Contents

1 Biological Background and Research Goal

The central dogma of molecular biology describes eukaryotic gene expression as the transcription of DNA into mRNA in the nucleus of the cell followed by the translation of mRNA to protein in the cytoplasm (Lodish *et al.*, 2000). This process of gene expression is controlled by proteins called transcription factors that bind to regulatory DNA sequences near their target genes. There are two types of transcription factors: activators, which increase gene expression, and repressors, which decrease gene expression (Chen & Rajewsky, 2007). Eukaryotes exhibit combinatorial transcriptional regulation in that the expression of a single gene is often controlled by multiple activators and repressors. Furthermore, transcription factors are also encoded by genes and thus regulate each other (Lemon & Tjian, 2000).

Researchers have identified the target genes of transcription factors encoded in *Saccharomyces cerevisiae*, or budding yeast, using genome-wide location analysis. In this analysis, individual yeast strains were created that had a Myc epitope coding sequence inserted into genetic sequences that encode each of the 203 transcriptional regulators, which were identified by searching for all known and predicted transcription factors and nucleic acid binding proteins in the YPD and MIPS databases (Harbison *et al.*, 2004). PCR was used to confirm the tagged genetic sequences and Western blot analysis was used to verify that the transcription regulatory proteins contained the Myc epitope tag (Harbison *et al.*, 2004). The tagged proteins in individual strains were cross-linked to DNA using formaldehyde and then immunoprecipitated using a anti-myc antibody. After the cross-linking was reversed, the DNA was fluorescently labeled and hybridized to DNA microarrays containing DNA from intergenic regions in order to identify the genes bound by the transcriptional regulators (Harbison *et al.*, 2004; Lee *et al.*, 2002). This work increases our understanding of which transcription factors regulate which target genes and can be applied to understanding how yeast regulate gene expression under different conditions like in response to environmental change, for example (Lee *et al.*, 2002).

All organisms must be able to respond to environmental stresses, such as changes in temperature. The study of organisms' responses to temperature changes has had applications in biotechnology for culturing and preserving cells as well as in medicine for the preservation of tissues (Al-Fageeh & Smales, 2006). Organisms respond to temperature stress by changing gene expression (Aguilera, Randez-Gil, & Prieto, 2007). The heat shock response is one of the best characterized stress responses in prokaryotes and eukaryotes. To cope with high temperatures, both prokaryotes and eukaryotes activate the expression of a conserved group of proteins termed the heat shock proteins that relieve cellular stress (Al-Fageeh & Smales, 2006; Thieringer, Jones, & Inouye, 1998). In contrast, such a highly conserved set of proteins has not been identified to respond to cold stress. However, it is known that cold stress causes a decrease in membrane fluidity and diffusion as well as protein synthesis (Al-Fageeh & Smales, 2006). In various prokaryotes, cold stress induces the expression of genes involved in transcription, translation, DNA replication, ribosome assembly, and nucleic acid structure (Al-Fageeh & Smales, 2006; Aguilera, Randez-Gil, & Prieto, 2007). However, further research is necessary into the cold shock response in yeast.

Studies investigating the cold stress response in eukaryotic cells have focused on *Saccharomyces cerevisiae*, or budding yeast, an important model organism. The beer- and wine-making industry performs cold fermentation for certain products, such as lager beer and white wine, between $10\,^{\circ}\mathrm{C}$ and $12\,^{\circ}\mathrm{C}$. As this is a suboptimal temperature for yeast, which grow optimally at $30\,^{\circ}\mathrm{C}$,

understanding the cold shock response in yeast may lead to improvement in the fermentation process (Hernández-López *et al.*, 2011). Budding yeast has several characteristics which facilitate studies of its cold stress response. It is a unicellular eukaryote that can be grown on different media allowing researchers to study the organism's response under different chemical and physical conditions (Goffeau *et al.*, 1996). Yeast can easily be grown at low cost as they have a short reproduction time. In addition, it is easy to manipulate the yeast genome using standard molecular biology techniques. The genome of approximately 6000 genes has been completely sequenced (Goffeau *et al.*, 1996). An international consortium has generated and made available strains deleted for any annotated yeast gene (Winzeler *et al.*, 1999). The availability of these deletion strains facilitates investigation into the effect of cold shock on gene expression.

Yeast respond to cold shock by changing gene expression (Al-Fageeh & Smales, 2006; Aguilera, Randez-Gil, & Prieto, 2007). Two types of cold shock responses have been identified in budding yeast: the early cold response (ECR) that occurs within 2 hours of cold exposure and the late cold response (LCR) that occurs after 12 hours of exposure (Schade *et al.*, 2004). More is known about transcriptional regulation during the late cold response than the early cold response. The Msn2p and Msn4p transcription factors control the late cold response by up-regulating environmental stress response genes. Interestingly, some of the genes that encode heat shock proteins, specifically *HSP12, HSP26, HSP42, HSP104, YRO2,* and *SSE2*, are induced during the late cold response (Schade *et al.*, 2004). Furthermore, genes involved in ubiquitin-dependent protein catabolism, fatty acid, lipid and phospholipid metabolism, and transport are induced while those associated with amino acid and protein biosynthesis, ribosome assembly, mitosis, and cell cycle are repressed during the late cold response. However, the transcription factors that regulate gene expression during the early cold response are unknown.

The goal of my research is to determine the relative contribution of transcription factors in a network that regulates the cold shock response in budding yeast. Unlike previous work in the field, this work does not seek to develop a topological model linking the regulation of each gene to one of the 203 transcription factor based on genome wide location analysis data (Harbison *et al.*, 2004; Lee *et al.*, 2002). Furthermore, this research does not involve creating a biochemical model for a single gene such as those that compute the binding constants and rates for specific transcription factors (Gibson & Mjolsness, 2001). Instead, I present a "medium-scale" model of the dynamics of a network of 21 genes described by a system of nonlinear differential equations. This is a dynamic rather than static model of the transcriptional gene regulatory network controlling the cold shock response. Unlike a static model, a dynamic model of a gene regulatory network (GRN) describes change in gene expression over time (Filkov, 2006). Thus, a dynamic model can quantitatively characterize interactions between the nodes in a network and offer predictions about network behavior. In contrast, static models are more appropriate to study network topology (Filkov, 2006). Dynamic models have been created using an array of deterministic, such as Boolean and Hill functions, as well as probabilistic mathematical models, but I will describe models based on a sigmoid function and Michaelis-Menten kinetics (Wilczynski & Furlong, 2010). In the sections that follow, I will give an introduction to DNA microarray technology and the cold shock experiment used to generate data for the model. I will describe the statistical analysis conducted on this data to determine significant changes in gene expression during cold shock. I will explain the method and results of estimating the parameters of the model as well as changes that have been made to the model in order to more accurately represent the dynamics of gene expression. Lastly, I will describe limitations of the model and future directions.

2 Measuring Global Gene Expression Using Microarray Data

Microarrays measure the amount of mRNA transcribed from a gene for thousands of genes simultaneously (Tanay, Sharan, & Shamir, 2006). This technology is the high-throughput version of Southern blots developed by Ed Southern over thirty years ago with some important differences (Ball & Sherlock, 2006). In a Southern blot, a nucleic acid of interest is detected using a radioactive probe of known sequence among nucleic acids fixed on a membrane. In contrast, microarrays involve hybridizing a population of fluorescently labeled nucleic acids of unknown sequence to nucleic acids of known sequence immobilized on a solid support (Ball & Sherlock, 2006). This technology is typically used to determine differences in gene expression between cells in a control condition and experimental condition over a time course or in response to a perturbation to the cell (Filkov, 2006).

There are three types of microarray platforms that differ in the synthesis of the nucleic acids that are fixed to a solid support. Microarrays based on the design principles developed by Steve Fodor and colleagues at Affymetrix contain short oligonucleotides that are synthesized *in situ* by photolithography (Ball & Sherlock, 2006). Since these oligonucletoides may hybridize to several genes due to their short length, these types of microarrays must contain multiple match and mismatch probes for each gene so that users can gauge the amount of non-specific background hybridization (Chou, 2006). Unlike the short oligo (20-25 base pairs) arrays, there are microarrays that are made for all or a subset of genes of an organisms using oligonucleotides of 50-75 base pairs in length. It is sufficient to synthesize only one probe for each gene because the length of the oligonucleotide results in high specificity for a specific target (Chou, 2006). Unfortunately, oligonucleotide arrays carry the risk of non-specific hybridization, if for some reason the *in situ* synthesis of the oligonucleotides fails at some step (Ball & Sherlock, 2006). The research presented here utilizes microarrays spotted with PCR products of full-length genes or oligonucleotide arrays. These arrays are created by robots with multi-tipped print heads that spot DNA onto a glass microscope slide (Ball & Sherlock, 2006; Chou, 2006). Unfortunately, DNA microarrays spotted with full-length genes do not have the same specificity as oligonucletoide arrays because it is difficult to find a temperature at which it is possible to isolate correct hybridization of the target to the probe from nonspecific binding for all genes (Chou, 2006).

While the aforementioned microarrays differ in their design, these microarray platforms are used similarly to measure changes in gene expression. In the experiment described in this paper, total RNA samples collected from a control and experimental condition are reverse transcribed into cDNA with an oligo dT primer which selects for mRNA. The cDNA is linearly amplified by synthesizing second strand DNA and then performing *in vitro* transcription to synthesize aRNA using an amino allyl modified UTP nucleotide. Samples are fluorescently labeled, mixed, and washed over a slide in a competitive hybridization. A scanning laser microscope creates an image of the slide, which is analyzed by software that measures the fluorescence of each spot (Ball & Sherlock, 2006). As these microarray experiments produce large datasets containing information for thousands of genes per sample analyzed, it has been necessary to develop databases that can store these datasets and make them available to scientists for further analysis and comparative studies. Several microarray databases, such as ArrayExpress and Gene-Expression Omnibus, have emerged that store gene expression data for different organisms (Anderle *et al.*, 2003).

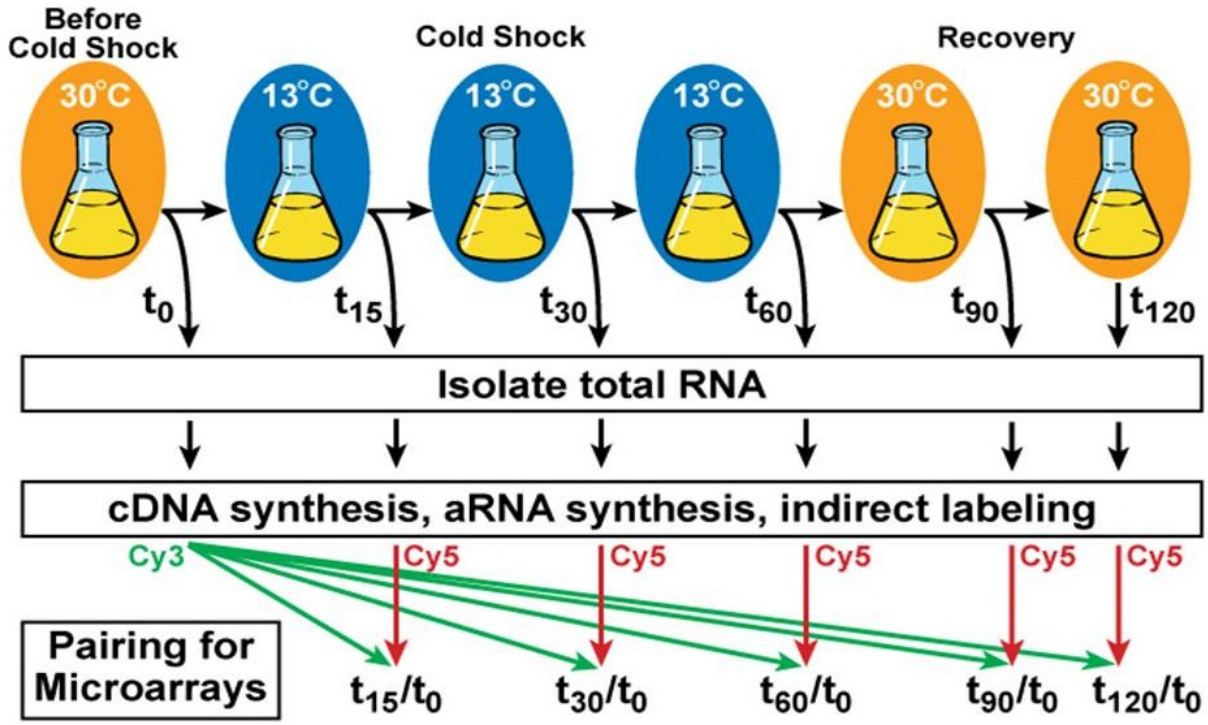

Figure 1. Preparation of microarrays from total RNA isolated from each strain at six time points over the course of the cold shock and recovery experiment. In half of the replicates, the t_0 time point was labeled with Cy3 and the other time points with Cy5. In the other half of the replicates, the orientation of the dyes was reversed.

