Genetic and Molecular Dissection of the Integration of Galactose and Glucose Signaling in Saccharomyces Cerevisiae Strains

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th>Citation</th>
<th>Escalante Chong, Renan Antonio. 2015. Genetic and Molecular Dissection of the Integration of Galactose and Glucose Signaling in Saccharomyces Cerevisiae Strains. Doctoral dissertation, Harvard University, Graduate School of Arts &amp; Sciences.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:17465312">http://nrs.harvard.edu/urn-3:HUL.InstRepos:17465312</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Genetic and molecular dissection of the integration of galactose and glucose signaling in *Saccharomyces cerevisiae* strains

A dissertation presented

by

Renan Antonio Escalante Chong

to

The Committee on Higher Degrees in Systems Biology

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in the subject of

Systems Biology

Harvard University
Cambridge, Massachusetts

May 2015
Genetic and molecular dissection of the integration of galactose and glucose signaling in *Saccharomyces cerevisiae* strains

Abstract

Cells need to sense the environment in order to survive, in particular they need to detect nutrients which will provide different building blocks and energy for the cell. This task is complicated by the fact that there can be multiple sources for the same type of nutrient available for the cell. A classical example of how cells sense multiple signals is given by carbon catabolite repression in the budding yeast *S. cerevisiae*. In this model the preferred carbon source glucose represses the genes used to metabolize an alternative source such as galactose. This means that the preferred carbohydrate glucose is thought to inhibit the induction of galactose genes when above a threshold concentration. Instead, we show that galactose metabolic genes (GAL) induction depends on the ratio of galactose and glucose. Surprisingly, we find that a critical portion of information processing occurs upstream of the canonical components of the GAL pathway. We then explore how cells choose between different responses to the environment. Specifically, we set out to characterize the variability in the response to combinations of galactose and glucose between several natural yeast isolates.

To elucidate the genetic basis of this phenotypic variation we use QTL mapping on these strains. Our study reveals that a signal transducer *GAL3* plays a central role in establishing variation in GAL gene induction. Lastly, we focus on the control of transcription in the cell. Many promoters in the cell produce both a coding transcript and a divergent transcript. To identify mutants that affect transcriptional directionality we use a bidirectional fluorescent protein reporter in the yeast nonessential gene deletion
collection. We determine that chromatin assembly can regulate divergent transcription. Moreover, mutations in the chromatin assembly factor CAF-I can lead to genome wide derepression of nascent divergent transcription.
I first want to start by thanking my advisor Michael Springer for pushing my limits. He offered many hours of his time to help me understand concepts and ideas, he never stopped asking me questions to guide me throughout this journey.

I want to thank as well my collaborators in the different projects I worked on. I learned so much from the interaction with these brilliant and creative minds. Among them I want to specially thank Yonatan Savir, Sebastian Marquardt, Kayla Lee, John Ingraham, Anna Green, Jue Wang, Steve Buratowski and Sean M. Carroll. I also want to express my deepest appreciation for my lab mates who were fundamental in helping me shape my thinking as a scientist: Sarah Boswell, Christine DeGenaro, Bo Hua, Ilan Wapinski and Jacky Jui-yu Chou. You all taught me a lot in my time at the lab and helped me grow through the years. I specially want to thank Meriem El Karoui and Raúl Fernández López, who were always glad to help and provide guidance during the first year of my PhD, when I was the only graduate student in the lab.

My time at Harvard was greatly facilitated by Fundación Harvard Mexico, who first sponsored me during an internship in Pamela Silver's lab, and then afterwards when I became a graduate student. I want to specially thank Julie Gallardo and Felipe Ortiz Monasterio who have helped me and others materialize our dreams through this foundation. They both kept an eye on me through the years and I am very grateful for all their help. In addition, I want to thank CONACYT for providing funding and assistance during my graduate career.

All of my work could not have been done without the help of my chosen family; friends who stood by me through good and bad times: Consuelo Dayzú Quinto, Patricia Sánchez Martinez, Daniela Robles, Mariana Gómez, Ben Kertman, Rosa Fernández, François Legoux, Aude Étrillard, Joao Duarte, Lilia Rodríguez, Jordan Jarvis, Alex Dorsk, Natalie Gyenes and Alexis de Boisset. Many thanks to my fellow dragon boaters who taught me the discipline to be strong and persist through hardship while working together for a specific goal in mind. I also want to thank Aidan Kostenko for all his love and support.

Last but not least, I want to thank four special people, my greatest inspiration of them all, my family: my mother Gloria Chong Yong and my sisters Lizbeth, Fabiola and Liliana. You had always been there for me and I cannot put in words how much your love, support and care have meant to me throughout my whole life.
## Table of contents

Chapter 1: Introduction .................................................................................................................. 1

Chapter 2: Galactose metabolic genes in yeast respond to a ratio of galactose and glucose
............................................................................................................................................. 5

2.1 Introduction .......................................................................................................................... 6

2.2 Results .................................................................................................................................. 7

2.2.1 GAL Metabolic Genes Respond to the Ratio of Glucose and Galactose ........ 7

2.2.2 Ratio Sensing Is Generated Upstream of the Canonical Gal Pathway .......... 10

2.2.3 The Galactose Transporter Gal2p Is Not Required for Ratio Sensing ........ 13

2.2.4 Galactose Uptake Depends on the Ratio of Extracellular Sugars
Concentrations ......................................................................................................................... 15

2.2.5 Ratio Sensing Can Provide a Fitness Advantage ...................................................... 16

2.3 Discussion ............................................................................................................................ 19

2.4 Methods ............................................................................................................................... 22

2.4.1 Growth Conditions and Media ..................................................................................... 22

2.4.2 Flow Cytometry .............................................................................................................. 22

2.4.3 Microfluidics and Live Cell Imaging ............................................................................ 23

2.4.4 Strain Construction ......................................................................................................... 23

2.4.5 Data Analysis ................................................................................................................. 23

2.4.6 GC-Mass Spectrometry ................................................................................................. 24

2.5 References ........................................................................................................................... 24

Chapter 3: Genetic dissection of the natural variation of galactose and glucose signaling in *S. cerevisiae* .................................................................................................................. 29

3.1 Introduction .......................................................................................................................... 29

3.2 Results .................................................................................................................................. 32

3.2.1 Measuring the variation in responses of the GAL pathway between strains 32

3.2.2 The set point of induction is a continuous, quantitative trait between natural yeast strains ................................................................................................................................. 35

3.2.3 Population and source of strains loosely match the variation in the set point of induction ................................................................................................................................. 35

3.2.4 Bulk segregant analysis reveals a single locus underlying the differences in set point between BC187 and YJM978 .................................................................................... 36
3.2.5 Dissection of the candidate locus by reciprocal hemizygous hybrids suggests that SOK1 is not implicated in the trait difference ................................................................. 40
3.2.6 The BC187 and YJM978 GAL3 alleles contribute a large fraction of the difference in set point between strains .................................................................................. 41
3.2.7 GAL3 alleles from various natural isolates fall in at least two categories .... 43
3.2.8 Both GAL3 and genetic background contribute to the set point across a panel of natural isolates ................................................................. 44
3.3 Discussion ............................................................................................................. 45
3.4 Methods ............................................................................................................... 47
3.4.1 Strains and media ............................................................................................ 47
3.4.2 Flow cytometry and sugar titrations ................................................................ 48
3.4.3 Meiotic segregants phenotyping ..................................................................... 49
3.4.4 DNA isolation and genomic library preparation .............................................. 49
3.4.5 Bioinformatic analysis ...................................................................................... 50
3.5 References .......................................................................................................... 51

Chapter 4: A Chromatin-Based Mechanism for Limiting Divergent Noncoding Transcription ..................................................................................................................... 55
4.1 Introduction .......................................................................................................... 56
4.2 Results .................................................................................................................. 58
4.2.1 Divergent Transcription Is Independent of Transcript Identity .................... 58
4.2.2 CAF-I Affects Divergent Promoter Directionality ........................................ 61
4.2.3 CAF-I Represses Divergent Noncoding Transcription Genome Wide ....... 66
4.2.4 CAF-I Repression of Divergent Transcription Is Distinct from Early Termination and RNA Degradation Pathways ................................................. 71
4.3 Discussion .......................................................................................................... 80
4.4 Experimental Procedures .................................................................................... 84
4.4.1 Yeast Strains .................................................................................................. 84
4.4.2 Northern Blotting ........................................................................................... 84
4.4.3 Chromatin Immunoprecipitation ................................................................... 84
4.4.4 Reverse Genetic Screen Using FPRs .............................................................. 85
4.4.5 Ranking of FPR Screen Candidate Deletion Mutations ............................... 85
4.4.6 NET-Seq Experiments and Analysis ............................................................... 86
4.4.7 Segmentation of H3K56ac and Histone Exchange Data ............................... 86
4.5 References .......................................................................................................... 86

Chapter 5: Conclusions ............................................................................................ 92
Appendix A. Supporting Information for Chapter 2 .................................................. 95
A.1. Strain construction ........................................................................................................... 95
A.2. Yeast strains and plasmids used in this study ................................................................. 96
A.3. Growth conditions and media ......................................................................................... 97
A.4. The effect of inoculum size ............................................................................................ 98
A.5. Metric for cell response ................................................................................................. 100
A.6. Growth and imaging conditions in a microfluidics chamber ...................................... 101
A.7. Mig1p localization .......................................................................................................... 102
A.8. Ratio sensing as a result of competitive binding ............................................................ 102
A.8.1. The linear case ........................................................................................................... 104
A.8.2. The Michaelian case ................................................................................................ 105
A.8.3. Ratio sensing input .................................................................................................. 106
A.8.4. Sharp response function ............................................................................................ 107
A.9. Fitness experiments ......................................................................................................... 107
A.10. The dependence of U-13C glucose and 12C galactose on external carbon concentrations ................................................................. 108
A.11. Ratio as a result of inverse enzyme catalysis ................................................................. 110
A.12. Considerations to study the potential advantage of ratio sensing .............................. 112
A.13. References ................................................................................................................ 113
A.14. Supplementary figures ............................................................................................... 114
Appendix B. Supporting information for Chapter 3 ................................................................. 124
B.1. Strains and plasmids used in this study ........................................................................... 124
B.2. Supplementary figures ................................................................................................... 130
B.3. References .................................................................................................................... 135
Appendix C. Supporting information for Chapter 4 ............................................................... 136
C.1. Extended Experimental Procedures .............................................................................. 136
C.1.1. Construction of yeast strains .................................................................................... 136
C.1.2. Yeast Media ............................................................................................................. 136
C.1.3. Plasmid construction ............................................................................................... 137
C.1.4. NET-seq analysis ..................................................................................................... 138
C.1.5. Visualization ............................................................................................................. 139
C.1.6. Post-processing ........................................................................................................ 139
C.1.7. Calculating NETseq coverage for CUTs and SUTs ..................................................... 140
C.1.8. Definition of divergent ORF pairs sharing a 5’NFR ..................................................... 140
C.1.9. Quantitative PCR (qPCR) ........................................................................................ 140
C.1.10. α-factor arrest ................................................................. 140
C.1.11. Flow cytometry ................................................................. 141
C.1.12. Targeted mutant versus wild-type FPR comparisons using FCM.... 141
C.1.13. Determining p Values with Multiple Hypotheses Testing Correction for Metagene Analysis in Figure 4-7B ......................................................... 142
C.2. Supplementary figures .................................................................. 156
C.3. Supplementary References ............................................................ 169
“The time you enjoy wasting is not wasted time.” - Bertrand Russell
Chapter 1: Introduction

The yeast *Saccharomyces cerevisiae* is a powerful model organism whose study has revealed a wealth of biological knowledge. Yeast is amenable to high throughput assays that can characterize hundreds of phenotypes from either naturally occurring genetic variants or human engineered mutations. This provides the opportunity of connecting phenotypes to specific genetic changes. In this thesis we exploit these properties to explore two intriguing aspects of biology. Chapters 2 and 3 explore how cells sense and integrate multiple signals in the environment and trigger an appropriate response. Chapter 4 examines how the directionality of transcription of genes in cells is regulated.

Cells need to sense the environment in order to survive. Signaling pathways in cells transfer information from the environment and trigger an appropriate response that often involves the expression of a subset of specific genes. A critical piece of information for survival is nutrient availability. Since nutrients can come from different sources cells have to detect a myriad of different nutrient signals. How cells monitor multiple signals and combine them to coordinate expression is not yet fully understood. This problem has been studied in a bacterial system where the multiplication of individual signals explained the combinatorial response. In this thesis, quantitative experimental and computational methods are combined to shed light on this process in one of the best studied eukaryotic signaling pathways. We use as our model nutrient signaling pathway the response of *S. cerevisiae* to the carbohydrate galactose.

Chapter 2 investigates how cells respond to the presence of two nutrients with opposing
effects on a nutrient signaling pathway. Galactose triggers the expression of the galactose metabolic genes (GAL genes) whereas glucose represses their expression. Traditionally glucose, the preferred carbon source, was thought to repress the expression of GAL genes at a single threshold concentration. This observation was based on sampling a scarce space of all possible combinations of both glucose and galactose. In our work we measured GAL metabolic gene expression in hundreds of combinations of galactose and glucose. Our observations show that, contrary to the traditional view, cells induce the expression of GAL genes following a ratio of galactose and glucose. We then set out to characterize potential mechanisms that underlie the ability to sense a ratio of two signals. A systematic genetic dissection of galactose and glucose pathway suggests that hexose transporters are the most likely molecular players where competition between carbon sources might lead to ratio sensing. This represents a previously unidentified layer of multiple signal processing in a component outside of the canonical galactose and glucose signaling pathways.

Signaling pathways connect the outside world and the inside of the cell. Therefore, mutations that affect how signaling pathways transfer information can also affect the ability of a cell to survive. How natural genetic variation impacts signaling pathways has remained a largely unexplored area. Chapter 3 explores how genetic changes in signaling underlie the variability in responses to galactose and glucose in a collection of domesticated and natural isolates of yeast. Experiments show that each natural isolate of *S. cerevisiae* has the ability to induce (or repress) GAL genes at a threshold concentration of inducer (or repressor) that widely differs between strains. For each strain we defined a new metric, the set point of induction, which corresponds to the concentration of inducer
(or repressor) that leads to half maximal induction of the GAL genes. The set point of induction is a highly variable trait across the *S. cerevisiae* strains, we set out to determine the genetic basis of this variability. Given the large number of components involved in the sensing of galactose and glucose we were expecting to find multiple genes underlying this trait. To our surprise, a quantitative genetics approach uncovered a single locus carrying the signal transducer *GAL3*. We show that *GAL3* contributes a large fraction of the phenotypic variability in different genetic backgrounds. Then we proceed to examine the contribution of other genes to the trait of interest.

Chapter 4 examines the regulation of transcription, a central mechanism in biology. Traditional knowledge states that gene expression in the genome arises from specific regions called open reading frames (ORFs) that are transcribed from promoters in a single direction. Recently, highly sensitive technologies that probe genome-wide transcripts have revealed that transcription does not only occur from promoters through ORFs. In fact, transcription can proceed in both directions of most promoters in the genomes of a wide range of species, ranging from yeast to humans. The canonical direction of transcription is known as the coding transcription whereas the transcription in the opposite direction is divergent. For years, the phenomenon of bidirectional transcription remained elusive because the most abundant and easily detectable transcripts in the cell are coding transcripts, whereas only a relatively small fraction are divergent transcripts. This suggests that mechanisms that reduce the number of divergent transcripts are in place in the cell. Most of the proposed mechanisms so far to explain this observation involve the degradation of divergent transcripts. To identify genes affecting the ratio of both types of transcripts, a bidirectional fluorescent protein reporter construct
was introduced into the yeast nonessential gene deletion collection. Screening this collection revealed chromatin assembly as an important regulator of divergent transcription. We then turned our attention to one of the hits in the screen, the CAF-I protein complex. Using nascent transcripts sequencing (NET-SEQ) we determine that CAF-I acts as a genome wide regulator that directly limits divergent transcription. Our work highlights the importance of chromatin-mediated effects in regulating the directionality of gene transcription.
Chapter 2: Galactose metabolic genes in yeast respond to a ratio of galactose and glucose

Renan Escalante-Chong*, Yonatan Savir*, Sean M. Carroll, John B. Ingraham, Jue Wang, Christopher J. Marx and Michael Springer


*equal contributions

Natural environments are filled with multiple, often competing, signals. In contrast, biological systems are often studied in “well-controlled” environments where only a single input is varied, potentially missing important interactions between signals. Catabolite repression of galactose by glucose is one of the best-studied eukaryotic signal integration systems. In this system, it is believed that galactose metabolic (GAL) genes are induced only when glucose levels drop below a threshold. In contrast, we show that GAL gene induction occurs at a constant external galactose:glucose ratio across a wide range of sugar concentrations. We systematically perturbed the components of the canonical galactose/glucose signaling pathways and found that these components do not account for ratio sensing. Instead we provide evidence that ratio sensing occurs upstream of the canonical signaling pathway and results from the competitive binding of the two sugars to hexose transporters. We show that a mutant that behaves as the classical model expects (i.e., cannot use galactose above a glucose threshold) has a fitness disadvantage compared with wild type. A number of common biological signaling motifs can give rise to ratio sensing, typically through negative interactions between opposing signaling molecules. We therefore suspect that this previously unidentified nutrient sensing paradigm may be common and overlooked in biology.


2.1 Introduction

The ability to integrate multiple cues about nutrient availability from the environment and coordinate uptake, metabolism, and regulatory networks is a major determinant of microbial cell fitness (Broach, 2012; Cai and Tu, 2012; Chubukov et al., 2014). The energy and building blocks needed for growth can come from many different sources, leading to a complex combinatorial signal integration problem. In an environment that contains a mixture of sugars, such as glucose and galactose, microbial cells regulate their response according to a carbon hierarchy mediated by catabolite repression. Galactose metabolic genes (GAL genes) are induced to a significant degree only after glucose-based catabolite repression is relieved, resulting in a lag in growth at the point of glucose exhaustion while GAL pathway proteins are produced (Broach, 2012; Cai and Tu, 2012; Chubukov et al., 2014; Dienert, 1900; Jacob and Monod, 1961). Recent studies of sugar integration in bacteria (Kaplan et al., 2008; Setty et al., 2003) suggested that in these organisms the combinatorial response results from the multiplication of individual responses to different sugars.

The response of Saccharomyces cerevisiae to galactose is one of the best-studied eukaryotic signaling pathways (Bhat, 2008; Chubukov et al., 2014; Dienert, 1900; Douglas and Hawthorne, 1966; Gancedo, 1998; Jacob and Monod, 1961; Ptashne and Gann, 2002a; Timson, 2007). The GAL response has become a canonical example for combinatorial signal integration based on a genetic switch (Bhat, 2008; Douglas and Hawthorne, 1966; Johnston, 1987; Ptashne and Gann, 2002b). All GAL genes are induced by the activator Gal4p in response to galactose (H C Douglas, 1964) but repressed by Mig1p when glucose is present (Nehlin et al., 1991). The inhibition of GAL
genes by glucose is thought to occur at a threshold concentration, with signal integration occurring at promoters (Ptashne and Gann, 2002a). These conclusions rest on a limited sampling of combinations of concentrations of glucose and galactose (Figure A-6) (Acar et al., 2005; 2010; Avendaño et al., 2013; Bennett et al., 2008; Biggar and Crabtree, 2001; Venturelli et al., 2012). Our goal, therefore, was to use modern high-throughput techniques that allow us to characterize the GAL genes’ metabolic response in detail.

2.2 Results

2.2.1 GAL Metabolic Genes Respond to the Ratio of Glucose and Galactose

We grew cells in ~500 combinations of glucose and galactose (Figure 2-1 A and B) spanning a ~1,000-fold range of glucose and galactose concentrations. We monitored the expression of a GAL1 promoter yellow fluorescent protein fusion (GAL1pr-YFP) in a derivative of the laboratory strain S288C (Appendix A, sections I and II and Table A-1).

Gal1p, a galactokinase that catalyzes the first step in the Leloir pathway (Bhat, 2008) is induced in the presence of galactose. We grew cells at low density so that the extracellular sugar concentrations are nearly constant throughout the course of the experiment, even at low sugar levels (Appendix A, sections III and IV). Previous studies have used the average population expression level from a GAL promoter as a metric for response, which can obscure low but significant expression. To identify the decision to express GAL genes in a manner that is less dependent on absolute expression level, we used flow cytometry to quantify the percentage of cells expressing a GAL1pr-YFP reporter above basal levels (Figure 2-1). We define basal levels to be the response of cells grown in 2% (wt/vol) glucose in the absence of galactose (Figure 2-1A and Appendix A).
Contrary to the classic view, we found that GAL genes do not respond simply to a threshold concentration of glucose; the decision to induce GAL genes instead depends on the ratio of glucose and galactose (Figure 2-1 B-D; replicates on Figure A-1). The response was not simply a multiplicative combination of the independent behavior of cells in glucose or galactose (Figure A-2). The value of the ratio was nearly constant over at least a 50-fold range of glucose and galactose concentrations (Figure 2-1 B). Below a glucose concentration of \( \sim 0.006\% \), cells responded solely to a threshold of galactose (Figure 2-1 B). The result was insensitive to decreases in starting inoculum density, confirming that nutrient depletion is not significant in our experiments (Appendix A, A.4 and Figure A-3 A–C). Furthermore, modeling shows that nutrient depletion would not create the appearance of ratio sensing (Appendix A, A.4 and Figure A-3 D and E). We directly verified that the ratio-sensing behavior was a steady-state, depletion-independent, single-cell phenomenon by monitoring the kinetics of induction for 8 h at several glucose and galactose concentrations in a microfluidic device with constant nutrient replenishment (Figure 2-1, Right and Appendix A, A.4 and Figure A-4). The onset of the decision occurs within 1 h (Appendix A, Figure A-4). Most cells are induced by 4 h, and steady-state is reached by 6 h (Appendix A, Figure A-4). This behavior was also observed in two other strains, BC187 and YJM978, isolated from a vineyard and a clinical sample, respectively (Figure 2-1E), showing that ratio sensing is not an aberrant behavior in a single laboratory strain. Furthermore, the existence of a ratio is robust to dosage perturbation of GAL genes (Appendix A, Figure A-5).

How can our results be reconciled with previous work that did not report ratio sensing? All previous studies examined a relatively small range of concentrations, such that
deviations from the expected threshold behavior were easily interpreted as noise; our study used a concentration range that was approximately 10-fold larger than previous studies (Appendix A, Figure A-6 D). Many studies also sampled sparsely in the concentration range that they used, obscuring the differences between ratio and threshold sensing (Appendix A, A.3 and Figure A-6 D). The metric we use here, which deconvolves expression level from the decision (Figure 2-1 and Appendix A, A.5 and Figure A-6), also helps to show the behavior clearly, because it is more responsive at low concentrations where individual cells begin to induce than at high concentrations of sugar where induction is nearly saturated. With a large enough concentration range, however, the ratio-sensing behavior would have been readily observed independent of which metric was used (Appendix A, Figure A-6).

Ratio sensing has not been previously described for carbohydrates but has been phenomenologically described for the sensing of NADH/NAD$^+$, ATP/ADP, and X vs. autosomal chromosome levels (Atkinson and Walton; Berg et al., 2009; James E Madl, 1979). In the case of ATP/ADP, ratio sensing was proposed to result from mutually exclusive binding to the $\gamma$ subunit of AMPK (Scott et al., 2004), but clarity regarding the mechanism is still lacking. In the GAL pathway, an obvious hypothesis would be that ratio sensing might be accomplished at the $GAL1$ or other GAL promoters (Ptashne and Gann, 2002a). Glucose and galactose signals converge on these promoters through Mig1p and Gal4p, respectively (Figure 2-2 A). Alternatively, ratio sensing could occur upstream of either the canonical glucose or galactose signaling pathways.
The galactose pathway responds to the ratio of galactose and glucose.

All experiments are in at least triplicate; replicates in Appendix A, Figure A-1. (A) Schematic of experiment and metric to measure steady-state GAL pathway response in S. cerevisiae in hundreds of combinations of glucose and galactose. The induced fraction (IF, hashed area) is computed (Appendix A, A.5) by estimating the fluorescence probability distribution for a given well (black curve) and taking the fraction of area outside the probability distribution of cells grown in glucose alone (green curve). (B) (Left) Flow cytometry (FCM) of response. The decision front is a linear fit to the concentrations at which 20% of all cells in the population show induction (IF >0.2). (Right) Comparison of cells monitored by live microscopy to FCM at three sugar mixtures, denoted by numbered squares (Appendix A, Figure A-4). (C) Fraction of inducing cells as a function of the ratio of galactose and glucose concentrations. Each well in B is represented by a single dot. The line is a 1D sigmoidal curve that depends mainly on the ratio of galactose and glucose. (D) Comparison of models of signal integration (Appendix A) by threshold sensing (Upper) and ratio sensing (Lower), displayed as in B and C. (E) Decision fronts, calculated as in B, for three strains of S. cerevisiae.

2.2.2 Ratio Sensing Is Generated Upstream of the Canonical Gal Pathway

To identify the mechanism for ratio sensing, we first tested whether glucose signaling is independent of galactose levels, by measuring the fraction of cells with Mig1p-GFP in the nucleus or cytoplasm in different galactose/glucose combinations (Figure 2-2 B and
Appendix A, Figure A-7). Mig1p localizes to the nucleus in the presence of glucose and to the cytoplasm in the absence of glucose (De Vit et al., 1997). As expected, Mig1p-GFP localization is independent of galactose concentration (Figure 2-2 B). To further confirm the independence of the glucose branch from galactose we measured the response of a gal80Δ strain. Gal80p is a repressor of Gal4p, which in turn induces GAL1; in a gal80Δ background GAL1pr is constitutive (i.e., galactose independent) (Torchia et al., 1984). Indeed, in this background the ratio sensor is broken; the response is converted into a threshold sensor that depends mainly on glucose (Figure 2-2 C; a quantitative comparison of glucose thresholds is shown in Appendix A, Figure A-7). With respect to glucose inhibition, a gal80Δ strain therefore mimics the classic threshold expectation.

To test whether ratio sensing occurs in the canonical GAL pathway (i.e., downstream or at Gal3p) (Figure 2-2A), we monitored the activity of GAL1pr-YFP in a mig1Δ mutant. Because Mig1p mediates the repression of the GAL pathway by glucose (Gancedo, 1998; Timson, 2007), a mig1Δ mutant should be sensitive only to galactose levels, responding as a galactose threshold sensor regardless of whether ratio sensing through Mig1p is achieved directly or indirectly. Surprisingly, we found that even in a mig1Δ strain GAL1pr-YFP expression is still sensitive to the ratio of galactose and glucose (Figure 2-2D). The ratio sensing ability of the mig1Δ strain is not due to the action of other transcription factors, because a gal80Δ mig1Δ strain is constitutively active for GAL1pr-YFP expression; that is, the activation of the GAL1 promoter is not dependent on either glucose or galactose in this strain (Figure 2-2E and Appendix A). These results are consistent with previous observations showing that glucose represses GAL1pr expression even in the absence of Mig1p (Biggar and Crabtree, 2001). These results imply that either
an intracellular mode by which glucose regulates the galactose pathway has been missed, or that ratio sensing is achieved neither at the GAL1 promoter nor in the canonical GAL pathway, but upstream of Gal3p.

Figure 2-2 The GAL pathway senses the ratio of glucose and galactose upstream of known glucose regulation. (A) The GAL regulatory network. (B) Mig1p localization as a function of glucose and galactose concentrations. Cells expressing Mig1p-GFP were grown under the same conditions as in Figure 2-1 B and imaged after 8 h (Appendix A, A.7); steady-state localization was typically achieved in minutes. Images show representative cells at the indicated sugar concentration. Each concentration is the result of at least 20 cells. The number of cells with nuclear Mig1p-GFP decreases with glucose levels in a galactose-independent manner. (C) In a gal80Δ background, the ratio response is converted to a threshold response (i.e., in the absence of Gal80p the response is galactose independent). Experiment performed in duplicate. Data for no glucose conditions is not shown for clarity (Methods). (D) In a mig1Δ background, cells continue to respond to the galactose:glucose ratio. Experiment performed in duplicate. Solid line represents the decision front of the mig1Δ; dashed line represents the decision front of the wild-type strain (from Figure 2-1 B). (E) In a mig1Δ gal80Δ strain, the response is constitutive and does not dependent on either glucose or galactose.
2.2.3 The Galactose Transporter Gal2p Is Not Required for Ratio Sensing

Because Gal3p directly senses internal galactose levels, ratio sensing upstream of Gal3p suggests a role for sugar transport in ratio sensing. When the GAL pathway is induced, the majority of galactose is imported through the Gal2p transporter, which transports both glucose and galactose with high affinity ($K_m \approx 1$ mM) (Ramos et al., 1989; Reifenberger et al., 1997). Gal2p is part of the GAL pathway; Gal2p levels are low in glucose media. Nevertheless, it is possible that even the low levels of Gal2p expressed in high glucose are important for ratio sensing. This would not be unprecedented: in the case of Lac induction in Escherichia coli, stochastic low-level expression of transporters is critical for the response (Choi et al., 2008). We therefore measured $GAL1pr$-YFP in a $gal2\Delta$ strain. Similar to the results with the $mig1\Delta$ mutant, a $gal2\Delta$ mutant does not “break” the ratio sensor (Figure 2-3 A); in both cases the mutation affects the ratio sensor, but neither mutant eliminates the ratio-sensing behavior. We interpret these results as strong evidence that the mechanism responsible for ratio sensing involves components outside the canonical galactose sensing pathway (Figure 2-2 A).
Figure 2-3 Galactose uptake depends on the ratio of galactose and glucose even in the absence of Gal2p.

(A) Gal2p and many of the Hxt1-17p family of hexose transporters import glucose and galactose but have differing relative affinities for the two sugars. Deletion of GAL2 does not eliminate ratio sensing. Black and red lines are the decision front (Figure 2-1 E). (B) Incorporation of $^{12}$C-galactose and $^{13}$C-glucose into amino acids as measured by GC-MS in a gal2Δ gal80Δ mig1Δ strain (Appendix A). This strain constitutively expresses the GAL pathway. Error bars are the SD from three biological replicates (duplicates for the highest glucose to galactose ratio). The slope of the fitted line (black line) is 1/170; the expectation based on literature uptake measurements is 1/250 (28, 29). (Inset) Breakdown of the data into two different galactose concentrations.

In a gal2Δ strain, the family of hexose transporters [Hxt1-17p or Mal11p, Mph2p, and Mph3p (Wieczorke et al., 1999)] are likely to be the main transporters of galactose and a likely source of ratio sensing (Figure 2-3 A). The HXT members transport glucose with various affinities ($K_m$ from ~1 mM to 100 mM) (Ramos et al., 1989; Reifenberger et al., 1997), and some also import galactose, albeit with significantly lower affinity ($K_m$ ~250 mM) (Ramos et al., 1989); Hxt14p can even support growth on galactose in a strain where all other hexose transporters have been deleted (Wieczorke et al., 1999). Thus,
ratio sensing might result from competition between the sugars during uptake. In a competitive uptake regime, the intracellular galactose concentration would depend on the ratio of the extracellular galactose and glucose concentrations (Appendix A, section VIII and Figure A-8 and Figure A-9).

2.2.4 Galactose Uptake Depends on the Ratio of Extracellular Sugars Concentrations

To directly test whether uptake of galactose through the hexose transporters depends on the extracellular ratio of galactose and glucose concentrations, we measured galactose uptake in mixtures of U-$^{13}$C-glucose and $^{12}$C-galactose. Because intracellular carbohydrates are rapidly metabolized, measuring the incorporation of $^{13}$C and $^{12}$C into amino acids using gas chromatography mass spectrometry (Zamboni et al., 2009) provides information on uptake rates; the ratio of incorporated $^{12}$C and $^{13}$C is equal to the ratio of galactose and glucose uptake rates (Appendix A, section X). To distinguish the role of hexose transporters from the effects of intracellular regulation, we constructed a $gal2\Delta\ gal80\Delta\ mig1\Delta$ strain. This strain is not responsive to glucose or galactose but constitutively expresses GAL genes (Figure 2-2E). Incorporation of $^{13}$C and $^{12}$C thus depends solely on the relative sugar uptake rates and not on the induction of the GAL pathway. We pregrew this strain in U-$^{13}$C-glucose medium and transferred it into media containing mixtures of U-$^{13}$C-glucose and $^{12}$C-galactose for two doublings (Appendix A, Table A-3). We found that the ratio between $^{12}$C and $^{13}$C incorporated into amino acids, and hence galactose uptake, increases as extracellular galactose:glucose ratio is
increased (Figure 2-3B). The $^{12}$C:$^{13}$C ratio increases as extracellular galactose is increased but decreases as extracellular glucose is increased (Figure 2-3B).

