate/dihydroxyacetone phosphate sn-1 acyltransferase
YOL151W	GRE2	3-methylbutanal reductase and NADPH-dependent methylglyoxal reductase
YML048W	GSF2	ER membrane protein that may promote hexose transporter secretion
YDR174W	HMO1	Chromatin associated high mobility group (HMG) family member
YDR158W	HOM2	Aspartic beta semi-aldehyde dehydrogenase
YNL004W	HRB1	Poly(A+) RNA-binding protein
YBR072W	HSP26	Small heat shock protein (sHSP) with chaperone activity
YCR021C	HSP30	Negative regulator of the H(+)-ATPase Pma1p
YER065C	ICL1	Isocitrate lyase
YLR174W	IDP2	Cytosolic NADP-specific isocitrate dehydrogenase
YDR037W	KRS1	Lysyl-tRNA synthetase
YKL008C	LAC1	Ceramide synthase component
YNL239W	LAP3	Cysteine aminopeptidase with homocysteine-thiolactonase activity

YDR034C	LYS14	Transcriptional activator involved in regulating lysine biosynthesis
YNL297C	MON2	Peripheral membrane protein with a role in endocytosis and vacuole in
YKR052C	MRS4	Iron transporter of the mitochondrial carrier family
YDL085W	NDE2	Mitochondrial external NADH dehydrogenase
YGR043C	NQM1	Transaldolase of unknown function
YNR009W	NRM1	Transcriptional co-repressor of MBF-regulated gene expression
YOR222W	ODC2	Mitochondrial inner membrane transporter
YOR269W	PAC1	Involved in nuclear migration, part of the dynein/dynactin pathway
YHR071W	PCL5	Cyclin
YER053C	PIC2	Mitochondrial phosphate carrier
YKL163W	PIR3	O-glycosylated covalently-bound cell wall protein
YDR276C	PMP3	Small plasma membrane protein related to family of plant polypeptides
YER012W	PRE1	Beta 4 subunit of the 20S proteasome
YJL079C	PRY1	Sterol binding protein involved in the export of acetylated sterols
YKL039W	PTM1	Protein of unknown function
YIL121W	QDR2	Plasma membrane transporter of the major facilitator superfamily
YBR256C	RIB5	Riboflavin synthase
YIL119C	RPI1	Transcription factor, allelic differences between S288C and Sigma1278b
YPL193W	RSA1	Protein involved in the assembly of 60S ribosomal subunits
YKL117W	SBA1	Co-chaperone that binds and regulates Hsp90 family chaperones
YGL224C	SDT1	Pyrimidine nucleotidase
YOR021C	SFM1	SPOUT methyltransferase
YEL065W	SIT1	Ferrioxamine B transporter
YOL113W	SKM1	Member of the PAK family of serine/threonine protein kinases
YMR095C	SNO1	Protein of unconfirmed function
YMR096W	SNZ1	Protein involved in vitamin B6 biosynthesis
YMR107W	SPG4	Protein required for high temperature survival during stationary phase
YER150W	SPI1	GPI-anchored cell wall protein involved in weak acid resistance
YML034W	SRC1	Inner nuclear membrane protein
YPL092W	SSU1	Plasma membrane sulfite pump involved in sulfite metabolism
YLR375W	STP3	Zinc-finger protein of unknown function
YOL020W	TAT2	High affinity tryptophan and tyrosine permease
YJL052W	TDH1	Glyceraldehyde-3-phosphate dehydrogenase, isozyme 1
YLR178C	TFS1	Protein that interacts with and inhibits carboxypeptidase Y and Ira2p
YOR192C	THI72	Transporter of thiamine or related compound
YBR067C	TIP1	Major cell wall mannoprotein with possible lipase activity
YFL037W	TUB2	Beta-tubulin
YDL210W	UGA4	GABA (gamma-aminobutyrate) permease
YIL056W	VHR1	Transcriptional activator
YGR141W	VPS62	Vacuolar protein sorting (VPS) protein
YER072W	VTC1	Subunit of the vacuolar transporter chaperone (VTC) complex
YAR035W	YAT1	Outer mitochondrial carnitine acetyltransferase
YER024W	YAT2	Carnitine acetyltransferase

YHR138C	Protein of unknown function
YKR018C	Protein of unknown function
YDL124W	NADPH-dependent alpha-keto amide reductase
YGL114W	Putative protein of unknown function
YGL117W	Putative protein of unknown function
YJR012C	Essential protein of unknown function
YJR030C	Putative protein of unknown function
YLR173W	Putative protein of unknown function
YLR413W	Putative protein of unknown function
YML131W	Protein of unknown function
YMR144W	Putative protein of unknown function
YMR262W	Protein of unknown function
YPL257W	Putative protein of unknown function
YKL033W-A	Putative protein of unknown function

Table 3.8 List of genes expressing robustness in YPS2073 at the G1/S checkpoint.

