

Mathematical Modeling of Six Database-Derived Gene Regulatory Networks Identifies Key Regulators and Network Properties Controlling the Early Response to Cold Shock in *Saccharomyces cerevisiae*

Brandon J. Klein

Department of Biology, Loyola Marymount University, 1 LMU Drive, Los Angeles, CA 90045

ABSTRACT:

A gene regulatory network (GRN) is a group of transcription factors that control the level of expression of genes encoding other transcription factors. Dynamics of GRNs illustrate how expression in the network changes over time. GRNmap, a MATLAB software package, uses differential equations to model the dynamics of medium-scale GRNs. The software estimates production rates, expression thresholds, and regulatory weights for each transcription factor in the network based on microarray data. Microarray data was obtained from a cold shock experiment where wild-type budding yeast, *Saccharomyces cerevisiae*, and five strains from which the transcription factors Cin5, Gln3, Hap4, Hmo1, and Zap1 were deleted were subjected to cold shock at 13°C for 15, 30, and 60 minutes. Six related GRNs, which ranged from 14-17 genes and 25-35 edges, were constructed using data from the YEASTRACT database. GRNmap was then used to estimate production rates, expression thresholds, and regulatory weights for each of these GRNs. Forward simulation of the model showed a good fit to the experimental data, particularly in comparison to 30 random networks with the same genes and number of edges. Systematic analysis of edges identified repeated motifs in the database-derived networks. These motifs included an early coherent type I feed forward loop (FFL), an activating regulatory chain, and symmetrical incoherent type I FFLs terminating on the paralogs Yhp1 and Yox1. Multiple regression analysis of the database-derived and random networks demonstrated that model fit to experimental expression data was repeatedly correlated with the eigenvector

centrality of individual transcription factors. Better model fit to expression data was observed for transcription factors with high eigenvector centrality, such as Gcn4 and Yhp1. A consolidated network containing 15 genes and 34 edges was constructed from conserved motifs and high eigenvector centrality transcription factors. The consolidated network outperformed all random networks and all but one database-derived network. The consolidated network also featured significant overlap with the general environmental stress response, with early activation of Hmo1 inducing Msn2/Msn4 directly and Yhp1 indirectly.

Keywords: *Saccharomyces cerevisiae*, cold shock, gene regulatory network, transcription factors, dynamical systems modeling, eigenvector centrality, network motifs

INTRODUCTION:

All organisms require specific internal environmental conditions to optimize cellular functions, which must be maintained in the face of external perturbations. Sources of environmental stressors are ubiquitous and often diurnal in any ecological niche, encompassing fluctuations in nutrient or oxygen availability, temperature, salinity, osmolarity, and acidity (Wingfield, 2013). For unicellular organisms such as the budding yeast, *Saccharomyces cerevisiae*, responding to environmental stressors depends on the activation of regulatory programs that dynamically modulate gene expression. The exact mechanisms involved in the regulation of gene expression are still emerging through the study of gene regulatory networks (Karlebach and Shamir, 2008), post-translational modifications (Youn et al., 2017), and RNA-mediated epigenetic regulation (Holoch and Moazed, 2015).

Saccharomyces cerevisiae is an ideal eukaryotic model organism for systems biology research. Molecular genetic tools and datasets are widely available to aid in the study of *S.*

cerevisiae, including the yeast model organism database, SGD (Cherry et al., 2011), and the YEASTRACT transcriptional regulator database (Teixeira et al., 2017), to name a few. Further, the transcriptomic responses to a myriad of stress responses have been investigated in *S. cerevisiae*. Microarray experiments assessing the response of yeast to eleven environmental stressors defined a group of approximately 900 genes that consistently show similar expression profiles, referred to collectively as the general environmental stress response, or ESR (Gasch et al., 2000). The ESR is dependent on the transcription factors Msn2/Msn4 and Yhp1. Although the ESR is commonly induced, individual stressors have also been shown to cause unique transcriptional responses to occur in the cell.

Temperature shock is a universal environmental stressor that is easy to investigate in the lab. In *Saccharomyces cerevisiae*, the response to heat shock has been extensively characterized. The response to heat shock is part of the ESR. A detailed mechanistic understanding of the conserved heat shock proteins governing this process has emerged (Ingolia et al., 1982; Morano et al., 2012; Verghese et al., 2012). However, the transcriptional response to cold shock in yeast has received less attention, despite early documentation of cold-sensitive phenotypes in mutant yeast strains (Hampsey, 1997).

The molecular consequences of low temperature include decreased membrane fluidity, impairment of translation due to stabilization of RNA secondary structure, and reduction of enzymatic activities with no observed induction of conserved cold shock proteins (Al-Fageeh and Smales, 2006). Microarray experiments have demonstrated that approximately 25% of the genome is involved in the transcriptional response to low temperatures. Nearly one-third of all up-regulated genes encode ribosomal proteins (Sahara et al., 2002), suggesting that cold adaptation necessitates *de novo* ribosome biogenesis (Aguilera et al., 2007). A subsequent study

by Kandror et al. (2004) established that the response of *S. cerevisiae* to near-freezing temperatures between 0-10°C was Msn2/Msn4-dependent and resembles the ESR. Similarly, cold shock microarray experiments performed by Schade et al. (2004) distinguished an early cold shock response occurring within 2 hours of cold temperature exposure from a late cold response occurring after 12 hours that resembles the ESR. Although Schade et al. (2004) reported that the late cold response is Msn2/Msn4-dependent, no canonical transcription factors regulating the early response to cold shock have been characterized.

Understanding the regulatory mechanisms governing the genome-wide response to an environmental perturbation like cold shock requires identification not only of the transcription factors directly regulating genes with altered expression, but also of the factors regulating those transcription factors themselves. Together, networks of transcription factors controlling the expression of target sets of genes form gene regulatory networks (GRNs). Extensive research has aimed to empirically determine the topology of GRNs in *Saccharomyces cerevisiae*. Genome-wide location analyses, in which epitope-tagged transcription factors are isolated via chromatin immunoprecipitation and then hybridized to DNA microarrays (ChIP-chip), have been employed to experimentally determine the regulatory relationships between transcription factors and their targets in yeast (Lee et al., 2002; Harbison et al., 2004). Lee et al. (2002) gathered genome-wide location data for 106 transcription factors and systematically identified gene regulatory network motifs, which represent consistent patterns of interconnection that are found within complex biological networks (Milo et al., 2002). The specific motifs identified were autoregulation, multi-component loops, feedforward loops, single-input motifs, multi-input motifs, and regulator chains. Lee et al. (2002) assembled the identified motifs into a global GRN spanning 106 transcription factors. An updated study by Harbison et al. (2004) documented genome-wide

location data for an expanded set of 203 transcription factors in *S. cerevisiae*, including under various environmental conditions.

As comprehensive models of gene regulatory network topology have emerged, graph statistical analysis tools have been applied to quantitatively analyze the connectivity of GRNs (Pavlopoulos et al., 2011). Investigators have shown particular interest in isolating core transcription factors from large GRNs by computing different centrality metrics (Koschützki and Shreiber, 2008; Özgür et al., 2008; Narang et al., 2015). One common measure of centrality is betweenness, which indicates the degree to which a node is present along the shortest paths connecting other nodes within a network (Freeman, 1977). Nodes with high betweenness centrality are thought to exercise high degrees of control on communication throughout the network. An alternative measure is eigenvector centrality, which considers the influence of a node on network dynamics based on the centrality of those nodes to which it is connected (Bonacich and Lloyd, 2001). A study by Özgür et al. (2008) reported that mutations in transcription factors with high eigenvector centrality have a heightened probability of causing disease, suggesting the value of this metric in identifying high influence transcription factors within GRNs.

Following establishment of network topology, gene regulatory network dynamics can be mathematically modeled. GRN dynamics include the weights of the regulatory relationships connecting transcription factors, the production of those transcription factors, and the expression thresholds for the genes encoding transcription factors in the network. Previous authors have approached modeling the non-linear dynamics of gene expression within GRNs (Wang et al., 2009) using mass balance ordinary differential equation structures (Chen et al., 2005; Vu and Vohradsky, 2007). The Dahlquist Lab has applied this technique to the inverse problem of

estimating gene regulatory network parameters from time course gene expression data obtained from DNA microarray experiments. The estimation of high-dimensional parameters from noisy, temporally sparse microarray data has presented various mathematical and computational challenges.

To address these issues, the Dahlquist Lab has developed GRNmap, an open-source MATLAB software package that employs mass balance ordinary differential equations to estimate network parameters for small to medium-scale GRNs from microarray data (Dahlquist et al., 2015). GRNmap achieves efficient optimization of high-dimensional parameters by employing a penalized nonlinear least squares approach, which has generally been avoided in previous models of GRN dynamics due to its mathematical complexity. A key advantage of the penalized least squares approach to optimization over alternatives such as extended Kalman filtering (Lillacci and Khammash, 2010) is its capacity to estimate parameters from temporally sparse time course microarray data. Scaling of the penalty term to ensure efficient regularization is empirically determined based on the results of L-curve analysis, which compares the size of regularized solutions with their associated residuals (Hansen and O’Leary, 1993). During GRNmap model runs, estimated parameters can include transcription factor production rates, thresholds of expression, and regulatory weights. The weighted regulatory relationships connecting the individual transcription factors in our models can be represented either as activation or repression. To visualize the topology of both unweighted gene regulatory networks and the weighted GRNs modeled by GRNmap, we have also developed a companion web application, GRNsight. GRNsight automatically lays out networks that are input as adjacency matrices, allowing for the rapid visualization and analysis of GRN connectivity (Dahlquist et al., 2016).

The goal of this work was to identify and mathematically model the medium-scale gene regulatory network responsible for controlling the early response to cold shock in *Saccharomyces cerevisiae*. GRNmap and GRNsight were used to model and comparatively analyze a family of six related GRNs. The six networks were derived from microarray experiments, which subjected wild-type yeast and five transcription factor deletion mutants to cold shock at 13°C, and using data from the YEASTRACT database. By analyzing these networks, we aimed to determine which transcription factors control the early response to cold shock, what their relative levels of influence are in the GRNs, what motifs are formed, and how those motifs combine to form an overall gene regulatory network governing the cold shock response.

MATERIALS AND METHODS:

Statistical Analysis of DNA Microarray Data

Raw DNA microarray data was obtained from previous cold shock experiments performed by the Dahlquist Lab in which wild-type *Saccharomyces cerevisiae* and the transcription factor deletion strains $\Delta cin5$, $\Delta gln3$, $\Delta hap4$, $\Delta hmo1$, and $\Delta zap1$ were subjected to cold shock at 13°C for 15, 30, and 60 minutes. Expression data was available for 6189 genes. The complete expression data set is publicly available: https://github.com/kdahlquist/DahlquistLab/blob/master/GEO_submission/Dahlquist_GA_dual_ch_w_platf_20160613.xls?raw=true.

Prior to statistical analysis, within- and between-chip normalization of the raw data was performed in the open-source statistical software R (version 3.1.0, R Foundation for Statistical Computing, Vienna, Austria) using the limma v3.20.1 normalization package. The normalized expression data for each strain was imported into Microsoft Excel, and a modified ANOVA was performed to identify genes with a significant expression change at any time point. To limit the

false discovery rate associated with multiple testing, the Benjamini and Hochberg (B&H) *p*-value correction was applied to all ANOVA results (Benjamini and Hochberg, 1995). A corrected B&H *p*-value < 0.05 was considered significant.

Derivation of Candidate Gene Regulatory Networks from Microarray Data

The YEASTRACT database of regulatory relationships in *S. cerevisiae* (Teixeira et al., 2017) was queried to identify transcription factors controlling genes that exhibited significant expression changes in response to cold shock. For each strain tested, the list of genes with significant B&H ANOVA *p*-values was entered into YEASTRACT using the “Rank by TF” function. Transcription factors regulating these genes were retrieved from the YEASTRACT database using only the “DNA binding evidence” filter.

The list of the 35 most significant regulators for each strain was recorded. The transcription factors that were deleted in the cold shock experiments from which data was obtained—Cin5, Gln3, Hap4, Hmo1, and Zap1—were then added to the lists if not already present. The “Generate Regulation Matrix” utility in YEASTRACT was used to construct unweighted gene regulatory networks from these lists of transcription factors on the basis of DNA binding evidence. The networks were subsequently pared down by individually deleting the least significant transcription factors, along with any target genes that became disconnected from the GRNs, until approximately 15 transcription factors remained. Following this protocol, a family of six database-derived GRNs with 14-17 genes and 25-35 edges was obtained. The networks were exported from YEASTRACT as unweighted adjacency matrices for modeling using GRNmap.

Network Modeling and Parameter Estimation Using GRNmap

GRNmap is an open-source MATLAB software package developed by the Dahlquist Lab for the purpose of modeling gene regulatory network dynamics and estimating parameters from microarray data (Dahlquist et al., 2015). GRNmap uses mass balance ordinary differential equations to model the nonlinear expression of each gene within the network (Equation 1).

$$\frac{dx_i(t)}{dt} = \frac{P_i}{1 + \exp\left(-\left(\sum_{j=1}^{n_g} (w_{ij}x_j(t)) - b_i\right)\right)} - d_i x_i(t) \quad (1)$$

The change in expression of $x_i(t)$ over time is represented by a sigmoidal term indicating production rate, originally proposed in Vu and Vohradsky (2007), minus a linear term $d_i x_i(t)$ indicating degradation rate. In the production term, the numerator P_i is the maximal production rate, n_g is the number of genes in the network, the parameter w_{ij} is the interaction weight of the regulatory gene j controlling the production of the target gene i , and the parameter b_i is the expression threshold.