Quantitatively, this result is consistent with a “passive” model of competitive uptake of glucose and galactose by the transporter, which predicts that relative uptake depends on the extracellular sugar ratio multiplied by the relative affinity of the transporter for each sugar ($K_m$ ratio). Our measurements yield a $K_m$ ratio of 170 (Figure 2-3B), similar to the $K_m$ ratio of 250 calculated from literature reports (Ramos et al., 1989; Reifenberger et al., 1997). The concentration of glucose at which the response changes from a ratio sensor to a galactose threshold sensor, 0.006% as measured in Figure 2-1B, is close to the $K_m$ of the high affinity hexose transporter for glucose $\sim$1 mM, or 0.002% glucose. This is consistent with a competitive uptake model: glucose concentrations below the $K_m$ of the transport have quickly diminishing effects on the uptake of galactose, thereby making galactose uptake glucose independent at low glucose concentration.

2.2.5 Ratio Sensing Can Provide a Fitness Advantage

It is possible that ratio sensing in the sugar metabolism pathways in yeast evolved to compensate for an inevitable lack of perfect discrimination between different sugars in the hexose transporter. Because of substrate competition for the transporter, high galactose will partially inhibit glucose uptake, and cells that do not induce GAL gene expression cannot compensate for the decreased glucose flux by metabolizing galactose.

On the other hand, it is also possible that ratio sensing is desirable for other reasons (for example, to allow the cell to sense when using multiple sugars is a better decision than using only a single sugar) and that the lack of discrimination of the hexose transporters is
in itself a selected trait. Consistent with the latter possibility, there is a wide variation in
the selectivity of HXT family transporters for glucose relative to other sugars, and many
do not sustain growth on medium with galactose as the sole carbon source (Wieczorke et
al., 1999). Thus, a cell could evolve to express only highly selective hexose transporters
if ratio sensing were undesirable.

The biological advantage of ratio sensing is most likely during a dynamic process such as
depletion of glucose in mixed sugar environments. However, no mutant currently exists
whose only defect is to convert the ratio response to a threshold response (e.g., a gal80Δ
has a fitness disadvantage in many media). Therefore, to establish whether ratio sensing
can offer a selective advantage relative to a threshold sensing response, we compared the
fitness of a gal4Δ strain to that of a wild-type strain in two conditions: glucose only, and
a glucose/galactose mixture (Figure 2-4A). A gal4Δ strain cannot mount a transcriptional
response to galactose and therefore behaves in a glucose/galactose mixture as if it were in
glucose alone (Gill and Ptashne, 1988), a behavior that phenocopies a threshold sensing
strain in this media regime (Figure 2-4A). When co-cultured in 0.016% glucose the wild-
type and gal4Δ strains grew comparably Figure 2-4A); 0.016% glucose is above the
concentration of glucose at which ratio sensing is observed (Figure 2-1B). When 2%
(wt/vol) galactose is added to the 0.016% glucose medium the wild-type strain has a
significant fitness advantage of 0.1% per hour ± 0.01% SE of mean, with a P value of
0.01 (two-tailed t test; Figure 2-4A). At this concentration all wild-type cells induce the
GAL pathway maximally. Given the steady-state advantage to the ratio response we
observe here, it is likely that if a true threshold-sensing strain could be constructed we
would find that it is at a disadvantage compared with the ratio-sensing strain in dynamic
environments as well.

Figure 2-4 Biological implications, implementations, and regimes for ratio sensing. (A) The ability to use galactose even in the presence of glucose gives cells a fitness advantage. A wild-type strain was competed against a gal4Δ in two concentrations of carbon: (i) 0.016% glucose or (ii) 0.016% glucose and 2% (wt/vol) galactose. The 0.016% glucose concentration places cells in the ratio sensing regime (Left, green and black squares). Ten independent replicates were grown until gene expression reached steady state (8 h), then samples were taken every 2 h for 10 h to calculate cellular fitness. Error bars are the SEM. (B) Ratio output can be generally achieved by a simple module in which two input molecules, an activator (green) and a repressor (blue), bind to an integrator molecule—a promoter, transporter, scaffold protein, etc. Mutual inhibition, e, is necessary for a robust ratio response. (C) Futile cycles, such as phosphorylation and dephosphorylation or acetylation and deacetylation, can also create ratio sensors. Ratio sensing is achieved when the enzymes are not saturated.
2.3 Discussion

We measured the response of yeast to hundreds of mixtures of glucose, the preferred carbon source, and galactose, a less preferred carbon source. Although glucose above a certain concentration threshold is commonly thought to repress other carbohydrate metabolism pathways, our results show that the GAL metabolic genes respond to the ratio of galactose and glucose. We show that ratio sensing is not generated from the interaction of transcription factors on the promoter, as one might expect from the literature, but instead is achieved upstream of the canonical GAL pathway. Moreover, we show that the intracellular galactose concentration depends on the ratio of galactose and glucose, highlighting an information processing step at the level of import.

Our results show that ratio sensing can be achieved even in the absence of the high-affinity galactose transporter Gal2p (Figure 2-3A). Relative uptake of glucose and galactose suggest that competitive binding of glucose and galactose to hexose transporters is responsible for setting the initial ratio response. GAL2 is induced as part of the GAL response (Bhat, 2008; Timson, 2007). The affinity of Gal2p for both galactose and glucose is approximately 1 mM; Gal2p thus has significantly higher affinity for galactose than the HXT transporters but should still exhibit competitive transport. Hence, even after significant Gal2p expression, cells would be expected to respond to the ratio of glucose and galactose but at a shifted “setpoint.” This is consistent with our observation that a gal2Δ strain responds to the ratio of glucose and galactose but the setpoint of this ratio is shifted several-fold from a wild-type cell (Figure 2-3A). These results raise an
intriguing hypothesis: because the setpoint of the ratio response can be modulated solely by changing the relative expression of transporters with different affinities, the large number of different hexose transporters may be involved in an intricate and physiologically tunable information processing layer.

Yeast cells respond both to extracellular cues, using transmembrane sensors such as Snf3p and Rgt2p that bind extracellular glucose, and to intracellular cues, using internal sensors such as Gal3p and Hxk2p (Ptashne and Gann, 2002a) that bind to galactose and glucose, respectively. The fact that galactose is sensed intracellularly is crucial for our model; competitive transport will only affect an internal sensor. Our data thus identify a previously unidentified layer of regulation above the regulation represented by previous mechanistic models of glucose and galactose integration. Because intracellular sensing is common in all domains of life from bacteria through mammals (Schlax et al., 1995; Yamamoto, 1997), information processing through competitive transport could be a common mechanism.

This phenomenon of ratio sensing through competitive binding at a transporter can be generalized to any situation in which an activator and a repressor both bind to a third molecule and affect its function, provided that there is some mutual exclusion (Figure 2-4B and Appendix A, A.8 and Figure A-8 and Figure A-9). Mutual exclusion through competitive binding can occur in almost any biological system and can be mediated by any substrate (e.g., DNA, RNA, or protein), opening up the possibility that ratio sensing could occur in a wide range of situations. Another common biological process that can create a ratio sensor is a futile cycle (Figure 2-4C), in which two reactions run simultaneously in opposite directions, burning energy. When the enzymes are unsaturated
the steady-state amount of modified substrate is a function of the ratio of the activities of the kinase and phosphatase. In both classes of modules the molecule being bound or modified needs to be limiting (Appendix A, A.9 and Figure A-8). These are common and simple motifs that can give rise to ratio sensing, although many other methods of creating a ratio sensor may exist (Daniel et al., 2013).

Although we found that ratio sensing gives cells a fitness advantage in a specific steady-state environment, in natural settings the ratio response is more likely to be relevant in the context of dynamic processes, such as the depletion of glucose in a mixed sugar environment. We observed a ratio response in multiple strains (Figure 2-1 E), but the specific ratio of galactose and glucose at which a given strain starts to induce GAL1 varies between strains. These differences might affect the duration of the diauxic (Monod, 1949); strains that induce at a higher galactose to glucose ratio might be expected to begin to induce galactose genes before glucose is fully depleted. Indeed, strains that respond at a lower galactose to glucose ratio (BC187 and S288C in Figure 2-1E) have a shorter diauxic lag than a strain that responds at a higher galactose to glucose ratio (YJM978; Figure 2-1E) (Wang et al., 2015).

Our results add a previously unidentified layer to the standard description of glucose repression in yeast and highlight the possibility that ratio sensing is frequent, and potentially useful, in biology. In many situations ratios may be more biologically robust than absolute concentrations. As techniques allowing for multidimensional analyses are becoming more readily available, we anticipate that ratio sensing will be identified in many other settings. Our work suggests that a critical portion of information processing in a major metabolic decision is made upstream of the canonical signaling network and
highlights the dual role transporters can play in both nutrient uptake and signal integration.

2.4 Methods

2.4.1 Growth Conditions and Media

Cells were grown for ~14–16 h in synthetic minimal media with 2% (wt/vol) raffinose, to an OD of ~0.3, and then diluted 1:100 in mixtures of glucose and galactose. Cells were grown for 8 h at 30 °C in flasks or 96-well plates and then washed twice in TE (10 mM Tris, 1 mM EDTA, pH 7.5) in preparation for flow cytometry. Cells to be imaged by microscopy were transferred to a micro-well plate (Matriplate from Metrical Bioscience) coated with concanavin A grown in synthetic media with different mixtures of glucose and galactose. A gal80Δ strain growing in raffinose constitutively activates GAL genes. Therefore, the gal80Δ strain was pregrown in 2% (wt/vol) glucose. When switched from glucose to no glucose plus very low galactose media, the gal80Δ strain does not grow and marginally induces. Hence, we do not show data for no glucose in this strain.

2.4.2 Flow Cytometry

Samples were measured by flow cytometer (LSRII with high-throughput sampler; Beckton Dickinson) as previously described (Springer et al., 2010).

Fluorescence Microscopy. Images were captured with a Ti Eclipse inverted Nikon microscope using Micromanager. A Hamamatsu Orca-R2 camera was used to capture fluorescent images and a Scion CFW-1612M for bright-field images. Nuclear and
cytoplasmic of Mig1p-GFP localization was analyzed using custom MATLAB software.

**2.4.3 Microfluidics and Live Cell Imaging**

Cells grown in 2% (wt/vol) raffinose to an OD of 0.3 were loaded into a Y04C yeast microfluidic plate (CellASIC ONIX system). Cells were maintained in 0.011% glucose and 0.0039% galactose; 0.011% glucose and 0.0221% galactose; 0.0625% glucose and 0.0039% galactose; or 0.0625% glucose and 0.0221% galactose for ~12 h with images taken, as described above, every 15 min.

**2.4.4 Strain Construction**

The reporter \textit{GAL1}pr-YFP, and the constitutively expressed fluorophore TDH3pr-mCherry, TDH3pr-BFP, or TDH3pr-TagBFP2 was transformed into the HO locus (Homothallic switching endonuclease) of the prototrophic S288C, BC187, and/or YJM978 strains. Hemizygous deletion strains were made by mating a haploid deletion strain containing the reporter \textit{GAL1}pr-YFP and a haploid S288C containing TDH3pr-mCherry or a TDH3pr-TagBFP2. Other deletion strains (Appendix A, Table A-1) were constructed by integrating a KanMX cassette in each of the loci using standard PCR and yeast transformation protocols.

**2.4.5 Data Analysis.**

Analysis of flow cytometry data were performed using custom-written MATLAB code (available upon request). A 2D Gaussian mixture model fit to the mCherry and BFP or the mCherry and nonfluorescent populations was used to segment the different competing populations.
2.4.6 GC-Mass Spectrometry

Cells were grown for 16 h in 1% U-$^{13}$C-glucose, then diluted into fresh 1% U-$^{13}$C-glucose until cells reached exponential phase. After 6 h cells were transferred to different combinations of U-$^{13}$C-glucose and $^{12}$C-galactose and grown until the cells had doubled once. Cells were spun down and pellets collected for processing in the mass spectrometer as described by Zamboni et al. (Zamboni et al., 2009).

Fitness Measurements. A diploid gal4A strain tagged with a fluorescent mCherry (or BFP) was grown with the wild-type strain S288C tagged with BFP (or mCherry) in synthetic 2% (wt/vol) raffinose media for 14 h to an OD of ~0.3. Cells were washed three times in synthetic media with no carbon and then inoculated in synthetic media with 0.0156% glucose and 2% (wt/vol) galactose or with 0.0156% glucose. Samples were collected every 2 h in TE (pH 7.5) plus 0.1% sodium azide and read by flow cytometry.

2.5 References


Author contributions

R.E.-C., Y.S., and M.S. designed the study; R.E.-C. and Y.S. performed the experiments; R.E.-C., Y.S., and M.S. analyzed the data; S.M.C. and C.J.M. helped develop and perform the mass spectrometry experiments; J.B.I. helped develop the numerical framework; J.W. helped develop the experimental setup; Y.S. and M.S. developed the analytical framework; and R.E.-C., Y.S., and M.S. wrote the paper.
Chapter 3: Genetic dissection of the natural variation of galactose and glucose signaling in *S. cerevisiae*

Renan Escalante-Chong, Kayla B. Lee and Michael Springer

Connecting genotypic variation to phenotypic diversity remains a central problem in biology. Of special interest are the genetic networks that process information from the environment and combine them to trigger a specific response. Here, we measure how the model organism *Saccharomyces cerevisiae* triggers galactose metabolic genes in mixed carbon environments of galactose and glucose. Specifically, we characterize the variability in gene expression in response to mixed carbons among natural isolates. To elucidate the genetic basis and potential mechanisms underlying the observed variation we used bulk segregant analysis to identify quantitative trait locus (QTL) analysis between two natural isolates BC187 and YJM978. Our results suggest that ~50\% of the variation between these strains can be explained by variation in a locus containing the galactose signal transducer *GAL3*. We also show that background genes in different strains do not seem to affect the effect of *GAL3* alleles, suggesting no epistatic interactions between *GAL3* and other genes.

3.1 Introduction

Organisms in nature harbor an enormous amount of phenotypic diversity for various traits. Most of these traits vary in magnitude rather than in kind, giving rise to a continuum of phenotypic variants. Examples of this type of variation are observed in
traits such as height, weight, color patterning, disease susceptibility and differences in sensitivity to drugs. This phenotypic variation often involves the combined action of multiple genes and interactions with the environment, making it challenging to pinpoint mechanisms that connect genotype and phenotype (Hill, 2010; Mackay et al., 2009). Quantitative genetic methods seek to determine how variation in specific regions of the genome underlie differences in phenotypic traits (Liti and Louis, 2012). Here, we apply this approach to better understand the genetic basis of variation in a physiological response to the environment.

The budding yeast *S. cerevisiae*, the workhorse of many genetic studies, has recently emerged as a powerful model to study phenotypic natural variation and its underlying genetic basis. *S. cerevisiae* was thought to be a completely domesticated species but actually lives in many ecological niches with contrasting growth conditions (Fay and Benavides, 2005). Natural isolates have been isolated from ripe fruits, oak trees, soil, sources related to alcoholic beverage fermentation and even from immunosuppressed patients (Liti et al., 2009; Schacherer et al., 2009). Perhaps reflecting this diversity of ecological niches, *S. cerevisiae* strains also show a large diversity in growth phenotypes (lag, rate, efficiency) molecular phenotypes (metabolite abundance, RNA and protein levels) and morphological traits (cell size, shape) (Liti et al., 2009; Nogami et al.; Skelly et al., 2013; Warringer et al., 2011). The genome sequences of these strains reveals a reservoir of standing genetic variability (Bergström et al., 2014; Liti et al., 2009; Skelly et al., 2013; Warringer et al., 2011). Overall strains fall into subgroups with unique genetic variation, defining clean genetic lineages (Liti et al., 2009; Schacherer et al., 2009).

Many quantitative genetic studies have sought to link the large phenotypic variability in
S. cerevisiae to its genetic variability. Some of the examined phenotypes include traits such as expression levels (Brem et al., 2002), small-molecule sensitivity (Perlstein et al., 2006), cell morphology (Nogami et al.), mitochondrial instability, heat resistance (Steinmetz et al., 2002), growth rates in 23 different environments (Cubillos et al., 2011) and resistance to sodium chloride and caffeine (Treusch et al., 2015). Most of these studies measure a quantitative trait in several pre-determined conditions. In contrast, our approach seeks to determine conditions of the environment that lead to a specific value of the quantitative trait. Although more labor intensive, this approach has the potential of better describing the physiological range of the response to a specific stimulus, in our case nutrients.

Nutrient signaling pathways in cells translate information from the environment to trigger an appropriate response. In yeast, induction of the galactose metabolic genes (GAL genes) is required in order to metabolize the carbohydrate galactose. The galactose regulatory pathway is an ideal system to study how genetic variation affects gene induction. The key components that participate in the induction of the GAL genes are very well characterized (Edwards and Gifford, 2012; Lohr et al., 1995; Timson, 2007). Additionally, the GAL pathway has a number of interesting quantitative features that could be affected by genetic variation. These features include the ability to access multiple steady states (Acar et al., 2005), the presence of a transcriptional memory (Kundu and Peterson, 2010) and the ability to sense a ratio of carbons (Escalante-Chong et al., 2015). In a previous study, we determined that different strains induce GAL genes following a ratio of galactose and glucose, but the specific ratio of induction can dramatically change between strains.
The combined actions of Gal4p, Gal80p and Gal3p control the expression of the GAL genes (Timson, 2007). Glucose sensing is more complex and involves at least five different pathways that monitor glucose levels in the cell (Zaman et al., 2009). Several interactions between glucose and galactose sensing have been described. For instance, the glucose induced repressor Mig1p represses the galactose transporter Gal2p, the galactose pathway activator Gal4p and several of the GAL genes including Gal1p and Gal3p (Johnston et al., 1994). Galactose has been suggested to activate the repressor of main glucose transporters Hxt via Mth1. Given the complex interplay between these two pathways it is hard to pinpoint specific changes that could explain some of the differences in the specific ratios at which different strains induced gene expression.

Here we characterized the variation in the response to combinations of galactose and glucose in 43 natural isolates of S. cerevisiae. We set out to determine the genetic basis of the variation between strains. Surprisingly, we identified a single major contributor to phenotypic differences between strains, the signal transducer GAL3.

3.2 Results

3.2.1 Measuring the variation in responses of the GAL pathway between strains

To characterize the variability in the induction of GAL genes between different strains of S. cerevisiae in combinations of galactose and glucose we monitored a GAL1 promoter fusion to a fluorescent protein YFP (Figure 3-1A). GAL1 is the first metabolic gene of the galactose catabolic pathway. Previously, we characterized the induction of GAL genes in large double gradients spanning ~500 combinations of sugars. Similar information about the induction of GAL genes can be derived by measuring expression in a titration of
glucose in a constant galactose concentration (or conversely a titration of galactose in constant glucose concentration, Figure B-1). To quantify the differences in GAL gene induction between different strains we defined a “set point of induction”. In a titration of glucose at a constant galactose concentration, the set point of induction is defined as the concentration of glucose where induction is half maximal (Figure 3-1B, see methods). An analogous definition can be applied to single galactose gradients at a constant glucose concentration. Overall the set points of induction in titrations of galactose and glucose are highly correlated (R=−0.83) (Figure B-1). Thus, to maximize our ability to characterize a large number of strains we used glucose titrations at a constant galactose concentration.

To better control for variability in the assay across experiments and compare the induction of strains more accurately, we co-cultured a query strain of interest in the presence of a reference strain. We then computed a “normalized set point Sₙ” as log₂ ratio of the set point of the query and the set point of the reference strain (Figure 3-1B). The normalized set point then represents the fold change difference between a query and a reference strain. The previously characterized strain BC187 (Escalante-Chong et al., 2015; Wang et al., 2015) induces the GAL genes over a large range of galactose and glucose concentrations making it this strain suitable as a reference strain that can be used to normalize gene induction.

**Figure 3-1 Induction of the GAL metabolic genes in galactose and glucose is a continuous trait among lab and natural isolates**
(A) The galactose metabolic genes (GAL genes) respond to galactose and glucose. GAL genes are activated by galactose and repressed by glucose. A promoter fusion of the first gene in the pathway to a fluorescent protein YFP (GAL1pr-YFP) reports the induction of the pathway in different combinations of both carbon sources.
(B) Schematic of experimental setup for GAL gene induction in a glucose gradient. Each query strain (gray) was co-cultured with the reference strain BC187 (blue), which expresses a constitutive fluorescent marker BFP, in a glucose gradient at a constant
galactose concentration (0.125% galactose) for 8 hours (see methods). Cells were then measured by flow cytometry.

(C) Illustration of set point metric calculation. Mean YFP values were computed and normalized for each strain (see methods) and a raw set point metric is computed for both query and reference strain. The normalized set point SN is the log2 ratio between the query and reference strain set points.

(D) Example of induction curves for YJM978 and BC187 strains. At a constant galactose concentration (0.125%) GAL gene expression is half-maximally repressed at very low glucose concentrations (~0.007%) in the strain YJM978 (red) whereas in BC187 (blue) half-maximal repression is achieved at a 9-fold higher glucose concentration (~0.063%). Median is shown as a black line. The 25th and 75th percentile are shown as dotted red lines.

(E) The normalized set point of induction SN for different lab and natural isolates. 37 natural isolates of S. cerevisiae were measured using the assay described in B and in methods. The set points for strains YJM978 and BC187 are shown in red and blue. A 9-fold difference in D is shown here as 3.23 difference ($2^{3.23}=\approx9$).
3.2.2 The set point of induction is a continuous, quantitative trait between natural yeast strains

Using our co-culture induction assay we characterized the induction of the GAL genes in ~45 natural isolates. These strains were isolated from 5 different continents and represent very diverse ecological niches (see Table B-1). West African strains have been previously shown to carry a GAL3 stop codon mutation which renders them unable to grow using galactose as a sole carbon source (Warringer et al., 2011). Consistent with this observation, we detected no induction in these strains (Figure B-2). Additionally, we observed strains with a high basal expression even in the highest glucose concentration (strain 378604X, Figure B-2). Overall, the set point of induction across 37 strains varies over a 32-fold range ($2^{4.19} = 31.19$). Many strains have similar set points, in fact 50% of them have set points contained in 2-fold interval (between ~0.31 and ~1.3, Figure 3-1E).

3.2.3 Population and source of strains loosely match the variation in the set point of induction

Previous studies showed that the diversity in growth phenotypes (lag, rate, yield) between strains in different environments followed population boundaries. Additionally, very few traits were specific to a source (e.g. clinical, baking, fermentation) (Warringer et al., 2011). Equivalently, in this study we find that the North American group (strains YPS128, YPS606, YPS163) had very similar set points (mean -0.78+- 0.09 s.e.m.) However, other groups like the sake and the Wine/European lineages have strains with much more variable set points (Figure B-2). A group of strains from a clinical source YJM975, YJM978 and YJM981 also show very similar set points of induction (-
2.6+0.34) (Figure B-2). The set point of induction loosely matches the population and source of the strains suggesting that variation in this trait is not tightly linked to population history, and that genetic variants leading to differences in this trait could have occurred multiple times in the history of the species.

3.2.4 Bulk segregant analysis reveals a single locus underlying the differences in set point between BC187 and YJM978

Among all the strains in our study, we focused our attention on BC187 and YJM978 whose behavior in combinations of galactose and glucose has been previously characterized (Escalante-Chong et al., 2015; Wang et al., 2015). The strain BC187, a wine fermentation isolate, represses the GAL genes at relatively high concentrations of glucose (Figure 3-1 right side of the plot in blue) whereas, YJM978, a clinical isolate, represses induction at much lower glucose concentrations. Quantitatively, there is 9-fold difference between the set points of these two strains.

To determine the genetic basis of the variability in set points between strains we generated a hybrid of BC187 and YJM978. The hybrid was then sporulated to generate ~1000 F2 meiotic segregants (Figure 3-2A). To facilitate large-scale phenotyping, the meiotic segregants were then phenotyped by measuring mean YFP fluorescence at a single glucose (2^−4) and galactose (2^−3) concentration. At this concentration BC187 is fully induced (mean YFP ~4), whereas the more repressible strain YJM978 is not induced (mean YFP ~1.5). Thus, the assay cleanly distinguished between the two parental phenotypes (Figure 3-2B). The meiotic segregants mostly fall into two categories, either
**Figure 3-2** Bulk segregant analysis identifies a single locus responsible for the differences in set point between strains YJM978 and BC187.

(A) Schematic description of experimental design for bulk segregant analysis. Parental strains are crossed to create a hybrid that is then sporulated to obtain haploids. The resulting meiotic segregants contain contributions of both parental genomes and are grouped based on phenotype.

(B) YFP distributions of parental strains BC187 and YJM978 as well as meiotic segregants in the assay. BC187 (blue) is fully induced whereas YJM978 (red) is not. Meiotic segregants fall into mainly fully induced or not induced populations.

(C) LOD-score for across chromosomes reveal a single causative locus. The LOD score values were

(D) Distribution of meiotic segregant mean YFP values. The distribution is bimodal with an almost 1:1 to ratio suggesting a single locus that underlies the differences in set point of induction.

fully induced or not induced. Additionally, there are very few intermediate distributions of meiotic segregants, resulting in a bimodal distribution of mean YFP levels (Figure 3-2D). The ratio of these two phenotypes is close to 1:1, suggesting the presence of a single locus controlling the differences between the studied strains.

To maximize our ability to detect QTLs we selected 96 meiotic segregants in each
extreme of the mean YFP distribution (Lander and Botstein, 1989) (Figure 3-2D), to make 2 pools: the YJM978-like meiotic segregants which show low mean YFP values and BC187 which had high YFP values. We then sequenced both the parental strains as well as the pools of meiotic segregants (see Materials and Methods). We then identify parental single nucleotide polymorphisms (SNPs) in the segregant pools. In general, SNPs that are not associated with the trait of interest will segregate randomly (frequency 50%) in both meiotic segregant pools. Deviations from this 50% frequency indicate linkage to a specific selected trait in the pool. Moreover, since the meiotic segregants pools have opposite phenotypes, causative alleles of interest not only will deviate from the 50% frequency but they will do so in opposite directions in each of the pools, that is a different parental allele will be enriched in each of the pools (Figure B-3). To detect loci of interest and provide statistical support for each of them, we computed maximum likelihood estimates of the causal locus location (Figure 3-2C) as well as log of odds scores (LOD) using MULTIPOOL (see methods, (Edwards and Gifford, 2012)).

We expected at least one non-causative locus to be enriched in our meiotic segregant pools. This locus corresponds to the left arm of chromosome IV near coordinates 47 000 and 54 000 which corresponds to the HO locus where our reporter GAL1pr-YFP was inserted (see methods). Meioitic segregants were selected to carry a reporter for induction for subsequent phenotypic analysis. An unexpected non-causative linked allele was detected on chromosome VII (Figure 3-2, Figure B-3 and B-4) approximately between coordinates 310 000 and 326 000 reflecting a potential selection bias when meiotic segregants were picked for further analysis. Using the Saccharomyces Genome Database (SGD) we found that yeast retrotransposon elements Ty are abundant in this region of the
chromosome. Ty elements are important elements that shape the evolution of the yeast genome which in some cases have been shown to confer a fitness advantage and might be also be biasing our selection (Wilke and Adams, 1992).

The only causal locus detected in the genome with a LOD score of 29.6 is located on chromosome IV, the 90% confidence interval spans a region of ~26 kB centered around the coordinate ~465 000 (Figure 3-2C.). This region contains 9 verified ORFs YRB1, RCR2, RAD57, MAF1, SOK1, TRP1, GAL3, SNQ2 listed in Table 3-1. The S. cerevisiae genome is a highly annotated so we could find verified functions for the genes in the candidate region (Table 3-1). Most of the genes (7 out of 9) had completely unrelated functions to nutrient signaling (Table 3-1) and were not prioritized in the follow up analysis. One gene is central for GAL gene induction, the galactose signal transducer GAL3. GAL3p directly senses the levels of galactose in the cell and activates GAL gene induction as a result (Yano and Fukasawa). Another gene potentially related to nutrient sensing in the candidate locus is SOK1. This gene has been implicated in cAMP signaling (Ward and Garrett, 1994). cAMP signaling is linked to cell growth and is activated by glucose (Tamaki, 2007). Given that both GAL3 and SOK1 participate in glucose and galactose signaling we decided to examine their potential role in establishing different set points in more detail.
Table 3.1. Genes, coordinates and functions of genes identified by QTL mapping.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Coordinates on chromosome IV</th>
<th>Annotation from SGD</th>
</tr>
</thead>
<tbody>
<tr>
<td>YRB1</td>
<td>453045-453650</td>
<td>Ran GTPase binding protein</td>
</tr>
<tr>
<td>RCR2</td>
<td>454122-454754</td>
<td>Vacuolar protein; presumably functions within the endosomal-vacuolar trafficking pathway.</td>
</tr>
<tr>
<td>RAD57</td>
<td>455201-456583</td>
<td>Protein that stimulates strand exchange; stimulates strand exchange by stabilizing the binding of Rad51p to single-stranded DNA.</td>
</tr>
<tr>
<td>MAF1</td>
<td>458017-456836</td>
<td>Highly conserved negative regulator of RNA polymerase III.</td>
</tr>
<tr>
<td>SOK1</td>
<td>461247-458542</td>
<td>Protein of unknown function; involved in cAMP-mediated signaling.</td>
</tr>
<tr>
<td>TRP1</td>
<td>461842-462516</td>
<td>Phosphoribosylanthranilate isomerase; catalyzes the third step in tryptophan biosynthesis.</td>
</tr>
<tr>
<td>GAL3</td>
<td>463434-464996</td>
<td>Transcriptional regulator; involved in activation of the GAL genes in response to galactose.</td>
</tr>
<tr>
<td>SNQ2</td>
<td>465919-470424</td>
<td>Plasma membrane ATP-binding cassette (ABC) transporter</td>
</tr>
</tbody>
</table>

3.2.5 **Dissection of the candidate locus by reciprocal hemizygous hybrids suggests that SOK1 is not implicated in the trait difference**

We used reciprocal hemizygosity analysis (Steinmetz et al., 2002) to characterize the contribution of *SOK1* and *GAL3* in the set point of induction. Hybrid strains of BC187 and YJM878 were created where only one allele of the gene of interest (*SOK1* or *GAL3*) was deleted (see Figure B-5) producing a hybrid carrying only the BC187 or YJM78 allele. For the hybrid and each of the hemizygous hybrids of *SOK1* (BC187-only and YJM978-only allele) we could not reject the null hypothesis that these strains have the same set point of induction (Figure B-5) suggesting that the set point of induction is not affected by the allele of *SOK1*.

There is ~ 4-fold difference between the hybrid and the hemizygous hybrid carrying only the *GAL3* BC187 allele. Moreover, there is a ~9 fold difference between the hybrid and
the \textit{GAL3} YJM978. These results suggest that the allele of \textit{GAL3} has a strong effect on determining the set point of induction. A quantitative assessment of the effects of each \textit{GAL3} allele is complicated in this case since the set point of induction of the pathway is dosage dependent (Acar et al., 2010). Here, the hybrid carries two copies of \textit{GAL3} whereas each of the hemizygous hybrids carries only one, resulting in a combined effect of each allele and gene dosage.

