Spring 5-15-2015

Molecular and Computational Methods for Cellular State Control

Drew Groves Michael
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Molecular Cell Biology

Dissertation Examination Committee:
Michael Brent, Chair
Thomas Baranski
Barak Cohen
Gautam Dantas
James Havranek
Joseph Jez

Molecular and Computational Methods for Cellular State Control

by
Drew Groves Michael

A dissertation presented to the
Graduate School of Arts & Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2015
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, I am deeply grateful to my advisor, Michael Brent. He has given me support and mentorship over the past five years and inspired my interest in network biology. Michael took a chance on a molecular biologist over five years ago and shepherded me through the long process of learning to think from a quantitative perspective. He has given me the freedom to succeed in this Ph.D. and the guidance to make sure I got through.

My committee has been invaluable for keeping me on track and I am thankful to each member. Barak Cohen's *cis*-regulation paper of 2009 inspired me to switch from immunology and his advice on the art of science has made this dissertation possible. Gautam Dantas gave me fantastic scientific advice and his scientific creativity has inspired entire figures within this dissertation. Joe Jez's passion for science and love of teaching and mentorship has inspired me during this Ph.D. program. Thomas Baranski kept me grounded and encouraged me to think not only about what was working, but also why other things might not be. Jim Havranek led me through the difficult DNA cloning that consumed the middle years of this Ph.D.

I have many fond memories of time spent in the Center for Genome Sciences and I am grateful to Dr. Jeff Gordon who has built a wonderful, collaborative environment.

I owe a great deal to my best friends Zeke Maier and Brian Haynes. They took me from buzzwords to Python and changed how I think about data. My conversations and work with Holly Brown was a highlight of this Ph.D.

My parents, Don and Jan, inspired my love of science and the natural world. They have unfailingly supported me through every moment of this life. Without them, I would not be here.

My wife Amanda has sacrificed greatly and loved me more. We flew half-way across a continent and drove half-way across a state to be together. Thank you Amanda for your support and love.

Drew Michael

Washington University in St. Louis

May 2015

To my wife Amanda Grace, I wrote your name into the very DNA of this dissertation.

ABSTRACT OF THE DISSERTATION

Molecular and Computational Methods for Cellular State Control

by
Drew Groves Michael
Doctor of Philosophy in Biology and Biomedical Sciences
(Molecular Cell Biology)
Washington University in St. Louis, 2015
Professor Michael Brent, Chair

The control of cellular state has many promising applications, including stem cell biology and regenerative medicine, biofuel production, and gene therapy. This dissertation demonstrates a comprehensive approach to cellular state control at the transcriptional level. We introduce a novel algorithm, NetSurgeon, which utilizes genome-wide gene regulatory networks to identify interventions that will force a cell toward a desired expression state. Following extensive *in silico* validation, we applied NetSurgeon to *S. cerevisiae* biofuel production, generating interventions designed to promote a fermentative state during xylose catabolism. Our selected interventions successfully promoted a fermentative transcriptional state and generated strains with higher xylose import rates, improved xylose integration and increased ethanol production rates. We then step down to a single gene level and exhibit a *cis*-engineering strategy that enables precise expression control. We demonstrate that synthetic promoters can be functionally decomposed into individual components that can be characterized in isolation and used to train a composite model capable of predicting the action of the full system. These findings represent significant progress towards the insertion of orthogonal control circuits into the cell for the control of gene expression. Taken together, this dissertation represents an integrative process of quantitative measurement, modeling, and intervention that comprehensively examines methods for cellular state control at the genome-wide and gene levels.

Chapter 1: Introduction

The advent of microarray technology in 1995 fundamentally changed how biological research is conducted (Schena, Shalon, Davis, & O., 1995). Researchers previously limited to studying dozens of genes could now examine entire genomes in a single experiment. This radical expansion of scope created the need for the discipline of systems biology. Systems biologists have invested massive resources in genome sequencing and systematic connectivity mapping projects, leading to an improved understanding of the cell as a complete system (Gerstein, et al., 2012). However, these maps are primarily focused on a single regulatory level and relatively little research has focused on using these network models for the prediction and manipulation of cellular behavior (Chuang, Hofree, & Ideker, 2010). Now, technical advances within DNA sequencing and synthesis are ushering in a new era of integrative systems and synthetic biology. We are rapidly developing the capacity to read cellular state at multiple levels and to integrate this information into systems level models of cellular connectivity. As these capacities develop, molecular geneticists will be faced with the challenge of using this multi-tiered information to understand how natural variation and interventions within one regulatory level will generate change across the entire system.

This dissertation presents research that manipulates cellular state at the genome wide and individual gene levels. We use genome wide gene regulatory networks to algorithmically identify regulatory interventions that will force a cell towards a desired state. As transcription factors can regulate many genes across the genome, these regulatory interventions offer a highly efficient method to manipulate gene expression. We demonstrate that a single regulatory

deletion can be rationally selected to manipulate the expression of hundreds of genes in the desired direction. These regulatory interventions offer the capacity to access both novel and evolutionarily optimized transcriptional states on a genome wide level. We then integrate our transcriptional measurements with a model of central carbon metabolism to evaluate how changes in transcriptional state cascade into changes in metabolic state. At the gene level we manipulate the *cis*-regulatory DNA of a synthetic promoter library to allow accurate expression prediction with substantially fewer experimental measurements. This *cis*-engineering component represents progress towards a future when the expression of a synthetic DNA construct can be precisely controlled across changes in cellular state and environment.

In Chapter 2 we demonstrate that a gene regulatory network model can be used to rationally engineer cellular transcriptomes. In this situation, we start with an origin expression state, a goal expression state and a gene regulatory network model describing the connectivity of all transcriptional regulators within the genome. We constructed an algorithm, NetSurgeon, to use this data to identify regulatory interventions that will enforce a desired cellular state. The NetSurgeon algorithm simulates the effects of regulator deletion and over-expression to generate a rank ordered list of interventions predicted to force the system towards the desired state. Following extensive *in silico* validation, we applied NetSurgeon to the problem of *S. cerevisiae* biofuel production, identifying interventions designed to promote a fermentative transcriptional state during xylose catabolism. 75% of the evaluated NetSurgeon interventions successfully forced the cells towards a fermentative transcriptional state. In addition, our intervention strains exhibited 31% higher rates of ethanol production and improved xylose import rates by 120%, demonstrating success in the transcriptional and phenotypic levels. This work represents the first quantitative assessment of transcriptome engineering, demonstrating the current state of the field

and integrating transcriptional data with metabolic measurements to provide a comprehensive view of the cellular response to transcriptional interventions.

In chapter 3 we demonstrate that synthetic promoters can be functionally decomposed into individual components which can be characterized in isolation and used to train a composite model capable of predicting the action of the full system. Working within a synthetic promoter library designed to respond to a repressor and an activator, we find that a composite model trained on the two transcription factors in isolation is able to predict 72.8% of the variance in expression within a novel genotype where the two factors are expressed simultaneously. The decomposition process provides evidence for the independent action of the two input transcription factors on promoter output, enabling a significant reduction in the number of measurements required to generate an expression model by avoiding the need to fit interaction terms. This work represents significant progress towards the goal of enabling synthetic promoters to be characterized in isolation and assembled into complex systems according to basal expression and response function requirements.

1.1 Contributions

1. **Application of NetSurgeon, an algorithm enabling rational transcriptome engineering.** We present an algorithm designed to rationally manipulate transcriptomes by simulating the effects of regulator deletion or overexpression within a gene regulatory network. We applied this algorithm to *S. cerevisiae* biofuel production, producing strains with higher rates of xylose consumption and ethanol production. This work is presented in Chapter 2.
2. **Measurement and integrative analysis of the cellular transcriptome and metabolism secondary to transcriptional interventions.** We develop an integrated picture of

cellular response to transcriptional interventions at the transcriptional and metabolic level through RNA sequencing and analytical chemistry. This work is presented in Chapter 2.

3. **Functional decomposition of a synthetic promoter architecture that exhibits independence and allows accurate expression prediction.** We demonstrate a synthetic promoter architecture that enables accurate expression modeling with reduced measurement requirements. This contribution is presented in Chapter 3.

1.2 Collaborative Effort Statement:

The work contained within Chapter 2 is a manuscript that Ezekiel Maier and I contributed equally to and are sharing first authorship. Ezekiel and I jointly designed the research and wrote the paper. I performed the molecular biology, designed the experiments, and analyzed the metabolic data. Ezekiel implemented the NetSurgeon algorithm and analyzed the RNA sequencing data. Ezekiel, Holly Brown and I completed the metabolic data acquisition process.

The work contained within Chapter 3 is a manuscript of which I am the primary author. This work is co-authored with Holly Brown and Prof. Michael Brent. I have executed the majority of the genetic engineering, experiments and modeling within this project. Holly Brown contributed to the cloning of the synthetic promoter library.

Chapter 2: Transcriptome Engineering Promotes a Fermentative Transcriptional State

2.1 Abstract

The rational manipulation of transcriptomes offers the possibility to engineer the cell as a collective unit toward specified goals, revolutionizing medicine and bioengineering. Progress in transcriptome engineering has primarily consisted of experimental approaches that are iterative, slow, and expensive. We have developed a novel algorithm, NetSurgeon, which utilizes genome-wide gene regulatory networks to identify interventions that will force a cell toward a desired expression state. Following extensive *in silico* validation, we applied NetSurgeon to *S. cerevisiae* biofuel production, generating interventions designed to promote a fermentative state during xylose catabolism. Our selected interventions successfully promoted a fermentative transcriptional state in the absence of glucose and generated strains with 120% higher xylose import rates, improved xylose integration into central carbon metabolism by 303%, and increased ethanol production rates by 31%. We conclude by presenting an integrated model of transcriptional regulation and metabolic flux that will enable metabolic engineering efforts to prioritize functional regulators of central carbon metabolism.

2.2 Introduction

The central promise of regulatory systems biology is that a map of the cell's global connectivity will enable us to understand, predict, and rationally manipulate cellular behavior. The manipulation of cellular state has many promising applications, including stem cell biology and regenerative medicine, biofuel production, and gene therapy. Fundamental progress toward the

goal of cellular state control has been advanced via systems biology - the study of cellular behavior as a complete unit, and synthetic biology - a rapidly advancing discipline which aims to design regulatory and effector molecules with defined behaviors. In systems biology, immense resources have been invested in genome sequencing, systematic deletion collections, and massively parallelized data acquisition, leading to network maps and improved understanding of the cell as a complete system (Gerstein, et al., 2012). However, relatively little research has focused on using these network models for the prediction and manipulation of cellular behavior (Chuang, Hofree, & Ideker, 2010). Synthetic biology has focused on creating molecular components that can be placed into a system to modify the transcriptional state of a small number of genes. However, genome-scale regulatory engineering is still rare, with most systems restricted to a small number of regulators and a limited set of controlled targets (Cameron, Bashor, & Collins, 2014). Bridging the gap between these two disciplines, we demonstrate that the integration of functional transcriptional network mapping, gene expression profiling, and computational modeling can be used to rationally engineer cellular state.

Transcriptome engineering focuses on the manipulation of extant cellular networks and regulatory systems to enforce a state associated with a desired cellular phenotype. The use of native cellular regulatory mechanisms and network models enables the investigator to access evolutionarily optimized states and avoid the extensive iteration often associated with the integration of a synthetic regulatory circuit into a host system (Cardinale & Arkin, 2012) (Litcofsky, Afeyan, Krom, Khalil, & Collins, 2012). The majority of transcriptome engineering thus far has taken place within the context of developmental stem cell engineering, with the generation of induced pluripotency being the best example (Takahashi & Yamanaka, 2006).

Since the development of induced pluripotent stem cells, many transcriptional interventions have been identified that move cells at various developmental stages along a specified lineage (Morris & Daley, 2013). However, current strategies for direct lineage conversion are often unable to fully convert cells to the state of the goal cell fate (Cahan, Li, Morris, da Rocha, Daley, & Collins, 2014) (Marro, et al., 2011) (Feng, et al., 2008).

The CellNet algorithm was developed in response to current deficiencies in cellular engineering. CellNet is a network-guided algorithm for determining how completely an engineered cell recapitulates a target cell state and identifying transcriptional interventions to guide further engineering (Cahan, Li, Morris, da Rocha, Daley, & Collins, 2014). CellNet identifies sub-networks within mouse and human cell-type-specific regulatory networks whose expression state is predictive of the cell type. These predictive sub-networks are used as features for classifying novel gene expression profiles according to the cell type they most resemble. In addition, CellNet selects transcription factor (TF) interventions for transcriptome engineering by computing a Network Influence Score for each TF which is the sum of two components: the dysregulation of the regulator weighted by its expression level and the dysregulation of its targets weighted by their expression levels. This approach for target selection intervention was used to guide B cell to macrophage conversion by knocking down B cell regulators (Morris, Cahan, Zhao, San Roman, Shivdasani, & Collins, 2014). The generalizability of this method remains unclear due to the limited number of interventions and evaluations performed. These are exciting demonstrations of the power of transcriptome engineering, but studies in these complex developmental systems are limited by incomplete transcriptional network maps, complex cell culture requirements, and a lack of quantitative phenotypes directly linked to molecular effectors.

These issues have thus far prevented a quantitative assessment of transcriptome engineering efforts.

In order to quantitatively assess the current state of transcriptome engineering and establish benchmarks, we utilized *S. cerevisiae* as a model system. 196 of the 209 transcription factors with an annotated DNA-binding domain in the *S. cerevisiae* genome possess a known DNA binding specificity (Spivak & Stormo, 2012) (Weirauch, et al., 2014) and the genome-wide effect of TF removal on expression has been quantitated through microarray profiling (Hu, Killion, & Iyer, 2007) (Kemmeren, et al., 2014). These data provide us with the ability to generate an accurate network model and to validate our algorithmic approaches. The simplicity of *S. cerevisiae* culture enables quantitative modeling and assessment, with the input/output metabolic function measurable by high-performance liquid chromatography (HPLC) and the transcriptional state of the cell quantitated by RNA sequencing.

We identified the industrially relevant fermentation of the pentose carbohydrate xylose as a prototype application that met all our criteria for the quantitative assessment of transcriptome engineering. Xylose is a component of hemicellulose, a polymer that represents approximately 23% of lignocellulosic biomass and is not efficiently fermented by *S. cerevisiae* into ethanol (Chandel & Singh, 2011). Biochemical research has identified all enzymes required for the integration of xylose into the cell's central carbon metabolism. However, recombinant yeast strains expressing these enzymes, and grown in mixed glucose/xylose cultures, rapidly ferment all available glucose and then undergo a diauxic shift into a respiratory metabolic state. Salusjarvi et al demonstrated through transcriptional and proteomic analysis that cells grown on

xylose exist in a hybrid fermentative/respiratory state (Salusjärvi, Kankainen, Soliymani, Pitkänen, Pentti, & Ruohonen, 2008). The abundance of systems-level data, known metabolic pathways, clear regulatory constraint and quantitative phenotypes enabled us to utilize the transcriptome engineering of xylose metabolism to evaluate the current state of transcriptome engineering.

In this work we present a novel algorithm, NetSurgeon, which is designed to enable transcriptome engineering. We ran this algorithm over genome-wide gene regulatory network maps (GRNs), and assessed its performance at selecting TFs whose deletion or overexpression will move the transcriptional state of the cell toward a desired goal. Following algorithmic validation, we applied the algorithm to engineer a fermentative xylose transcriptional state and assessed global cellular response to our transcriptional interventions by using analytical chemistry. Our results demonstrate that transcriptome engineering can be efficiently guided using network models and reveal the degree of transcriptional control over a quantitative multi-factorial phenotype.

2.3 Results

2.3.1 Algorithmic approach to transcriptome engineering

Our transcriptome engineering method, NetSurgeon, simulates interventions on a transcriptional network model to prioritize those that are likely to move the transcriptional state towards a goal state. Our transcriptome engineering efforts consisted of three steps. First, a map of the network of direct, functional regulation is built (Fig. 2.1.A). Second, starting and goal transcriptional states are defined and NetSurgeon, searches through the all possible interventions (deletion or overexpression) to identify interventions that are likely to move the transcriptional state towards

the goal state (Fig. 2.1.B). Finally, strains containing the predicted best interventions are created and RNA-seq and HPLC are used to quantitatively assay their transcriptional and metabolic state (Fig. 2.1.C).

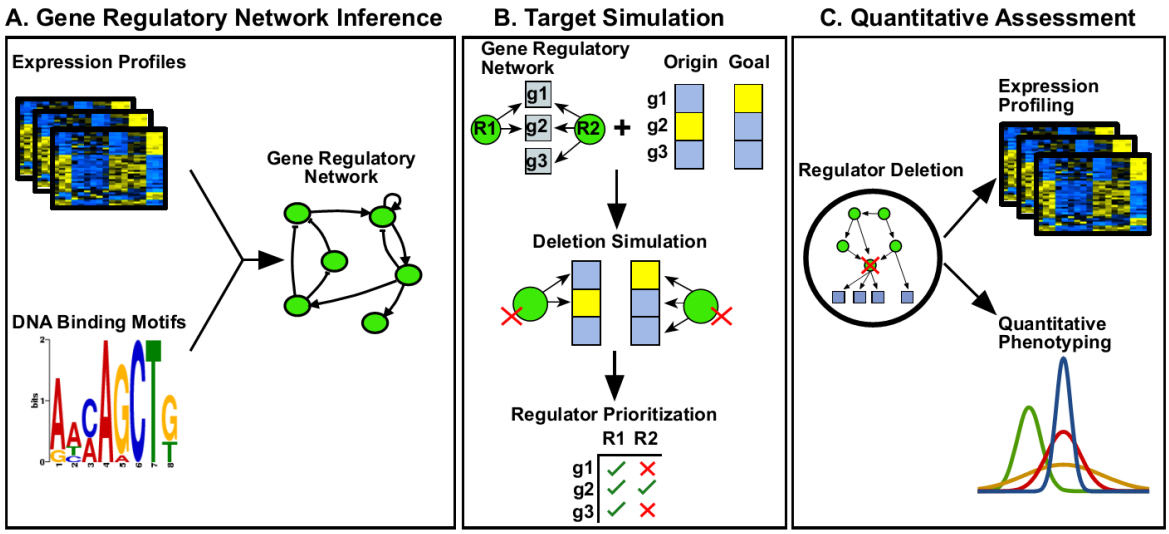

Figure 2.1. Overview of the computational and experimental approaches for rational control of transcriptional state. Panel A: Approach for generation of a gene regulatory network model from DNA binding specificity information and gene expression profiling. Panel B: Approach for target selection through intervention simulation and regulator prioritization. Panel C: Approach for quantitative assessment of intervention effect via RNA sequencing and HPLC metabolite profiling and modeling.

We built an integrated gene regulatory network map by building and combining separate functional and physical maps using NetProphet (Haynes, Maier, Kramer, Wang, Brown, & Brent, 2013). NetProphet is a state-of-the-art GRN mapping algorithm that combines a differential expression (DE) analysis and a co-expression analysis. The physical map was built using a combination of TF binding information from both chromatin immunoprecipitation (ChIP) implicated TF target interactions (Abdulrehman, et al., 2010) (Balaji, Babu, Iyer, Luscombe, & Aravind, 2006) (Harbison, et al., 2004) (Lee, et al., 2002) and TF binding potential estimated by scanning a collection of position weight matrix (PWM) models over all yeast promoters (Spivak & Stormo, 2012). An integrated functional and physical GRN map was built by assigning a score to each TF-target gene pair that was equal to the geometric mean of the

scores assigned to it in the functional and physical networks. The geometric mean ensures that high scoring interactions are supported by both binding and expression evidence.

To select interventions that will shift transcriptional state toward the goal state, we applied NetSurgeon. This algorithm assigns a score to each possible intervention representing its confidence that the intervention will yield a shift toward the goal state. The score assigned to each intervention is based on the number of targets of the regulator that are predicted to move toward the goal state and degree to which the initial and goal states differ for the regulator and targets. Deletion of a TF is predicted to increase the expression of targets it represses and decrease the expression of targets it activates. Conversely, overexpression of a TF is predicted to decrease expression targets that the TF represses and increase expression targets it activates. High-scoring interventions are those that are predicted to change many genes in the right direction, with greater weight given to targets that are the most significantly differentially expressed genes between the initial state and goal state.

2.3.2 Network models can efficiently guide transcriptome engineering efforts

To assess the ability of NetSurgeon to select interventions that will move the initial transcriptional state toward the goal state, we used NetSurgeon to select regulator interventions for single regulator intervention goal states from publically available gene expression datasets. We choose to use independent single regulator intervention expression profiles for validation goal states, rather than randomly generated expression states, because randomly generated expression states may not be biologically achievable. We constructed the GRN used for validation in a similar fashion as previously described, except the functional network was

inferred from only one of the three gene expression datasets previously used, the dataset consisting of 269 regulator deletion strains grown in YPD (Hu, Killion, & Iyer, 2007).

We initially examined NetSurgeon's performance on goal states that we knew could be achieved by a single TF deletion mutant growing in synthetic complete medium (SC). This medium was different from the rich medium (YPD) in which the expression profiles used to build the network were obtained, but the two media featured the same sugar: 2% glucose. The goal states were an independent set of expression profiles from regulator deletion mutants (Kemmeren, et al., 2014). For each of the 245 goal states, NetSurgeon used the NetProphet+PWM network to assign scores to all 320 possible regulator deletions. We plotted the number of goal states for which NetSurgeon ranked the best intervention (the one that actually produced the goal state profile) at or above each rank (Fig 2.2.A, green). We compared this to a random assignment of rankings for each of the deletion goal states, by running NetSurgeon on 100 random networks of the same topology (Fig. 2.2.A, gray). We found that NetSurgeon is able to assign higher scores to the correct interventions compared with ranks assigned by running NetSurgeon over randomly generated networks (Mann-Whitney *U* test $P < 10^{-46}$). Further, we observed that NetSurgeon performed at random chance levels using the permuted networks, indicating network structural accuracy is critical for NetSurgeon performance. We also assessed the ability of NetSurgeon to identify the best intervention within the top 5 scoring interventions, a reasonable number of interventions to test experimentally. NetSurgeon ranks the best intervention in the top 5 for 91 goal states, which is 29-times better than random networks scores ($P<10^{-165}$).

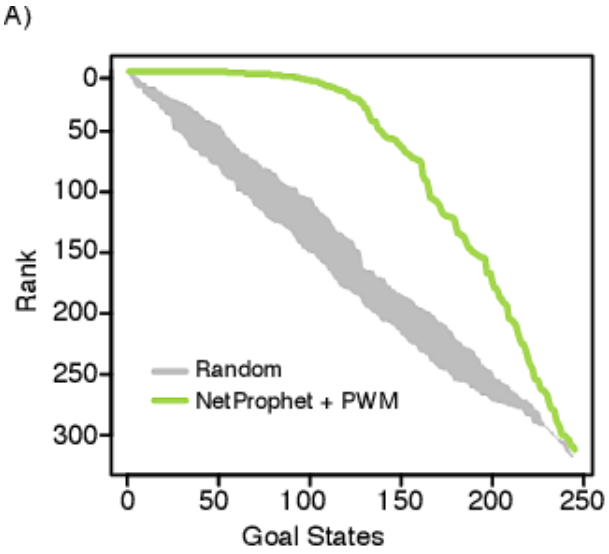

Figure 2.2.A. *In silico* assessment of NetSurgeon using 245 deletion mutant expression profiles grown in synthetic complete medium. Plotted curves show the number of goal states for which NetSurgeon ranked the best intervention at or above each rank (green), compared with random ranks (gray).

We also evaluated the ability of NetSurgeon to identify interventions in cells cultured in conditions even further from those used to construct the GRN. The goal states consisted of 63 expression profiles obtained from regulator overexpression strains grown in selective synthetic medium supplemented with 2% galactose (Chua, et al., 2006). We assessed the scores assigned to the best regulator for each overexpression goal state and compared the outcome to scores generated using random networks (Fig. 2.2.B). We found that NetSurgeon is able to assign higher scores to the correct interventions compared with random network generated scores (Mann-Whitney *U* test $P < 10^{-6}$). NetSurgeon is also able to assign the best intervention a top 5 rank for 8 of the 63 goal states (13%), a 10 fold improvement over the mean of the random network.

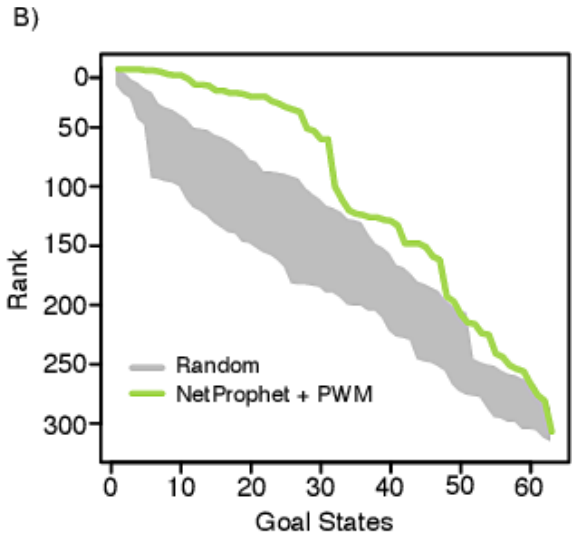

Figure 2.2.B. *In silico* assessment of NetSurgeon using 63 overexpression strains grown in grown in selective synthetic medium supplemented with 2% galactose. Plotted curves show the number of goal states for which NetSurgeon ranked the best intervention at or above each rank (green), compared with random ranks (gray).

In order to evaluate the effect of network accuracy on NetSurgeon performance, we applied NetSurgeon to GRNs inferred from the same expression data sets by CLR (Faith, et al., 2007), regression (Bonneau, et al., 2006), NetProphet (Haynes, Maier, Kramer, Wang, Brown, & Brent, 2013), and NetProphet integrated with PWM scores. We first evaluated the structural accuracy of the five GRNs by determining the level of ChIP support for high confidence interactions in each GRN. We then evaluated the performance of NetSurgeon when using each of these five GRNs on our two test data sets: the TF-deletion in SC glucose and TF-overexpression in SC galactose. We plotted the structural accuracy of each of the five GRNs against the NetSurgeon's accuracy when using that GRN (Fig. 2.2.C). We observed a clear pattern of improved NetSurgeon performance with more structurally accurate GRNs. A level-log regression model was fit to test this observation (Multiple $R^2 = 0.853$, P=0.00014) and forecasted a maximum NetSurgeon intervention recovery of 0.85 AUC with a perfect network model, which is a 33% improvement over current NetSurgeon performance.