2.1 Cold Shock Experiment Design

Global gene expression was measured in the *S. cerevisiae* BY4741 wild type strain as well as strains deleted for the genes *CIN5*, *GLN3*, *HMO1*, and *ZAP1*, using two different DNA microarrays: GCAT chips containing 70-mer oligonucleotides obtained by the Genome Consortium for Active Teaching and Yeast 6.4K Array containing full-length PCR products obtained from the UNH Microarray Centre ("Ontario" chips) (Campbell, 2009; UNH Microarray Centre, 2012). The GCAT chips were used to collect data for the wild type strain before switching to Ontario chips for the remainder of the experiments. After being allowed to grow to early log phase at its optimal temperature of $30\,^{\circ}\mathrm{C}$, cultures of each strain in YEPD medium were subjected to cold shock at $13\,^{\circ}\mathrm{C}$ for one hour and then allowed to recover at $30\,^{\circ}\mathrm{C}$ for one hour (Figure 1). Samples were harvested at 0 (t_0), 15 (t_{15}), 30 (t_{30}), 60 (t_{60}), 90 (t_{90}), and 120 (t_{120}) minutes into the experiment. Total RNA was purified and first and second strand cDNA was synthesized, which was then used as a template to synthesize amino allyl-aRNA. Samples were fluorescently labeled with either a green dye called Cy3 or a red dye called Cy5. In half of the experiments, the samples from the t_0 time point were labeled with Cy3 and the other samples were labeled with Cy5. The dye orientation between the control time point t_0 and the other time points were swapped in the other half of the experiments to mitigate potential gene expression differences between the time points as a result

of inconsistencies between the dyes. Microarrays were hybridized with labeled aRNA from the t_0 time point and one of the other time points in the experiment (Figure 1). Each microarray was then scanned using an Axon GenePix 4000B Scanner after which GenePix Pro 6.1 software was used to determine the ratio of red to green dye for each spot. These ratios, which are henceforth referred to as fold change ratios, were transformed by the log base 2 in order to make the data more symmetric. Three to five replicates (usually four) of the cold shock and recovery experiment were performed for each time point for each strain for a total of 103 microarrays (Table S1). Only data from the cold shock time points was used in the model.

2.2 Microarray Data Normalization

Microarray data must be normalized before it is analyzed because there are biases inherent to the technology. A common source of error comes from differences in the intensity of the fluorescent dyes. These differences stem from several factors, such as physical properties of the dyes, dye incorporation efficacy, and the scanner settings (Yang *et al.*, 2002). In addition, microarrays are also prone to spatial biases due to differences in the length or opening of the print tips used to spot the microarrays with DNA (Yang *et al.*, 2002). Dye intensity biases become visible when creating scatter plots (henceforth called *MA* plots) of the log fold change ratio versus the intensity of each spot (Quackenbush, 2002). In the name *MA* plot, *M* refers to the log fold change computed as the log base 2 of the ratio of the red to green dye intensity ($M = \log_2 R/G$), while *A* refers to the intensity of each microarray spot calculated as the log base 2 of the product of the red and green dye intensities ($M = \log_2 R \cdot G$). Spatial biases can be observed in box plots of the log fold changes for each spot on a chip. Within chip normalization corrects for dye intensity biases by adjusting the *M*-values within each array while scale normalization addresses the fact that there are scaling differences between chips due to variation between the print tips used to create the arrays and any technical differences between experiments (Smyth *et al.*, 2005).

Loess normalization and median absolute deviation scaling were used to perform within array and scaling normalization, respectively. Loess normalization is a robust local linear regression, fitting a curve through each point using a subset of the data around each point. The normalized values are obtained by subtracting the slope of the loess fit from each log fold change (Yang *et al.*, 2002). In effect, the data in the *MA* plot becomes centered at 0. Median absolute deviation scaling subtracts the median log fold change from each point and then finds the new median of the chip. The normalized values are obtained by dividing each log fold change by this new median (Yang *et al.*, 2002). This procedure normalizes the spread of the log fold changes across all of the microarrays to make sure that no one chip has undue weight on the data analysis. Both types of normalization were performed using the R Statistical Software 2.7.2 (R Development Core Team, 2005) and the limma package (Smyth *et al.*, 2005). The protocol is available on OpenWetWare (Sherbina, 2014a). Table S1 gives a record of all 103 arrays hybridized and normalized in this fashion.

Normalization is different from standard statistical approaches to analyzing data. Many statistical tests aim to draw inferences about the mean or median between different treatments. Analysis of variance, or ANOVA, is one such statistical analysis that tests for the equality of means between different conditions measuring a certain response variable (Larson, 2008). However, in normalizing microarray data, we are instead interested in detecting departures from the mean or, in other words, outliers that represent the genes that have a significant change in expression between the control and experimental conditions. However, in order to draw accurate conclusions about gene expression

levels, it is necessary to normalize microarray data in order to remove low-quality measurements (Quackenbush, 2002).

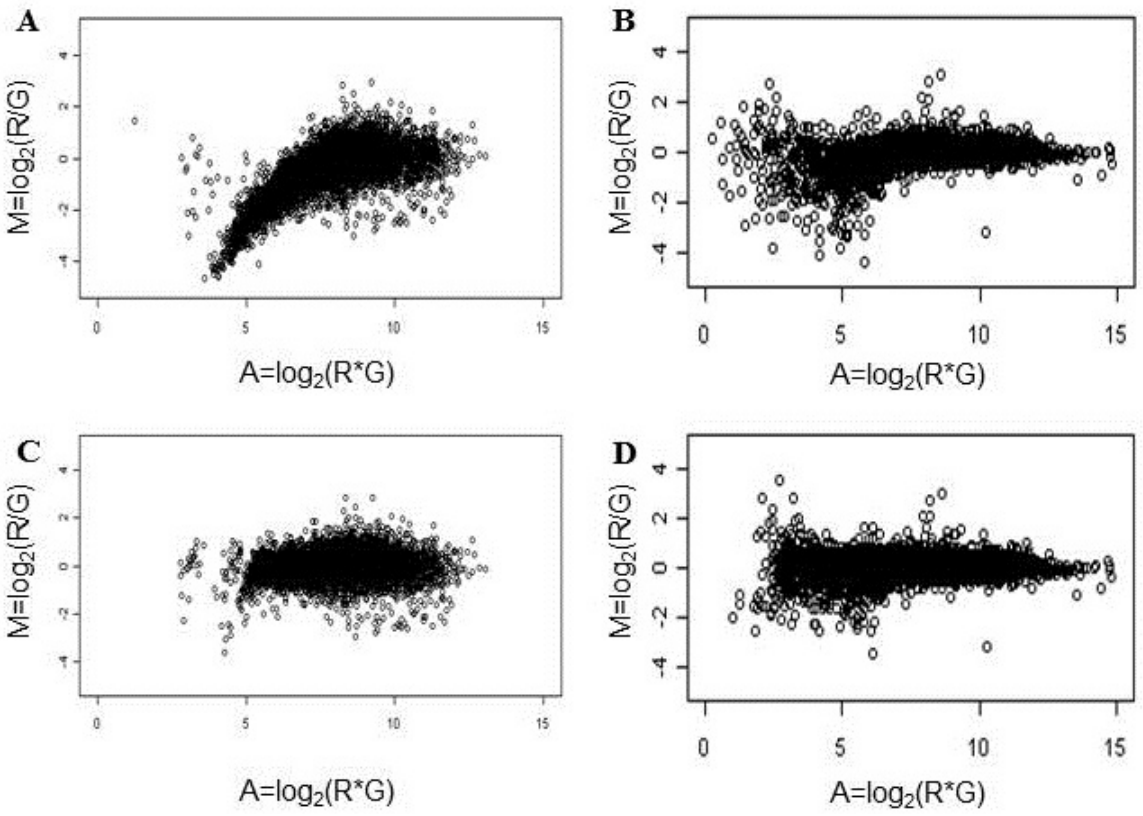

Figure 2. *MA* plots graph the log fold change (*M*) versus the total intensity (*A*) of each spot. In the formulas for*M* and *A* on the axis labels, the *R* refers to the red dye (Cy5) intensity and *G* refers to the green dye (Cy3) intensity. Plots **A** and **B** represent the wild type strain data before and after loess normalization, respectively. Plots **C** and **D** represent the Δ*zap1* strain data before and after loess normalization, respectively.

2.2.1 Visual Comparison of Data Before and After Normalization

MA plots created for each strain and replicate before and after loess normalization show the effectiveness of the normalization procedure that I used. Before normalization, there is a noticeable bias in the data for dim green spots, which have a low intensity and negative log fold change (Figure 2A, 2B, S1, S3, S5, S7, S9, S11). This is likely a result of the shorter half-life of the Cy5 dye in comparison to the Cy3 dye (Sabanayagam & Lakowicz, 2007). Loess normalization removes this bias, centering all of the data at a log fold change of 0 (Figure 2C, 2D, S2, S4, S6, S8, S10, S12). This change in the median of the data is acceptable as the underlying assumption of this normalization is that a majority of genes in the yeast genome do not change expression in the cold shock response.

Box plots created for each strain and replicate before normalization, after loess normalization, and after median absolute deviation scaling both reinforce the aforementioned observations with

regard to loess normalization and demonstrate the success of between array normalization by median absolute deviation scaling. As seen in the box plots, the median and spread of the data is different between the replicate chips for each strain (Figure 3A, S13A, S14A, S15A, S16A, S17A). However, after loess normalization, the median log fold change for all of the replicates for each strain is normalized to 0 (Figure 3B, S13B, S14B, S15B, S16B, S17B). After median absolute deviation scaling, the spread of the data in terms of the interquartile range across the replicates for each strain is normalized (Figure 3C, S13C, S14C, S15C, S16C, S17C). Both the box plots and the *MA* plots demonstrate the effectiveness of loess normalization and median absolute deviation scaling to resolve dye and spatial biases in both GCAT and Ontario microarrays.

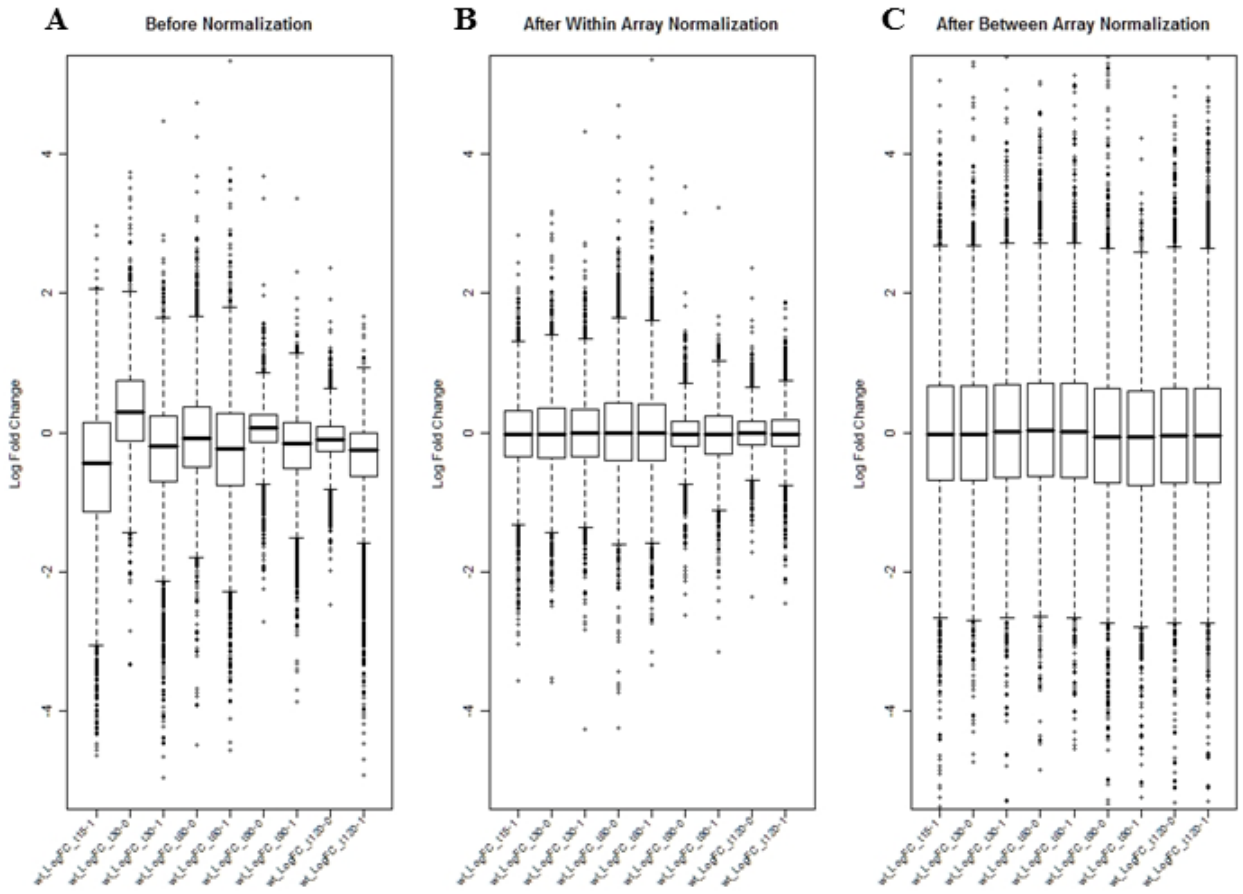

Figure 3. The effects of within array and between array normalization on the log fold change of spots on the GCAT DNA microarrays collected for the wild type strain. **(A)** Before normalization. **(B)** After within array normalization by loess. **(C)** After between array normalization using median absolute deviation scaling. In plots **A**, **B**, and **C**, the box plots represent the following sequence of wild type data: t_{15} flask 1, t_{30} flask 0, t_{30} flask 1, t_{60} flask 0, t_{60} flask 1, t_{90} flask 0, t_{90} flask 1, t_{120} flask 0, and t_{120} flask 1.