3.2.6 The BC187 and YJM978 \textit{GAL3} alleles contribute a large fraction of the difference in set point between strains

To quantitatively characterize the effect of \textit{GAL3} allele we transferred the BC187 allele into YJM978 background, as well as the YJM978 allele into BC187. We started from a \textit{gal3}Δ strain in both YJM978 and BC187 background strains. We used a heterologous locus yercΔ8 on chromosome V to insert each of the \textit{GAL3} alleles. This locus has been shown not to affect growth rate (Flagfeldt et al., 2009). We inserted both the promoter, ORF and 3’ UTR of each allele of \textit{GAL3} (Figure 3-3A). Thus, our swaps encompass both regulatory and ORF changes that could mediate the different sensitivities to glucose and galactose.

We first tested here if expression of \textit{GAL3} from a heterologous locus could affect the set point of induction. We conclude that we cannot distinguish the haploid and the strain expressing the endogenous \textit{GAL3} from the heterologous locus (ttestYJM978 allele p-value=0.13 and t-test BC187 allele p-value=0.15)( Figure 3-3B.).

To quantify the effect of \textit{GAL3} we computed the difference between the set points of a strain carrying its endogenous \textit{GAL3} (e.g. \textit{GAL3} YJM978 in a YJM978 background) and
the same background strain carrying a different allele of \textit{GAL3} (e.g. \textit{GAL3} BC187 in a YJM978 background). We then express this quantity as a percent of “full conversion”, assuming that a full conversion corresponds to the difference between each strain carrying its endogenous \textit{GAL3} allele (see methods). In this manner, we observe that the YJM978 carrying \textit{GAL3} BC187 is converted \(~45\%\) and a BC187 haploid carrying YJM978 converts 57\% (Figure 3-3B). We do not however achieve a full conversion by solely swapping \textit{GAL3} alleles suggesting that other genes that we failed to identify contribute to the set point of induction.

**Figure 3-3 Allele swaps of \textit{GAL3} between BC187 and YJM978 show the contribution of \textit{GAL3} to the differences in set point**

(A) Schematic representation of the experimental setting for allele swaps. (B) Set points of induction of different allele swaps. To assess the potential effects of expression \textit{GAL3} in a heterologous locus we compared the effect of the swapping the endogenous allele to the locus yerc\(\Delta\)8. We could not reject the null hypothesis that the haploids (YJM978 and BC187) were equal to the haploids carrying a single copy of the endogenous allele in the heterologous locus yerc\(\Delta\)8 (GAL3-YJM978 (YJM978) and GAL3-BC187 (BC187)): t-test.
3.2.7  *GAL3* alleles from various natural isolates fall in at least two categories

Given the relatively large effect size of the *GAL3* allele between BC187 and YJM978 we set out to determine how other naturally occurring allelic variants of this gene from other backgrounds affect the set point of induction. To measure the contribution of the different alleles of *GAL3* isolated from the effects of different background genes between strains, we transferred different *GAL3* alleles to the same background strain YJM978 (Figure 3-4A).

We found that different allelic variants significantly change the set point of induction compared to the endogenous allele (YJM978) (Figure 3-4B). We were not able to completely phenocopy the set points of each natural isolate although an overall qualitative agreement can be observed. (Figure 3-4B) In general, we observe at least two types of behavior: *GAL3* alleles that decrease the set point of induction and make strains more sensitive to glucose repression and alleles that make the strain less sensitive to glucose repression (Figure 3-4B). For instance, strains that are highly sensitive to glucose repression (YJM421, DBVPG1106, Y9-WashU) also carry allelic variants of *GAL3* that make the background strain YJM978 more repressible by glucose. Other *GAL3* alleles with intermediate effects on the set point of induction might exist. These variants could have been missed by the resolution of our current measurements, or might not have been sampled in our experiments.

Since the different allelic variants of *GAL3* partially capture the natural variation in set points we conclude that in most strains other genes are involved in establishing the set point. Such genes could have smaller effects compared to *GAL3* or might be acting epistatically.
Figure 3-4 Different naturally occurring alleles of GAL3 affect the set point of induction in the background strain YJM978.

(A) Schematic representation of the effect of the insertion of natural variants into YJM978. (B) Effect of different allelic variants of GAL3 in the set point of induction. In red the natural allelic variants and in gray the natural isolate set point. In most instances the set point of the natural isolate and the strain YJM978 carrying the GAL3 allele do not overlap.

3.2.8 Both GAL3 and genetic background contribute to the set point across a panel of natural isolates

To further characterize the relative contribution of GAL3 versus background genes we transformed the BC187 and YJM978 alleles into 7 different natural isolate backgrounds (Figure 3-5A). Overall, we observe two groups, where despite the genetic background the GAL3- YJM978 carrying strains decreases the set point (more sensitive to repression, Figure 3-5 blue dots) and the GAL3-BC187 strains increases the set point (less sensitive, Figure 3-5 red dots) relative to the natural isolate set point alone. The relative fold change difference between the two alleles for a given strain is consistent across this panel of
isolates (~4 fold). Thus, the directionality and the magnitude of the allelic effect remained constant, suggesting little to no epistatic interactions of \textit{GAL3} and background genes.

![Diagram of experimental set up and data representation]

**Figure 3-5** Effect of 2 alleles of \textit{GAL3} across a panel of different strain backgrounds
(A) Schematic representation of the experimental set up. (B) Percent of the continuous variation that can be explained by \textit{GAL3} alleles across a panel of 7 different backgrounds.

### 3.3 Discussion

Signaling pathways play a central role in biology since they interphase between the environment and the cell. Only recently we had begun to explore how genetic variation affects signaling. Treusch et al. examined how variation in the MAPK signaling pathway impacts resistance to stress conditions (growth in caffeine and sodium chloride) between strains of \textit{S. cerevisiae}. In this study, we explore a different set of signaling pathways triggered in response to the presence of the carbohydrates galactose and glucose. Most studies that have explored variable responses to carbohydrates have surveyed macroscopic phenotypes such as growth or viability. In contrast, we focused on
a molecular phenotype: the induction of a pathway. Studying this type of trait has the potential of revealing genetic variants that are specific to the variation in signaling and do not necessarily result in large growth defects.

The \textit{GAL3} locus can explain about \textasciitilde 50\% of the variation in set points of induction between BC187 and YJM978. This suggests that other loci in the genome contribute to the differences in set point between strains. To determine the effect of other loci on the set point across strains we transferred the BC187 and YJM978 \textit{GAL3} alleles to other genetic backgrounds. In the tested strains, the effect of different alleles of \textit{GAL3} displayed the same directionality and magnitude suggesting little to no epistatic interactions. Other genes affecting the set point in the genome might therefore not be acting through \textit{GAL3} and could be instead changing other important parameters for \textit{GAL} gene induction such as the internal concentration of each carbohydrate, or targeting the much more complex regulatory networks that sense glucose. Other techniques such as targeted backcrossing mapping (Sinha et al., 2008) could help identify the missing genes of smaller effects.

Other studies have examined the variation in the induction of the \textit{GAL} pathway between \textit{S. paradoxus} and \textit{S. cerevisiae} (Peng et al., 2015). In this case, interspecies variation has been shown to be driven by the regulation of \textit{GAL80}. \textit{GAL80} encodes a negative regulator of \textit{GAL} genes that directly interacts with GAL3p. Signal transduction is therefore emerging as an important driver of the variability in the induction of the pathway within and between species. In the case of \textit{S. paradoxus} and \textit{cerevisiae} differences in the promoters of \textit{GAL80} seem to explain the differences in the induction of the pathway between species. It will be interesting to expand our study to determine if
differences in the regulatory regions, the ORF or both are the drivers of the variability in the pathway within a species.

We are only beginning to understand how genetic variation can affect signaling. Our study identified a single major locus involved in the sensing of galactose and glucose for a single quantitative trait: the set point of induction. Other aspects of the response to nutrients in this system can also be systematically characterized (e.g. heterogeneity in the response, speed, sharpness). The quantitative study of the genetic basis of these traits, all related to the output of a specific network, has the potential of identifying different genes that independently tune different aspects of a response. This approach might uncover novel design principles of genetic networks.

3.4 Methods

3.4.1 Strains and media

Strains used in this study can be found in supplementary Table C-1 and plasmids used in this study are shown in Supplementary Table C-2. Natural isolate yeast strains were obtained from multiple sources. 23 strains were part of the Saccharomyces Genome Resequencing Project and were obtained from the National Collection of Yeast Cultures (Liti et al., 2009); 18 strains were obtained from the Fay lab at Washington University (Cromie et al., 2013). Strain Bb32 was obtained from the Broad Institute (Mortimer et al., 1994).

All experiments were performed in synthetic minimal media, which contains 1.7 g/L yeast nitrogen base (YNB) (BD Difco) and 5 g/l ammonium sulfate (EMD Millipore), plus carbon sources. In glucose titration experiments, the synthetic minimal medium base
was supplemented with a constant concentration of 0.125% galactose and a two-fold titration of glucose starting at 1.0% (1.0%, 0.50%, 0.25%, 0.13%, 0.063%, 0.031%, 0.016%, 0.008%, 0.004%, 0.002%, 0%). Cultures were grown in a humidified incubator (Infor Multitron) at 30°C with rotary shaking at 230 rpm (tubes and flasks) or 999 rpm (deep 96-well plates).

*GAL3* deletions were made in haploid strains deleting 950 bp upstream of *GAL3* and 950 bp downstream of the ORF.

Plasmids containing the *GAL3* ORF as well as 890 bp upstream and 900 bp downstream were generated.

### 3.4.2 Flow cytometry and sugar titrations

We assayed the gene expression of the natural isolates and constructed strains in a glucose titration by inoculating from single colonies into liquid YPD, incubating for 16 h, mixing 1:5 (reference:query) by volume and then dilution 1:5000-1:2500 into 2% raffinose and growing for 14-16 h to ~ (0.2-0.5 OD). The raffinose cultures were pelleted by centrifugation, washed once, and then resuspended in synthetic media without carbon to an OD of 0.10-0.250 and diluted 1:125 in the glucose gradient. Cultures were incubated in 96 deep well plates at 30°C with shaking for 8 h. After the incubation, cultures were pelleted by centrifugation, washed once and resuspended in TE (pH 7.5). Plates were read immediately on a Stratedigm S1000EX cytometer.

Flow cytometry data were analyzed using custom MATLAB code. All experiments were co-cultured and a 2-D Gaussian mixture model (using the gmdistribution class MATLAB) using a fluorescent channel (mCherry or BFP) and side scatter channel (SSC)
were used for segmentation. *GAL1*pr-YFP expression for each segmented population was collected and a mean value for each concentration of sugars was computed. Cells were then measured by flow cytometry. Mean YFP values were computed and normalized to the maximum and minimum levels of expression of each strain to compute normalized YFP values for both query and reference strain. The normalized set point $S_N$ is the log$_2$ ratio between the query and reference strain set points (Figure 2-1D).

### 3.4.3 Meiotic segregants phenotyping

Strains were sporulated and random spore segregation was used to isolate haploids. Haploids of BC187 and YJM978 were mated and a single hybrid diploid was sporulated. Random spore segregation was used to isolate meiotic segregants and a robotic colony picker was used for transfer of colonies from solid to 96 well plate format.

The resulting haploids were phenotyped for their sensitivity to glucose and galactose mixtures. Haploid segregants were directly compared to their parental phenotypes by growing the parental strain and the haploid segregants in the same media. The parents were segmented using their fluorescent constitutive promoter BFP or mCherry, the haploid strain was identified by its lack of fluorescence. Haploid segregants were classified as BC187-like strains or YJM978-like strains depending on their induction distribution at 8 hours after inoculation.

### 3.4.4 DNA isolation and genomic library preparation

Genomic DNA was extracted from parental strains BC187, YJM978 and meiotic segregants using the QIAGEN genomic extraction kit. Libraries were made with the nextera kit. The libraries were QC by qPCR. Sequencing of libraries was performed in a
HiSeq 2000 (100-bp paired-end reads)

### 3.4.5 Bioinformatic analysis

Sequences for the SGRP strains were downloaded from SGRP website and sequences for the Wash-U strains were derived by local BLAST and manual assembly of contig hits in each strain.

Sequences can be downloaded from:

[https://www.sanger.ac.uk/research/projects/genomeinformatics/sgrp.html](https://www.sanger.ac.uk/research/projects/genomeinformatics/sgrp.html)

Strain phylogeny was constructed using Phylomatch (Ramazzotti et al., 2012)
3.5 References


Kundu, S., and Peterson, C.L. (2010). Dominant role for signal transduction in the


Skelly, D.A., Merrihew, G.E., Riffle, M., Connelly, C.F., Kerr, E.O., Johansson, M.,


Author contributions

R.E.-C., and M.S. designed the study; R.E.-C. and K.L. performed the experiments; R.E.-C., K.L., and M.S. analyzed the data; R.E.-C. and M.S. wrote this manuscript.
Chapter 4: A Chromatin-Based Mechanism for Limiting Divergent Noncoding Transcription

Sebastian Marquardt, Renan Escalante-Chong, Nam Pho, Jue Wang, L. Stirling Churchman, Michael Springer, and Stephen Buratowski


In addition to their annotated transcript, many eukaryotic mRNA promoters produce divergent noncoding transcripts. To define determinants of divergent promoter directionality, we used genomic replacement experiments. Sequences within non-coding transcripts specified their degradation pathways, and functional protein-coding transcripts could be produced in the divergent direction. To screen for mutants affecting the ratio of transcription in each direction, a bidirectional fluorescent protein reporter construct was introduced into the yeast nonessential gene deletion collection. We identified chromatin assembly as an important regulator of divergent transcription. Mutations in the CAF-I complex caused genome-wide derepression of nascent divergent noncoding transcription. In opposition to the CAF-I chromatin assembly pathway, H3K56 hyperacetylation, together with the nucleosome remodeler SWI/SNF, facilitated divergent transcription by promoting rapid nucleosome turnover. We propose that these chromatin-mediated effects control divergent transcription initiation, complementing downstream pathways linked to early termination and degradation of the noncoding RNAs.
4.1 Introduction

Divergent transcription from RNA polymerase II (Pol II) promoters is seen in many eukaryotic organisms (Wei et al., 2011). The majority of yeast noncoding RNAs (ncRNA) (Neil et al., 2009) and at least 60% of human and murine long ncRNAs found in embryonic stem cells originate from bidirectional transcription (Sigova et al., 2013). In yeast, initiation typically occurs near the boundaries of nucleosome-free regions (NFRs) that characterize promoters (Xu et al., 2009). Two distinct transcription preinitiation complexes (PICs) are observed, so divergent transcription likely results from two separate basal promoters rather than a single truly bidirectional element (Rhee and Pugh, 2012). Whereas the two PICs might compete for factors, divergent long ncRNA (IncRNA)/mRNA pairs in human embryonic stem cells (Sigova et al., 2013) and yeast (Xu et al., 2009) show coordinated expression changes. No clear function for divergent ncRNAs has yet been identified.

Generally, steady-state levels of divergent ncRNAs are low relative to coding transcripts. This is at least partly due to weaker transcription in the noncoding direction (Churchman and Weissman, 2011; Rhee and Pugh, 2012). Promoter elements can also influence transcript decay, perhaps contributing to differential stability of the divergent transcripts (Bregman et al., 2011). Additionally, ncRNAs are susceptible to early termination linked to rapid degradation. In humans, promoter-proximal polyA sites (PAS) within divergent ncRNAs trigger termination and subsequent degradation by the nuclear exosome (Almada et al., 2013; Flynn et al., 2011; Ntini et al., 2013; Preker et al., 2009). In mRNAs, early PASs are suppressed by U1 small nuclear ribonucleoprotein (snRNP) recognition sites. These RNA motifs show a biased distribution, with a higher frequency
of PASs and fewer U1 sites in the noncoding direction (Almada et al., 2013; Ntini et al., 2013).

In yeast, the helicase Sen1 and RNA-binding proteins Nrd1 and Nab3 (the NNS system) terminate small noncoding RNAs and cryptic unstable transcripts (CUTs) (Mischo and Proudfoot, 2013; Wei et al., 2011). NNS recruits the nuclear exosome for 3' trimming of small nucleolar RNAs (snoRNAs) or degradation of CUTs (Vasiljeva and Buratowski, 2006; Wyers et al., 2005). Many divergent ncRNAs are CUTs, and Nrd1 and Nab3 sites are enriched in the ncRNA direction but depleted in the mRNA direction, so mutating NNS factors increases divergent ncRNA levels (Schulz et al., 2013). A second class of ncRNAs produced by divergent transcription, the stable uncharacterized transcripts (SUTs), are also partially affected by nuclear exosome but are mostly degraded by cytoplasmic decapping and the 5' to 3' exonuclease Xrn1 (Marquardt et al., 2011; van Dijk et al., 2011; Xu et al., 2009). The short length of divergent CUTs and SUTs might suggest that productive elongation is only possible in the coding direction of a divergent promoter (Seila et al., 2008). However, 3' extended SUT and CUT transcripts are detected when the early termination pathway is compromised, indicating elongation in the noncoding direction is not intrinsically limited but instead subject to early termination (Marquardt et al., 2011). Finally, gene looping of promoter and terminator sequences via the Ssu72 phosphatase has been proposed to favor transcription in the coding direction (Tan-Wong et al., 2012). However, it should be noted that Ssu72 is also necessary for the yeast NNS pathway (Steinmetz and Brow, 2003).

Failure to efficiently terminate divergent transcription can produce extended ncRNAs that overlap the neighboring gene. Yeast noncoding transcription can be inhibited by
chromatin pathways within coding gene boundaries (Cheung et al., 2008; Churchman and Weissman, 2011; Keogh et al., 2005; Venkatesh et al., 2012; Whitehouse et al., 2007). In particular, methylation of H3K36 by Pol II-associated Set2 targets the Rpd3S histone deacetylase (HDAC) complex, which inhibits both initiation and elongation in downstream regions of protein-coding genes (Carrozza et al., 2005; Keogh et al., 2005). This pathway also suppresses histone exchange associated with incorporation of H3K56 acetylated histones (Venkatesh et al., 2012).

Here, we first establish that functional protein-coding transcripts can be produced from the noncoding direction of a divergent promoter. Therefore, termination and RNA decay pathways are specified by transcript sequences rather than the promoter. Exploiting this fact to create a high-throughput genetic screen based on fluorescent reporters, we identified regulators of transcriptional directionality. The top-ranking class of mutants affects chromatin assembly and turnover. In particular, divergent noncoding transcription is repressed genome wide by the chromatin assembly factor I (CAF-I) and associated factors that incorporate H3K56ac-marked histones. Failure to incorporate histones, or increased nucleosome turnover caused by elevated H3K56ac, increases divergent noncoding transcription. In contrast, loss of the chromatin remodeler SWI/SNF blocks these H3K56ac-mediated effects and decreases divergent ncRNA transcription.

4.2 Results

4.2.1 Divergent Transcription Is Independent of Transcript Identity

PPT1 and SUT129 transcripts are divergently transcribed from a shared promoter region (Marquardt et al., 2011; Xu et al., 2009; Figure 4-1A). Northern blot analysis shows that
levels of both transcripts decrease in galactose and increase in glucose media (Figure 4-1B and Figure C-1). These effects do not reflect increased RNA stability, as similar results were obtained in a mutant lacking the cytoplasmic Xrn1 pathway (Figure C-1B and Figure C-1C). Yeast lncRNA species are distinguished by their RNA decay pathway, with CUTs primarily degraded by the nuclear exosome and SUTs by Xrn1. To determine whether the RNA decay pathway is specified by the promoter or the transcribed region, we conducted "body-swap" experiments. The transcribed region of CUT60, normally expressed divergently from the MED2 promoter (Marquardt et al., 2011; Xu et al., 2009), was used to replace the SUT129 transcribed region (sut129D::CUT60). To allow detection of the ectopic Cut60 transcript, the endogenous CUT60 locus was deleted (cut60-D::URA3) and transcripts were analyzed by northern blotting (Figure 4-1C). As expected, both Sut129 and a Sut477 control were increased in xrn1D. In contrast, Cut60 accumulated in rrp6D, even when under control of the SUT129 promoter and 3' sequences. Therefore, the choice of decay pathway for lncRNAs produced by the divergent PPT1-SUT129 promoter is specified by the transcript body.

Figure 4-1 Characterization of Divergent Transcription at the PPT1 Promoter
A) Schematic of divergent transcription showing noncoding (black) and mRNA (white) transcripts. Xrn1 and Rrp6 RNA represent the primary decay pathways for SUT or CUT ncRNA species, respectively.
(B) Northern blot analysis of PPT1 and SUT129 transcripts in cells grown in raffinose (Raf) media, then galactose (Gal) media for 60 or 120 min, and finally glucose (Glu) media for 6, 9, or 20 min. Wild-type (left) and PPT1/SUT129 promoter inversion strains (right) were analyzed; asterisk denotes the ScR1 loading control. See Figure 5-16 for additional information.
(C) The ncRNA "body" specifies its RNA decay pathway. Selected ncRNA transcripts (left) produced in the indicated RNA decay pathway mutants (top) were analyzed by northern blot. cut60D::URA3 replaces the CUT60 locus with URA3, whereas sut129D::CUT60 replaces the Sut129 transcribed region with that of Cut60. ScR1 is a loading control.
(D) A functional protein-coding transcript can be expressed in the noncoding direction. sut129D::URA3 replaces the SUT129 transcribed region with the URA3 open reading frame.
frame and 3’ region. RNA was analyzed as in (B; left). Spot growth assay (right) shows 3-fold serial dilutions of indicated strains tested for growth on media lacking uracil (SC-URA).

**A**

![Diagram showing divergent transcription]

**B**

<table>
<thead>
<tr>
<th></th>
<th>wild-type</th>
<th>PPT1/SUT129 promoter inversion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PPT1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SUT129</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Xrn1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Rrp6</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**C**

<table>
<thead>
<tr>
<th></th>
<th>cut60Δ::URA3</th>
<th>cut60Δ::CUT60</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>wild-type</td>
<td>wild-type</td>
</tr>
<tr>
<td></td>
<td>xrn1Δ</td>
<td>xrn1Δ</td>
</tr>
<tr>
<td></td>
<td>rrp6Δ</td>
<td>rrp6Δ</td>
</tr>
</tbody>
</table>

**D**

<table>
<thead>
<tr>
<th></th>
<th>Δsut129::URA3</th>
<th>wild-type</th>
<th>SC-URA</th>
<th>YPD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Δsut129::URA3</td>
<td>wild-type</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>wild-type</td>
<td>wild-type</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>wild-type</td>
<td>wild-type</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>wild-type</td>
<td>wild-type</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We next asked whether a protein-coding mRNA could be expressed in the ‘‘noncoding’’ direction of a divergent promoter. The PPT1/SUT129 promoter was inverted so that PPT1 mRNA is produced by the SUT129 promoter and vice versa. Although higher levels of Sut129 were observed, response of both transcripts to changing carbon source was maintained (Figure 4-1B and Figure C-1C). As a further test, the SUT129 transcribed region was replaced with the URA3 open reading frame (sut129D::URA3). The response to carbon source was intact in both directions, showing that promoter function was not disrupted (Figure 4-1D). Importantly, sut129D::URA3 supports growth in media lacking uracil. Therefore, functional protein-coding mRNAs can be expressed from the noncoding direction of a divergent promoter.

**4.2.2 CAF-I Affects Divergent Promoter Directionality**

Given the independence between transcript identity and promoter, a genetic screen for genes affecting divergent promoter directionality was designed using a fluorescent protein reporter (FPR) strain (Figure 4-2 A). The PPT1 open reading frame was replaced with mCherry sequences, and yellow fluorescent protein (YFP) was expressed from the SUT129 promoter, allowing quantitation of expression in both directions by flow cytometry. A library of roughly 4,500 viable gene deletions was combined with the FPR using synthetic genetic array (SGA) technology (Figure 4-2 B; Tong et al., 2001). The median filtered mCherry and YFP values for each strain were plotted against each other to visualize PPT1/SUT129 promoter ‘‘correlation’’ and ‘‘directionality’’ (Figure 4-2 C and C-2A and D). Whereas any mutant can cause a decrease in overall transcription, there may also be promoter-specific effects on directionality. The mCherry/YFP (or coding/ncRNA) ratio normalizes for any general effects and will therefore be used for
figures in this paper (absolute values for each direction are shown in supplemental figures). An overall positive correlation between mCherry and YFP was seen with most mutants (Figure 4-2 C and C-2C), indicating that their effects were equal in both directions. In contrast, some mutants significantly deviated from the correlation line to suggest changes in directionality (Figure C-2D).

**Figure 4-2 A Genetic Screen Identifies CAF-I as a Regulator of PPT1/ SUT129 Promoter Directionality**

(A) Schematic representation of the fluorescent protein reporter (FPR) locus. (B) FPR is crossed into the library of viable KanMx-marked deletion mutants using SGA technology and YFP and mCherry fluorescence determined by flow cytometry. (C) Histogram of YFP versus mCherry levels, with individual distributions shown at each axis. Each point represents a deletion mutant, with red marking those with GO annotations for nucleosome assembly or chromatin assembly. One diagonal measures correlation between mCherry and YFP, whereas the other provides a measure of directionality. See Figure C-2 for related information.
Whereas no Gene Ontology (GO) terms were enriched in the list of mutants with increased mCherry/YFP ratio, genes associated with ‘‘nucleosome assembly’’ and ‘‘chromatin assembly’’ were significantly enriched in the top 250 mutants having
reduced mCherry/YFP (p values of $3 \times 10^{-6}$ and $9 \times 10^{-5}$, respectively; Figure 4-2 C and Figure C-3E). These include multiple histone genes (HHT1, HTA1, HTA2, and HTB2) and subunits of the HIR and CAF-I chromatin assembly complexes (Figure C-3A and Figure C-3B). The effect of CAF-I was confirmed in two ways. First, the FPR was transformed de novo into strains deleted for each of the three CAF-I subunits: CAC1/RLF2/YPR018W, CAC2/YML102W, and CAC3/MSI1/YBR195C (Figure 4-2 D, Figure C-3C, and Figure C-3G). Conversely, the CAF-I deletion alleles were transformed into the FPR parent strain (Figure C-3D). As an internal control, cacD/FPR cells were cocultured with CAF-I wild-type cells expressing both the FPR reporter and blue fluorescent protein (BFP). The BFP strain was used to normalize mCherry and YFP values (Figure 4-2 D and Figure C-3F). All three CAF-I sub- unit deletions consistently reduced the mCherry/YFP ratio (Figure 4-2D, Figure C-3B, Figure C-3C, Figure C-3F, and Figure C-3G). Quantitative PCR (qPCR) analysis showed that these directionality changes resulted from altered RNA expression (Figure C-3E).

To test whether CAF-I depletion affects divergent transcription from the endogenous PPT1/SUT129 locus, transcripts from cacD strains lacking the FPR were analyzed. All three deletions decreased the coding/ncRNA ratio by increasing Sut129 and slightly reducing PPT1 expression (Figure 4-3A and Figure C-3H). As RNA levels reflect both transcription and degradation, Pol II was analyzed by chromatin immunoprecipitation (ChIP) of Rpb3. In cells lacking CAF-I, Pol II ChIP increased at SUT129 and reduced at PPT1 (Figure 4-3B and Figure C-3 I). Therefore, CAF-I changes the directionality of transcription for the endogenous PPT1/SUT129 locus as well as the FRP reporter.
Figure 4-3 CAF-I Represses Transcription in the Noncoding Direction at Endogenous PPT1/SUT129

(A) Northern blot analysis of PPT1 and SUT129 in wild-type and cacD cells (left). ScR1 transcript serves as a loading control. Blots were quantitated by phosphorimager, and the mean PPT1/SUT129 ratio of six biological repeats is graphed ± SEM (individual values are shown in Figure C-3H).

(B) RNA Pol II ChIP at PPT1/SUT129 in CAF-I mutants. Rpb3 ChIP was performed in the indicated genotypes and analyzed by qPCR using primer locations shown in schematic at top. Quantification of PPT1/SUT129 ratio is shown below normalized to wild-type. Means ± SEM of four biological repeats are shown. Asterisks denote statistical significance of means from wild-type with p < 0.05. See also Figure C-3
4.2.3 CAF-I Represses Divergent Noncoding Transcription Genome Wide

To determine whether the CAF-I effect was specific to PPT1/ SUT129 or more general, transcription elongation complexes were mapped by native elongating transcript sequencing (NET-seq) (Churchman and Weissman, 2011) in cells lacking or containing CAF-I (Figure 4-4). This technique allows a direct measure of transcription, independent of RNA or protein stability. Excellent reproducibility was seen between deletion mutants of the three individual CAF-I subunits and their two biological replicates (Figure C-4B). Metagene analysis of all yeast-coding genes showed increased Pol II density in the divergent antisense direction but little change in the sense direction upon CAF-I deletion (Figure 4A).

Directionality at individual promoters was calculated as a ratio of coding direction to divergent noncoding sequence read counts (see Experimental Procedures). A large number of promoters showed a reduced coding/noncoding ratio in CAF- I deletion mutants, whereas a much smaller number showed an increase (Figure 4-4B, Figure C-4C, and Figure C-4E). CAF-I-affected promoters (1,234 in total) were conservatively defined as those where the directionality score is reduced by at least 1.5-fold (red line in Figure 4-4 B, Figure C-4C, and Figure C-4E) in at least four of the six NET-seq experiments. Interestingly, the greatest ratio changes upon CAF-I deletion (Figure 4-4B, Figure C-4C, and Figure C-4E, y axis) are seen at promoters with the largest coding/noncoding ratio (x axis), arguing that CAF-I is an essential factor in this differential. To determine whether the reduced coding/noncoding ratio in CAF-I deletions reflects decreased coding transcription, increased noncoding transcription, or a combination, sequence reads were normalized using the reads per kilobase per million (RPKM) method and plotted in
coding and noncoding directions separately. Little difference in coding reads was seen between CAF-I affected or unaffected genes (red and gray circles, respectively, in left panels of Figure 4-4C, Figure C-4D, and Figure C-4F). In contrast, noncoding reads were strikingly elevated upon deletion of CAF-I subunits (right panels). Increased noncoding direction transcription was responsible for the directionality effect over a wide range of threshold values (Figure 4-4 D).

Of several hundred previously identified CUTs or SUTs at divergent promoters (Xu et al., 2009), similar percentages (35%–40%) were affected by CAF-I (Figure C-5A). We note that much of the divergent transcription revealed by NET-seq at CAF-I-affected promoters was previously unannotated. Therefore, the CAF-I effect is not specific to a particular class of ncRNA. We also analyzed data for divergent coding pairs. Of 659 coding/coding pairs with no obvious additional ncRNA (Xu et al., 2009), only seven show a directionality change that meets the criteria used for ncRNAs (Figure C-5B). For these, loss of CAF-I preferentially increased transcription of the divergent open reading frame (ORF) with lower read density.

Interestingly, the NET-seq directionality change for PPT1- Sut129 upon CAF-I deletion was below the 1.5-fold cutoff, consistent with the relatively modest changes seen by FPR and northern blot (Figure C-4A). The genome-wide NET-seq showed that the vast majority of bidirectional promoters have much stronger responses. For subsequent figures, we analyze two more representative loci: MUB1 as a “moderate” responder and KIP1 as a “strong” responder. The NET-seq directionality change at MUB1 is close to the 1.5-fold cutoff, whereas KIP1 promoter shows a roughly 5-fold effect, both caused by an increase in divergent transcription (Figure 4-4 E and F). Northern blotting showed a
corresponding increase in MUB1 divergent (MUB1div) and KIP1 divergent (KIP1div)
transcripts upon deletion of CAF-I (Figure 4-4G, H, Figure C-4G, and Figure C-4H).
MUB1div overlaps and likely corresponds to the previously annotated Sut730 (Xu et al.,
2009). No CUT or SUT was previously annotated for KIP1div, yet at least three distinct
transcripts were observed, likely resulting from alternative termination (Figure 4-4B).
Figure 4-4 NET-Seq Analysis Reveals a Widespread Effect of CAF-I on Promoter Directionality

(A) Metagene analysis. Annotated coding genes were aligned by transcription start site (TSS), and scaled NET-seq reads were mapped in the sense (+strand) or antisense (-strand) directions. (B) NET-seq coding/noncoding ratios in cac2Δ versus wild-type cells. For each gene, directionality was calculated as the ratio of coding to divergent noncoding reads (the mean of two biological NET-seq replicates). The CAF-I effect was measured as the ratio of directionality in the mutant divided by that in wild-type cells, and this was plotted against the wild-type directionality using log2 scale. Changes in the mutant will appear as deviations from the zero value on the y axis. A 1.5-fold cutoff marking “CAF-I affected” loci is marked by dashed red line. Red dots mark MUB1 and KIP1.