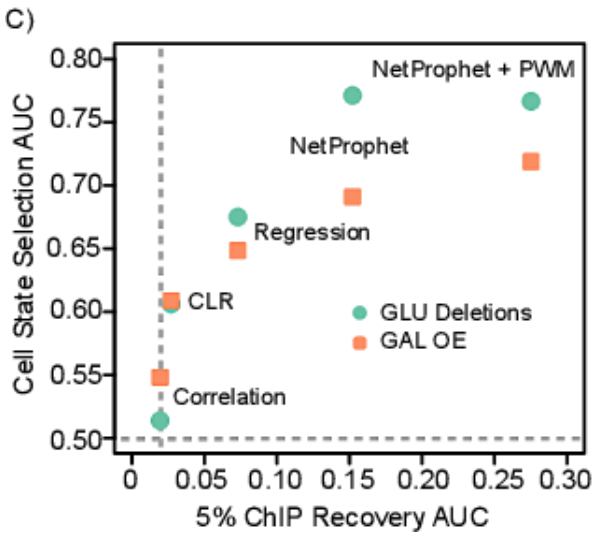

Figure 2.2.C. *In silico* assessment of the effect of network structural accuracy on NetSurgeon target intervention selection accuracy. Network structural accuracy of five GRNs, summarized by area under the precision recall curve at 5% ChIP recovery (x-axis), is compared with NetSurgeon intervention target selection accuracy, summarized by area under the curve of the number of goal states for which NetSurgeon ranked the best intervention at or above each rank (y-axis). Gray dotted lines indicate chance 5% ChIP recovery AUC and cell state selection AUC.

To assess the practicality of NetSurgeon-guided engineering, we ran NetSurgeon on the NetProphet+PWM network and computed the median number of interventions needed to identify the first, the best, and all deletion genotypes that reduce the distance between the wild-type cells and the goal by at least 10% (Fig. 2.2.D). A median of 12, 22 and 51 mutant strains were required to recover the first, the best, and all interventions (10-, 7-, and 4-fold better than random, respectively).

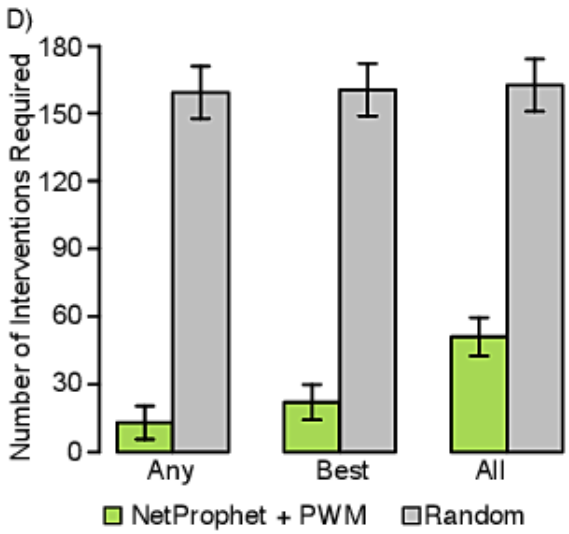

Figure 2.2.D. *In silico* assessment of the median number of NetSurgeon interventions required to generate any strain, the best strain, or all strains, that will converge expression state at least 10% towards the goal state (green), compared with random ranking (gray).

2.3.3 Application of transcriptome engineering to ethanol fermentation

Following the successful *in silico* validation of our approach for transcriptome engineering, we applied the algorithm to the industrially relevant problem of ethanol production in a mixed glucose-xylose co-culture. Principle components analysis of RNA-seq data from *S. cerevisiae* cells grown with xylose as the sole carbon source indicated that the system was in a hybrid transcriptional state with some characteristics of cells grown in 2% glucose, a fermentative state, and some characteristics of cells grown in 1.3% ethanol, a respiratory state (Fig. 2.3A). As *S. cerevisiae* cells do not natively consume xylose, we hypothesized that the system was unable to recognize the pentose carbohydrate as a fermentable carbon source and therefore entered into a transcriptional state that was non-optimal for fermentative metabolism.

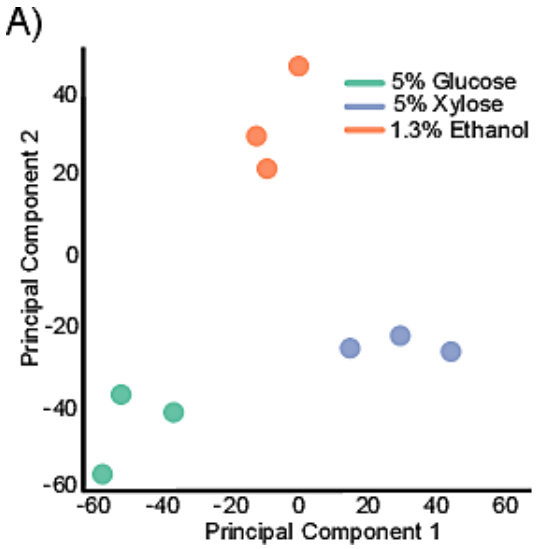

Figure 2.3.A. Principal component analysis of RNA expression profiles reveals a state transition between cells grown on 5% glucose (green), 5% xylose (blue), and 1.3% ethanol (red).

We therefore sought to identify interventions that would shift the system from the xylose-only transcriptional state (origin state) to the high-glucose state (goal state). In order to apply NetSurgeon to this problem, we generated the integrative, genome-wide network map described above and attempted to optimize the expression of the 445 genes involved in central carbon metabolism. Using this map, NetSurgeon produced a rank-ordered list of regulators whose deletion was predicted to force the system toward the 2% glucose transcriptional state. From this rank ordered list, we selected the top eight predicted interventions for biological validation via PCR-mediated genetic deletion of the selected regulators in the H2217-7 yeast strain. In order to assess the combinatoric effect of the predicted deletions, we generated an additional three strains carrying deletions in two of the NetSurgeon-selected regulators (*cat8/hap4*, *cat8/adr1*, *cat8/aft2*). For a limited comparison of our algorithmically selected deletions with expert intuition, we deleted the master regulator *SNF1*, the yeast ortholog of AMP kinase and a critical regulator responsible for glucose repression and other features of fermentative metabolism.

WILD-TYPE	SINGLE KO	DOUBLE KO
H2217-7	*snf1*	*cat8/adr1*
	adr1	*cat8/hap4*
	cat8	*cat8/aft2*
	usv1	
	gis1	
	msn2	
	hap4	
	msn4	
	aft2	

Table 2.1. Wild-type and deletion mutant strains profiled

We found that the NetSurgeon-selected targets were supported by existing literature. Cat8 and Hap4 are respiratory factors active in the general cellular response to xylose and deletion of *HAP4* was recently shown to improve cellobiose consumption rates (Salusjärvi, Kankainen, Soliymani, Pitkänen, Pentti, & Ruohonen, 2008) (Lin, et al., 2014). *MSN2* and *MSN4,* encoding stress associated factors, were observed to be highly upregulated in xylose and their transcriptional targets misregulated (Matsushika, Goshima, Hoshino, & others, 2014). Usv1, Gis1 and Aft2 were all found to have clear roles in the yeast transcriptional response to non-fermentable carbon sources and general stress response (Hlynialuk, Schierholtz, Vernooy, & der, 2008) (Pedruzzi, Bürckert, Egger, & Virgilio, 2000) (Blaiseau, Lesuisse, & Camadro, 2001).

Aerobic batch fermentations were used to assess the outcome of our transcriptome interventions at the transcriptional and metabolic levels. Cells were inoculated into synthetic complete medium supplemented with 2% glucose and 5% xylose at an OD600 of 1.0+/- 0.2 and grown for 48 hours. Samples were taken for RNA-sequencing at 4 hours and 24 hours, representative of the glucose-xylose and xylose-only metabolic states. Aliquots were acquired for HPLC metabolite

analysis across the 48 hour fermentation (Fig. 3B). Using this data, we examined the NetSurgeon's ability to control transcriptional state and quantitatively assessed the effect of transcriptome transcriptional state change on a complex phenotype.

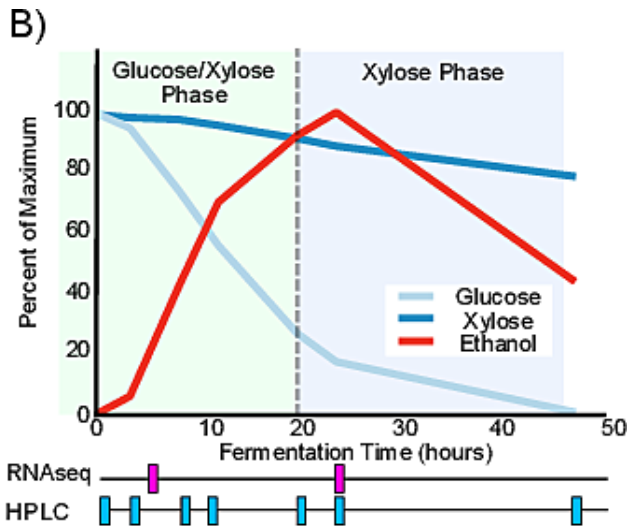

Figure 2.3.B. Top: Glucose (light blue), xylose (dark blue), and ethanol (red) metabolite concentration profiles from the fermentation of the wild-type H2217-7. Bottom: Overview of RNA-seq (magenta) and HPLC (turquoise) sampling strategy for aerobic batch fermentations used in this study.

2.3.4 Transcriptome engineering successfully promotes a fermentative state

Differential expression analysis revealed that 2,887 genes are differentially expressed by at least two fold in wild-type cells as a result of glucose depletion (42% of all genes. Fig. 2.4.A). Six of the eight NetSurgeon-selected interventions lowered the number of differentially expressed genes. The *cat8* mutant was the best, preventing the change in expression of 1,182 of 2,887 DE genes while creating only 526 new DE genes, for a net reduction of 656 DE genes. Notably, the deletion of *CAT8* reduced differential expression better than the deletion of *SNF1*, a master regulator of the *S. cerevisiae* glucose repression system.

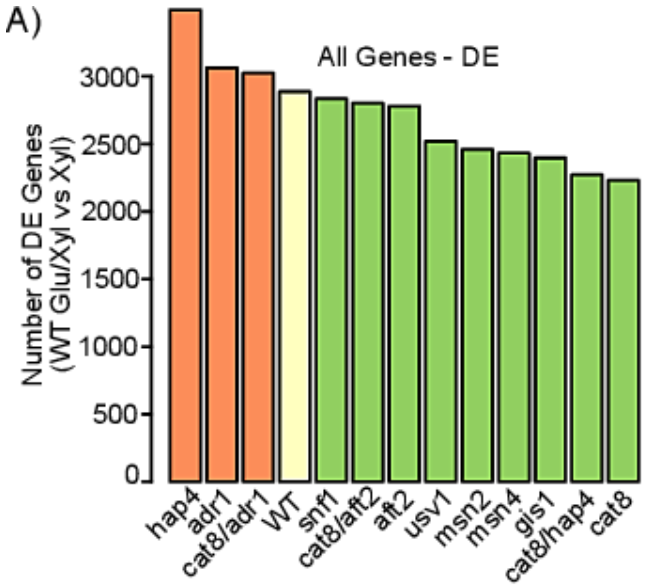

Figure 2.4.A. Number of 2-fold or greater differentially expressed genes between the wild-type strain in the fermentative state and each strain in the respiratory state. Green and red bars indicate strains with less and more differentially expressed genes than wild-type respectively.

Next, we calculated the Euclidean distance between the wild-type expression state in the glucose-xylose phase and the deletion strain's expression state in the xylose-only phase (Fig. 2.4.B). Six of the eight NetSurgeon interventions lowered the Euclidean distance between the two phases. The single deletion mutant *cat8* reduced the genome-wide expression distance between the glucose-xylose phase and the xylose-only phase by 28.4%. The mean reduction in Euclidean distance of the six successful NetSurgeon selected interventions was 20.8%. As in the DE analysis, the deletion of *HAP4* and *ADR1* increased the total distance between the two state vectors. None of the double mutant strains (*cat8/hap4*, *cat8/adr1*, *cat8/aft2)* reduced the distance between the two states more than the single deletion of *CAT8*.

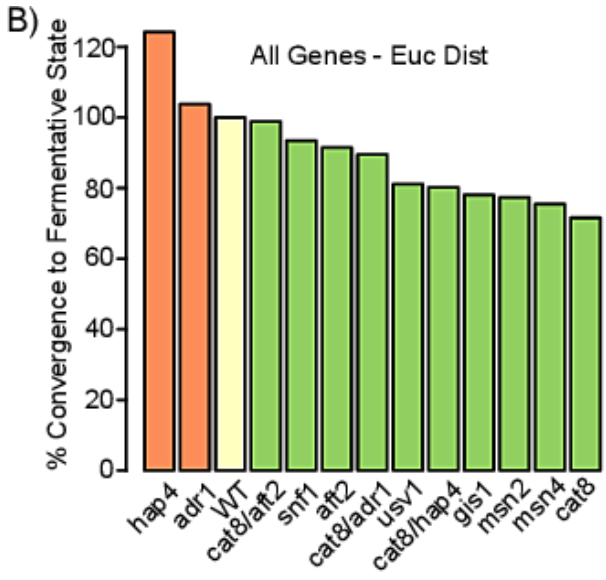

Figure 2.4.B. Euclidean distance between the full expression profile of wild-type strain in the fermentative state, and the full expression profiles of all strains in the respiratory state. Green and red indicate reduced and increased Euclidean distance compared with wild-type respectively.

The NetSurgeon-selected interventions were specifically targeted at optimizing the expression state of 445 genes involved in carbon metabolism. Among these 445 genes, the *CAT8* deletion reduced the Euclidean distance by 36% (Fig. 2.4.C). On average, the six successful NetSurgeon-selected interventions reduced the Euclidean distance between the two state vectors by 24%.

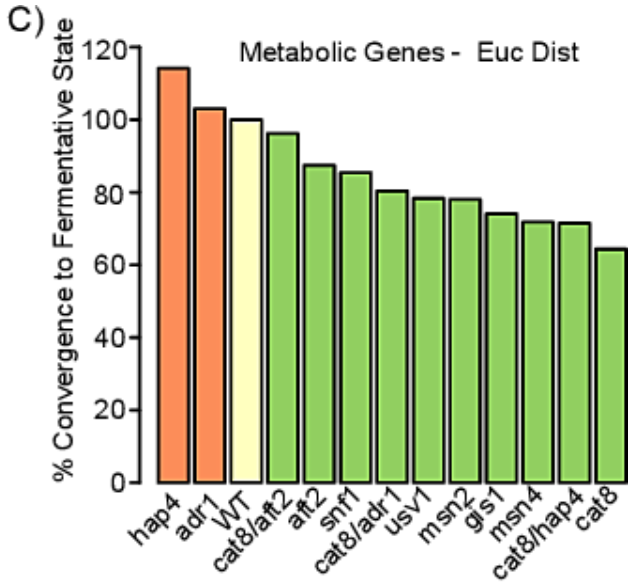

Figure 2.4.C. Euclidean distance between the expression 445 metabolically active computationally optimized genes in the fermentative state of the wild-type strain, and the matching optimized gene expression profiles of all strains in the respiratory state. Green and red indicate reduced and increased Euclidean distance compared with wild-type respectively.

Each of the eight NetSurgeon selected transcription factors had known roles in the regulation of the cellular stress response or respiratory processes. We evaluated the ability of each transcription factor to promote a fermentative state across specific metabolic pathways (Fig. 2.4.D). With the exception of *adr1*, each deletion mutant affected the expression of genes across many of the metabolic pathways in central carbon metabolism. Seven of the eight NetSurgeon-selected interventions lowered the Euclidean distance in at least one of the central carbon metabolism pathways evaluated. All six of the interventions that reduced differential expression and global Euclidean distance moved the expression of glycolytic genes toward a fermentative state.

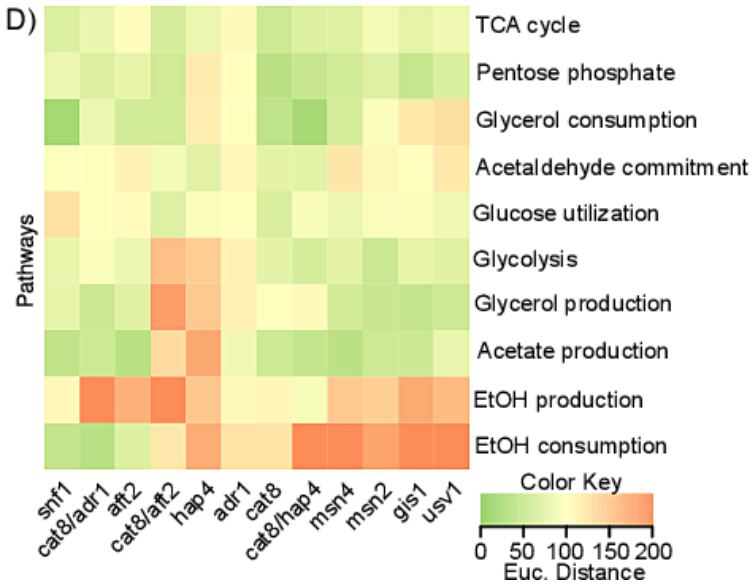

Figure 2.4.D. Euclidean distance between the expression profiles of central carbon metabolic pathways in the fermentative state of the wild-type strain, and the matching central carbon metabolic pathway expression profiles of all strains in the respiratory state. Green and red indicate reduced and increased Euclidean distance compared with wild-type respectively.

Three of these interventions shifted the expression of TCA cycle genes toward a fermentative state. Deletion of *CAT8* promoted a fermentative state in many metabolic pathways essential for xylose fermentation, including genes involved in glucose utilization, the pentose phosphate pathway, glycolysis, the TCA cycle, and acetate/glycerol production. All of the TCA cycle genes were moved toward the expression level associated with fermentative metabolism (Fig. 2.4.E). Deletion of *CAT8* also reduced the Euclidean distance of all TCA genes from the fermentative state by 60%. These observations highlight the power of transcriptome level interventions to modulate the expression of many more genes than is feasible by traditional, one-gene-at-a-time genetic engineering.

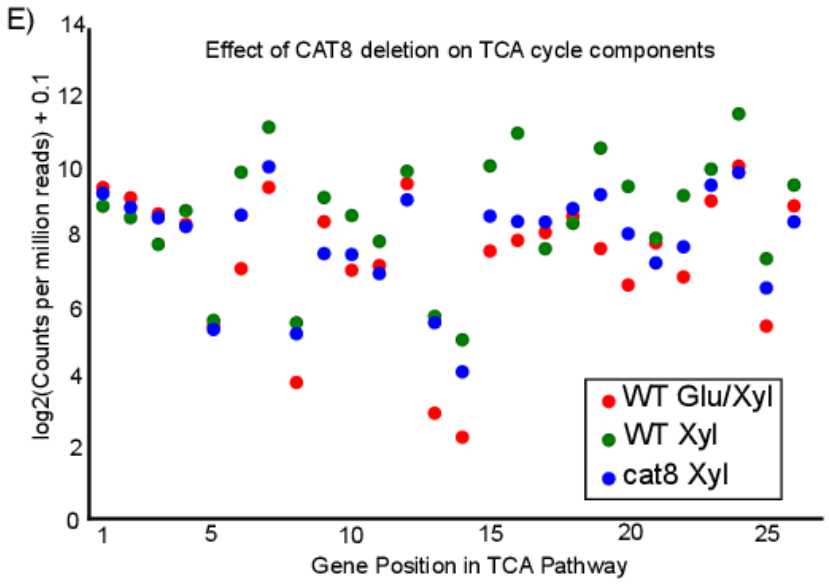

Figure 2.4.E. Comparison of the expression TCA cycle genes between the fermentative state of the wild-type strain (red), the respiratory state of the wild-type strain (green), and the respiratory state of the cat8 deletion mutant strain (blue).

2.3.5 Identification of transcriptional states associated with improved fermentative performance

In order to assess the change in cellular metabolic behavior following our transcriptional interventions, we profiled metabolic intake and output via HPLC. HPLC analysis identified three metabolic states associated with high glucose, low glucose, and respiratory metabolic phases. We focused our downstream analyses on the high glucose and respiratory phases of the fermentation during which we had carried out RNA-seq.

To examine the ability of the selected transcriptional interventions to control the metabolic state of the cell, we calculated the percentage of input carbon that end up in each of the major carbon fates in each phase (Fig. 2.5.A, 2.5.B). Carbon import rates significantly declined in the absence of high glucose, with a mean reduction of import across all assayed genotypes by 86%. In addition to changes in import rate, the cells significantly upregulated their commitment of carbon to respiratory processes in the xylose-only phase. Carbon commitment to respiration changed from a mean of 24% in the glucose-xylose phase to 89% in the xylose-only phase. This indicated that the metabolism of all strains had shifted into a respiratory mode during the xylose-only phase (Fig. 2.5.B).

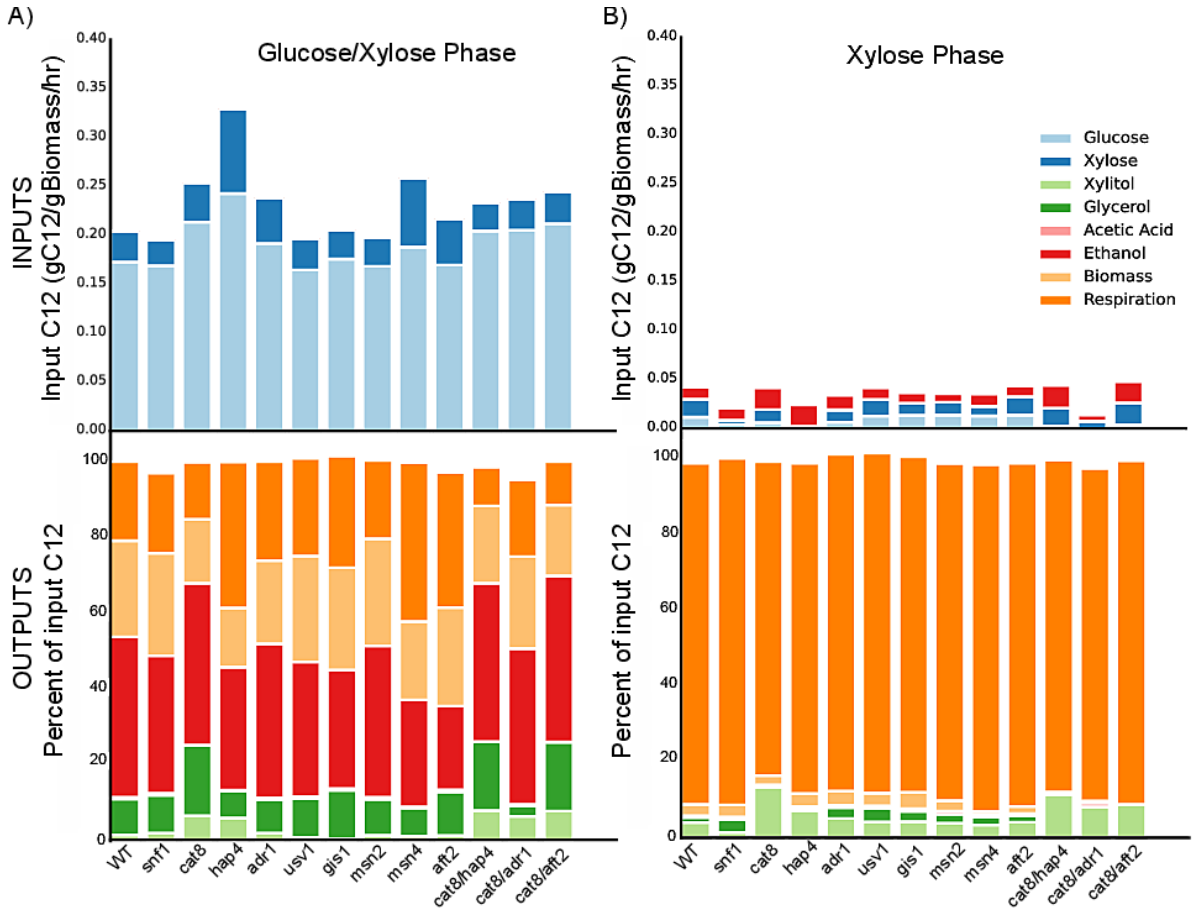

Figure 2.5.A-B. Transcriptome interventions alter carbon intake rates and commitment, but do not prevent a transition to a respiratory metabolism. Panel A: Cellular metabolic input and output across profiled genotypes during the glucose-xylose phase of aerobic fermentation. Panel B: Cellular metabolic input and output within the xylose phase.

Although the tested interventions did not prevent the transition to a respiratory metabolic state, they did affect the cell's commitment of carbon to output metabolites significantly. We observed 41 statistically significant changes in carbon commitment across the 13 profiled genotypes ($p<0.05$, t-test, Benjamini-Hochberg corrected). 28 of these changes were within the glucose-xylose phase. Carbon commitment to all of the profiled metabolites and phenotypes was altered in at least one of our transcriptional interventions, indicating that changes in transcriptional state have the power to impact all dimensions of cellular metabolism. Carbon commitment to xylitol was significantly increased in transcriptional interventions associated with respiratory processes,

a potential side effect of the respiratory factors regulating the promoters of the *XYL1*, *XYL2*, and *XKS1* enzymes required for xylose integration into central carbon metabolism. Interestingly, all significant changes in carbon commitment to ethanol and biomass were reductions. The deletion of *SNF1*, *HAP4*, *USV1*, *GIS1*, *MSN4* and *AFT1* significantly reduced carbon commitment to ethanol, with a mean reduction in carbon flux by 26%. Deletions involving *CAT8* or *HAP4* significantly reduced carbon commitment to biomass by 33% and 38%, respectively, in the glucose-xylose phase of the fermentation. We found that the *cat8/hap4* and *cat8/aft2* double mutant strains exhibited similar carbon commitment phenotypes as the *CAT8* single. However, the cat8/adr1 strain deviated from the *cat8* single mutant by committing wild type level of carbon to acetic acid, while lowering carbon commitment to glycerol by 70% vs the wild type strain (85% lower vs. *cat8*).