2.3 Determining Significant Changes in Gene Expression with Statistical Analysis

Two types of statistical analyses were carried out to determine which genes had a significant change in gene expression during the cold shock and recovery experiment. First, an ANOVA statistic was computed for each gene to determine which genes had a significant change in expression during at least one time point in the cold shock and recovery experiment for each yeast strain. The null hypothesis for each test was that the average log fold change of the gene g for time point j is 0: $\mu_j^g = 0$. In other words, the null hypothesis states that the gene does not have a significant change in expression over the course of the experiment at any of the time points. The statistic is computed as follows

$$F = \frac{n-p}{q} \frac{SS_{full} - SS_{hyp}}{SS_{full}} \tag{1}$$

where n is the number of data points for a gene, p is the number of time points, and $q = p \times (\text{number of strains} - 1)$. SS_{full} is the sum of the square of the error between the log fold change for each gene g at time point j for replicate i (Y_{ij}^g) and the average log fold change for a gene g at time point j ($\bar{Y}_j^g$):

$$SS_{full} = \sum_{i,j} (Y_{ij}^g - \bar{Y}_j^g)^2 \tag{2}$$

SS_{hyp} is the sum of the square of the error between the log fold change for each gene g at time point j for replicate i (Y_{ij}^g) and the average log fold change for a gene g under the null hypothesis:

$$SS_{hyp} = \sum_{i,j} (Y_{ij}^g - 0)^2 \tag{3}$$

The subscript $full$ refers to the model based on the alternative hypothesis that the average log fold change of a gene is not zero while the subscript hyp refers to the model based on the null hypothesis. For the data collected, n was either 23 or 20 depending on whether the wild type or a deletion strain was analyzed, respectively; p was 5; and q was also 5. A p value was computed for each gene using the F distribution corresponding to the computed ANOVA statistic; a significance level α of 0.05, 0.01, or 0.001; and $n - p$ and q degrees of freedom (Graybill, 1976).

A second statistical analysis was carried out using a general linear model to test the hypothesis that each gene shows the same pattern of gene expression over the time course of the experiment in each strain. Comparisons were performed between the wild type strain and each of the deletion strains as well as between pairs of deletion strains in order to test the null hypothesis that the average log fold change of a gene g for a time point j in one strain s is the same as that in another strain s': $\mu_j^{g,s} = \mu_j^{g,s'}$. An F statistic is computed for each gene using equation 1, where the SS_{full} and SS_{hyp} are computed as follows where $full$ refers to the model based on the alternative hypothesis that genes do not follow the same dynamics between strains and hyp refers to the model based on the null hypothesis:

$$SS_{hyp} = \sum_{i,j,s} (Y_{ij}^{g,s} - \bar{Y}_j^{g,s})^2, \text{where } s \text{ refers to strain} \tag{4}$$

$$= (Y - X_{full}\beta_{full})^T (Y - X_{full}\beta_{full}) \tag{5}$$

Y is a vector of the log fold changes of a gene for each time point and each replicate in the cold shock and recovery experiment, X_{full} and X_{hyp} are both matrices representing the independent variables, and β_{full} and β_{hyp} are vectors of population constants that must be estimated. X_{full} is an $n \times p$ matrix, where n is the number of data points and p is the number of time points multiplied by the number of strains while X_{hyp} is an $n \times q$ matrix where $q = p/$number of strains. Each matrix is composed of 0's and 1's in which there is a 1 in every cell that represents data available for a gene in a given strain for a certain time point and replicate. The parameters β are estimated by solving a linear system of equations in the following form:

$$Y = X_{full}\beta_{full} \tag{6}$$

$$Y = X_{hyp}\beta_{hyp} \tag{7}$$

A p value was computed for each gene using the F distribution corresponding to the computed F statistic, an α of 0.05, and $n - p$ and q degrees of freedom.

2.3.1 More Genes Have a Significant Change in Expression than Expected at a Particular Significance Level

In each of the statistical analyses, the number of genes with a significant change in expression was much higher than expected by the α level. With $\alpha = 0.05$, approximately 309 genes are expected to have a nonzero change in expression due to chance. However, for both statistical analyses described in Section 2.3, there are more than 309 genes with significant differential expression (Table 1 and Table 2). Similarly, the number of genes exhibiting a significant change in expression at $\alpha = 0.01$ is higher than expected, which is approximately 62 genes (Table 1). While only about 6 genes are expected to have significant changes in expression with $\alpha = 0.001$, only the Δ*zap1* strain met the expectation not having any genes with a $p < 0.001$ (Table 1). Therefore, differential expression was demonstrated in all strains at all significance levels with a false positive rate of α, with the exception of Δ*zap1* at $\alpha = 0.001$. Of the genes showing differential expression at the 0.05 significance level, 5% of these genes were changed by chance but it is not known which changes in gene expression are actually false positives. Therefore, as described in the next section, this type of statistical analysis necessitates the use of a correction method for false positives. The use of the correction method does not eliminate the the problem, but only increases our confidence that a particular gene has a significant change in expression.

2.3.2 Challenges of Multiple Hypothesis Testing

Both of the statistical tests previously described are examples of multiple hypothesis testing since the null hypotheses is being tested for thousands of genes simultaneously. Multiple hypothesis testing increases the probability of introducing Type I errors, or false positives, and Type II errors, or false negatives. In the case of analyzing differential gene expression using microarray data, a Type I error refers to falsely determining that a gene has a significant change in expression. A Type II error refers to inaccurately ruling that a gene does not have a significant change in expression when the opposite is true (Dudoit, Shaffer, & Boldrick, 2003). There are two categories of Type I error rate that are addressed in this analysis: family-wise error rate (FWER) and false discovery rate (FDR). The FWER is the probability of committing at least one Type I error. The FDR is the

proportion of Type I errors among the number of correctly rejected null hypotheses (Dudoit, Shaffer, & Boldrick, 2003).

To mitigate the Type I error rate, two multiple testing corrections were applied on the unadjusted p values after performing the statistical analysis described in Section 2.3: the single-step Bonferroni correction for the FWER and the Benjamini & Hochberg correction for the FDR. The single-step Bonferroni correction rejects an hypothesis if

$$P \leq \frac{\alpha}{m}, \tag{8}$$

where P is the unadjusted *p* value, α is the significance level, and m is the total number of hypotheses tested (Dudoit, Shaffer, & Boldrick, 2003). The Benjamini & Hochberg correction requires that all of the unadjusted *p* values are sorted in increasing order. This correction rejects a hypothesis if

$$P^{(i)} \leq \frac{\alpha \times i}{m} \tag{9}$$

where P is the unadjusted *p* value of rank i, α is the significance level, and m is the total number of hypotheses tested (Benjamini & Hochberg, 1995). The advantage of the Benjamini & Hochberg correction is that it is less stringent and thus has more power than the single-step Bonferroni correction (Benjamini & Hochberg, 1995). This is an important distinction between these two correction methods. The goal of a multiple testing procedure is to minimize the Type I error rate while maximizing the power, or minimizing the Type II error rate (Dudoit, Shaffer, & Boldrick, 2003).

The analysis of the significance of gene expression changes within each strain was performed using both Excel and MATLAB. The resulting *p* values associated with each ANOVA statistic were adjusted using the single-step Bonferroni and the Benjamini & Hochberg corrections to compare the stringency of those methods. Only the Benjamini & Hochberg correction was applied to the *p* values generated by the statistical analysis in MATLAB. The general linear model to determine if any genes had significant changes in expression between strains was implemented in MATLAB only. The *p* values resulting from this analysis were adjusted using the Benjamini & Hochberg correction. Detailed protocols and code for both analyses can be found on OpenWetWare (Sherbina, 2014b).

2.3.3 Strain Differences in Number of Genes with Significant Changes in Expression

Statistical analysis of significant gene expression changes within each strain of yeast demonstrates the differences in stringency between the single-step Bonferroni and Benjamini & Hochberg correction methods. Prior to these Type I Error correction methods, the Δ*cin5* and Δ*gln3* strains show the greatest number of genes with a significant change in expression with a *p* value of less than 0.05, 0.01, and 0.001 (Table 1). After the single-step Bonferroni correction, only the Δ*cin5* and Δ*zap1* strains showed any genes with significant changes in expression, specifically only 0.02% of the genes in the genome (Table 1). However, after the Benjamini & Hochberg correction, the Δ*gln3* strain in addition to the Δ*cin5* and Δ*zap1* strains showed some significant changes in gene expression with theΔ*cin5* and Δ*gln3* strains showing a greater portion of genes with significant changes in expression with respect to the total genome than the Δ*zap1* strain (Table 1). These results

Table 1. Summary statistics of the number and percent of genes with a significant change in gene expression under three α values (0.05, 0.01, 0.001) and the following two Type I error corrections: the single-step Bonferroni correction and the Benjamini & Hochberg (B & H) correction. The statistical analysis was performed using normalized microarray data for a total of 6,189 genes.

	wild type	Δ*cin5*	Δ*gln3*	Δ*hmo1*	Δ*zap1*
$p < 0.05$	1095 (17.7%)	1591 (25.7%)	1810 (29.3%)	760 (12.3%)	512 (8.27%)
$p < 0.01$	312 (5.04%)	645 (10.4%)	697 (11.3%)	228 (3.68%)	117 (1.89%)
$p < 0.001$	34 (0.55%)	115 (1.86%)	120 (1.94%)	33 (0.53%)	0
Bonferroni $p < 0.05$	0	1 (0.02%)	0	0	1 (0.02%)
B & H $p < 0.05$	0	102 (1.65%)	12 (0.19%)	0	1 (0.02%)

Table 2. Summary statistics of the number and percent of genes with a significantly different expression profile (p value < 0.05) determined by comparing expression data for pairs of strains using a general linear model with a Benjamini & Hochberg correction ($B\&H$) refers to the The statistical analysis was performed using normalized microarray data for a total of 6,189 genes.

Strains Compared	$p < 0.05$	$p_{B\&H} < 0.05$
wild type vs Δ*cin5*	507 (8.19%)	4 (0.06%)
wild type vs Δ*cin5*	697 (11.3%)	38 (0.61%)
wild type vs Δ*hmo1*	520 (8.40%)	5 (0.08%)
wild type vs Δ*zap1*	545 (8.81%)	30 (0.48%)

corroborate earlier discussion regarding the decreased stringency of the Benjamini & Hochberg correction in comparison to the single-step Bonferroni correction. Since there were no genes that showed up as significantly differentially expressed after the Benjamini & Hochberg correction, it seems that the deletion of *HMO1* severely impairs the yeast's ability to respond to cold shock (Table 1).

As with the aforementioned statistical analysis of the data for each strain, the general linear model to test differences in the gene expression profiles between strains shows that fewer genes appear to have a significant change in expression after the Benjamini & Hochberg correction. The Δ*gln3* strain shows the greatest number of genes with different expression profiles than the wild type strain while the Δ*cin5*, Δ*hmo1*, and Δ*zap1* strains have similar numbers of genes that have significantly different profiles than the wild type strain (Table 2).

3 Gene Regulatory Network Involved in the Cold Shock Response

A gene regulatory network consists of genes, regulators, and the regulatory connections between them (Filkov, 2006). The regulatory network presented in this paper consists of 21 nodes representing the gene, mRNA, and transcription factor that it encodes with the assumption that the gene is immediately translated after being transcribed (Figure 4). The nodes are connected by 50 edges, which represent gene activation or repression depending on the sign of the weight of the regulatory effect. Transcription factors were included in the network if they satisfied one of two criteria. First, transcription factors were included if their target genes were enriched in a list of genes with a significant change in expression in the microarray data denoted by a p value<0.05 after an ANOVA test and Benjamini & Hochberg correction. The second criterion was that the transcription factor is potentially involved in the cold shock response as suggested by other experimental evidence. Once the list of 21 transcription factors was generated, the list was submitted to the YEASTRACT database "generate regulation matrix" function using documented and direct relationships only (Teixeira *et al.*, 2014). This generated the 50 edges shown in Figure 4.