(C) CAF-I primarily suppresses divergent ncRNA transcription. Normalized reads (RPKM; mean of two biological replicates) for each coding gene (left) and corresponding divergent direction (right) were calculated. Values for cac2Δ versus wild-type were plotted, with CAF-I-affected loci in red. Inset histogram shows data distributions. (D) CAF-I suppresses divergent transcription at varying cutoff thresholds. Plots as shown in (C) were created at different threshold numbers. Slopes of best-fit lines for CAF-I-affected (red lines) and unaffected (black lines) loci were calculated for both coding (left) and divergent noncoding (right) NET-seq data.

(E) NET-seq for MUB1. Vertical bars represent NET-seq reads at that position, showing nascent transcription on the Watson strand in cac2Δ (green) and wild-type (gray) on the Crick strand in cac2Δ (blue) and wild-type (pink). MUB1 ORF is in white and the divergent MUB1div transcript in black. Red bar shows location of the northern probe for MUB1div. Mean RPKM values for MUB1 and MUB1div are graphed in right panel ± SEM; asterisks denote differences with p < 0.05 compared to wild-type.

(F) NET-seq data for KIP1; annotations as in (E). (G) Northern blot analysis of MUB1 and MUB1div in wild-type and cacD cells (left). ScR1 transcript controls loading. Blots were quantitated by phosphorimager and the MUB1/MUB1div ratio is graphed (±SEM; n = 3). Asterisks denote statistical significance with p < 0.05. (H) Northern blot analysis of KIP1 and KIP1div, as in (G). Asterisk (*) indicates the expected size of KIP1div based on the NET-seq analysis, and this band was used for quantitation, but longer species may represent readthrough of KIP1div into the downstream gene. See Figure C-4 and Figure C-5 for additional related data.
4.2.4 CAF-I Repression of Divergent Transcription Is Distinct from Early Termination and RNA Degradation Pathways

Divergent transcript elongation can be restricted by early termination. In agreement, we observed that mutations in four different NNS factors (Sen1, Nrd1, Nab3, and Ssu72) increased divergent ncRNA levels from MUB1div (Figure 4-5A). To determine functional connections between early termination and CAF-I, divergent transcription was analyzed in double mutants. Combining cac2Δ with sen1-1, nab3-11, or ssu72-2 increases MUB1div levels compared to the single mutations (Figure 4-5B and Figure C-5D). KIP1div was not affected by NNS pathway mutants (Figure C-5 C and Figure C-5D). Combining RNA degradation mutants (rrp6D or xrn1D) with a CAF-I deletion led to stronger accumulation of divergent ncRNA at both MUB1 and KIP1 (Figure 4-5C, Figure C-5E, and Figure C-5G). In contrast, when two different CAF-I subunit deletions are combined (cac1Δ cac2Δ), the effects are indistinguishable from the single mutants (Figure C-5F). Thus, CAF-I functions upstream of, or parallel to, early termination and RNA degradation in suppressing divergent transcription.
Figure 4-5 CAF-I Repression of Divergent Transcription Is Independent of NNS Termination or RNA Decay

(A) Northern blotting of MUB1 divergent transcription in NNS pathway mutants. Transcripts (left) and genotypes (top) are indicated. MUB1/MUB1div (gray) is quantitated below gel images as fold change compared to the corresponding isogenic wild-type strain. Means of three biological repeats are shown ± SEM; asterisks indicate statistically significant differences from wild-type or as indicated by brackets with p < 0.05. (B) Northern blotting in NNS pathway and cac2Δ double mutants. See (A) for annotations.
(C) Northern blotting in RNA degradation and cac2Δ double mutants. See (A) for annotations. See also Figure C-5.
The H3K56ac-Linked Nucleosome Incorporation Pathway Enforces Promoter Directionality
CAF-I incorporates newly synthesized histones on newly replicated or repaired DNA (Gaillard et al., 1996; Smith and Stillman, 1989). To test whether the CAF-I effect depends on DNA replication, divergent transcription was analyzed in yeast cells arrested in G1 with a-factor for 2 hr. Both the FPR reporter (Figure 4-A) and endogenous divergent KIP1 and MUB1 loci (Figure 4-6 B and Figure C-6A) were still affected, arguing that CAF-I enforcement of promoter directionality is not S phase specific.

Figure 4-6 Characterization of CAF-I Repression of Divergent Transcription
(A) Repression is not S phase specific. Wild-type and cacD cells were assayed for the FPR-based mCherry/YFP ratio before (cycling) and during G1 arrest with a-factor (a-T0). Data are normalized to the wild-type mCherry/YFP ratio; means ± SEM of four repeats are shown. (B) Northern blot of divergent transcription at KIP1 before or during G1 arrest by a-factor in wild-type (BY4741) and cac2Δ cells. (C) H3K56ac-linked nucleosome assembly represses KIP1div. Northern blot analysis of indicated transcripts (left) and genotypes (top). Mean quantification relative to the ScR1 loading control of three biological repeats is shown below ± SEM. (A and C) Asterisks denote significant differences from wild-type or as indicated by brackets with p < 0.05. See also Figure C-6.
Figure 4-6 (Continued)
Histone H3 deposited during replication by CAF-I is acetylated at lysine 56 (H3K56ac) (Li et al., 2008). We asked if divergent transcription repression involves other components of the H3K56ac-linked nucleosome incorporation pathway. Rtt106 functions redundantly with CAF-I in histone incorporation (Li et al., 2008). Accordingly, rtt106Δ has increased MUB1div (Figure C-6B). Combining rtt106Δ with CAF-I deletion further increased divergent transcription of both MUB1div and KIP1div, consistent with functional redundancy (Figure 4-6C, Figure C-6B, and Figure C-6C). Rtt109, together with Vps75, is the H3K56 acetyltransferase (Han et al., 2007), and deletion of either of these factors decreases the mCherry/ YFP ratio from the FPR. Finally, a nonacetylable H3K56A substitution also increased divergent transcripts and was additive with CAF-I deletion (Figure 4-6C Figure C-6B, and Figure C-6C). Thus, it appears multiple components of the H3K56ac-linked nucleosome incorporation pathway are important for repression. H3K56ac-Linked Nucleosome Remodeling by SWI/SNF Contributes to Promoter Directionality In addition to promoting replication-dependent incorporation, H3K56ac has more recently been shown to mark nucleosomes for exchange by one or more chromatin remodelers (Kaplan et al., 2008; Smolle et al., 2012; Watanabe et al., 2013). The functionally redundant Hst3 and Hst4 HDACs deacetylate H3K56 to repress this exchange (Celic et al., 2006; Kaplan et al., 2008). As CAF-I deletion reduces H3K56ac incorporation, we asked whether derepression of divergent transcription in cac2Δ cells would be suppressed by elevating H3K56ac with an hst3/hst4Δ double deletion or by an H3K56Q acetylation mimic mutant (Dai et al., 2008). Surprisingly, combining cac2Δ with these high- K56ac mutants further increased KIP1div and
MUB1\textsuperscript{div}, and hst3/4\textsuperscript{Δ} or H3K56Q mutants by themselves had elevated levels of divergent transcripts (Figure 4-6A, Figure C-6, and Figure -7A). Therefore, both incorporation of H3K56ac and its subsequent deacetylation may be critical for suppressing divergent transcription. This idea is supported by metagene analysis of ChIP data from Venkatesh et al. (2012) measuring H3K56ac and histone exchange (calculated as the ratio of an induced H3-flag to a constitutive H3- myc histone). Histone exchange and H3K56ac are highest not at the TSS or +1 nucleosome but 250–500 nt upstream where divergent transcription initiation and the -1 nucleosome would be (Figure 4-6B and Figure C-7C–Figure C-7F). Importantly, CAF-I-responsive promoters (as defined in Figure 4-4) have higher levels of histone exchange and H3K56ac in this region than nonaffected promoters (Figure 4-6B and Figure C-7C and F). Similarly, histone exchange and H3K56ac are higher when the NET-seq signal is stronger in the coding direction (Figure C-7G and Figure C-7H). These data suggest more rapid exchange of H3K56-acetylated nucleosomes on the less-preferred side of divergently transcribed promoters.

If elevated H3K56ac increases divergent transcription via a chromatin-remodeling complex, deletion of this factor should suppress the effect. To test this idea, deletions of candidate remodelers (Clapier and Cairns, 2009) were introduced into an hst3/hst4\textsuperscript{D} strain. Remarkably, the increases of KIP1\textsuperscript{div} and MUB1\textsuperscript{div} caused by hst3/hst4\textsuperscript{D} were reversed upon deletion of SNF5, a subunit of the SWI/SNF complex (Figure 4-6C, Figure I, and Figure C-7J). The MUB1 and KIP1 mRNAs were unaffected. Similarly, snf5\textsuperscript{Δ} also suppressed elevated divergent transcription in H3K56Q (Figure 4-6D and Figure C-7B) and cac1\textsuperscript{Δ} (Figure 4-6E and Figure C-7B) cells. Deletion of Swi/Snf subunit Snf11 strongly increased the mCherry/YFP ratio in the original FPR screen, as did fresh
transformation of snf5Δ into the FPR strain (Figure C7-K). In contrast, other remodeler deletions tested did not suppress increased divergent transcription caused by hst3/hst4D (Figure C7-I and Figure C7-J). Deletion of CHD1 or RSC1 had no effect on divergent transcripts, and deletion of SWR1 had little or no effect on MUB1div and KIP1div despite scoring in the original FPR screen. In contrast, deletion of RSC2, ISW2, or ARP8 (a component of the Ino80 complex) increased levels of MUB1div and KIP1div alone or additively with hst3/hst4D. Together, these results suggest a specific functional antagonism between the H3K56ac/CAF-I pathway and Swi/Snf-dependent nucleosome remodeling in limiting divergent noncoding transcription.
Figure 4-7 Elevated H3K56ac and SWI/SNF Promote Divergent Transcription

(A) Northern blotting of MUB1 divergent transcripts in indicated mutants (top). Samples were run on a single gel; gaps are where irrelevant lanes were removed. MUB1/MUB1div (gray) is quantified below gel images as fold change compared to the corresponding isogenic wild-type strain. Means of three biological repeats are shown ± SEM; asterisks indicate significant differences from wild-type or as indicated by brackets with p < 0.05.

(B) CAF-I-affected genes have higher H3K56ac and nucleosome exchange upstream of TSS. ChIP data from Venkatesh et al. (2012) were organized into ten bins as described in Experimental Procedures, and averaged values for CAF-I-affected genes (red; 1,234 genes as defined in Figure 4-4) and unaffected genes (gray) were plotted. Top panel shows histone turnover as measured by the ratio of inducible flag-H3 to constitutive myc-H3 (individual values in Figure C-7C and Figure C-7D). Bottom panel shows H3K56ac/H3 ratio (individual values in Figure C-7E and Figure C-7F). TSS marks transcription start site and TE the polyA site. Shading around line shows ± SEM; asterisks mark significant differences with p < 0.05.

(C) Deletion of SNF5 suppresses increased divergent transcription caused by hst3/hst4D. Northern blot analysis of MUB1 divergent transcripts in indicated strains (top). See Figure C-7B for similar analysis of KIP1.

(D) Deletion of SNF5 suppresses increased divergent transcription caused by H3K56Q. Northern blot analysis was performed as in (C). (E) Deletion of SNF5 suppresses increased divergent transcription caused by cac1Δ. Northern blot analysis was performed as in (C).

(F) Regulation of divergent transcription by -1 nucleosome dynamics. CAF-I incorporates an H3K56ac nucleosome that is deacetylated by Hst3/Hst4. Failure to deacetylate promotes remodeling by Swi/Snf. Loss of CAF-I or Hst3/Hst4 decreases nucleosome occupancy, leading to derepression. Conversely, loss of Swi/Snf decreases divergent transcription. See also Figure C-6.
Figure 4-7 (Continued)
4.3 Discussion

Bidirectional transcription from RNA Pol II promoters is widespread but poorly understood. As a major source of noncoding transcription, understanding its regulation is important. Divergent ncRNA levels can be suppressed by early termination and rapid degradation (Almada et al., 2013; Flynn et al., 2011; Ntini et al., 2013; Preker et al., 2009; Schulz et al., 2013). However, these postinitiation events cannot explain the asymmetry of RNA Pol II and transcription-coupled histone modifications that are preferentially seen in the coding gene direction. Finding that functional mRNAs can be produced from the noncoding direction of divergent promoters, we designed a genetic screen for effectors of promoter directionality based on fluorescent protein reporters, thus avoiding effects of transcript sequences on termination or RNA decay. We find that CAF-I, along with other factors affecting chromatin assembly and dynamics, represses divergent transcription at a large fraction of mRNA promoters.

CAF-I places histones onto newly synthesized DNA after replication or repair (Kaufman et al., 1997; Smith and Stillman, 1989). The partially redundant assembly factors CAF-I and Rtt106 preferentially bind newly synthesized H3/H4 tetramers that have been acetylated by Rtt109 at H3K56 (Han et al., 2007; Huang et al., 2005; Li et al., 2008). Once incorporated, this acetylation is removed by Hst3 and Hst4, and failure to do so leads to DNA damage and genome instability (Celic et al., 2006; Kaplan et al., 2008). This pathway is also needed for transcriptionally repressive yeast heterochromatin (Huang et al., 2005; Kaufman et al., 1997), perhaps because incomplete nucleosome assembly impedes chromatin compaction. Interestingly, outside of S phase, H3K56ac is enriched at the rapidly exchanging nucleosomes flanking NFRs, suggesting a replication-
independent role for K56ac (Dion et al., 2007; Kaplan et al., 2008; Rufiange et al., 2007). Despite the promoter enrichment, studies of H3K56ac and Rtt106 suggest functions in transcription elongation and repression of cryptic initiation within transcribed regions (Imbeault et al., 2008; Venkatesh et al., 2012). More specifically, the Set2- H3K36me3-Rpd3S deacetylation pathway, which suppresses gene-internal initiation, acts in opposition to H3K56ac-mediated nucleosome exchange (Venkatesh et al., 2012).

Here, we propose a model where divergent transcription is affected by the relative rates of both incorporation and eviction at the -1 nucleosome (Figure 4-6F). This two-step model is supported by the finding that both H3 K56A (which cannot be acetylated) and K56Q (to mimic hyperacetylation) lead to increased MUB1div and KIP1div expression. In the first step, acetylation of H3K56 by Rtt109 promotes nucleosome incorporation by CAF-I and Rtt106, so mutations in this assembly pathway reduce occupancy or positioning at the -1 nucleosome, thereby increasing divergent transcription. Supporting this idea, deletions of individual histone genes (yeast has two for each histone) scored strongly in the FPR screen. The gene for histone chaperone Nap1 scored just below our cutoff value. Whereas initial nucleosome incorporation is probably replication coupled, high H3K56ac at promoter-proximal nucleosomes (Venkatesh et al., 2012; Figure 4-6B) suggests additional turnover events, perhaps caused by remodelers or transcription initiation. We also note that deletions of two HIR complex subunits, needed for replication-independent chromatin assembly, also scored in the FPR assay. However, more work is needed to resolve whether HIR directly affects promoter directionality through nucleosome assembly or indirectly through its role in histone gene expression regulation (Fillingham et al., 2009).
In a second step after chromatin assembly, newly incorporated nucleosomes must be deacetylated to mediate repression. Combined deletion of the H3K56 HDACs Hst3 and Hst4 increases divergent transcription, as does the acetylation mimic H3K56Q. We propose that H3K56ac at the -1 nucleosome stimulates (directly or indirectly) eviction or remodeling by Swi/Snf. Indeed, H3K56ac promotes Swi/Snf function during transcription activation at mRNA promoters (Williams et al., 2008; Xu et al., 2005), and the same mechanisms presumably stimulate divergent ncRNA expression. In snf5Δ cells, -1 nucleosomes would be stabilized, suppressing the effects of deletion CAF-I or other H3K56ac-linked factors (Figure 4-7 and C-7).

H3K56ac also affects the balance between the Swr1 chromatin remodeler replacing H2A with histone variant H2A.Z and the Ino80 complex doing the reverse (Papamichos-Chronakis et al., 2011). Constitutive H3K56 acetylation reduces H2A.Z levels at the +1 nucleosome (Watanabe et al., 2013). At yeast telomeres, CAF-I promotes and H2A.Z antagonizes heterochromatic silencing (Kaufman et al., 1997; Meneghini et al., 2003). This antagonistic relationship may also help generate asymmetry between H3K56ac at -1 and H2A.Z at +1 nucleosomes, contributing to directionality. We identified swr1Δ in the FPR screen as having a decreased coding to noncoding ratio, although only minor effects were seen on MUB1 and KIP1. However, deletion of Ino80 complex subunit ARP8 increased divergent transcription at both these promoters (Figure C-7). Testing of other remodelers showed derepression of divergent transcription in cells lacking Isw1, Isw2, or Rsc2, but not Rsc1 or Chd1. Isw2 has previously been shown to suppress antisense transcription arising from mRNA 3’ ends (Whitehouse et al., 2007). However, only Swi/Snf activates divergent transcription.
Divergent ncRNA transcription in mammals is subject to rapid termination and degradation, a pathway suppressed at mRNAs by U1 snRNP at 5' splice sites (Almada et al., 2013; Ntini et al., 2013). Only a fraction of yeast genes contain introns, so additional mechanisms likely exist. For example, most divergent CUTs are terminated by the NNS pathway levels (Schulz et al., 2013). By swapping transcribed regions, we showed that the noncoding direction of divergent promoters can support productive elongation of coding mRNAs (Figure 4-1). As FPR lacks the sequences that would trigger early termination, our screen for mutants affecting promoter directionality was set up to uncover factors acting earlier than NNS, likely affecting initiation of divergent transcription. The CAF-I pathway acts upstream or in parallel to early termination, as shown by the synergistic increase in divergent transcripts upon combining CAF-I and Nrd1-Nab3-Sen1 pathway mutants (Figure 4-5).

CAF-I affects transcription from a small number of divergent coding/coding pairs, where the relatively weaker transcriptional direction is elevated in CAF-I mutants (Figure C-5B). At coding/noncoding pairs, NET-seq and ChIP results show transcription is generally weaker in the noncoding direction (Churchman and Weissman, 2011; Rhee and Pugh, 2012). A large fraction of these divergent promoters are affected by CAF-I, arguing that chromatin is a major enforcer of promoter directionality. Because the mechanism proposed here is independent of transcript identity, this differential is likely to be encoded in how the promoter sequences position nucleosomes. Although both directions are subject to chromatin remodeling and initiation events, rapid histone replacement and stabilization by CAF-I and the H3K56 cycle may counteract histone modifications such as acetylation and H3K4 trimethylation that promote increased
transcription in the stronger coding direction.

Further work will be required to fully elucidate any functional roles for divergent noncoding transcription or their resulting RNA products. Our work identifies specific chromatin factors as key modulators of transcription in the noncoding direction. The conservation of these factors and activities suggests that nucleosome assembly and disassembly pathways will also determine transcriptional directionality in multicellular organisms.

4.4 Experimental Procedures

4.4.1 Yeast Strains

Yeast strains used in this study are listed in Table C-1.

4.4.2 Northern Blotting

Twenty-five micrograms of total RNA was separated in a 1.5% 4-morpholinepropanesulfonic acid-formaldehyde agarose gel, transferred onto nylon membranes (GE10416296) using capillary electrophoresis, and incubated in hybridization buffer overnight with strand-specific $^{32}$P-labeled DNA probes as described in Marquardt et al. (2011). Primers for probe amplification by PCR and single-strand-specific labeling are listed in Table C-2. RNA levels were quantitated by Phosphorimager and normalized to a loading control.

4.4.3 Chromatin Immunoprecipitation

To analyze RNA Pol II, formaldehyde crosslinked chromatin fractions were immunoprecipitated with anti-Rpb3 antibodies (Neoclon; W0012) as described in
Keogh et al. (2005). Decrosslinked DNA was used as templates for quantitative PCR reactions. Rpb3 crosslinking to divergent transcript regions was normalized to an ACT1 control amplicon using the indicated primers (Table C-2).

4.4.4 Reverse Genetic Screen Using FPRs

Progenitor FPR strain (YSB2767) was generated by transforming MATa strain YF2133 with EcoRI-digested SB1629 (pCR-mCherry-PPT1/SUT129pro-YFP), replacing endogenous PPT1 and SUT129 with mCherry and YFP, respectively. YSB2767 was dispensed into sterile 384-well plates and mated with the S. cerevisiae nonessential yeast MATa deletion library using a 384-well pinning device. Diploids were selected for NatMx (FPR) and KanMx (deletion library). After sporulation, haploids containing both the FPR and deletion cassettes were selected for resistance to G418 and nourseothricin to select for the deletion and FPR, thiolysine and canavinine to eliminate diploids, and growth in the absence of leucine to ensure the haploid strains were MATa (Tong et al., 2001). Resulting strains were arrayed in 96-well plates and fluorescence in each well determined in midlogarithmic growth. Details of data collection and analysis are described in Supplemental Information.

4.4.5 Ranking of FPR Screen Candidate Deletion Mutations

The median mCherry and YFP values of each FPR-containing deletion strain were computed and normalized to the median value of the corresponding 96-well plate to account for plate-to-plate variation. Data were described by two axes of variation quantified by principal component analysis. The first component of variation (PC1) is the correlation of mCherry and YFP ($r = 0.51$; correlation axis in Figure 2C). The second
component of variation (PC2) measures the orthogonal distance from PC1 and hence is a measure for directionality (Figure S-2C and C-2D). Yeast deletion mutants were ranked ordered based on their PC2 directionality value (Figure C-2D). High directionality scores represent a reduced mCherry/YFP ratio and low scores a high mCherry/YFP ratio. Gene Ontology terms for top-ranking mutations (Figure C-2E) were derived using standard settings for the GO-term finder http://go.princeton.edu/cgi-bin/GOTermFinder (Boyle et al., 2004). A p value cutoff (p > 0.01) was chosen to identify significantly enriched terms.

4.4.6 NET-Seq Experiments and Analysis

NET-seq experiments were performed as described in Churchman and Weissman (2011, 2012) with minor modifications. See Supplemental Information for details.

4.4.7 Segmentation of H3K56ac and Histone Exchange Data

Genome-wide ChIP-chip data from Venkatesh et al. (2012) were organized into ten bins for each gene: two 250 nt regions upstream of the transcription start site, six bins scaled to the length of the gene, and two 250 nt regions downstream of the polyA site. CAF-I-affected loci based on NET-seq data were compared to the remaining loci. See Extended Experimental Procedures for more information.

4.5 References


regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. Cell 144, 200–213.


Author contributions

R.E.-C. analyzed the screen data. R.E-C. performed the GO term analysis. R.E-C. wrote custom MATLAB scripts to analyze flow cytometry data. R.E-C. and N.P. analyze NET-SEQ data. J.W. analyzed ChIP data. L.S.C helped designed NET-SEQ experiments. S.M. performed the experiments. S.M. and S.B. designed the experiments and wrote the manuscript.
Chapter 5: Conclusions

In this thesis we dissected at the molecular and genetic level the sensing of galactose and glucose in *S. cerevisiae*. We also used this powerful model organism to explore how the directionality of transcription is regulated.

In Chapter 2 we explored the integration of signals galactose and glucose. Our work shows that contrary to the canonical view, glucose does not simply repress GAL genes above a threshold concentration. Instead, cells sense a ratio of the two carbon sources. Ratio sensing can be generated as a result of two signals competing on the same substrate. Almost any biological substrate (DNA, RNA or protein) can be the target of competitive binding by multiple signaling molecules. Ratio sensing is therefore likely to occur in many settings in nature and might serve different and important physiological roles, which yet remain to be discovered. Additionally, our work highlights the importance of high throughput experiments that simultaneously measure many input signals in a system. This type of measurement has the potential of revealing a complex and surprising landscape of responses. These observations open up the possibility of discovering other molecular mechanisms, different from competitive binding, that underlie multiple signal integration. Another very interesting research direction is to explore the interactions between hexose transporters in yeast. There are more than 20 different transporters in the cell capable of importing hexoses with different affinities. Using a synthetic biology approach each transporter could be assayed individually or in different combinations to assess the contribution of each transporter to the establishment of molecular responses and cell physiology.
Chapter 3 revealed the large variation in the induction of GAL genes in strains of *S. cerevisiae*. Our study used a quantitative trait that corresponds to the molecular output of a specific network. Using molecular phenotypes for quantitative genetics studies has the potential of revealing specific “genetic tuning knobs”, that is, genes used in nature to control the variability in responses of a specific pathway to different environments. Exploring other quantitative aspects of the same response (e.g. the sharpness of the response, the heterogeneity in cells or the speed of a response) and using them as quantitative traits, should identify other “genetic tuning knobs”. The interactions between these genes could then reveal interesting principles of how specific features of a response are controlled in a network. Specifically, it would be interesting to determine the identity of the genes that control specific features of the response of a network. Perhaps this knowledge could guide the design of synthetic circuits where these features are tuned independently. These synthetic circuits could then be used to examine how tuning independent features of a network help a cell deal with specific environments, perhaps shedding light on the evolution of the network. This area of research will be greatly facilitated by the advent of new quantitative genetic methods that make use of new sequencing technologies and high-throughput phenotyping to more efficiently explore the variability in a species.

Recently, new and more sensitive technologies have revealed that transcription initiates in two opposite directions from almost every promoter in the genome. Chapter 4 illustrates how a previously unidentifed pre-initiation mechanism can limit divergent transcription through chromatin remodeling. Other mechanisms that enforce
directionality at different stages of transcription have been described such as gene looping and early termination linked to rapid degradation. To what extent these mechanisms are redundant and how their interactions maintain the global control of transcription has yet to be explored. The integration of these mechanisms will shed light into the intimate life cycle of a transcript, from initiation to termination and degradation. The abundance of mechanisms regulating transcription suggests that pervasive non-coding RNA expression is deleterious to the cell or that a complex regulation is needed for the proper function of these RNAs. Therefore an interesting avenue of research is to explore the roles of non-coding RNA in genetic networks, specifically how their regulation shapes the outputs of a network.

Many of the open questions presented in this thesis can be addressed by using the model organism yeast *S. cerevisiae*, which will continue to be an excellent tool to study eukaryotic biology. As we approach the frontier of our knowledge and seek out to better understand our own biology, the study of more complex organisms will be necessary to complete this quest.
Appendix A. Supporting Information for Chapter 2.

A.1. Strain construction

The strains used in this study are listed in Table A-1. The wild type strain background used for this paper was S288C, specifically FY4 (MAT a), FY5 (MAT a), and BY4741 (Brachmann et al., 1998). Multiple derivatives of this strain were created by standard transformation protocols. For example, SLYH14 was created by transforming a NotI digest of plasmid SLVD02, a HO-GAL1pr-YFP-hphNT1-HO integration cassette into FY4 and selecting for hygromycin resistance. FY5 was transformed with a NotI digest of A06V, a HO-TDH3pr-mCherry-natMX4-HO integration cassette, and selected for clonNAT to create SLH15. SLH9 was then generated by mating SLYH14 and SLYH15, and selecting for hygromycin and clonNAT resistant colonies. The strains BC1887 and YJM978 (Liti et al., 2009) were also used in this study. Strains were confirmed by colony PCR and fluorescence.

Prototrophic haploid deletion strains (SLYH25 to SLYH29) were generated by mating haploid deletion strains from the yeast deletion library (Winzeler et al., 1999) to a prototrophic haploid, followed by sporulation and selection on minimal media, followed by selection on G418.

A homozygous deletion strain of GAL2 (SLYH20, SLYH21) was created by mating a haploid gal2Δ carrying our GAL1pr-YFP reporter (SLYH31) to a haploid gal2Δ carrying a constitutive reporter (SLYH36) to give SLYH20. Analogous matings were performed to create homozygous and heterozygous deletes.
## A.2. Yeast strains and plasmids used in this study

All strains were made for this study, except SLYH13. SLYH13 is a strain from the GFP library (Huh et al., 2003)

<table>
<thead>
<tr>
<th>Table A-1 . Yeast strains used in this study.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strain</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>SLYH9</td>
</tr>
<tr>
<td>SLYH10</td>
</tr>
<tr>
<td>SLYH11</td>
</tr>
<tr>
<td>SLYH12</td>
</tr>
<tr>
<td>SLYH13</td>
</tr>
<tr>
<td>SLYH14</td>
</tr>
<tr>
<td>SLYH15</td>
</tr>
<tr>
<td>SLYH16</td>
</tr>
<tr>
<td>SLYH17</td>
</tr>
<tr>
<td>SLYH18</td>
</tr>
<tr>
<td>SLYH19</td>
</tr>
<tr>
<td>SLYH20</td>
</tr>
<tr>
<td>SLYH21</td>
</tr>
<tr>
<td>SLYH22</td>
</tr>
<tr>
<td>SLYH23</td>
</tr>
<tr>
<td>SLYH24</td>
</tr>
<tr>
<td>SLYH25</td>
</tr>
<tr>
<td>SLYH26</td>
</tr>
<tr>
<td>SLYH27</td>
</tr>
<tr>
<td>SLYH28</td>
</tr>
<tr>
<td>SLYH29</td>
</tr>
<tr>
<td>SLYH30</td>
</tr>
<tr>
<td>SLYH31</td>
</tr>
<tr>
<td>SLYH32</td>
</tr>
<tr>
<td>SLYH33</td>
</tr>
<tr>
<td>SLYH34</td>
</tr>
<tr>
<td>SLYH35</td>
</tr>
<tr>
<td>SLYH36</td>
</tr>
<tr>
<td>Plasmid number</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>SLVD02</td>
</tr>
<tr>
<td>SLVA63</td>
</tr>
<tr>
<td>SLVA06</td>
</tr>
<tr>
<td>SLVA19</td>
</tr>
<tr>
<td>SLVD13</td>
</tr>
</tbody>
</table>

### A.3. Growth conditions and media
Cells were grown in synthetic media S (Amberg et al., 2005) with 2% raffinose for 14 and 16 hours (the only exception being the gal80Δ strains which were grown in 2% glucose). Raffinose was chosen as the initial carbon source because it neither represses nor induces the GAL genes. SLYH13, the Mig1p-GFP fusion, is auxotrophic and therefore it was grown in synthetic complete media SC (Amberg et al., 2005).
We additionally illustrate on Figure CD measurements performed in previous studies where cells were grown in different concentrations of galactose and glucose. Note that the specific quantitative details of the reported thresholds of the GAL pathway are strain dependent.

Bennet and Pang (Bennett et al., 2008) examined two different strains YPH499 and K699 in a titration of pure galactose and observed the onset of induction of a GAL1-YFP reporter at ~0.03% galactose for K699 and ~0.1% galactose for YPH499. Biggar and Crabtree (Biggar and Crabtree, 2001) observed induction at around 0.05% in their YSB7 strain. For the strain W303 (Acar et al., 2005; Avendaño et al., 2013; Venturelli et al., 2012) it was observed the onset of expression at ~0.002% galactose. We observe a similar onset of expression for S288C.

Glucose at more than 0.2% repressed the GAL reporter for strains YPH499 and K699 growing in 0.2% galactose. Biggar and Crabtree (Biggar and Crabtree, 2001) observed repression of their GAL1-GFP reporter in an YSB7 strain growing in 2% galactose at concentrations as high as 2% glucose for a raffinose pre-grown culture. Consistent with our results, in both studies the authors observed that the GAL genes were repressed when the ratio of galactose to glucose was approximately lower than 1. In a related result, Acar et al. (Acar et al., 2005) observed the onset of GAL induction at 0.1% galactose in cells growing at 0.1% glucose.

A.4. The effect of inoculum size

To test experimentally the possibility that an OD of 0.003 might be too high for accurate studies of catabolite repression, we performed the same experiment with 4 different
initial ODs (Figure A-3 and B). Starting ODs of 1X, 1/5X and 1/25X our standard level of OD 0.003 all gave similar results, suggesting that nutrient depletion does not play a role in the response we see. In contrast, the response of strains grown at 5X initial OD was affected by depletion.