Systematic name	Standard name	Gene description
YCR088W	ABP1	Actin-binding protein of the cortical actin cytoskeleton
YBR145W	ADH5	Alcohol dehydrogenase isoenzyme V
YPL202C	AFT2	Iron-regulated transcriptional activator
YCL025C	AGP1	Low-affinity amino acid permease with broad substrate range
YFL055W	AGP3	Low-affinity amino acid permease
YER073W	ALD5	Mitochondrial aldehyde dehydrogenase
YDL192W	ARF1	ADP-ribosylation factor
YOL058W	ARG1	Arginosuccinate synthetase
YGL157W	ARI1	NADPH-dependent aldehyde reductase
YDR035W	ARO3	3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase
YGL202W	ARO8	Aromatic aminotransferase I
YPR145W	ASN1	Asparagine synthetase
YPR185W	ATG13	Regulatory subunit of the Atg1p signaling complex
YJR001W	AVT1	Vacuolar transporter
YJR148W	BAT2	Cytosolic branched-chain amino acid (BCAA) aminotransferase
YJR025C	BNA1	3-hydroxyanthranilic acid dioxygenase
YEL063C	CAN1	Plasma membrane arginine permease
YKL007W	CAP1	Alpha subunit of the capping protein heterodimer (Cap1p and Cap2p)
YHR122W	CIA2	Component of cytosolic iron-sulfur protein assembly (CIA) machinery
YDR119W-A	COX26	Putative protein of unknown function
YIL111W	COX5B	Subunit Vb of cytochrome c oxidase
YJR109C	CPA2	Large subunit of carbamoyl phosphate synthetase
YLR216C	CPR6	Peptidyl-prolyl cis-trans isomerase (cyclophilin)
YLR286C	CTS1	Endochitinase
YER088C	DOT6	Protein involved in rRNA and ribosome biogenesis
YER124C	DSE1	Daughter cell-specific protein
YNR067C	DSE4	Daughter cell-specific secreted protein with similarity to glucanases
YHL016C	DUR3	Plasma membrane transporter for both urea and polyamines
YDR539W	FDC1	Putative phenylacrylic acid decarboxylase
YLR342W	FKS1	Catalytic subunit of 1,3-beta-D-glucan synthase
YBR047W	FMP23	Putative protein of unknown function
YDR519W	FPR2	Membrane-bound peptidyl-prolyl cis-trans isomerase (PPIase)
YMR250W	GAD1	Glutamate decarboxylase
YAL044C	GCV3	H subunit of the mitochondrial glycine decarboxylase complex
YGL057C	GEP7	Protein of unknown function
YOR164C	GET4	Protein involved in inserting tail-anchored proteins into ER membranes
YCL040W	GLK1	Glucokinase
YKL152C	GPM1	Tetrameric phosphoglycerate mutase
YOL151W	GRE2	3-methylbutanal reductase and NADPH-dependent methylglyoxal reductase
YGR154C	GTO1	Omega-class glutathione transferase

YIL116W	HIS5	Histidinol-phosphate aminotransferase
YDR174W	HMO1	Chromatin associated high mobility group (HMG) family member
YDR158W	HOM2	Aspartic beta semi-aldehyde dehydrogenase
YNL004W	HRB1	Poly(A+) RNA-binding protein
YER048W-A	ISD11	Cysteine desulfurase (Nfs1p) activator
YML068W	ITT1	Protein that modulates the efficiency of translation termination
YDR037W	KRS1	Lysyl-tRNA synthetase
YNL239W	LAP3	Cysteine aminopeptidase with homocysteine-thiolactonase activity
YDR034C	LYS14	Transcriptional activator involved in regulating lysine biosynthesis
YLL061W	MMP1	High-affinity S-methylmethionine permease
YMR164C	MSS11	Transcription factor
YDL107W	MSS2	Peripherally bound inner membrane protein of the mitochondrial matrix
YGL122C	NAB2	Nuclear polyadenylated RNA-binding protein
YOR222W	ODC2	Mitochondrial inner membrane transporter
YOR130C	ORT1	Ornithine transporter of the mitochondrial inner membrane
YJL104W	PAM16	Subunit of the PAM complex and the TIM23 complex
YHR071W	PCL5	Cyclin
YBR296C	PHO89	Na+/Pi cotransporter, active in early growth phase
YER012W	PRE1	Beta 4 subunit of the 20S proteasome
YOR323C	PRO2	Gamma-glutamyl phosphate reductase
YKR093W	PTR2	Integral membrane peptide transporter
YIL121W	QDR2	Plasma membrane transporter of the major facilitator superfamily
YDL103C	QRI1	UDP-N-acetylglucosamine pyrophosphorylase
YJL217W	REE1	Cytoplasmic protein involved in the regulation of enolase (ENO1)
YFR032C-A	RPL29	Ribosomal 60S subunit protein L29
YDL097C	RPN6	Essential, non-ATPase regulatory subunit of the 26S proteasome lid
YGR118W	RPS23A	Ribosomal protein 28 (rp28) of the small (40S) ribosomal subunit
YHR065C	RRP3	Protein involved in rRNA processing
YOR001W	RRP6	Nuclear exosome exonuclease component
YGL224C	SDT1	Pyrimidine nucleotidase
YDL168W	SFA1	Bifunctional alcohol dehydrogenase and formaldehyde dehydrogenase
YPL047W	SGF11	Integral subunit of SAGA histone acetyltransferase complex
YEL065W	SIT1	Ferrioxamine B transporter
YMR095C	SNO1	Protein of unconfirmed function
YMR096W	SNZ1	Protein involved in vitamin B6 biosynthesis
YDR006C	SOK1	Protein whose overexpression suppresses growth defects of PKA mutants
YAL005C	SSA1	ATPase involved in protein folding and NLS-directed nuclear transport
YLR452C	SST2	GTPase-activating protein for Gpa1p
YPL092W	SSU1	Plasma membrane sulfite pump involved in sulfite metabolism
YOR027W	STI1	Hsp90 cochaperone
YJR130C	STR2	Cystathionine gamma-synthase, converts cysteine into cystathionine
YJL052W	TDH1	Glyceraldehyde-3-phosphate dehydrogenase, isozyme 1
YBR067C	TIP1	Major cell wall mannoprotein with possible lipase activity