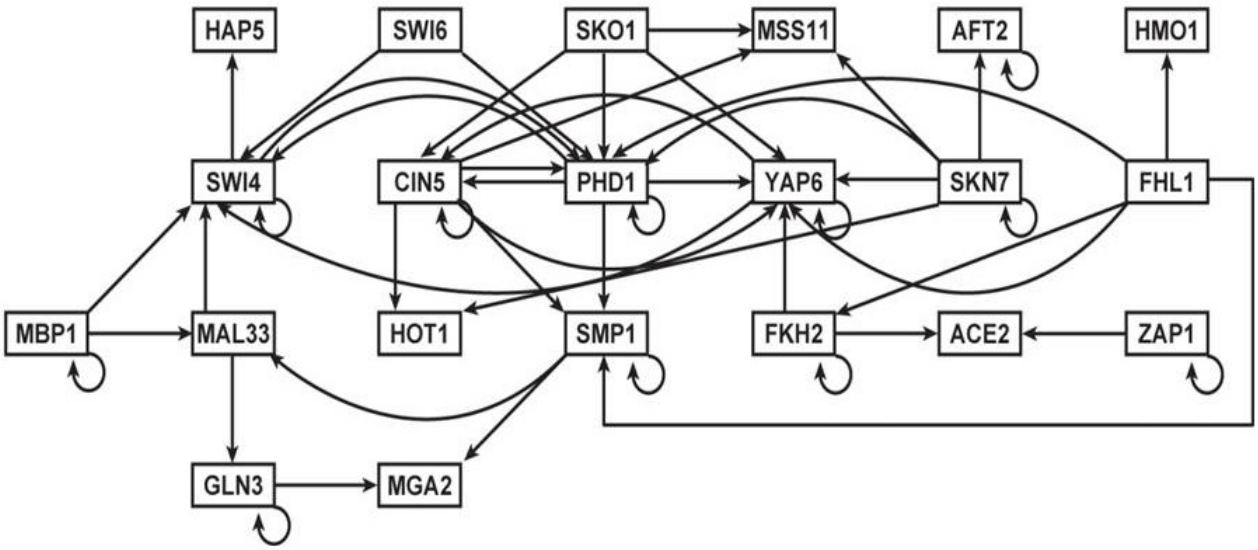

Figure 4. The gene regulatory network consists of 21 nodes representing the gene, mRNA, and transcription factor that it encodes, assuming that the gene is immediately translated after transcription. The nodes are connected by 50 edges which represent activation or repression depending on the sign of the weight of the regulatory effect.

4 Gene Expression is a Dynamic Balance Between Production and Degradation

The aforementioned gene regulatory network (GRN) is modeled by a system of nonlinear ordinary differential equations one for each gene in the network. This is a deterministic model in which the state of gene is given by a formula (Filkov, 2006). In the model that will be presented in this paper, the formula for the rate of change of expression for each gene i is given by a nonlinear ordinary differential equation (ODE) consisting of a production term and a degradation term:

$$\frac{dx_i}{dt} = p(\vec{x}) - \lambda_i x_i \tag{10}$$

In this model, the production and degradation rates for mRNA and protein are not separated for the sake of simplicity. The degradation of mRNA is represented by the linear function $\lambda_i x_i$ where x_i is the log fold change of a gene and λ_i is the degradation rate constant based on protein half-life data provided by Belle *et al.* (2006). The protein half-lives were converted to degradation rates for each mRNA i by the following equation:

$$\lambda_i = \frac{\ln(0.5)}{\text{half-life}_i} \tag{11}$$

In the case that a protein-half life for a corresponding gene was not available, the degradation rate was calculated by equation 11 using the median protein half-life of 202 transcription factors, which was 25.5 min (Table 3). The production of a gene i is given by $p(\vec{x})$ where $\vec{x}$ consists of the log fold changes of all of the transcription factors that regulate gene i. The production of a gene i is not only dependent upon the network of regulating genes whose transcription factors activate or repress but also a production rate constant. The initial guess of the production rate constant for each gene i was double the degradation rate corresponding to that gene (Table 3). Final production rates were then estimated by the model.

4.1 Sigmoidal Model

In the first iteration of this nonlinear differential equation model, gene production was modeled by a sigmoid function based on the model presented by Vu & Vohradsky (2007):

$$p(\vec{x}) = \frac{P_i}{1 + \exp(\sum_j -w_{ij}x_j + b_i)} \tag{12}$$

P_i is the production rate constant of a gene i the initial value of which is twice the gene's degradation rate. The production of a gene i is also influenced by the weighted w_{ij} concentrations of all transcription factors j that regulate it. The sign of w determines whether or not a target gene is activated or repressed: up-regulation or activation is denoted by a positive w while down-regulation or repression is denoted by a negative w (Figure 5). The magnitude of w corresponds to the strength of the regulatory effect where larger values of $|w|$ denote stronger activation or repression, or a steeper slope in the sigmoid curve (Figure 5). The position of the expression threshold of gene i is given by the constant b_i. This threshold corresponds to the midpoint of the change in gene expression from 0 to 1 in the case of activation or decrease from 1 to 0 in the case of repression, where 0 indicates that the expression of a gene is off while 1 indicates that gene expression is on.

Table 3. Degradation rates were computed for each transcription factor in the gene regulatory network in Figure 4 based on the protein half-life data provided by Belle *et al.* (2006). The initial production rates were computed by doubling the degradation rates. The production rates were then estimated in the model. For those proteins with an asterisk, half-life data was not available so the median half-life of 202 transcription factors was used for these proteins.

Protein	Half-Life (min)	Degradation Rate (min^{-1})	Production Rate (min^{-1})
ACE2	3	0.2310	0.4621
AFT2	35	0.0198	0.0396
CIN5*	25.5	0.0272	0.0544
FHL1	40	0.0173	0.0347
FKH2	26	0.0267	0.0533
GLN3	3	0.2310	0.4621
HAP5	300	0.0023	0.0046
HMO1	300	0.0023	0.0046
HOT1	3	0.2310	0.4621
MAL33*	25.5	0.0272	0.0544
MBP1	20	0.0347	0.0693
MGA2	31	0.0224	0.0447
MSS11	19	0.0365	0.0730
PHD1	14	0.0495	0.0990
SKN7	23	0.0301	0.0603
SKO1	18	0.0385	0.0770
SMP1*	25.5	0.0272	0.0544
SWI4	141	0.0049	0.0098
SWI6	17	0.0408	0.0815
YAP6	21	0.0330	0.0660
ZAP1	163	0.0043	0.0085

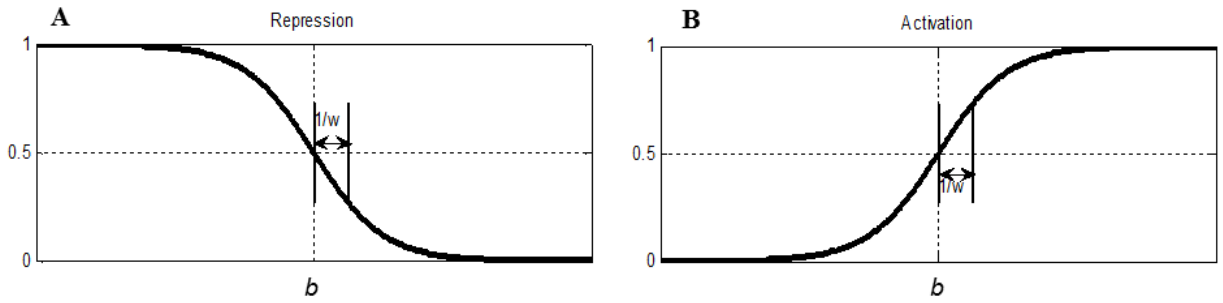

Figure 5. The gene regulatory network consists of 21 nodes representing the gene, mRNA, and transcription factor that it encodes, assuming that the gene is immediately translated after transcription. The nodes are connected by 50 edges which represent activation or repression depending on the sign of the weight of the regulatory effect.

4.2 Model Based on Michaelis-Menten Kinetics

The production term of the differential equation for gene expression was altered because the sigmoidal model does not accurately describe transcriptional regulation. The construction of the sigmoidal model suggests that there is a high level of initial transcription of a gene before the binding of a repressor, which may not be the case in a true biological system (Figure 5A). Furthermore, in the case of multiple regulators, the sigmoidal model does not accurately represent either an "AND" or an "OR" transcriptional gate type (Figure 6). An "AND" gate indicates that all transcription factors are required to regulate the target gene while an "OR" gate indicates that not all transcription factors of a target gene are required to regulate the transcription of that gene. An "OR" gate is more reflective of the combinatorial control expected in a biological system.

To address the drawbacks of the sigmoidal model, gene production was redefined using Michaelis-Menten kinetics. Michaelis-Menten kinetics describe the rate at which the product P of an enzymatic reaction is synthesized:

$$f([S]) = \frac{d[P]}{dt} = \frac{V_{max}[S]}{K_m + [S]} \tag{13}$$

where V_{max} is the maximum reaction rate, $[S]$ is the concentration of the substrate, and K_m is the substrate concentration at which the enzyme reaction rate is at 50%. There are several important characteristics of this function (Robeva & Yildirim, 2013):

- The right side of equation 13 is an increasing function in terms of the substrate concentration [S].
- $\lim_{[S]\to\infty} f([S]) = V_{max}$
- $f([S] = K_m) = \frac{V_{max}}{2}$

Khanin *et al.* (2007) used Michaelis-Menten kinetics to describe the regulation of a gene by one transcription factor. The Michaelis-Menten model for gene regulation describes the transcription rate as a function of the transcription factor concentration where K_m is the concentration of the

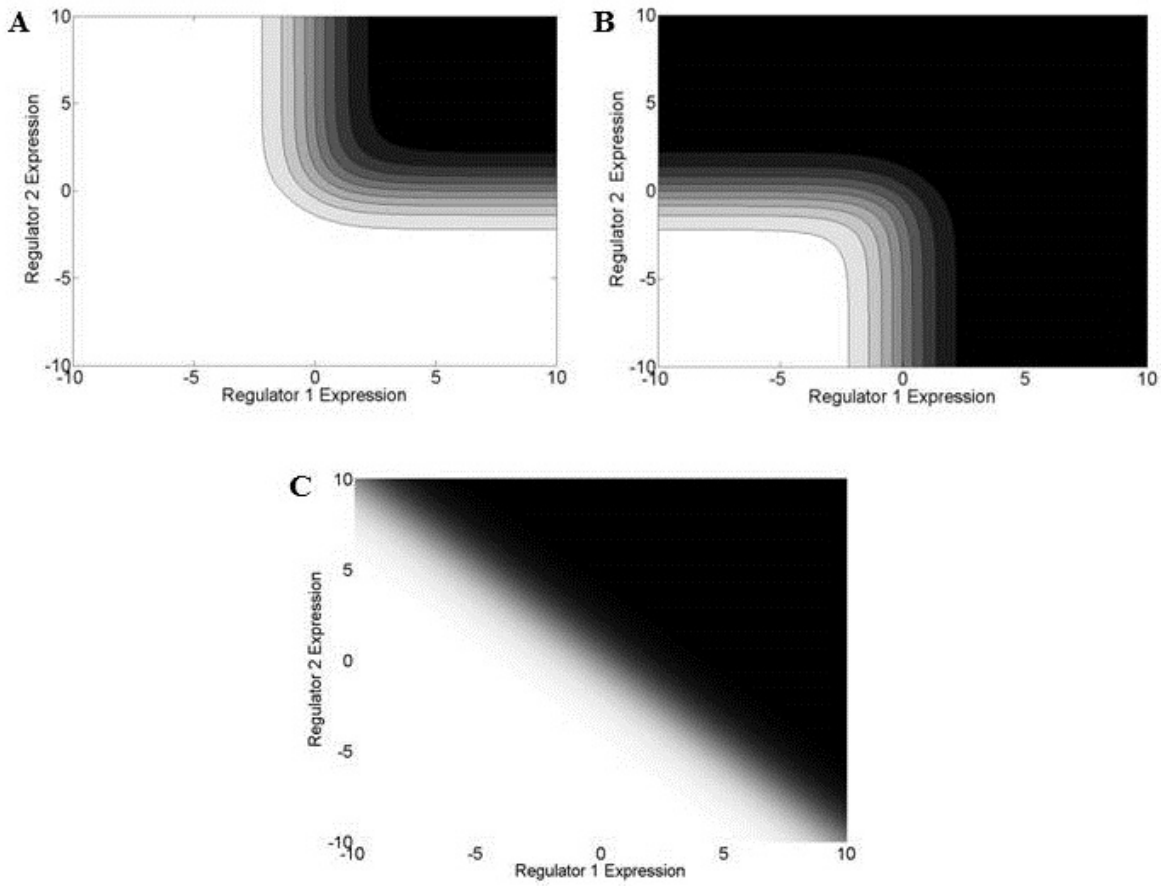

Figure 6. Representations of the regulation of a target gene controlled by two regulators in the case of **(A)** AND and **(B)** OR transcriptional gates in comparison to a **(C)** sigmoidal model. The x-axis denotes the log fold change of one regulator while the y-axis denotes the log fold change of the second regulator.

transcription factor at which the transcription rate is at 50% of its maximum rate V_{max} (Figure 7). This research uses the model by Khanin *et al.* (2007) to describe the combinatorial regulation of a gene by all of its transcription factors.