Assuming that a cell has a volume of $60 \mu m^3$ and a density of 1 g/mL, its total mass is $60 \times 10^{-12} g$. Since around 30% of the mass is dry mass and about 50% of the dry mass is carbon (Milo et al., 2010) the total carbon mass is $\sim 9 \times 10^{-12} g$, i.e. $7.5 \times 10^{-13}$ moles of carbon. Using a strict estimation in which each glucose molecule yields one carbon atom, each cell doubling requires $7.5 \times 10^{-13}$ moles of glucose. An OD600 of 1 is around $1 \times 10^{10}$ cells/L and therefore requires $7.5 \times 10^{-3}$ moles per liter of glucose, which corresponds to 7.5mM/OD. 10mM of glucose is $\sim 0.02\%$ so we estimate the glucose per doubling as $0.02\%$/OD. For example, the first doubling of a culture with initial OD of 0.003 will consume $6.1 \times 10^{-5}$ % of glucose. For exponentially growing cells the carbon concentration, $C$, and OD dynamics are given by:

\[ \frac{\partial}{\partial t} OD = OD(t) \cdot \lambda \]
\[ \frac{\partial}{\partial t} C = -OD(t) \cdot \delta \cdot \lambda \]

Eq. A-1 \[ \Rightarrow \]
\[ OD(t) = OD(0)e^{\lambda t} \]
\[ C(t) = C(0) - OD(0)\delta\left(e^{\lambda t} - 1\right) \]

where $\lambda$ is the growth rate and $\delta$ is the uptake rate per doubling. For example, a culture with initial OD of 0.003, a glucose concentration of 0.0156 % and a doubling time of 90 minutes will consume only 15% of the glucose after 8 hours. Figure A-3C illustrates the
effect of depletion on threshold response (Figure A-3D) and ratio response (Figure A-3E). The response effect was calculated as follows: 1. The depletion time for the cutoff concentration was calculated based on Eq. A-1 using a depletion rate of 0.02%/OD, doubling time of 90 minutes and an OD of 0.003. 2. If the depletion time, \( t_d \), is shorter than 8 hours (the time in which we perform our measurements) the response is the fraction between the time left for the end of the measurement (8 hours - depletion time) and 6 hours (which is the time in which the response attains its maximal value).

GAL gene induction was simulated at different concentrations of inoculum and assuming that the underlying sensor was either a threshold sensor or a ratio sensor. In the case of the threshold sensor, the induction caused by depletion of sugars alone does not match the observed ratio sensing regime.

In summary, these results show that the cell density used in our manuscript does not result in nutrient depletion. High inoculum density (5X) leads to glucose depletion causing induction in low glucose galactose mixtures.

### A.5. Metric for cell response

To quantify the percent of cells that respond in a given well, we first calculate the normalized probability density function of the \( i^{th} \) well, \( p_i \),

\[
\text{Eq. A-2} \quad p_i(x) = \frac{f_i(x)}{\int f_i(x) dx}
\]

where \( f_i \) is the measured signal distribution and \( x \) is the fluorescence in arbitrary units.
Next we compute \( \delta p_i \) by subtracting \( p_0 \) from \( p_i \) where \( p_0 \) is the signal distribution of cells grown in medium lacking galactose and high glucose (2%; the highest concentration used in any of our assays) and removing all negative values.

\[
\text{Eq. A-3} \quad \delta p_i(x) = \begin{cases} 
0 & p_i(x) < p_0(x) \\
p_i(x) - p_0(x) & p_i(x) \geq p_0(x)
\end{cases}
\]

The response, \( R_i \), is defined as:

\[
\text{Eq. A-4} \quad R_i = \int \delta p_i(x) dx
\]

We compared this metric to two other commonly used metrics (Figure A-6B): the mean of \( p_i \) and the density of \( p_i \) above some constant threshold. The main difference between the mean metric, the threshold metric, and area metric is that the mean metric does not saturate even when the support of \( p_i \) does not overlap with \( p_0 \) or when it is above the threshold.

**A.6. Growth and imaging conditions in a microfluidics chamber**

Time lapse microscopy of wild type cells grown in 4 combinations of glucose and galactose was performed using an Onix Cellasic system in an Y04C yeast microfluidic plate. Cells were grown overnight in synthetic minimal media 2% raffinose, then loaded into the device. Four different concentrations of glucose and galactose were used for the assay: 0.011% glucose and 0.0039% galactose; 0.011% glucose and 0.0221%
galactose; 0.0625% glucose and 0.0221% galactose; 0.0625% glucose and galactose 0.176% (Figure A-6)

A.7. Mig1p localization
A strain carrying a Mig1p-GFP fusion (Huh et al., 2003) was grown in SC with 2% raffinose and then switched to a double gradient of SC glucose and SC galactose, After 8 hours, cells were transferred to a glass bottom 96 plate treated with Concanavin A.

To quantify the localization, we counted the fraction of cells with observable nucleus localization of Mig1p (Figure 2-2 B).

We also performed an automatic estimation of localization. We performed two global image threshold using Otsu's method (MATLAB graythresh) one after the other. The first round segmented the cells and the second segmentation gave the most intense fluorescent area inside the cell. In cells grown in high glucose the fluorescence is concentrated in a small area, the nucleus, while in cells grown in no glucose the signal is uniform along the cell. We calculate the ratio between the segmented signal inside the cell and the area of the cell, \( f \). This ratio is typically 0.25 when the signal is localized and 0.5 in no glucose conditions. We define the cytoplasmic fraction (Figure A) as \( 1 – f_N \), where \( f_N \) is the normalized ratio \( f \).

Both methods resulted in similar results; the localization does not depend on galactose concentrations and exhibits a sharp transition between 0.0312% and 0.0078% glucose.

A.8. Ratio sensing as a result of competitive binding
Consider architecture in which two input molecules, an activator and a repressor, bind to
a regulating molecule (which could be on the DNA or on another protein) which has two binding sites. Once the regulator is bound only to the activator it can induce the desired function and we define this state as the output of the circuit. The levels of activator and repressor are controlled by external stimuli, different sugars in the case of the GAL pathway. The system can have complex interactions and feedback loops such as downregulation of the activator by the repressor or production of activator by the regulator function.

Although the activator and repressor dynamics could be complicated, the equations that describe the regulator occupation are:

\[
\begin{align*}
\partial_t AE &= -AE(k_r A + k_f R / \epsilon \cdot R) + EE \cdot A \cdot k_f A + AR \cdot k_r R \\
\partial_t ER &= -ER(k_r R + k_f A / \epsilon \cdot A) + EE \cdot R \cdot k_f R + AR \cdot k_r A \\
\partial_t AR &= -AR(k_r A + k_r R) + ER \cdot A \cdot k_f A / \epsilon + AE \cdot R \cdot k_f R / \epsilon
\end{align*}
\]

Eq. A-5

With the conservation, \( T_{\text{regulator}} = EE + EA + ER + AR \). The fraction of the binding sites which is bound by input molecule \( A \) is,

\[
f = \frac{1}{1 + \frac{K_A}{A(s_1, s_2)} + \frac{1}{\epsilon K_R} \frac{R(s_1, s_2)}{K_R A(s_1, s_2)}}
\]

Eq. A-6

where, \( K_i = \frac{k_r}{k_f} \). Note that Eq. A-5 is not enough to solve the steady state level of the system. To solve the system one must write the relevant model equations and then substitute the solution of \( R \) and \( A \) in Eq. A-5. We take a model independent approach and investigate what are the different functions of \( A(s_1, s_2) \) and \( R(s_1, s_2) \) that can yield a ratio sensor.
A.8.1. The linear case

We first analyze the simplest case in which \( R(s_i) = \alpha R s_i \) and \( A(s_2) = \alpha s_2 \). In this case the output is given by the fraction of bound regulator,

\[
\text{Eq. A-7} \quad f = \frac{1}{1 + \frac{C_A}{s_A} + \frac{1}{\varepsilon C_R} + \frac{C_A s_R}{C_R s_A}}
\]

where, \( C_i = K_i / \alpha_i \).

The output function in terms of the unitless variables, \( a = s_A / C_A \) and \( r = s_R / C_R \) is,

\[
\text{Eq. A-8} \quad f = \frac{1}{1 + \frac{1}{a} + \frac{1}{\varepsilon r} + \frac{r}{a}}
\]

The regime in which the output function is a ratio sensor, that is has the form \( 1/(1+r/a) \) when \( r/a \gg 1/a + r/\varepsilon \). Thus the condition for ratio sensing is,

\[
\text{Eq. A-9} \quad r \gg \frac{1}{1 - a/\varepsilon}.
\]

Eq. A-9 defines the ratio sensing regime (Figure A-8). The maximal response that could be achieved in the ratio sensing domain is,

\[
\text{Eq. A-10} \quad f_{max} = \frac{1}{1 + \frac{\min(r)}{\max(a)}} = \frac{1}{1 + 1/\varepsilon}
\]

104
Figure A-8 illustrates the dependence of the ratio sensing domain on the interaction between binding sites. In the case the binding sites are mutually exclusive, $\varepsilon \to \infty$, the output is a ratio sensor as long as $r \gg 1$ and the response has a full dynamic range from 0 to 1. As the interaction between the binding sites is reduced, both the phase-space and dynamical range of ratio sensing is reduced. When there is no interaction at all, $\varepsilon = 1$, there is no ratio sensor with full dynamical range.

**A.8.2. The Michaelian case**

A typical scenario is when the activator and repressor are Michaelian functions of the corresponding external cues,

\[
R = \alpha_R \frac{S_R}{S_R + K_{TR}}
\]

\[\text{Eq. A-11}\]

\[
A = \alpha_A \frac{S_A}{S_A + K_{TA}}
\]

This could be due to saturation in the transport of the external molecules or due to saturation of the production process of the activator and repressor. In term of the variables, $a = s_A / C_A$ and $r = s_R / C_R$ where $C_i = K_i/K_{ri} / \alpha_i$,

\[
\text{Eq. A-12} \quad \frac{R}{K_R} = \frac{\alpha_R}{K_R} \frac{C_Rr}{C_R r + K_{TR}} = \frac{\alpha_R}{K_R} \frac{r}{r + K_{TR} / C_R} = \frac{K_{TR}}{C_R} \frac{r}{r + K_{TR} / C_R}
\]

and in a similar way,

\[
\text{Eq. A-13} \quad \frac{A}{K_A} = \frac{K_{TA}}{C_A} \frac{r}{r + K_{TA} / C_A}
\]
In the non-saturating case, \( a \ll K_{iA} / C_A; r \ll K_{iR} / C_R \Leftrightarrow s_i \ll K_i \), the output function is the same as in Eq. A-8 (Figure A-8). When one of the external cues are saturating, the response approaches a constant.

**A.8.3. Ratio sensing input**

In this case the ratio sensing is done on the activator level,

\[
R = \alpha_R s_R
\]

**Eq. A-14**

\[
A = \frac{1}{\frac{s_R}{s_A} + \rho}
\]

When \( R \) is not saturating, in term of the variables, In term of the variables, \( a = s_A / C_A \)

and \( r = s_R / C_R \) where \( C_R = K_R / \alpha_R \), and \( C_A = K_A C_R / \alpha_A \),

\[
\frac{R}{K_R} = r
\]

**Eq. A-15**

\[
A = \frac{1}{\frac{1}{K_A} \frac{rC_R}{\alpha_A a C_A} + \rho} = \frac{1}{\frac{r}{a} + \frac{C_A}{C_R} \rho}
\]

And the output is,

\[
f = \frac{1}{1 + \left( \frac{r}{a} + \frac{C_A}{C_R} \rho \right) + \frac{1}{\epsilon} + \left( \frac{r}{a} + \frac{C_A}{C_R} \rho \right) r}
\]

**Eq. A-16**

\[
= \frac{1}{1 + \rho \frac{C_A}{C_R} + \frac{1}{\epsilon} \frac{C_A}{C_R} \rho + \frac{r}{a} (r + 1)}
\]
The output depends only on the ratio between the $a$ and a power of $r$ when (note that when $r \gg 1$, the output depends on the ratio between $r^2$ and $a$ rather than the ratio between $r$ and $a$) when, $r \gg a \left( \frac{1}{\epsilon} + \frac{C_A}{C_R} \rho \right) - 1$ or $s_R \gg s_A \left( \frac{K_A}{\alpha_A} \frac{1}{\epsilon} + \rho \right) - C_R$. It is evident that as the exclusion is higher the range of ratio sensing is increased (Fig. A-9).

A.8.4. Sharp response function

In the case where cooperativity takes place, the response function involves Hill function with coefficients larger than one. In that case the response function is,

$$ f = \frac{1}{1 + \left( \frac{K_A}{A(s_1, s_2)} \right)^{n_A} + \frac{1}{\epsilon} \left( \frac{R(s_1)}{K_R} \right)^{n_R} + \left( \frac{K_A}{A(s_1, s_2)} \right)^{n_A} \left( \frac{R(s_1)}{K_R} \right)^{n_R}} $$

Eq. A-17

where $K_i = \left( \frac{k_{ri}}{k_R} \right)^{n_i}$.

A.9. Fitness experiments

A $gal4\Delta$ strain (SLH42) and its congenic wild-type strain (SLH10), were co-cultured in synthetic media + 2% raffinose for 14 hours until they reached an OD of 0.3. Cells were washed 3 times in media with no carbon and then diluted to an OD600 of 0.0005 into synthetic media with 0.0156% glucose and 2% galactose or with 0.0156% glucose. The glucose-only media was supplemented with 2% sorbitol to maintain osmolarity. Cells
were grown 8 hours in media until they reached an OD of ~0.01, then cells were diluted 1:10 and sampled every 2 hours for 10 hours. Each sample was spun down, the supernatant removed, and cells were resuspended in TE + sodium 0.1% azide to stop growth – a procedure that takes approximately 25 minutes. Each time point was then measured by flow cytometry. Manual segmentation was used to identify the two populations. Data was then analyzed in MATLAB with custom scripts.

A.10. The dependence of U-\textsuperscript{13}C glucose and \textsuperscript{12}C galactose on external carbon concentrations.

To determine if the internal galactose levels are dependent on both the external concentration of glucose and galactose we monitored the incorporation of U-\textsuperscript{13}C glucose and U-\textsuperscript{12}C galactose into amino acids by gas chromatography mass spectrometry (GC-MS). To ensure that uptake is not being affected by regulation from the canonical signaling pathway, a \textit{mig1Δ gal80Δ gal2Δ} strain was used. A \textit{mig1Δ gal80Δ} strain constitutively expresses the genes required for galactose metabolism irrespective of the external concentration of glucose or galactose (Figure 2-2E). Gal2p is the sole galactose transporter upregulated in response to galactose; hence in a \textit{gal2Δ} galactose uptake occurs through HXT transporters without transcriptional feedback from galactose.

The intracellular sugar concentration is a balance of the uptake rate and the usage rate (Eq. A-18)

$$\frac{dGlc}{dt} = u_{glc} - \phi_{glc} Glc_i$$

$$\frac{dGal}{dt} = u_{gal} - \phi_{gal} Gal_i$$

Eq. A-18
where $Glc_i$ and $Gal_i$ are the intracellular glucose and galactose, respectively, $u$ denotes the corresponding uptake rate and $\phi$ denotes the rate at which $^{13}$C-glucose and $^{12}$C-galactose are phosphorylated into glucose 6-phosphate.

Since this process is much faster than the doubling time, the effect of dilution is negligible. In steady state, the flux to glucose 6-phosphate for each sugar is equal its respective uptake rate (Eq. A-19).

$$\begin{align*}
  u_{\text{glc}} &= \phi_{\text{glc}} G_{\text{glc} i} \\
  u_{\text{gal}} &= \phi_{\text{gal}} G_{\text{gal} i}.
\end{align*}$$

Eq. A-19

The $^{12}$C to $^{13}$C composition of amino acids, $AA(^{12}C)$ and $AA(^{13}C)$, is proportional to the $^{12}$C to $^{13}$C composition of glucose 6-phosphate that is directly proportional to the relative uptake rate of $^{13}$C-glucose and $^{12}$C-galactose,

$$\begin{align*}
  \frac{AA(^{13}C)}{AA(^{12}C)} &= \frac{^{13}C-G6P}{^{12}C-G6P} = \frac{\phi_{\text{glc}} G_{\text{glc} i}}{\phi_{\text{gal}} G_{\text{gal} i}} = \frac{u_{\text{glc}}}{u_{\text{gal}}}.
\end{align*}$$

Eq. A-20

Hence, by measuring the isotopic composition of carbons in amino acids we can measure the relative uptake rate of each sugar without knowing its intracellular concentration.

A saturated culture of the $mig1\Delta gal80\Delta gal2\Delta$ strain was diluted 1:1000 in 1% U-$^{13}$C glucose and grown to saturation. Cells were diluted 1:50 into fresh 1% U-$^{13}$C glucose and grown an additional 10 hours to an OD600 of ~0.6. Cells were then washed 3 times in synthetic media with no carbon. Cells were then resuspended in 100 uL of synthetic media with no carbon and split into 10 mL of 6 different combinations of glucose and galactose to obtain a final OD600 of between 0.07 and 0.1. Growth was monitored every 30 minutes or every hour until cells doubled twice. Cells were then collected by
centrifugation and washed in isotonic buffer 0.9% NaCl. Cell pellets were then manipulated as described by Zamboni et al (Zamboni et al., 2009).

Since $^{12}$C/$^{13}$C ratio is a dynamic quantity that depends on the number of doubling, we measured the number of doubling in $^{12}$C for each sampling based on OD600 (that is biomass doubling). (S3) The relation between the $^{12}$C/$^{13}$C uptake fraction, $f$, the measured $^{12}$C fraction after $n$ doublings, $f_n$, and the initial $^{12}$C fraction, $f_0$, is given by,

$$
\text{Eq. A-21} \quad f = \frac{f_n - (1/2)^n \cdot f_0}{1 - (1/2)^n}
$$

After taking into account growth effect we found that the $^{12}$C:$^{13}$C ratio increases as extracellular galactose is increased but decreases as extracellular glucose is increased.

Table A-3. Doubling times (Td) and number of divisions (n) for the mig1∆gal80gal2Δ strain

<table>
<thead>
<tr>
<th>Glucose (2x)</th>
<th>Replicate 1</th>
<th>Replicate 2</th>
<th>Replicate 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1% galactose</td>
<td>Td=3.21 n=0.94</td>
<td>Td=3.45 n=0.87</td>
<td>Td=3.11 n=1.93</td>
</tr>
<tr>
<td>2% galactose</td>
<td>Td=2.74 n=2.79</td>
<td>Td=2.74 n=1.93</td>
<td>Td=3.12 n=1.93</td>
</tr>
<tr>
<td>1% galactose</td>
<td>Td=2.74 n=2.79</td>
<td>Td=2.74 n=1.93</td>
<td>Td=3.05 n=1.97</td>
</tr>
<tr>
<td>2% galactose</td>
<td>Td=3.11 n=1.93</td>
<td>Td=3.12 n=1.93</td>
<td></td>
</tr>
</tbody>
</table>

A.11. Ratio as a result of inverse enzyme catalysis

Consider a process in which a substrate $S$ is being catalyzed by an enzyme $A$ into a product $P$, with the back reaction catalyzed by an enzyme $R$ (Figure A-10). An example
is phosphorylation and dephosphorylation.

This process is described by,

\[ \partial_t SA = -SA(kr_A + k\text{cat}_A) + S \cdot A \cdot kf_A \]

\[ \partial_t PR = -PR(kr_R + k\text{cat}_R) + P \cdot R \cdot kf_R \]

\[ \partial_t P = -P \cdot R \cdot kf_R + PR \cdot kr_R + SA \cdot k\text{cat}_A \]

such that the total concentration of substrate and product \( T = S + SA + PR + P \). In steady state, the fraction of product is given by,

\[ f = \frac{1}{1 + \frac{R K\text{eff}_A}{K_R k\text{cat}_A + k\text{cat}_R k\text{cat}_A + k\text{cat}_R} + \frac{r k\text{cat}_R}{a k\text{cat}_A} + r \left(1 + \frac{k\text{cat}_R}{k\text{cat}_A}\right)} = \frac{1}{1 + \frac{R v_A}{A v_R} + \frac{1}{\tau v_R}} \]

where, \( K\text{eff}_i = \frac{kr_i}{k\text{cat}_i + kf_i} \), \( r = R / K_{MR} \), \( a = A / K_{AA} \) and \( \tau = \left(\frac{1}{k\text{cat}_A} + \frac{1}{k\text{cat}_R}\right)^{-1} \). The response is a ratio sensor provided that

\[ \frac{v_A}{A} \gg \frac{1}{\tau} \Rightarrow A \ll v_A \tau = \frac{k\text{cat}_A + kr_A}{kf_A k\text{cat}_A + k\text{cat}_R} \leq K_A \]. In normalized variables,

\[ \frac{1}{a k\text{cat}_A} \gg 1 + \frac{k\text{cat}_R}{k\text{cat}_A} \Rightarrow a \ll \frac{k\text{cat}_R}{k\text{cat}_A} \leq 1 \]

\[ 1 + \frac{k\text{cat}_R}{k\text{cat}_A} \]
A.12. Considerations to study the potential advantage of ratio sensing

An interesting question is whether ratio sensing could constitute an optimal strategy of gene induction in mixed environments. This question can be thought of as a problem of optimal switching between an uninduced state to an induced state.

To address this problem several considerations need to be made. To begin with, the quantity that cells seek to optimize needs to be determined. One assumption could be that cells seek to maximize the total final biomass produced. However, in an environment with other potential competitors perhaps a more relevant quantity to minimize is the time to use up all nutrients.

The approach to this problem should take into account the dynamics of induction of the pathway. Induction of specific genes to metabolize alternative carbon sources is not instantaneous and it is accompanied by a temporary halt of cell divisions. Moreover, different carbon sources might be used by the cell with different efficiencies therefore directly affecting the growth rate of cells in each type of nutrient. Last, the concentration of nutrients in the environment dynamically varies depending on the onset of gene induction, as well as the number of individuals that are actively using the nutrient, which could also itself vary over time.

Understanding the potential advantages of ratio sensing will likely require a theoretical framework that takes into account all of these considerations and constitutes an active research direction out of the scope of this thesis that will be discussed somewhere else.
A.13. References


Figure A-1 Reproducibility of the data for S288C.

(A) Shows the same data as Figure 2-1B for additional 3 replicates. (B) Lines represent the decision front, a linear fit to the concentrations at which 20% of all cells in the population show induction, as in Figure 2-1B and Figure 2-1E for the replicates in panel A. (C) Fraction of inducing cells as a function of the ratio of glucose and galactose concentrations for each of the three replicates in panel a. Lines are a fit to the function $\rho^n / (\rho^n + (gal/gluc)^\rho^n)$; $n = 2.28 \pm 0.347$, $\rho = 0.64 \pm 0.086$. The error is standard deviation of the mean.
Figure A-2. Measured response compared to a multiplicative prediction for S288C. In our case the measured response is significantly different from the response predicted by multiplicative response (Kaplan et al., 2008; Winzeler et al., 1999). Measured response compared to a multiplicative prediction for (A) S288C and (B) BC187. Heatmaps are generated as in Figure 2-1B. The single response curves (left panels) are in the absence of glucose (top) and in 2% galactose (bottom).
Figure A-3 Nutrient depletion does not explain ratio sensing.

(A) The response of BC187 at 4 initial ODs: 1) 1x corresponding to OD 0.003, 2) 1/5x corresponding to OD 0.0006, 3) 1/25x corresponding to OD 0.00012, and 4) 5x corresponding to OD 0.015; all after 8 hours in a mixture of glucose and galactose. (B) Lines represent the decision front, a linear fit to the concentrations at which 20% of the cells in the population show induction. (C) Fraction of inducing cells as a function of the ratio of glucose and galactose concentrations for each of the three replicates. Lines are a fit to the function $\rho^\dagger / (\rho^\dagger + (gal / glu)^\circ)$. 1X: $n = 2.07$, $\rho = 1.77$; 1/25X: $n = 2.26$, $\rho = 2.37$; 1/5X: $n = 1.91$, $\rho = 2.34$; 5X: $n = 1.65$, $\rho = 0.38$. (D) Simulated effect of depletion on a threshold response. (E) Simulated effect of depletion on a ratio response.
Figure A-4 Response and growth in a microfluidics chamber.
Each condition was sampled in 4 different fields of view that had at least 20 single cell colonies at T = 0 per field of view. (A) Cells with a constitutive mCherry and a GAL1 promoter YFP fusion were grown in a microfluidics device for 8 hours. Cells were segmented using the mCherry signal and histograms of the YFP signal were plotted at different time points. The red histogram represents the YFP distribution at T = 0. (B) The evolution of the YFP distribution as a function of time in different concentrations. (C) Fraction of inducing cells as a function of time.
Figure A-5 Effect of heterozygous deletion of GAL genes.
Wild-type and congenic strains were co-cultured. Strains were segmented based on the expression of a constitutive fluorescent protein in each of the strains. (A) Heatmap of the response. (B) Fraction of inducing cells as a function of the ratio of glucose and galactose concentrations. Experiments were done in triplicate. (C) Response of a GAL1pr-YFP reporter in a mig1Δ mig2Δ double strain to mixtures of glucose and galactose. The black solid line is the decision front in mutant while the dashed line is the decision front in the wild type.
Figure A-6 Different metrics for quantifying the response of S288C.

(A) The same data as in Figure 2-1B in the main text (left panel) compared to a metric based on constant fluorescent expression threshold of $10^{2.5}$ (middle panel) or to a metric based on mean (right panel). (B) Contour lines at various fractional response levels, e.g. 0.5 is the contour when half of the cells respond. (C) Projection of the response to the galactose/glucose axis, as in Figure 2-1C. Note that the mean projection does not collapse as tightly onto a single curve as the other two metrics. (D) Concentrations of galactose and glucose used in 5 previous studies overlaid on the double gradient used in this manuscript. A blue stripe highlights titrations of galactose in the absence of glucose that have been extensively explored in the literature.
Figure A-7 Mig1 localization and quantification.
(4) Mig1p cytoplasmic fraction as quantified using an automated image processing algorithm (supporting information, section VII). This quantification generates similar results to the manual counting in Figure 2-2B in the main text (B) The fraction of cells with localized Mig1p (black) and the mean GAL1pr-YFP expression (red) as a function of glucose. The plots were generated by averaging over all galactose values.
Figure A-8 Ratio sensing modules and response functions.
(A) Schematic of ratio sensing module, consisting of an activator (green) and a repressor (blue) interacting with an integrator molecule. The system has 4 possible states (bottom). If the activator is bound, but not the repressor, the integrator carries out a function. In addition to the binding-unbinding kinetics, there is a negative interaction between the activator and repressor: once one of the regulators is bound it impedes the association constant of the other regulator – the activator impedes the repressor and the repressor impedes the activator. Negative interaction is denoted by $\varepsilon$.

(B) Response function as a function of extracellular sugar concentrations for the case in which the activator and the repressor are Michaelian functions of their corresponding sugars. The solid black lines define the boundary of the ratio sensing region. The dashed lines are $K_{TA} = C_A$ and $K_{TR} = C_R$ and denote the concentrations where saturation effects start to take place (see Supplementary Information). As the mutual interaction, $\varepsilon$, increases (left to right) the ratio sensing regime is increased according to Eq. A-9. The ratio between the saturation constant $K_{Ti}$ and the effective binding coefficient $C_i$ decreases from the top to bottom rows.
Figure A-9 Response functions where the activator is already assumed to depend on the ratio of input sugars or when activator and repressor are Hill functions.