We also observed 57 statistically significant changes in the specific rates of metabolite production or consumption across the glucose-xylose and xylose-only phases of the fermentation ($p < 0.05$, t-test, Benjamini-Hochberg corrected). Within the glucose-xylose phase, we identified industrially relevant changes in glucose and xylose consumption rates, acetic acid output and ethanol production. All of the profiled interventions on respiratory regulators (*cat8*, *hap4*, *adr1*) improved the specific rate of glucose consumption between 11% and 40% (Fig. 2.5.C). We found that *hap4* and *msn4* mutants improved the specific rate of xylose consumption by 170% and 120% respectively (Fig. 2.5.D). Acetic acid, a fermentation byproduct demonstrated to inhibit glycolysis, was also produced at 53%-83% lower specific rates in the *hap4* and *cat8* mutants (Fig. 2.5.E) (Pampulha & Loureiro-Dias, 1990). Importantly, the specific rate of ethanol production was significantly increased by 22% and 31% in the *cat8* and *hap4* mutants (Fig.

2.5.F). Within the set of stress-associated factors, we found that the deletion of *USV1*, *MSN2*, *MSN4* and *AFT2* significantly reduced the specific rate of ethanol production, with a mean rate reduction of 22% (Fig 2.5.F). Double mutant strains for *cat8/hap4* and *cat8/aft2* performed similarly to the cat8 single deletion strain. The double mutant for *cat8/adr1* exhibited wild type levels of acetic acid production while lowering production rates of glycerol vs. the wild type genotype. Taken together, these data demonstrate the ability of transcriptome engineering to generate significant changes in cellular behavior, even in the absence of complete phenotypic conversion.

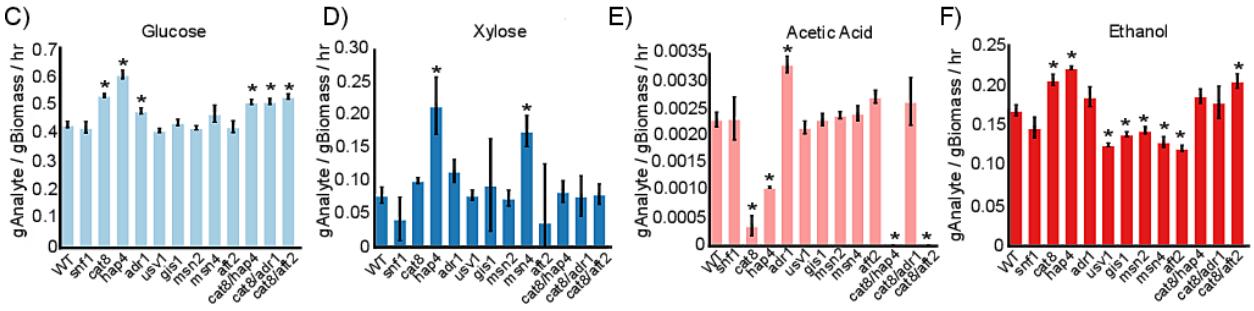

Figure 2.5.C-F. Transcriptome interventions alter specific rates of metabolite production or consumption in the glucose-xylose phase. Panel C: Specific rate of glucose consumption. Panel D: Specific rate of xylose consumption. Panel E: Specific rate of acetic acid production. Panel F: Specific rate of ethanol production.

2.3.6 An integrated model of transcriptional regulation and metabolic flux leads to novel, validated predictions of interventions to optimize biofuel production

The lack of data linking transcriptional state with metabolic phenotypes has prevented the use of transcriptional interventions for effective engineering of metabolism. In order to address this issue, we utilized our dataset to construct an integrated model of *S. cerevisiae* central carbon metabolic flux and expression. We identified regulators linked to flux by correlating their expression with pathway carbon flux. From this set of regulator-flux correlations, we identified regulators putatively controlling metabolic flux outcomes via network-predicted direct regulatory relationships (Fig. 2.6).

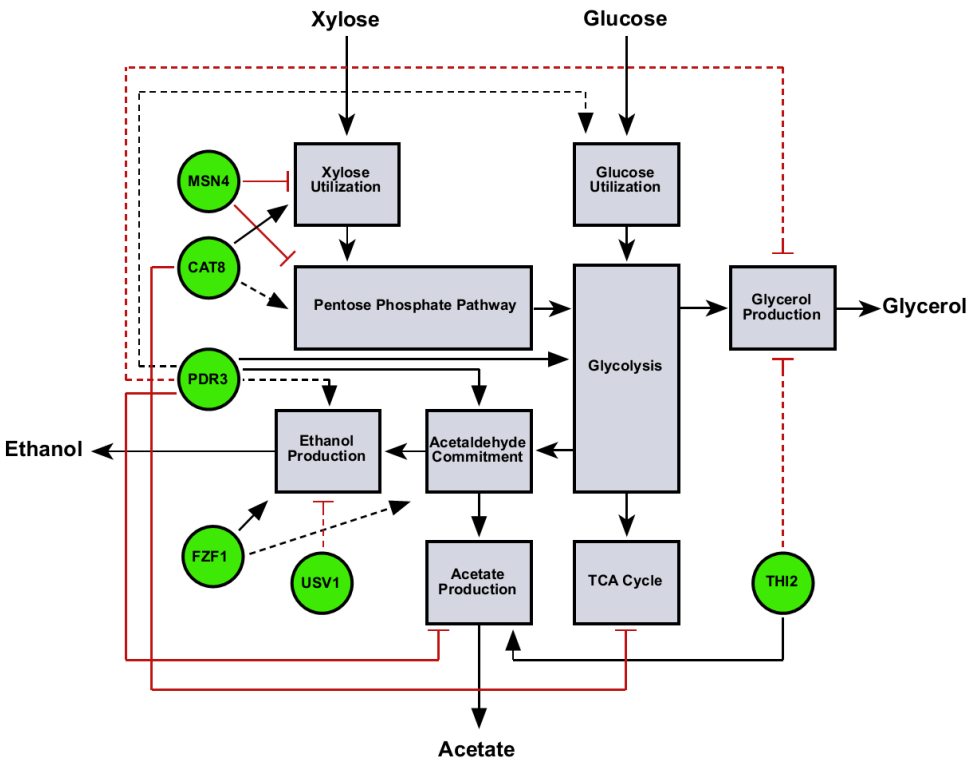

Figure 2.6. An integrated map relating transcription factors to central carbon metabolism flux. Blue rounded rectangles: pathways in central carbon metabolism. Green ovals: transcription factors. Links between transcription factors and pathways denotes transcription factor expression correlation with increased (black arrow headed link) or decreased (red circle headed link) flux through the pathway. Solid link lines: transcription factor directly regulates the expression of genes in the pathway.

This analysis revealed that three transcriptional regulators were deeply interconnected with biochemical pathways important for xylose metabolism and fermentation. *CAT8* expression was correlated with genes associated with xylose utilization, the pentose phosphate pathway, acetate production and the TCA cycle. Msn4 was predicted to directly regulate genes involved in xylose utilization, the pentose phosphate pathway, and the TCA cycle, and flux through these pathways was anti-correlated with *MSN4* expression. Pdr3 was revealed to be a regulator of glycolytic genes, and flux through these pathways was positively correlated with *PDR3* expression. This integrated model of transcriptional regulation and metabolic flux is an important step toward the rational engineering of *S. cerevisiae* metabolism.

2.4 Discussion

We have demonstrated that transcriptional network maps can be used to rationally manipulate cellular state by identifying the crucial regulators mediating a state transition and prioritizing them for genetic intervention. The formalization of this process of rational state manipulation is expected to enable future developments in personalized medicine, improve approaches to stem cell engineering, and reduce the costs associated with these efforts. Our work establishes quantitative benchmarks in this new field, enabling the rapid progress generally associated with clear benchmarks (Stolovitzky, Monroe, & Califano, 2007).

The availability of deletion and overexpression collections in *S. cerevisiae* has enabled us to assess the state of the art in network-guided transcriptome engineering. We found that NetSurgeon can identify the best intervention within a median of 22 guesses, a 7-fold improvement over random guessing. We observed that network maps built from data on one environmental condition can be successfully used to predict interventions in different conditions. This is important for applications that deviate from standard environmental conditions. Finally, we have demonstrated the utility of TF-network maps enriched with direct regulatory relationships; maps generated by NetProphet together with PWM models led to selections that were substantially better than those made by using maps expression correlation or CLR.

We applied NetSurgeon to optimizing yeast for ethanol production from glucose-xylose coculture. NetSurgeon selected critical regulators highlighted in the literature and six of the eight promoted a fermentative transcriptional state. Although the single deletions were insufficient to entirely prevent a state transition involving 43% of the yeast genome, it succeeded in

significantly changing the rate and ratio of cellular carbon commitment. We found that regulators associated with respiratory processes had significant metabolic effects in the fermentative phase of the culture. We also found that deletion of stress factors lowers the rate of production and the total ethanol yield. In addition, our dataset of 8,055 metabolic measurements with 73 matched RNA sequencing profiles across 14 genotypes will enable future engineering efforts to identify and rationally manipulate the critical regulators of metabolic flux in order to maximize biofuel production.

One of the advantages of transcriptome engineering is the possibility of accessing evolutionarily optimized states associated with specific phenotypes. The expression levels of genes within linear metabolic pathways such as glycolysis and the TCA cycle are highly regulated in order to maintain a correct ratio of enzyme products necessary for avoiding intermediate metabolite accumulation and allosteric inhibition of upstream processes. The engineering of optimal expression levels across entire pathways is a challenging problem that is often addressed through iterative selection strategies (Wang, et al., 2009). We observed that manipulation of regulator expression levels is a promising strategy to access pre-defined expression states across entire pathways. The effect of *CAT8* deletion on TCA gene expression is one example of an intervention reconfiguring the expression of an entire pathway toward a fermentative state. The TCA cycle within *S. cerevisiae* consists of twenty-six genes, making optimization of this pathway's expression level a difficult task through one-gene-at-a-time engineering. Cat8 was predicted by NetProphet to regulate four genes within the TCA cycle and the glyoxylate pathway, and removal of this factor was predicted to move the TCA cycle toward a fermentative expression configuration. We found that *CAT8* deletion moved all twenty-six genes of the TCA

cycle toward a fermentative state, providing evidence that naturally evolved transcriptional states can be leveraged for transcriptome engineering.

Our analysis of the double deletion strains highlighted the complexity of epistatic effects within gene regulatory networks. Although the generation of strains with multiple regulatory perturbations offers the possibility of large scale reconfiguration of cellular state, we observed that the three double deletion strains failed to reduce differential expression and Euclidean distance as much as their component single deletions. This non-additivity between genotypes indicates that a more sophisticated approach to modeling the effect of multiple regulator perturbations will be required to expand target selection approaches multiple perturbations.

2.5 Materials and Methods

2.5.1 Network guided target selection

To rank possible regulator interventions for convergence toward a goal expression state, NetSurgeon uses a GRN to simulate interventions for all regulators, and for each simulated regulator intervention a score is assigned representing the confidence that the regulator intervention will converge the expression state toward the goal state. The score for a simulated regulator intervention is based on the enrichment of the regulator's simulated intervention effects to fix the total dysregulation of all genes between the initial and goal expression states, where the total dysregulation of all genes is quantified by the sum of the negative log pvals of significance of differential expression. Specifically the NetSurgeon network intervention score for a regulator is:

NetSurgeon network intervention score (R_i) = max($-\log_{10}$ (hypergeometric distribution(X_{ij} * (W/D), W, U-W, (X_{ij}+Y_{ij}) * (W/D) + C_{ij} - Z_{ij}))) for network cutoff j = 500, …, 40,000

where U is the total number of genes in the network, W is the number of dysregulated genes, D is total amount of dysregulation, X_{ij} is the total amount of dysregulation that the intervention of regulator R_i will remove when considering only the top j interactions in the network, Y_{ij} is the total amount of dysregulation that the intervention of regulator R_i will make worse when considering only the top j interactions in the network, C_{ij} is the total number of genes regulated by regulator R_i when considering only the top j interactions in the network, and Z_{ij} is the total number of dysregulated genes regulated by regulator R_i when considering only the top j interactions in the network.

2.5.2 Strain engineering

The xylose metabolizing strain VTT-C-99318 (CEN.PK2-1D ura3::XYL1 XYL2 his3::XKS1 kanMX) was acquired from Salusjarvi et al. and used as the base strain for all experiments in this study (Salusjärvi, Kankainen, Soliymani, Pitkänen, Pentti, & Ruohonen, 2008). The At5g17010 xylose transporter from *A. thaliana* was transformed into the VTT-C-99318 strain and maintained through the use of dropout media (Hector, Qureshi, Hughes, & Cotta, 2008). The genetic deletion of algorithmically selected transcription factors was accomplished through PCR amplification and targeting of drug cassettes to the selected ORF via the addition of 45 base pairs of homologous sequence to the 5'/3' amplifying oligos (Baudin, Ozier-Kalogeropoulos, Denouel, Lacroute, & Cullin, 1993). Prior to use in experimentation, all strains were freshly plated onto selectable media from frozen stocks.

2.5.3 *S. cerevisiae* fermentations

All *S. cerevisiae* strains were grown aerobically in 60 mL of synthetic complete at 30℃ in 250 mL baffled erlenmeyer culture flask shaken at 225 RPM. Cultures for identification of differential expression associated with carbon sources were grown in triplicate in either 50 g/L

glucose, 50 g/L xylose or 1.3 g/L ethanol for 8 hours prior to collection of biomass for RNA sequencing. Culture of cells for evaluation of the impact of transcriptional interventions were performed in triplicate and initiated by inoculating 1.0+/- 0.2 OD600 units of biomass into 60 mL of synthetic complete media supplemented with 20 g/L glucose and 50 g/L xylose. Samples taken for RNA-seq analysis were aliquoted from the primary culture, spun down at 3000xg and frozen in liquid nitrogen prior to downstream analysis. The supernatant of samples for HPLC was collected by centrifugation of culture samples at 12,000xg for 3 minutes prior to snap freezing for storage in a dry ice/ethanol bath. All samples were stored at -80℃. At least two independent experiments of three biological replicates was performed for each genotype evaluated by HPLC. Cellular density was quantitated through analysis of culture turbidity at 600 nm.

2.5.4 Metabolite analysis:

The concentration of input and output cellular metabolites was analyzed using HPLC. Supernatant solutions were stored at -80℃ and filtered through the use of 0.22 um syringe prior to HPLC analysis. Metabolites were eluted from an Aminex HPX-87H column maintained at 65℃ and peaks detected by refractive index. Identified peaks were quantified through integration and interpolated against serial dilutions of standards for glucose, xylose, xylitol, glycerol, acetic acid and ethanol. Analysis of HPLC data was performed on a per biological replicate basis, with metabolic input/output relationships quantified across each fermentation and pooled into a single distribution based on genotype. Turbidity measurements were converted into units of g/biomass based on the turbidity to biomass conversion factor published (Hector, Qureshi, Hughes, & Cotta, 2008). Calculations of analyte rate and specific rate of change were performed across steady states identified in the ethanol dimension. In order to evaluate internal carbon flux, a

system of linear equations was developed to describe central carbon metabolism in *S. cerevisiae*. The system of equations was fit to experimentally measured parameters of carbon import and export for each genotype across the glucose/xylose and xylose-only phases of each fermentation.

2.5.5 RNA sequencing and analysis:

Total mRNA was isolated using the yeast RiboPure kit (Life Technologies, Carlsbad CA). Libraries for RNA-Seq were prepared as in (Haynes, et al., 2011). Briefly, poly(A) RNA was selected from the total RNA isolated as above using the mRNA Catcher Plus Kit (Life Technologies) with an epMotion 5075 automated pipettor (Eppendorf). The poly(A) RNA was subsequently sheared by incubating in TURBO DNA-free buffer at 75°C for 10 minutes and purified with the QIAquick PCR Purification Kit (Qiagen). First strand cDNA synthesis was performed using random hexameric primers and SuperScript III Reverse Transcriptase, followed by treatment with E. coli DNA ligase, DNA polymerase I, and RNase H for second-strand synthesis, all using standard methods. The cDNA libraries were end-repaired with a Quick Blunting kit and A-tailed using Klenow exo- with dATP (New England Biolabs). Illumina adapters were ligated to the cDNA and fragments ranging from 150-250 bp in size were selected using gel electrophoresis. The libraries were enriched and indexed in a 10-cycle PCR using Phusion Hot Start II High-Fidelity DNA (Fermentas), purified, and pooled in equimolar ratios for multiplex sequencing on an Illumina HiSeq 2500.

2.5.6 RNA/metabolic data integration and analysis:

We utilized two different complementary methods for integrating RNA expression profiles and metabolic data in order to gain a better understanding of the molecular mechanisms controlling metabolic phenotypes. First, we used the limma software package (Ritchie, et al., 2015) to identify differentially expressed genes within the fermentative and respiratory states between the

wild-type strain and each deletion strain. We then used this differential expression analysis to putatively link genes mechanistically to metabolic analyte outcomes by identifying differentially expressed genes in metabolic pathways linked to each metabolic analyte.

In addition to differential expression analysis, we also identified genes linked to metabolic outcomes by identifying genes whose expression significantly correlates with carbon flux. For each gene and each metabolic pathway we computed the Pearson correlation coefficient between the gene's expression profile, and the computed carbon flux through the pathway. We then generated a null distribution of correlation coefficients between gene expression and pathway flux by randomly generating 10,000 expression vectors by sampling per condition from the expression of all genes within the condition with replacement. These null distributions of expression correlation with pathway flux were then used to assign false discovery rate corrected p-values to the significance of each gene's expression correlation with flux measurements of a metabolic pathway.

2.6 Supplemental Data

2.6.1 High glucose phase rates of metabolite change:

Genotype	GLUmean	GLUsem	GLU_t_pval
WT	-0.8640	0.04068	NA
snf1	-0.8830	0.03323	0.8440
cat8	-0.8415	0.0121	0.7723
hap4	-1.317	0.03588	1.82E-05
adr1	-1.297	0.1274	0.009629
usv1	-0.9007	0.01453	0.6935
gis1	-0.9811	0.02502	0.2283
msn2	-0.9082	0.01736	0.4717
msn4	-0.9382	0.03238	0.4728
aft2	-0.8673	0.03078	0.9833
cat8/hap4	-0.8808	0.01033	0.8841
adr1/cat8	-1.959	0.01241	1.55E-08
cat8/aft2	-0.8936	0.01969	0.7587
Genotype	XYLmean	XYLsem	XYL_t_pval

WT	-0.1633	0.0287	NA
snf1	-0.09280	0.06998	0.4728
cat8	-0.1617	0.008607	0.9869
hap4	-0.4809	0.04761	0.0003854
adr1	-0.3100	0.04937	0.04074
usv1	-0.1477	0.009625	0.8637
gis1	-0.1950	0.1421	0.8637
msn2	-0.1627	0.02697	0.9982
msn4	-0.3855	0.04770	0.006759
aft2	-0.06599	0.1808	0.6297
cat8/hap4	-0.1465	0.03033	0.8486
adr1/cat8	-0.1050	0.07373	0.5462
cat8/aft2	-0.1359	0.02614	0.7104
Genotype	XYLITOLmean	XYLITOLsem	XYLITOL_t_pval
WT	0.01504	0.001955	NA
snf1	0.01799	0.006156	0.7587
cat8	0.06486	0.006472	9.88E-07
hap4	0.1059	0.003404	3.55E-12
adr1	0.02693	0.001828	0.001984
usv1	0.006838	0.004180	0.1331
gis1	0.008630	0.003863	0.2209
msn2	0.01200	0.002770	0.5416
msn4	0.01561	0.001403	0.9238
aft2	0.01438	0.001193	0.9175
cat8/hap4	0.08079	0.001039	1.02E-11
adr1/cat8	0.1272	0.009137	7.30E-10
cat8/aft2	0.08068	0.0006325	7.30E-10
Genotype	GLYmean	GLYsem	GLY_t_pval
WT	0.1001	0.006579	NA
snf1	0.09914	0.00716	0.9693
cat8	0.1881	0.002016	1.57E-08
hap4	0.1451	0.005586	0.002191
adr1	0.1277	0.01063	0.08131
usv1	0.1040	0.004458	0.8579
gis1	0.1350	0.003150	0.01654
msn2	0.1023	0.002893	0.8637
msn4	0.1006	0.003264	0.9863
aft2	0.1157	0.006877	0.2885
cat8/hap4	0.1920	0.001974	5.64E-07
adr1/cat8	0.07439	0.008366	0.08275
cat8/aft2	0.1903	0.001928	6.76E-07
Genotype	ACETICmean	ACETICsem	ACETIC_t_pval
WT	0.004685	0.0003660	NA
snf1	0.004943	0.0008489	0.8637

cat8	0.0005829	0.0002893	9.64E-07
hap4	0.002423	1.34E-05	0.01327
adr1	0.009178	0.001232	0.003930
usv1	0.004734	0.0002487	0.9756
gis1	0.005000	8.99E-05	0.7473
msn2	0.005123	0.0001408	0.4713
msn4	0.005301	0.0002399	0.4418
aft2	0.005568	0.0001715	0.22833
cat8/hap4	4.62E-06	0	1.26E-06
adr1/cat8	0.009605	0.001838	0.009461
cat8/aft2	4.62E-06	0	1.26E-06
Genotype	ETOHmean	ETOHsem	ETOH_t_pval
WT	0.3403	0.01810	NA
snf1	0.3083	0.02564	0.4654
cat8	0.3222	0.01188	0.6297
hap4	0.4901	0.01425	0.0008515
adr1	0.5172	0.07528	0.05013
usv1	0.2763	0.004821	0.1490
gis1	0.2790	0.009706	0.1217
msn2	0.3123	0.01122	0.36099
msn4	0.2855	0.01558	0.1506
aft2	0.2461	0.006381	0.03550
cat8/hap4	0.3227	0.005091	0.7459
adr1/cat8	0.6423	0.08348	0.001027
cat8/aft2	0.3470	0.01571	0.9036
Genotype	BIOMASSmean	BIOMASSsem	BIOMASS_t_pval
WT	0.3747	0.01888	NA
snf1	0.3811	0.01456	0.8964
cat8	0.2375	0.007063	6.22E-05
hap4	0.4279	0.01569	0.1744
adr1	0.5245	0.08661	0.1504
usv1	0.4085	0.01083	0.4037
gis1	0.3864	0.01879	0.8257
msn2	0.4248	0.00664	0.05645
msn4	0.3875	0.006663	0.7808
aft2	0.3740	0.006049	0.9951
cat8/hap4	0.2722	0.004360	0.01347
adr1/cat8	0.8211	0.01212	4.86E-08
cat8/aft2	0.2697	0.001422	0.01194

Table 2.2: High glucose phase measured rates of metabolite change. Units in grams analyte/hour.

2.6.2 Low glucose phase rates of metabolite change:

Genotype	GLUmean	GLUsem	GLU_t_pval
WT	-0.76000	0.02587	NA
snf1	-0.5639	0.09095	0.07733
cat8	-0.9529	0.02526	0.0002985
hap4	-0.7996	0.005789	0.5641
adr1	-0.6028	0.08063	0.1183
usv1	-0.6308	0.004378	0.04090
gis1	-0.6287	0.01001	0.01194
msn2	-0.5929	0.01442	8.13E-05
msn4	-0.5989	0.007970	0.002705
aft2	-0.6717	0.02808	0.1183
cat8/hap4	-0.9891	0.008234	0.0003133
adr1/cat8	-0.04606	0.006369	7.30E-10
cat8/aft2	-0.9579	0.002997	0.001140
Genotype	XYLmean	XYLsem	XYL_t_pval
WT	-0.30811	0.03079	NA
snf1	-0.2401	0.1379	0.7444
cat8	-0.3110	0.03465	0.9805
hap4	-0.5234	0.06040	0.01194
adr1	-0.3757	0.05140	0.4027
usv1	-0.1727	0.04210	0.05730
gis1	-0.1261	0.04237	0.01654
msn2	-0.1187	0.02508	0.0008600
msn4	-0.1858	0.04959	0.1044
aft2	-0.5588	0.2099	0.2205
cat8/hap4	-0.2619	0.009227	0.5748
adr1/cat8	-0.3455	0.06256	0.7229
cat8/aft2	-0.3014	0.01133	0.9629
Genotype	XYLITOLmean	XYLITOLsem	XYLITOL_t_pval
WT	0.03124	0.004564	NA
snf1	0.01565	0.003949	0.06187
cat8	0.2021	0.02013	1.67E-07
hap4	0.2151	0.009082	4.10E-11
adr1	0.05618	0.008336	0.03554
usv1	0.02012	0.008886	0.3848
gis1	0.02392	0.007061	0.5462
msn2	0.02499	0.001931	0.3895
msn4	0.01521	0.001134	0.07112
aft2	0.01809	0.002766	0.1532
cat8/hap4	0.2422	0.004232	7.15E-13
adr1/cat8	0.1301	0.004710	3.21E-08
cat8/aft2	0.2233	0.001818	5.51E-11
Genotype	GLYmean	GLYsem	GLY_t_pval

WT	0.07131	0.003119	NA
snf1	0.04807	0.002550	0.000175841
cat8	0.1360	0.001975	2.73E-10
hap4	0.02332	0.0007066	4.54E-06
adr1	0.05335	0.01065	0.17089
usv1	0.06736	0.004306	0.6550
gis1	0.09382	0.007308	0.01435
msn2	0.06634	0.003552	0.4654
msn4	0.05605	0.0008246	0.04535
aft2	0.05642	0.007501	0.1045
cat8/hap4	0.1385	0.002602	5.08E-09
adr1/cat8	0.001222	0.0008510	7.07E-08
cat8/aft2	0.1396	0.001274	9.92E-08
Genotype	ACETICmean	ACETICsem	ACETIC_t_pval
WT	0.005455	0.0004036	NA
snf1	0.007101	0.0008818	0.1654
cat8	0.0003676	0.0004408	8.11E-07
hap4	0.002770	0.0004088	0.003318
adr1	0.01014	0.001942	0.04090
usv1	0.006227	0.0002950	0.4218
gis1	0.006832	0.0002242	0.1630
msn2	0.006638	0.0004214	0.1299
msn4	0.005399	5.89E-05	0.9756
aft2	0.005403	0.0002528	0.9756
cat8/hap4	0.0008132	0.0003617	1.28E-05
adr1/cat8	0.01200	0.002458	0.008565
cat8/aft2	0.001540	6.53E-05	0.0001269
Genotype	ETOHmean	ETOHsem	ETOH_t_pval
WT	0.1580	0.02454	NA
snf1	0.0310	0.04200	0.03721
cat8	0.1444	0.06895	0.9175
hap4	0.08498	0.02998	0.2564
adr1	0.002097	0.07833	0.1080
usv1	0.1799	0.06412	0.8257
gis1	0.09483	0.01527	0.3110
msn2	0.01142	0.02772	0.003222
msn4	0.06948	0.01741	0.1044
aft2	0.1583	0.03766	0.9988
cat8/hap4	0.2582	0.005461	0.09193
adr1/cat8	-0.2485	0.03299	1.26E-06
cat8/aft2	0.2481	0.02454	0.08610
Genotype	BIOMASSmean	BIOMASSsem	BIOMASS_t_pval
WT	0.2379	0.01495	NA
snf1	0.2519	0.04480	0.8523

cat8	0.2219	0.02070	0.6935
hap4	0.1362	0.01832	0.003894
adr1	0.1831	0.009282	0.03189
usv1	0.1779	0.03296	0.1532
gis1	0.1812	0.01042	0.06257
msn2	0.1914	0.01184	0.06257
msn4	0.1675	0.02452	0.05645
aft2	0.1693	0.02661	0.07095
cat8/hap4	0.2089	0.01964	0.4277
adr1/cat8	0.14416	0.008730	0.00316
cat8/aft2	0.2095	0.008854	0.3734

Table 2.3: Low glucose phase measured rates of metabolite change. Units in grams analyte/hour.