The differential equation models of gene expression are fit to experimental microarray data using a penalized least squares approach, which minimizes the discrepancy between model outputs and observed data (Equation 2).

$$E = \frac{1}{n_g n_t} \sum_{i=1}^{n_g} \sum_{k=1}^{n_t} [z^d(t_r) - z^c(t_r)]^2 + \alpha \|\theta\|^2 \quad (2)$$

The least squares error (LSE), represented by the first term of equation 2, is reached by minimizing the total sum of squares of the differences between the observed \log_2 fold changes (z^d) and the estimated \log_2 fold changes (z^c), derived from equation 1, for each gene in the

network at each time point. The total sum of squares is then divided by the product of the number of genes in the GRN (n_g) and the number of time points (n_t) to yield the LSE. The optimization problem is constrained by addition of a penalty term, which consists of the square of θ , the parameter vector, multiplied by α , which is used for weighting. An efficient α value for medium-scale GRN parameter estimation, 0.002, was empirically derived by varying α during successive modeling of a 15-gene sample GRN and analyzing the resultant L-curve plot comparing the size of the regularized solution against the size of the associated residual (Hansen and O'Leary, 1993). Following optimization, GRNmap outputs the LSE for the model run as well as the minimum theoretical LSE (minLSE). The minLSE is determined by dividing the total sum of squares of the differences between the observed and average \log_2 fold changes by the total number of data points ($n_g \cdot n_t$). By comparing observed expression to average expression rather than estimated expression, minLSE captures the degree of variance present in the input microarray data. Thus, LSE:minLSE ratio values allow for the comparison of model performance between runs where different expression data was input into GRNmap.

Whereas the least squares error provides a measure of model fit to experimental data for the entire GRN, the mean square error (MSE) indicates model fit to expression data for individual genes in the network (Equation 3).

$$\text{MSE} = \frac{1}{n_t} \sum_{k=1}^{n_t} [z^d(t_r) - z^c(t_r)]^2 \quad (3)$$

MSE represents the summation of the squared differences between observed \log_2 fold changes (z^d) and estimated \log_2 fold changes (z^c) at each time point for a single gene divided by the total number of time points (n_t). The minimum theoretical MSE (minMSE) is determined by substituting the estimated \log_2 fold changes in Equation 3 with the average \log_2 fold changes at

each time point. MinMSE values indicate the variance in gene expression over time. MSE:minMSE ratios are calculated to allow for the comparison of model fit to actual expression between genes from the same GRN or from different GRNs.

The unweighted adjacency matrices representing the six GRNs derived from YEASTRACT were prepared for GRNmap modeling by creating input sheets in Microsoft Excel. Detailed instructions for formatting GRNmap input sheets are available online: https://github.com/kdahlquist/GRNmap/wiki/How-to-format-the-input-file-for-GRNmap-v1.4-and-above#optimization_parameters_sheet. In brief, a series of worksheets were created in a single Excel workbook, which contain input data for the GRNmap software. The worksheets supplied the following information: microarray expression data; initial guesses for the production rate (p), degradation rate (d), and expression threshold (b) parameters; optimization parameters (e.g. the value of α); and an adjacency matrix representing the connectivity of the unweighted GRN to be modeled.

In our experiments, degradation rates were derived from the RNA half-lives reported in Neymotin et al. (2014). If the RNA half-life for a particular gene was not reported in this data set, then the median half-life among the set of 203 transcription factors reported in Harbison et al. (2004) was used. Assuming a steady state, initial guesses for production rates were input as twice the degradation rate values. Microarray expression data obtained at the 15-, 30-, and 60-minute time points for all replicates was included. Missing values were filled by averaging the expression values of the other replicates at the same time point. Wild-type strain data was always included, and additional deletion strain expression data was incorporated when the deleted transcription factors were represented in the candidate GRN. During optimization, the sigmoidal production function described above was applied, and regulatory weights and expression

thresholds were estimated. For a complete description of the optimization parameters utilized in these modeling runs, visit the following web page:

https://openwetware.org/wiki/Dahlquist:Microarray_Data_Analysis_Workflow#Step_11:_GRNmap.

Following construction of input sheets for each of the six database-derived networks, GRNmap modeling was performed. The GRNmap source code and executable can be found at the following site: <http://kdahlquist.github.io/GRNmap/downloads/>. All modeling was performed in MATLAB (version 2014b, MathWorks, Natick, Massachusetts) by running the GRNmap v1.4.4 executable.

Post Hoc Analysis of GRNmap Estimated Parameters and Network Connectivity

A suite of custom R scripts was created for the semi-automated post hoc analysis of gene regulatory networks following GRNmap modeling: https://github.com/kdahlquist/DahlquistLab/tree/master/R_scripts. The Degree-Distribution-Generator.R script creates degree distribution charts from input unweighted adjacency matrices. The L-Curves.R script generates L-curve plots from compiled lists of least squares error (E , or LSE), alpha (α), and penalty values derived from GRNmap output workbooks. A third script, Random-Matrix.R, generates random networks with specified numbers of genes and edges for comparison to database-derived networks. The code for these scripts and instructions for their use can be found on the Dahlquist Lab GitHub page, linked above.

In addition to these R scripts, a Microsoft Excel workbook including prespecified equations for the automated generation of summary statistics and heat maps comparing estimated weight parameters was also created and can be downloaded at the following address:

https://github.com/kdahlquist/DahlquistLab/blob/master/data/Spring2017/15-gene_networks_analysis/Regulatory-Relationships_Six-Networks_BK20170313.xlsx.

Manual statistical analysis of estimated GRN parameters was performed in SPSS (version 22.0, IBM Statistics, Chicago, Illinois). For multiple regression analysis, all models were selected by backward selection and verified through forward selection. P values < 0.05 were considered significant. Graph statistical analysis of GRN connectivity was performed using the open-source software Gephi (Bastian et al., 2009).

Visualization of Gene Regulatory Networks with GRNsight

Unweighted and weighted diagrams of each network modeled in GRNmap were created using our companion web application, GRNsight v2.3.0 (<http://dondi.github.io/GRNsight/>). GRNsight automatically lays out GRNs represented by adjacency matrices in GRNmap input or output workbooks (Dahlquist et al., 2016). For each GRN diagram, the boxes representing transcription factors were manually arranged into a predetermined grid formation, which allowed for the consistent placement and alphabetization of the transcription factors. When modeling motifs and a single consolidated network, boxes were instead arranged to minimize overlapping edges and enhance overall clarity.

RESULTS:

Six Related Gene Regulatory Networks Were Derived from Cold Shock Microarray Data

Statistical analysis of microarray data from wild-type *Saccharomyces cerevisiae* and four deletion strains subjected to cold shock for one hour was performed. A modified ANOVA was used to identify genes with expression changes that significantly differed from zero at any time point. Following Benjamini and Hochberg correction to account for the multiple testing problem,

the number of genes that exhibited significant expression changes in each strain were totaled (Table 1). In wild-type *S. cerevisiae*, 31% of genes were involved in the transcriptional response to cold shock, which accords with earlier reports (Sahara et al., 2002). In the four deletion strains, responses were of a similar magnitude, with 28-30% of genes showing significant changes in expression at any time point.

Table 1. Within-strain ANOVA results for wild-type *S. cerevisiae* and four deletion strains indicate a widespread transcriptional response to cold shock. Expression changes were considered significant for genes with Benjamini and Hochberg corrected ANOVA *p*-values < 0.05.

Strain	Wild-type	$\Delta cin5$	$\Delta gln3$	$\Delta hap4$	$\Delta zap1$
Significant Genes	1936 (31%)	1683 (28%)	1683 (28%)	1794 (29%)	1859 (30%)

YEASTRACT was used to identify transcription factors (TFs) whose targets were overrepresented in the list of genes that exhibited significant expression changes in each strain based on DNA binding evidence. The 35 most significant regulators were retrieved for each strain, and the TFs Cin5, Gln3, Hap4, and Hmo1 were added to that list if not already present. Using YEASTRACT, these lists of significant TFs for each strain were connected through known regulatory relationships to yield candidate gene regulatory networks (GRNs). The five GRNs were then pared down in size through serial deletion of the least significant TFs until a set of six medium-scale networks with 14-17 genes each was obtained. Two networks were derived from the $\Delta cin5$ strain data, as deletion of Mcm1 from the larger 17-gene network disconnected Zap1 and its target Ace2, yielding a smaller 14-gene network. Because deletion strain data was available for Zap1, the larger 17-gene network including Zap1 was analyzed in addition to the 14-gene network. The final GRNs were named db1-db6. A total of 27 unique TFs were present in the family of networks (Table 2). Notably, Msn2 and Swi4 appeared in all six GRNs and are known to play roles in regulating the response to low temperatures in yeast (Kandror et al., 2004; Córcoles-Sáez et al., 2012). All of the transcription factors that had been deleted in the cold

shock experiments were represented in the six networks with the exception of Zap1, which was absent from db2 and db4. Other common TFs appearing in five of the six related networks included Sfp1, Yhp1, Yox1, and Gcr2.

Table 2. Six medium-scale candidate GRNs were derived from cold shock microarray data. A total of twenty-seven unique transcription factors were found in the networks.

Network	db1	db2	db3	db4	db5	db6
Strain Data Used	Wild-type	$\Delta cin5$	$\Delta cin5$	$\Delta gln3$	$\Delta hap4$	$\Delta zap1$
Genes, Edges	16, 26	14, 25	17, 32	14, 35	15, 28	16, 27
Gene List	<i>ABF1</i>					<i>ABF1</i>
	<i>ACE2</i>		<i>ACE2</i>		<i>ACE2</i>	<i>ACE2</i>
	<i>AFT2</i>					
	<i>ASF1</i>					
	<i>ASH1</i>				<i>ASH1</i>	
	<i>CIN5</i>	<i>CIN5</i>	<i>CIN5</i>	<i>CIN5</i>	<i>CIN5</i>	<i>CIN5</i>
						<i>CST6</i>
				<i>CYC8</i>		
	<i>GCN4</i>					<i>GCN4</i>
		<i>GCR2</i>	<i>GCR2</i>	<i>GCR2</i>	<i>GCR2</i>	<i>GCR2</i>
	<i>GLN3</i>	<i>GLN3</i>	<i>GLN3</i>	<i>GLN3</i>	<i>GLN3</i>	<i>GLN3</i>
	<i>HAP4</i>	<i>HAP4</i>	<i>HAP4</i>	<i>HAP4</i>	<i>HAP4</i>	<i>HAP4</i>
	<i>HMO1</i>	<i>HMO1</i>	<i>HMO1</i>	<i>HMO1</i>	<i>HMO1</i>	<i>HMO1</i>
						<i>HSF1</i>
			<i>MCM1</i>			<i>MCM1</i>
		<i>MGA2</i>	<i>MGA2</i>			<i>MGA2</i>
	<i>MSN2</i>	<i>MSN2</i>	<i>MSN2</i>	<i>MSN2</i>	<i>MSN2</i>	<i>MSN2</i>
				<i>MSN4</i>		<i>MSN4</i>
		<i>RDS3</i>	<i>RDS3</i>			
	<i>SFP1</i>	<i>SFP1</i>	<i>SFP1</i>	<i>SFP1</i>	<i>SFP1</i>	
		<i>STB5</i>	<i>STB5</i>		<i>STB5</i>	
	<i>SWI4</i>	<i>SWI4</i>	<i>SWI4</i>	<i>SWI4</i>	<i>SWI4</i>	<i>SWI4</i>
		<i>SWI5</i>	<i>SWI5</i>	<i>SWI5</i>	<i>SWI5</i>	
				<i>TEC1</i>		
	<i>YHP1</i>	<i>YHP1</i>	<i>YHP1</i>	<i>YHP1</i>	<i>YHP1</i>	
	<i>YOX1</i>	<i>YOX1</i>	<i>YOX1</i>	<i>YOX1</i>	<i>YOX1</i>	
	<i>ZAP1</i>		<i>ZAP1</i>		<i>ZAP1</i>	<i>ZAP1</i>

Forward Simulation of the Six Networks Using GRNmap Fit Expression Data Well

Each of the six database-derived networks was modeled in GRNmap. Estimated parameters included thresholds of expression (b), production rates (p), and regulatory weights (w). B and p values were estimated for each gene, whereas w values were estimated for each edge. The total number of parameters (θ) estimated in these model runs is therefore represented by the equation

$\theta = (2 \times n_g) + n_e$, where n_g is the number of genes, and n_e is the number of edges in the network. To determine an efficient penalty term value for medium-scale GRN modeling and parameter estimation, sixteen model runs were performed for each network. Unique values of alpha ranging from 0.8000-0.0001 were used for each run to variably scale the penalty term. Following the sixteen runs, L-curve plots of the least squares error (LSE) versus the penalty term value were generated (Figure 1). Efficient regularization was observed for db1-db6 at $\alpha=0.002$. This alpha value was consequentially selected use in all subsequent model runs, and the modeling results generated for db1-db6 at $\alpha=0.002$ were analyzed.

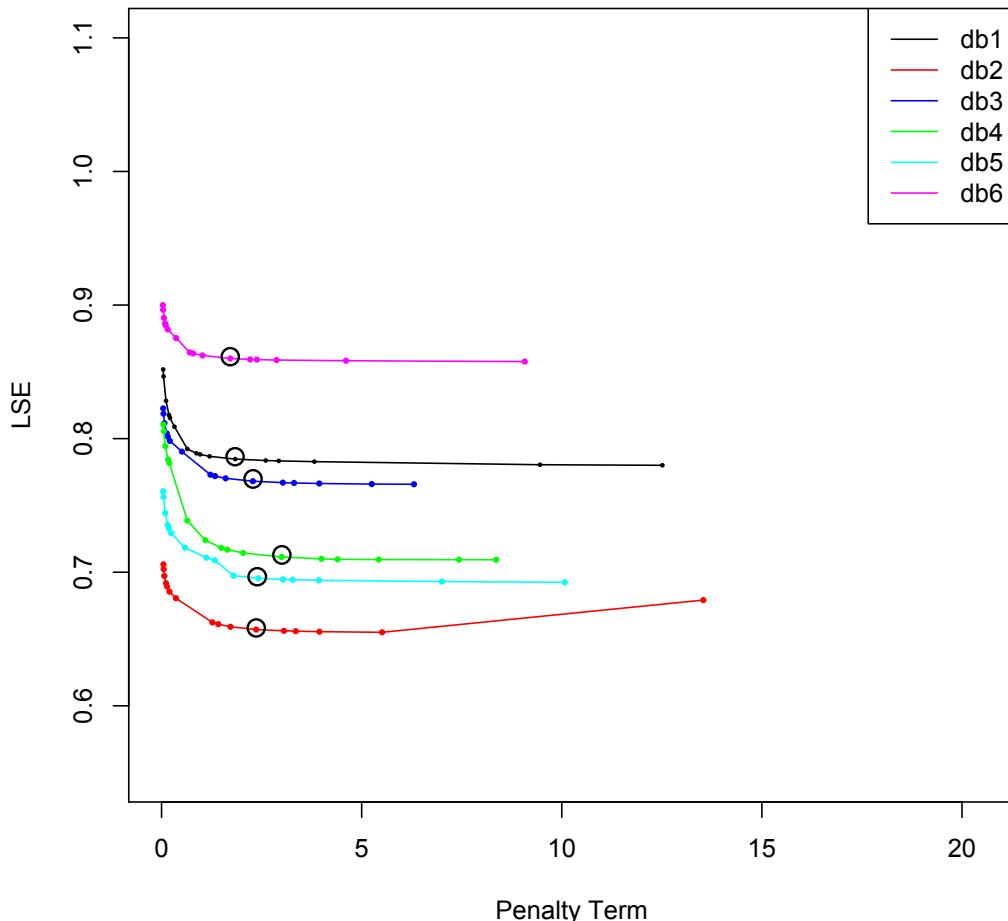


Figure 1. L-curve plots of least squares error (LSE) versus penalty term value for db1-db6 show efficient regularization at $\alpha=0.002$ (corresponding runs are circled in black). Individual points correspond to model runs conducted with unique α values, which ranged from 0.8000-0.0001. Sixteen runs were completed for each network. The L-Curves.R script was used to generate this graph.