YOR010C	TIR2	Putative cell wall mannoprotein
YER175C	TMT1	Trans-aconitate methyltransferase
YMR271C	URA10	Minor orotate phosphoribosyltransferase (OPRTase) isozyme
YJL154C	VPS35	Endosomal subunit of membrane-associated retromer complex
YER072W	VTC1	Subunit of the vacuolar transporter chaperone (VTC) complex
YDR451C	YHP1	Homeobox transcriptional repressor
YOR172W	YRM1	Zinc-finger transcription factor involved in multidrug resistance
YHR138C		Protein of unknown function
YKR018C		Protein of unknown function
YDL124W		NADPH-dependent alpha-keto amide reductase
YDR262W		Putative protein of unknown function
YDR341C		Arginyl-tRNA synthetase
YGL114W		Putative protein of unknown function
YGL117W		Putative protein of unknown function
YJR030C		Putative protein of unknown function
YLL058W		Putative protein of unknown function with similarity to Str2p
YLR179C		Protein of unknown function with similarity to Tfs1p
YLR413W		Putative protein of unknown function
YML131W		Protein of unknown function
YMR262W		Protein of unknown function
YOL162W		Putative protein of unknown function
YPL257W		Putative protein of unknown function
YKL033W-A		Putative protein of unknown function

Table 3.9 List of genes expressing robustness in YPS2073 at the M/G1 checkpoint.

Systematic name	Standard name	Gene description
YCR088W	ABP1	Actin-binding protein of the cortical actin cytoskeleton
YBR145W	ADH5	Alcohol dehydrogenase isoenzyme V
YCL025C	AGP1	Low-affinity amino acid permease with broad substrate range
YER073W	ALD5	Mitochondrial aldehyde dehydrogenase
YOR175C	ALE1	Broad-specificity lysophospholipid acyltransferase
YDL192W	ARF1	ADP-ribosylation factor
YGL157W	ARI1	NADPH-dependent aldehyde reductase
YDR035W	ARO3	3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase
YJR148W	BAT2	Cytosolic branched-chain amino acid (BCAA) aminotransferase
YJR025C	BNA1	3-hydroxyanthranilic acid dioxygenase
YPL111W	CAR1	Arginase
YML042W	CAT2	Carnitine acetyl-CoA transferase of both mitochondria and peroxisomes
YHR122W	CIA2	Component of cytosolic iron-sulfur protein assembly (CIA) machinery
YIL111W	COX5B	Subunit Vb of cytochrome c oxidase
YJR109C	CPA2	Large subunit of carbamoyl phosphate synthetase
YLR216C	CPR6	Peptidyl-prolyl cis-trans isomerase (cyclophilin)
YPR017C	DSS4	Guanine nucleotide dissociation stimulator for Sec4p
YHL016C	DUR3	Plasma membrane transporter for both urea and polyamines
YLR299W	ECM38	Gamma-glutamyltranspeptidase
YKR076W	ECM4	Omega class glutathione transferase
YLR342W	FKS1	Catalytic subunit of 1,3-beta-D-glucan synthase
YBR047W	FMP23	Putative protein of unknown function
YDR519W	FPR2	Membrane-bound peptidyl-prolyl cis-trans isomerase (PPIase)
YMR250W	GAD1	Glutamate decarboxylase
YKR058W	GLG1	Glycogenin glucosyltransferase
YOL151W	GRE2	3-methylbutanal reductase and NADPH-dependent methylglyoxal reductase
YIL116W	HIS5	Histidinol-phosphate aminotransferase
YDR174W	HMO1	Chromatin associated high mobility group (HMG) family member
YDR158W	HOM2	Aspartic beta semi-aldehyde dehydrogenase
YNL004W	HRB1	Poly(A+) RNA-binding protein
YBR072W	HSP26	Small heat shock protein (sHSP) with chaperone activity
YCR021C	HSP30	Negative regulator of the H(+)-ATPase Pma1p
YER065C	ICL1	Isocitrate lyase
YLR174W	IDP2	Cytosolic NADP-specific isocitrate dehydrogenase
YDR037W	KRS1	Lysyl-tRNA synthetase
YNL239W	LAP3	Cysteine aminopeptidase with homocysteine-thiolactonase activity
YDR034C	LYS14	Transcriptional activator involved in regulating lysine biosynthesis
YOL064C	MET22	Bisphosphate-3'-nucleotidase
YGL087C	MMS2	Ubiquitin-conjugating enzyme variant
YDL079C	MRK1	Glycogen synthase kinase 3 (GSK-3) homolog