2.6.3 Respiratory phase rates of metabolite change:

Genotype	GLUmean	GLUsem	GLU_t_pval
WT	-0.1491	0.01330	NA
snf1	-0.04858	0.01287	0.0002607
cat8	-0.05351	0.01058	0.0002269
hap4	7.53E-05	7.35E-05	0.0001152
adr1	-0.07057	0.01970	0.01194
usv1	-0.1620	0.001016	0.7308
gis1	-0.1453	0.003358	0.9389
msn2	-0.1661	0.003271	0.3609
msn4	-0.1423	0.003716	0.8397
aft2	-0.1469	0.0001007	0.9756
cat8/hap4	-0.02512	0.001392	0.0005899
adr1/cat8	0.0001488	0	2.44E-05
cat8/aft2	-0.03455	0.002254	0.0001088
Genotype	XYLmean	XYLsem	XYL_t_pval
WT	-0.2247	0.0317	NA
snf1	-0.03776	0.05824	0.02693
cat8	-0.1238	0.02082	0.05752
hap4	-0.0433	0.04537	0.01570
adr1	-0.1044	0.06913	0.2115
usv1	-0.1905	0.01198	0.6997
gis1	-0.1599	0.02209	0.3666
msn2	-0.1679	0.03185	0.3734
msn4	-0.1139	0.02407	0.1501
aft2	-0.1999	0.01390	0.7825
cat8/hap4	-0.1777	0.02773	0.5768
adr1/cat8	-0.1588	0.03047	0.4220
cat8/aft2	-0.2100	0.07106	0.9118
Genotype	XYLITOLmean	XYLITOLsem	XYLITOL_t_pval
WT	0.02406	0.002801	NA

snf1	0.004538	0.001899	0.0002260
cat8	0.05806	0.001988	5.89E-07
hap4	0.01938	0.003988	0.51248
adr1	0.02763	0.002729	0.5462
usv1	0.02567	0.001713	0.8597
gis1	0.02191	0.002204	0.7653
msn2	0.01972	0.002118	0.3840
msn4	0.01734	0.002403	0.3040
aft2	0.02819	0.001806	0.6550
cat8/hap4	0.04450	0.002613	0.006099
adr1/cat8	0.01934	0.004564	0.5462
cat8/aft2	0.04692	0.005475	0.004074
Genotype	GLYmean	GLYsem	GLY_t_pval
WT	0.01080	0.002991	NA
snf1	0.01048	0.001648	0.9756
cat8	0.003694	0.001527	0.1532
hap4	0.002914	0.002155	0.1980
adr1	0.006534	0.001086	0.4654
usv1	0.01082	0.002363	0.9988
gis1	0.01267	0.003509	0.8314
msn2	0.01212	0.002161	0.8434
msn4	0.01043	0.002149	0.9756
aft2	0.007961	0.0007762	0.7819
cat8/hap4	-0.001448	0.001398	0.09284
adr1/cat8	-0.0009731	0.0007099	0.06690
cat8/aft2	-0.005282	0.005142	0.03857
Genotype	ACETICmean	ACETICsem	ACETIC_t_pval
WT	0.002646	0.0001809	NA
snf1	0.001816	0.001151	0.5809
cat8	0.0005004	0.0001094	2.02E-07
hap4	0.002029	2.51E-05	0.1581
adr1	0.003177	0.0002555	0.1898
usv1	0.002669	7.75E-05	0.9756
gis1	0.003224	0.0002226	0.1541
msn2	0.003390	0.0002091	0.04019
msn4	0.002788	0.0001332	0.7742
aft2	0.002839	3.39E-05	0.7587
cat8/hap4	0.0002233	4.50E-05	3.80E-06
adr1/cat8	0.003420	0.0003505	0.1044
cat8/aft2	0.0001506	7.39E-05	2.97E-06
Genotype	ETOHmean	ETOHsem	ETOH_t_pval
WT	-0.1099	0.01905	NA
snf1	-0.1168	0.02537	0.9118
cat8	-0.1554	0.006761	0.1878

hap4	-0.1812	0.006568	0.08275
adr1	-0.1559	0.02827	0.3110
usv1	-0.1256	0.005367	0.7819
gis1	-0.09323	0.004446	0.7700
msn2	-0.08174	0.01266	0.4081
msn4	-0.1063	0.01731	0.9668
aft2	-0.08977	0.003533	0.7587
cat8/hap4	-0.1707	0.007069	0.1878
adr1/cat8	-0.07971	0.01374	0.5583
cat8/aft2	-0.1572	0.02818	0.3110
Genotype	BIOMASSmean	BIOMASSsem	BIOMASS_t_pval
WT	0.01788	0.007420	NA
snf1	0.008518	0.007927	0.5641
cat8	0.01175	0.006568	0.7171
hap4	0.0200	0.001792	0.9389
adr1	0.02547	0.002284	0.6218
usv1	0.02062	0.005903	0.9036
gis1	0.01583	0.008078	0.9389
msn2	0.01768	0.002857	0.9951
msn4	0.01625	0.009072	0.9561
aft2	0.01152	0.0005007	0.8092
cat8/hap4	0.00525	0.003329	0.4654
adr1/cat8	-0.003194	0.01224	0.2578
cat8/aft2	-0.007291	0.001755	0.1606

Table 2.4: Respiratory phase measured rates of metabolite change. Units in grams analyte/hour.

2.6.4 High glucose phase specific rates of metabolite change:

Genotype	GLUmean	GLUsem	GLU_t_pval
WT	-0.4317	0.01115	NA
snf1	-0.4223	0.01952	0.8034
cat8	-0.5339	0.006630	5.51E-06
hap4	-0.6068	0.01507	1.15E-06
adr1	-0.4794	0.01080	0.0220
usv1	-0.4114	0.007343	0.3584
gis1	-0.4393	0.01094	0.8424
msn2	-0.4207	0.007133	0.5559
msn4	-0.4700	0.02988	0.2810
aft2	-0.4241	0.02071	0.8486
cat8/hap4	-0.51118	0.008711	0.001128
adr1/cat8	-0.5127	0.01045	0.001213
cat8/aft2	-0.5296	0.009417	0.0002006
Genotype	XYLmean	XYLsem	XYL_t_pval
WT	-0.07901	0.01243	NA
snf1	-0.04307	0.03305	0.4060

cat8	-0.1011	0.004600	0.2965
hap4	-0.2150	0.04367	0.005094
adr1	-0.1159	0.01740	0.1797
usv1	-0.07915	0.007689	0.9970
gis1	-0.09444	0.07032	0.8803
msn2	-0.07475	0.01209	0.9004
msn4	-0.1766	0.02368	0.0068785
aft2	-0.03800	0.08872	0.6881
cat8/hap4	-0.08467	0.01658	0.9004
adr1/cat8	-0.07779	0.03094	0.9918
cat8/aft2	-0.08063	0.01554	0.9825
Genotype	XYLITOLmean	XYLITOLsem	XYLITOL_t_pval
WT	0.007286	0.0008039	NA
snf1	0.008860	0.003082	0.7318
cat8	0.04057	0.003150	8.63E-09
hap4	0.04904	0.002704	6.35E-11
adr1	0.01041	0.0009437	0.05078
usv1	0.003106	0.001899	0.06694
gis1	0.004278	0.001913	0.2000
msn2	0.005607	0.001286	0.4060
msn4	0.007073	0.0005546	0.9386
aft2	0.007023	0.0004801	0.9193
cat8/hap4	0.04771	0.001379	4.74E-13
adr1/cat8	0.03695	0.001361	7.77E-10
cat8/aft2	0.04686	0.0004497	6.06E-14
Genotype	GLYmean	GLYsem	GLY_t_pval
WT	0.04989	0.003097	NA
snf1	0.04755	0.003956	0.786786056
cat8	0.1194	0.001946	1.28E-11
hap4	0.06797	0.0005997	0.007887
adr1	0.05069	0.006095	0.9605
usv1	0.04746	0.001932	0.8042
gis1	0.06443	0.001011	0.04956
msn2	0.04761	0.001470	0.6881
msn4	0.04584	0.001625	0.5388
aft2	0.05628	0.002676	0.32728
cat8/hap4	0.1132	0.001129	2.16E-07
adr1/cat8	0.02137	0.003028	0.0001747
cat8/aft2	0.1128	0.0005845	4.37E-09
Genotype	ACETICmean	ACETICsem	ACETIC_t_pval
WT	0.002308	0.0001335	NA
snf1	0.002327	0.0003932	0.9911
cat8	0.0003702	0.0001837	4.20E-07
hap4	0.001070	1.79E-05	0.0006723

adr1	0.003311	0.0001508	0.0004690
usv1	0.002161	0.0001166	0.7122
gis1	0.002308	0.0001088	0.9996
msn2	0.002386	6.86E-05	0.7789
msn4	0.002421	0.0001380	0.7531
aft2	0.002722	0.0001198	0.1349
cat8/hap4	2.73E-06	5.23E-08	3.20E-08
adr1/cat8	0.002635	0.0004355	0.5167
cat8/aft2	2.78E-06	1.54E-08	1.17E-06
Genotype	ETOHmean	ETOHsem	ETOH_t_pval
WT	0.1691	0.006968	NA
snf1	0.1476	0.01312	0.2513
cat8	0.2071	0.006927	0.006346
hap4	0.2223	0.002148	0.001128
adr1	0.1861	0.01233	0.3408
usv1	0.1260	0.002189	0.01231
gis1	0.1396	0.003067	0.07458
msn2	0.1443	0.004583	0.02413
msn4	0.1299	0.006671	0.008868
aft2	0.1219	0.003876	0.00770
cat8/hap4	0.1870	0.008833	0.2565
adr1/cat8	0.1792	0.01999	0.7122
cat8/aft2	0.2057	0.009022	0.02096
Genotype	BIOMASSmean	BIOMASSsem	BIOMASS_t_pval
WT	0.1222	0.003283	NA
snf1	0.1204	0.006021	0.8816
cat8	0.09935	0.002307	0.0002274
hap4	0.12972	0.001676	0.2564
adr1	0.1224	0.009142	0.9970
usv1	0.1230	0.002806	0.9528
gis1	0.1228	0.004476	0.9770
msn2	0.1298	0.001888	0.1199
msn4	0.1151	0.0002426	0.3684
aft2	0.1179	0.0007409	0.6242
cat8/hap4	0.1055	0.0009395	0.03412
adr1/cat8	0.1386	0.003421	0.03042
cat8/aft2	0.1063	0.001000	0.01315

Table 2.5: High glucose phase measured specific rates of metabolite change. Units in grams analyte/grams biomass/hour.

2.6.5 Low glucose phase specific rates of metabolite change:

Genotype	GLUmean	GLUsem	GLU_t_pval
WT	-0.2042	0.007456	NA
snf1	-0.1402	0.01838	0.00783
cat8	-0.3333	0.007709	6.76E-09
hap4	-0.1962	0.001803	0.7122
adr1	-0.1489	0.02838	0.1013
usv1	-0.1532	0.002747	0.006878
gis1	-0.1707	0.006701	0.03150
msn2	-0.1512	0.003967	1.93E-05
msn4	-0.1587	0.005588	0.004345
aft2	-0.1869	0.0004651	0.3368
cat8/hap4	-0.3239	0.008456	3.66E-07
adr1/cat8	-0.006967	0.0008907	9.84E-10
cat8/aft2	-0.3133	0.004322	3.66E-07
Genotype	XYLmean	XYLsem	XYL_t_pval
WT	-0.08113	0.006543	NA
snf1	-0.1014	0.02055	0.3999
cat8	-0.1081	0.009998	0.06855
hap4	-0.1387	0.01347	0.002545
adr1	-0.08139	0.01188	0.9954
usv1	-0.04567	0.01167	0.03135
gis1	-0.03499	0.01246	0.009241
msn2	-0.03106	0.006784	0.0002193
msn4	-0.06715	0.006330	0.3999
aft2	-0.1469	0.05378	0.1980
cat8/hap4	-0.08542	0.003269	0.8435
adr1/cat8	-0.05265	0.009390	0.09195
cat8/aft2	-0.09707	0.01095	0.3272
Genotype	XYLITOLmean	XYLITOLsem	XYLITOL_t_pval
WT	0.008094	0.001089	NA
snf1	0.003814	0.0008607	0.02548
cat8	0.07616	0.002982	2.02E-13
hap4	0.05708	0.003749	6.15E-10
adr1	0.01148	0.0006106	0.05509
usv1	0.005142	0.002135	0.3065
gis1	0.006288	0.001784	0.5228
msn2	0.006351	0.0004519	0.2817
msn4	0.004064	0.0003970	0.05343
aft2	0.006290	0.0004941	0.5167
cat8/hap4	0.08243	0.001031	2.02E-13
adr1/cat8	0.01908	0.0007882	0.00039658
cat8/aft2	0.07569	0.001122	1.00E-14
Genotype	GLYmean	GLYsem	GLY_t_pval

WT	0.01950	0.001532	NA
snf1	0.01240	0.0006667	0.004962
cat8	0.04854	0.002495	5.97E-08
hap4	0.005921	0.0002787	0.0002926
adr1	0.01370	0.003112	0.1703
usv1	0.01656	0.0007795	0.4372
gis1	0.02730	0.0007553	0.03627
msn2	0.01688	0.0007377	0.2602
msn4	0.01430	0.0004788	0.1553
aft2	0.01600	0.002545	0.3584
cat8/hap4	0.04488	0.001206	1.31E-07
adr1/cat8	0.0001936	0.0001382	3.77E-05
cat8/aft2	0.04556	0.001667	1.84E-07
Genotype	ACETICmean	ACETICsem	ACETIC_t_pval
WT	0.001434	7.19E-05	NA
snf1	0.001784	0.0001647	0.1015
cat8	0.0001682	0.0001645	2.41E-06
hap4	0.0007509	0.0001330	0.0007309
adr1	0.002003	0.0001727	0.01123
usv1	0.001641	7.49E-05	0.2102
gis1	0.001803	6.94E-05	0.03997
msn2	0.001676	9.38E-05	0.1083
msn4	0.001393	2.50E-05	0.8726
aft2	0.001514	0.0001148	0.7078
cat8/hap4	0.0002675	0.0001193	1.27E-06
adr1/cat8	0.001913	0.0003221	0.1372
cat8/aft2	0.0005049	1.79E-05	5.51E-06
Genotype	ETOHmean	ETOHsem	ETOH_t_pval
WT	0.04252	0.006765	NA
snf1	0.02754	0.01594	0.5008
cat8	0.0515	0.02312	0.8091
hap4	0.02250	0.008148	0.2524
adr1	0.009062	0.01639	0.1115
usv1	0.04558	0.01579	0.9147
gis1	0.02054	0.006656	0.1505
msn2	0.003376	0.006975	0.002699
msn4	0.01827	0.004474	0.09872
aft2	0.04278	0.009321	0.9954
cat8/hap4	0.0833	0.007024	0.006572
adr1/cat8	-0.04261	0.00668	1.59E-05
cat8/aft2	0.08147	0.008095	0.01045
Genotype	BIOMASSmean	BIOMASSsem	BIOMASS_t_pval
WT	0.04196	0.002417	NA
snf1	0.04113	0.005895	0.9583

cat8	0.05026	0.003402	0.1161
hap4	0.02332	0.002644	0.001049
adr1	0.02775	0.003001	0.0057006
usv1	0.02962	0.004656	0.04773
gis1	0.03233	0.001678	0.04710
msn2	0.03203	0.001790	0.01094
msn4	0.02868	0.003491	0.01944
aft2	0.03037	0.003970	0.04710
cat8/hap4	0.04788	0.002436	0.2832
adr1/cat8	0.01454	0.0009660	2.65E-05
cat8/aft2	0.04533	0.001486	0.5148

Table 2.6: Low glucose phase measured specific rates of metabolite change. Units in grams analyte/grams biomass/hour.

2.6.6 Respiratory phase specific rates of metabolite change:

Genotype	GLUmean	GLUsem	GLU_t_pval
WT	-0.03126	0.002965	NA
snf1	-0.009996	0.002940	0.0004690
cat8	-0.01508	0.003442	0.007830
hap4	1.61E-05	1.61E-05	0.0002006
adr1	-0.01563	0.004504	0.02362
usv1	-0.03142	0.0003164	0.9950
gis1	-0.03357	0.001901	0.7759
msn2	-0.03440	0.001090	0.4545
msn4	-0.03181	0.001280	0.9583
aft2	-0.03440	0.0008461	0.6607
cat8/hap4	-0.006350	0.0002588	0.001236
adr1/cat8	2.06E-05	2.71E-07	0.0002006
cat8/aft2	-0.009061	0.0005444	0.0003965
Genotype	XYLmean	XYLsem	XYL_t_pval
WT	-0.04620	0.006630	NA
snf1	-0.006371	0.01362	0.03402
cat8	-0.03297	0.005493	0.2653
hap4	-0.01045	0.01019	0.02385
adr1	-0.02545	0.01473	0.30650
usv1	-0.04238	0.002539	0.8609
gis1	-0.03002	0.008117	0.2676
msn2	-0.03503	0.006736	0.38362
msn4	-0.02479	0.004632	0.1703
aft2	-0.04696	0.004116	0.9883
cat8/hap4	-0.04519	0.006741	0.9804
adr1/cat8	-0.02309	0.004625	0.13729
cat8/aft2	-0.05515	0.01804	0.7216
Genotype	XYLITOLmean	XYLITOLsem	XYLITOL_t_pval

WT	0.004916	0.0005541	NA
snf1	0.0007903	0.0003068	7.15E-05
cat8	0.014932	0.001083	3.66E-07
hap4	0.004300	0.0008342	0.7008
adr1	0.005101	0.0005095	0.9004
usv1	0.005514	0.0002121	0.71225
gis1	0.004940	0.0004550	0.9954
msn2	0.004032	0.0004177	0.3408
msn4	0.0038882	0.0005999	0.4322
aft2	0.006715	0.0003738	0.25646
cat8/hap4	0.01131	0.0005758	0.0001526
adr1/cat8	0.002812	0.0005521	0.09195
cat8/aft2	0.01342	0.0008718	1.39E-05
Genotype	GLYmean	GLYsem	GLY_t_pval
WT	0.002370	0.0007170	NA
snf1	0.002157	0.0003382	0.9004
cat8	0.0009957	0.0004074	0.2564
hap4	0.0006162	0.0004681	0.2301
adr1	0.001142	0.0001288	0.3408
usv1	0.002877	0.0006724	0.8022
gis1	0.001676	0.0004747	0.7500
msn2	0.002488	0.0004480	0.9583
msn4	0.00233147	0.0005062	0.9950
aft2	0.001827	0.0001195	0.8435
cat8/hap4	-0.0003786	0.0003556	0.1095
adr1/cat8	-5.55E-05	7.16E-05	0.15531
cat8/aft2	-0.001391	0.001295	0.04250
Genotype	ACETICmean	ACETICsem	ACETIC_t_pval
WT	0.0005442	2.88E-05	NA
snf1	0.0002910	0.0002078	0.2965
cat8	0.0001269	2.46E-05	3.40E-08
hap4	0.0004351	1.70E-05	0.05343
adr1	0.0005853	4.30E-05	0.5707
usv1	0.0006105	2.04E-05	0.2968
gis1	0.0007239	3.27E-05	0.00594
msn2	0.000695	3.70E-05	0.01234
msn4	0.0005949	1.85E-05	0.4934
aft2	0.0006544	1.10E-05	0.1703
cat8/hap4	5.63E-05	1.05E-05	2.51E-07
adr1/cat8	0.0005144	4.03E-05	0.7122
cat8/aft2	3.83E-05	1.87E-05	2.19E-07
Genotype	ETOHmean	ETOHsem	ETOH_t_pval
WT	-0.02291	0.003945	NA
snf1	-0.02355	0.004636	0.9726

cat8	-0.04240	0.002100	0.008263
hap4	-0.04213	0.0008925	0.04250
adr1	-0.02847	0.005803	0.5707
usv1	-0.02296	0.002609	0.9970
gis1	-0.02070	0.001200	0.87268
msn2	-0.01657	0.002490	0.3352
msn4	-0.02374	0.004149	0.9605
aft2	-0.02070	0.0006198	0.8931
cat8/hap4	-0.04456	0.002333	0.02548
adr1/cat8	-0.01183	0.001150	0.2510
cat8/aft2	-0.04147	0.007428	0.06449
Genotype	BIOMASSmean	BIOMASSsem	BIOMASS_t_pval
WT	0.002538	0.001030	NA
snf1	0.0008083	0.001010	0.39000
cat8	0.002165	0.001183	0.9004
hap4	0.001957	0.0007562	0.8435
adr1	0.003206	0.0003044	0.7820
usv1	0.002935	0.0008657	0.9004
gis1	0.002381	0.001223	0.9770
msn2	0.002387	0.0003701	0.9583
msn4	0.002459	0.001362	0.9918
aft2	0.001752	1.05E-05	0.8435
cat8/hap4	0.000928	0.0005724	0.4943
adr1/cat8	0.0005621	0.0007871	0.3972
cat8/aft2	-0.001263	0.0003105	0.1199

Table 2.7: Respiratory phase measured specific rates of metabolite change. Units in grams analyte/grams biomass/hour.

2.6.7 High glucose phase carbon commitment ratios:

Genotype	GLUmean	GLUsem	GLU_t_pval
WT	0	0	NA
snf1	0	0	NA
cat8	0	0	NA
hap4	0	0	NA
adr1	0	0	NA
usv1	0	0	NA
gis1	0	0	NA
msn2	0	0	NA
msn4	0	0	NA
aft2	0	0	NA
cat8/hap4	0	0	NA
adr1/cat8	0	0	NA
cat8/aft2	0	0	NA
Genotype	XYLmean	XYLsem	XYL_t_pval

WT	0	0	NA
snf1	0	0	NA
cat8	0	0	NA
hap4	0	0	NA
adr1	0	0	NA
usv1	0	0	NA
gis1	0	0	NA
msn2	0	0	NA
msn4	0	0	NA
aft2	0	0	NA
cat8/hap4	0	0	NA
adr1/cat8	0	0	NA
cat8/aft2	0	0	NA
Genotype	XYLITOLmean	XYLITOLsem	XYLITOL_t_pval
WT	0.01352	0.001159	NA
snf1	0.01798	0.006360	0.6007
cat8	0.06383	0.005148	1.60E-08
hap4	0.05779	0.001890	3.37E-10
adr1	0.01805	0.002427	0.1581
usv1	0.006776	0.004142	0.1003
gis1	0.002302	0.002291	0.001625
msn2	0.01198	0.002720	0.70595
msn4	0.01002	0.001548	0.17361
aft2	0.01148	0.003382	0.6312
cat8/hap4	0.07738	0.002978	4.09E-12
adr1/cat8	0.06229	0.004966	1.13E-08
cat8/aft2	0.07632	0.002879	2.10E-12
Genotype	GLYmean	GLYsem	GLY_t_pval
WT	0.09462	0.006691	NA
snf1	0.1000	0.01060	0.7567
cat8	0.1860	0.003726	1.60E-08
hap4	0.07209	0.002692	0.1523
adr1	0.08730	0.01397	0.7247
usv1	0.1033	0.007475	0.6183
gis1	0.1287	0.01808	0.09638
msn2	0.09577	0.003405	0.9166
msn4	0.07439	0.003687	0.2014
aft2	0.11525	0.02015	0.3603
cat8/hap4	0.18255	0.007486	1.21E-05
adr1/cat8	0.02747	0.002899	0.0003756
cat8/aft2	0.1823	0.007601	4.61E-06
Genotype	ACETICmean	ACETICsem	ACETIC_t_pval
WT	0.004386	0.0001036	NA
snf1	0.004995	0.0009385	0.6183

cat8	0.0006000	0.0002980	7.34E-10
hap4	0.001417	0.0001174	1.64E-10
adr1	0.005610	0.0002493	0.0004755
usv1	0.004539	0.0002024	0.6312
gis1	0.004586	0.0007391	0.8019
msn2	0.004747	0.0002387	0.27212
msn4	0.003437	0.0004443	0.03294
aft2	0.005373	0.0005067	0.04521
cat8/hap4	4.74E-06	2.55E-07	2.33E-13
adr1/cat8	0.004527	0.0008047	0.87744
cat8/aft2	4.53E-06	1.89E-07	2.33E-13
Genotype	ETOHmean	ETOHsem	ETOH_t_pval
WT	0.4239	0.01551	NA
snf1	0.3637	0.01128	0.03796
cat8	0.4281	0.01037	0.8947
hap4	0.3259	0.009744	0.01130
adr1	0.4075	0.02054	0.6591
usv1	0.3561	0.01203	0.04443
gis1	0.3133	0.01155	0.005537
msn2	0.4006	0.02470	0.5973
msn4	0.2821	0.004013	0.0007962
aft2	0.2217	0.009443	6.85E-06
cat8/hap4	0.4189	0.009825	0.8942
adr1/cat8	0.4118	0.01767	0.7567
cat8/aft2	0.4395	0.009425	0.6492
Genotype	BIOMASSmean	BIOMASSsem	BIOMASS_t_pval
WT	0.2546	0.01477	NA
snf1	0.2714	0.0175	0.6270
cat8	0.1687	0.004961	0.0005967
hap4	0.1564	0.005801	0.007645
adr1	0.2200	0.01434	0.2050
usv1	0.2796	0.001897	0.5041
gis1	0.2707	0.01626	0.6680
msn2	0.2831	0.01029	0.2186
msn4	0.2072	0.01080	0.1799
aft2	0.26051	0.03459	0.9025
cat8/hap4	0.2039	0.01158	0.08722
adr1/cat8	0.2430	0.01896	0.7516
cat8/aft2	0.1873	0.008329	0.02207
Genotype	RESIDUALmean	RESIDUALsem	RESIDUAL_t_pval
WT	0.2087	0.02893	NA
snf1	0.2111	0.04785	0.9683
cat8	0.1508	0.01276	0.2293
hap4	0.3852	0.01605	0.006322

adr1	0.2614	0.02917	0.3462
usv1	0.2584	0.01632	0.4436
gis1	0.2947	0.07261	0.3185
msn2	0.2066	0.03563	0.9683
msn4	0.4204	0.01779	0.004579
aft2	0.35680	0.06945	0.07375
cat8/hap4	0.1025	0.02677	0.07378
adr1/cat8	0.20287	0.01042	0.9304
cat8/aft2	0.1144	0.02092	0.1002

Table 2.8: High glucose phase carbon commitment ratios. Mean values are expressed on a per genotype basis as a fraction of total input carbon detected in analyte output.