Assessment of optimization diagnostics indicated that the six networks consistently fit the input expression data well, with the ratios between the observed least squares error (LSE) and theoretical minimum least squares error (minLSE) values, or LSE:minLSE ratios, ranging from 1.3000 for db4 to 1.4263 for db5 (Table 3). Given that this goodness-of-fit measure is normalized based on the theoretical minimum least squares error possible given the variation in input microarray data, the LSE:minLSE ratio provides a fair comparison between networks of varying size.

Table 3. Optimization parameters output following forward simulation in GRNmap demonstrated consistent and effective modeling of the six database-derived networks.

Parameter	db1	db2	db3	db4	db5	db6
Penalty Term	2.5923	1.8830	1.7570	1.7895	2.3443	1.8276
LSE	0.8194	0.6634	0.7864	0.6994	0.6919	0.8602
minLSE	0.5768	0.4885	0.5449	0.5379	0.4851	0.6156
LSE:minLSE Ratio	1.4206	1.3580	1.4100	1.3000	1.4263	1.3973
Iteration Count	109,718	53,862	118,921	78,124	62,139	76,769

Assessment of individual expression plots generated by GRNmap demonstrated good modeling of gene expression changes over time in the six GRNs. A sampling of plots from db1 showcases four common transcriptional profiles (Figure 2). MSE:minMSE ratios, which ranged from 0.2209-4.2798 in db1-db6, were low for each of the TFs indicating good model fit to actual expression data. TF within-strain ANOVA *p*-values are included for reference. Interestingly, they were not correlated with MSE:minMSE ratio (Pearson's R, r=0.0579, n=16, p=0.8313).

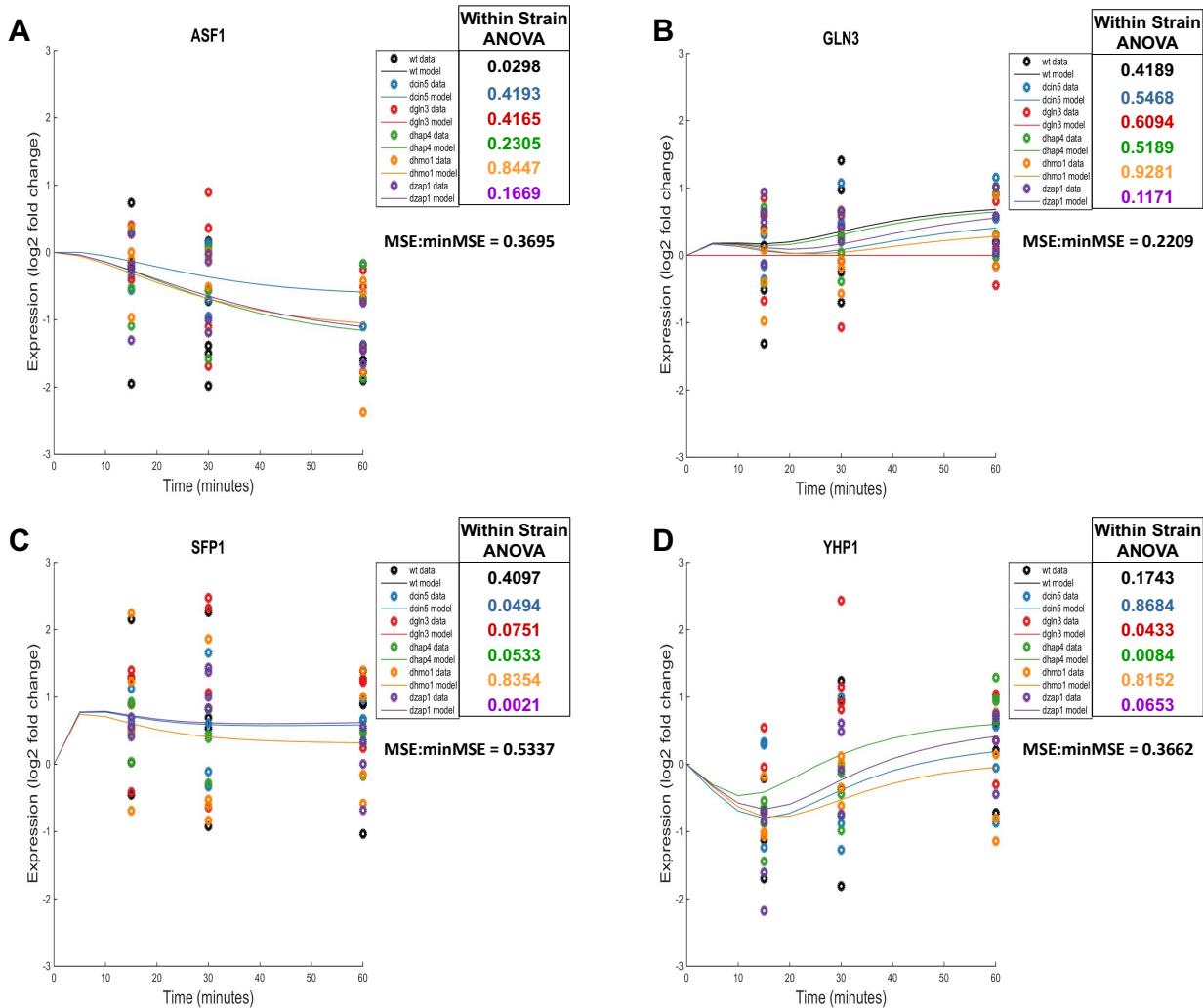


Figure 2. Forward simulation of db1 in GRNmap accurately modeled common expression profiles. Within-strain ANOVA *p*-values are overlaid onto the expression plots automatically output by GRNmap, with color coding indicating individual strains from which input data was derived. MSE:minMSE ratio values averaged across all six strains are also reported, indicating the goodness-of-fit of the model data (lines) to the actual expression of each gene (circles) in each strain. The complete set of expression plots generated from this model run can be accessed here: https://github.com/kdahlquist/DahlquistLab/blob/master/data/Spring2017/bouldardii2_GRNmap_inputs_outputs/GRNmap_db1-through-db6_input_output/GRNmap_db1-through-db6_figures/DB1_wt_figures_NEW.zip

Estimated parameters varied between the six gene regulatory networks. Figure 3 shows the compiled optimized *p* and *b* values output for each of the 27 unique TFs found in db1-db6. Although both parameters exhibited wide ranges, consistent estimation across multiple GRNs was observed for several TFs. Production rates for Gcr2, Swi4, and Zap1 and net thresholds for

Gln3 were similar across each GRN. Further, both the estimated p and b values for Ace2 and Hmo1 were consistently modeled over a short range, representing the conserved dynamics of these TFs in their respective GRNs. In contrast, the TFs Msn2 and Yhp1 exhibited highly variable estimated p and b values in db1-db6. In general, variability was associated with high magnitude parameter estimates.

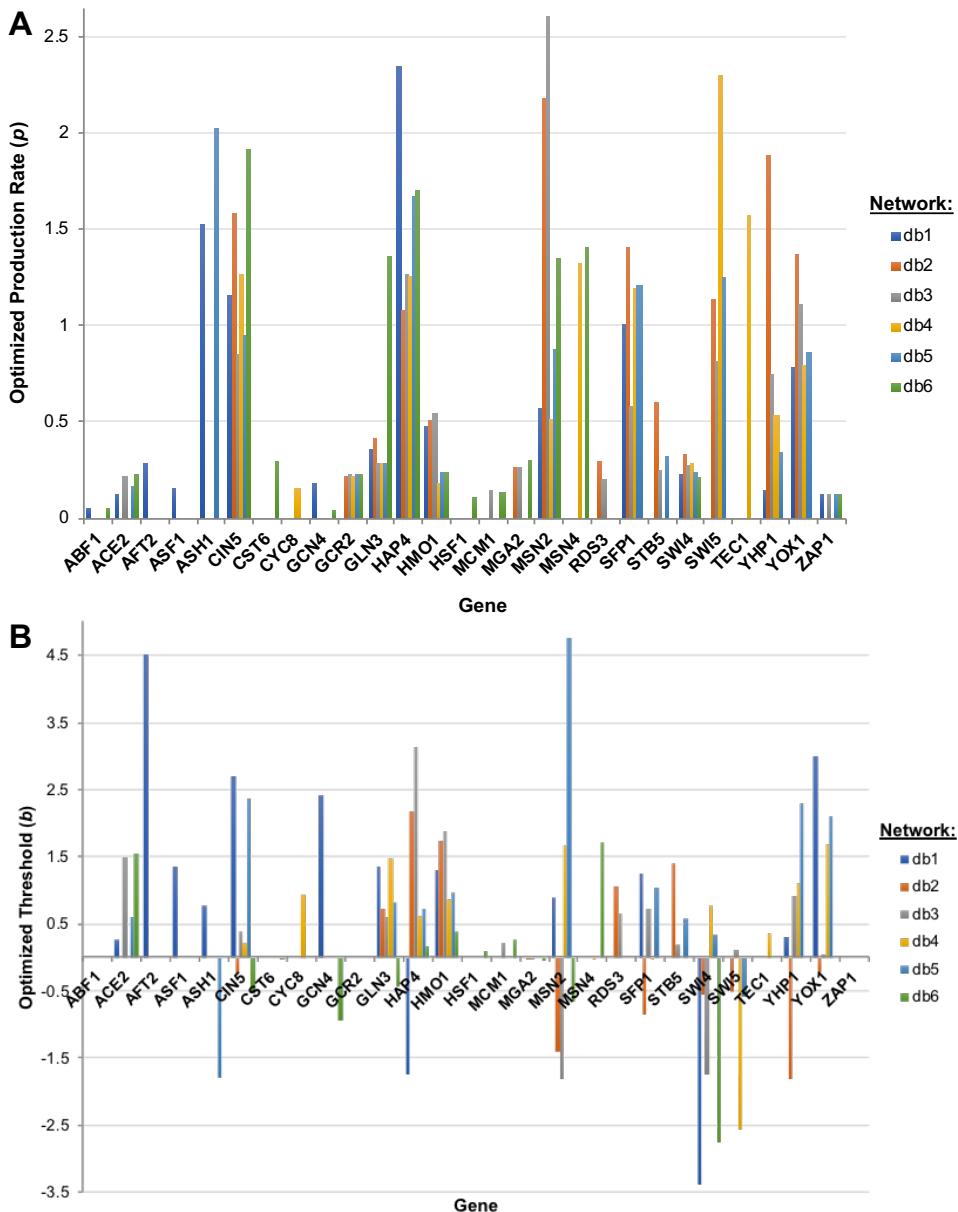


Figure 3. Production rates (A) and thresholds (B) estimated by GRNmap for db1-db6 were consistent for some TFs and variable for others. Bars are color coded based on GRN number as indicated in the legends.

Db5 Outperforms 30 Random Networks Containing Identical Genes and the Same Number of Edges

Db5 was selected as a representative candidate network for further study. Thirty random GRNs were generated with the same genes and numbers of edges as db5, but with varying connectivity. Modeling results for the random networks demonstrated that db5 generally provided a better fit to actual expression data, with only 5 random networks exhibiting lower LSE:minLSE ratios when compared to db5 (Figure 4). Further, the average LSE:minLSE ratio for the six database-derived networks (1.3853) indicated that these GRNs exhibited better fits than even the best performing random network (LSE:minLSE=1.3880). These findings support the validity of the database-derived networks, which appear better suited to the modeling of the transcriptional response to cold shock in yeast.

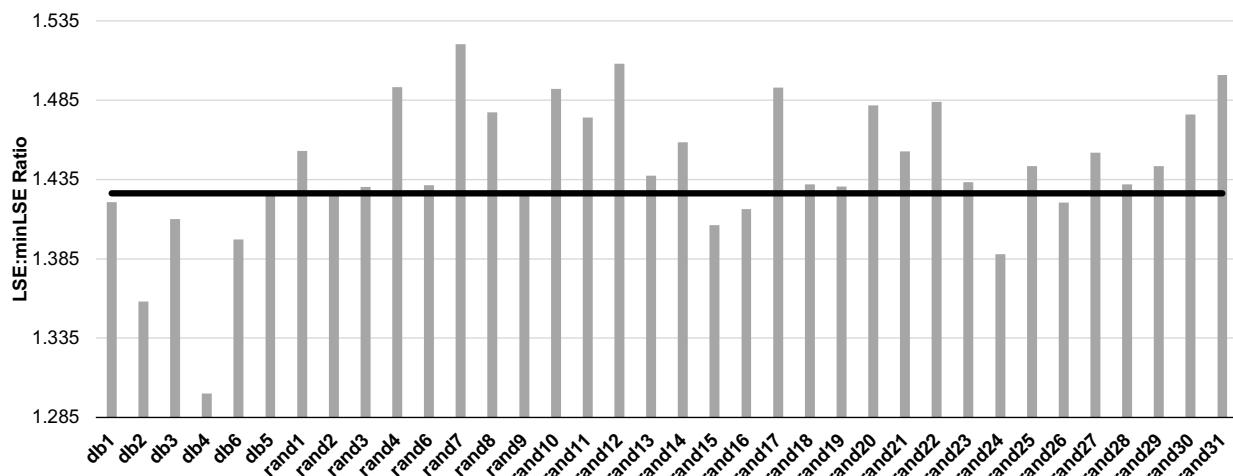


Figure 4. The network derived from $\Delta hap4$ strain data, db5, generally outperformed 30 random networks composed of the same transcription factors and containing the same number of edges. The LSE:minLSE ratios for db1-db6 are reported on the left and for the thirty random networks on the right. A horizontal line is drawn demarcating the LSE:minLSE ratio of db5 for comparison.