YML128C	MSC1	Protein of unknown function
YDL085W	NDE2	Mitochondrial external NADH dehydrogenase
YGR043C	NQM1	Transaldolase of unknown function
YOR130C	ORT1	Ornithine transporter of the mitochondrial inner membrane
YHR071W	PCL5	Cyclin
YNR070W	PDR18	Putative transporter of the ATP-binding cassette (ABC) family
YER053C	PIC2	Mitochondrial phosphate carrier
YER012W	PRE1	Beta 4 subunit of the 20S proteasome
YJL079C	PRY1	Sterol binding protein involved in the export of acetylated sterols
YGR170W	PSD2	Phosphatidylserine decarboxylase of the Golgi and vacuolar membranes
YMR137C	PSO2	Nuclease required for DNA single- and double-strand break repair
YKL039W	PTM1	Protein of unknown function
YKR093W	PTR2	Integral membrane peptide transporter
YPR191W	QCR2	Subunit 2 of ubiquinol cytochrome-c reductase (Complex III)
YIL121W	QDR2	Plasma membrane transporter of the major facilitator superfamily
YDL103C	QRI1	UDP-N-acetylglucosamine pyrophosphorylase
YBR256C	RIB5	Riboflavin synthase
YLR185W	RPL37A	Ribosomal 60S subunit protein L37A
YGR214W	RPS0A	Ribosomal 40S subunit protein S0A
YHR065C	RRP3	Protein involved in rRNA processing
YOL142W	RRP40	Exosome non-catalytic core component
YOR001W	RRP6	Nuclear exosome exonuclease component
YHR154W	RTT107	Protein implicated in Mms22-dependent DNA repair during S phase
YGL224C	SDT1	Pyrimidine nucleotidase
YDR078C	SHU2	Component of the Shu complex, which promotes error-free DNA repair
YMR175W	SIP18	Phospholipid-binding hydrophilin
YEL065W	SIT1	Ferrioxamine B transporter
YMR095C	SNO1	Protein of unconfirmed function
YMR096W	SNZ1	Protein involved in vitamin B6 biosynthesis
YDR006C	SOK1	Protein whose overexpression suppresses growth defects of PKA mutants
YMR107W	SPG4	Protein required for high temperature survival during stationary phase
YER150W	SPI1	GPI-anchored cell wall protein involved in weak acid resistance
YPL092W	SSU1	Plasma membrane sulfite pump involved in sulfite metabolism
YJL052W	TDH1	Glyceraldehyde-3-phosphate dehydrogenase, isozyme 1
YBR067C	TIP1	Major cell wall mannoprotein with possible lipase activity
YMR271C	URA10	Minor orotate phosphoribosyltransferase (OPRTase) isozyme
YIL056W	VHR1	Transcriptional activator
YGR065C	VHT1	High-affinity plasma membrane H+-biotin (vitamin H) symporter
YER072W	VTC1	Subunit of the vacuolar transporter chaperone (VTC) complex
YAR035W	YAT1	Outer mitochondrial carnitine acetyltransferase
YER024W	YAT2	Carnitine acetyltransferase
YLR120C	YPS1	Aspartic protease
YBR046C	ZTA1	NADPH-dependent quinone reductase

YHR138C	Protein of unknown function
YKR018C	Protein of unknown function
YDL124W	NADPH-dependent alpha-keto amide reductase
YGL114W	Putative protein of unknown function
YGL117W	Putative protein of unknown function
YLR413W	Putative protein of unknown function
YML131W	Protein of unknown function
YMR262W	Protein of unknown function
YOR385W	Putative protein of unknown function
YPL257W	Putative protein of unknown function
YKL033W-A	Putative protein of unknown function

Table 3.10 Gene Ontology (GO) terms overrepresented in genes with robust expression in natural strains at both checkpoints. Blue shading indicates GO terms involving biological processes.