2.6.8 Low glucose phase carbon commitment ratios:

Genotype	GLUmean	GLUsem	GLU_t_pval
WT	0	0	NA
snf1	0	0	NA
cat8	0	0	NA
hap4	0	0	NA
adr1	0	0	NA
usv1	0	0	NA
gis1	0	0	NA
msn2	0	0	NA
msn4	0	0	NA
aft2	0	0	NA
cat8/hap4	0	0	NA
adr1/cat8	0	0	NA
cat8/aft2	0	0	NA
Genotype	XYLmean	XYLsem	XYL_t_pval
WT	0	0	NA
snf1	0	0	NA
cat8	0	0	NA
hap4	0	0	NA
adr1	0	0	NA
usv1	0	0	NA
gis1	0	0	NA
msn2	0	0	NA
msn4	0	0	NA
aft2	0	0	NA
cat8/hap4	0	0	NA
adr1/cat8	0	0	NA
cat8/aft2	0	0	NA
Genotype	XYLITOLmean	XYLITOLsem	XYLITOL_t_pval
WT	0.02780	0.003196	NA
snf1	0.01457	0.003052	0.02869

cat8	0.16331	0.003734	2.33E-13
hap4	0.17308	0.01469	8.40E-09
adr1	0.05507	0.008892	0.01421
usv1	0.01149	0.004092	0.04390
gis1	0.03291	0.01079	0.68394
msn2	0.03152	0.004508	0.6474
msn4	0.01942	0.001170	0.2678
aft2	0.01891	0.004944	0.2307
cat8/hap4	0.1917	0.002876	3.46E-12
adr1/cat8	0.1453	0.01175	1.60E-08
cat8/aft2	0.1829	0.006955	2.79E-12
Genotype	GLYmean	GLYsem	GLY_t_pval
WT	0.06749	0.005184	NA
snf1	0.04833	0.006769	0.08472
cat8	0.1071	0.006651	0.0008233
hap4	0.01705	0.001473	0.0001221
adr1	0.05291	0.01253	0.3777
usv1	0.08258	0.01107	0.2838
gis1	0.11876	0.01962	0.01307
msn2	0.08993	0.007987	0.06165
msn4	0.07101	0.006263	0.7976
aft2	0.06206	0.01809	0.8025
cat8/hap4	0.1047	0.003931	0.001185
adr1/cat8	0.0006359	0.0003789	2.89E-05
cat8/aft2	0.1083	0.0008388	0.002393
Genotype	ACETICmean	ACETICsem	ACETIC_t_pval
WT	0.005081	0.0002723	NA
snf1	0.008840	0.0009790	0.002232
cat8	0.0005912	0.0002948	1.36E-08
hap4	0.002298	0.0004116	0.0001716
adr1	0.009934	0.001869	0.02323
usv1	0.007704	0.0006988	0.002556
gis1	0.007227	0.001620	0.1618
msn2	0.009443	0.001501	0.02500
msn4	0.007295	0.0006187	0.005313
aft2	0.005625	0.0012908	0.6944
cat8/hap4	0.0006153	0.0002729	2.00E-07
adr1/cat8	0.01641	0.004337	0.005834
cat8/aft2	0.001208	6.05E-05	1.82E-06
Genotype	ETOHmean	ETOHsem	ETOH_t_pval
WT	0.19175	0.02640	NA
snf1	0.1373	0.04812	0.4434
cat8	0.1787	0.04781	0.8774
hap4	0.1091	0.04585	0.1997

adr1	0.09888	0.04054	0.1213
usv1	0.196	0.07429	0.9535
gis1	0.1577	0.01737	0.6313
msn2	0.07782	0.02287	0.01261
msn4	0.1567	0.04345	0.6312
aft2	0.1630	0.01867	0.6312
cat8/hap4	0.2706	0.009417	0.1523
adr1/cat8	0	0	0.001625
cat8/aft2	0.2568	0.02024	0.2181
Genotype	BIOMASSmean	BIOMASSsem	BIOMASS_t_pval
WT	0.1617	0.01388	NA
snf1	0.2097	0.02276	0.1440
cat8	0.1093	0.01163	0.03246
hap4	0.07627	0.009949	0.003680
adr1	0.1275	0.01162	0.1628
usv1	0.1104	0.01911	0.1457
gis1	0.19231	0.01009	0.3699
msn2	0.1770	0.01292	0.5973
msn4	0.1510	0.02201	0.7701
aft2	0.09440	0.005475	0.02428
cat8/hap4	0.1132	0.01050	0.08075
adr1/cat8	0.1329	0.01676	0.3487
cat8/aft2	0.1187	0.006969	0.1083
Genotype	RESIDUALmean	RESIDUALsem	RESIDUAL_t_pval
WT	0.5461	0.02659	NA
snf1	0.5551	0.03727	0.8947
cat8	0.4049	0.04625	0.02869
hap4	0.6166	0.04234	0.2659
adr1	0.6555	0.03184	0.03907
usv1	0.5407	0.05908	0.9416
gis1	0.5173	0.02756	0.6677
msn2	0.5727	0.03318	0.6677
msn4	0.6513	0.01583	0.09656
aft2	0.6484	0.02641	0.06443
cat8/hap4	0.3590	0.007799	0.005313
adr1/cat8	0.6520	0.03064	0.06359
cat8/aft2	0.3210	0.002893	0.001444

Table 2.9: High glucose phase carbon commitment ratios. Mean values are expressed on a per genotype basis as a fraction of total input carbon detected in analyte output.

2.6.9 Respiratory phase carbon commitment ratios:

Genotype	GLUmean	GLUsem	GLU_t_pval
WT	0	0	NA
snf1	0	0	NA

cat8	0	0	NA
hap4	0	0	NA
adr1	0	0	NA
usv1	0	0	NA
gis1	0	0	NA
msn2	0	0	NA
msn4	0	0	NA
aft2	0	0	NA
cat8/hap4	0	0	NA
adr1/cat8	0	0	NA
cat8/aft2	0	0	NA
Genotype	XYLmean	XYLsem	XYL_t_pval
WT	0	0	NA
snf1	0	0	NA
cat8	0	0	NA
hap4	0	0	NA
adr1	0	0	NA
usv1	0	0	NA
gis1	0	0	NA
msn2	0	0	NA
msn4	0	0	NA
aft2	0	0	NA
cat8/hap4	0	0	NA
adr1/cat8	0	0	NA
cat8/aft2	0	0	NA
Genotype	XYLITOLmean	XYLITOLsem	XYLITOL_t_pval
WT	0.04064	0.004802	NA
snf1	0.01417	0.004595	0.004036
cat8	0.1334	0.01838	0.0001477
hap4	0.07079	0.01882	0.1140
adr1	0.05091	0.01013	0.4510
usv1	0.04221	0.007740	0.9100
gis1	0.04265	0.004003	0.8774
msn2	0.03887	0.004474	0.8769
msn4	0.03377	0.008468	0.6149
aft2	0.04116	0.007351	0.9683
cat8/hap4	0.1130	0.01312	8.28E-05
adr1/cat8	0.08181	0.02430	0.08075
cat8/aft2	0.08859	0.01220	0.002313
Genotype	GLYmean	GLYsem	GLY_t_pval
WT	0.01302	0.003898	NA
snf1	0.03495	0.008821	0.06247
cat8	0.006011	0.002096	0.2904
hap4	0.004370	0.004138	0.3462

adr1	0.02885	0.008357	0.1722
usv1	0.03643	0.01203	0.09023
gis1	0.02757	0.007196	0.1409
msn2	0.02391	0.005037	0.1935
msn4	0.02077	0.006424	0.4283
aft2	0.01717	0.001671	0.6970
cat8/hap4	0.001787	0.001787	0.1889
adr1/cat8	0.0003550	0.0002205	0.09068
cat8/aft2	0	0	0.1239
Genotype	ACETICmean	ACETICsem	ACETIC_t_pval
WT	0.005007	0.0002636	NA
snf1	0.006249	0.002377	0.6862
cat8	0.0007528	6.90E-05	1.38E-08
hap4	0.007984	0.0005475	0.000580
adr1	0.005997	0.001095	0.4383
usv1	0.005609	0.0002703	0.3462
gis1	0.007296	0.0008452	0.01692
msn2	0.007769	0.0009707	0.03853
msn4	0.006619	0.0006016	0.03073
aft2	0.00592	0.0001432	0.1313
cat8/hap4	0.0006521	0.0001575	4.67E-07
adr1/cat8	0.01038	0.001735	0.003529
cat8/aft2	6.98E-05	6.72E-05	5.87E-07
Genotype	ETOHmean	ETOHsem	ETOH_t_pval
WT	0	0	NA
snf1	0	0	NA
cat8	0	0	NA
hap4	0	0	NA
adr1	0	0	NA
usv1	0	0	NA
gis1	0	0	NA
msn2	0	0	NA
msn4	0	0	NA
aft2	0	0	NA
cat8/hap4	0	0	NA
adr1/cat8	0	0	NA
cat8/aft2	0	0	NA
Genotype	BIOMASSmean	BIOMASSsem	BIOMASS_t_pval
WT	0.02974	0.008777	NA
snf1	0.03157	0.01196	0.9248
cat8	0.02466	0.007199	0.7701
hap4	0.03434	0.008993	0.8321
adr1	0.03777	0.01038	0.6839
usv1	0.03430	0.008727	0.8321

gis1	0.04316	0.004223	0.5680
msn2	0.02752	0.005803	0.8942
msn4	0.008362	0.005840	0.3144
aft2	0.01884	0.0009880	0.6816
cat8/hap4	0.005601	0.003559	0.1898
adr1/cat8	0.004691	0.004691	0.2260
cat8/aft2	0	0	0.147962972
Genotype	RESIDUALmean	RESIDUALsem	RESIDUAL_t_pval
WT	0.9020	0.013690	NA
snf1	0.9166	0.02057	0.6750
cat8	0.8296	0.02243	0.02502
hap4	0.8733	0.02992	0.4677
adr1	0.8906	0.02015	0.7452
usv1	0.8988	0.01164	0.9248
gis1	0.8865	0.01700	0.6474
msn2	0.8913	0.01854	0.7516
msn4	0.9171	0.02347	0.6816
aft2	0.9072	0.009879	0.8942
cat8/hap4	0.8775	0.01226	0.4944
adr1/cat8	0.8791	0.03868	0.6427
cat8/aft2	0.9084	0.01321	0.8774

Table 2.10: Respiratory phase carbon commitment ratios. Mean values are expressed on a per genotype basis as a fraction of total input carbon detected in analyte output.

Chapter 3: Functional decomposition of synthetic promoters allows accurate expression prediction

3.1 Abstract

Quantitative control over gene expression is a fundamental requirement for the design of complex biological systems. We demonstrate that synthetic promoters can be functionally decomposed into components that can be characterized in isolation and used to train a composite model capable of predicting the behavior of the complete promoter. We characterized response functions for two types of components individually, and then constructed cells expressing both types, and used them to drive a fluorescent reporter in *S. cerevisiae*. Our model predicted expression in this novel genotype in response to simultaneous changes in concentrations of activating and repressive *trans*-factors. The model explained 72.8% of observed variance in expression and provided evidence of independence between modules of the synthetic promoter architecture used within this study. We utilize this property of modular independence to reduce the number of measurements required for expression modeling by up to 81.2% versus non-independent architectures.

3.2 Introduction

An understanding of the quantitative relationship between *cis*-regulatory DNA sequence, *trans*-factor concentration, and mRNA output will be required for engineering complex gene regulation systems. Attempts to quantify these relationships fall into two categories – top-down and bottom-up. The top-down strategy uses large numbers of synthetic *cis*-regulatory sequences within a limited number of *trans* factor states to learn general principles of cis-regulation and to

train mathematical models designed to predict expression from novel promoters (Gertz, Siggia, & Cohen, 2009) (Sharon, et al., 2012) (Mogno, Kwasnieski, & Cohen, 2013) (Kheradpour, et al., 2013). The alternative bottom-up approach focuses on a small number of synthetic *cis*-sequences while measuring output at numerous *trans*-factor input concentrations to train models focused on enabling the engineering of multi-component systems (Ellis, Wang, & Collins, 2009) (Guido, et al., 2006) (Murphy, Balázsi, & Collins, 2007). Both approaches have been successfully utilized to predict the expression and behavior of biological systems, but the immense effort and expense required to train a predictive model has restricted the widespread application of quantitative expression modeling and synthetic biology.

One of the primary challenges preventing the widespread adoption of expression modeling is the large number of expression measurements required to learn the complex interaction rules describing expression output for a library of *cis*-sequences when numerous *trans* factors are present across a range of concentrations. Characterizing this space by brute force requires observing the effect of each *trans* factor in isolation and in interaction with all other *trans*-factors that act on each promoter, leading to an explosion in the number of experiments required to generate a model. This explosion results from the potential for non-independent interactions between concentrations of different TFs and binding-site configurations. Transcription factors (TFs) have evolved to form protein-protein interactions, allowing DNA-bound TFs to recruit additional TFs and components of the basal transcriptional machinery (Fig. 3.1.A) (Ptashne & Gann, 1997). This results in cooperativity and non-linear response functions. Also, changes within the *cis* component cannot be assumed to exhibit independence; alterations in DNA sequence can generate new sites for alternative TFs or alter local three dimensional structure, thereby altering TF affinity for binding sites without changing the binding site sequences

(Matthew, Zhou, L, Dantas-Machado, Gordân, & Rohs, 2014). These complex interactions make it difficult to accurately predict the effects of sequence changes on how promoters respond to various TF concentrations (Fig. 3.1.B).

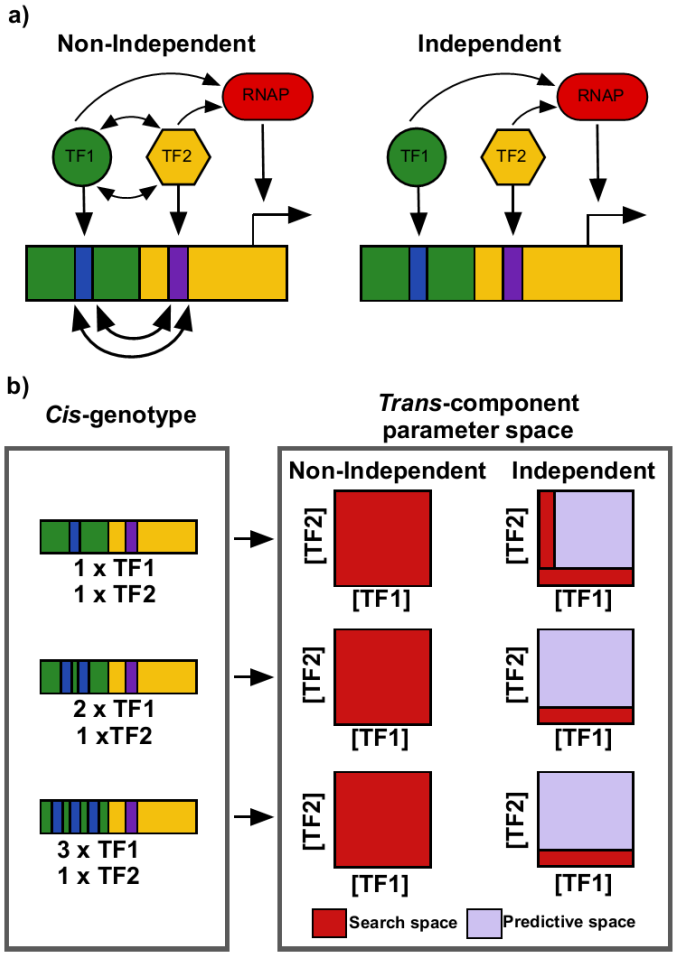

Figure 3.1. Panel A: Non-independent TF architectures exhibit physical TF:TF interactions and functional DNA:DNA interactions. Independent architectures minimize TF:TF, DNA-DNA interactions. Panel B: Independence within a modular promoter library reduces the measurement space required for modeling by avoiding the need to characterize interactions. The TF1 sites vary across the library (blue, 1-3 sites). TF2 sites are held constant (purple, 1 site). Due to independence the TF2:DNA function is only characterized once, while the TF1:DNA function is generated for each new genotype.

Efficient design of complex transcriptional circuits depends on the availability of *cis* and *trans* components that can be characterized independently and combined to produce a system with predictable behavior. There have been initial attempts to characterize the responses of multiple

promoters to changes in TF activity, but to our knowledge there have been no demonstrations of independence between promoter components and trans factor concentrations (Ellis, Wang, & Collins, 2009). Working with a combinatoric synthetic promoter library in *S. cerevisiae*, Mogno et al. observed that the TATA box acts as a modular and noise-less scaling device, demonstrating that the transition from weak to strong TATA boxes increased expression without altering the rules of upstream *cis*-regulation (Mogno, Vallania, Mitra, & Cohen, 2010). On the other hand, Ellis et al. demonstrated that changes in the nucleotides surrounding the TATA box of the *S. cerevisiae GAL1* promoter were sufficient to alter both basal expression and response to the LacI repressor, even though the TATA box and LacI operators were not modified (Ellis, Wang, & Collins, 2009). The demonstration that 86% of single nucleotide substitutions in the mammalian *Rho* promoter generated significant changes in expression casts doubt on whether the functions of any sites are independent of sequence changes to other parts of the promoter (Kwasnieski, Mogno, Myers, Corbo, & Cohen, 2012).

We set out to develop promoter architectures in which sites that are responsive to different *trans* components can be characterized separately and combined to produce a system with predictable behavior. We therefore constructed a small promoter library with one or more sites for the ZifH and LacI DNA binding domains. We studied the responses of these promoters to varying concentrations of *E. coli* LacI and to a fusion of the ZifH DNA-binding domain to the VP64 activation domain or the *SSN6* repression domain (Ellis, Wang, & Collins, 2009) (Ajo-Franklin, et al., 2007). We selected these TFs because we thought that their origin in highly diverged species and their distinct mechanisms of transcriptional activation/repression would minimize the likelihood of direct TF-TF interactions. We also designed the promoter library to localize the activity of each TF to specific *cis*-regulatory modules (CRM), each containing one or more sites

for that TF. Each promoter was composed of two CRMs and a basal region required for RNA polymerase (RNAP) recruitment. We then characterized the response of each CRM to titration of the TF it was designed to interact with in a strain lacking the other TF. We also measured the basal expression level from each promoter in the absence of either TF. The component response functions were combined to predict expression in a novel genotype in which both TFs were expressed across a range of concentrations under control of a single drug. The full system containing both CRMs and both TFs demonstrated two types of independence. First, the responses of promoters to LacI titration were independent of the ZifH CRM. Second, the responses to simultaneous titration of both TFs could be predicted from the responses to titration of each TF alone.

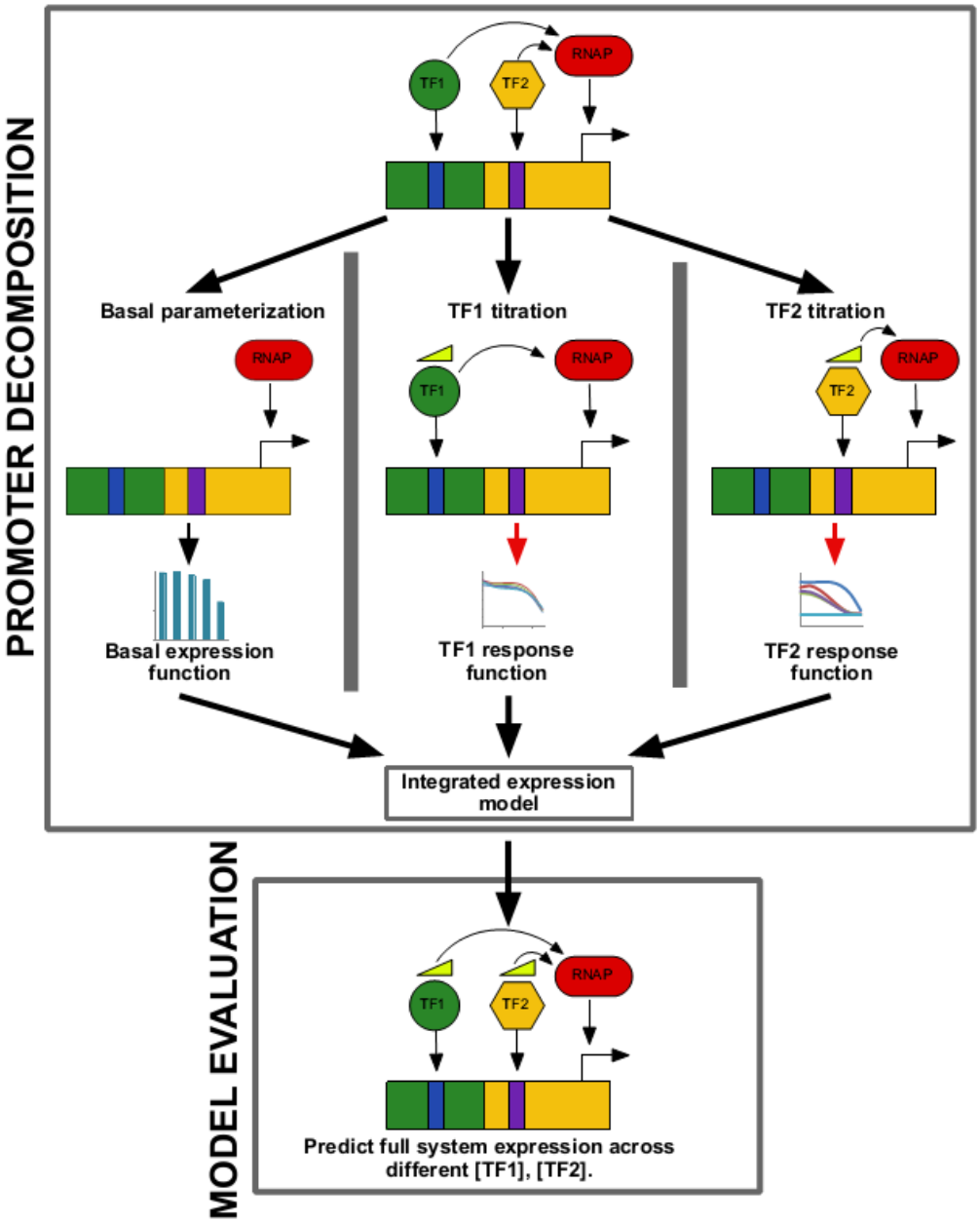

Figure 3.2. Promoter decomposition and evaluation process. Response functions for each of the components of the synthetic promoter are fitted in isolation. Basal expression is characterized in the absence of all *trans* factors. TF1&TF2 response functions are characterized in isolation through titration across a range of concentrations. The model is evaluated on its ability to predict a novel genotype in which both TFs are expressed.

3.3 Results

3.3.1 Construction of synthetic promoter libraries and TF titration systems.

A synthetic promoter library was generated through the assembly of two modular components. Each synthetic promoter consisted of a ZifH module containing 1-5 consensus ZifH binding sites, spaced 8 nucleotides apart and one LacI core (Fig. 3.3). In ZifH modules containing less than 5 binding sites, the extra upstream sites were replaced with a randomly generated sequence held constant for each ZifH position across the library. Although the LX core from Ellis et al. was used in this study for all synthetic promoters, we also evaluated the assembly of alternative cores with the ZifH upstream sequence (Ellis, Wang, & Collins, 2009). Pilot evaluations of this assembly indicated that not all LacI cores were structurally compatible with the upstream ZifH modules. Specifically, the assembled DNA was unstable, leading to consistent, large segmental deletions within the inserted cores. We found the LX (wild-type *GAL1* sequence with inserted LacO sites) core to be compatible with all tested upstream ZifH modules and utilized this core for all further experiments and analyses. The ZifH modules and LX cores were assembled via overlap extension PCR, cloned into plasmids downstream of the *GAL1* UAS and integrated at the URA3 locus. We use ZiLX to describe the promoter genotypes generically. When referring to a specific member of the promoter library, we substitute *i* for a number that indicates the number of ZifH sites within the ZifH responsive CRM.

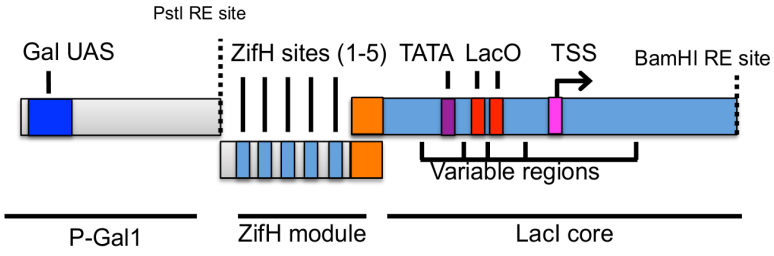

Figure 3.3. Synthetic promoter architecture used within this study. Each promoter consists of two modules. The ZifH module contains 1-5 ZifH consensus sites. The LacI core module consists of two consensus Lac operators,

cloned downstream of the GAL1 TATA box (Ellis, Wang, & Collins, 2009). Synthetic promoters are cloned downstream of the *GAL1* UAS, enabling galactose induction and glucose repression.