(A) The curved black line defines one boundary of the regime in which the response depends on the ratio between activator and some power of the repressor. Below the horizontal line the response is based on the ratio between the activator and the repressor to the power one. Thus, the ratio sensing region is bottom and left of the two drawn boundaries. The parameters that were used in the figure are: $\rho = 1$; $C_A / C_R = 0.01$. (B) The figure illustrates the effect of increasing the Hill coefficient in the case of mutual inhibition (large $\varepsilon$). The slope of the decision front in a log-log plot depends on the ratio between the Hill coefficients of the repressor and activator, $n_R$ and $n_A$, respectively.
Figure A-10 Ratio sensing with enzymes of opposite functions.
(A), A futile cycle involving enzymes with opposite functions. In the schema, the activator and the repressor catalyze inverse reactions: the activator binds a substrate and catalyzes the formation of a product while the repressor binds the product and catalyzes the formation of the substrate. An example of such a process is phosphorylation catalyzed by kinase and dephosphorylation catalyzed by phosphatase. (B), Response, in this case product fraction, as a function of input enzymes. The vertical lines are the upper limit on the activator for ratio sensing to occur. The horizontal lines are the repressor value that gives a response of half maximal when the activator is saturating.
### B.1. Strains and plasmids used in this study

Table B-1. Yeast strains used in this study

<table>
<thead>
<tr>
<th>Strain ID</th>
<th>Background</th>
<th>Mating type</th>
<th>Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>RYE01</td>
<td>BC187</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE02</td>
<td>Bb32</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE03</td>
<td>CLIB215</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE04</td>
<td>CLIB324</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE05</td>
<td>CLIB382</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE06</td>
<td>DBVPG1106</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE07</td>
<td>DBVPG1373</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE08</td>
<td>DBVPG1788</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE09</td>
<td>DBVPG1853</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE10</td>
<td>DBVPG6040</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE11</td>
<td>DBVPG6765</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE12</td>
<td>FL100</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE13</td>
<td>IL-01</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE14</td>
<td>I-14</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE15</td>
<td>L-1374</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE16</td>
<td>L-1528</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE17</td>
<td>M22</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE18</td>
<td>NC-02</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE19</td>
<td>S288C</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE20</td>
<td>T7</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE21</td>
<td>UC5</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE22</td>
<td>UWOPS03-461.4</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE23</td>
<td>UWOPS87-242.1</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE24</td>
<td>WE372</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE25</td>
<td>Y12-SGRP</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE26</td>
<td>Y12-WashU</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE27</td>
<td>Y9-WashU</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE28</td>
<td>YJM421</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE29</td>
<td>YJM428</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE30</td>
<td>YJM653</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE31</td>
<td>YJM975</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE32</td>
<td>YJM978</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE33</td>
<td>YJM981</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE34</td>
<td>YPS1009</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE35</td>
<td>YPS128</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE36</td>
<td>YPS163</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE37</td>
<td>YPS606</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYE38</td>
<td>YPS166</td>
<td>MAT a/α</td>
<td>HO/hoΔ::GAL1pr-YFP-hphNT1</td>
</tr>
<tr>
<td>RYB22</td>
<td>BC187/YJM978</td>
<td>MAT a/x</td>
<td>BC187/YJM978 MATa/x hoΔ::GAL1pr-YFP-kanMX4/hoΔ::TDH3pr-mCherry-natMX4</td>
</tr>
<tr>
<td>RYB23</td>
<td>BC187/YJM978</td>
<td>MAT a/x</td>
<td>BC187/YJM978 MATa/x gal3Δ::hphNT1/GAL3-YJM978 hoΔ::GAL1pr-YFP-kanMX4/hoΔ::TDH3pr-mCherry-natMX4</td>
</tr>
<tr>
<td>RYB24</td>
<td>BC187/YJM978</td>
<td>MAT a/x</td>
<td>BC187/YJM978 MATa/x gal3Δ::hphNT1/GAL3-BC187 hoΔ::TDH3pr-mCherry-natMX4/hoΔ::GAL1pr-YFP-kanMX4</td>
</tr>
<tr>
<td>RYB3</td>
<td>YJM978</td>
<td>MAT a/x</td>
<td>YJM978/YJM978 MATa/x gal3Δ::hphNT1/GAL3-YJM978 hoΔ::GAL1pr-YFP-kanMX/hoΔ::TDH3pr-mCh-natMX</td>
</tr>
<tr>
<td>RYB42</td>
<td>BC187</td>
<td>MAT a/x</td>
<td>gal3Δ::hphNT1/GAL3-BC187 hoΔ::GAL1pr-YFP-kanMX4/hoΔ::TDH3pr-BFP-natMX4</td>
</tr>
<tr>
<td>RYC68</td>
<td>YJM978</td>
<td>MAT a/x</td>
<td>sokIΔ::hphNT1/SOK1-YJM978 hoΔ::GAL1pr-YFP-kanMX4/hoΔ::TDH3pr-mCherry-natMX</td>
</tr>
<tr>
<td>RYC69</td>
<td>BC187/YJM978</td>
<td>MAT a/x</td>
<td>sokIΔ::hphNT1/SOK1-BC187 hoΔ::GAL1pr-YFP-kanMX4/hoΔ::TDH3pr-mCherry-natMX</td>
</tr>
<tr>
<td>RYD40</td>
<td>BC187/YJM978</td>
<td>MAT a/α</td>
<td>sokIΔ::hphNT1/SOK1-YJM978 hoΔ::GAL1pr-kanMX4/hoΔ::TDH3pr-mCherry-natMX</td>
</tr>
</tbody>
</table>

Table B-1 (Continued)
<table>
<thead>
<tr>
<th>Strain</th>
<th>Strain</th>
<th>MAT Code</th>
<th>Genotype Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RYD41</td>
<td>BC187</td>
<td>MAT a/a</td>
<td>(\text{sok1}\Delta::\text{hphNT1/SOK1-BC187 ho}\Delta::\text{GAL1pr-kanMX4}/\text{ho}\Delta::\text{TDH3pr-mCherry-natMX4})</td>
</tr>
<tr>
<td>RYA16</td>
<td>YJM978</td>
<td>MAT a</td>
<td>(\text{ho}\Delta::\text{GAL1pr-YFP-kanMX})</td>
</tr>
<tr>
<td>RYB28</td>
<td>BC187</td>
<td>MAT a</td>
<td>(\text{ho}\Delta::\text{GAL1pr-YFP-kanMX4})</td>
</tr>
<tr>
<td>RYB53</td>
<td>YJM978</td>
<td>MAT a</td>
<td>(\text{gal3}\Delta::\text{hphNT1 ho}\Delta::\text{GAL1pr-YFP-kanMX4 yerc}\Delta8\Delta::\text{GAL3-BC187-natMX4})</td>
</tr>
<tr>
<td>RYB59</td>
<td>YJM978</td>
<td>MAT a</td>
<td>(\text{ho}\Delta::\text{GAL1pr-YFP-hphNT1 gal3}\Delta::\text{kanMX4 yerc}\Delta8\Delta::\text{GAL3-YJM978-natMX4})</td>
</tr>
<tr>
<td>RYB65</td>
<td>BC187</td>
<td>MAT a</td>
<td>(\text{gal3}\Delta::\text{hphNT1 ho}\Delta::\text{GAL1pr-YFP-kanMX4 yerc}\Delta8\Delta::\text{GAL3-BC187-natMX4})</td>
</tr>
<tr>
<td>RYB66</td>
<td>BC187</td>
<td>MAT a</td>
<td>(\text{gal3}\Delta::\text{hphNT1 ho}\Delta::\text{GAL1pr-YFP-kanMX4 yerc}\Delta8\Delta::\text{GAL3-YJM978-natMX4})</td>
</tr>
<tr>
<td>RYC44</td>
<td>YJM978</td>
<td>MAT a</td>
<td>(\text{gal3}\Delta::\text{hphNT1 ho}\Delta::\text{GAL1pr-YFP-kanMX4 yerc}\Delta8\Delta::\text{GAL3NCYC110-natMX4})</td>
</tr>
<tr>
<td>RYC45</td>
<td>YJM978</td>
<td>MAT a</td>
<td>(\text{gal3}\Delta::\text{hphNT1 ho}\Delta::\text{GAL1pr-YFP-kanMX4 yerc}\Delta8\Delta::\text{GAL3-L-1528-natMX4})</td>
</tr>
<tr>
<td>RYC48</td>
<td>YJM978</td>
<td>MAT a</td>
<td>(\text{gal3}\Delta::\text{hphNT1 ho}\Delta::\text{GAL1pr-YFP-kanMX4 yerc}\Delta8\Delta::\text{GAL3_NC-02-natMX4})</td>
</tr>
<tr>
<td>RYC49</td>
<td>YJM978</td>
<td>MAT a</td>
<td>(\text{gal3}\Delta::\text{hphNT1 ho}\Delta::\text{GAL1pr-YFP-kanMX4 yerc}\Delta8\Delta::\text{GAL3_YPS606-natMX4})</td>
</tr>
<tr>
<td>RYC50</td>
<td>YJM978</td>
<td>MAT a</td>
<td>(\text{gal3}\Delta::\text{hphNT1 ho}\Delta::\text{GAL1pr-YFP-kanMX4 yerc}\Delta8\Delta::\text{GAL3_I-14-natMX4})</td>
</tr>
<tr>
<td>RYC51</td>
<td>YJM978</td>
<td>MAT a</td>
<td>(\text{gal3}\Delta::\text{hphNT1 ho}\Delta::\text{GAL1pr-YFP-kanMX4 yerc}\Delta8\Delta::\text{GAL3_Y9-WashU-natMX4})</td>
</tr>
<tr>
<td>Strain</td>
<td>Source</td>
<td>MAT</td>
<td>Details</td>
</tr>
<tr>
<td>------------</td>
<td>-----------</td>
<td>-------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>RYC52</td>
<td>YJM978</td>
<td>MATa</td>
<td>gal3Δ::hphNT1 ho::GAL1pr-YFP-kanMX4 yercΔ8Δ::GAL3_UC-5-natMX4</td>
</tr>
<tr>
<td>RYC59</td>
<td>YJM978</td>
<td>MATa</td>
<td>hoΔ::GAL1pr-YFP-hphNT1 gal3Δ::kanMX4 yercΔ8Δ::GAL3_M22-natMX4</td>
</tr>
<tr>
<td>RYC60</td>
<td>YJM978</td>
<td>MATa</td>
<td>hoΔ::GAL1pr-YFP-hphNT1 gal3Δ::kanMX4 yercΔ8Δ::GAL3_YPS163-natMX4</td>
</tr>
<tr>
<td>RYC62</td>
<td>YJM978</td>
<td>MATa</td>
<td>hoΔ::GAL1pr-YFP-hphNT1 gal3Δ::kanMX4 yercΔ8Δ::GAL3_CLIB324-natMX4</td>
</tr>
<tr>
<td>RYC63</td>
<td>YJM978</td>
<td>MATa</td>
<td>hoΔ::GAL1pr-YFP-hphNT1 gal3Δ::kanMX4 yercΔ8Δ::GAL3_L-1374-natMX4</td>
</tr>
<tr>
<td>RYB92</td>
<td>YJM978</td>
<td>MATa</td>
<td>hoΔ::GAL1pr-YFP-hphNT1 gal3Δ::kanMX4 yercΔ8Δ::GAL3-S288C-natMX44</td>
</tr>
<tr>
<td>RYC72</td>
<td>YJM978</td>
<td>MATa</td>
<td>gal3Δ::hphNT1 hoΔ::GAL1pr-YFP-kanMX4 yercΔ8Δ::GAL3-FL100-natMX4</td>
</tr>
<tr>
<td>RYD25</td>
<td>YJM978</td>
<td>MATa</td>
<td>hoΔ::GAL1pr-YFP-hphNT1 gal3Δ::kanMX4 yercΔ8Δ::GAL3_CLIB382-natMX4</td>
</tr>
<tr>
<td>RYD26</td>
<td>YJM978</td>
<td>MATa</td>
<td>hoΔ::GAL1pr-YFP-hphNT1 gal3Δ::kanMX4 yercΔ8Δ::GAL3_CLIB382-natMX4</td>
</tr>
<tr>
<td>RYD27</td>
<td>YJM978</td>
<td>MATa</td>
<td>hoΔ::GAL1pr-YFP-hphNT1 gal3Δ::kanMX4 yercΔ8Δ::GAL3_DBVPG1106-natMX4</td>
</tr>
<tr>
<td>RYD28</td>
<td>YJM978</td>
<td>MATa</td>
<td>hoΔ::GAL1pr-YFP-hphNT1 gal3Δ::kanMX4 yercΔ8Δ::GAL3_DBVPG1373-natMX4</td>
</tr>
</tbody>
</table>
Table B-1 (Continued)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RYD29</td>
<td>YJM978</td>
<td>MAT a</td>
<td>hoΔ::GAL1pr-YFP-hphNT1 gal3Δ::kanMX4 yercΔ8Δ::GAL3_Y12-WasU-natMX4</td>
</tr>
<tr>
<td>RYD30</td>
<td>YJM978</td>
<td>MAT a</td>
<td>hoΔ::GAL1pr-YFP-hphNT1 gal3Δ::kanMX4 yercΔ8Δ::GAL3_YJM428-natMX4</td>
</tr>
<tr>
<td>RYD42</td>
<td>S288C</td>
<td>MAT x</td>
<td>hoΔ::GAL1pr-YFP-hphNT1 gal3Δ::kanMX4 yercΔ8Δ::GAL3_BC187-natMX4</td>
</tr>
<tr>
<td>RYB89</td>
<td>S288C</td>
<td>MAT x</td>
<td>hoΔ::GAL1pr-YFP-hphNT1 gal3Δ::kanMX4 yercΔ8Δ::GAL3_YJM978-natMX4</td>
</tr>
<tr>
<td>RYD01</td>
<td>Y9-WashU</td>
<td>MAT a</td>
<td>hoΔ::GAL1pr-YFP-hphNT1 gal3Δ::kanMX4 yercΔ8Δ::GAL3_BC187-natMX4</td>
</tr>
<tr>
<td>RYD02</td>
<td>Y9-WashU</td>
<td>MAT a</td>
<td>hoΔ::GAL1pr-YFP-hphNT1 gal3Δ::kanMX4 yercΔ8Δ::GAL3_YJM978-natMX4</td>
</tr>
<tr>
<td>RYD03</td>
<td>I_14</td>
<td>MAT a</td>
<td>hoΔ::GAL1pr-YFP-hphNT1 gal3Δ::kanMX4 yercΔ8Δ::GAL3_BC187-natMX4</td>
</tr>
<tr>
<td>RYD04</td>
<td>I_14</td>
<td>MAT a</td>
<td>hoΔ::GAL1pr-YFP-hphNT1 gal3Δ::kanMX4 yercΔ8Δ::GAL3_YJM978-natMX4</td>
</tr>
<tr>
<td>RYD05</td>
<td>YPS163</td>
<td>MAT a</td>
<td>hoΔ::GAL1pr-YFP-hphNT1 gal3Δ::kanMX4 yercΔ8Δ::GAL3_BC187-natMX4</td>
</tr>
<tr>
<td>RYD06</td>
<td>YPS163</td>
<td>MAT a</td>
<td>hoΔ::GAL1pr-YFP-hphNT1 gal3Δ::kanMX4 yercΔ8Δ::GAL3_YJM978-natMX4</td>
</tr>
<tr>
<td>RYD12</td>
<td>M22</td>
<td>MAT a</td>
<td>hoΔ::GAL1pr-YFP-hphNT1 gal3Δ::kanMX4 yercΔ8Δ::GAL3_BC187-natMX4</td>
</tr>
<tr>
<td>RYD13</td>
<td>M22</td>
<td>MAT a</td>
<td>hoΔ::GAL1pr-YFP-hphNT1 gal3Δ::kanMX4 yercΔ8Δ::GAL3_YJM978-natMX4</td>
</tr>
<tr>
<td>RYD14</td>
<td>CLIB382</td>
<td>MAT a</td>
<td>hoΔ::GAL1pr-YFP-hphNT1 gal3Δ::kanMX4 yercΔ8Δ::GAL3_BC187-natMX4</td>
</tr>
</tbody>
</table>
Table B-1 (Continued)

<table>
<thead>
<tr>
<th>ID</th>
<th>Cloning site</th>
<th>MAT</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RYD15</td>
<td>CLIB382</td>
<td>α</td>
<td>hoΔ::GAL1::YFP-hphNT1 gal3Δ::kanMX4 yerαΔ8::GAL3_YJM978-natMX4</td>
</tr>
</tbody>
</table>

Table B-2  Plasmids used in the study

<table>
<thead>
<tr>
<th>ID</th>
<th>GAL3 allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>RB2</td>
<td>GAL3-BC187</td>
</tr>
<tr>
<td>RB3</td>
<td>GAL3_YJM978</td>
</tr>
<tr>
<td>RB21</td>
<td>GAL3-S288C</td>
</tr>
<tr>
<td>RB41</td>
<td>GAL3_L-1528</td>
</tr>
<tr>
<td>RB45</td>
<td>GAL3_YPS606</td>
</tr>
<tr>
<td>RB47</td>
<td>GAL3_114</td>
</tr>
<tr>
<td>RB48</td>
<td>GAL3_NC-02</td>
</tr>
<tr>
<td>RB49</td>
<td>GAL3_Y9_WashU</td>
</tr>
<tr>
<td>RB50</td>
<td>GAL3_UC5</td>
</tr>
<tr>
<td>RB54</td>
<td>GAL3-M22</td>
</tr>
<tr>
<td>RB55</td>
<td>GAL3-yps163</td>
</tr>
<tr>
<td>RB57</td>
<td>GAL3-UWOPS034614</td>
</tr>
<tr>
<td>RB58</td>
<td>GAL3-CLIB324</td>
</tr>
<tr>
<td>RB62</td>
<td>GAL3-CLIB382</td>
</tr>
<tr>
<td>RB63</td>
<td>GAL3-DBVPG1106</td>
</tr>
<tr>
<td>RB64</td>
<td>GAL3-DBVPG1373</td>
</tr>
<tr>
<td>RB65</td>
<td>GAL3-Y12-WashU</td>
</tr>
<tr>
<td>RB66</td>
<td>GAL3_YJM421</td>
</tr>
<tr>
<td>RB67</td>
<td>GAL3_YJM428</td>
</tr>
</tbody>
</table>
B.2. Supplementary figures

Figure B-1 The set point of induction is reproducible and the measurements in glucose and galactose are correlated.

(A) Reproducibility of the metric (Correlation coefficient, R=0.96) (B) Correlation of set points in galactose and galactose. (Correlation coefficient R=0.83). Best fit in red line.
Figure B-2. Induction curves of all natural isolates in a titration of galactose at a constant glucose. Every panel represents the induction curves for each natural isolated listed in the title. The x-axis is log2 % galactose.
Figure B-2 (Continued)

Additional strains (West African and 378604X)
Figure B-3 Mapping of set points of induction to phylogeny and source of different strains of the strains

Neighbor joining tree was constructed using Phylomatch with genes YPR152C, YJL099W, YJL057C (Ramazzotti et. al 2012). Clean lineages are labeled by gray boxes. Source of the strains is color coded. Set point of induction from Figure 3-1 are represented as bars on the left panel.
Figure B-4 Frequency plots across chromosomes for each meiotic segregant pool

(A) Frequency of the allele YJM978 in both the BC187-like and YJM978-like pools. Most of the alleles are found at a 50% across all chromosomes.

(B) Specific regions of the genome that deviate from the expected random frequency. Deviations in the same direction are linked to selection biases in our study.
Figure B-5 Reciprocal hemizygous analysis of SOK1 and GAL3 suggests that SOK1 is not involved in establishing the set point of induction

(A) Schematic illustrating the construction of the hybrid and reciprocal hemizygous carrying either a BC187 or YJM978 allele only.

(B) SOK1 hemizygous reciprocal analysis. The hybrid (left), the hemizygous carrying a BC187 allele (center) and the hemizygous hybrid carrying a YJM978. A t-test could not reject the null hypothesis that the hybrid and each of the respective hemizygous hybrids are equal.

(C) GAL3 hemizygous reciprocal analysis. The hybrid (left), the hemizygous carrying a BC187 allele (center) and the hemizygous hybrid carrying a YJM978. A t-test could not reject the null hypothesis that the hybrid and each of the respective hemizygous hybrids are equal.

B.3. References


Appendix C. Supporting information for Chapter 4.

C.1. Extended Experimental Procedures

C.1.1. Construction of yeast strains

$sut129Δ::URA3$ was generated by yeast transformation of BY4741 with a PCR product containing $URA3$ with the oligonucleotides 2565 and 2566 to contain $SUT129$ flanking sequences. $URA3$ (in $sut129Δ::URA3$) was replaced with $CUT60$ by transformation with a PCR product of $CUT60$ containing $SUT129$ flanking sequences using 2801 and 2801, 5-FOA counter-selection was used to follow loss of $URA3$ (non-coding transcript coordinates were taken from (Xu et al., 2009)).

YSB2757 carrying the $PPT1/SUT129$ promoter replacement with the delitto-perfetto CORE sequence using 2805 and 2806 on pCORE template (Storici and Resnick, 2006).

YSB2764 carrying the $PPT1/SUT129$ promoter inversion was generated by replacing CORE with a PCR product inverting the orientation of the promoter amplified using 2807 and 2808 from genomic BY4741 DNA with appropriate 5’ overhang sequences. Loss of CORE can be detected by counter selection for loss of the URA3 marker using 5-FOA, the promoter inversion was confirmed by genotyping and sequencing. $SUT129$ start coordinate was taken from (Xu et al., 2009).

C.1.2. Yeast Media

Standard yeast media were used. YPD media plates containing 200 µg/ml G418 Sulfate (Geniticin, American Bioanalytical) to select for KanMx or/and 100 µg/ml ClonNat (Werner Bioagents) to select for NatMx. SC media containing uracil was used for
selection for growth in presence of 1 mg/ml 5-Fluoro Orotic Acid (5-FOA, US Biological). Cells with functional URA3 do not grow on 5-FOA. Additional media used for FPR screen as described in (Tong et al., 2001).

C.1.3. Plasmid construction

SB1603 was generated by cloning a PCR product generated from genomic YSB2737 template with the oligonucleotides 2570 and 2571 into pCR-TOPO-blunt (invitrogen) as described in the instruction manual.

SB1705 was generated by cloning a PCR product generated from genomic YF2292 template with the oligonucleotides 3038 and 3039 into pCR-TOPO-blunt (invitrogen) as described in the instruction manual.

F1047, F1048 and F1092 were obtained from Michael Springer.

SB1629 was constructed using one step isothermal assembly of four overlapping dsDNA fragments as described (Gibson, 2011). The backbone fragment including PPT1 and SUT129 3’ sequences for homologous recombination were amplified using SB1603 template with the oligonucleotides 2810 and 2815. The mCherry-NAT fragment was amplified using F1047 template with the oligonucleotides 2809 and 2870. The YFP fragment was amplified using F1048 template with the oligonucleotides 2816 and 2871. The PPT1/SUT12p promoter was amplified using genomic BY4741 yeast DNA template with the oligonucleotides 2846 and 2847. Assembly and identity was confirmed by sequencing. SUT129 flanking coordinates were taken from (Xu et al., 2009).
C.1.4. NET-seq analysis

Eight NET-seq libraries were generated: two biological repeats for wild-type and each CAF-I subunit deletion mutant. Nascent transcript sequencing libraries were generated as described with minor modifications (Churchman et al., 2011, 2012). DNA linker ligation was performed in a reaction mixture of 8 µl PEG 8000 (50%), 2 µl DMSO, 2 µl 10x T4 RNA Ligase 2 buffer (NEB), 1 µl (200 U) purified truncated T4 RNA ligase 2, 1 µl RNAselIN (Promega), 1 µl pre-adenylated Linker-1 (1 µg/µl), 5 µl purified nascent RNA (200 ng/µl). See supplementary Table C-2 for primers as part of NET-seq protocol. Libraries were sequenced using the Illumina Hiseq platform at the Harvard biopolymers facility using 50 bp single-end sequencing. Between 12.5-50 million reads were obtained for the eight libraries sequenced.

The post-sequencing files in FASTQ format were processed using the NET-Seq computational pipeline (code available at https://github.com/churchmanlab/NET-seq/). The primer sequence of LSC007 common to all libraries (ATCTCGTATGCCGTCTTCTGCTTG) was first stripped from the reads. Next, structural non-coding RNA (tRNA, rRNA, and snoRNA) were filtered out of the data using the their respective coordinates obtained from SGD. Alignment at this and all subsequent steps was provided by bowtie-0.12.8 (Langmead et al., 2009) and the structural non-coding RNAs removed have bowtie indexes included with the code (prefixed as sc_tRNA_snoRNA and rrna). The filtered reads were aligned to the *S. cerevisiae* genome annotation whose sequence and bowtie index are provided (prefixed as sc_sgd_gff_20091011). The resulting SAM files were then both processed to WIG format for immediate visualization as well as post-processed.
C.1.5. Visualization
The SAM files produced by the NET-seq computational pipeline were converted into two .wig files, one for each DNA strand, for visualization purposes using Mochiview (Homann and Johnson, 2010).

C.1.6. Post-processing
The SAM files produced by the NET-Seq computational pipeline were converted to BAM and indexed using the SAMTools package (Li et al., 2009). Then the coding to non-coding ratio (CD/NC) was calculated. To capture the reads representing the coding direction, the protein coding gene annotation was determined using TSS annotations and gene models provided by the literature (Xu et al., 2009). To capture reads representing divergent non-coding transcription, reads were quantified in a window of up to 600 nt size. We counted reads between 100 nt downstream and up to 500 nt upstream of the TSS on the complementary strand of an annotated gene, unless that would overlap with a coding annotation on either strand, in which case the non-coding interval used in the calculation stopped 10 bp before the annotated gene boundary. To account for differences in the number of reads between the sequencing libraries the data was normalized using the reads per kilobase per million (RPKM) method. Normalized coding and divergent non-coding values were obtained genome-wide and plotted for visualization.

CAF-I affected loci were defined by identifying loci whose ratio of coding to divergent non-coding transcript was at least log2 (1.5) fold smaller in the mutant compare to the wild-type (visualized by the red line in Figure 4C, Figure C-4 C,E). We required each CAF-I affected loci to have this relationship in at least 4 of the 6 CAF-I mutant NET-seq libraries.
C.1.7. Calculating NETseq coverage for CUTs and SUTs.
CUT and SUT annotation were obtained (Xu et al., 2009) and the RPKM value calculated for each transcript.

C.1.8. Definition of divergent ORF pairs sharing a 5’NFR
Definition of divergently transcribed ORF pairs sharing a 5’ NFR was obtained (Xu et al., 2009) and the associated RPKM value used for calculation.

C.1.9. Quantitative PCR (qPCR)
Yeast RNA was extracted using the Phenol/Chloroform method. RNA was and digested with DNase TURBO (Ambion) and DNase removed according to manufacturers’ instructions. cDNA was generated with gene specific primers from DNase treated yeast RNA using SuperScript III (Invitrogen) following the instruction manual. qPCR reactions were performed using diluted ChIP DNA or cDNA samples with specific qPCR primers (see oligonucleotide table) in a Roche Lightcycler 480. A home-made qPCR reaction mix was used as described in (Geisberg and Struhl, 2005). Quantification relative to reference genes was performed using the using the ΔΔcT method (Schmittgen and Livak, 2008).

C.1.10. α-factor arrest
Cycling yeast cells (OD=0.3) were incubated in YPD media containing 4.5 μg/ml α-factor peptide (gift of Hannah Mischo) twice for 60 minutes, arrest of cell division was controlled by yeast cell morphology. Media containing α-factor was washed away to analyze the effect in synchronized yeast cultures exiting G1.
**C.1.11. Flow cytometry**

A Beckton Dickinson LSRII flow cytometer with a high-throughput resolution sampler was used to quantitate BFP, YFP, and mCherry fluorescence. YFP was excited at 488 nm and fluorescence collected through a 550/30 band-pass and 525 LP emission filter. mCherry was excited at 593 nm and assayed with a 630/22 band pass filter. BFP was excited with a 405 nm laser and fluorescence collected with a 450/50 band pass filter. Between 10000 and 100000 events were collected for each well. Flow cytometry data was exported from the acquisition program (FACSDiva, Beckton Dickinson, San Jose, CA, USA) in the FCS3.0 format. Matlab was used to process the data, importing using a custom modified version of FCSread.m (Robert Hanson, available at Matlab Central) which makes it compatible with the LSRII data. To compare cells with the same average sizes and filter out cellular debris and aggregates, the median side scatter (SSC) was used. Data within +/- 10% of the median SSC value were used for all the comparisons. The first third of events and the last ninth of data in time were removed to minimize errors due to unstable sample flow through the flow cytometer. Any well that had fewer than 700 counted cells was excluded from the analysis to remove wells with failed inoculations (see also Figure C-2A-C).

**C.1.12. Targeted mutant versus wild-type FPR comparisons using FCM**

To confirm candidate mutants identified in the screen, deletions strains were re-created de novo and grown to mid-logarithmic phase for analysis by FCM. The median mCherry and YFP fluorescence values were determined using Matlab. Yeast strains without FPR served as background control in FCM measurements. Background corrected mCherry,
YFP, and mCherry/YFP values of several independently generated strains of the same genotype were used to calculate their means. The data was then represented normalized to co-assayed isogenic wild-type FPR control strains.

C.1.13. Determining p Values with Multiple Hypotheses Testing Correction for Metagene Analysis in Figure 4-7B