GOID	GO terms	*p*-value	Genes annotated to the term
6520	cellular amino acid metabolic process	0.022790251	ADH5, ARO3, BAT2, BNA1, CPA2, GAD1, HOM2, KRS1, LAP3, LYS14, SNO1
6082	organic acid metabolic process	0.032738607	ADH5, ARO3, BAT2, BNA1, CPA2, GAD1, HOM2, KRS1, LAP3, LYS14, SNO1, VTC1
43436	oxoacid metabolic process	0.04770005	ADH5, ARO3, BAT2, BNA1, CPA2, GAD1, HOM2, KRS1, LAP3, LYS14, SNO1, VTC1

Table 3.11 Gene Ontology (GO) terms overrepresented in genes with robust expression in natural strains at the G1/S checkpoint. Orange shading indicates GO terms involving molecular functions.

GOID	GO terms	*p*-value	Genes annotated to the term
71944	cell periphery	0.026072525	ABP1, AGP1, CTS1, DSE4, DUR3, FDC1, FKS1, HOM2, MMP1, PHO89, PTR2, QDR2, SSU1, TDH1, TIP1, YDL124W, YLR413W

Table 3.12 Gene Ontology (GO) terms overrepresented in genes with robust expression in natural strains at the M/G1 checkpoint. Blue shading indicates GO terms involving biological processes; green shading indicates GO terms involving cellular components.

GOID	GO terms	*p*-value	Genes annotated to the term
44281	small molecule metabolic process	2.18E-03	ADH5, ALD5, ARO3, BAT2, BNA1, CAT2, CPA2, DUR3, GAD1, GRE2, HOM2, HSP30, ICL1, IDP2, KRS1, LAP3, LYS14, NDE2, NQM1, RIB5, SDT1, SNO1,SNZ1, VTC1, YAT1, YAT2, YDL124W
6082	organic acid metabolic process	2.96E-03	ADH5, ALD5, ARO3, BAT2, CAT2, CPA2, GAD1, HOM2, ICL1, IDP2, KRS1, LAP3, LYS14, SNO1, VTC1, YAT1, YAT2
43436	oxoacid metabolic process	3.55E-03	ADH5, ALD5, ARO3, BAT2, CAT2, CPA2, GAD1, HOM2, ICL1, IDP2, KRS1, LAP3, LYS14, SNO1, VTC1, YAT1, YAT2
6577	amino-acid betaine metabolic process	5.67E-03	CAT2, YAT1, YAT2
9437	carnitine metabolic process	5.67E-03	CAT2, YAT1, YAT2
19752	carboxylic acid metabolic process	5.87E-03	ADH5, ALD5, ARO3, BAT2, CAT2, CPA2, GAD1, HOM2, ICL1, IDP2, KRS1, LAP3, LYS14, SNO1, YAT1, YAT2
44710	single-organism metabolic process	6.71E-03	ALD5, ARO3, BAT2, BNA1, CAT2, COX5B, CPA2, DUR3, GAD1, GLG1, GRE2, HOM2, HSP30, ICL1, IDP2, KRS1, LAP3, LYS14, NDE2, NQM1, RIB5, SDT1, SNO1, SNZ1, TDH1, VTC1, YAT1, YAT2, YDL124W, YML131W
72524	pyridine-containing compound metabolic process	2.30E-02	ADH5, BNA1, NDE2, NQM1, SNO1, SNZ1
55114	oxidation-reduction process	3.68E-02	ADH5, ALD5, BNA1, COX5B, GLG1, GRE2, HOM2, ICL1, IDP2, NDE2, NQM1, RIB5, TDH1, YDL124W, YML131W
4092	carnitine O-acetyltransferase activity	0.006083609	CAT2, YAT1, YAT2
16406	carnitine O-acyltransferase activity	0.006083609	CAT2, YAT1, YAT2

REFERENCES

[1] Winzeler, E. A., Shoemaker, D. D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J. D., Bussey, H., et al. (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* ***285***, 901–906.

[2] Mnaimneh, S., Davierwala, A. P., Haynes, J., Moffat, J., Peng, W.-T., Zhang, W., Yang, X., Pootoolal, J., Chua, G., Lopez, A., et al. (2004) Exploration of essential gene functions via titratable promoter alleles. *Cell* ***118***, 31–44.

[3] Wagner, A. (2007) Robustness and evolvability in living systems.

[4] Ancel, L. W. and Fontana, W. (2000) Plasticity, evolvability, and modularity in RNA. *J. Exp. Zool.* ***288***, 242–283.