To determine whether the best performing random networks had similar connections to the database-derived parent network, the number of edges shared between each of the thirty random networks and db5 was determined. Plotting the LSE:minLSE ratio of individual random

networks versus the number of regulatory relationships they shared with db5 revealed an interesting pattern (Figure 5).

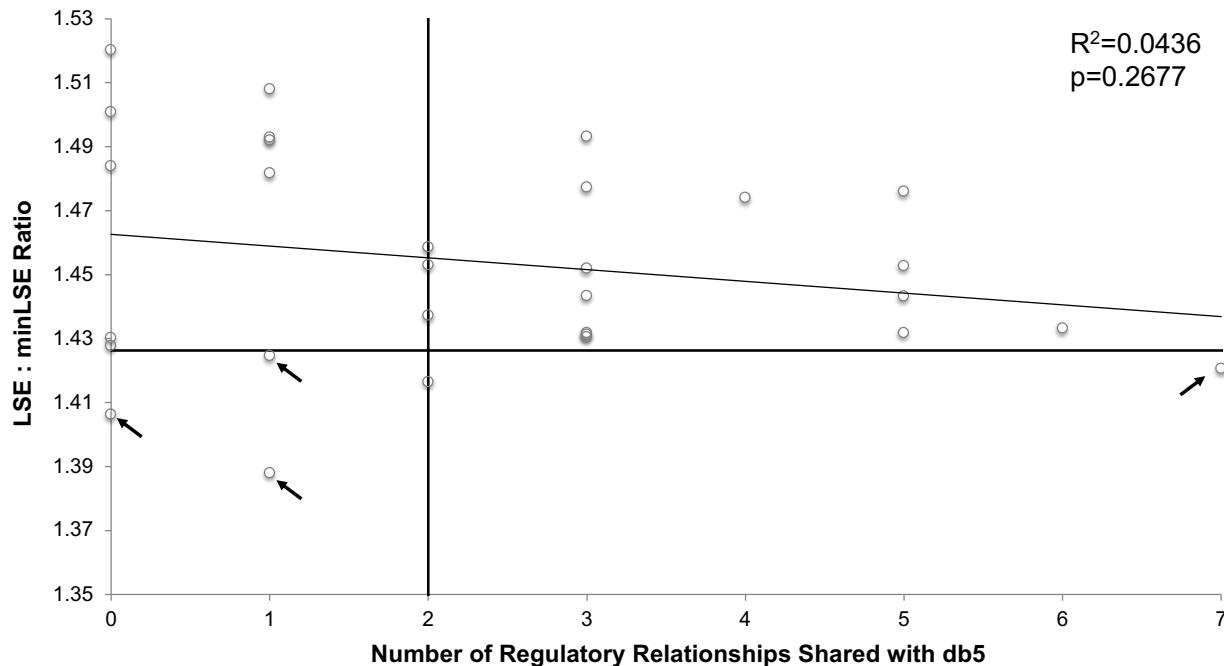


Figure 5. LSE:minLSE ratio of random networks versus the number of regulatory relationships shared with db5. The vertical line indicates the median number of regulatory relationships shared with db5 (2). The horizontal line marks the LSE:minLSE ratio of db5 (1.4263). A linear regression line was fit to the data, with its goodness-of-fit value reported in the top right corner. Notable outliers are indicated with black arrows.

Overall, a weak negative correlation between LSE:minLSE ratio and the number of regulatory relationships random networks shared with db5 was observed (Pearson's R, $r=-0.2089$, $n=30$, $p=0.2677$). In general, networks that shared increasing numbers of regulatory relationships with db5 approached the model fit obtained for db5, with only one random network sharing greater than the median number of relationships exceeding the database-derived network's fit to experimental data (Figure 5, arrow pointing right). However, three random networks sharing only one or no regulatory relationships outperformed db5 (Figure 5, arrows pointing left). These results suggest that although better model fit could be attained by sharing edges with the database-derived network, random networks that exhibited extensive rewiring

without regard to DNA binding evidence could occasionally outperform the database-derived network.

GRNsight Visualization and Analysis of Database-Derived Network Connectivity

To investigate the connectivity of db1-db6, the network weights were visualized using the GRNsight web application (Figure 6). GRNsight models provide visual aids that facilitate the interpretation of complex networks by highlighting common patterns of connectivity. In the case of db1-db6, it can be seen that *ZAPI* is only connected to the networks through its regulation of *ACE2*, which in db3 and db6 does not have any further outputs. The poor connectedness of *ZAPI* to the GRN controlling the response to cold shock in yeast may explain why the networks that excluded *ZAPI*, db2 and db4, showed the best fit to experimental expression data (Table 3). Another common feature highlighted in the GRNsight models is the positive autoregulation exhibited by *HMO1*. *HMO1* further activates several other genes such as *MSN2* and *CIN5* but itself does not receive any inputs in db2, db3, db4, or db5. This suggests that *HMO1* expression is activated early in response to low temperature stress and provides multiple activating signals to the GRN. A final qualitative observation readily derived from GRNsight visualization is the high connectedness of *MSN2* to the database-derived networks.

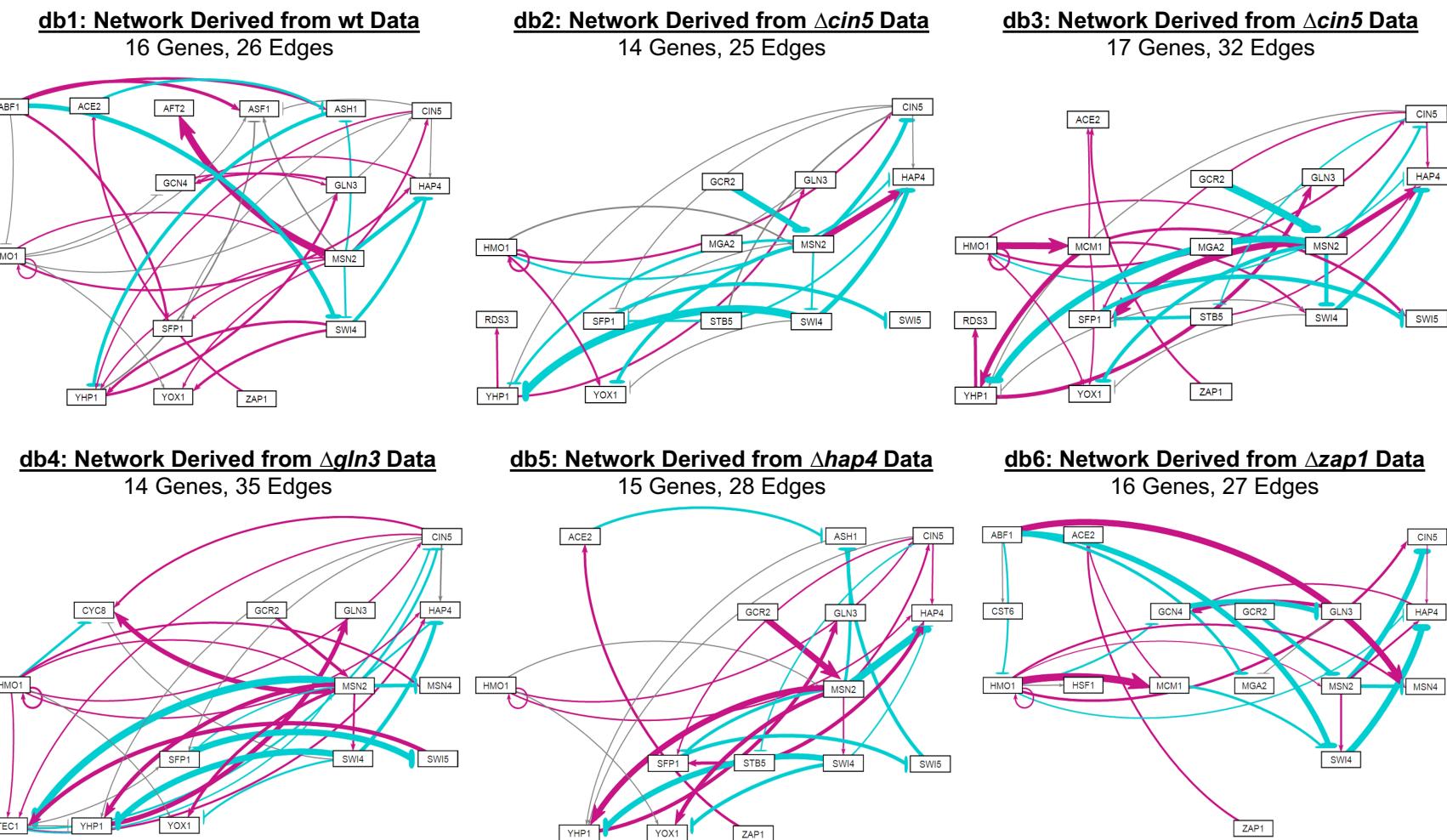


Figure 6. GRNsight visualization highlights salient features of db1-db6. Transcription factors are represented by boxes and arranged in a consistent array. Regulatory relationships are represented by arrows directed from regulator to target. Pointed arrowheads and magenta coloration indicate activation, whereas flat arrowheads and cyan coloration indicate repression. Edges colored in gray exhibit small influence, meaning their associated weights are <5% of the highest magnitude edge weight in the network.

Quantitative analyses of node degree and edge weights were performed to more precisely investigate the connectivity of the six database-derived gene regulatory networks. Transcription factor in-degrees and out-degrees most commonly ranged from 0-2 in db1-db6, although outliers exhibiting high out-degrees ≥ 6 were present (Figure 7). No nodes with degrees greater than four were present in db2, the smaller of the two $\Delta cin5$ -derived networks, which exhibited the smallest number of genes and edges out of the six GRNs. Similarly, the networks with the largest number of genes (db1) and edges (db6) contained the nodes with the highest overall degree.

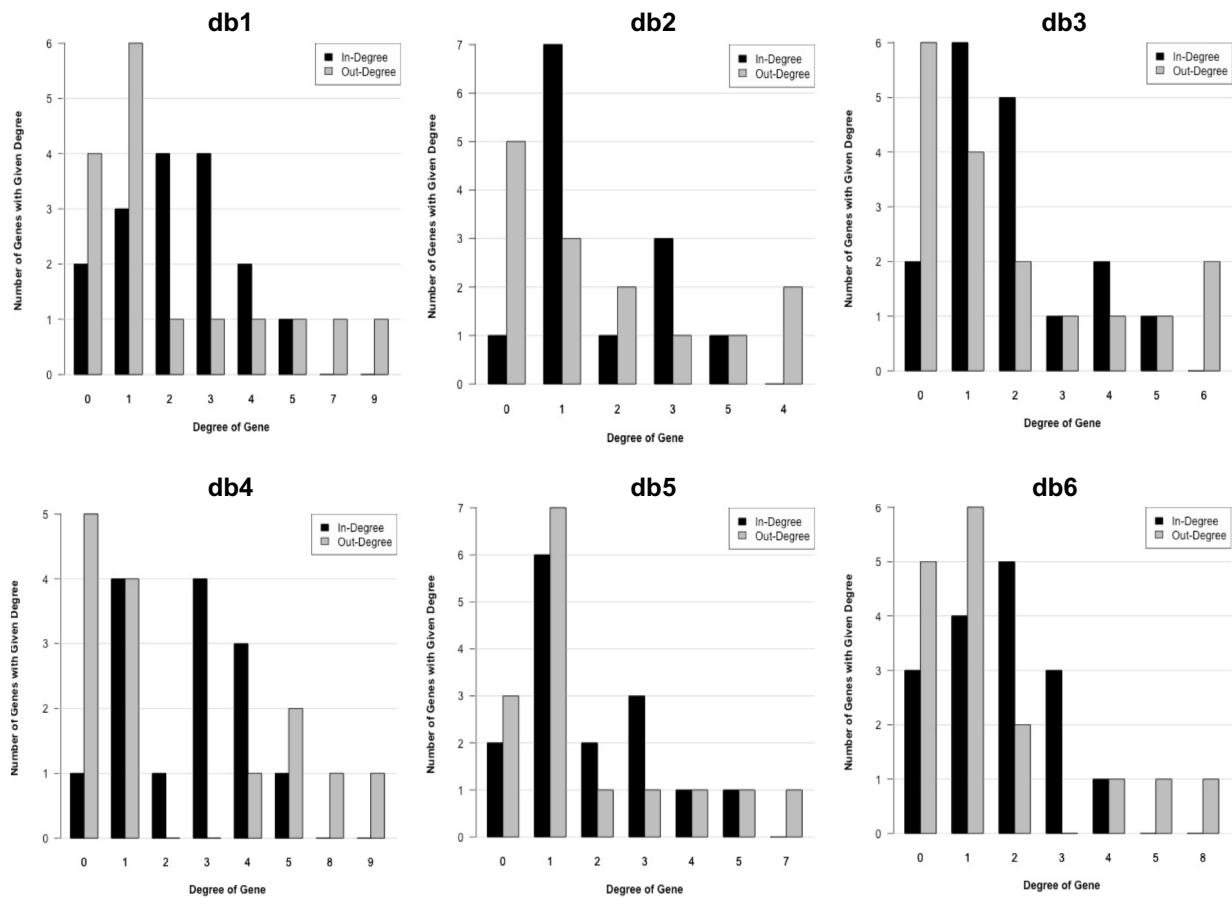


Figure 7. The degree distributions of db1-db6 are similar, with most genes exhibiting low in-degrees (black) and out-degrees (gray). Individual charts and their axes were automatically generated using the Degree-Distribution-Generator.R script.