Characterizing the response functions of the components required the ability to titrate *trans*-factor concentrations. We utilized the tet-VP16, Tet-Off system for all titrations (Fig. 3.4) (Gossen & Bujard, 1992). This system was integrated at the TRP1 locus, controlling expression of the titrated *trans*-component as a function of anhydrotetracycline (ATc) concentration. Both engineered variants of the ZifH transcription factor were titrated against the synthetic promoter library (Ajo-Franklin, et al., 2007) (Bellí, Piedrafita, Aldea, & Herrero, 1998). The yeast codon optimized LacI from Ellis et al. was used for LX core repression. All constructs were sequence confirmed following integration. See table 3.1 for all strains used within this study.

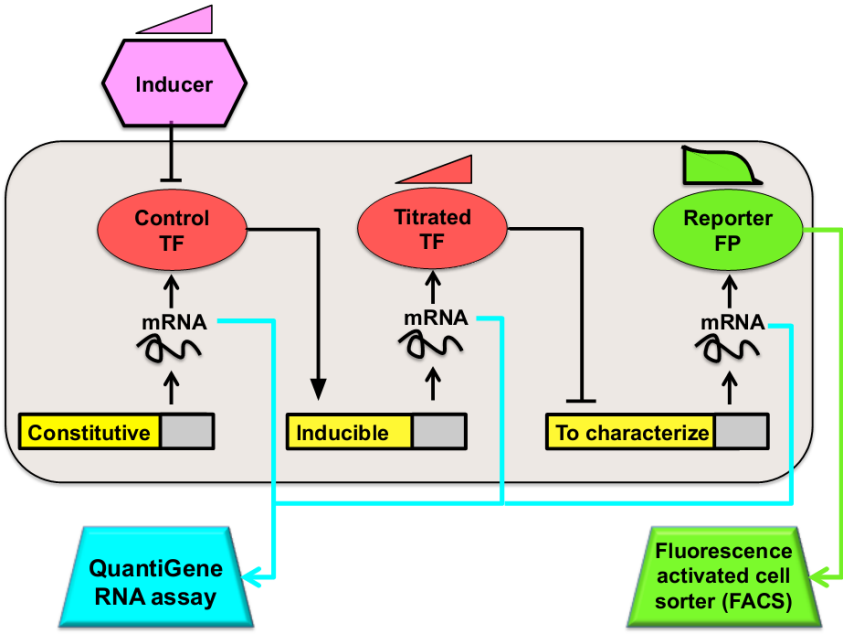

Figure 3.4. Synthetic promoter characterization scheme used in this study. Promoter libraries are integrated into the S. cerevisiae genome and the TF:promoter response function is generated by titrating ZifH or LacI. TF titration is achieved through ATc modulation of the constitutively expressed tet-VP16 system (Gossen & Bujard, 1992). mRNA levels of all system components are read out via the Quantigene RNA assay and promoter library output characterized by flow cytometry.

Strain ID	Promoter	LacI	ZifH-Vp64	ZifH-SSN6	Function
sDM79	Z1LX	-	-	-	Basal
sDM81	Z2LX	-	-	-	Basal
sDM83	Z3LX	-	-	-	Basal

sDM85	Z4LX	-	-	-	Basal
sDM87	Z5LX	-	-	-	Basal
sDM31	Z1LX	+	-	-	LacI
sDM42	Z2LX	+	-	-	LacI
sDM44	Z3LX	+	-	-	LacI
sDM47	Z4LX	+	-	-	LacI
sDM27	Z5LX	+	-	-	LacI
sDM48	Z1LX	-	+	-	ZifH-VP64
sDM49	Z2LX	-	+	-	ZifH-VP64
sDM50	Z3LX	-	+	-	ZifH-VP64
sDM51	Z4LX	-	+	-	ZifH-VP64
sDM52	Z5LX	-	+	-	ZifH-VP64
sDM58	Z1LX	-	-	+	ZifH-SSN6
sDM59	Z2LX	-	-	+	ZifH-SSN6
sDM60	Z3LX	-	-	+	ZifH-SSN6
sDM61	Z4LX	-	-	+	ZifH-SSN6
sDM62	Z5LX	-	-	+	ZifH-SSN6
sDM68	Z1LX	+	+	-	LacI/ZifH-VP64
sDM69	Z2LX	+	+	-	LacI/ZifH-VP64
sDM70	Z3LX	+	+	-	LacI/ZifH-VP64
sDM71	Z4LX	+	+	-	LacI/ZifH-VP64
sDM72	Z5LX	+	+	-	LacI/ZifH-VP64

Table 3.1. Strains used within this study.

3.3.2 The tet-VP16 system allows precise titration of transcription factor concentration across a physiologically relevant range

After constructing strains with synthetic promoters and the TF titration machinery, we wanted to know whether the changes in TF concentration achieved with ATc titration were comparable in magnitude to the changes in TF concentration during natural transcriptional responses. We therefore analyzed RNA-seq data collected from an *S. cerevisiae* CEN.PK2-1D strain grown in synthetic complete medium supplemented with five different carbon sources: 5% glucose, 5% galactose, 5% glycerol, 1.3% ethanol, 5% xylose. This showed that transcription factors were expressed at a broad range of levels, from zero expression to the 99th percentile of all genes, with a median expression in the 29th percentile of all genes (Fig. 3.5.A). For each gene, we calculated the maximum fold change observed between any pair of carbon sources. This analysis indicated

a median fold change of 2.3 for both TFs and non-TFs (Fig. 3.5.B).

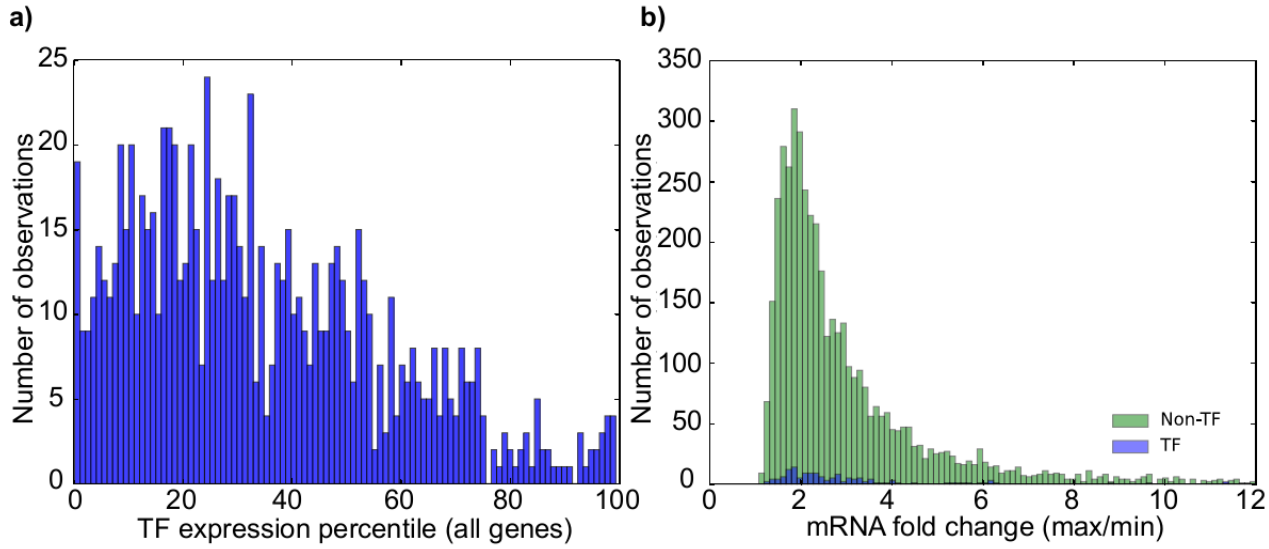

Figure 3.5. Panel A: Combined expression percentiles of all *S. cerevisiae* transcription factors across six environmental carbon source conditions. Panel B: Maximum mRNA fold change of Non-TFs and TFs across six environmental carbon source conditions.

We found the the tetVP16 system was able to generate a maximum 1.9 fold change in LacI mRNA levels when tested within a strain titrating LacI against the Z5LX synthetic promoter. This was similar to the median fold change observed for native TFs in response to carbon source variation and generated a 10.1 fold change in GFP reporter mRNA expression (Fig. 3.6).

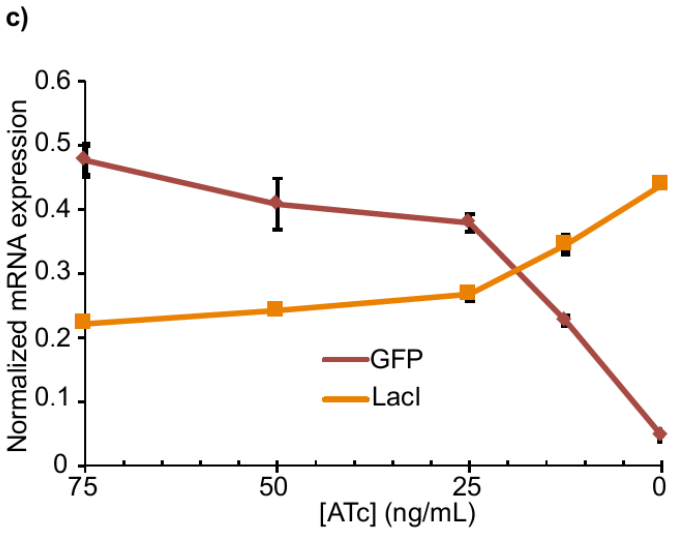

Figure 3.6. Quantification of mRNA expression levels during the titration of LacI on the Z5LX synthetic promoter controlling GFP expression.

3.3.3 Characterization of synthetic promoter basal expression and transcription factor response functions

We characterized the behavior each component of the system (basal level, LacI response, and ZifH response) by titrating each trans-factor in isolation and measuring expression of GFP by flow cytometry. Promoter expression in the absence of all TFs was calculated by measuring GFP expression during the ATc titration in a strain that did not express any of the TFs. Originally, we used this as the basal expression level in our model. However, we found that slightly better predictive results were obtained when using a basal level calculated by averaging the reporter expression levels in the two strains containing a single TF, measured at minimal TF expression (250 ng/mL ATc). Since this approach also reduces the number of strains and measurements needed to construct the model, we use it hereafter. Relative to the Z1LX promoter, we observed a 43% reduction in basal GFP expression for the Z5LX promoter (Fig. 3.7).

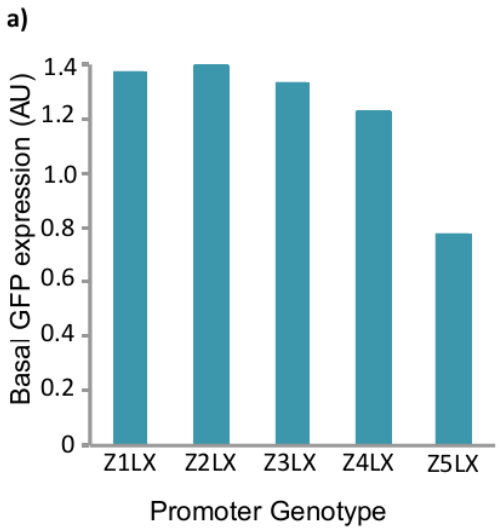

Figure 3.7: Expression of the ZiLX library members in the absence of *trans*-inputs. Expression values were calculated from the mean of GFP expression observed at maximal repression of LacI or ZifH within the single TF titration strains (referred to as 'basal expression' hereafter in the text).

The titration of LacI exhibited a clear repressive effect that did not vary greatly across the promoter genotypes (Fig. 3.8.A). On average, repression saturated at 62% reduction from basal expression. In light of the different basal expressions of the promoter library, the relatively constant LacI response functions indicate successful construction of a system with a LacI response that is independent of the specific 5' ZifH CRM chosen.

The ZifH-SSN6 titration indicated that this chimeric protein was a strong repressor, capable of generating functional changes in promoter expression even when maximally repressed by ATc (Fig. 3.8.B). We observed a clear monotonic effect of genotype within the titration, with the addition of each site increasing ZifH-SSN6 repression efficiency. Repression saturated at mean of 71% repression of basal expression when averaged across all genotypes. Notably, the Z5LX genotype exhibited maximal repression at all points in the titration curve. We observed a significant reduction in the cellular growth rate at maximal expression of the ZifH-SSN6 construct (data not shown). This observation generated the potential for non-independent action

of ZifH-SSN6 on other components and we therefore did not include ZifH-SSN6 in downstream modeling. The ZifH-VP64 activator also exhibited a monotonic response to the first three additional ZifH binding sites, saturating at a 61% increase in expression at maximal ZifH-VP64 expression in the Z5LX genotypes (Fig. 3.8.C). Unlike the ZifH-SSN6 construct, a reliable effect of ZifH-VP64 on promoter output was not observed until maximal ZifH-VP64 expression.

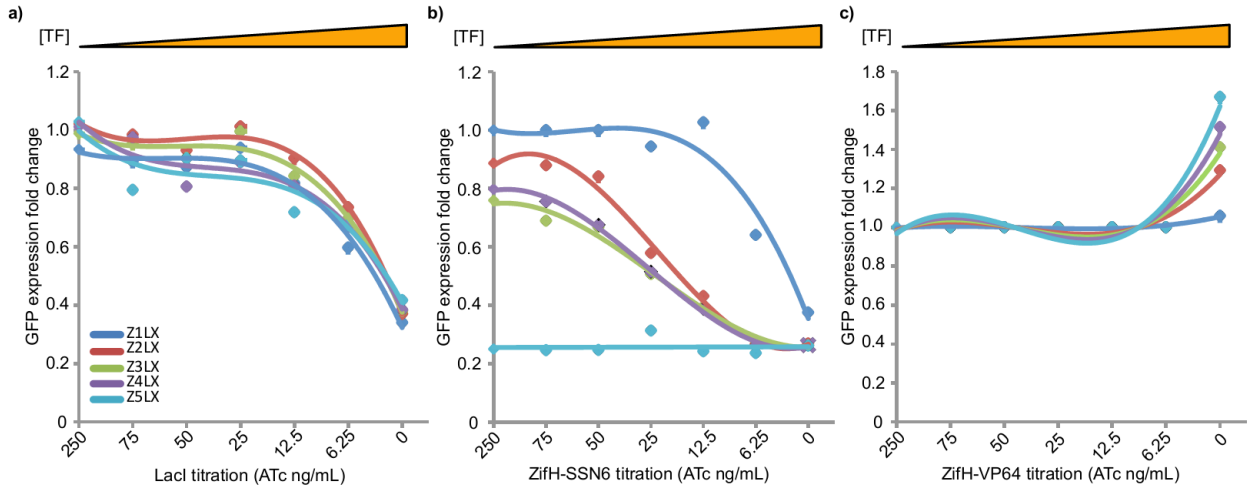

Figure 3.8. Fitted transcription factor response functions. Each line is a polynomial or linear function of TF concentration. Panel A: LacI response function. Panel B: ZifH-SSN6 response function. Panel C: ZifH-VP64 response function.

3.3.4 The LacI response function exhibits *cis*-independence:

Following TF titration, we evaluated the system for the presence of *cis* and *trans* independence using the analytical strategy outlined in Figure 3.9. For LacI *cis* independence, changes in the ZifH CRM had no impact on the LacI response function. This enables the characterization of the LacI response function with one ZifH CRM to predict the response with other ZifH CRMs (Fig. 3.9.A). In the case of *trans* independence, the simultaneous effect of LacI and ZifH on promoter output can be predicted from response functions of the two components characterized in isolation (Fig. 3.9.B and Fig. 3.11B). The presence of *cis* and *trans* independence enables further

reduction in measurements by using the *cis* independence of LacI to characterize a response function on one promoter genotype and predict the LacI response within the other ZiLX library members (Fig. 3.9.C and Fig. 3.11.C).

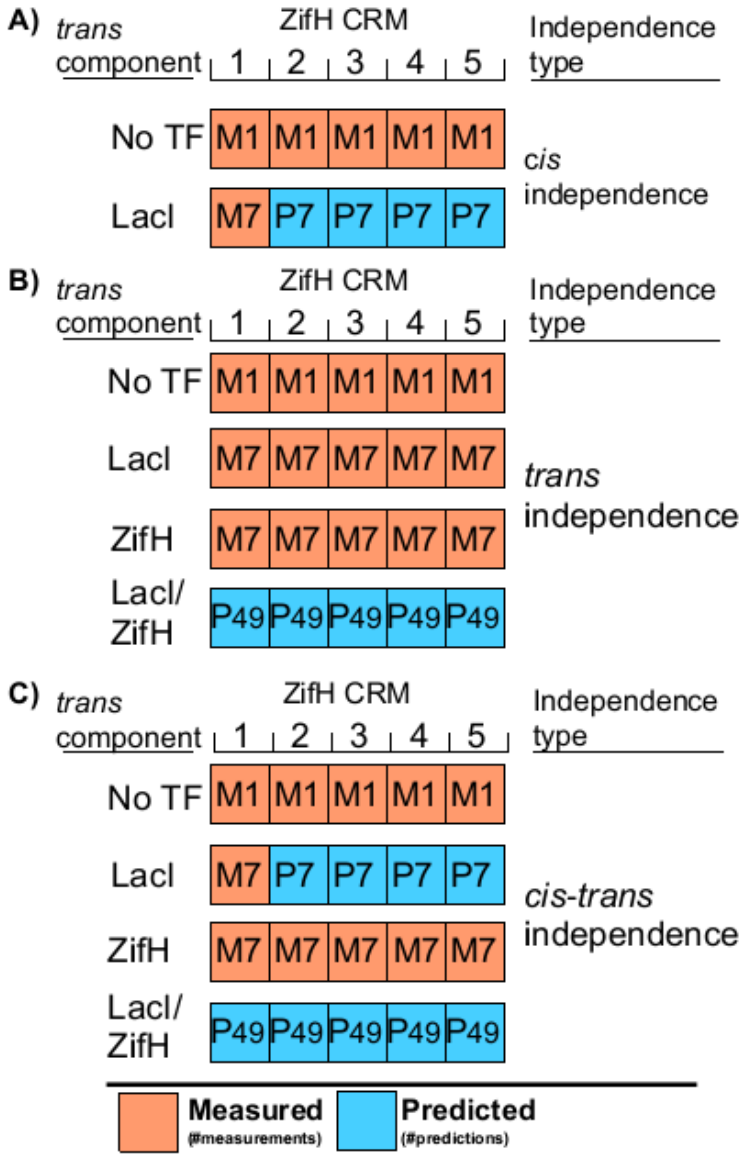

Figure 3.9. Analytical strategy for evaluation of *cis* and *trans* independence within the ZiLX system. Each element in the No-TF column represents a measurement of promoter expression in the absence of *trans*-inputs. Elements of the LacI and ZifH vectors represent a response function. Elements in the LacI/ZifH category represent a 2D matrix of LacI and ZifH concentrations (Fig. 3.11.A-C). Orange boxes are measured and blue boxes predicted. Panel A: *cis* independence of LacI response from changes in the ZifH CRM. The LacI response function does not change

significantly across the ZiLX library, enabling a response function trained using one promoter genotype to predict the response of the other four genotypes to LacI. Panel B: *trans* independence of the ZifH and LacI inputs. Independence between the ZifH-VP64 and LacI enables response functions for ZifH and LacI trained in isolation to predict the simultaneous titration of the two inputs. Panel C: Use of *cis* and *trans* independence to further decrease characterization requirements. The *cis* independence of LacI enables a simultaneous titration of LacI and ZifH to be predicted without characterizing a LacI response function for each ZifH CRM.

We hypothesized that the *cis* independence of LacI from genotype could be used to minimize the number of samples required to generate an accurate model of LacI response. We utilized a five-fold cross-validation approach, training a LacI response function on data from one promoter genotype (i.e. one ZifH CRM) and predicting the other four (Fig. 3.9.A). The average variance explained across the cross-validated models of each promoter's LacI response was 95%, indicating that response function trained on information from one member of the library were highly predictive of the LacI response across the others (Fig. 3.10). As a control, we attempted a similar cross-validation strategy on the ZifH-SSN6 response function, a function expected to be genotype-dependent. The genotype-independent ZifH-SSN6 models on average recovered only 51% of total variance and presented with a highly variable degree of predictive power (standard deviation of variance explained = 26%). Our independence result for LacI represents a 70% reduction in the number of measurements required for characterizing the LacI response function.

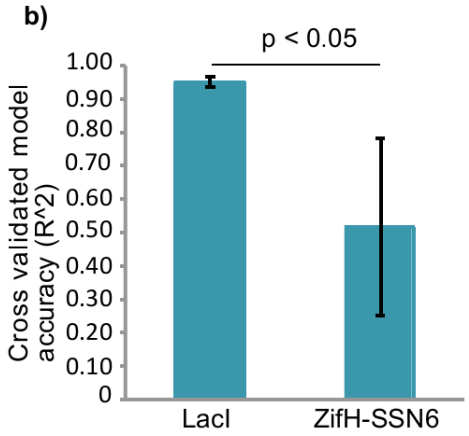

Figure 3.10. LacI *cis* independence enables LacI response functions trained on one genotype to be highly predictive of the LacI effect on the other ZiLX genotypes.

3.3.5 *trans*-independence between LacI and ZifH enables accurate prediction of ZiLX library gene expression

We assessed the ability of a composite model predicated on *trans*-independence between LacI and ZifH to predict gene expression of a novel genotype in which both TFs were simultaneously titrated across a range of concentrations. In this evaluation, the response functions generated from the individual titrations of LacI and ZifH are used to predict the expression from the ZiLX library during the simultaneous titration of LacI and ZifH across a range of concentrations (Fig. 3.9.B, Fig. 3.11.B). In order to test the performance of each component function, we generated expression models that were trained only on basal expression (Fig. 3.12.A), basal expression and LacI response (Fig. 3.12.B) and basal expression, LacI response and ZifH-VP64 response (Fig. 3.12.C). This process revealed a successive improvement in predictive power as each function was added to the composite model, ranging from 52% of variance explained for the basal-only trained model, to 72.8% of total variance explained with the composite model trained on all response functions. The variance explained by the full composite model is consistent with the most accurate thermodynamic models of promoter libraries currently available (Mogno, Vallania, Mitra, & Cohen, 2010). The use of *trans* independence reduced the total number of measurements required for expression modeling by 70% compared to an architecture that requires characterization of TF:TF interactions (Fig. 3.11.A vs. Fig. 3.11.B)

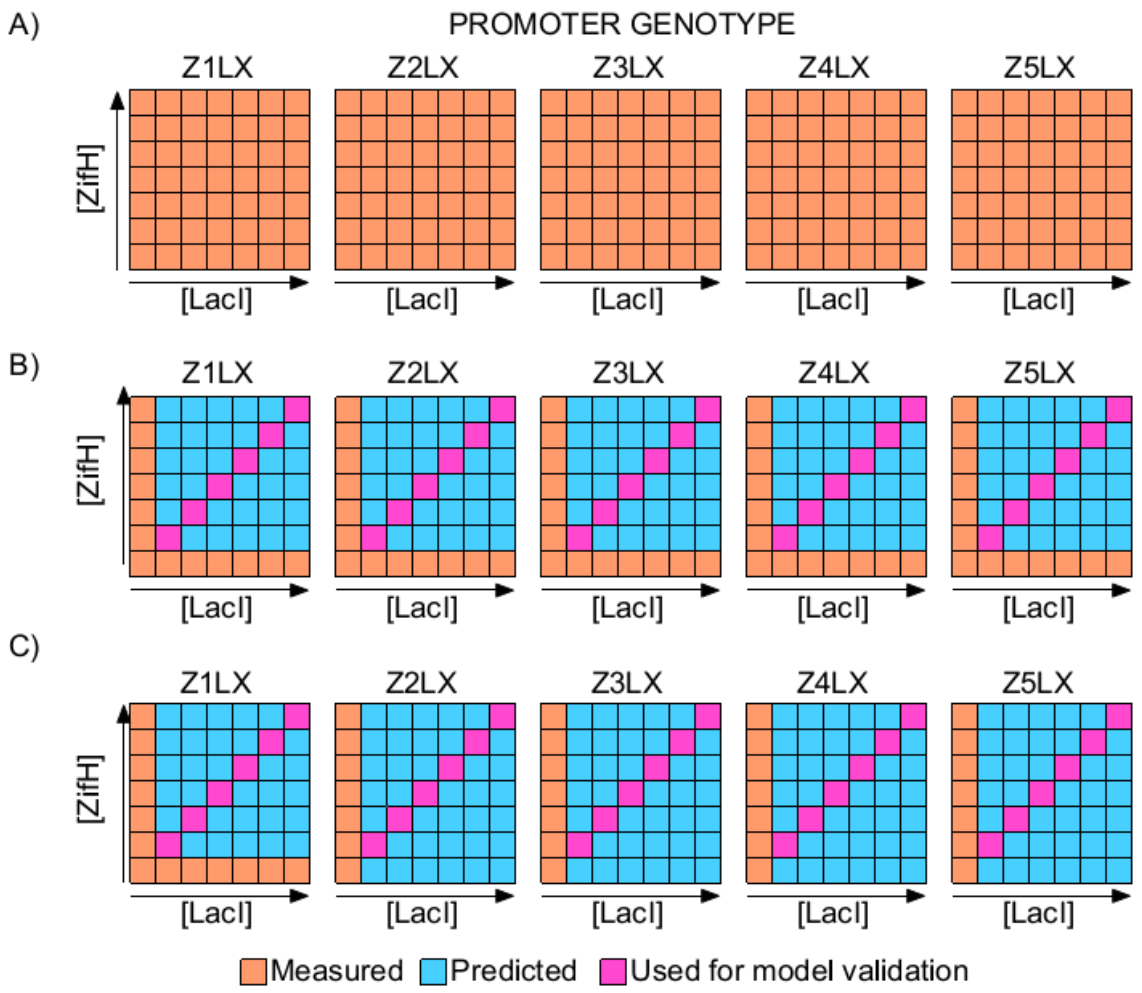

Figure 3.11. *trans* independence enables response functions generated from LacI and ZifH titrated alone to predict the effect of simultaneous ZifH and LacI expression. Panel A: Promoters which exhibit non-independence between LacI and ZifH requires measurement of TFs expressed across many different concentrations for model training. Panel B: *trans* independence between ZifH and LacI enables response functions characterized from the two factors expressed in isolation to predict expression of both. Panel C: *cis* and *trans* independence enables a further reduction in measurements by training a response function for LacI in one genotype and predicting the LacI effect in the other genotypes. (Orange boxes: Measurements used to characterize response function. Blue boxes: Predictions generated from response functions. Purple boxes: Predictions tested with ZifH/LacI simultaneous expression.)