[5] Visser, J. A. G. M., Hermisson, J., Wagner, G. P., Meyers, L. A., Bagheri-Chaichian, H., Blanchard, J. L., Chao, L., Cheverud, J. M., Elena, S. F., Fontana, W., et al. (2003) Perspective: evolution and detection of genetic robustness. *Evolution* ***57***, 1959–1972.

[6] Proulx, S. R., Nuzhdin, S. and Promislow, D. E. L. (2007) Direct selection on genetic robustness revealed in the yeast transcriptome. *PLoS ONE* ***2***, e911.

[7] Lenski, R. E. and Travisano, M. (1994) Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proceedings of the National Academy of Sciences* ***91***, 6808–6814.

[8] Montville, R., Froissart, R., Remold, S. K., Tenaillon, O. and Turner, P. E. (2005) Evolution of mutational robustness in an RNA virus. *PLoS Biol.* ***3***, e381.

[9] Esberg, A., Muller, L. A. H. and McCusker, J. H. (2011) Genomic structure of and genome-wide recombination in the *Saccharomyces cerevisiae* S288C progenitor isolate EM93. *PLoS ONE* ***6***, e25211.

[10] Mortimer, R. K. and Johnston, J. R. (1986) Genealogy of principal strains of the yeast genetic stock center. *Genetics* ***113***, 35–43.

[11] Sniegowski, P. D., Dombrowski, P. G. and Fingerman, E. (2002) *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* coexist in a natural woodland site in North America and display different levels of reproductive isolation from European conspecifics. *FEMS Yeast Research* ***1***, 299–306.

[12] Simola, D. F., Francis, C., Sniegowski, P. D. and Kim, J. (2010) Heterochronic evolution reveals modular timing changes in budding yeast transcriptomes. *Genome Biol.* ***11***, R105.

[13] Ben Lehner. (2010) Genes confer similar robustness to environmental, stochastic, and genetic perturbations in yeast. *PLoS ONE* ***5***, e9035.

[14] Ben Lehner. (2010) Conflict between noise and plasticity in yeast. *PLoS Genet.* ***6***, e1001185.

[15] Seringhaus, M. (2006) Predicting essential genes in fungal genomes. *Genome Res.* ***16***, 1126–1135.

[16] Bähler, J. (2005) Cell-cycle control of gene expression in budding and fission yeast. *Annu. Rev. Genet.* ***39***, 69–94.

[17] Hahn, S. and Young, E. T. (2011) Transcriptional regulation in *Saccharomyces cerevisiae*: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics* ***189***, 705–736.

[18] Rhee, H. S. and Pugh, B. F. (2012) Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* ***483***, 295–301.

[19] Matangkasombut, O., Auty, R. and Buratowski, S. (2004) Structure and function of the TFIID complex. *Proteins in Eukaryotic Transcription*, pp 67–92.

[20] Huisinga, K. L. and Pugh, B. F. (2004) A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae. *Molecular Cell* ***13***, 573–585.

[21] Young, R. A., Lee, T. I., Causton, H. C., Holstege, F. C. P., Shen, W.-C., Hannett, N., Jennings, E. G., Winston, F. and Green, M. R. (2000) Redundant roles for the TFIID and SAGA complexes in global transcription. *Nature* ***405***, 701–704.

[22] Basehoar, A. D., Zanton, S. J. and Pugh, B. F. (2004) Identification and distinct regulation of yeast TATA box-containing genes. *Cell* ***116***, 699–709.

[23] Landry, C. R., Lemos, B., Rifkin, S. A., Dickinson, W. J. and Hartl, D. L. (2007) Genetic properties influencing the evolvability of gene expression. *Science* ***317***, 118–121.

[24] Jeong, H., Mason, S. P., Barabási, A. L. and Oltvai, Z. N. (2001) Lethality and centrality in protein networks. *Nature* ***411***, 41–42.

[25] Levy, S. F. and Siegal, M. L. (2008) Network hubs buffer environmental variation in *Saccharomyces cerevisiae*. *PLoS Biol.* ***6***, e264.

[26] Stark, C. (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Research* ***34***, D535–D539.

[27] Lee, D. C. and Chapman, D. (1987) The effects of temperature on biological membranes and their models. *Symp Soc Exp Biol* ***41***, 35–52.

[28] Hazel, J. R. (1995) Thermal adaptation in biological membranes: is homeoviscous adaptation the explanation? *Annual Review of Physiology* ***57***, 19–42.

[29] Quinn, P. J. (1988) Effects of temperature on cell membranes. *Symp Soc Exp Biol* ***42***, 237–258.

[30] Sokal, R. R. and Rohlf, F. J. (1995) Biometry.

[31] Saxton, A. M.SAS Institute. (2004) Genetic analysis of complex traits using SAS.