The frequencies of regulatory relationships, the sums of activation and repression relationship weights, and the maximum as well as minimum weight values all showed

considerable variation in the six related networks (Table 4). The wild-type derived network, db1, differed from the deletion strain derived networks due to the high sum and frequency of its activation relationships. Although db3-db6 contained larger numbers of activation relationships than repression relationships, the overall magnitude of repression either was similar to or exceeded that of activation in these networks. Further, the small *Δcin5*-derived network was dominated by repression relationships, of which several exhibited high magnitudes. Among the related GRNs, db1 and db2 possessed the largest numbers of small influence regulatory relationships. In GRNsight v2.3.0, regulatory relationships were considered small influence when their weight values were less than 5% of the highest magnitude edge weight in the network. In db1 and db2, normalization to individual high magnitude weights resulted in overrepresentation of small influence edges. In db1, the estimated weight of *MSN2* regulating *AFT2*, which only appears in db1 and receives no other inputs, was 5.9424 and far exceeded any other positive weight in db1-db6. In db2, the second highest magnitude weight for a regulatory relationship, -4.2947, was observed between *SWI4* and *YHP1*. Together, these outliers drove down the weights of other regulatory relationships following normalization in GRNsight, which resulted in the classification of many edges as small influence in db1 and db2.

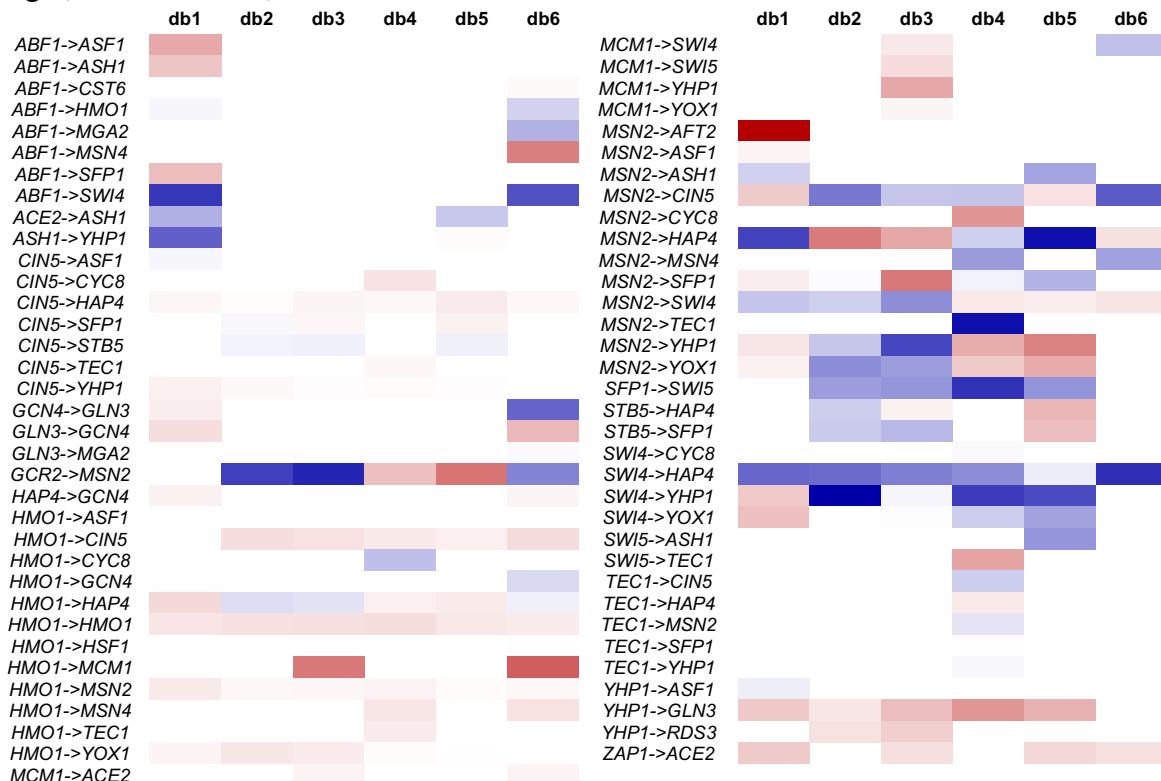
Table 4. The six database-derived networks exhibit varying degrees of activation and repression. Counts of edge type are provided based on the color coding criteria used in GRNsight (Figure 6).

	db1	db2	db3	db4	db5	db6
Activation Relationships	19	6	18	16	14	13
Repression Relationships	7	11	10	12	10	11
Small Influence Relationships	10	8	4	7	4	3
Sum of Activation Weights	23.4929	6.9575	18.8606	17.1864	17.7493	13.5913
Sum of Repression Weights	-15.3497	-20.1966	-17.2608	-19.7614	-16.4508	-19.4790
Maximum Weight	5.9424	3.0815	3.1810	2.4940	3.2167	3.7631
Minimum Weight	-3.3601	-4.2947	-3.6707	-4.0281	-4.0374	-3.5141

Identification of Conserved Motifs in the Database-Derived GRNs

Although the weight statistics reported above indicate overall regulation in db1-db6, assessment of individual edges is critical to understanding network connectivity. Thus, a heat map indicating the relative weights of each edge in db1-db6 was created to identify conserved relationships and compare their modes of regulation in different networks (Table 5).

Table 5. Heat map visualizing regulatory weights for each of the sixty-nine edges represented in db1-db6. Positive weights are indicated by red and negative weights by blue. Darker shades correspond to higher magnitude weights. All values were normalized to the largest magnitude weight, *MSN2*→*AFT2*, in db1.



Through this method, multiple activation relationships appearing in five or more networks were identified: *HMO1*→*HMO1*, *HMO1*→*CIN5*, *HMO1*→*MSN2*, *HMO1*→*YOX1*, *CIN5*→*HAP4*, *CIN5*→*YHP1*, and *YHP1*→*GLN3*. As indicated by qualitative assessment of GRNsight models, *HMO1* consistently up-regulates its own expression as well as that of *CIN5*, *MSN2*, and *YOX1* in the database-derived networks. Additionally, two of the conserved activation relationships not

directly involving *HMO1* form a regulatory chain initiated with the up-regulation of *CIN5* expression by *HMO1* (Figure 8A).

Another gene that was consistently activated by *HMO1* is *MSN2*, which itself regulates two genes in the chain, *CIN5* and *YHP1*, in addition to *YOXI* in five of the GRNs. However, the estimated weight parameters for the edges involving *MSN2* vary from strong activation to strong repression. To simplify subsequent modeling, these edges were represented by their most frequent mode of regulation (activation or repression) in db1-db6. Upon addition of the regulatory relationships involving *MSN2* and the consistent activation relationship *HMO1* → *YOXI* to the regulatory chain, a conserved system of three feedforward loops is observed (Figure 8B). This includes a coherent type 1 (C1) feed forward loop (FFL) terminating on *YOXI* and an incoherent type 1 (I1) FFL converging on *CIN5*, which are the most common types of FFLs found in gene regulatory networks, including in *S. cerevisiae* (Mangan and Alon, 2003). The third feed forward loop ending on *YHP1* is a rare incoherent type 4 FFL.

MSN2 also regulates *SWI4* and *YOXI*, a paralog of *YHP1*, in five of the six networks (Gitter et al., 2009). These edges create additional feedforward loops when including connections to *YHP1*, although their estimated regulatory weights vary between networks. Mapping of the connection between these four genes with modes of regulation determined by the most frequently appearing edges reveals a system of two symmetrical I1-FFLs that terminate on the paralogs *YHP1* and *YOXI* (Figure 8C). Interestingly, the gene that appears only in the five GRNs derived from deletion strain microarray data, *GCR2*, exhibits a single output to *MSN2*. Thus, we speculate that *GCR2* may play a role in a genetic backup circuit that signals through *MSN2* in mutant strains. Adding this symmetrical pair of feedforward loops to the previously described FFLs results in a conserved system of five FFLs present at the core of five out of six

database-derived GRNs (Figure 8D). The only edges appearing in 5+ GRNs that are not represented in this system are those originating from *HMO1*, *MSN2*, *CIN5*, and *SWI4* that terminate on *HAP4*, the previously described connection between *GCR2*→*MSN2*, and a final edge connecting *MSN2*→*SFP1*. It is worth noting that the only repression relationships that are consistently estimated to have similar weight values are *SWI4*→*HAP4* and *SFP1*→*SWI5*, which only appears in four of the GRNs.

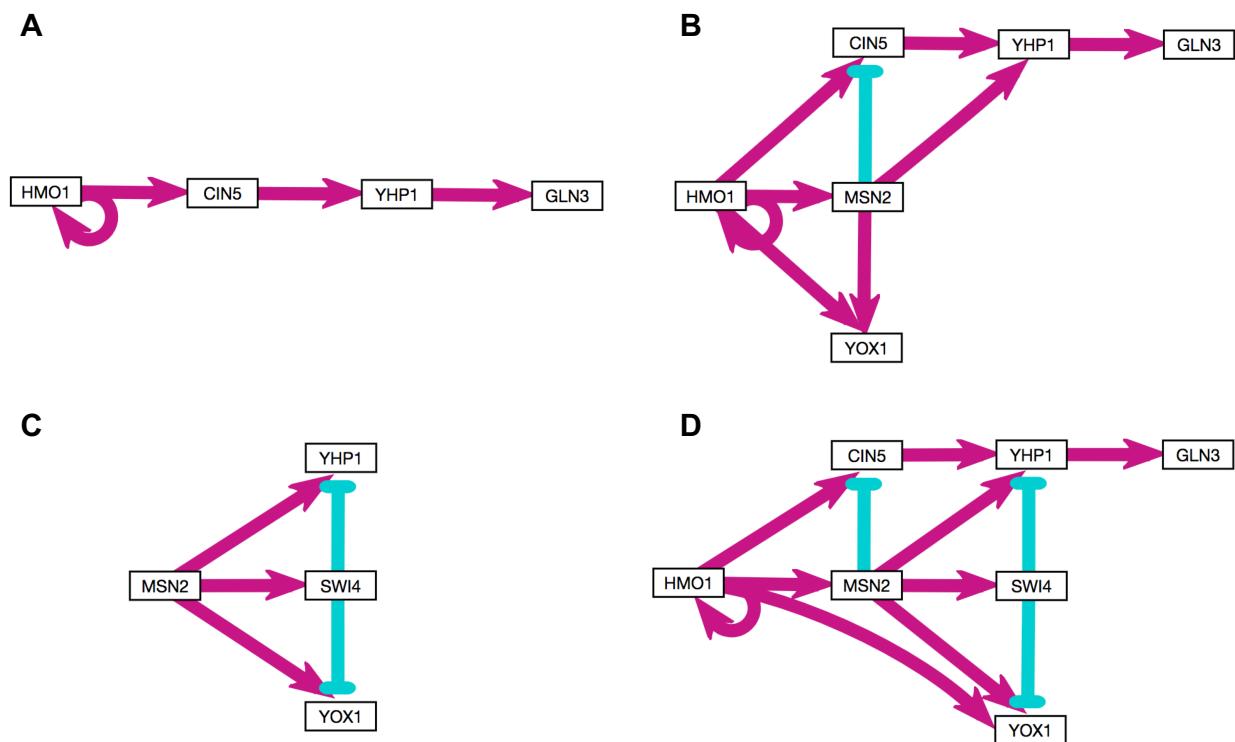


Figure 8. Conserved motifs in the six database-derived GRNs include regulatory chains and feedforward loops (FFLs). A regulatory chain comprised of *HMO1*, *CIN5*, *YHP1*, and *GLN3* exhibited consistent activation in five GRNs (A). Addition of *MSN2*, *YOX1*, and their associated edges (which were weighted based on frequency) to the regulatory chain yielded a system of 3 FFLs (B). *MSN2* also forms symmetrical FFLs mediated by *SWI4* that terminate on the paralogs *YHP1* and *YOX1*, which most frequently appear as incoherent type 1 FFLs (C). Addition of the described motifs (A-C) yields a conserved system containing 5 FFLs and a regulatory chain (D). Boxes indicate genes and arrows indicate edges, with pointed arrowheads representing activation (magenta) and flat arrowheads repression (cyan). Line thicknesses are uniform as edge weights were all estimated to be either 1 or -1.

Centrality Measures Indicate Genes Important to Overall Network Structure

To identify individual genes that are important to overall GRN structure, betweenness centrality and eigenvector centrality were computed for each node in db1-db6 using the open-source software Gephi (Figure 9). Betweenness centrality indicates the degree to which a node is present within the shortest paths connecting other nodes in the network (Freeman, 1977), providing an indication of the extent to which a gene controls communication in the GRN. Eigenvector centrality is a relative metric that increases when a gene is connected to other central genes in the network (Bonacich and Lloyd, 2001), which captures the influence a gene has on overall network dynamics.

The two genes with consistently high betweenness centrality measures in the database-derived networks, *MSN2* and *YHP1* (Figure 9A), were previously identified as mediators in conserved regulatory motifs (Figure 8). This supports the use of betweenness centrality to identify genes that shape overall network connectivity by connecting subcomponents of the network. Eigenvector centrality also indicated that *YHP1* is a high influence factor, while suggesting important roles for less examined genes including *SFP1* and *GCN4* (Figure 9B). Notably, *GCN4* possesses an eigenvector centrality of 1 in both networks where it appears, db1 and db6. In these networks, *GCN4* creates a negative feedback loop with *GLN3* (Figure 6), which may add stability to the networks (Becskei and Serrano, 2000). Further, eigenvector centrality highlights nodes at the end of regulatory chains, such as *GLN3* and *HAP4*, by virtue of their connection to high influence upstream regulators, including *YHP1* and *YOXI*.