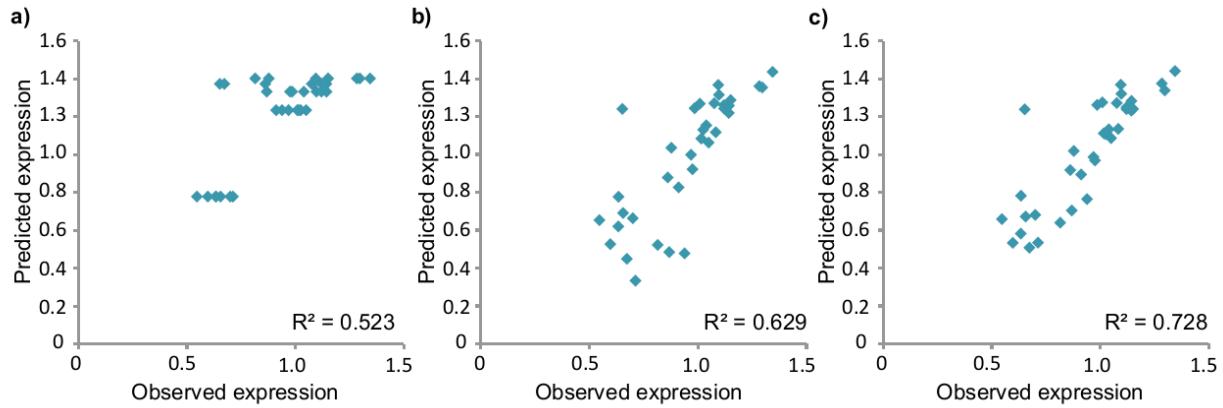

Figure 3.12. Composite models of ZiLX library expression within a novel genotype expressing ZifH and LacI across a range of different concentrations. Panel A: Expression predictions generated from a model trained only on genotype basal expression. Panel B: Expression predictions generated from a composite model using basal expression and LacI response functions. Panel C: Expression predictions generated from a composite model trained on basal expression, LacI and ZifH response functions.

3.3.6 Expression models built using *cis* and *trans* independence require fewer measurements

We evaluated the ability of models built using *cis* and *trans* independence to predict the behavior of the ZifH/LacI double titration. We again utilized a five-fold cross validation approach, training the LacI response function from one genotype and testing model predictions on the other ZiLX library members (Fig. 3.9.C, Fig. 3.11.C). Mean variance explained within the cross-validated composite models was 64.5%, exhibiting 88.5% of the accuracy observed in a composite model trained on all titration LacI and ZifH titration data (Fig. 3.13). This represents an 11.5% reduction in accuracy for an additional 37.3% reduction in measurements required for expression prediction over the composite model utilizing only *trans* independence and 81.2% fewer measurements required vs a non-independent architecture.

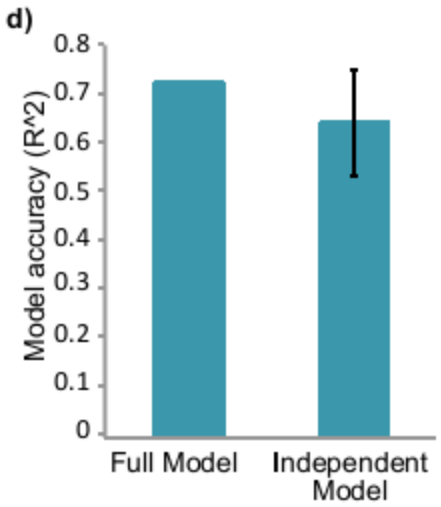

Figure 3.13. Comparison of the total variance explained by the *trans* independence composite model and cross validated models built using *cis* and *trans* independence.

3.4 Discussion

We have attempted to functionally decompose a synthetic promoter into individual functions, characterize each function, and predict the expression of the full system. This approach, predicated upon *trans* independence between the basal, LacI and ZifH modules, successfully explained 72.8% of total variance in observed gene expression. Significantly, we found that an independent promoter architecture allows a composite model fitted only to examples of components titrated in isolation to predict the behavior of the complete system. This is a departure from other modeling strategies which are commonly fitted to a subset of the complete system and used to predict the behavior of a hold-out test data set. By avoiding the need to characterize interactions between components, we were able to significantly reduce the sample space required to train a predictive model. The independent promoter architecture also enabled a simple mathematical formalism to accurately predict gene expression, potentially expanding the scope of quantitative gene expression modeling beyond specialist laboratories.

The small number of genotypes in our promoter library limits our ability to predict the generalizability of the observations above. Due to the technical limitation of the single titration system used in this study, LacI and ZifH were expressed at equivalent levels during the combined titration. It is possible that a combined titration in which the two components are expressed at non-equal levels may generate interactions between the two factors, requiring characterization of these interactions. We selected LacI and LacI-VP64 for the combined titration as the ZifH-SSN6 construct exhibited growth defects during maximal expression. The *SSN6* growth defect is a phenotype potentially indicative of sequestration of transcriptional components by *SSN6*, a process that would generate non-independent effects during a combined titration. Although the use of the ZifH-VP64 response function in the composite model significantly improved predictive accuracy, we note that ZifH-VP64 only induces a significant change on promoter output within a limited range of the ATc concentrations used within our analysis. We observed clear positional effects of the number of ZifH sites within the promoter. These effects lead to a reduction in the basal expression from the Z5LX genotype relative to the Z1LX genotype. Significantly, changes in basal activity did not substantially alter the LacI response function and we found the effect of additional binding sites for ZifH-VP64 and ZifH-SSN6 to monotonically increase the effect of each transcription factor on promoter output.

We have also demonstrated that accurate characterization of the LacI response function can be achieved through titration of LacI against the LX core module in one promoter genotype. A cross-validation strategy was applied to successfully predict the effect of LacI in isolation and during a combined titration with ZifH. This observation promotes a model in which the genotype of the 5' ZifH module and the functional effect of LacI are independent.

This work represents a step towards a future in which independence can be utilized to generate modular libraries of *cis* and *trans* components that can be characterized in isolation and combined to produce complex biological systems. We have taken the initial steps towards this future by characterizing the interaction between the ZifH CRM, the LX core and two *trans* inputs. Future research that expands on this work to include variation in the core module of the promoter could allow modification of promoter basal expression and identify the expression regimes in which independence between CRMs exists.

We attempted to extend same modular approach that we took with the CRMs into the *trans* components of the transcriptional machinery. We demonstrated with ZifH-VP64/*SSN6* domain swaps that the activity of a DNA binding domain and linked CRM can be modulated through changes of the TF regulatory domain. Although we could not expand on this observation due to the cell-growth defect generated by high levels of *SSN6* expression, regulatory domain swaps offer a promising approach to maximize the modularity and utility of transcriptional regulatory components.

3.5 Materials and Methods

3.5.1 Titration strain construction and synthetic promoter assembly

DNA for the ZifH DNA binding domain and the VP64 transactivator was acquired from the Ajo-Franklin lab (Ajo-Franklin, et al., 2007). The SSN6 ORF as described in Bellí et al. was provided by the Herrero lab (Bellí, Piedrafita, Aldea, & Herrero, 1998). DNA for the yeast codon optimized LacI repressor from Ellis et al. was provided by the Ellis lab (Ellis, Wang, & Collins, 2009). The ZifH DNA binding domain was fused in frame to VP64 and SSN6 ORFs via overlap extension PCR and cloned into the pMB008 expression plasmid downstream of the tet-O2. LacI was PCR amplified and cloned into the pMB008 expression plasmid. Following

sequence confirmation, the resulting transcription factor expression plasmids were digested with PvuII, gel purified and integrated into the *TRP1* locus of *S. cerevisiae* BY4741 using the lithium acetate/single-stranded carrier DNA method (Geitz & Schiestl, 2007). All integrated constructs were sequence confirmed.

DNA for the 1-5 ZifH and LX modules ZiLX library was synthesized by Blue Heron (Seattle, WA). Individual modules were PCR amplified with overlapping primers, gel isolated and the purified modules combined into full length promoters via overlap extension PCR. Following PCR assembly, the modules were gel purified, BamH1/Pst1 digested and cloned into a BamH1/Pst1 restricted pLVGI plasmid (Ellis, Wang, & Collins, 2009) via ligation and transformed into chemically competent STBL2 *E. coli* (Life Technologies, Carlsbad, CA). Following sequence confirmation, the synthetic promoters were amplified by PCR, fused to the hphMX drug resistance cassette via overlap extension PCR and gel purified. Purified promoter-hphMX constructs were integrated into the genome of the transcription factor titration strains at the *URA3* locus using the lithium acetate/single-stranded carrier DNA method. All integrated promoters were sequence confirmed.

3.5.2 Transcription factor titrations and flow cytometry

Titration strains containing the integrated transcription factor (s) and synthetic promoters were activated from frozen stocks and grown on YPD + Hygromycin B for three days at 30°C. Individual colonies were selected and grown overnight in YPD. Aliquots of the overnight cultures were transferred into 1 mL of synthetic complete medium supplemented with 2% galactose and ATc and grown for 22 hours within deep well 96 well plates. GFP quantitation was performed on a Beckman Coulter Cell Lab Quanta SC cytometer and 10,000 events captured per well. Fluorescence measurements were gated by forward and side scatter to remove doublets

and outliers prior to normalization. We observed a linear decay in fluorescent signal due to removal of the cells from galactose induction during flow cytometry. A linear model was fit to internal plate controls at the start and end of each 96 well plate to remove this effect. Across plate variation in fluorescence was normalized by dividing the mean fluorescence measurement of each well by the mean of plate normalization controls, providing a normalized mean expression value for each well.

3.5.3 Expression modeling

Expression was modeled according to the equation: $Exp_{ik} = \alpha_i \prod_{j=1}^{j=2} \theta_{ijk}$. Where expression of genotype *i* in ATc concentration, *k,* is equal to the basal expression of genotype *i* (α), multiplied by the response function (θ) return of transcription factor *j,* on genotype *i,* in ATc concentration *k.* A polynomial response function (θ) for each transcription factor was fit to each member of the library by titrating each transcription factor input in isolation against the library member. To avoid over fitting, transcription factor response functions were constrained to adhere to the molecular mechanism of each transcription factor and the effect of ZifH-VP64 was thresholded to the final point in the titration series.

3.5.4 Expression analysis of LacI titration on the Z5LX promoter

The LacI titration system and Z5LX promoter were integrated into BY4741 as described in section 3.5.1. Titration of LacI onto the Z5LX promoter was performed as described in section 3.5.2, with the exception of the experiment being run in 3 mL culture volumes and not 96 well plates. Following titration, cultures were centrifuged at 10,000 x G for 1 minute, and pellets snap frozen in liquid nitrogen and stored at -80°C. mRNA expression levels for LacI and GFP were quantitated from cell lysates using the Quantigene Plex 2.0 platform (Affymetrix, Santa

Clara, CA) with the manufacturer's recommended protocol. Background subtracted expression values were normalized to the geometric mean of *ACT1*, *PDA1* and *UBC1*.

3.5.5 Transcription factor percentile expression and fold change analysis

Normalized RNA expression values for *S. cerevisiae* CEN.PK2-1D grown for 12 hours in 5% glucose, 5% galactose, 5% glycerol, 1.3% ethanol, 5% xylose were used from Chapter 2 of this dissertation. The expression percentile of each transciption factor in the *S. cerevisiae* genome across each environmental condition was calculated and plotted for all environmental conditions. The maximal fold change for transcription factors and non-transcription factors across all environmental conditions was calculated as max(median expression) / min(median expression) for each gene. All analysis and data visualization was performed using Python.

3.6 Supplemental Data

3.6.1 Transcription Factor Titration Data

Genotype	Titration	[ATc/Dox]	Plate	Index	Mean	StDev
Z1LX	LacI	100 ng/mL Dox	AA1046	A2	94.8	31.6
Z1LX	LacI	250 ng/mL ATC	AA1046	B2	95.8	31.6
Z1LX	LacI	75 ng/mL ATC	AA1046	C2	90.4	29.9
Z1LX	LacI	50 ng/mL ATC	AA1046	D2	87.3	28.6
Z1LX	LacI	25 ng/mL ATC	AA1046	E2	93.2	30.6
Z1LX	LacI	12.5 ng/mL ATC	AA1046	F2	79.7	25.8
Z1LX	LacI	6.25 ng/mL ATC	AA1046	G2	55.6	19.6
Z1LX	LacI	0 ng/mL	AA1046	H2	27.8	9.26
Z2LX	LacI	100 ng/mL Dox	AA1046	A3	105	35.5
Z2LX	LacI	250 ng/mL ATC	AA1046	B3	107	37.2
Z2LX	LacI	75 ng/mL	AA1046	C3	102	33.2

		ATC				
Z2LX	LacI	50 ng/mL ATC	AA1046	D3	95.3	31
Z2LX	LacI	25 ng/mL ATC	AA1046	E3	103	33.9
Z2LX	LacI	12.5 ng/mL ATC	AA1046	F3	90.4	29.5
Z2LX	LacI	6.25 ng/mL ATC	AA1046	G3	71.5	25.2
Z2LX	LacI	0 ng/mL	AA1046	H3	31.7	10.5
Z3LX	LacI	100 ng/mL Dox	AA1046	A4	99.6	35.3
Z3LX	LacI	250 ng/mL ATC	AA1046	B4	98.2	33.2
Z3LX	LacI	75 ng/mL ATC	AA1046	C4	93.6	31.2
Z3LX	LacI	50 ng/mL ATC	AA1046	D4	87.6	31.6
Z3LX	LacI	25 ng/mL ATC	AA1046	E4	95.8	32.8
Z3LX	LacI	12.5 ng/mL ATC	AA1046	F4	79.4	28.2
Z3LX	LacI	6.25 ng/mL ATC	AA1046	G4	63.6	23.9
Z3LX	LacI	0 ng/mL	AA1046	H4	30.6	10.4
Z4LX	LacI	100 ng/mL Dox	AA1046	A5	96.4	33.9
Z4LX	LacI	250 ng/mL ATC	AA1046	B5	92.8	33.5
Z4LX	LacI	75 ng/mL ATC	AA1046	C5	88.3	29.6
Z4LX	LacI	50 ng/mL ATC	AA1046	D5	71.6	27
Z4LX	LacI	25 ng/mL ATC	AA1046	E5	78.4	28.2
Z4LX	LacI	12.5 ng/mL ATC	AA1046	F5	69.9	25.4
Z4LX	LacI	6.25 ng/mL ATC	AA1046	G5	55.9	21.3
Z4LX	LacI	0 ng/mL	AA1046	H5	28	9.71
Z5LX	LacI	100 ng/mL Dox	AA1046	A6	50.4	20.4
Z5LX	LacI	250 ng/mL ATC	AA1046	B6	58.7	20.8
Z5LX	LacI	75 ng/mL	AA1046	C6	43.9	18.2

		ATC				
Z5LX	LacI	50 ng/mL ATC	AA1046	D6	49.2	18.6
Z5LX	LacI	25 ng/mL ATC	AA1046	E6	47.8	19.5
Z5LX	LacI	12.5 ng/mL ATC	AA1046	F6	36.4	15.7
Z5LX	LacI	6.25 ng/mL ATC	AA1046	G6	32.7	14.2
Z5LX	LacI	0 ng/mL	AA1046	H6	16.5	6.58
Z1LX	ZifH-VP64	100 ng/mL Dox	AA1046	A7	113	37.9
Z1LX	ZifH-VP64	250 ng/mL ATC	AA1046	B7	109	33.8
Z1LX	ZifH-VP64	75 ng/mL ATC	AA1046	C7	94.7	32.3
Z1LX	ZifH-VP64	50 ng/mL ATC	AA1046	D7	106	37.8
Z1LX	ZifH-VP64	25 ng/mL ATC	AA1046	E7	98.4	32.8
Z1LX	ZifH-VP64	12.5 ng/mL ATC	AA1046	F7	71.5	36.9
Z1LX	ZifH-VP64	6.25 ng/mL ATC	AA1046	G7	90.9	31.3
Z1LX	ZifH-VP64	0 ng/mL	AA1046	H7	102	33
Z2LX	ZifH-VP64	100 ng/mL Dox	AA1046	A8	112	36.7
Z2LX	ZifH-VP64	250 ng/mL ATC	AA1046	B8	102	37.6
Z2LX	ZifH-VP64	75 ng/mL ATC	AA1046	C8	108	35.2
Z2LX	ZifH-VP64	50 ng/mL ATC	AA1046	D8	114	38.2
Z2LX	ZifH-VP64	25 ng/mL ATC	AA1046	E8	108	35.9
Z2LX	ZifH-VP64	12.5 ng/mL ATC	AA1046	F8	105	35.3
Z2LX	ZifH-VP64	6.25 ng/mL ATC	AA1046	G8	104	35.4
Z2LX	ZifH-VP64	0 ng/mL	AA1046	H8	129	42.5
Z3LX	ZifH-VP64	100 ng/mL Dox	AA1046	A9	99.4	32.6
Z3LX	ZifH-VP64	250 ng/mL ATC	AA1046	B9	100	31.2
Z3LX	ZifH-VP64	75 ng/mL	AA1046	C9	71.6	30

		ATC				
Z3LX	ZifH-VP64	50 ng/mL ATC	AA1046	D9	69	28.9
Z3LX	ZifH-VP64	25 ng/mL ATC	AA1046	E9	51.8	34.5
Z3LX	ZifH-VP64	12.5 ng/mL ATC	AA1046	F9	89.8	28.9
Z3LX	ZifH-VP64	6.25 ng/mL ATC	AA1046	G9	91.2	29.3
Z3LX	ZifH-VP64	0 ng/mL	AA1046	H9	134	43.2
Z4LX	ZifH-VP64	100 ng/mL Dox	AA1046	A10	100	33.1
Z4LX	ZifH-VP64	250 ng/mL ATC	AA1046	B10	90.2	31.6
Z4LX	ZifH-VP64	75 ng/mL ATC	AA1046	C10	81.5	29.5
Z4LX	ZifH-VP64	50 ng/mL ATC	AA1046	D10	90.2	32.1
Z4LX	ZifH-VP64	25 ng/mL ATC	AA1046	E10	96.7	38.2
Z4LX	ZifH-VP64	12.5 ng/mL ATC	AA1046	F10	89.5	30.1
Z4LX	ZifH-VP64	6.25 ng/mL ATC	AA1046	G10	88.9	30.3
Z4LX	ZifH-VP64	0 ng/mL	AA1046	H10	133	44.7
Z5LX	ZifH-VP64	100 ng/mL Dox	AA1046	A11	64.2	24.4
Z5LX	ZifH-VP64	250 ng/mL ATC	AA1046	B11	54.8	22.8
Z5LX	ZifH-VP64	75 ng/mL ATC	AA1046	C11	52.5	20.6
Z5LX	ZifH-VP64	50 ng/mL ATC	AA1046	D11	50.9	23
Z5LX	ZifH-VP64	25 ng/mL ATC	AA1046	E11	52.2	21.9
Z5LX	ZifH-VP64	12.5 ng/mL ATC	AA1046	F11	51.1	21.1
Z5LX	ZifH-VP64	6.25 ng/mL ATC	AA1046	G11	49.4	21.4
Z5LX	ZifH-VP64	0 ng/mL	AA1046	H11	89.5	36.5
E1	Plate Cntl	NA	AA1046	A1	62.5	40.5
E1	Plate Cntl	NA	AA1046	B1	88.1	30.2
E1	Plate Cntl	NA	AA1046	C1	77.3	28.7
E1	Plate Cntl	NA	AA1046	F12	71.3	26
E1	Plate Cntl	NA	AA1046	G12	62.9	24.6

E1	Plate Cntl	NA	AA1046	H12	68.9	25.3
Sample	Titration	[ATc/Dox]	Plate	Index	Mean	StDev
Z1LX	ZifH-SSN6	100 ng/mL Dox	AA1047	A2	95.9	34.8
Z1LX	ZifH-SSN6	250 ng/mL ATC	AA1047	B2	97.4	31.7
Z1LX	ZifH-SSN6	75 ng/mL ATC	AA1047	C2	94.1	32.6
Z1LX	ZifH-SSN6	50 ng/mL ATC	AA1047	D2	94.7	32.8
Z1LX	ZifH-SSN6	25 ng/mL ATC	AA1047	E2	67.8	24.7
Z1LX	ZifH-SSN6	12.5 ng/mL ATC	AA1047	F2	72	25.7
Z1LX	ZifH-SSN6	6.25 ng/mL ATC	AA1047	G2	36.4	18.8
Z1LX	ZifH-SSN6	0 ng/mL	AA1047	H2	11	15.7
Z2LX	ZifH-SSN6	100 ng/mL Dox	AA1047	A3	58	30.4
Z2LX	ZifH-SSN6	250 ng/mL ATC	AA1047	B3	71.2	34.4
Z2LX	ZifH-SSN6	75 ng/mL ATC	AA1047	C3	67.6	32.1
Z2LX	ZifH-SSN6	50 ng/mL ATC	AA1047	D3	61.6	31.5
Z2LX	ZifH-SSN6	25 ng/mL ATC	AA1047	E3	36.7	23.1
Z2LX	ZifH-SSN6	12.5 ng/mL ATC	AA1047	F3	21.5	17.4
Z2LX	ZifH-SSN6	6.25 ng/mL ATC	AA1047	G3	4.6	7.83
Z2LX	ZifH-SSN6	0 ng/mL	AA1047	H3	1.77	8.05
Z3LX	ZifH-SSN6	100 ng/mL Dox	AA1047	A4	47.8	30.8
Z3LX	ZifH-SSN6	250 ng/mL ATC	AA1047	B4	60.2	33.6
Z3LX	ZifH-SSN6	75 ng/mL ATC	AA1047	C4	51.3	31.7
Z3LX	ZifH-SSN6	50 ng/mL ATC	AA1047	D4	46.7	30.8
Z3LX	ZifH-SSN6	25 ng/mL ATC	AA1047	E4	30.4	24.3
Z3LX	ZifH-SSN6	12.5 ng/mL ATC	AA1047	F4	18.1	17.9
Z3LX	ZifH-SSN6	6.25 ng/mL	AA1047	G4	5.33	10.2

		ATC				
Z3LX	ZifH-SSN6	0 ng/mL	AA1047	H4	0.906	5.05
Z4LX	ZifH-SSN6	100 ng/mL Dox	AA1047	A5	43.9	28.7
Z4LX	ZifH-SSN6	250 ng/mL ATC	AA1047	B5	63.2	30.4
Z4LX	ZifH-SSN6	75 ng/mL ATC	AA1047	C5	56.6	29.1
Z4LX	ZifH-SSN6	50 ng/mL ATC	AA1047	D5	47.1	29
Z4LX	ZifH-SSN6	25 ng/mL ATC	AA1047	E5	30.8	21.3
Z4LX	ZifH-SSN6	12.5 ng/mL ATC	AA1047	F5	17.3	16.1
Z4LX	ZifH-SSN6	6.25 ng/mL ATC	AA1047	G5	3.56	7.62
Z4LX	ZifH-SSN6	0 ng/mL	AA1047	H5	0.712	2.81
Z5LX	ZifH-SSN6	100 ng/mL Dox	AA1047	A6	8.9	8.85
Z5LX	ZifH-SSN6	250 ng/mL ATC	AA1047	B6	16.7	13
Z5LX	ZifH-SSN6	75 ng/mL ATC	AA1047	C6	13.3	11
Z5LX	ZifH-SSN6	50 ng/mL ATC	AA1047	D6	10.6	10
Z5LX	ZifH-SSN6	25 ng/mL ATC	AA1047	E6	13.5	11.9
Z5LX	ZifH-SSN6	12.5 ng/mL ATC	AA1047	F6	4.72	5.52
Z5LX	ZifH-SSN6	6.25 ng/mL ATC	AA1047	G6	1.2	2.52
Z5LX	ZifH-SSN6	0 ng/mL	AA1047	H6	0.289	0.345
Z1LX	LacI/ZifH	100 ng/mL Dox	AA1047	A7	86.6	29.6
Z1LX	LacI/ZifH	250 ng/mL ATC	AA1047	B7	86.7	28.2
Z1LX	LacI/ZifH	75 ng/mL ATC	AA1047	C7	47.9	24.4
Z1LX	LacI/ZifH	50 ng/mL ATC	AA1047	D7	84.7	27.7
Z1LX	LacI/ZifH	25 ng/mL ATC	AA1047	E7	83.7	27.7
Z1LX	LacI/ZifH	12.5 ng/mL ATC	AA1047	F7	75.9	25.4
Z1LX	LacI/ZifH	6.25 ng/mL	AA1047	G7	54.2	19.9

		ATC				
Z1LX	LacI/ZifH	0 ng/mL	AA1047	H7	35.3	12.1
Z2LX	LacI/ZifH	100 ng/mL Dox	AA1047	A8	96.5	33.7
Z2LX	LacI/ZifH	250 ng/mL ATC	AA1047	B8	109	35.5
Z2LX	LacI/ZifH	75 ng/mL ATC	AA1047	C8	102	33.7
Z2LX	LacI/ZifH	50 ng/mL ATC	AA1047	D8	98.1	33.5
Z2LX	LacI/ZifH	25 ng/mL ATC	AA1047	E8	79.4	27.1
Z2LX	LacI/ZifH	12.5 ng/mL ATC	AA1047	F8	81.4	27.5
Z2LX	LacI/ZifH	6.25 ng/mL ATC	AA1047	G8	55.3	20.5
Z2LX	LacI/ZifH	0 ng/mL	AA1047	H8	47	16.1
Z3LX	LacI/ZifH	100 ng/mL Dox	AA1047	A9	88.9	30.8
Z3LX	LacI/ZifH	250 ng/mL ATC	AA1047	B9	87.9	28.8
Z3LX	LacI/ZifH	75 ng/mL ATC	AA1047	C9	86.8	29.1
Z3LX	LacI/ZifH	50 ng/mL ATC	AA1047	D9	86	29.6
Z3LX	LacI/ZifH	25 ng/mL ATC	AA1047	E9	69.9	30.2
Z3LX	LacI/ZifH	12.5 ng/mL ATC	AA1047	F9	71.7	23.6
Z3LX	LacI/ZifH	6.25 ng/mL ATC	AA1047	G9	63.3	22.1
Z3LX	LacI/ZifH	0 ng/mL	AA1047	H9	51.2	18.2
Z4LX	LacI/ZifH	100 ng/mL Dox	AA1047	A10	82	30.3
Z4LX	LacI/ZifH	250 ng/mL ATC	AA1047	B10	80.2	28.5
Z4LX	LacI/ZifH	75 ng/mL ATC	AA1047	C10	78.5	28.7
Z4LX	LacI/ZifH	50 ng/mL ATC	AA1047	D10	75	27.4
Z4LX	LacI/ZifH	25 ng/mL ATC	AA1047	E10	75.1	26.4
Z4LX	LacI/ZifH	12.5 ng/mL ATC	AA1047	F10	65.5	23.2
Z4LX	LacI/ZifH	6.25 ng/mL	AA1047	G10	57.6	21.4