[32] Harris, R. S. Improved pairwise alignment of genomic DNA. *Ph.D. Thesis, The Pennsylvania State University*.

[33] Pamilo, P. and Bianchi, N. O. (1993) Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. *Mol. Biol. Evol.* ***10***, 271–281.

[34] Li, W.-H. (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J Mol Evol* ***36***, 96–99.

[35] Kimura, M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* ***16***, 111–120.

[36] Kohn, M. H. (2003) Inference of positive and negative selection on the 5' regulatory regions of *Drosophila* genes. *Mol. Biol. Evol.* ***21***, 374–383.

[37] Osada, N., Hirata, M., Tanuma, R., Kusuda, J., Hida, M., Suzuki, Y., Sugano, S., Gojobori, T., Shen, C. K. J., Wu, C.-I., et al. (2005) Substitution rate and structural divergence of 5′UTR evolution: comparative analysis between human and cynomolgus monkey cDNAs. *Mol. Biol. Evol.* ***22***, 1976–1982.

[38] Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. A., Davey, R. P., Roberts, I. N., Burt, A., Koufopanou, V., et al. (2009) Population genomics of domestic and wild yeasts. *Nature* ***458***, 337–341.

CHAPTER 4

CONCLUSIONS

Transcriptional regulation relies on inputs from outside environments to properly direct developmental programs in unicellular eukaryotes. The relative timing of the cell cycle events and gene expression of key factors has been a subject of substantial consideration [1-5]. A recent study has shown that natural yeast strains display significant variability in the length of cell cycle and in the dynamics of gene expression throughout the cell cycle [6]. If a perturbation results in homogeneously expression responses in these natural strains, it may lead to similar changes in the relative timing of the cell cycle events.

This dissertation aimed to investigate the effects of temperature on the dynamics of cell cycle-dependent gene expression in the budding yeasts to address two fundamental aspects: variability and robustness of gene expression. It revealed that evolution of cell cycle-dependent expression variability to temperature is specific in budding yeast strains, not only in gene identity but also in expression pattern. However, a simple multivariate response involving a majority of the yeast transcriptome is found to consistent with temperature gradient, suggesting potential of global regulation of the yeast transcriptome under temperature perturbations.

First, we dissected the variability of cell cycle-dependent gene expression in the face of temperature perturbation. Perturbations tend to alter the immediate pattern of gene expression in a cell and dependent on the degree of the perturbation may cause long-term

effects as well. While transient expression responses to perturbations serve as immediate protective mechanisms, long-term expression responses represent adaptive mechanisms to allow cell growth under these perturbations. Using a generalized linear model to take into account genetic differences from three strains of *Saccharomyces cerevisiae*, we demonstrated that steady-state expression responses (that is, gene expression responses at two cell cycle transitions) are distinct from the transient expression responses that mainly serve as immediate protective mechanisms. We also showed that different subsets of genes in each strain responds differently to temperature perturbation. When we examined the potential mechanisms leading to the observed divergence of expression responses among the three strains, we found no evidence for evolution of *cis*-regulatory elements that may correlate with this divergence. We presented that evolution of *trans*-factors is likely to involve in the regulation of gene expression against temperature perturbation. However, this divergence of expression is insufficient to explain the consistent progression at the G1/S checkpoint under temperature perturbation. Therefore, we suspected that coordinated regulation of the whole-transcriptome would be likely involved in this function. When we analyzed the global structure of yeast transcriptional space, we found a coordinated whole-transcriptome response to temperature that involves majority of the genes. This suggests that the global regulation of gene expression is important to cell cycle progression under growth-permissive temperatures.

One arguably fundamental question in biological systems is the role of global transcriptional machinery to the regulation of gene expression. Conventional studies of gene expression in budding yeasts focused on transcriptional responses that are

specifically regulated across different environment conditions [7,8]. However, it is difficult to uncover how much the global factors of cellular states have contributed to the transcriptional responses across these conditions. Several attempts analyzing synthetically constructed promoters in bacteria to decouple the global transcriptional regulation from specific regulation programs have shown that the global transcriptional machinery dominates the transcriptional responses during the cell growth [9,10]. A recent study of the promoter activity of 859 genes in the budding yeast has shown that the global factor is a major determinant of genome-wide gene expression profiles across conditions [11]. All these recent findings are consistent with our report of the global coordinated responses of gene expression to temperature perturbation. Further understanding of potential mechanisms controlling the global transcriptional regulation could manifest the coupling architecture of global and specific gene expression programs across different conditions.

Much efforts has been expended on studying the effects of *cis*-elements and *trans*-factors to regulation of gene expression. Recent studies using comparative analysis of gene expression between parental strains and their F1 hybrids have revealed that *trans*-acting variants has played at least partial roles in driving divergence of gene expression [12-14]. These findings were consistent with our results that *trans*-factor are likely to involve in the divergence of gene expression among three yeast strains.