As in the sigmoid model, the rate of gene expression is described as the difference between the production and degradation of the gene product. However, in this new model, the production of a gene explicitly takes into account the "OR" nature of activation and repression:

$$p(\vec{x}) = P_i \cdot (\sum_j ([\frac{w_{ij}x_j}{\sum_k w_{ik}x_k}] \cdot [\frac{w_{ij}x_j}{1 + w_{ij}x_j}] I(w_{ij} > 0)) \tag{14}$$

The first bracketed factor represents the relative weight of transcription factor j on gene i. The second bracketed factor represents the Michaelis-Menten reaction rate for transcription. The third term models the effect of repression. If the regulatory weight is positive, the production of a gene product is modeled by the Michaelis-Menten production term. However, if the regulatory weight is negative, then neither mRNA nor protein is produced.

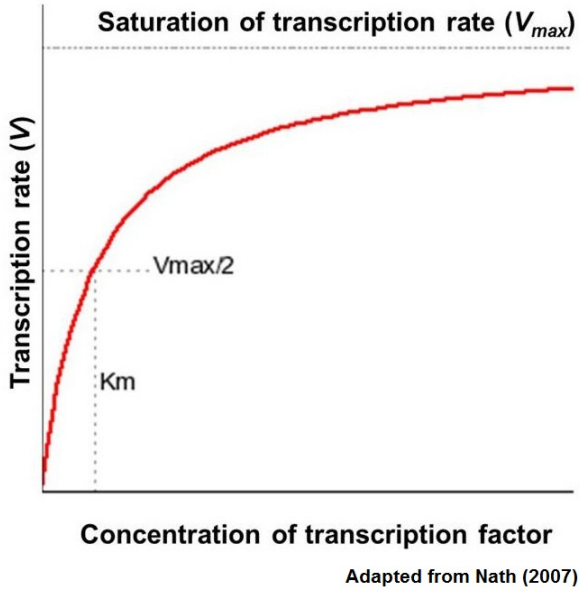

Figure 7. The transcription rate is a function of the transcription factor concentration. K_m is the concentration of the transcription factor at which the transcription rate is at 50% of its maximum (V_{max}) and $V_{max}/2$ is the transcription rate at this half-saturation point.

4.3 Inverse and Forward Problems in Creating the Model

The expression profile of each gene in the GRN is the solution of its representative differential equation 10 in which the production of a gene is dependent upon the regulatory weights w, production rates, and, in the case of the sigmoidal model, the thresholds b. However, the true values of these parameters are not known. Therefore, there are two issues to tackle in modeling the expression patterns of each gene in the GRN: estimating the unknown parameters, which is the inverse problem, and numerically solving the ODE, which is the forward problem.

To address the inverse problem, the unknown parameters are estimated using a method that minimizes the least squares of error. Initially, we know the degradation rates λ for each gene, initial guesses for the production rates, and that production is represented by either the function 12 or 14 depending on whether the model is based on a sigmoid function or Michaelis-Menten kinetics, respectively. We have also experimentally observed the fold change $\hat{x}(t_k)$ for times t_k where $k = 1, ..., n_t$ for a total number of n_t time points. However, we need to find the regulatory weights w, production rates, and, in the case of the sigmoidal model, the thresholds b, all of which are contained in the vector θ. These parameters are found by minimizing the sum of the square of the error between the data and the model $x(t, \theta)$:

$$E = \frac{1}{n_g} \sum_{k=1}^{n_t} |\hat{x}(t_k) - x(t, \theta)|^2 \tag{15}$$

where n_g is the number of genes in the GRN. However, this requires numerically finding the solution to the differential equation model for different values of the parameters until the best fit to the data is discovered. Unfortunately, error estimates using equation 15 are inaccurate in this case due to

the large number of unknown model parameters with respect to those that are known. As a result, equation 15 must be modified as follows to obtain more accurate estimates (Fitzpatrick, 1991):

$$E = \frac{1}{n_g} \sum_{k=1}^{n_t} |\hat{x}(t_k) - x(t, \theta)|^2 + \alpha \|\theta\|^2 \tag{16}$$

The term α is a regularization parameter that represents the ratio of the error in the observed data to the error in the estimates of the model parameters (Fitzpatrick, 1999). The value of α is determined using an L-curve discussed by Hansen & O'Leary (1993). To this end, the parameters of the sigmoidal model and Michaelis-Menten model were estimated for several values of α. The L-curve was created by plotting the least squares of error versus a penalty term for using a regularization parameter in the minimization procedure for each α value. The α value chosen for each model represents the point of maximum curvature in the L-curve, which occurs with $\alpha = 0.01$ for the sigmoidal model and $\alpha = 0.04$ for the Michaelis-Menten model.

Once the parameters θ are known, the forward problem involves finding the numerical solution to the differential equation using these estimated parameters. In addition to the estimated parameters θ, we also know the degradation rates λ and experimentally derived fold changes for each gene. As in the inverse problem, production is defined by either the function 12 or 14 depending on whether the model is based on a sigmoid function or Michaelis-Menten kinetics, respectively. The goal of the forward problem is then to simulate the expression profile, or the fold changes, of a gene over some time course given degradation rates λ, a production function, and estimates of the parameters θ.

4.3.1 Numerical Solution to ODE and Parameter Estimation in MATLAB

The MATLAB simulation to estimate parameters and numerically determine the solution to the differential equation requires an Excel workbook that provides information on each gene and the parameters of the differential equation model to be estimated. The input workbook for both the sigmoidal and Michaelis-Menten models contains sheets with the following names in the following order:

1. The sheet "production_rates" contains the initial guesses of the production rates as described in Section 4.

2. The sheet "degradation_rates" contains the degradation rates computed as described in Section 4.

3. The next five sheets labeled "wt", "dcin5", "dgln3", "dhmo1", and "dzap1" contain all of the log fold change data for each gene in the GRN for each replicate of the t_{15}, t_{30}, and t_{60} cold shock time points collected for the strain denoted by the sheet label.

4. The sheet "concentration_sigmas" contains the standard deviation of the replicate log fold changes for each gene in the GRN for t_{15}, t_{30}, and t_{60} time points.

5. The sheets "network" and "network_weights" contain an adjacency matrix that denotes the nodes in the network and the edges between them. The columns of the matrix correspond to the transcription factors while the rows correspond to the target genes regulated by the

transcription factors. A cell contains a 1 if there is an edge between the corresponding transcription factor and target gene and a 0 if no such edge exists.

6. The sheet "optimization_parameters" lists the optimization parameters for the model, including α, maximum number of iterations of the least square estimation procedure, the time points at which the microarray data was collected, and the strains for which data was collected.

7. The sheet "simulation_times" list the time points at which the solution to the differential equation model will be used to calculate the log fold change of each gene in each network in the forward simulation.

8. The sigmoidal model requires one last sheet labeled "network_b" that contains the initial estimates for the threshold b, which is 0 for each gene.

There are a total of 50 weights and 21 production rates that must be estimated in order to find the solution of the both differential equation models and an additional 21 thresholds that must be estimated for the sigmoidal model. The MATLAB function ODE45 solves the differential equation model using a Runge-Kutta 4th/5th order method (Howard, 2009). The MATLAB function FMINCON compares the parameters of the model to the microarray data to estimate the weights, thresholds, and production rates so as to minimize the sum of the square of the error between the model and the data. The deletion strains are simulated by removing the parameters pertaining to the gene deleted in a given strain. In addition, fits for both models were compared to the data for each strain individually and all strains simultaneously.

There are two types of output generated by these MATLAB simulations. The first is an Excel workbook that contains the following sheets from the input Excel workbook and additional sheets regarding the simulation output in the following order:

1. The sheet "log2_concentrations" contains the log fold change data from the input Excel workbook.

2. The sheet "log2_optimized_concentrations" contains the log fold changes for each gene in the network for each of the simulation times specified in the input Excel workbook calculated by numerically solving the differential equation model.

3. The sheet "degradation_rates" contains the degradation rates computed as described in Section 4.

4. The sheet "production_rates" contains the optimized production rates.

5. The sheet "measurement_times" contains the cold shock experiment time points.

6. The sheet "network" contains the adjacency matrix that denotes the nodes in the network and the edges between them taken from the input Excel workbook.

7. The sheet "network_weights" contains a copy of the adjacency matrix that is in the sheet "network".

8. For the sigmoidal model, a sheet called "network_optimized_b" is created that contains the optimized threshold b for each gene.

9. The last sheet for both the sigmoidal and Michaelis-Menten models is called "network_optimized_weights" that contains the optimized weight of each edge in the gene regulatory network.

The second type of output is a set of plots, one for each gene in the network, of the change in gene expression over time. The weights estimated for the sigmoidal and Michaelis-Menten models are given in Tables 4 and and 7, respectively. Tables 5 and 6 list the estimated thresholds and production rates for the sigmoidal model, respectively. Table 8 lists the estimated production rates for the Michaelis-Menten model.

Table 4. Estimated weights for the sigmoidal model fitted to the data for each strain individually (columns labeled "wt", "*Δcin5*", "*Δgln3*", "*Δhmo1*", and "*Δzap1*") and all strains simultaneously (column labeled "All Strain Simulation").

Transcription Factor	Target Gene	wt	*Δcin5*	*Δgln3*	*Δhmo1*	*Δzap1*	All Strain Simulation
AFT2	AFT2	-0.00877	-0.05508	0.075404	0.02217	-0.32395	-0.16845
CIN5	CIN5	0.716147	0.004214	0.045607	-0.24614	0.507387	0.301127
CIN5	HOT1	-0.2414	0.004214	-0.6092	0.672226	-0.26365	0.136401
CIN5	MSS11	0.154892	0.004214	0.155487	0.408683	0.349581	0.092872
CIN5	PHD1	-0.37312	0.004214	-0.65948	-0.27773	-0.49771	1.687921
CIN5	SMP1	0.251672	0.004214	0.079783	0.582773	0.095493	0.751061
CIN5	YAP6	0.214074	0.004212	-0.32439	0.389309	-0.3483	-0.0802
FHL1	FKH2	0.013375	0.096988	-0.00116	0.003329	-0.00588	0.096254
FHL1	HMO1	0.006775	0.024618	-0.01618	0.004074	-0.00401	0.068037
FHL1	PHD1	-0.06083	0.037335	0.024582	-0.03869	0.011719	0.015823
FHL1	SMP1	0.028294	0.209469	-0.0002	0.049377	0.0036	0.118648
FHL1	YAP6	0.078496	0.07348	-0.01342	0.046826	-0.01981	0.053987
FKH2	ACE2	-0.00496	-0.20882	-1.44399	-0.00228	-0.00761	-0.66093
FKH2	FKH2	-0.03281	-0.45538	-0.00361	0.004162	-0.00522	-0.2313
FKH2	YAP6	-0.05919	-0.1804	0.024815	0.111088	-0.01887	-0.08527
GLN3	GLN3	0.444648	0.00736	-0.00527	0.153193	0.01836	-0.24562
GLN3	MGA2	0.366739	0.003902	-0.00527	0.028937	0.097003	-0.04344
MAL33	GLN3	0.88291	0.008515	-0.00527	0.532947	0.040977	-0.67335
MAL33	SWI4	-0.00032	0.068522	0.007805	-0.05809	-0.10437	-0.41035
MBP1	MAL33	-1.90324	-2.27984	-3.19024	-2.18444	-2.93663	-4.65269
MBP1	MBP1	-0.02696	0.905627	0.14491	0.79267	-0.46727	-0.00641
MBP1	SWI4	-0.00032	-0.0368	0.00086	-0.01302	0.101441	0.403591
PHD1	CIN5	0.03714	0.004214	0.057134	-0.01562	0.1259	0.111884
PHD1	PHD1	-0.00902	-0.05421	0.477094	-0.13926	-0.1035	0.196469
PHD1	SMP1	-0.03454	-0.55841	0.00065	0.127925	0.020694	-2.02619
PHD1	SWI4	-0.00032	0.009254	-0.00137	-0.00427	-0.02335	-0.76448
PHD1	YAP6	0.154412	-0.12389	-0.17831	0.060912	0.452114	-0.26582
SKN7	AFT2	-0.07611	-1.17284	-1.97232	-0.0634	-2.24168	-2.40829
SKN7	HOT1	-0.17191	0.949254	-0.85333	0.483376	-0.50122	-0.734
SKN7	MSS11	0.081352	0.079622	0.143001	0.283612	0.090213	0.47199
SKN7	PHD1	-0.23542	-0.24507	0.556468	-0.1974	-0.23272	-0.98489
SKN7	SKN7	0.551294	-1.31685	-1.42223	-0.08927	-0.40376	-1.60111
SKN7	YAP6	0.24443	0.351758	-0.38091	0.291111	0.088012	0.292308
SKO1	CIN5	0.078855	0.004214	0.005204	-0.14222	0.121368	0.104972
SKO1	MSS11	0.004591	0.027403	0.020401	0.254185	0.048012	0.301096
SKO1	PHD1	-0.04738	0.103938	-0.05489	-0.1857	-0.09869	-0.1757