		ATC				
Z4LX	LacI/ZifH	0 ng/mL	AA1047	H10	56.9	19.7
Z5LX	LacI/ZifH	100 ng/mL Dox	AA1047	A11	46.2	22.5
Z5LX	LacI/ZifH	250 ng/mL ATC	AA1047	B11	48.3	20.6
Z5LX	LacI/ZifH	75 ng/mL ATC	AA1047	C11	47.1	18.7
Z5LX	LacI/ZifH	50 ng/mL ATC	AA1047	D11	48.1	18.5
Z5LX	LacI/ZifH	25 ng/mL ATC	AA1047	E11	32.4	14.1
Z5LX	LacI/ZifH	12.5 ng/mL ATC	AA1047	F11	36.9	16.5
Z5LX	LacI/ZifH	6.25 ng/mL ATC	AA1047	G11	30.8	15
Z5LX	LacI/ZifH	0 ng/mL	AA1047	H11	37.5	14.8
E1	Plate Cntl	0 ng/mL	AA1047	A1	80.9	31.9
E1	Plate Cntl	0 ng/mL	AA1047	B1	81	30.6
E1	Plate Cntl	0 ng/mL	AA1047	C1	84.7	30
E1	Plate Cntl	0 ng/mL	AA1047	F12	62.9	26
E1	Plate Cntl	0 ng/mL	AA1047	G12	64.4	25.2
E1	Plate Cntl	0 ng/mL	AA1047	H12	61.9	24.3
Sample	Titration	[ATc/Dox]	Plate	Index	Mean	StDev
E1	Plate Cntl	0 ng/mL	AA1048	A1	68.6	28.3
Z1LX	No-TF	0 ng/mL	AA1048	A2	78	30
E1	Plate Cntl	0 ng/mL	AA1048	B1	65.6	28
Z1LX	No-TF	0 ng/mL	AA1048	B2	73.2	29.5
E1	Plate Cntl	0 ng/mL	AA1048	C1	58.1	24.6
Z1LX	No-TF	0 ng/mL	AA1048	C2	65.8	26.6
Z1LX	No-TF	0 ng/mL	AA1048	D2	65	25.3
Z1LX	No-TF	0 ng/mL	AA1048	E2	59.2	24.5
Z1LX	No-TF	0 ng/mL	AA1048	F2	57.5	23
Z1LX	Plate Cntl	0 ng/mL	AA1048	F12	54.9	22.6
Z1LX	No-TF	0 ng/mL	AA1048	G2	56.1	23
Z1LX	Plate Cntl	0 ng/mL	AA1048	G12	52.8	22.1
Z1LX	No-TF	0 ng/mL	AA1048	H2	53.9	22.3
Z1LX	Plate Cntl	0 ng/mL	AA1048	H12	52.9	21.8
Z2LX	No-TF	0 ng/mL	AA1048	A3	94.6	34.8
Z2LX	No-TF	0 ng/mL	AA1048	B3	82.1	31.9
Z2LX	No-TF	0 ng/mL	AA1048	C3	81.3	30
Z2LX	No-TF	0 ng/mL	AA1048	D3	82.4	30.1
Z2LX	No-TF	0 ng/mL	AA1048	E3	85.8	30.3
Z2LX	No-TF	0 ng/mL	AA1048	F3	73.4	27.5
Z2LX	No-TF	0 ng/mL	AA1048	G3	73.2	26.5

Z2LX	No-TF	0 ng/mL	AA1048	H3	73.6	27.6
Z3LX	No-TF	0 ng/mL	AA1048	A4	92.9	35.9
Z3LX	No-TF	0 ng/mL	AA1048	B4	80.7	31.9
Z3LX	No-TF	0 ng/mL	AA1048	C4	80.9	31
Z3LX	No-TF	0 ng/mL	AA1048	D4	82.4	31
Z3LX	No-TF	0 ng/mL	AA1048	E4	80	29.9
Z3LX	No-TF	0 ng/mL	AA1048	F4	75.8	28.6
Z3LX	No-TF	0 ng/mL	AA1048	G4	76.4	28.5
Z3LX	No-TF	0 ng/mL	AA1048	H4	72.2	27.6
Z4LX	No-TF	0 ng/mL	AA1048	A5	92.5	33.6
Z4LX	No-TF	0 ng/mL	AA1048	B5	84.5	32.3
Z4LX	No-TF	0 ng/mL	AA1048	C5	84.4	30.9
Z4LX	No-TF	0 ng/mL	AA1048	D5	99.2	33.5
Z4LX	No-TF	0 ng/mL	AA1048	E5	94.6	32.3
Z4LX	No-TF	0 ng/mL	AA1048	F5	74.6	27.9
Z4LX	No-TF	0 ng/mL	AA1048	G5	68.2	26.1
Z4LX	No-TF	0 ng/mL	AA1048	H5	71.5	26.5
Z5LX	No-TF	0 ng/mL	AA1048	A6	65.1	27.4
Z5LX	No-TF	0 ng/mL	AA1048	B6	55	25
Z5LX	No-TF	0 ng/mL	AA1048	C6	55.6	23.9
Z5LX	No-TF	0 ng/mL	AA1048	D6	54	23.3
Z5LX	No-TF	0 ng/mL	AA1048	E6	53.1	22.3
Z5LX	No-TF	0 ng/mL	AA1048	F6	44.5	20
Z5LX	No-TF	0 ng/mL	AA1048	G6	42.2	19.8
Z5LX	No-TF	0 ng/mL	AA1048	H6	34.7	20.9

Table 3.2: ZiLX library transcription factor titration data.

Chapter 4: Discussion

4.1 Conclusions

This dissertation demonstrates a comprehensive approach to cellular state control at the transcriptional level. We harness the power of gene regulatory networks (GRNs) to identify regulatory interventions that broadly manipulate cellular state on a genome wide level. We then step down to a single gene level and exhibit a *cis*-engineering strategy that enables the precise expression control of a limited number of genes. We believe that both approaches are complementary in nature, enabling an investigator to globally reconfigure the cell with regulatory interventions and then optimize expression of particularly important targets via *cis* regulatory engineering. Having introduced these two mechanisms of cellular state control, we take the next step and successfully apply the genome wide approach to the industrially relevant problem of biofuel production.

In Chapter 2 we introduce NetSurgeon, an algorithmic approach to identify regulatory interventions that enforce a desired transcriptional state. We show *in silico* that NetSurgeon is capable of effectively identifying all regulatory interventions that are generating differential expression between two states, theoretically detailing the exact set of regulatory interventions required to completely reconfigure a cell towards a desired state. We then apply NetSurgeon to the complex problem of *S. cerevisiae* biofuel production. 75% of the NetSurgeon selected interventions moved transcriptional state in the desired direction, each intervention moving

hundreds of genes towards the correct expression level. These results demonstrate that complex gene regulatory networks can be rationally manipulated, enabling an investigator to broadly control cellular state. The application of this algorithmic approach to a complex problem also highlights the present challenges faced in moving beyond a single transcriptional intervention: none of the three double regulatory interventions outperformed the single deletion of CAT8 in enforcing a desired cellular state. This observation highlights the need for more research on understanding the response of the cell to multiple transcriptional interventions.

We also measured and analyzed the cellular response to transcriptional interventions. This is a rare display of integrating data across multiple cellular regulatory levels to produce a coherent picture of how interventions at one level cascade across the cell. We generated a dataset of over eight thousand metabolic measurements, painstakingly generated through low-throughput analytical chemistry by technical necessity. We used this information to contextualize the success of our regulatory interventions at the transcriptional level with the difficulty of rationally creating a desired change in a complex quantitative phenotype. We find that although our interventions promoted the desired cellular state at the transcriptional level, cellular metabolism – the phenotype we were attempting to control, still transitioned into a respiratory mode and started consuming ethanol. Taking the first steps towards a solution to this problem of phenotype control, we generated an integrated map of central carbon metabolism and transcriptional regulation linking a flux balance model with RNA expression measurements. This integrated map identified regulators that functionally regulate carbon flux through central carbon metabolism and will enable more sophisticated approaches to metabolic engineering in the future.

The difficulty in moving beyond a single regulatory intervention and enforcing control over metabolism from purely transcriptional interventions highlights the need for a precise *cis*-engineering capability. Integrative maps of cellular metabolism and transcriptional state can theoretically be used to identify specific regulators and effectors whose expression can be altered to optimize a phenotype. *Cis*-engineering of the promoters of these genes would enable an investigator to modify expression levels of a few genes while minimizing off-target impacts. Working towards this end, we demonstrate that a synthetic promoter library can be decomposed into a set of independent functions can be characterized in isolation and used to predict expression from the full system. We also observed evidence of independent action within the *trans*-elements of this synthetic promoter architecture, an important property that will reduce the number of measurements required for model training and enables a simple expression model to accurately predict system output. This work represents significant progress towards the insertion of orthogonal control circuits into the cell for the precise control of gene expression.

In summary, this dissertation represents a thorough examination of approaches to control cellular transcriptional state. We demonstrate success at genome wide transcriptome engineering and exhibit a method of *cis*-engineering that enables the precise expression control of a single gene. We also go one step further and begin the process of learning how transcriptional state controls the complex quantitative phenotype of cellular metabolism. This work is an example of an integrative process of quantitative measurement, modeling and intervention that will become increasingly common and powerful as our understanding of the molecular physiology of the cell continues to develop. Looking beyond the direct application of our work within biofuel production, we hope that the advances presented by this dissertation will contribute to the process of advancing human health and treatment of disease.

4.2 Future Directions

The rational manipulation of complex biological phenotypes is likely to require an integrated understanding of the molecular processes that generate the phenotype at all levels of cellular information flow. With the constant expansion and increasing accuracy of transcriptional regulatory and protein-protein interaction networks it is rapidly becoming feasible to form an integrated network that connects quantitative phenotypes to all layers of the central dogma and cellular signal transduction system. As these integrated networks become available, perturbation of the system within one level offers the potential for the investigator to form an integrated picture of the connections that link the regulatory layers. These perturbation studies also offer the potential to uncover the optimal regulatory layer for intervention. Once the optimal regulatory layer for intervention is selected, network connectivity maps could be weighted by information gained from the perturbation studies to select interventions to control the phenotype.

We are able to conclusively demonstrate that GRNs can be used to efficiently select the 1st regulatory intervention that moves the system towards a desired state. Our work also highlights the difficulty of selecting the correct second intervention to move the system even further towards the goal state. The problem of selecting addition interventions beyond the first is not simple to solve. We found that the removal of the first factor can be effectively modeled by looking at the network connectivity and considering the likely up and down regulation of the targets of each regulator. With the removal of the second regulator, we increase the potential of altering the *cis*-regulatory landscape, opening up previously occupied sites on promoters to competing transcription factors and altering transcriptional output. The successful prediction of multiple regulatory interventions may require the use of sophisticated models that include *cis*-combinatoric regulation to effectively predict system steady state.

We have also highlighted the difficulty in controlling a complex phenotype by transcriptional regulatory interventions. At the beginning of this dissertation work, we lacked an understanding of the connectivity between the cellular transcriptional state and metabolism. This gap in our understanding required us to hypothesize that fixing the cell into a glucose transcriptional state would force a cell consuming xylose towards a fermentative state. Now that we have generated an extensive set of linked metabolic and transcriptional measurements, we can begin to understand the how changes in transcriptional state will alter metabolic phenotypes. From this perspective, the deletion of transcription factors acted as a large scale transcriptional state perturbation study, in which we altered the system at one level and observed changes in metabolic space. When this perturbation information is combined with the flux balance modeling performed, we can start to identify functional regulators of transcriptional state that generate effective changes of metabolic behavior. The prediction that PDR3 over expression will increase glycolytic flux and direct more carbon to ethanol is an ideal starting place for further investigation.

The general improvement of xylose fermentation is likely to require the consideration of post-translational regulation. We attempted to modify cellular post-translational state in a limited fashion within by deleting *SNF1*, but other interventions that modify cellular post-translational state need to be evaluated. During the course of strain construction, we attempted to constitutively activate the glucose sensing signal transduction machinery through the deletion of *BCY1* within the *S. cerevisiae* CEN.PK2-1D background. All attempts at generating these strains failed. Similar attempts to constitutively activate the Ras pathway via *Ras2* G19V integrations failed (Fedor-Chaiken, Deschenes, & Broach, 1990). These pilot evaluations of post-translational state modification indicate that this problem is non-trivial and needs to be carefully considered.

In theory, interventions that modify cellular post-translational state synergize well with transcriptome interventions. The transcriptome interventions change the cell's interpretation of the DNA, inducing the expression of protein machinery that is then activated by interventions that alter post-translational state.

On the *cis*-engineering side of cellular transcriptional state control, we present one example of an architecture that exhibits *trans*-input independence. These examples are limited to demonstrating the independence of the LX core from changes in the upstream ZifH binding sites and do not attempt to examine the independence of ZifH activity from changes in the promoter core module. A full ensemble of titration strains and promoters were assembled towards this end, fusing the L18 core to the 5 different ZifH modules (Ellis, Wang, & Collins, 2009). These strains were never evaluated due to time constraints. It is likely that different expression regimes exhibit different rules for interaction with activators and repressors – it may be easier to repress a highly expressed gene than activate further expression, and vice versa for a lowly expressed gene. These expression regimes and their interaction with upstream modules could be evaluated using the different promoter cores generated by Ellis et al, but the difficulties in assembling upstream regions with certain cores needs to be considered (Ellis, Wang, & Collins, 2009).

The promoter architecture demonstrated in Chapter 3 used a small artificial regulatory network to titrate in non-native transcription factors against the ZiLX promoter library. These non-native TFs were selected to maximize the probability of achieving independence between the transcription factor inputs. However, the use of non-native transcription factors drastically increases the amount of genetic engineering required to control expression from the promoter library as the non-native transcription factors must be integrated into the genome to control

expression. The use of native transcription factors would lessen this engineering burden. Examination of different native transcription factor families and mechanisms of action for independence represents a logical extension to this research.

References

Abdulrehman, D., Monteiro, P. T., Teixeira, M. C., Mira, N. P., Lourenço, A. B., dos, S. C., et al. (2010). YEASTRACT: providing a programmatic access to curated transcriptional regulatory associations in Saccharomyces cerevisiae through a web services interface. *Nucleic acids research*, gkq964.

Ajo-Franklin, C., Drubin, D., Eskin, J., Gee, E., Landgraf, D., Phillips, I., et al. (2007). Rational design of memory in eukaryotic cells. *Genes Development*, 21(18):2271-6.

Balaji, S., Babu, M., Iyer, L., Luscombe, N., & Aravind, L. (2006). Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast. *Journal of molecular biology, 360*(1), 213-227.

Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F., & Cullin, C. (1993). A simple and efficient method for direct gene deletion in Saccharomyces cerevisiae. *Nucleic acids research, 21*(14), 3329.

Bellí, G. G., Piedrafita, L., Aldea, M., & Herrero, E. (1998). An activator/repressor dual system allows tight tetracycline-regulated gene expression in budding yeast. *Nucleic Acids Research*, 26(4): 942–947.

Blaiseau, P.-L., Lesuisse, E., & Camadro, J.-M. (2001). Aft2p, a novel iron-regulated transcription activator that modulates, with Aft1p, intracellular iron use and resistance to oxidative stress in yeast. *Journal of Biological Chemistry, 276*(36), 34221-34226.

Bonneau, R., Reiss, D. J., Shannon, P., Facciotti, M., Hood, L., Baliga, N. S., et al. (2006). The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. *Genome biology, 7*(5), R36.

Cahan, P., Li, H., Morris, S. A., da Rocha, E. L., Daley, G. Q., & Collins, J. J. (2014). CellNet: network biology applied to stem cell engineering. *Cell*, 158(4):903-915.

Cameron, D. E., Bashor, C. J., & Collins, J. J. (2014). A brief history of synthetic biology. *Nature Reviews Microbiology, 12*(5), 381-390.

Cardinale, S., & Arkin, A. P. (2012). Contextualizing context for synthetic biology--identifying causes of failure of synthetic biological systems. *Biotechnology journal, 7*(7), 856-866.

Chandel, A. K., & Singh, O. V. (2011). Weedy lignocellulosic feedstock and microbial metabolic engineering: advancing the generation of 'Biofuel'. *Applied microbiology and biotechnology, 89*(5), 1289-1303.

Chua, G., Morris, Q. D., Sopko, R., Robinson, M. D., Ryan, O., Chan, E. T., et al. (2006). Identifying transcription factor functions and targets by phenotypic activation. *Proceedings of the National Academy of Sciences, 103*(32), 12045-12050.

Chuang, H.-Y., Hofree, M., & Ideker, T. (2010). A decade of systems biology. *Annual review of cell and developmental biology, 26*, 721.

Dong, J., Fan, P., & Frizzel, R. A. (2008). Quantitative Analysis of the Packaging Capacity of Recombinant Adeno-Associated Virus. *Human Gene Therapy*, 7(17): 2101-2112.

Ellis, T., Wang, X., & Collins, J. J. (2009). Diversity-based, model guided construction of synthetic gene networks with predicted functions. *Nature Biotechnology*, 27, 465 - 471.

Faith, J. J., Hayete, B., Thaden, J. T., Mogno, I., Wierzbowski, J., Cottarel, G., et al. (2007). Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS biology, 5*(1), e8.

Fedor-Chaiken, M., Deschenes, R. J., & Broach, J. R. (1990). SRV2, a gene required for RAS activation of adenylate cyclase in yeast. *Cell*, 329–340.

Feng, R., Desbordes, S. C., Xie, H., Tillo, E. S., Pixley, F., Stanley, E. R., et al. (2008). PU.1 and C/EBPalpha/beta convert fibroblasts into macrophage-like cells. *Proceedings of the National Academy of Sciences*, 105(16):6057-62.

Gasch, A. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., Storz, G., et al. (2000). Genomic expression programs in the response of yeast cells to environmental changes. *Molecular biology of the cell, 11*(12), 4241-4257.

Geitz, R. D., & Schiestl, R. H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature Protocols*, 2(1):31-4.

Gerstein, M. B., Kundaje, A., Hariharan, M., Landt, S. G., Yan, K.-K., Cheng, C., et al. (2012). Architecture of the human regulatory network derived from ENCODE data. *Nature, 489*(7414), 91-100.

Gertz, J., Siggia, E., & Cohen, B. (2009). Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature*, 215-218.

Gonçalves, D. L., Matsushika, A., Belisa, B., Goshima, T., Bon, E. P., & Stambuk, B. U. (2014). Xylose and xylose/glucose co-fermentation by recombinant Saccharomyces cerevisiae strains expressing individual hexose transporters. *Enzyme and microbial technology, 63*, 13-20.

Gossen, M., & Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *PNAS*, 89(12): 5547–5551.

Guido, N., Wang, X., Adalsteinsson, D., McMillen, D., Hasty, J., Cantor, C., et al. (2006). A bottom-up approach to gene regulation. *Nature*, 856-60.

Harbison, C. T., Gordon, D. B., Lee, T. I., Rinaldi, N. J., Macisaac, K. D., Danford, T. W., et al. (2004). Transcriptional regulatory code of a eukaryotic genome. *Nature, 431*(7004), 99-104.

Haynes, B. C., Maier, E. J., Kramer, M. H., Wang, P. I., Brown, H., & Brent, M. R. (2013). Mapping functional transcription factor networks from gene expression data. *Genome research, 23*(8), 1319-1328.

Haynes, B. C., Skowyra, M. L., Spencer, S. J., Gish, S. R., Williams, M., Held, E. P., et al. (2011). Toward an integrated model of capsule regulation in Cryptococcus neoformans. *PLoS pathogens, 7*(12), e1002411.

Hector, R. E., Qureshi, N., Hughes, S. R., & Cotta, M. A. (2008). Expression of a heterologous xylose transporter in a Saccharomyces cerevisiae strain engineered to utilize xylose improves aerobic xylose consumption. *Applied microbiology and biotechnology, 80*(4), 675-684.

Hlynialuk, C., Schierholtz, R., Vernooy, A., & der, G. v. (2008). Nsf1/Ypl230w participates in transcriptional activation during non-fermentative growth and in response to salt stress in Saccharomyces cerevisiae. *Microbiology, 154*(8), 2482-2491.

Hu, Z., Killion, P. J., & Iyer, V. R. (2007). Genetic reconstruction of a functional transcriptional regulatory network. *Nature genetics, 39*(5), 683-687.

Kemmeren, P., Sameith, K., van, L. A., Benschop, J. J., Lenstra, T. L., Margaritis, T., et al. (2014). Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. *Cell, 157*(3), 740-752.

Kheradpour, P., Ernst, J., Melnikov, A., Rogov, P., Wang, L., Zhang, X., et al. (2013). Systematic dissection of regulatory motifs in 2000 predicted enhancers using a massively parallel reporter assay. *Genome Research*, :800-11.

Kwasnieski, J., Mogno, I., Myers, C., Corbo, J., & Cohen, B. (2012). Complex effects of nucleotide variants in a mammalian cis-regulatory element. *Proc Natl Acad Sci USA*, 19498-503.

Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., et al. (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. *science, 298*(5594), 799-804.

Lin, Y., Chomvong, K., Acosta-Sampson, L., Estrela, R., Galazka, J. M., Kim, S. R., et al. (2014). Leveraging transcription factors to speed cellobiose fermentation by Saccharomyces cerevisiae. *Biotechnology for biofuels, 7*(1), 126.

Litcofsky, K. D., Afeyan, R. B., Krom, R. J., Khalil, A. S., & Collins, J. J. (2012). Iterative plug-and-play methodology for constructing and modifying synthetic gene networks. *Nature methods, 9*(11), 1077-1080.

Marro, S., Pang, Z. P., Yang, N., Tsai, M. C., Qu, K., Chang, H. Y., et al. (2011). Direct lineage conversion of terminally differentiated hepatocytes to functional neurons. *Cell Stem Cell*, 9(4):374-82.

Matsushika, A., Goshima, T., Hoshino, T., & others. (2014). Transcription analysis of recombinant industrial and laboratory Saccharomyces cerevisiae strains reveals the molecular basis for fermentation of glucose and xylose. *Microbial cell factories, 13*(1), 16.

Matthew, S., Zhou, T., L, Y., Dantas-Machado, A., Gordân, R., & Rohs, R. (2014). Absence of a simple code: how transcription factors read the genome. *Trends in Biochemical Sciences*, 381-399.

Mogno, I., Kwasnieski, J., & Cohen, B. (2013). Massively parallel synthetic promoter assays reveal the in vivo effects of binding site variants. *Genome Research*, 1908-1915.

Mogno, I., Vallania, F., Mitra, R., & Cohen, B. (2010). TATA is a modular component of synthetic promoters. *Genome Research*, 1391-7.

Morris, S. A., & Daley, G. Q. (2013). A blueprint for engineering cell fate: current technologies to reprogram cell identity. *Cell research, 23*(1), 33-48.

Morris, S. A., Cahan, P. L., Zhao, A. M., San Roman, A. K., Shivdasani, R. A., & Collins, J. J. (2014). Dissecting engineered cell types and enhancing cell fate conversion via CellNet. *Cell*, 158(4):889-902.

Murphy, K., Balázsi, G., & Collins, J. (2007). Combinatorial promoter design for engineering noisy gene expression. *Proc Natl Acad Sci USA*, 12726-31.

Pampulha, M. E., & Loureiro-Dias, M. C. (1990). Activity of glycolytic enzymes of Saccharomyces cerevisiae in the presence of acetic acid. *Applied Microbiology and Biotechnology, 34*(3), 375-380.

Pedruzzi, I., Bürckert, N., Egger, P., & Virgilio, C. D. (2000). Saccharomyces cerevisiae Ras/cAMP pathway controls post-diauxic shift element-dependent transcription through the zinc finger protein Gis1. *The EMBO Journal, 19*(11), 2569-2579.

Ptashne, M., & Gann, A. (1997). Transcriptional activation by recruitment. *Nature*, 569-577.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*, gkv007.

Salusjärvi, L., Kankainen, M., Soliymani, R., Pitkänen, J.-P., Pentti, M., & Ruohonen, L. (2008). Regulation of xylose metabolism in recombinant Saccharomyces cerevisiae. *Microbial cell factories, 7*(1), 18.

Schena, M., Shalon, D., Davis, R. W., & O., B. P. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 270(5235):467-70.

Sharon, E., Kalma, Y., Sharp, A., Raveh-Sadka, T., Levo, M., Zeevi, D., et al. (2012). Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nature Biotechnology*, 521–530.

Spivak, A. T., & Stormo, G. D. (2012). ScerTF: a comprehensive database of benchmarked position weight matrices for Saccharomyces species. *Nucleic acids research, 40*(D1), D162--D168.

Stolovitzky, G., Monroe, D., & Califano, A. (2007). Dialogue on Reverse-Engineering Assessment and Methods. *Annals of the New York Academy of Sciences, 1115*(1), 1-22.

Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *cell, 126*(4), 663-676.

Temme, K., Zhao, D., & Voigt, C. (2012). Refactoring the nitrogen fixation gene cluster from Klebsiella oxytoca. *Proc Natl Acad Sci USA*, 7085-90.

Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., et al. (2009). Programming cells by multiplex genome engineering and accelerated evolution. *Nature, 460*(7257), 894-898.

Weirauch, M. T., Yang, A., Albu, M., Cote, A. G., Montenegro-Montero, A., Drewe, P., et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. *Cell, 158*(6), 1431-1443.