To assess the significance of the difference between CAF-I affected and CAF-I nonaffected genes in histone exchange or H3K56 acetylation, we used bootstrapping to compute the probability, among 107 randomly permuted pairs of ‘‘candidate’’ and ‘‘noncandidate’’ gene sets, that their means would differ in magnitude at any bin by as much or more as the observed difference at a given bin. This p value represents a two-tailed, multiple-testing-corrected p value for the observed difference at each bin. For any bin where $p < 0.05$, the p value is indicated in Figure 4-7B with an asterisk. Table of p values For histone exchange
<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Genotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>BY4741</td>
<td>wild-type BY4741</td>
<td>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0</td>
<td>(Winzeler et al., 1999)</td>
</tr>
<tr>
<td>YSB2375</td>
<td>rrp6Δ</td>
<td>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; rrp6Δ::KANMX</td>
<td>(Marquardt et al., 2011)</td>
</tr>
<tr>
<td>YF1694</td>
<td>xrn1Δ</td>
<td>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; xrn1Δ::KANMX</td>
<td>(Winzeler et al., 1999)</td>
</tr>
<tr>
<td>YSB2757</td>
<td>ppt1/sut129proΔ::CORE E</td>
<td>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; ppt1/sut129proΔ::CORE</td>
<td>This study</td>
</tr>
<tr>
<td>YSB2764</td>
<td>PPT1/SUT129 promoter inversion</td>
<td>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; ppt1/sut129proΔ::CORE [inverted]</td>
<td>This study</td>
</tr>
<tr>
<td>YSB2755</td>
<td>cut60Δ::URA3</td>
<td>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; cut60Δ::URA3</td>
<td>This study</td>
</tr>
<tr>
<td>YSB2756</td>
<td>xrn1Δ cut60Δ::URA3</td>
<td>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; cut60Δ::URA3; xrn1Δ::KANMX</td>
<td>This study</td>
</tr>
<tr>
<td>YSB3057</td>
<td>rrp6Δ cut60Δ::URA3</td>
<td>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; cut60Δ::URA3; rrp6Δ::KANMX</td>
<td>This study</td>
</tr>
<tr>
<td>YSB3056</td>
<td>cut60Δ::URA3 sut129Δ::CUT60</td>
<td>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; sut129Δ::CUT60; cut60Δ::URA3</td>
<td>This study</td>
</tr>
<tr>
<td>YSB3057</td>
<td>xrn1Δ cut60Δ::URA3 sut129Δ::CUT60</td>
<td>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; sut129Δ::CUT60; cut60Δ::URA3; xrn1Δ::KANMX</td>
<td>This study</td>
</tr>
<tr>
<td>YSB3058</td>
<td>rrp6Δ cut60Δ::URA3 sut129Δ::CUT60</td>
<td>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; sut129Δ::CUT60; cut60Δ::URA3; rrp6Δ::KANMX</td>
<td>This study</td>
</tr>
<tr>
<td>YSB2759</td>
<td>sut129Δ::URA3</td>
<td>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; sut129Δ::URA3</td>
<td>This study</td>
</tr>
<tr>
<td>YSB3059</td>
<td>sut129Δ::CUT60</td>
<td>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; sut129Δ::CUT60</td>
<td>This study</td>
</tr>
<tr>
<td>YSB2737</td>
<td>sut129Δ::URA3 ppt1Δ::HIS3</td>
<td>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; sut129Δ::URA3; ppt1Δ::HIS3</td>
<td>This study</td>
</tr>
<tr>
<td>YM5196</td>
<td>SGA mat-α parent</td>
<td>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; can1Δ::STE2pr-spHIS5; lyp1ΔSTE3pr-LEU2; cyh2</td>
<td>(Schuldiner et al., 2006)</td>
</tr>
<tr>
<td>YSB2767</td>
<td>FPR screen parent</td>
<td>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; can1Δ::STE2pr-spHIS5; lyp1ΔSTE3pr-LEU2; cyh2; ppt1Δ::mCherry::NAT; sut129Δ::YFP</td>
<td>This study</td>
</tr>
<tr>
<td>YSB2770</td>
<td>FPR wild-type control</td>
<td>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; ppt1Δ::mCherry::NAT; sut129Δ::YFP</td>
<td>This study</td>
</tr>
<tr>
<td>Strain</td>
<td>Description</td>
<td>Genotype</td>
<td>Reference</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>YSB3060</td>
<td>BFP+ FPR wild-type control</td>
<td>MATa; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; ppt1Δ::mCherry::NAT; sut129Δ::YFP; hoΔ::BFP::KAN</td>
<td>This study</td>
</tr>
<tr>
<td>YF1365</td>
<td>cac1Δ</td>
<td>MATa; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; cac1Δ::KANMX</td>
<td>(Winzeler et al., 1999)</td>
</tr>
<tr>
<td>YF2292</td>
<td>cac2Δ</td>
<td>MATa; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; cac2Δ::KANMX</td>
<td>(Winzeler et al., 1999)</td>
</tr>
<tr>
<td>YF1370</td>
<td>cac3Δ</td>
<td>MATa; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; cac3Δ::KANMX</td>
<td>(Winzeler et al., 1999)</td>
</tr>
<tr>
<td>YSB3015</td>
<td>FPR cac1Δ</td>
<td>MATa; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; ppt1Δ::mCherry::NAT; sut129Δ::YFP; cac1Δ::KANMX</td>
<td>This study</td>
</tr>
<tr>
<td>YSB3016</td>
<td>FPR cac2Δ</td>
<td>MATa; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; ppt1Δ::mCherry::NAT; sut129Δ::YFP; cac2Δ::KANMX</td>
<td>This study</td>
</tr>
<tr>
<td>YSB3014</td>
<td>FPR cac3Δ</td>
<td>MATa; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; ppt1Δ::mCherry::NAT; sut129Δ::YFP; cac3Δ::KANMX</td>
<td>This study</td>
</tr>
<tr>
<td>YSB3061</td>
<td>Promoter inversion cac2Δ</td>
<td>MATa; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; ppt1/sut129pro [inverted]; cac2Δ::KANMX</td>
<td>This study</td>
</tr>
<tr>
<td>YSC0001</td>
<td>Rpb3-FLAG</td>
<td>MATa; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; RPB3Δ::RPB3-3xFLAG::NatMX</td>
<td>(Churchman et al., 2011)</td>
</tr>
<tr>
<td>YSB3062</td>
<td>Rpb3-FLAG cac1Δ</td>
<td>MATa; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; RPB3Δ::RPB3-3xFLAG::NatMX; cac1Δ::KANMX</td>
<td>This study</td>
</tr>
<tr>
<td>YSB3063</td>
<td>Rpb3-FLAG cac2Δ</td>
<td>MATa; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; RPB3Δ::RPB3-3xFLAG::NatMX; cac2Δ::KANMX</td>
<td>This study</td>
</tr>
<tr>
<td>YSB3064</td>
<td>Rpb3-FLAG cac3Δ</td>
<td>MATa; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; RPB3Δ::RPB3-3xFLAG::NatMX; cac3Δ::KANMX</td>
<td>This study</td>
</tr>
<tr>
<td>LRB535</td>
<td>SSU72</td>
<td>MATa; ura3-52; leu2-3,112; his3Δ200</td>
<td>(Pappas and Hampsey, 2000)</td>
</tr>
<tr>
<td>YZ584</td>
<td>ssu72-2</td>
<td>MATa; ura3-52; leu2-3,112; his3Δ200; ssu72-2</td>
<td>(Pappas and Hampsey, 2000)</td>
</tr>
<tr>
<td>1971</td>
<td>SEN1</td>
<td>MATx; ura3-52; leu2-3,112; pep4-3</td>
<td>(Pappas and Hampsey, 2000)</td>
</tr>
<tr>
<td>FWY1</td>
<td>sen1-1</td>
<td>MATx; ura3-52; leu2-3,112; pep4-3; sen1-1</td>
<td>(Pappas and Hampsey, 2000)</td>
</tr>
<tr>
<td>W303-1B</td>
<td>NAB3 (W303-1B)</td>
<td>MATx; ura3-1; leu2-3,112; trp1-1; his3-11,15; ade2-1; can1-100</td>
<td>Claire Moore</td>
</tr>
<tr>
<td>Yeast Name</td>
<td>Mutation</td>
<td>First Column</td>
<td>Description</td>
</tr>
<tr>
<td>------------</td>
<td>----------</td>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>YPN103</td>
<td>nab3-11</td>
<td>MATx; ura3-1; leu2-3,112; trp1-1; his3-11,15; ade2-1; can1-100; nab3-11</td>
<td>(Conrad et al., 2000)</td>
</tr>
<tr>
<td>YSB2064</td>
<td>nrd1Δ6-150</td>
<td>MATa; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; nrd1Δ6-150</td>
<td>(Vasiljeva et al., 2006)</td>
</tr>
<tr>
<td>YSB3065</td>
<td>SSU72 cac2Δ</td>
<td>MATa; ura3-52; leu2-3,112; his3Δ200; cac2Δ::KANMX</td>
<td>This study</td>
</tr>
<tr>
<td>YSB3066</td>
<td>ssu72-2 cac2Δ</td>
<td>MATa; ura3-52; leu2-3,112; his3Δ200; ssu72-2; cac2Δ::KANMX</td>
<td>This study</td>
</tr>
<tr>
<td>YSB3067</td>
<td>SEN1 cac2Δ</td>
<td>MATa; ura3-52; leu2-3,112; pep4-3; cac2Δ::KANMX</td>
<td>This study</td>
</tr>
<tr>
<td>YSB3068</td>
<td>sen1-1 cac2Δ</td>
<td>MATa; ura3-52; leu2-3,112; pep4-3; sen1-1; cac2Δ::KANMX</td>
<td>This study</td>
</tr>
<tr>
<td>YSB3069</td>
<td>NAB3 cac2Δ</td>
<td>MATa; ura3-52; leu2-3,112; trp1-1; his3-11,15; ade2-1; can1-100; cac2Δ::KANMX</td>
<td>This study</td>
</tr>
<tr>
<td>YSB3070</td>
<td>nab3-11 cac2Δ</td>
<td>MATa; ura3-52; leu2-3,112; trp1-1; his3-11,15; ade2-1; can1-100; nab3-11; cac2Δ::KANMX</td>
<td>This study</td>
</tr>
<tr>
<td>YSB2710</td>
<td>rrp6Δ-NAT</td>
<td>MATa; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; rrp6Δ::NATMX</td>
<td>(Hazelbaker et al., 2013)</td>
</tr>
<tr>
<td>YSB3071</td>
<td>rrp6Δ cac2Δ</td>
<td>MATa; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; rrp6Δ::NATMX; cac2Δ::KANMX</td>
<td>This study</td>
</tr>
<tr>
<td>YSB3072</td>
<td>xrn1Δ-NAT</td>
<td>MATa; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; xrn1Δ::NATMX</td>
<td>This study</td>
</tr>
<tr>
<td>YSB3073</td>
<td>xrn1Δ cac2Δ</td>
<td>MATa; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; xrn1Δ::NATMX; cac2Δ::KANMX</td>
<td>This study</td>
</tr>
<tr>
<td>YSB30 3</td>
<td>cac1Δ-NAT</td>
<td>MATa; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; cac1Δ::NATMX</td>
<td>This study</td>
</tr>
<tr>
<td>YSB3075</td>
<td>cac1Δ cac2Δ</td>
<td>MATa; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; cac1Δ::NATMX; cac2Δ::KANMX</td>
<td>This study</td>
</tr>
<tr>
<td>YF2291</td>
<td>rtt106Δ</td>
<td>MATa; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; rtt106Δ::KANMX</td>
<td>(Winzeler et al., 1999)</td>
</tr>
<tr>
<td>YSB3076</td>
<td>rtt106Δ-NAT</td>
<td>MATa; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; rtt106Δ::NATMX</td>
<td>This study</td>
</tr>
<tr>
<td>YSB3077</td>
<td>rtt106Δ cac2Δ</td>
<td>MATa; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; rtt106Δ::NATMX; cac2Δ::KANMX</td>
<td>This study</td>
</tr>
<tr>
<td>YF2206</td>
<td>H3</td>
<td>MATa; ura3Δ0; leu2Δ0; trp1Δ63; his3Δ200; met15Δ0; lys2Δ0; can1::MFA1pr-HIS3; hht1-hhf1::NatMX4; hht2-hhf2::[HHTS-HHFS]-URA3</td>
<td>(Dai et al., 2008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td><strong>Table C-1 (Continued)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>YF2208</strong></td>
<td><strong>H3K56A</strong></td>
<td><strong>MATα; ura3Δ0; leu2Δ0; trp1Δ63; his3Δ200; met15Δ0; lys2Δ0; can1::MFA1pr-HIS3; hht1- hhf1::NatMX4; hht2- hhf2::[hhts(K56A)-HHFS]-URA3</strong></td>
<td></td>
</tr>
<tr>
<td><strong>YF2210</strong></td>
<td><strong>H3K56Q</strong></td>
<td><strong>MATα; ura3Δ0; leu2Δ0; trp1Δ63; his3Δ200; met15Δ0; lys2Δ0; can1::MFA1pr-HIS3; hht1- hhf1::NatMX4; hht2- hhf2::[hhts(K56Q)-HHFS]-URA3</strong></td>
<td></td>
</tr>
<tr>
<td><strong>YSB3078</strong></td>
<td><strong>H3 cac2Δ</strong></td>
<td><strong>MATα; ura3Δ0; leu2Δ0; trp1Δ63; his3Δ200; met15Δ0; lys2Δ0; can1::MFA1pr-HIS3; hht1- hhf1::NatMX4; hht2- hhf2::[HHTS- HHFS]-URA3; cac2Δ::KANMX</strong></td>
<td></td>
</tr>
<tr>
<td><strong>YSB3079</strong></td>
<td><strong>H3K56A cac2Δ</strong></td>
<td><strong>MATα; ura3Δ0; leu2Δ0; trp1Δ63; his3Δ200; met15Δ0; lys2Δ0; can1::MFA1pr-HIS3; hht1- hhf1::NatMX4; hht2- hhf2::[hhts(K56A)-HHFS]-URA3; cac2Δ::KANMX</strong></td>
<td></td>
</tr>
<tr>
<td><strong>YSB3080</strong></td>
<td><strong>H3K56Q cac2Δ</strong></td>
<td><strong>MATα; ura3Δ0; leu2Δ0; trp1Δ63; his3Δ200; met15Δ0; lys2Δ0; can1::MFA1pr-HIS3; hht1- hhf1::NatMX4; hht2- hhf2::[hhts(K56Q)-HHFS]-URA3; cac2Δ::KANMX</strong></td>
<td></td>
</tr>
<tr>
<td><strong>YF1616</strong></td>
<td><strong>hst4Δ</strong></td>
<td><strong>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; hst4Δ::KANMX</strong></td>
<td></td>
</tr>
<tr>
<td><strong>YSB3081</strong></td>
<td><strong>hst3Δ hst4Δ-KAN</strong></td>
<td><strong>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; hst4Δ::KANMX; hst3Δ::HIS3</strong></td>
<td></td>
</tr>
<tr>
<td><strong>YSB3082</strong></td>
<td><strong>hst3Δ hst4Δ</strong></td>
<td><strong>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; hst4Δ::NATMX; hst3Δ::HIS3</strong></td>
<td></td>
</tr>
<tr>
<td><strong>YSB3083</strong></td>
<td><strong>hst3Δ hst4Δ cac2Δ</strong></td>
<td><strong>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; hst4Δ::NATMX; hst3Δ::HIS3; cac2Δ::KANMX</strong></td>
<td></td>
</tr>
<tr>
<td><strong>YF538</strong></td>
<td><strong>snf5Δ</strong></td>
<td><strong>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; snf5Δ::KANMX</strong></td>
<td></td>
</tr>
<tr>
<td><strong>YSB3084</strong></td>
<td><strong>hst3Δ hst4Δ snf5Δ</strong></td>
<td><strong>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; hst4Δ::NATMX; hst3Δ::HIS3; snf5Δ::KANMX</strong></td>
<td></td>
</tr>
<tr>
<td>Strain</td>
<td>PhenoType</td>
<td>Genetic Composition</td>
<td>Source</td>
</tr>
<tr>
<td>------------</td>
<td>---------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>YSB3085</td>
<td>H3 snf5Δ</td>
<td>MATα; ura3Δ0; leu2Δ0; trp1Δ63; his3Δ200; met15Δ0; lys2Δ0; can1::MFA1pr-HIS3; hht1-hhf1::NatMX4; hht2-hhf2::[HHTS-HHFS]-URA3; snf5Δ::KANMX</td>
<td>This study</td>
</tr>
<tr>
<td>YSB3087</td>
<td>H3K56Q snf5Δ</td>
<td>MATα; ura3Δ0; leu2Δ0; trp1Δ63; his3Δ200; met15Δ0; lys2Δ0; can1::MFA1pr-HIS3; hht1-hhf1::NatMX4; hht2-hhf2::[hhts(K56Q)-HHFS]-URA3; snf5Δ::KANMX</td>
<td>This study</td>
</tr>
<tr>
<td>YSB3088</td>
<td>eac1Δ snf5Δ</td>
<td>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; cac1Δ::NATMX; snf5Δ::KANMX</td>
<td>This study</td>
</tr>
<tr>
<td>YF1362</td>
<td>rsc1Δ</td>
<td>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; rsc1Δ::KANMX</td>
<td>(Winzeler et al., 1999)</td>
</tr>
<tr>
<td>YSB3089</td>
<td>hst3Δ hst4Δ rsc1Δ</td>
<td>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; hst4Δ::NATMX; hst3Δ::HIS3; rsc1Δ::KANMX</td>
<td>This study</td>
</tr>
<tr>
<td>YF503</td>
<td>chd1Δ</td>
<td>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; chd1Δ::KANMX</td>
<td>(Winzeler et al., 1999)</td>
</tr>
<tr>
<td>YSB3090</td>
<td>hst3Δ hst4Δ chd1Δ</td>
<td>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; hst4Δ::NATMX; hst3Δ::HIS3; chd1Δ::KANMX</td>
<td>This study</td>
</tr>
<tr>
<td>YSB3074</td>
<td>FPR snf5Δ</td>
<td>MATα; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; ppt1Δ::mCherry::NAT; sut129Δ::YFP; snf5Δ::KANMX</td>
<td>This study</td>
</tr>
<tr>
<td>ID</td>
<td>Name</td>
<td>Sequence 5' to 3'</td>
<td>Use in study</td>
</tr>
<tr>
<td>------</td>
<td>------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>2805</td>
<td>PSP_DP_F</td>
<td>TCCTTCGTTCTTTTTTCAAG TGCCCTTGCGCGATCTGCTG TGTGTTTTGACATGAGCT CGTTTTTCGACACTGG</td>
<td>Replacing PPT1/SUT129 promoter with delitto perfetto CORE element to generate YSB2757.</td>
</tr>
<tr>
<td>2806</td>
<td>PSP_DP_R</td>
<td>AATTCTGGCTGAGACAATTA TCAGGAACACCTTCTAGTCCTC AAAAGTCTCTCTACATTTC TACCATTAAAGTGC</td>
<td>Replacing PPT1/SUT129 promoter with delitto perfetto CORE element to generate YSB2757.</td>
</tr>
<tr>
<td>2807</td>
<td>Pinv1F</td>
<td>CATTTCCTTCGTTCTTTTCCTA AAGTGCCCTTGCGCGATCTG CTGCTGTGGGTGTTGACATTT CTTCTTTATGAAA</td>
<td>Replacing CORE element of YSB2757 with inverted PPT1/SUT129 promoter using FOA counter selection to generate YSB2764.</td>
</tr>
<tr>
<td>2808</td>
<td>Pinv1R</td>
<td>TCTGGCTGAGACAATTTATCA GGAACACCTCTCTAGTCCTCAA AAGTTCCTCCTATTTCGA GATATTTGATTAACAGGCA</td>
<td>Replacing CORE element of YSB2757 with inverted PPT1/SUT129 promoter using FOA counter selection to generate YSB2764.</td>
</tr>
<tr>
<td>2803</td>
<td>e60URA3F</td>
<td>CGTATGCGATGCAACTAA AAGGATACCAATCTAAAT ATTATAATATTATTTATC TTAAAGGAAACGAAGATAAATCATGTCGA</td>
<td>Replacing CUT60 with URA3.</td>
</tr>
<tr>
<td>2804</td>
<td>e60URA3R</td>
<td>GACGGCCCCCTCTAGACGA CTCCATAAAGCGCTTTGTATTG GCTGGGCTATCTATTTTCTGTA AAGTTAGTTTGGCTGCGATCCTTTCT</td>
<td>Replacing CUT60 with URA3.</td>
</tr>
<tr>
<td>2565</td>
<td>S129toUR A3fwd</td>
<td>ACGGAAAAGAATAACATAC ATACAATGATTGAAAATAATTTTCTAAAGAAGAAGAAAGGAAACGAGATAAACATGTCGA</td>
<td>Replacing SUT129 with URA3 as in YSB2759.</td>
</tr>
<tr>
<td>2566</td>
<td>S129toUR A3rev</td>
<td>AGGAAATAGAGTGCTACA AAAAAATGACGATTACTTCTTTATTCGTTCTATTATAGTTTTGCTG GCCGCATCTTTCT</td>
<td>Replacing SUT129 with URA3 as in YSB2759.</td>
</tr>
<tr>
<td>Line</td>
<td>Description</td>
<td>Sequence</td>
<td>Note</td>
</tr>
<tr>
<td>------</td>
<td>------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>2801</td>
<td>S129toC60 F</td>
<td>ACGGAAAAAGAAATACATACATACAAATGATTTGAAAAATATTTTCTATAAAGAAAGACTATTACACGAAAGACAAAGTTGGAATTGGAAT</td>
<td>Replacing UT129 with CUT60 as in sut129Δ::URA3 strains using FOA counter selection.</td>
</tr>
<tr>
<td>2802</td>
<td>S129toC60 R</td>
<td>AGGAAAAATGAGTGTGCTACAAAAATGACGATTACTTTTTTTCGCTATGATGATTACAGGCGTTGACCAGGAGGTTGTAAT</td>
<td>Replacing SUT129 with CUT60 as in sut129Δ::URA3 strains using FOA counter selection.</td>
</tr>
<tr>
<td>2563</td>
<td>PPT1toHIS 3Fwd</td>
<td>CCCACAGCTTTTCTGCTTTTGATTAAATCTAGCAAAATAGAAAGGACATGACTACATAAGAAGACCTTTTGG</td>
<td>Replacing PPT1 with HIS3 as in YSB2737.</td>
</tr>
<tr>
<td>2564</td>
<td>PPT1toHIS 3rev</td>
<td>TTAGAGGCAAGAAATAACGCTTGCCTTGGTTATAATCAATAATAATCTCGAATATGACAAGCAGGAAAGG</td>
<td>Replacing PPT1 with HIS3 as in YSB2737.</td>
</tr>
<tr>
<td>2570</td>
<td>SUT129_term_exch_clo_R</td>
<td>AAGTCGACTCGCTCTCTGCCACAGCAGGAAGGAGAACCTTTTGG</td>
<td>Amplification of genomic region containing sut129Δ::URA3 ppt1Δ::HIS3 replacements of YSB2737. Cloning of SB1603.</td>
</tr>
<tr>
<td>2571</td>
<td>PPT1-3utr_top</td>
<td>CGGTATTACTCGAGCCGTA</td>
<td>Amplification of genomic region containing sut129Δ::URA3 ppt1Δ::HIS3 replacements of YSB2737. Cloning of SB1603.</td>
</tr>
<tr>
<td>3169</td>
<td>HST3delHI S3-F</td>
<td>TGGTGAAAAACTGAAGGGAATTAATTCATACAAGTAGATCCATCTTCTCAAAAATGAGAGGACAAAAAGGCTCTTGG</td>
<td>Replacing HST3 with HIS3 as in hst3Δ::HIS3.</td>
</tr>
<tr>
<td>3170</td>
<td>HST3delHI S3-R</td>
<td>AGGGGTGTCTCTGCTATCTCCGATTTATCGCTAAACTCAAATTTATAATGTAAATGCTACATAAGAGACACCTGGTTG</td>
<td>Replacing HST3 with HIS3 as in hst3Δ::HIS3.</td>
</tr>
<tr>
<td>2261</td>
<td>MX Switch Forward</td>
<td>GCGCATGGAGGCCGCAAAATACAGATCC</td>
<td>Switching between KANMX and NATMX markers in double mutant construction</td>
</tr>
<tr>
<td>2262</td>
<td>MX Switch Reverse</td>
<td>GGCGCGGTATATCATGACGTTTATTAGATCAGTTTATAGCTTTAAAATCAGTTTACAGTTTATTTAACATTTTATTAGCTTTTATAC ACC</td>
<td>Switching between KANMX and NATMX markers in double mutant construction</td>
</tr>
<tr>
<td>2471</td>
<td>S129/PPT1 P-F</td>
<td>TACCAAGGCAAGCGTTATTTC</td>
<td>Genotyping PPT1/SUT129 promoter inversion</td>
</tr>
<tr>
<td>2476</td>
<td>S129_3'-R</td>
<td>ATATCGCCCGCTCATTGTTT</td>
<td>Genotyping PPT1/SUT129 promoter inversion</td>
</tr>
<tr>
<td>Position</td>
<td>Tid</td>
<td>Primer/Sequence</td>
<td>Notes</td>
</tr>
<tr>
<td>----------</td>
<td>-----</td>
<td>-----------------</td>
<td>-------</td>
</tr>
<tr>
<td>266</td>
<td>URA3qPC R-F</td>
<td>GCAGAATGGGCAGACATTAC</td>
<td>Genotyping CUT60 to URA3 replacement with 2452.</td>
</tr>
<tr>
<td>2472</td>
<td>S129/PPT1 P-R</td>
<td>TTCGTTAAACGCTAAACC</td>
<td>Genotyping SUT129 replacements with 2475.</td>
</tr>
<tr>
<td>2475</td>
<td>S129_3'-F</td>
<td>GGCGGGCGTAAAATAAAATAC</td>
<td>Genotyping SUT129 replacements with 2472.</td>
</tr>
<tr>
<td>3166</td>
<td>HST3-A</td>
<td>AAAGTTCATCACCTTCTTAGCACC</td>
<td>Genotyping HST3</td>
</tr>
<tr>
<td>3168</td>
<td>HST3-D</td>
<td>TTTTGTAGACTGTTAAAGAGC</td>
<td>Genotyping HST3</td>
</tr>
<tr>
<td>2877</td>
<td>Xrn1-A</td>
<td>ACAAAACAAGAAGGTTAGAAGCA</td>
<td>Amplification and genotyping of xrn1Δ::KANMX</td>
</tr>
<tr>
<td>2879</td>
<td>Xrn1-D</td>
<td>AATGAGATCAATGAGAAGAGTGC</td>
<td>Amplification and genotyping of xrn1Δ::KANMX</td>
</tr>
<tr>
<td>1055</td>
<td>RRP6_up</td>
<td>GCAGAATTCACAGAAAACACTTGAG</td>
<td>Amplification and genotyping of rrp6Δ::KANMX</td>
</tr>
<tr>
<td>1056</td>
<td>RRP6_down</td>
<td>GAGCAAGCTTTCTTCTACTTAAGTGGAG</td>
<td>Amplification and genotyping of rrp6Δ::KANMX</td>
</tr>
<tr>
<td>3036</td>
<td>RLF2-5'</td>
<td>ATCAAGTGGTTCCCTACCTG</td>
<td>Amplification and genotyping of cac1Δ::KANMX</td>
</tr>
<tr>
<td>3037</td>
<td>RLF2-3'</td>
<td>CGGCGAACTAGCATTAAATT</td>
<td>Amplification and genotyping of cac1Δ::KANMX</td>
</tr>
<tr>
<td>3038</td>
<td>CAC2-5'</td>
<td>GTGGGGCCTTTTTGTCTGTTT</td>
<td>Amplification and genotyping of cac2Δ::KANMX. Cloning of SB1705.</td>
</tr>
<tr>
<td>3039</td>
<td>CAC2-3'</td>
<td>AACCGTACAGGATGCCAACCG</td>
<td>Amplification and genotyping of cac2Δ::KANMX. Cloning of SB1705.</td>
</tr>
<tr>
<td>3022</td>
<td>MSI1-5'</td>
<td>ATGTGAACAAATCGCTCTTA</td>
<td>Amplification and genotyping of cac3Δ::KANMX</td>
</tr>
<tr>
<td>3023</td>
<td>MSI1-3'</td>
<td>GAAACTCGGTATAAGAAGTAT</td>
<td>Amplification and genotyping of cac3Δ::KANMX</td>
</tr>
<tr>
<td>3197</td>
<td>rtt106-A</td>
<td>GATTGAAAACATTCAACAAAAGGAAT</td>
<td>Amplification and genotyping of rtt106Δ::KANMX</td>
</tr>
<tr>
<td>3198</td>
<td>rtt106-D</td>
<td>ATGACTCGAACTTGATAATGAAAG</td>
<td>Amplification and genotyping of rtt106Δ::KANMX</td>
</tr>
<tr>
<td>3160</td>
<td>HST4-A</td>
<td>CAGGATTTTCTTCTGACTTACGTAT</td>
<td>Amplification and genotyping of hst4Δ::KANMX</td>
</tr>
<tr>
<td>3162</td>
<td>HST4-D</td>
<td>GGAATAATAGATTTGTCTAAGGGA</td>
<td>Amplification and genotyping of hst4Δ::KANMX</td>
</tr>
<tr>
<td>3226</td>
<td>SNF5-A</td>
<td>CTCTAAATCTCTCTCAACCATTGAA</td>
<td>Amplification and genotyping of snf5Δ::KANMX</td>
</tr>
<tr>
<td>3227</td>
<td>SNF5-D</td>
<td>AGTTTGTGTCTCTAGTTCTCGTCCT</td>
<td>Amplification and genotyping of snf5Δ::KANMX</td>
</tr>
</tbody>
</table>
Table C-2 (Continued)

<p>| 1333 | Rsc1 F (809) | ATTGCTCCCTACCCCTCTCTCGTC | Amplification and genotyping of rsc1Δ::KANMX |
| 1334 | Rsc1 R (3934) | CTGGACCGCGTAGATAAATAGTAAGC | Amplification and genotyping of rsc1Δ::KANMX |
| 3212 | CHD1-A | CTGACCACGGAAACATTGTGAG | Amplification and genotyping of chd1Δ::KANMX |
| 3213 | CHD1-D | GTAAACAGAAACCTCTTCTTAGTGG | Amplification and genotyping of chd1Δ::KANMX |
| 2453 | SUT129_F | TGTGGAAACCCCAACAAATTCTC | SUT129 northern probe amplification |
| 2454 | SUT129_R | ATCGCCGCTCATTGTTCTATC | SUT129 northern probe amplification and labeling |
| 2896 | CUT060_2 F | CTATTACACGAAAGACAAGGTGTGAAATG | CUT60 northern probe amplification |
| 2897 | CUT060_2 R | TACAAATTAGCTATTATTACGTTTGAAATG | CUT60 northern probe amplification and labeling |
| 2509 | SUT477_F | AAAACAGCTGGGCTCTCTTCTCTCTC | SUT477 northern probe amplification and labeling |
| 2510 | SUT477_R | TAAGGCGCAATCTACCAGGA | SUT477 northern probe amplification |
| 3181 | Kip1-divF | GATGGCGCTATGGTAAGGTCTCT | KIP1div northern probe amplification |
| 3182 | Kip1-divR | GACCGTGCTCGCTATTCAGAA | KIP1div northern probe amplification and labeling |
| 3185 | MUB1-divF | TGAACACCCATACAATACAT | MUB1div northern probe amplification |
| 3186 | MUB1-divR | TGTTATCAGTAATCGGACTGCTACTGGTAATG | MUB1div northern probe amplification and labeling |
| 2985 | SCR1_F | AGGCTGTAATGGCTTCTGGTGGA | ScR1 northern probe amplification |
| 2986 | SCR1_R | ACGGGCTGGGCAAGCATGGA | ScR1 northern probe amplification and labeling |
| 2470 | PPT1_5'-R | TCGAATATGTCACACCTCTCTA | PPT1 northern probe amplification |
| 2467 | PPT1a_F | GCCTCAGATAGCTTTTCTAAAA | PPT1 northern probe amplification and labeling |
| 2586 | URA3-668-PstI | TATCTGCAAGAAAGCCATTACGCCAAGTAAAT | URA3 northern probe amplification |
| 2588 | URA3-230-XbaI-F | TATCTGCAAGAAAGCCATTACGTTGGAATAAACGC | URA3 northern probe amplification and labeling |
| 3200 | KIP1-F2 | TGGGATGCACACTTAAGGC | KIP1 northern probe amplification |
| 3201 | KIP1-R2 | TGTACATTTGTACACCGTGGCAACAT | KIP1 northern probe amplification and labeling |</p>
<table>
<thead>
<tr>
<th>Line</th>
<th>amplicon Name</th>
<th>Forward Sequence</th>
<th>Reverse Sequence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3187</td>
<td>MUB1-F</td>
<td>TCGGAAGAATTGTCCCCAAA CAAGA</td>
<td>MUB1 northern probe amplification</td>
<td></td>
</tr>
<tr>
<td>3188</td>
<td>MUB1-R</td>
<td>AGAACGGGACAGTGGTTATT ACATAT</td>
<td>MUB1 northern probe amplification and labeling</td>
<td></td>
</tr>
<tr>
<td>3006</td>
<td>ACT-qPCR-F</td>
<td>TGGATTCGCGTGATGGTTT</td>
<td>ACT1 control amplicon for ChIP assays</td>
<td></td>
</tr>
<tr>
<td>3007</td>
<td>ACT-qPCR-R</td>
<td>AAATGGCGTGAGGAGAG AAAC</td>
<td>ACT1 control amplicon for ChIP assays</td>
<td></td>
</tr>
<tr>
<td>2461</td>
<td>PPT1_3'F</td>
<td>ATCTGGGTGCTCAACAGCTT</td>
<td>PPT1 amplicon for ChIP assays</td>
<td></td>
</tr>
<tr>
<td>2462</td>
<td>PPT1_3'R</td>
<td>GCAATTGAGGCGAGGGTTA</td>
<td>PPT1 amplicon for ChIP assays</td>
<td></td>
</tr>
<tr>
<td>2475</td>
<td>S129_3'-F</td>
<td>GGGCCGGCTAAATTTTAAATA C</td>
<td>SUT129 amplicon for ChIP assays</td>
<td></td>
</tr>
<tr>
<td>2476</td>
<td>S129_3'-R</td>
<td>ATATCGCCGCTCAATTGTTCT</td>
<td>SUT129 amplicon for ChIP assays</td>
<td></td>
</tr>
<tr>
<td>518</td>
<td>ADH1_844</td>
<td>TTCAACCAAGTGCAGTCAAGTCC ATCTCTA</td>
<td>ADH1 reference amplicon for qPCR</td>
<td></td>
</tr>
<tr>
<td>519</td>
<td>ADH1_1013</td>
<td>ATTTGACCCCTTTCCATCTTT TCGTAA</td>
<td>ADH1 reference amplicon for qPCR</td>
<td></td>
</tr>
<tr>
<td>2983</td>
<td>mCherry_qPCR_F</td>
<td>AAGACCATGGGCTGGGA</td>
<td>mCherry amplicon for qPCR</td>
<td></td>
</tr>
<tr>
<td>2868</td>
<td>mCherry_5'_out</td>
<td>TCAGCTTCAGCCTCTGCT</td>
<td>mCherry amplicon for qPCR</td>
<td></td>
</tr>
<tr>
<td>2989</td>
<td>YFP_qPCR_F2</td>
<td>AGTCCATGGCCAACCTTGAAG</td>
<td>YFP amplicon for qPCR</td>
<td></td>
</tr>
<tr>
<td>2990</td>
<td>YFP_qPCR_R2</td>
<td>TGACTTCAGCTCGTCTGTTAT</td>
<td>YFP amplicon for qPCR</td>
<td></td>
</tr>
<tr>
<td>2810</td>
<td>PJ1R</td>
<td>AACGCCGCCATCCAGTGTGCG AACCTAGGTGCTAGCCTTTTCT ATTTTGC</td>
<td>Amplification of backbone containing flanking PPT1 and SUT129 3’ sequences with 2815, used for ITA cloning of FPR.</td>
<td></td>
</tr>
<tr>
<td>2815</td>
<td>PJ4T</td>
<td>TCCGCTAGGGATAACAGGGT AACCCAGGGTGAGGAACTAGGAAAG</td>
<td>Amplification of backbone containing flanking PPT1 and SUT129 3’ sequences with 2810, used for ITA cloning of FPR.</td>
<td></td>
</tr>
<tr>
<td>2871</td>
<td>YFP_start</td>
<td>ATGTCTAAAAGGTGAAGAATT ATTCAC</td>
<td>Amplification of YFP with overlap to SUT129 5’ and 3’ sequences using 2816 for ITA cloning of FPR.</td>
<td></td>
</tr>
<tr>
<td>2816</td>
<td>PJ4B</td>
<td>TTACTTTTTTTTCTGTCTACC CTGGGGTACCCTTTATCCCT AGCGGA</td>
<td>Amplification of YFP with overlap to SUT129 5’ and 3’ sequences using 2871 for ITA cloning of FPR.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td>2809</td>
<td>PJ1T</td>
<td>GCAAAATAGAAAAAGGCATGACTGGATGG CGGCGTT</td>
<td>Amplification of mCherry with overlap to PPT1 5’ and 3’ sequences using 2870 for ITA cloning of FPR.</td>
<td></td>
</tr>
<tr>
<td>2870</td>
<td>mCherry_start</td>
<td>ATGGTGAGCAAGGCGAGGAGGATGG</td>
<td>Amplification of mCherry with overlap to PPT1 5’ and 3’ sequences using 2809 for ITA cloning of FPR.</td>
<td></td>
</tr>
<tr>
<td>2846</td>
<td>mCherry-PPT1/SUT1 29 pro</td>
<td>GAACCTCGTATGGGATGCGCATGGACCCATGGCCCTTGTC ACCATTTCTTTTATTATTATTGTGTTATACCATCATC</td>
<td>Amplification of PPT1/SUT129 promoter with overlap to mCherry and YFP using 2847 for ITA cloning of FPR.</td>
<td></td>
</tr>
<tr>
<td>2847</td>
<td>YFP-PPT1/SUT1 29 pro</td>
<td>CACCACAGGAGAACCCGACCGTACCATC TTTCTTTACATTTAATTTTTTACT</td>
<td>Amplification of PPT1/SUT129 promoter with overlap to mCherry and YFP using 2846 for ITA cloning of FPR.</td>
<td></td>
</tr>
<tr>
<td>2452</td>
<td>CUT_60_R</td>
<td>CCCCTTCTAGACGACTCCA</td>
<td>Genotyping CUT60 to URA3 replacement with 2666.</td>
<td></td>
</tr>
<tr>
<td>Linker-1</td>
<td>AppCTGTAGGCGACCATCAAT/3d dC</td>
<td>Pre-adenylated DNA linker oligo for ligation to nascent RNA in NET-seq protocol.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>oGAB11</td>
<td>agu cac uua ggc aug uac acu gac ugu g</td>
<td>RNA oligonucleotide. Used as control in NET-seq protocol.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSC007</td>
<td>5’/5Phos/ATCTCGTATGGCCTGCTCTTCTGCTTG/iSp18/CACCTA/iSp18/TCGACGATCATGATGGTG CCTACAG 3’</td>
<td>RT-reaction primer used in construction of NET-seq libraries. iSp18 are 18-carbon spacers.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>oNTI231</td>
<td>CAAGCAGAAGACGCCATACG A</td>
<td>For PCR amplification of circularized NET-seq libraries with one of the illumina barcoding primers.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barcode</td>
<td>Sequence</td>
<td>Amplification and Barcoding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
<td>-----------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barcode1</td>
<td>AATGATACGGCGACCACCGA GATCTACAGATCGGAAGAGC ACACGTCTGAACCTCCAGTCAC ATCACGTCCGACGATCATTTGA TGG</td>
<td>For amplification and barcoding of circularized NET-seq sequencing libraries by PCR with oNTI231.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barcode2</td>
<td>AATGATACGGCGACCACCGA GATCTACAGATCGGAAGAGC ACACGTCTGAACCTCCAGTCAC CGATGTTCCGACGATCATTTGA TGG</td>
<td>For amplification and barcoding of circularized NET-seq sequencing libraries by PCR with oNTI231.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barcode3</td>
<td>AATGATACGGCGACCACCGA GATCTACAGATCGGAAGAGC ACACGTCTGAACCTCCAGTCAC TTAGGCTCCGACGATCATTTGA TGG</td>
<td>For amplification and barcoding of circularized NET-seq sequencing libraries by PCR with oNTI231.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barcode4</td>
<td>AATGATACGGCGACCACCGA GATCTACAGATCGGAAGAGC ACACGTCTGAACCTCCAGTCAC TGACCATCCGACGATCATTTGA TGG</td>
<td>For amplification and barcoding of circularized NET-seq sequencing libraries by PCR with oNTI231.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barcode5</td>
<td>AATGATACGGCGACCACCGA GATCTACAGATCGGAAGAGC ACACGTCTGAACCTCCAGTCAC ACAGTGTTCCGACGATCATTTGA TGG</td>
<td>For amplification and barcoding of circularized NET-seq sequencing libraries by PCR with oNTI231.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barcode6</td>
<td>AATGATACGGCGACCACCGA GATCTACAGATCGGAAGAGC ACACGTCTGAACCTCCAGTCAC GCCAATTCGACGATCATTTGA TGG</td>
<td>For amplification and barcoding of circularized NET-seq sequencing libraries by PCR with oNTI231.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barcode7</td>
<td>AATGATACGGCGACCACCGA GATCTACAGATCGGAAGAGC ACACGTCTGAACCTCCAGTCAC CAGATCTCCGACGATCATTTGA TGG</td>
<td>For amplification and barcoding of circularized NET-seq sequencing libraries by PCR with oNTI231.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barcode8</td>
<td>AATGATACGGCGACCACCGA GATCTACAGATCGGAAGAGC ACACGTCTGAACCTCCAGTCAC ACTTGATCCGACGATCATTTGA TGG</td>
<td>For amplification and barcoding of circularized NET-seq sequencing libraries by PCR with oNTI231.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table C-3: Plasmids used in this study.