Transcription Factor	**Target Gene**	**wt**	***Δcin5***	***Δgln3***	***Δhmo1***	***Δzap1***	**All Strain Simulation**
SKO1	YAP6	0.088514	0.25746	-0.08181	0.233024	0.022988	0.230358
SMP1	MAL33	0.4319	1.201528	0.651056	0.096883	-0.26354	3.190825
SMP1	MGA2	-0.06104	-0.00542	0.006294	-0.00433	0.08952	0.008761
SMP1	SMP1	0.002628	-0.52004	-0.00225	-0.00506	0.006064	-0.26353
SWI4	HAP5	-0.00471	0.002088	-0.00455	0.004613	0.001763	-0.00463
SWI4	PHD1	0.020505	0.129184	0.13931	-0.062	0.025855	0.138632
SWI4	SWI4	-0.00032	-0.01529	-0.00837	0.003385	-0.01353	0.02906
SWI6	PHD1	0.012701	0.052966	0.057819	-0.02709	0.02002	0.046726
SWI6	SWI4	-0.00032	-0.00297	-0.00823	0.011363	-0.0079	-0.04107
YAP6	CIN5	-0.1018	0.004214	0.009701	0.04085	-0.01725	-0.0265
YAP6	SWI4	-0.00032	-0.0039	-0.00898	0.031478	-0.0354	-0.08634
YAP6	YAP6	-0.07603	-0.02785	-0.01397	-0.09008	0.040257	-0.03518
ZAP1	ACE2	0.007639	0.564999	-0.88664	-0.00212	0.001901	0.78233
ZAP1	ZAP1	1.215402	-0.93062	0.003594	-0.00157	0.001901	-0.01893

Table 5. Estimated thresholds for the sigmoidal model fitted to the data for each strain individually (columns labeled "wt", "*Δcin5*", "*Δgln3*", "*Δhmo1*", and "*Δzap1*") and all strains simultaneously (column labeled "All Strain Simulation").

Gene	All Strain Simulation	wt	*Δcin5*	*Δgln3*	*Δhmo1*	*Δzap1*
ACE2	0.050995	-4.58982	0.073841	-1.49114	-3.94065	-3.89508
AFT2	-2.79696	-1.85191	-1.49744	-3.10527	-1.3681	-2.90919
CIN5	-0.5637	1.756819	0	-3.07889	-1.79785	0.916647
FHL1	0	0	0	0	0	0
FKH2	-0.69582	-1.0231	0.464833	-2.48184	-2.20056	-2.60308
GLN3	-2.81911	3.194069	-6.12359	0	0.444135	-4.32915
HAP5	0.596735	0.248591	0.212487	0.062199	0.144587	0.127247
HMO1	-2.20512	-3.06954	-3.61275	-4.29233	0	-3.9257
HOT1	-0.32915	-1.26417	3.942528	-2.08177	3.002032	-0.9263
MAL33	-0.86154	-1.33569	0.13316	-2.03215	-1.31223	-3.50084
MBP1	-4.37443	-2.68224	1.773716	-1.80301	1.690395	-2.0736
MGA2	-4.32697	-0.36645	-3.18369	-3.88644	-2.37178	-1.24639
MSS11	1.652348	-0.45911	-2.09822	-0.63189	1.858951	1.266427
PHD1	-1.75485	-2.29138	-0.812	0.992262	-1.87416	-2.31005
SKN7	-2.76501	0.734238	-3.02722	-3.55268	-2.26166	-1.71936
SKO1	0	0	0	0	0	0
SMP1	0.143177	0.135202	0.590446	-1.75784	2.062052	-1.71822
SWI4	0.729422	0.000699	-0.74874	0.043502	-0.68104	-0.03975
SWI6	0	0	0	0	0	0
YAP6	0.885299	1.738062	1.744646	-1.88156	2.739425	0.030512
ZAP1	-1.73455	3.644845	0.077868	0.959476	-0.22933	0

Table 6. Estimated production rates for the sigmoidal model fitted to the data for each strain individually (columns labeled "wt", "*Δcin5*", "*Δgln3*", "*Δhmo1*", and "*Δzap1*") and all strains simultaneously (column labeled "All Strain Simulation").

Gene	All Strain Simulation	wt	*Δcin5*	*Δgln3*	*Δhmo1*	*Δzap1*
ACE2	0.37754	0.229946	0.298348	0.587592	0.198373	0.134069
AFT2	0.381315	0.032721	0.185457	0.416634	0.02326	0.410626
CIN5	0.123548	0.209213	0	0.122351	0.086471	0.231661
FHL1	0.041086	0.055095	0.039742	0.034898	0.040184	0.035761
FKH2	0.030676	0.015607	0.044646	0.016736	0.040336	0.028744
GLN3	0.420632	0.448105	0.309686	0	0.323773	0.31564
HAP5	0	0	0	0	0	0
HMO1	0.088421	0.085089	0.071659	0.074867	0	0.099678
HOT1	0.098783	0.036326	0.095047	0.177306	0.115139	0.140811
MAL33	1.12027	0.592438	0.399222	0.772273	0.465574	0.692213
MBP1	0.060848	0.069774	0.150594	0.067102	0.165443	0.124575
MGA2	0.044411	0.090809	0.040839	0.045174	0.049384	0.039363
MSS11	0.07376	0.03242	0.036676	0.050959	0.103901	0.079065
PHD1	0.083397	0.104198	0.058736	0.193231	0.119114	0.202793
SKN7	0.25718	0.147159	0.336943	0.276154	0.066247	0.1408
SKO1	0.125667	0.106875	0.117309	0.120567	0.137777	0.149346
SMP1	0.125063	0.036112	0.081165	0.026149	0.078091	0.035965
SWI4	0.042632	0	0.01461	0.000523	0.014961	0.006075
SWI6	0.075049	0.055087	0.104441	0.060876	0.085097	0.078015
YAP6	0.054492	0.04374	0.068984	0.071	0.063028	0.0473
ZAP1	0.009598	0.252466	0.088297	0	0.005201	0

Table 7. Estimated weights for the Michaelis-Menten model fitted to the data for each strain individually (columns labeled "wt", "*Δcin5*", "*Δgln3*", "*Δhmo1*", and "*Δzap1*") and all strains simultaneously (column labeled "All Strain Simulation").

Transcription Factor	Target Gene	wt	dCIN5	dGLN3	dHMO1	dZAP1	All Strain Simulation
AFT2	AFT2	0.3457	0.0522	-0.0002	0.3537	0.614	0.5922
CIN5	CIN5	0.4925	0.1926	-0.0001	0.028	0.5351	0.0095
CIN5	HOT1	0.4459	0.1926	0.6238	0.2987	0.0359	0.6765
CIN5	MSS11	0.1672	0.1926	0.4385	0.3251	0.3947	0.3011
CIN5	PHD1	0.0237	0.1926	-0.0431	0.0028	-0.1893	0.7017
CIN5	SMP1	0.0074	0.1926	0.1668	0.2047	0.0381	0.3452
CIN5	YAP6	0.3767	0.1926	-0.0555	0.324	-0.0789	-0.043
FHL1	FKH2	0.3249	0.3113	0.3732	0.4222	0.4242	0.194
FHL1	HMO1	0.5757	0.7125	0.5187	0.0086	0.6509	0.7623
FHL1	PHD1	-0.0005	0.4946	0.0001	0.4012	0.0006	0.0175
FHL1	SMP1	0.3662	0.4362	0.0078	0.1319	0.3919	0.2987
FHL1	YAP6	-0.0029	0.2691	0.4158	0.0179	0.0249	0.0961
FKH2	ACE2	-0.0002	0.5063	0.0001	0.8516	0.6891	0.3994
FKH2	FKH2	0.2726	-0.3527	-0.0216	0.0066	-0.0034	0.0474
FKH2	YAP6	0.0461	0.1705	-0.0003	0.0395	0.0015	0.012
GLN3	GLN3	-0.0003	0.3881	0.0469	0.0005	0.0036	0.0043
GLN3	MGA2	0.5349	0.47	0.0154	0.5033	0.3892	0.4389
MAL33	GLN3	0.5285	0.976	0.0154	0.6432	0.7924	0.6819
MAL33	SWI4	0.0091	0.3537	0.0658	0.0213	0.1295	0.2633
MBP1	MAL33	-0.0516	0.0996	0.6298	0.0239	-0.3858	-0.7424
MBP1	MBP1	0.5095	1.0242	0.4061	0.4804	0.4407	0.5999
MBP1	SWI4	-0.0081	0.1746	-0.0719	0.0127	0.0388	0.0033
PHD1	CIN5	0.0024	0.1926	0.6405	0.008	-0.0046	0.1371
PHD1	PHD1	-0.0025	0.2518	0.0001	0.0015	-0.0044	-0.0131
PHD1	SMP1	0.1894	0.1306	0.4295	0.1328	0.0296	0.0409
PHD1	SWI4	0.0283	0.1399	0.0168	0.3099	0.0189	-0.2297
PHD1	YAP6	-0.0155	0.1368	0.4148	0.0377	0.3627	-0.0903
SKN7	AFT2	0	0.7671	0.8166	0.041	0.5516	-0.1975
SKN7	HOT1	0	0.2296	0.0002	-0.0099	0.5596	0.3214
SKN7	MSS11	-0.0042	0.1641	0.0068	0.0167	0.1284	0.2775
SKN7	PHD1	0.0125	0.3444	-0.0002	0.0237	-0.0151	-0.4114
SKN7	SKN7	0.525	1.3963	1.1862	0.5008	0.717	1.3568
SKN7	YAP6	0.1159	0.3962	0.017	0.1071	-0.0029	0.0502
SKO1	CIN5	-0.004	0.1926	0.2902	0.4393	0.0012	0.7614
SKO1	MSS11	0.3971	0.6862	0.0001	-0.014	0.0082	0.2951

Transcription Factor	**Target Gene**	**wt**	***Δcin5***	***Δgln3***	***Δhmo1***	***Δzap1***	**All Strain Simulation**
SKO1	PHD1	0.4909	0.3242	-0.0001	-0.0549	0	0.0323
SKO1	YAP6	-0.0208	0.2038	0.0377	0.0274	0.0108	0.3819
SMP1	MAL33	0.7581	0.7407	0	0.5828	0.7187	0.9101
SMP1	MGA2	0.001	0.7058	0.4752	0.0031	0.0075	0.5587
SMP1	SMP1	0.0002	0.0858	0.0001	0.0146	0.0007	0.0086
SWI4	HAP5	0.0045	0	0.0327	0.0265	0	0
SWI4	PHD1	0.5476	0.3286	0.6687	-0.0046	0.7597	0.722
SWI4	SWI4	0.127	0.1596	0.0283	-0.031	0.0319	-0.0304
SWI6	PHD1	0.1106	0.2139	0	0.45	0.0002	0.0185
SWI6	SWI4	-0.0825	0.1713	0.0784	0.0456	0.0377	-0.0313
YAP6	CIN5	0.0213	0.1926	0.0094	0.0178	-0.0052	0.6335
YAP6	SWI4	-0.0346	0.1594	0.0598	0.0659	0.0236	-0.0286
YAP6	YAP6	0.0158	0.1437	-0.033	0.0469	0.0467	-0.0071
ZAP1	ACE2	0.8569	0.9122	0.9757	0.192	0.0105	0.8175
ZAP1	ZAP1	0.2363	0.3564	0.1508	0.1816	0.0105	0.1518

Table 8. Estimated production rates for the Michaelis-Menten model fitted to the data for each strain individually columns labeled "wt", "*Δcin5*", "*Δgln3*", "*Δhmo1*", and "*Δzap1*") and all strains simultaneously (column labeled "All Strain Simulation").