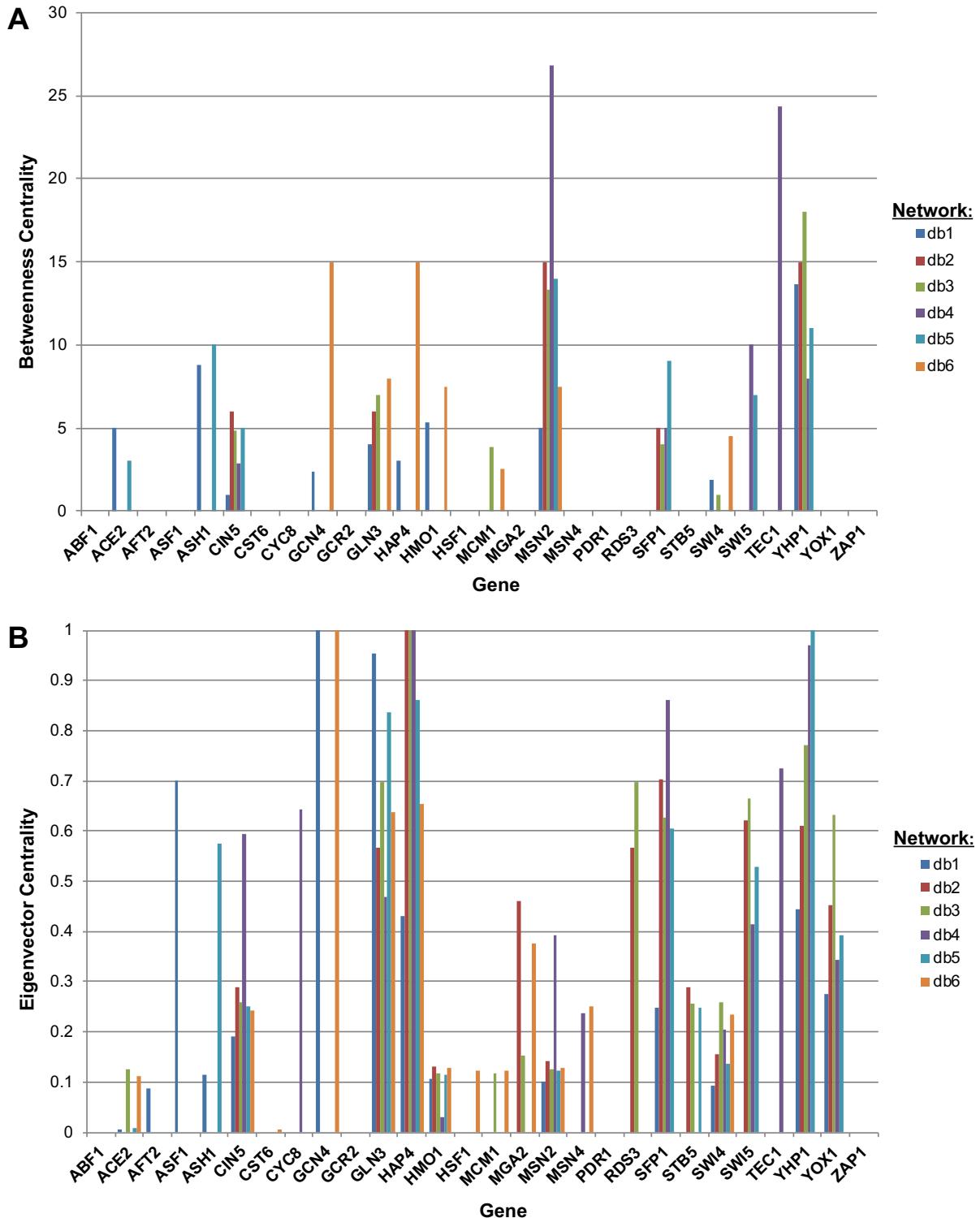


Figure 9. The graph statistics betweenness centrality (A) and eigenvector centrality (B) highlight transcription factors that exhibit high influence on overall network connectivity. Bars are color coded based on GRN number as indicated in the keys.

Multiple Regression Analysis Identifies Node Characteristics Important to Fitting the Expression of Individual Genes

Multiple regression models of node MSE:minMSE ratio were constructed to identify GRNmap model inputs, outputs, or graph statistics that were significantly correlated to better model fit (Table 6). Modeling was performed for all six database-derived networks as well as for the three best and worst performing db5-derived random networks based on LSE:minLSE ratio. Possible predictors that were not consistently correlated to MSE:minMSE ratio included ANOVA Benjamini and Hochberg *p*-value, average log₂ expression at 30 minutes, and betweenness centrality. In contrast, the only consistent predictors of model fit to gene expression data across models were average log₂ expression at 15 minutes and eigenvector centrality.

	Standardized Model Coefficients											
	db1*	db2	db3	db4*	db5	db6*	rand7	rand12	rand15	rand16	rand24	rand31
ANOVA B&H p-value												
Average Expression at 15	-0.507	-0.624	-0.479	-0.552	-0.528	-0.626	-0.838					-0.656
Average Expression at 30												
Average Expression at 60								-0.503				
Degradation Rate									-0.505		-0.611	
Optimized Production Rate	0.578			0.777		0.437						
Optimized Threshold b				0.883								
Weighted In-degree						0.617						
Weighted Out-degree							0.433		-0.627			
Eccentricity												
Closeness Centrality									-0.443			
Betweenness Centrality							0.399			-0.609		
Clustering		0.522	0.473		0.501							
Eigen Centrality	-0.384			-0.533		-0.328		-0.685			-0.598	-0.831
Model Adjusted R²	0.290	0.525	0.360	0.441	0.350	0.196	0.500	0.692	0.436	0.552	0.257	0.437

Table 6. Multivariate models of node MSE:minMSE ratio incorporating GRNmap inputs, outputs, and network graph statistics as predictors were created for db1-db6 as well as the three best (rand15, 16, 24) and worst (rand7, 12, 31) performing db5-derived random networks. The standardized coefficients and goodness-of-fit metrics of the models are provided. All models were significant with the exception of those labeled with an asterisk (*), which fell marginally short of the *p*<0.05 threshold for significance.

The significant negative correlation between expression at 15 minutes and MSE:minMSE ratio demonstrates that early up-regulation of gene expression is more easily modeled by GRNmap than the early down-regulation. This observation reflects the ability of GRNmap to

model up-regulation through increasing the estimated production rate (p) of a gene or increasing the activation of production through modulating expression threshold (b) or regulatory weight (w) estimates. In contrast, although GRNmap may decrease production through modulating estimates of p , w , and b , degradation rate is the only value that directly contributes to down-regulation. Further, expression at this earliest 15-minute time point was more significant to model fit than expression at later time points, as indicated by the fact that expression at 30 minutes was not significantly correlated to MSE:minMSE ratio in any multiple regression model. Eigenvector centrality was the second independent variable that exhibited a frequent negative correlation with MSE:minMSE ratio. Because genes with high eigenvector centrality exhibit high influence on overall network dynamics, it is possible that modeling their expression well is important to overall network fit during optimization.

Construction and Analysis of a Consolidated GRN Controlling the Early Response to Cold Shock

A consolidated, medium-scale gene regulatory network synthesizing the key genes and regulatory relationships identified in the family of six database-derived networks was constructed. The network of core motifs conserved across five out of the six database-derived GRNs (Figure 8D), which contained 7 genes and 11 edges, was selected as a starting point. Additional transcription factors were considered for inclusion in the consolidated network if they were conserved across multiple GRNs, exhibited high eigenvector centrality values in the database-derived networks, or if supporting literature substantiating their inclusion was present. Transcription factors were excluded from consideration if they exhibited any of the following: no connections to the consolidated network, only connections of negligible magnitude to the consolidated network, or negligible \log_2 expression changes in the cold shock microarray data.

The genes *HAP4* and *SFP1* were selected for inclusion on the basis of their appearance in five database-derived GRNs and consistent connections to the core network pictured in Figure 8D. Further, *SFP1* exhibited a high eigenvector centrality >0.6 in four of the GRNs. Although *GCR2* also was present in five GRNs, it was excluded due to its minimal \log_2 expression change in the microarray data. Next, genes appearing in four of the database-derived GRNs were considered for inclusion. *ZAPI* was excluded, as it did not exhibit any connections to the consolidated network. Although the sole target of *ZAPI* in the six database-derived networks, *ACE2*, was later included in the network, other lines of evidence support the exclusion of *ZAPI*. In particular, a single edge deletion experiment in db5 showed that deletion of $ZAPI \rightarrow ACE2$ lowered the LSE:minLSE ratio of the modified network moreso than any other edge deletion, demonstrating that inclusion of *ZAPI* impairs overall model fit as was previously hypothesized (Dahlquist Lab, unpublished data). *ACE2* and *SWI5* were also represented in four of the database-derived GRNs and are paralogs. Although *ACE2* was not connected to the network, it was noted that *MCM1* activates *ACE2/SWI5* and is highly connected to other factors in the network of core motifs. For instance, the MADS box protein Mcm1 is known to interact with Yhp1/Yox1 (Pramila et al., 2002), which it activates in db3. *MCM1* is also among the genes activated by *HMO1* in the GRNs. Thus, *ACE2* and *SWI5* were connected to the consolidated network with the addition of *MCM1*, which formed two additional I1-FFLs terminating on *YHPI/YOX1*. The list of unique TFs represented in the family of six GRNs was screened for any additional paralogs (Table 2). The only other set identified was *MSN2/MSN4*, with the latter gene appearing in two of the GRNs. Like *MSN2* and *YHPI*, *MSN4* is also important to induction of the general environmental stress response (Gasch et al., 2000) and is activated by *HMO1* in the

GRNs. Given its consistency with patterns already observed in the consolidated network, *MSN4* was selected for inclusion as the thirteenth gene in the network.

Next, the genes *MGA2* and *STB5* that were present in three of the database-derived GRNs were screened for inclusion. *MGA2* was only connected to the network through its regulation by *GLN3*—an edge of negligible weight in the GRNs—so it was excluded. In contrast, *STB5* exhibited multiple connections to the consolidated network, which included its consistent repression by *CIN5*. Further, $\Delta stb5$ mutants are known to exhibit a cold-sensitive phenotype (Akache and Turcotte, 2001). Thus, *STB5* was included in the network. Finally, genes with high eigenvector centrality in the database-derived networks were considered on the basis of their high influence on network dynamics and suggested importance to overall model fit (Table 6). Notably, *GCN4* exhibited the maximum eigenvector centrality value of 1 in both of the GRNs where it was represented (Figure 9), so it was included in the consolidated network. With this addition, a medium-scale consolidated network consisting of 15 genes and 34 edges had been reached.

The consolidated network was titled db7 and subsequently analyzed. It was hypothesized that the types of regulatory relationships existing between genes in db7 would reflect the most frequent modes of regulation for these relationships represented in the family of six database-derived networks (Table 5). Thus, an adjacency matrix representing db7 was constructed in which initial guesses of 1 and -1 were entered for relationships that were hypothesized to exhibit either activation or repression, respectively. The hypothesized relationships in db7 were then modeled using GRNsight (Figure 10). Several key observations were made based on the GRNsight visualization. First, the core network of motifs previously described (Figure 8D) was visible at the center of db7, but with the addition of two addition I1-FFLs initiated by *MCM1* that

are mediated by *SWI4* and terminate on *YHP1/YOX1*. Interestingly, these I1-FFLs represent a mirror image of those initiated by *MSN2* at the core of the network. Further, *MCM1* was noticed to activate two sets of paralogs: *YHP1/YOX1* and *ACE2/SWI5*. *MCM1* itself is up-regulated by initial activation of *HMO1*, which also activates *MSN2/MSN4*, *CIN5*, *HAP4*, and *YOX1*. Finally, the induction of *MSN2/MSN4* and *YHP1* in db7 indicated overlap with the general ESR.

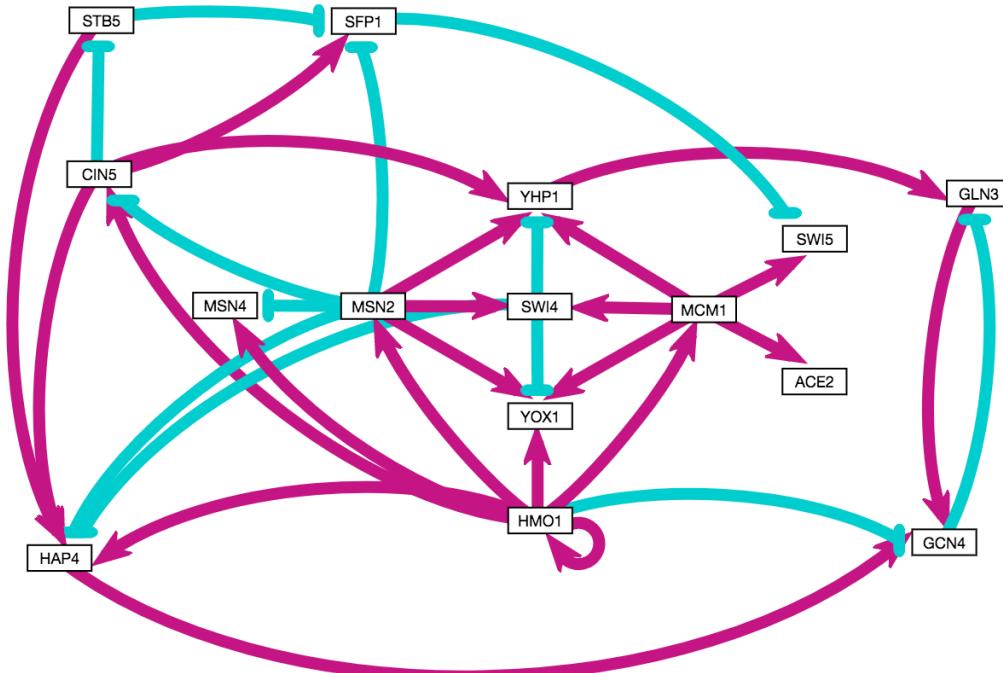


Figure 10. GRNsight visualization of the hypothesized regulatory relationships connecting db7. Boxes indicate genes and arrows indicate edges, with pointed arrowheads representing activation (magenta) and flat arrowheads repression (cyan). Line thicknesses are uniform as edge weights were all estimated to be either 1 or -1.

Db7 was modeled using GRNmap to assess goodness-of-fit to experimental expression data, estimate parameters, and test regulatory hypotheses. Notably, the LSE:minLSE ratio for db7, 1.3418, indicated a better fit to input microarray data than was observed for any database-derived network except db4 or any random network. It is possible that db4 performed better due to the inclusion of *TEC1*, which exhibited an in-degree of 4, an out-degree of 5, and an eigenvector centrality of 0.7237. *TEC1* was excluded from db7 despite its high eigenvector

centrality because it did not appear in more than one database-derived GRN. The estimated edge weights, production rates, and expression thresholds for db7 are provided in Table 7.

Table 7. Estimated edge weights (w), production rates (p), and expression thresholds (b) for db7. Edges with weights that differed from the hypothesized regulation types for db7 are marked with asterisks (*).