Next, we examined that the variation of gene expression against two types of perturbations: genetic differences and temperature changes. We observed a coupling between two types of variation of gene expression in a genome-wide scale. We

speculated that this coupling is likely due to pleiotropic effects of genes and therefore genes with higher likelihood of pleiotropy may have higher degree of coupling between two types of variation. When we dissected the coupling of variation of gene expression in four functional classes of genes, we found a significantly higher degree of coupling in these four classes. Our results on coupling of variation of gene expression suggests a potential for congruent evolution of gene expression variability. We then looked for evidence of natural selection for reduced expression variability in the natural strains of yeast using the laboratory yeast strain as a reference. We found thirty eight genes in the natural strains that display significantly reduced expression variability to temperature perturbation. Therefore, we concluded that evolutionary selection for robustness of gene expression to perturbations occurs in natural populations.

Fine tuning of the developmental progress with the changes in the environment is critical for unicellular eukaryotes like yeasts that can not avoid adverse environmental conditions. Environmental perturbations alter activity of multiple kinases, which modulate the downstream sets of transcription factors that orchestrate the expression of target genes. In addition to the specific tuning of gene expression to particular environments, a recent report has shown the importance of global effects on the expression of many genes across conditions [11]. This global effect may achieve via a simple strategy like a combinatorial scheme of several transcription factors that individually regulate the stoichiometric expression of different groups of genes [15]. This strategy may help eliminate the extensive promoter tuning, which is possibly required to coordinate the fine tuning of the developmental with the changes in the environment. Further understanding of this

scheme of regulation of gene expression would essentially enhance our understanding about the mechanisms of global effects on gene expression.

REFERENCES

[1] Nurse, P. and Bissett, Y. (1981) Gene required in G1 for commitment to cell cycle and in G2 for control of mitosis in fission yeast. *Nature* ***292***, 558–560.

[2] Futcher, B., Zhu, G., Spellman, P. T., Volpe, T., Brown, P. O., Botstein, D. and Davis, T. N. (2000) Two yeast forkhead genes regulate the cell cycle and pseudohyphal growth. *Nature* ***406***, 90–94.

[3] Breeden, L. L. (2003) Periodic Transcription: A Cycle within a Cycle. *Current Biology* ***13***, R31–R38.

[4] Bähler, J. (2005) Cell-cycle control of gene expression in budding and fission yeast. *Annu. Rev. Genet.* ***39***, 69–94.

[5] Bertoli, C., Skotheim, J. M. and de Bruin, R. A. M. (2013) Control of cell cycle transcription during G1 and S phases. *Nat Rev Mol Cell Biol* ***14***, 518–528.

[6] Simola, D. F., Francis, C., Sniegowski, P. D. and Kim, J. (2010) Heterochronic evolution reveals modular timing changes in budding yeast transcriptomes. *Genome Biol.* ***11***, R105.

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

[8] Causton, H. C., Ren, B., Koh, S. S., Harbison, C. T., Kanin, E., Jennings, E. G., Lee, T. I., True, H. L., Lander, E. S. and Young, R. A. (2001) Remodeling of yeast genome expression in response to environmental changes. *Molecular Biology of the Cell* ***12***, 323–337.

[9] Gerosa, L., Kochanowski, K., Heinemann, M. and Sauer, U. (2013) Dissecting specific and global transcriptional regulation of bacterial gene expression. *Mol Syst Biol* ***9***.

[10] Berthoumieux, S., de Jong, H., Baptist, G., Pinel, C., Ranquet, C., Ropers, D. and Geiselmann, J. (2013) Shared control of gene expression in bacteria by transcription factors and global physiology of the cell. *Mol Syst Biol* ***9***.

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

[12] McManus, C. J., Coolon, J. D., Duff, M. O., Eipper-Mains, J., Graveley, B. R. and Wittkopp, P. J. Regulatory divergence in Drosophila revealed by mRNA-seq. *Genome Res.* ***10***, 816–825.

[13] Goncalves, A., Leigh-Brown, S., Thybert, D., Stefflova, K., Turro, E., Flicek, P., Brazma, A., Odom, D. T. and Marioni, J. C. (2012) Extensive compensatory cis-trans regulation in the evolution of mouse gene expression. *Genome Res.* ***22***, 2376–2384.

[14] Schaefke, B., Emerson, J. J., Wang, T. Y., Lu, M. Y. J., Hsieh, L. C. and LI, W. H. (2013) Inheritance of gene expression level and selective constraints on *trans*- and *cis*-regulatory changes in yeast. *Mol. Biol. Evol.* ***30***, 2121–2133.

[15] Stewart-Ornstein, J., Nelson, C., DeRisi, J., Weissman, J. S. and El-Samad, H. (2013) Msn2 coordinates a stoichiometric gene expression program. *Current Biology* ***23***, 1–10.