Gene	All Strain Simulation	wt	*Δcin5*	*Δgln3*	*Δhmo1*	*Δzap1*
ACE2	0.5106	0.3497	0.4168	0.4896	0.4293	0.3152
AFT2	0.1153	0.0602	0.0385	0.1077	0.0779	0.0884
CIN5	0.2377	0.1724	0.0014	0.2821	0.1935	0.2804
FHL1	0.0213	0.0551	0.0253	0.0179	0.02	0.0188
FKH2	0.1336	0.0215	0.1975	0.0639	0.1178	0.0882
GLN3	0.457	0.3503	0.5324	0	0.3802	0.4939
HAP5	0.0002	0	0.0151	0	0	0.0013
HMO1	0.1778	0.141	0.1695	0.209	0	0.2366
HOT1	0.0536	0.0572	0.0827	0.0453	0.0893	0.0803
MAL33	0.4916	0.6522	0.2348	0.3315	0.2857	0.37
MBP1	0.1247	0.1297	0.0732	0.1175	0.1303	0.1973
MGA2	0.13	0.1052	0.1548	0.142	0.1426	0.0972
MSS11	0.102	0.0377	0.0933	0.0807	0.1218	0.0892
PHD1	0.2036	0.1464	0.1486	0.2426	0.2819	0.4074
SKN7	0.1252	0.1158	0.1337	0.1418	0.1352	0.1516
SKO1	0.0628	0.1065	0.0583	0.06	0.068	0.0734
SMP1	0.0594	0.0352	0.1041	0.069	0.1042	0.1064
SWI4	0.0179	0	0.1	0	0.0475	0.0129
SWI6	0.0381	0.0551	0.053	0.0304	0.0425	0.0388
YAP6	0.0971	0.0266	0.0792	0.1044	0.0641	0.0892
ZAP1	0.0546	0.0632	0.0867	0	0.0169	0

4.4 Comparison of the Sigmoidal and Michaelis-Menten Model Between the Wild Type and Deletion Strains

In looking at the results of both the sigmoidal and Michaelis-Menten models, there are distinct differences between the expression profiles between the wild type and a deletion strain for the genes that are regulated by the deleted transcription factor. This is seen for the genes *PHD1*, *HOT1*, and *SMP1* which have different expression profiles between the wild type and the $\Delta cin5$ strains (Figure 8). Since *Cin5* is one of the transcription factors regulating the expression of these genes, it is expected that the deletion of *CIN5* would affect the expression of *PHD1*, *HOT1*, and *SMP1*. Interestingly, the difference in the expression profile between the wildtype and $\Delta cin5$ strains is exacerabted for *SMP1* (Figure 8). However, for some genes and deletion strains, it is unclear whether the Michaelis-Menten model or sigmoidal model better fit the data. For instance, the deletion of *GLN3* does not significantly alter the expression of *MGA2* compared to the wild type strain even though *GLN3* regulates *MGA2* (Figure S18b and S20b). For those deletion strains where the deleted transcription factor does not control a given gene, the expression profile for that gene is not different between the wild type strain and that deletion strain, which is expected.

Some interesting expression patterns emerge for some genes between the strains for the sigmoidal and Michaelis-Menten models. In the sigmoidal model, the production rate for *HAP5* is 0 for simulations run with all data simultaneously and the data for each strain individually (Table 6). This would suggest that *HAP5* is strongly repressed. In fact, this hypothesis is supported by the majority of the simulations for the sigmoidal and Michaelis-Menten models for different strains, which estimate a negative regulatory weight of the *Swi4* transcription factor on *HAP5* (Tables 4 and 7). However, in all simulations, this regulatory effect is very small, which suggests that there may be other transcription factors that regulate *HAP5* that are not in our model.

FHL1, *SWI6*, and *SKO1* are three genes in the GRN that do not have any inputs and, as a result, the parameters estimated for these genes are similar. For the sigmoidal model, the threshold for each of these genes is 0 (Table 5). Furthermore, neither *FHL1* nor *SWI6* seem to be expressed according to the sigmoidal and Michaelis-Menten models (Figures S18a, S18d, S19a, S19d, S20a, S20d, S21a, S21d, S22a, & S22d).

For some genes, the sigmoidal and Michaelis-Menten models better fit the log fold change data for a gene when using the microarray data for all strains rather than individual strains in the simulation. This is evident in the expression profile for *SWI6*, which is centered at 0 for both the sigmoid and Michaelis-Menten models fit with the data from all strains. This is expected for *SWI6* as this gene is not regulated by any of the transcription factors in the proposed GRN (Figure 9). Interestingly, the sigmoidal and Michaelis-Menten models of expression for a gene are similar when all of the data is used. In contrast, these models produce different expression profiles when data for only one strain is used (Figure 9).

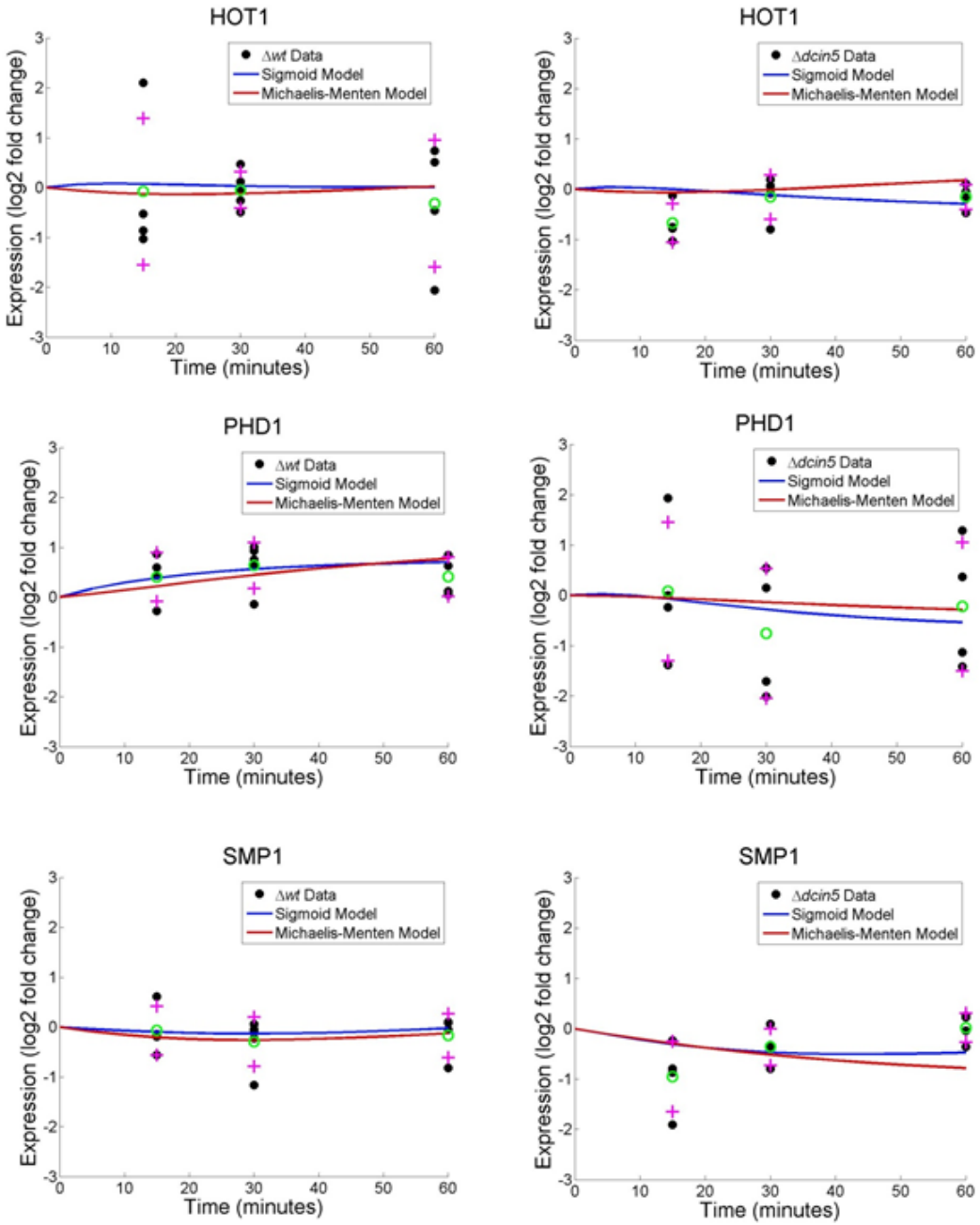

Figure 8. The wild type and $\Delta cin5$ forward simulations for *HOT1*, *PHD1*, and *SMP1*. The models fitted by both the sigmoid function (blue line) and Michaelis-Menten function (red line) are shown. The green circles designate the average log fold change for each time point in the cold shock experiment. The magenta pluses designate ± 1 standard deviation in the log fold change for each time point in the cold shock experiment.

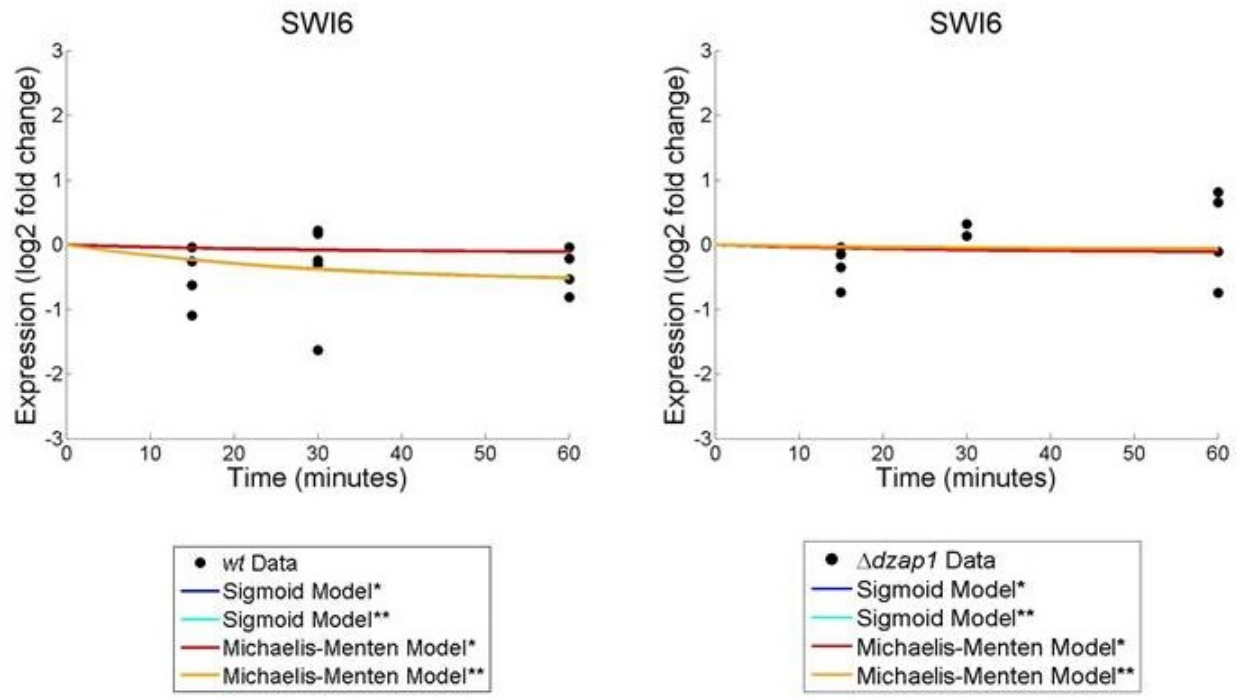

Figure 9. The wild type and $\Delta zap1$ forward simulations for SWI6 based on Michaelis-Menten kinetics and the sigmoid function. The models were fit using data for all strains (*) and for individual strains (**) in the simulation.

5 Discussion and Future Work

Mathematical models of gene regulation can be used to elucidate the dynamics of environmental stress responses in organisms. The goal of this research has been to determine a dynamic model of transcriptional regulation in the cold shock response in budding yeast to determine the relative influence of transcription factors that regulate this response.

The nonlinear differential equation model presented in this paper was developed using DNA microarray data collected for the wild type and four deletion strains in a cold shock experiment. Loess normalization and median absolute deviation scaling was implemented in the R statistical environment using the limma package to mitigate dye and spatial biases that affect the data. I showed that these normalization methods mitigated dye intensity and spatial biases in the microarray data. Statistical analysis was used to determine if any genes had a significant change in gene expression in at least one time point in the cold shock experiment and if any genes in a pair of strains have similar gene expression profiles. Finally, the MATLAB ODE45 and FMINCON functions were used to solve the differential equation model describing gene expression as a dynamic balance between production and degradation by minimizing the sum of the square of the error between the model and the data. As a result, I was able to successfully optimize the model parameters in order to obtain the best possible fits to the data.

There is still work to be done to validate the current models and to improve the fits of the models to the data. One problem with the current models arises from the fact that there is little data to allow for a good model fit. There are 92 parameters to estimate for the sigmoidal model and 71 parameters to estimation for the Michaelis-Menten model, but only four data points for each cold shock time point for each gene. A solution to this problem is to find a way to incorporate the recovery phase data. However, this involves investigation into whether the dynamics of recovery are similar or different from those of cold shock. Since there is a temperature difference between the two phases, it is expected that the production and degradation rates will be different. In addition, it would be beneficial to validate the current models both experimentally and by performing sensitivity analysis to determine what parameters if perturbed have the greatest effect on the output. Performing sensitivity analysis for both the sigmoidal and Michaelis-Menten models may yield insightful information in further comparisons of the models.