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Name</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>SB1603 (SMC12)</td>
<td>pCR-HIS3-PPT1/SUT129pro-URA3</td>
<td>PPT1/SUT129 promoter driving HIS3 in the PPT1 direction and URA3 in SUT129 direction, Zeocin&lt;sub&gt;R&lt;/sub&gt;, Kan&lt;sub&gt;R&lt;/sub&gt;.</td>
</tr>
<tr>
<td>F1048 (MSP79)</td>
<td>HO-TDH3p-YFP-ADH1t-TEFp-KAN-TEFt-HO</td>
<td>hoΔ::YFP::KANMX, Kan&lt;sub&gt;R&lt;/sub&gt; YFP is codon optimized for yeast</td>
</tr>
<tr>
<td>F1047 (MSP80)</td>
<td>HO-TDH3p-mCherry-ADH1t-TEFp-NAT-TEFt-HO</td>
<td>hoΔ::mCherry::NATMX, Kan&lt;sub&gt;R&lt;/sub&gt; mCherry is codon optimized for yeast</td>
</tr>
<tr>
<td>SB1629 (SMC50)</td>
<td>pCR-mCherry-PPT1/SUT129pro-YFP (FPR)</td>
<td>PPT1/SUT129 promoter driving mCherry::NAT in PPT1 direction and YFP in SUT129 direction, Zeocin&lt;sub&gt;R&lt;/sub&gt;, Kan&lt;sub&gt;R&lt;/sub&gt;.</td>
</tr>
<tr>
<td>F1092 (SMC60) (MSP-A17V).</td>
<td>HO-TDH3p-BFP-ADH1t-TEFp-KAN-TEFt-HO</td>
<td>hoΔ::BFP::KANMX, AMP&lt;sub&gt;R&lt;/sub&gt;</td>
</tr>
<tr>
<td>SB1705 (SMC70)</td>
<td>pCR-cac2Δ::KANMX</td>
<td>cac2Δ::KANMX, Kan&lt;sub&gt;R&lt;/sub&gt;</td>
</tr>
<tr>
<td>SB1603 (SMC12)</td>
<td>pCR-HIS3-PPT1/SUT129pro-URA3</td>
<td>PPT1/SUT129 promoter driving HIS3 in the PPT1 direction and URA3 in SUT129 direction, Zeocin&lt;sub&gt;R&lt;/sub&gt;, Kan&lt;sub&gt;R&lt;/sub&gt;.</td>
</tr>
<tr>
<td>F1048 (MSP79)</td>
<td>HO-TDH3p-YFP-ADH1t-TEFp-KAN-TEFt-HO</td>
<td>hoΔ::YFP::KANMX, Kan&lt;sub&gt;R&lt;/sub&gt; YFP is codon optimized for yeast</td>
</tr>
<tr>
<td>F1047 (MSP80)</td>
<td>HO-TDH3p-mCherry-ADH1t-TEFp-NAT-TEFt-HO</td>
<td>hoΔ::mCherry::NATMX, Kan&lt;sub&gt;R&lt;/sub&gt; mCherry is codon optimized for yeast</td>
</tr>
<tr>
<td>SB1629 (SMC50)</td>
<td>pCR-mCherry-PPT1/SUT129pro-YFP (FPR)</td>
<td>PPT1/SUT129 promoter driving mCherry::NAT in PPT1 direction and YFP in SUT129 direction, Zeocin&lt;sub&gt;R&lt;/sub&gt;, Kan&lt;sub&gt;R&lt;/sub&gt;.</td>
</tr>
<tr>
<td>F1092 (SMC60) (MSP-A17V).</td>
<td>HO-TDH3p-BFP-ADH1t-TEFp-KAN-TEFt-HO</td>
<td>hoΔ::BFP::KANMX, AMP&lt;sub&gt;R&lt;/sub&gt;</td>
</tr>
<tr>
<td>SB1705 (SMC70)</td>
<td>pCR-cac2Δ::KANMX</td>
<td>cac2Δ::KANMX, Kan&lt;sub&gt;R&lt;/sub&gt;</td>
</tr>
</tbody>
</table>
C.2. Supplementary figures

**Figure C-1. Experimental strategy of carbon source shift experiments and quantification of divergent transcription presented in Figure 4-1**

(A) Schematic representation of growth regime used in Figure 4-1 and Figure C-1B. Initial growth was in raffinose (Raf) media to midlogarithmic growth, then in galactose (Gal) media, then in glucose (Glu) media. Samples for RNA analysis were taken at the indicated times.

(B) Northern blot analysis of divergent transcripts from the PPT1/SUT129 promoter in xrn1Δ cells. Cells were grown as in Figure 4-1B and part A. Asterisk (*) denotes the ScR1 loading control. (C) Divergent transcript levels at each time point were normalized to the corresponding loading control and expressed as relative to the repressed Gal 120 time point. Strains analyzed are indicated at top, time points and carbon source indicated below the graph. PPT1 is shown in red, SUT129 in orange and URA3 in gray. Means of 3 repeats are shown ± SEM. Note that the black arrow indicating transcription in the PPT1 direction drives SUT129 in the inverted promoter strain schematic. Left two panels correspond to Figure 4-1B, third panel to Figure C-1B, and last panel to Figure 4-1D.
Figure C-2. FACS data processing and systematic analysis of FPR screen

(A) Data filtering by time. Fluorescence of mCherry was plotted against the time passed from sample injection until detection of the cellular fluorescence event to show cells selected for further analysis (red). Data of the remaining cells (black circles) are discarded. (B) Filtering by cell shape and size. A histogram of the forward scatter against the side scatter is shown. The forward scatter correlates with cell shape, the side scatter correlates with cell size. The red area was chosen based on the median side scatter value of each individual analyzed 96-well plate in the screen. Cells in the red area were selected for further analysis and the remaining cells discarded (black). (C) Representation of data points remaining after filtering to calculate the median fluorescence values used for analysis. Scatter plot of YFP fluorescence against mCherry fluorescence. The red circles represent cells remaining after filtering to determine the median YFP and mCherry fluorescence values of a given well, black circles represent discarded cells. (D) Principal component analysis of screen data. Two principal components capture the main axis of variation of the data. PC1 (green) represents the correlation of transcription driven divergently from PPT1/SUT129. PC2 describes a vector orthogonal to PC1. PC2 values describe deviation from the correlated expression and were used to isolate and rank candidate outlier mutations. As in Figure 4-2 C, the data from the PCA analysis were used to generate this plot. However the distribution was rotated to have PC1/correlation as the x axis to better visualize deviations from this axis. The distance from PC1 gives the PC2 value (directionality) and is used for ranking candidate mutations. A positive directionality score indicates a reduced mCherry/YFP ratio, a negative directionality score indicates an increased mCherry/YFP ratio. (E) The top two gene ontology terms for candidate lists of varying size were identified based on directionality score. The indicated number of genes ranked by directionality score was used as input for GO-term analysis (http://go.princeton.edu/cgi-bin/GOTermFinder). For example, “highest 150” indicates the 150 gene identifiers with the highest directionality score. Note the term “nucleosome assembly” is the most significantly enriched term for all tested lists with a high directionality score. No consistent enrichment of any GO term is found for various list sizes of gene deletions with a low directionality score.
Figure C-2 (Continued)

A. Filtering by time

B. Filtering by size and shape of cells

C. Filtering of data

D. Screen data rotated to visualize deviations from the correlation axis based on the directionality score.

E. Table of selected data after filtering:

<table>
<thead>
<tr>
<th>Number of deletions</th>
<th>GO ID rank 1</th>
<th>GO ID rank 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TERM</td>
<td>p-value</td>
</tr>
<tr>
<td>highest 150</td>
<td>nucleosome assembly</td>
<td>0.00565399</td>
</tr>
<tr>
<td>highest 200</td>
<td>nucleosome assembly</td>
<td>0.0029221</td>
</tr>
<tr>
<td>highest 227 (5%)</td>
<td>nucleosome assembly</td>
<td>2.36E-05</td>
</tr>
<tr>
<td>highest 250</td>
<td>nucleosome assembly</td>
<td>3.33E-06</td>
</tr>
<tr>
<td>highest 300</td>
<td>nucleosome assembly</td>
<td>2.02E-05</td>
</tr>
<tr>
<td>lowest 150</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>lowest 200</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>lowest 227 (5%)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>lowest 250</td>
<td>single-organism phenotypes</td>
<td>3.0039210</td>
</tr>
<tr>
<td>lowest 300</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Figure C-3. CAF-I is a strong candidate in the FPR screen and re-combining mutations in CAF-I with FPR confirms the findings of the original screen. (A) The “highest 250” list of candidate deletion mutants (Figure C-2 E) was subjected to GO-term analysis (http://go.princeton.edu/cgi-bin/GOTermFinder) and the GO-term space containing enriched categories at different statistical significances was color coded ($10^{-2} > p \text{ value} > 10^{-4}$ in blue, $10^{-4} > p \text{ value} > 10^{-6}$ in light blue). Grey boxes display the yeast gene identifiers for chromatin-related factors among the highest 250 that are found within the box above. Common names appear below the gray boxes. (B) Data from the original high-throughput reverse genetic screen showing the reduction of mCherry/YFP in CAF-I mutants. The cac3Δ yeast deletion corresponds to well C2 on deletion collection plate 10. The cac1Δ yeast deletion corresponds to well H8, the cac2Δ yeast deletion corresponds to well C9 on deletion collection plate 34. The fluorescence data for each well are given relative to the average of the entire plate containing that deletion. The SGA-parent strain represents the parental FPR strain before crossing to the deletion library via meiotic recombination. The median values for the mCherry/YFP ratio (gray bars, log2 scale) are given. Note that the mCherry/YFP ratio is also reduced in cac2Δ, although this deletion did not meet the cut-off chosen for candidate mutants in part A. (C) The FPR construct was freshly transformed into the indicated CAF-I deletion mutant strains to confirm results of the high-throughput screen. BY4741 is an isogenic wild-type control. The median values for mCherry (red) and YFP (orange) are graphed in the right panel, and the mCherry/YFP ratios (gray, log2 scale) are shown in left panel. Means of at least 4 independent transformants are shown ± SEM. (D) The cacD::KanMx deletion constructs were freshly transformed into a strain containing the FPR construct to confirm results from the high-throughput screen. The median values for mCherry (red), YFP (orange) and mCherry/YFP ratio (gray) are plotted as in part C. Means of at least 4 independent transformants are shown ± SEM. (E) Analysis of FPR mRNA upon deletion of CAF-I (cac3Δ). Levels of mCherry and YFP mRNAs were determined by qPCR, and data are expressed relative to the isogenic wild-type control. The mean values for mCherry (red), YFP (orange), and the mCherry/YFP expression ratio (gray) are graphed as above, with error bars representing ± SEM of at least 3 biological repeats. (F) Separation of cocultured wild-type and CAF-I mutant strains based on BFP expression. Histogram of a representative well used to compute Figure 4-2D. The two populations of cells are distributed around the log10 fluorescence value of 3.75 (BFP-negative) and 4.5 (BFP-positive). The red vertical line illustrates the cutoff between BFP-positive controls (right of line) from BFP-negative cells (left of line). (G) Relative mCherry and YFP fluorescence levels for data shown in Figure 4-2 D. (H) Quantitation of transcript levels shown in Figure 4-3 A and five biological repeats. Bands from Northern blots were quantitated by phosphorimager, normalized to the loading control (ScR1), and expressed relative to the isogenic wild-type control. (I) Quantitation of RNA pol II ChIP data used for the ratios shown in Figure 4-3 B. Crosslinked and sheared chromatin preparations from the indicated strains were immunoprecipitated with antibodies against RNA pol II subunit Rpb3. Levels were normalized to a nontranscribed region and expressed relative to the wild-type control strain. Asterisks marking bar graphs in Figure (B-I) denote statistical significant differences from the isogenic wild-type controls with $p < 0.05$. 
Figure C-3 (Continued)

- Northern blot quantification
- **p-value:** 0.0
- **G1/S transition** of mitotic cell transcription regulation of
  involved in positive cycle SWI4
  \[ \Delta HPC2 YAL040C YJR140C (1e-4 to 1e-2) \]
  = PPT1
  = \[ \Delta HPC2 YAL040C YJR140C \]
  = SUT129
  = YFP BY4741 
  = \[ \Delta cac1 \]
  Nucleosome assembly or
  replication-dependent
  \[ \Delta cac2 \]
  Silencing
  \[ \Delta cac3 \]
  \[ \Delta HHT1 YDR225W \]
  Histone exchange
  \[ \Delta HHT1 YDR225W \]
  = mCherry
  BY4741
  = \[ \Delta cac1 \]
  \[ \Delta cac2 \]
  \[ \Delta cac3 \]
  Blue fluorescent signal from cells used for segmentation
  log2 mCherry/YFP fluorescence
  -1.5
  -1.0
  -0.5
  0.0
  0.5
  1.0
  1.5
  2.0
  2.5
  BY4741 cac1\( ^{\Delta} \) cac2\( ^{\Delta} \) cac3\( ^{\Delta} \)
  BY4741 cac1\( ^{\Delta} \) cac2\( ^{\Delta} \) cac3
  BY4741 cac1\( ^{\Delta} \) cac2\( ^{\Delta} \) cac3
  Blue fluorescent signal from cells used for segmentation
  log10 BFP fluorescence
  -1.0
  -0.5
  0.0
  0.5
  1.0
  BY4741 cac1\( ^{\Delta} \) cac2\( ^{\Delta} \) cac3
  BY4741 cac1\( ^{\Delta} \) cac2\( ^{\Delta} \) cac3
  BY4741 cac1\( ^{\Delta} \) cac2\( ^{\Delta} \) cac3
  \[ \Delta cac1 \]
  \[ \Delta cac2 \]
  \[ \Delta cac3 \]

- Reporter gene expression after freshly introducing FPR into mutant and wild-type
- p-value: 1e-6 to 1e-4
- 1e-4 to 1e-2
- p > 0.1

- Fresh introduction of PPR into mutants and wild-type
- Fresh introduction of CAF-I deletions into FPR strain
Figure C-4 Additional NET-Seq Data, Related to Figure 4-4

(A) NET-Seq of the PPT1 locus (screenshot from Mochiview). Vertical bars show density of NET-Seq reads. Color code for Watson strand: cac2Δ (green) and wildtype (gray), the PPT1 div ncRNA (Sut129, black) is transcribed from this strand. Color code for the Crick strand: cac2Δ (blue) and wild-type (pink), the PPT1 coding transcript (white) is transcribed from this strand. A tRNA downstream of SUT129 is annotated which may interfere with detecting NET-Seq signal specific to SUT129. (B) MA-plots illustrate the reproducibility of the biological repeat NET-Seq library preparations and sequencing results. The data of all four genotypes cluster around the straight mean ratio line (red line) between ± two standard deviations (black dashed lines) in the MA-plots, indicating very good reproducibility. (C) NET-Seq coding/noncoding ratios in cac1Δ versus wild-type cells. For each gene, directionality was calculated as the ratio of coding to divergent noncoding reads (the mean of two biological NET-Seq replicates). The CAF-I effect was measured as the ratio of directionality in the mutant divided by that in wild-type cells, and this was plotted against the wild-type directionality using log2 scale. Changes in the mutant appear as deviations from the zero value on the y axis. The data for cac1Δ are shown; note the downward shift of the loci in the bottom right corner indicating a reduction of the coding/noncoding ratio in cac1Δ. A 1.5-fold cutoff (red dashed line) was chosen to identify loci with a reduced coding/noncoding ratio in cac1Δ. A 1.5-fold cutoff (red dashed line) was chosen to identify loci with a reduced coding/noncoding ratio in cac1Δ. The red dots indicate the positions of MUB1 and KIP1 loci in the data set. (D) Comparison of NET-Seq reads in the coding and noncoding direction in cac1Δ and wild-type cells. Normalized reads (RPKM, mean of two biological replicates) for each coding gene (left panel) and corresponding divergent direction (right panel) were calculated. Values for cac1Δ versus wild-type were plotted, with CAF-I affected loci in red. Inset histogram shows data distributions. (E) NET-Seq coding/noncoding ratios in cac3Δ versus wild-type cells. Analysis was done as in part (C). (F) Comparison of NET-Seq reads in the coding and noncoding direction in cac3Δ and wild-type cells. Analysis was done as in part (D). (G) Quantification of MUB1 (red) and MUB1 div (orange) transcripts relative to the ScR1 loading control. Data are represented as fold change relative to the isogenic wild-type control. Data correspond to image shown in Figure 4E. Means of three biological repeats are shown ± SEM, asterisks (*) denote values that were statistically significantly different from the wild-type with p < 0.05. (H) Quantification of KIP1 (red) and KIP1 div (orange) transcripts relative to the ScR1 loading control, corresponding to the image shown in F. Data are represented as fold change relative to the isogenic wild-type control. Means of three biological repeats are shown ± SEM, asterisks (*) denote values that were statistically significantly different from the wild-type with p < 0.05.
Figure C-4 (Continued)

A

B

C

D

E

F

G

H

Supplemental figure 4

All genes

cac1∆
cac3∆
cac2∆

−6 −4 −2 0 2 4 6 8 10 12

−4 −2 −0 1 2 3 4

wild-type coding

log2 mean

wild-type non-coding

cac1∆ coding

cac1∆ non-coding

Mean

Mean + 2 SD

0HDQ−26'

wild-type biorep 1 (cd/nc)
wild-type biorep 2 (cd/nc)
(wild-type biorep 1 (cd/nc) + wild-type biorep 2 (cd/nc))
wild-type biological repeat 1 vs repeat 2

Mean

Mean + 2 SD

0HDQ−26'

cac3∆ biorep 1 (cd/nc)
cac3∆ biorep 2 (cd/nc)
(cac3∆ biorep 1 (cd/nc) + cac3∆ biorep 2 (cd/nc))
cac3∆ biological repeat 1 vs repeat 2

Mean

Mean + 2 SD

0HDQ−26'

cac1∆ biorep 1 (cd/nc)
cac1∆ biorep 2 (cd/nc)
(cac1∆ biorep 1 (cd/nc) + cac1∆ biorep 2 (cd/nc))
cac1∆ biological repeat 1 vs repeat 2

Mean

Mean + 2 SD

0HDQ−26'

cac2∆ biorep 1 (cd/nc)
cac2∆ biorep 2 (cd/nc)
(cac2∆ biorep 1 (cd/nc) + cac2∆ biorep 2 (cd/nc))
cac2∆ biological repeat 1 vs repeat 2

Mean

Mean + 2 SD

0HDQ−26'

Northern blot quantification of Fig. 4E relative to wild-type

Northern blot quantification of Fig. 4F relative to wild-type

 Reads per 10^6 sequences

PPT1

SUT129/PPT1div

tl(AAU)G

= MUB1

= MUB1div

= KIP1

= KIP1div

*
Figure C-5 CAF-I Mutants Increase Transcription at Divergent Promoters Independent of the Transcript Type and in Parallel to RNA Decay or Early Termination Pathways, Related to Figures 4 and 5

(A) Divergent CUTs and SUTs are similarly affected by CAF-I mutants. Annotations of CUTs or SUTs 50 divergent to a protein coding gene were taken from (Xu et al., 2009). Density of normalized NET-seq reads (RPKM) was computed for each library and the fraction of CUTs (black) or SUTs (gray) increased at least 1.5-fold in CAF-I mutants compared to wild-type is given relative to the total number of 50 NFR divergent CUTs or SUTs. (B) Divergent coding-coding promoters with the configuration shown in schematic at top (as annotated in Xu et al., 2009) were analyzed for directionality using NET-seq data. Normalized NET-seq reads (RPKM, the mean of two biological repeats) across the genes were used to determine the preferred direction (A > B, light orange; or B > A, light blue). Loci that showed at least a 1.5-fold difference in directionality in the cac1Δ, cac2Δ, or cac3Δ mutants are shown in red (C) Northern blotting of KIP1 divergent transcription in NNS pathway mutants. Transcripts (left) and genotypes (top) are indicated. Quantification of KIP1/KIP1 div is shown below the gel images as fold change compared to the corresponding isogenic wild-type strain after normalization to the ScR1 loading control. Means of three biological repeats are shown ± SEM. (D) Northern blotting of KIP1 divergent transcription in ssu72-2. Analysis as in (C). (E) Northern blotting of KIP1 divergent transcription in double mutants between RNA decay pathway mutants and cac2Δ. See (C) for annotations. A shorter exposure for KIP1 div is shown to aid visualization of the effect in xrn1D cac2Δ. In addition, KIP1 (red) and KIP1 div (orange) are quantified relative to the ScR1 loading control in bottom panel, and the ratio is shown above that (gray bars) as fold change relative to the isogenic wild-type control. Note that all KIP1/KIP1 div ratio and KIP1 div means are statistically different from wild-type with p < 0.05. (F) Divergent transcript analysis by Northern blotting in double mutants between CAF-I subunits. Annotations as in (C). (G) Quantification of MUB1 (red) and MUB1 div (orange) transcripts relative to the ScR1 loading control. Data are represented as fold change relative to the isogenic wild-type control. Data correspond to image shown in Figure 4-5C. Means of three biological repeats are shown ± SEM. Note that all MUB1 div means are statistically different from wild-type with p < 0.05. Statistical analyses in Figure C-5 (C-G): asterisks (*) denote values that were statistically significantly different from the wild-type with p < 0.05. Statistically significant differences with p < 0.05 comparing means of single to double mutants are indicated by brackets and marked by asterisks.
Figure C-5 (Continued)

A. Diagram showing the percentage of 5' divergent SUTs and 5' divergent CUTs.

B. Chart illustrating the percentage of 5' divergent SUTs and 5' divergent CUTs.

C. Northern blot quantification of KIP1 locus relative to wild-type.

D. Northern blot quantification of KIP1 locus relative to wild-type.

E. Northern blot quantification of KIP1 locus relative to wild-type.

F. Northern blot quantification of KIP1 locus relative to wild-type.

G. Northern blot quantification of MUB1 and MUB1div.

The A/B ratio of 7 of 659 5' NFR sharing coding/coding pairs is decreased by at least 1.5 fold in all individual CAF-I mutants.

NET-seq reads: A > B in wild-type

The A/B ratio of 7 of 659 5' NFR sharing coding/coding pairs is decreased by at least 1.5 fold in all individual CAF-I mutants.
Figure C-6 The H3K56 Acetylation-Dependent Nucleosome Incorporation Pathway Enforces Promoter Directionality at the MUB1/MUB1 div Divergent Promoter in a Non-S-Phase-Specific Manner, Related to Figure 4-6 and Figure 4-7

(A) Analysis of the MUB1/MUB1 div ratio following G1 arrest by a-factor. Expression of indicated transcripts (left) from the designated strains (top) was determined by Northern blotting. Cells were monitored before a-factor arrest (cycling) or during G1-arrest (a-T0). (B) Disruption of the H3K56 acetylation-dependent nucleosome incorporation pathway increases divergent transcription at the MUB1/MUB1 div promoter. Northern blot analysis of indicated divergent transcripts (left) in indicated genotypes (top). Bottom panels show quantification of MUB1 (red) and MUB1 div (orange) relative to the ScR1 loading control. MUB1/MUB1 div ratio (gray) is shown below the images as fold change compared to the corresponding isogenic wild-type strain (BY4741 or H3). Means of three biological repeats are shown ± SEM. (C) Quantification of KIP1 (red) and KIP1 div (orange) transcripts relative to the ScR1 loading control, corresponding to the image shown in Figure 6C. Data are represented as fold change relative to the isogenic wild-type control. Means of three biological repeats are shown ± SEM. (D) Quantification of KIP1 (red) and KIP1 div (orange) transcripts relative to the ScR1 loading control, corresponding to the image shown in Figure 4-6A. Means of three biological repeats are shown ± SEM. Statistical analyses in Figure C-6 (B-D): asterisks (*) denote values that were statistically significantly different from the wild-type with p < 0.05. Statistically significant differences with p < 0.05 comparing means of single to double mutants are indicated by brackets and marked by asterisks.
Figure C-6 (Continued)

A

B

C

D

Supplemental figure 6

BY4741 cac2 Δ

BY4741 cac2 Δ

cycling α-T0

MUB1div

BY4741

cac2 Δ

rtt106 Δ

H3

H3

cac2 Δ

H3K56A

H3K56A

MUB1

ScR1

0

1

2

3

4

5

BY4741

cac2 Δ

rtt106 Δ

H3

H3

cac2 Δ

H3K56A

H3K56A

MUB1

ScR1

log2 MUB1/MUB1div

log2 MUB1/MUB1div

Northern blot quantification of MUB1 locus relative to wild-type

* = MUB1

*= MUB1div

Northern blot quantification of Fig. 7A relative to wild-type

* = KIP1

*= KIP1div

Northern blot quantification of Fig. 6C relative to wild-type

* = KIP1

*= KIP1div

Northern blot quantification of Fig. 7A relative to wild-type

* = KIP1

*= KIP1div

BY4741

cac2 Δ

rtt106 Δ

H3

H3

cac2 Δ

H3K56A

H3K56A

MUB1

ScR1

0

1

2

3

4

5

6

7

BY4741

cac2 Δ

rtt106 Δ

H3

H3

cac2 Δ

H3K56A

H3K56A

MUB1

ScR1

log2 MUB1/MUB1div

log2 MUB1/MUB1div

Northern blot quantification of MUB1 locus relative to wild-type

* = MUB1

*= MUB1div

Northern blot quantification of Fig. 7A relative to wild-type

* = KIP1

*= KIP1div

Northern blot quantification of Fig. 6C relative to wild-type

* = KIP1

*= KIP1div

Northern blot quantification of Fig. 7A relative to wild-type

* = KIP1

*= KIP1div
Figure C-7 Effects of H3K56 Acetylation-Mediated Nucleosome Exchange and SWI/SNF on Promoter Directionality, Related to Figure 4-7

(A) Deletion of H3K56 deacetylases or H3K6Q elevates divergent noncoding transcription in parallel to CAF-I mutants. Left panel shows Northern blot analysis of indicated divergent transcripts (left) in indicated genotypes (top). Samples were run on a single gel; gaps are where irrelevant lanes were removed. Quantification relative to the ScR1 loading control is shown in right panels. Individual KIP1 (red) and KIP1 div (orange) values are shown at bottom and and KIP1/KIP1 div ratio (gray) is expressed as fold change compared to the corresponding isogenic wild-type strain (BY4741 or H3). Means of three biological repeats are shown ± SEM. (B) The SWI/SNF component Snf5 is required for H3K56 acetylation linked nucleosome exchange-mediated increase of divergent KIP1 div transcription. Top panels show Northern blot analysis of indicated divergent transcripts (left) in indicated genotypes (top). Quantification of KIP1/KIP1 div ratio (gray) is given below the images as fold change compared to the corresponding isogenic wild-type strain. Means of three biological repeats are shown ± SEM. Note the increased KIP1/KIP1 div ratio in snf5Δ hst3/4Δ compared to hst3/4Δ; H3K56Q/snf5Δ compared to H3K56Q, and snf5Δ cac1Δ compared to cac1Δ. (C) Metagene analysis of newly incorporated histones (H3-flag) from (Venkatesh et al., 2012), separated for CAF-I affected (red) and unaffected (gray) loci as determined in B. Shading around lines shows ± SEM. ChIP-chip data were split up into 10 bins as described in Methods, transcription start (TSS) and transcript end (TE) represent bin 3 and bin 8, respectively. These data were used to calculate the ratios in Figure 4-6B. (D) Metagene analysis of pre-existing histones (H3-flag), as in part C. (E) Metagene analysis of H3K56ac, as in part C. (F) Metagene analysis of total H3, as in part C. (G) Metagene analysis of histone exchange (as measured by flag-H3 to myc-H3 ratio) separated into divergent promoters with the protein coding side as preferred direction (log coding/noncoding > 0 from NET-seq data, black) or the noncoding side as preferred direction (log coding/noncoding < 0, green). Error bars represent ± SEM, data representation as in C-F. (H) Metagene analysis of H3K56ac/H3 ratio, analyzed as in (G). (I) Analysis of divergent KIP1 and KIP1 div expression in chromatin remodeling mutants. Northern blot analysis of divergent transcripts (left) in indicated strains (top). Quantification of KIP1/KIP1 div ratio (gray) is given below the images as fold change (log2 scale) compared to the corresponding isogenic wild-type strain. Means of three biological repeats are shown ± SEM, asterisks (*) denote values that were statistically significantly different from the wild-type with p < 0.05.; asterisks above indicate statistically significant differences from hst3/4Δ strain with p < 0.05. (J) Analysis of divergent MUB1 and MUB1 div expression as in (I). (K) Transformation of snf5Δ::KanMx and cac2Δ::KanMx deletion constructs into FPR parental strain. The median values for mCherry (red) and YFP (orange) are shown in right panel. mCherry/YFP ratio (gray) was calculated relative to wild-type and shown in left panel. Means of at least four biological repeats are shown ± SEM. Statistical analyses in Figure C-7: asterisks (*) denote values that were statistically significantly different from the wild-type with p < 0.05. Statistically significant differences with p < 0.05 comparing means of single to double mutants are indicated by brackets and marked by asterisks.
C.3. Supplementary References