Edge	w	Gene	p	b
<i>CIN5</i> → <i>HAP4</i>	0.7011	<i>ACE2</i>	0.2097	0.7535
<i>CIN5</i> → <i>SFP1</i>	0.3487	<i>CIN5</i>	1.1807	3.7825
<i>CIN5</i> → <i>STB5</i>	-0.5721	<i>GCN4</i>	0.2216	2.8230
<i>CIN5</i> → <i>YHP1</i>	0.2240	<i>GLN3</i>	0.3229	0.6674
<i>GCN4</i> → <i>GLN3</i>	-1.0833	<i>HAP4</i>	1.5752	0.7067
<i>GLN3</i> → <i>GCN4</i>	0.8479	<i>HMO1</i>	0.4869	1.2551
<i>HAP4</i> → <i>GCN4</i>	0.4985	<i>MCM1</i>	0.1887	0.2538
<i>HMO1</i> → <i>CIN5</i> *	-0.2762	<i>MSN2</i>	0.5680	0.8260
<i>HMO1</i> → <i>GCN4</i>	-0.0391	<i>MSN4</i>	2.1222	-0.8784
<i>HMO1</i> → <i>HAP4</i> *	-0.2015	<i>SFP1</i>	0.5820	-0.1529
<i>HMO1</i> → <i>HMO1</i>	0.4980	<i>STB5</i>	0.1472	-1.4572
<i>HMO1</i> → <i>MCM1</i>	1.3387	<i>SWI4</i>	1.7181	0.8277
<i>HMO1</i> → <i>MSN2</i>	0.4946	<i>SWI5</i>	0.9660	0.0131
<i>HMO1</i> → <i>MSN4</i>	1.4124	<i>YHP1</i>	0.2373	3.8721
<i>HMO1</i> → <i>YOX1</i>	0.4154	<i>YOX1</i>	0.8584	2.9780
<i>MCM1</i> → <i>ACE2</i>	0.4250			
<i>MCM1</i> → <i>SWI4</i>	3.5546			
<i>MCM1</i> → <i>SWI5</i>	0.8853			
<i>MCM1</i> → <i>YHP1</i>	1.8926			
<i>MCM1</i> → <i>YOX1</i>	0.5814			
<i>MSN2</i> → <i>CIN5</i> *	2.2913			
<i>MSN2</i> → <i>HAP4</i>	-0.5291			
<i>MSN2</i> → <i>MSN4</i>	-3.6916			
<i>MSN2</i> → <i>SFP1</i> *	3.0566			
<i>MSN2</i> → <i>SWI4</i> *	-5.6459			
<i>MSN2</i> → <i>YHP1</i> *	-1.2556			
<i>MSN2</i> → <i>YOX1</i> *	-0.7453			
<i>SFP1</i> → <i>SWI5</i>	-2.1215			
<i>STB5</i> → <i>HAP4</i>	2.1285			
<i>STB5</i> → <i>SFP1</i>	-1.5301			
<i>SWI4</i> → <i>HAP4</i>	-4.8168			
<i>SWI4</i> → <i>YHP1</i> *	1.5471			
<i>SWI4</i> → <i>YOX1</i> *	1.6049			
<i>YHP1</i> → <i>GLN3</i>	2.1261			

Twenty-five of the thirty-four hypothesized regulation types for db7 were supported by edge weight estimates. Of the 9 regulatory relationships that were incorrectly predicted, 8 exhibited inconsistent edge weights that were variously modeled as either activation or repression in the database-derived GRNs. The only exception, the edge *HMO1*→*CIN5*, was of low magnitude in

the database-derived GRNs and in db7. Further, it was noted that 5 of the 9 incorrectly predicted regulatory relationships involved the FFLs mediated by *SWI4* and terminating on *YHP1/YOX1*.

The estimated edge weights for db7 following GRNmap modeling were visualized using GRNsight (Figure 11). Notably, the revised modeling of five regulatory relationships involved in the four central feedforward loops regulating *YHP1/YOX1* changed the structure of these FFLs. It was hypothesized that each was an incoherent type I FFL (Figure 10). Instead, GRNmap modeled the feedforward loops initiated by *MCM1* as coherent type I FFLs, resulting in the overall activation of *YHP1/YOX1*. Conversely, the feedforward loops initiated by *MSN2* were modeled as coherent type II (C2) FFLs, resulting in the net repression of *YHP1/YOX1*. As a result, *YOX1/YHP1* expression is regulated more dynamically than was hypothesized, with *MCM1* and *MSN2* inputting activating and inhibitory signals, respectively.

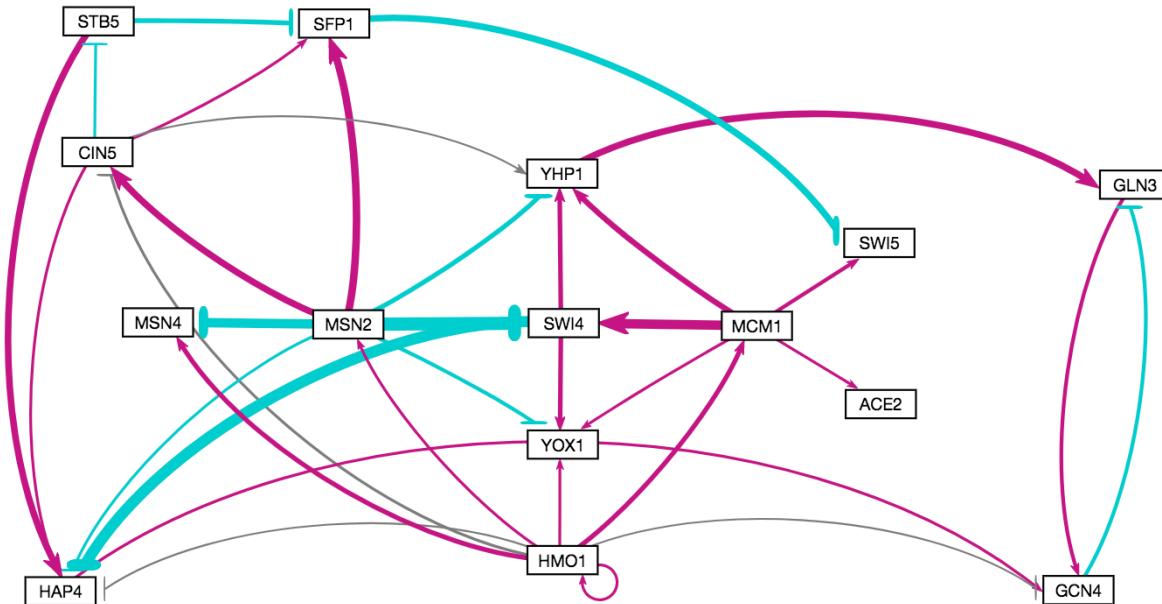


Figure 11. GRNsight visualization of db7 showing edge weights estimated by GRNmap. Boxes indicate genes and arrows indicate edges, with pointed arrowheads representing activation (magenta) and flat arrowheads repression (cyan). Edges colored in gray exhibit small influence, meaning their associated weights are <5% of the highest magnitude edge weight in the network, *MSN2*→*SWI4*.

To determine which transcription factors in db7 exhibit the highest influence on network dynamics, eigenvector centrality values were computed for each transcription factor (Table 8). *GCN4* was noted to once again exhibit the highest eigenvector centrality value of 1, as in the database-derived networks. *GLN3* also exhibited a high eigenvector centrality value in db7 (0.9258) as well as in the database-derived networks, suggesting this gene may play an important role in the early response to cold shock in yeast.

Table 8. Eigenvector centrality values for the 15 transcription factors represented in db7.

Gene	Eigenvector Centrality
<i>ACE2</i>	0.0594
<i>CIN5</i>	0.1188
<i>GCN4</i>	1.0000
<i>GLN3</i>	0.9258
<i>HAP4</i>	0.4415
<i>HMO1</i>	0.0594
<i>MCM1</i>	0.0594
<i>MSN2</i>	0.0594
<i>MSN4</i>	0.1188
<i>SFP1</i>	0.2697
<i>STB5</i>	0.1125
<i>SWI4</i>	0.1188
<i>SWI5</i>	0.2853
<i>YHP1</i>	0.3437
<i>YOX1</i>	0.2907

DISCUSSION:

We have demonstrated the use of GRNmap to model the dynamics of a family of six database-derived gene regulatory networks controlling the early response to cold shock in *Saccharomyces cerevisiae*. Microarray data from cold shock experiments performed in wild-type yeast and in five transcription factor deletion strains at 13°C was obtained, and lists of genes that exhibited significant log₂ expression changes in each experiment were determined to allow for derivation of the networks. The six networks, db1-db6, were constructed by identifying transcription factors (TFs) that had their targets enriched in these lists using the YEASTRACT database and then connecting the TFs into GRNs. Db1-db6 ranged in size from 14-17 genes and 25-35 edges.

Following initial GRNmap modeling, L-curve analyses were used to determine that efficient regularization was achieved at $\alpha=0.002$ for these medium-scale GRNs. Using this value of alpha, GRNmap was successfully used to estimate network parameters for db1-db6 by fitting differential equation models to the experimental \log_2 expression data using a penalized least squares approach. In addition, thirty random networks with the same genes as db5 but with randomized edges were modeled in GRNmap. In general, the database-derived network outperformed the random networks, validating our experimental approach.

Comparative analysis of db1-db6 allowed for the identification of TFs and inference of network properties important for regulating the early response to low temperatures in yeast. The most prominent conserved element of the database-derived networks was initial activation of the genes *CIN5*, *MSN2*, and *YOX1* by Hmo1, which further activates itself while receiving no other inputs in most networks (Figure 6). In the cell, Hmo1 activity is necessary for TORC1-dependent regulation of ribosome biogenesis (Berger et al., 2007; Xiao and Grove, 2009). Due to mRNA and ribosome stability at low temperatures, there is an immediate need for *de novo* ribosome synthesis during cold shock, which explains the initial activating role of *HMO1* in the database-derived GRNs (Al-Fageeh and Smales, 2006; Aguilera et al., 2007). Further lines of evidence supporting the role of *HMO1* in this network include the observations that $\Delta hmo1$ mutants are severely growth-impaired at 13°C (Dahlquist Lab, unpublished data), that $\Delta hmo1$ mutants are highly susceptible to inviability due to freeze-thaw cycles (Kasahara et al., 2008; Ando et al., 2006), and that Hmo1 expression may confer environmental stress resistance through functioning as a linker histone (Panday and Grove, 2016).

Db1-db6 consistently featured induction of Msn2, a key regulator of the general environmental stress response (ESR) described by Gasch et al. (2000). In the cell, Msn2 aids in

the induction of the ESR through recognizing the stress response element promotor sequence (Martinez-Pastor et al., 1996). In the database-derived networks, Hmo1 induces expression of *MSN2* and its paralog *MSN4*. *MSN2* itself exhibits a high out-degree of between 6 and 9 in the GRNs, forming feedforward loops (FFLs) with *HMO1* that terminate on *CIN5*, *YOX1*, and *HAP4* while also initiating symmetrical FFLs with *SWI4* that end on the homeodomain protein-encoding genes *YHP1* and *YOX1* (Pramila et al., 2002). Interestingly, when present in the GRNs, the MADS box protein-encoding gene *MCM1* also is activated by *HMO1* and forms symmetrical FFLs with *SWI4* terminating on the genes *YHP1* and *YOX1* (Grueneberg et al., 1992). Together, *MSN2* and *MCM1* form four incoherent type I FFLs mediated by *SWI4* that regulate *YHP1* and *YOX1* activity while also composing two coherent type I FFLs ending on *YOX1*. Finally, *MCM1* alone upregulates the paralogs *ACE2* and *SWI5*. The high level of redundancy demonstrated by symmetrical motifs and multiple paralogs suggests that backup circuitry exists to support the important function of these transcription factors (Kafri et al., 2005; Gitter et al., 2009), which in fact all regulate the cell cycle with the exception of *Msn2* (Simon et al., 2001).

Mcm1-mediated cell cycle control appears as another important feature of the GRN controlling the early response to cold shock in yeast. *Mcm1* itself controls the G2/M transition (Althoefer et al., 1995), whereas its targets *Ace2* and *Swi5* regulate the M/G1 transition (Simon et al., 2001). Finally, *Swi4*, *Yhp1*, and *Yox1* time the G1/S transition, with *Swi4* activity serving as the rate-limiting step (Pramila et al., 2001). In the cold shock expression data analyzed in this study, it was observed that *HMO1* induces steady up-regulation of *MCM1*, which subsequently derepresses *SWI4*, *SWI4*, *YHP1*, and *YOX1*. It is possible that this could represent a compensatory phase of growth following cold adaptation through *de novo* ribosome synthesis. However, many of these cell cycle TFs are multifunctional and are known to also regulate genes

important to cold shock. The fact that $\Delta swi4$ mutants exhibit a cold-sensitive phenotype underscores this point, although the exact mechanism by which Swi4 promotes cold tolerance is unknown (Córcoles-Sáez et al., 2012). The secondary functions of Yhp1 and Yox1 at low temperatures have been better characterized, with Yox1 known to regulate genes involved with translation and Yhp1 with cell wall deposition (Horak et al., 2002). The protective effect of Yhp1 and Yox1 activity at low temperatures has been indirectly suggested by the fact that exposure of yeast to cryoprotectants such as trehalose induces expression of those genes (Momse et al., 2010).

Beyond the subnetwork described above lie several other transcription factors that are noteworthy for their contributions to cold resistance, network properties, and novelty. The only other TF appearing in db1-db6 whose deletion results in a documented cold-sensitive phenotype is Stb5 (Akache and Turcotte, 2001), which is better known for the role it plays in conferring resistance to oxidative stress (Larochelle et al., 2006). *STB5* appears in three of the GRNs, where it is repressed by *CIN5*, activates *HAP4*, and represses *SFP1*. *SFP1* is found in five of the database-derived GRNs, where it is also regulated by *MSN2* and consistently represses *SWI5*. Like Hmo1, Sfp1 is also involved in TORC1 signaling, where it contributes to the downstream feedback regulation of ribosome biogenesis in response to low temperature (Lempäinen et al., 2009). Of note, Sfp1 is one of five TFs to exhibit an eigenvector centrality >0.7 in two or more GRNs, suggesting its importance to overall network dynamics. Interestingly, the transcription factor Gcn4, which activates synthesis of amino acid biosynthetic enzymes in response to stress, possesses the maximal eigenvector centrality value of 1 each time it appears (Hinnebusch and Natarajan, 2002). However, this transcription factor has been extensively studied and its regulatory relationships might therefore be overrepresented in the YEASTRACT database.

compared to other TFs. Finally, the remaining TFs Gln3 and Hap4 were studied by the Dahlquist Lab through transcription factor deletion strains and have not been implicated in cold shock until now.