Currently, the models presented in this paper are weakened by the fact that the proposed GRN may not be the complete network involved in the cold shock response as genome-wide location analysis has not been performed under cold-shock conditions. Future directions in this research could include performing these experiments as well as investigating methods by which to infer the GRN from the microarray data available. The work presented here provides a solid foundation for these future directions in elucidating the dynamics of the cold shock response in budding yeast.

6 Acknowledgments

I would like to thank Dr. Kam D. Dahlquist and Dr. Ben G. Fitpatrick for all of the guidance they have given me through out this research project, including their input in writing this thesis. I would like to specifically acknowledge Dr. Dahlquist for providing the data to normalize, the current regulatory network, and other biological background information regarding the experiment. Furthermore, I would like to specifically acknowledge Dr. Fitzpatrick for all of his assistance with R and MATLAB code, understanding the deterministic model and the statistical tests, and all other mathematical background on the research.

This research would also have not been possible without the work of several former Loyola Marymount University students. Specifically, I would like to thank Nicholas A. Rohacz for all of his contributions to creating the microarray normalization protocol and both the sigmoidal and Michaelis-Menten differential equation models. In addition, I would like to thank Andrew Herman and Katie Hornick for the laboratory work they conducted for this research as well as past lab members for contributing the following data:

- wild type: Wesley Citti, Elizabeth Liu, Heather King, Matt Meijia, Olivia Sakhon
- $\Delta cin5$: Kevin Entzminger, Stephanie Kuelbs, Kenny Rodriguez
- $Deltagln3$: Kenny Rodriguez, Andrew Herman, Alondra Vega
- $Deltahmo1$: Cybele Arsan, Andrew Herman, Alondra Vega
- $Deltazp1$: Andrew Herman, Stephanie Kuelbs, Lauren Kubeck, Kenny Rodriguez

This work was made possible by the NSF-RUI grant 0921038.

7 References

Aguilera, J., Randez-Gil, F., & Prieto, J.A. (2007). Cold response in *Saccharomyces cerevisiae*: new functions for old mechanisms. *FEMS Microbiology Reviews*, *31*(3), 327-341. doi:10.1111/j.1574-6976.2007.00066.x.

Al-Fageeh, M.B. & Smales, C.M. (2006). Control and regulation of the cellular response to cold shock: the response in yeast and mammalian systems. *Biochemical Journal*, *397*, 247-259. doi: 10.1042/BJ20060166.

Anderle, P., Duval, M., Draghici, S., Kuklin, A., Littlejohn, T.G., Medrano, J.F., Vilanova, D., & Roberts, M.A. (2003). Gene Expression Databases and Data Mining. *BioTechniques*, *34*, S36-S44.

Ball, C.A. & Sherlock, G. (2006). Microarray Data: Annovation, Stoarge, Retrieval, and Communication. In A. Aluru (Ed.), *Handbook of Computational Molecular Biology*, Boca Raton, Florida: Chapman & Hall/CRC, pp. 23-1 - 23-10

Belle, A., Tanay, A., Bitincka, L., Shamir, R., & O'Shea, E.K. (2006). Quantification of protein half-lives in the budding yeast proteome. *Proceedings of the National Academy of Sciences*, *103*(35), 130004-130009.

Benjamini, Y. & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289-300. doi: 10.1073/pnas.0605420103.

Campbell, A.M. (2009). GCAT-Chip. <http://www.bio.davidson.edu/GCAT/GCATchip.html>. Accessed 10 April 2014.

Chen, K. & Rajewsky, N. (2007). The evolution of gene regulation by transcription factors and microRNAs. *Nature Review Genetics*, *8*(2), 93-103. doi:10.1038/nrg1990.

Chou, H. (2006). Computational Methods for Microarray Design. In A. Aluru (Ed.), *Handbook of Computational Molecular Biology*, Boca Raton, Florida: Chapman & Hall/CRC, pp. 24-1 - 24-22.

Dudoit, S., Shaffer, J.P., & Boldrick, J.C. (2003). Multiple Hypothesis Testing in Microarray Experiments. *Statistical Science*, *18*(1), 71-103. doi:10.1214/ss/1056397487.

Filkov, V. (2006). Identifying Gene Regulatory Networks from Gene Expression Data. In A. Aluru (Ed.), *Handbook of Computational Molecular Biology*, Boca Raton, Florida: Chapman & Hall/CRC, 27-1 - 27-29.

Fitzpatrick, B.G. (1991). Bayesian analysis in inverse problems. *Inverse Problems*, *7*, 675-702.

Gibson, M. A. & Mjolsness, E. (2001). *Computational modeling of genetic and biochemical networks*. In J. M. Bower & H. Bolouri (Eds.), Cambridge, Massachusetts: The MIT Press, pp. 1-48.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettlin, T., & Oliver, S. G. (1996). Life with 6000 genes. *Science*, 546-567.

Graybill, F. A. (1976). *Theory and Application of the Linear Model*, Duxbury, North Scituate.

Hansen, P.C. & O'Leary, D.P. (1993). The Use of the L-Curve in the Regularization of Discrete Ill-Posed Problems. *SIAM Journal on Scientific Computing*, *14*(6), 1487-1503.

Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N.J., Macisaac, K.D., Danford, T.W., Hannett, N.M., Tagne, J., Reynolds, D.B., Yoo, J., Jennings, E.G., Zeitlinger, J., Pokholok, D.K., Kellis, M., Rolfe, P.A., Takusagawa, K.T., Lander, E.S., Gifford, D.K., Fraenkel, E., & Young, R.A. (2004). Transcriptional regulatory code of a eukaryotic genome. *Nature*, *431*, 99-104. doi:10.1038/nature02886.

Hernández-López, M.J., García-Marqués, S., Randez-Gil, F., & Prieto, J.A. (2011). Multicopy Suppression Screening of *Saccharomyces cerevisiae* Identifies the Ubiquitination Machinery as a Main Target for Improving Growth at Low Temperatures. *Applied and Environmental Microbiology*, *77*(21), 7517-7525. doi:10.1128/AEM.00404-11.

Howard, P. (2009) Solving ODE in MATLAB. <http://www.math.tamu.edu/ phoward/m401/matode.pdf>. Accessed 19 April 2014.

Khanin, R., Vinciotti, V., Mersinias, V., Smith, C. P., & Wit, E. (2007). Statistical Reconstruction of Transcription Factor Activity Using Michaelis-Menten Kinetics. *Biometrics*, *63*(3), 816-823. doi:10.1111/j.1541-0420.2007.00757.x.

Larson, M.G. (2008). Statistical Primer for Cardiovascular Research. *Circulation*, *117*, 115-121. doi: 10.1161/CIRCULATIONAHA.107.654335

Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson1, C.M., Simon, I., Zeitlinger, J., Jennings, E.G., Murray, H.L., Gordon, D.B., Ren1, B., Wyrick, J.J., Tagne, J., Volkert, T.L., Fraenkel, E., Gifford, D.K., Young, R.A. (2002). Transcriptional Regulatory Networks in *Saccharomyces cerevisiae*. *Science*, textit298(5594), 799-804. doi: 10.1126/science.1075090.

Lemon, B. & Tjian, R. (2000). Orchestrated response: a symphony of transcription factors for gene control. *Genes & Development*, *14*, 2551-2569. doi:10.1101/gad.831000.

Lodish, H., Book, A., Zipursky, S.L., Matsudaira, P., Baltimore, D., & Darnell, J. (2000).

The Dyanmic Cell. In H. Lodish, A. Book, S.L. Zipursky, P. Matsudaira, D. Baltimore, & J. Darnelle (Eds.), *Molecular Cell Biology*, 4th Edition, New York: W.H. Freeman. <http://www.ncbi.nlm.nih.gov/books/NBK21475/>

Nath, A. (2007). Michaelis-Menten Kintecits and Briggs-Haldane Kinetics. <http://depts.washington.edu/wmatkins/kinetics/michaelis-menten.html>. Accessed 19 April 2014.

Quackenbush, J. (2002). Microarray data normalization and transformation. *Nature Genetics*, *32*, 496-501. doi:10.1038/ng1032.

R Development Core Team (2005). R: A language and environment for statistical computing, reference index version 2.13.0. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

Robeva, R. & Yildirim, N. (2013). Bistability in the Lactose Operaon of *Escherichia coli*: A Comparison of Differential Equation and Boolean Network Models. In R. Robeva & T. Hodge (Eds.), *Mathematical Concepts and Methods in Modern Biology*, London, UK: Academic Press, pp. 37-74.

Sabanayagam, C.R. & Lakowicz, J.R. (2007). Increasing the sensitivity of DNA microarrays by metal-enhanced fluorescence using surface-bound silver nanoparticles. *Nucleic Acids Research*, *35*(2), e13. doi: 10.1093/nar/gkl1054.

Schade, B., Jansen, G., Whiteway, M., Entian, K.D., & Thomas, D.Y. (2004). Cold Adaptation in Budding Yeast. *Molecular Biology of the Cell*, *15*(12), 5492-5502. doi: 10.1091/mbc.E04-03-0167.

Sherbina, K. (2014a) Dahlquist:Microarray Data Processing in R. <http://openwetware.org/wiki/ Dahlquist:Microarray_Data_Processing_in_R. Accessed 15 April 2014.

Sherbina, K. (2014b). Dahlquist:Modified ANOVA and p value Corrections for Microarray Data. Retrieved from http://openwetware.org/wiki/ Dahlquist:Microarray_Data_Processing_in_R. Accessed 4 May 2014.

Smyth, G. K. (2005). Limma: linear models for microarray data. *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber (eds.), Springer, New York, 397-420.

Strassburg, K., Walther, D., Takahashi, H., Kanaya, S., & Kopka, J. (2010). Dynamic Transcriptional and Metabolic Responses in Yeast Adapting to Temperature Stress. *OMICS*, *14*(3), 249-259. doi: 10.1089ōmi.2009.0107.

Tanay, A., Sharan, R. & Shamir, R. (2006). Biclustering Algorithms: A Survey. In A. Aluru (Ed.), *Handbook of Computational Molecular Biology*, Boca Raton, Florida: Chapman & Hall/CRC, pp. 26-1 - 26-17.

Teixeira, M.C., Monteiro, P.T., Guerreiro, J.F., Gonçalves, J.P, Mira, N.P., dos Santos, S.C., Cabrito, T.R., Palma, M., Costa, C., Francisco, A.P., Madeira, S.C., Oliveira, A.L., Freitas, A.T., & Sá-Correia, A. (2014). The YEASTRACT database: an upgraded information system for the analysis of gene and genomic transcription regulation in *Saccharomyces cerevisiae*. *Nucleic Acids Research*, *42*(D1), D161-D166. doi: 10.1093/nar/gkt1015. Thieringer, H. A., Jones, P. G., & Inouye, M. (1998). Cold shock and adaptation. *Bioessays*, *20*(1), 49-57. doi:10.1002/(SICI)1521-1878(199801)20:1<49::AID-BIES8>3.0.CO;2-N.

UNH Microarray Centre. (2012). Microarray: Types. <http://www.microarrays.ca/products/microarrays_types.html>. Accessed 18 April 2014.

Vu, T. T., & Vohradsky, J. (2007). Nonlinear differential equation model for quantification of transcriptional regulation applied to microarray data of *Saccharomyces cerevisiae*. *Nucleic Acids Research*, *35*(1), 279-287. doi:10.1093/nar/gkl1001.

Wilczynski, B. & Furlong, E.E.M. (2010). Challenges for modeling global gene regulatory networks during development: Insights from *Drosophila*. *Developmental Biology*, *340*, 161-169. doi:10.1016/j.ydbio.2009.10.032.

Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J.D., Bussey, H., Chu, A.M., Connelly, C., Davis, K., Dietrich, F., Dow, S.W., Bakkoury, M.E., Foury, F., Friend, S.H., Gentalen, E., Giaever, G., Hegemann, J.H., Jones, T., Laub, M., Liao, H., Liebundguth, N., Lockhart, D.J., Lucau-Danila, A., Lussier, M., M'Rabet, N., Menard, P., Mittmann, M., Pai, C., Rebischung, C., Revuelta, J.L., Riles, L., Roberts, C.J., Ross-MacDonald, P., Scherens, B., Snyder, M., Sookhai-Mahadeo, S., Storms, R.K., Véronneau, S., Voet, M., Volckaert, G., Ward, T.R., Wysocki, R., Yen, G.S., Yu, K., Zimmermann, K., Philippsen, P., Johnston, M., & Davis, R.W. (1999). Functional Characterization of the *S. cerevisiae* Genome by Gene Deletion and Parallel Analysis. *Science*, *285*, 901-906. doi:10.1126/science.285.5429.901.

Yang, Y., Dudoit, S., Luu, P., Lin, D. M., Peng, V., Ngai, J. , & Speed, T. P. (2002). Normalization for cDNNA microarray data: A robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Research*, 30(4), e15. doi:10.1093/nar/30.4.e15