Consistent determinants of goodness-of-fit for transcription factors in both the database-derived and random networks were identified through multiple regression analysis. Despite assessing numerous model inputs, outputs, and graph statistics, only high expression at the 15 minute time point and high eigenvector centrality were correlated repeatedly with low MSE:minMSE ratio. The former case illustrates that the current iteration of GRNmap more easily models up-regulation rather than down-regulation. This is because overall production of a gene can be modulated through estimates of production rate, expression threshold, and regulatory weights, whereas degradation is modeled solely through a fixed degradation rate. The latter case is noteworthy given that eigenvector centrality was consistently correlated to TF goodness-of-fit, but its correlates, which include in-degree and betweenness centrality, were not. Thus, TFs with high eigenvector centrality appear to have been modeled well due to their high influence on overall GRN dynamics, rather than because of their degree or relative position. This is consistent with the observation that compared to other graph statistics, eigenvector centrality is well-suited to the identification of genes that cause pathological phenotypes when mutated (Özgür et al., 2008; Davis et al., 2010). In contrast, other factors including experimental indicators of node significance such as Benjamini and Hochberg corrected ANOVA *p*-values did not correlate to node MSE:minMSE ratio (Pearson's R, r=0.0579, n=16, p=0.8313). Overall, these multiple regression results indicate that high quality model fit was dependent on inclusion of relevant TFs and their high influence regulators, rather than on other incidental variables.

A consolidated GRN, db7, was inferred from comparative analysis of db1-db6, graph statistics, and published literature (Figure 11). Db7 supports our previous conclusion that the ESR overlaps with the early response to cold shock (Dahlquist et al., 2015), and not only the late response to low temperature (Schade et al., 2004; Kandror et al., 2004). Specifically, early induction of Hmo1 was shown to activate the Msn2/Msn4-dependent environmental stress response. However, the general ESR results in repression of genes involved in mRNA and ribosome biogenesis (Gasch et al., 2000), as well as slowing of the cell cycle (Brauer et al., 2008). Db7 appears to circumvent this problem by preceding Msn2/Msn4 up-regulation with strong Hmo1-dependent TORC1 signaling and initial *de novo* ribosome biogenesis, resulting in cold adaptation. This is followed by parallel signaling through *MCM1* and *MSN2//MSN4*. Although *MSN2* induces the repression of the cell cycle regulators *YHP1/YOX1* through two coherent type II FFLs, the *MCM1*-axis opposes this signaling by derepressing *YHP1/YOX1* through two coherent type I FFLs. *MCM1* also activates the cell cycle regulators *ACE2/SWI5*. The activity of *MCM1* may result in both a compensatory resumption of growth following ribosome biogenesis and activation of the secondary functions of *YHP1/YOX1* that confer cold tolerance.

The consolidated network proposed here suggests several fruitful opportunities for future research. One direction could involve investigating the role played by the paralogs *YHP1/YOX1* and *ACE2/SWI5* in regulating the response to cold shock through the study of double deletion mutants. The $\Delta yhp1\Delta yox1$ mutant strain is known to be viable (Pramila et al., 2002), for instance, although we are not aware of any previous studies that have investigated the growth of this strain at low temperatures. Similarly, multiple stressors experiments could be implemented to tease out the secondary functions of these important cell cycle regulators in stress responses, or of

regulators in this network that have predominantly been associated with regulating responses to other types of stressors (e.g. Stb5 and oxidative stress). Finally, it would be interesting to repeat the cold shock experiments of Schade et al. (2004) with the $\Delta msn2\Delta msn4$ double deletion strain in more detail, now that we have presented two separate lines of evidence suggesting that Msn2/Msn4-dependent signalling is important to the early cold response in yeast.

In conclusion, GRNmap was used to reliably model and systematically assess six database-derived networks regulating the early response to cold shock in yeast. Qualitative assessment and visual modeling of motifs in GRNsight was implemented alongside quantitative and graph statistical analysis to infer intrinsic properties of the network. A final, consolidated GRN exhibited complex motifs, such as multiple instances of symmetrical signaling through paralogs, and harmonized well with previous reports of the molecular mechanisms underlying the cold response in yeast. Further, the GRN highlighted novel aspects of the early response to cold shock, including overlap with the ESR mediated through Msn2/Msn4-dependent signaling and the implication of new transcription factors in cold shock. Overall, these results demonstrate how modeling through GRNmap and GRNsight offer powerful and broadly applicable *in silico* tools that can lead to the development of testable biological hypotheses for *in vivo* study.

ACKNOWLEDGMENTS:

I am grateful to Dr. Kam D. Dahlquist and Dr. Ben G. Fitzpatrick for their consistent mentorship, guidance, and support throughout the course of this investigation. I would like to specifically thank Dr. Dahlquist for introducing me to the GRNmap and GRNsight projects, providing biological background information, and sharing best practices for the generation and curation of open-source data. I also would like to thank Dr. Fitzpatrick for providing advice on the use of the

statistical software R, explaining the mathematical background information important to this project, and helping in the interpretation of statistical results.

This research was a collaborative effort that was made possible thanks to the contributions of current and former members of the Dahlquist Lab. Of the various groups that compose the Dahlquist Lab, I first would like to thank my fellow members of the Data Analysis team. In particular, I am grateful to Natalie Williams for her mentorship and collaboration on the initial analysis of the database-derived and random networks; and to Maggie O’Neil for her development of the graph statistical analysis protocol used in this project. I would also like to thank Lauren Kelly, Tessa Morris, K. Grace Johnson, and Kristen Horstmann for their contributions to the analyses presented in this study. Within the GRNmap coding team, I would like to thank Juan Carrillo, Trixie Roque, Eddie Azinge, Justin Kyle Torres, and John Lopez for developing and constantly improving the GRNmap modeling software upon which this study is based. I am further grateful to Anindita Varshneya, Nicole Anguiano, Eileen Choe, Mihir Samdarshi, and Jen Shin for their work on the GRNsight web application. Finally, I would like to acknowledge Monica Hong, Nika Vafadari, and Katherine Scheker for their work performing the cold shock microarray experiments that were analyzed in this study. All of the students mentioned above have made invaluable contributions to this work.

REFERENCES:

- Aguilera, J., Randez-Gil, F., and 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.
- Akache, B., Wu, K., and Turcotte, B. (2001). Phenotypic analysis of genes encoding yeast zinc cluster proteins. *Nucleic Acids Research*, 29(10), 2181-2190. DOI: 10.1093/nar/29.10.2181.
- Al-Fageeh, M.B., and Smales, C.M. (2006). Control and regulation of the cellular responses to cold shock: the responses in yeast and mammalian systems. *Biochemical Journal*, 397(2), 247-259. DOI: 10.1042/BJ20060166.

- Althofer, H., Schleiffer, A., Wassmann, K., Nordheim, A., and Ammerer, G. (1995). Mcm1 is required to coordinate G2-specific transcription in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*, 15(11), 5917-5928. DOI: 10.1128/MCB.15.11.5917.
- Ando, A., Nakamura, T., Murata, Y., Takagi, H., and Shima, J. (2006). Identification and classification of genes required for tolerance to freeze-thaw stress revealed by genome-wide screening of *Saccharomyces cerevisiae* deletion strains. *FEMS Yeast Research*, 7(2), 244-253. DOI: 10.1111/j.1567-1364.2006.00162.x.
- Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. *Icwsm*, 8, 361-362.
- Becskei, A., and Serrano, L. (2000). Engineering stability in gene networks by autoregulation. *Nature*, 405(6786), 590-593. DOI: 10.1038/35014651.
- Benjamini, Y., and 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)*, 289-300.
- Berger, A.B., Decourty, L., Badis, G., Nehrbass, U., Jacquier, A., and Gadal, O. (2007). Hmo1 is Required for TOR-Dependent Regulation of Ribosomal Protein Gene Transcription. *Molecular and Cellular Biology*, 27(22), 8015-8026. DOI: 10.1128/MCB.01102-07.
- Bonacich, P., & Lloyd, P. (2001). Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*, 23(3), 191-201. DOI: 10.1016/S0378-8733(01)00038-7.
- Brauer, M.J., Huttenhower, C., Airoldi, E.M., Rosenstein, R., Matese, J.C., Gresham, D., ... & Botstein, D. (2008). Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast. *Molecular Biology of the Cell*, 19(1), 352-367. DOI: 10.1091/mbc.E07-08-0779.
- Chen, K.C., Wang, T.Y., Tseng, H.H., Huang, C.Y.F., & Kao, C.Y. (2005). A stochastic differential equation model for quantifying transcriptional regulatory network in *Saccharomyces cerevisiae*. *Bioinformatics*, 21(12), 2883-2890. DOI: 10.1093/bioinformatics/bti415.
- Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., ... and Fisk, D.G. (2011). Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Research*, 40(D1), D700-D705. DOI:10.1093/nar/gkr1029.
- Córcoles-Sáez, I., Ballester-Tomas, L., Maria, A., Prieto, J.A., and Randez-Gil, F. (2012). Low temperature highlights the functional role of the cell wall integrity pathway in the regulation of growth in *Saccharomyces cerevisiae*. *Biochemical Journal* 446(3), 477-488. DOI: 10.1042/BJ20120634.
- Costanzo, M., VanderSluis, B., Koch, E.N., Baryshnikova, A., Pons, C., Tan, G., ... and Pelechano, V. (2016). A global genetic interaction network maps a wiring diagram of cellular function. *Science*, 353(6306), aaf1420. DOI: 10.1126/science.aaf1420.
- Dahlquist, K.D., Dionisio, J.D.N., Fitzpatrick, B.G., Anguiano, N.A., Varshneya, A., Southwick, B.J., and Samdarshi, M. (2016). GRNsight: a web application and service for visualizing models of small-

to medium-scale gene regulatory networks. *PeerJ Computer Science*, 2, e85. DOI: 10.7717/peerj-cs.85.

Dahlquist, K.D., Fitzpatrick, B.G., Camacho, E.T., Entzinger, S.D., and Wanner, N.C. (2015). Parameter Estimation for Gene Regulatory Networks from Microarray Data: Cold Shock Response in *Saccharomyces cerevisiae*. *Bulletin of Mathematical Biology*, 77(8), 1457-1492. DOI: 10.1007/s11538-015-0092-6.

Davis, N. A., Crowe Jr, J. E., Pajewski, N. M., and McKinney, B. A. (2010). Surfing a genetic association interaction network to identify modulators of antibody response to smallpox vaccine. *Genes and Immunity*, 11(8), 630-636. DOI: 10.1038/gene.2010.37.

Freeman, L.C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 35-41.

Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., ... and Brown, P.O. (2000). Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes. *Molecular Biology of the Cell*, 11(12), 4241-4257. DOI: 10.1091/mbc.11.12.4241.

Gitter, A., Siegfried, Z., Klutstein, M., Fornes, O., Oliva, B., Simon, I., and Bar-Joseph, Z. (2009). Backup in gene regulatory networks explains differences between binding and knockout results. *Molecular Systems Biology*, 5(1), 276. DOI: 10.1038/msb.2009.33.

Grueneberg, D.A., Natesan, S., Alexandre, C., and Gilman, M.Z. (1992). Human and *Drosophila* Homeodomain Proteins that Enhance the DNA-Binding Activity of Serum Response Factor. *Science*, 257(5073), 1089-1095. DOI: 10.1126/science.257.5073.1089.

Hampsey, M. (1997). A Review of Phenotypes in *Saccharomyces cerevisiae*. *Yeast*, 13(12), 1099-1133. DOI: 10.1002/(SICI)1097-0061(19970930)13:12<1099::AID-YEA177>3.0.CO;2-7.

Hansen, P.C., and 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. DOI: 10.1137/0914086.

Harbison, C. T., Gordon, D. B., Lee, T. I., Rinaldi, N. J., Macisaac, K. D., Danford, T. W., ... and Jennings, E. G. (2004). Transcriptional regulatory code of a eukaryotic genome. *Nature*, 431(7004), 99-104. DOI: 10.1038/nature02800.

Hinnebusch, A G., and Natarajan, K. (2002). Gcn4p, a Master Regulator of Gene Expression, is Controlled at Multiple Levels by Diverse Signals of Starvation and Stress. *Eukaryotic Cell*, 1(1), 22-32. DOI: 10.1128/EC.01.1.22-32.2002.

Holoch, D., and Moazed, D. (2015). RNA-mediated epigenetic regulation of gene expression. *Nature Reviews Genetics*, 16(2), 71-84. DOI: 10.1038/nrg3863.

Horak, C.E., Luscombe, N.M., Qian, J., Bertone, P., Piccirillo, S., Gerstein, M., and Snyder, M. (2002). Complex transcriptional circuitry at the G1/S transition in *Saccharomyces cerevisiae*. *Genes and Development*, 16(23), 3017-3033. DOI: 10.1101/gad.1039602.

Ingolia, T.D., Slater, M.R., and Craig, E.A. (1982). *Saccharomyces cerevisiae* contains a complex multigene family related to the major heat shock-inducible gene of *Drosophila*. *Molecular and Cellular Biology*, 2(11), 1388-1398. DOI: 10.1128/MCB.2.11.1388.

- Kafri, R., Bar-Even, A., and Pilpel, Y. (2005). Transcription control reprogramming in genetic backup circuits. *Nature Genetics*, 37(3), 295. DOI: 10.1038/ng1523.
- Kandror, O., Bretschneider, N., Kreydin, E., Cavalieri, D., and Goldberg, A.L. (2004). Yeast Adapt to Near-Freezing Temperatures by STRE/Msn2, 4-Dependent Induction of Trehalose Synthesis and Certain Molecular Chaperones. *Molecular Cell*, 13(6), 771-781. DOI: 10.1016/S1097-2765(04)00148-0.
- Karlebach, G., and Shamir, R. (2008). Modelling and analysis of gene regulatory networks. *Nature Reviews Molecular Cell Biology*, 9(10), 770-780. DOI: 10.1038/nrm2503.
- Kasahara, K., Ki, S., Aoyama, K., Takahashi, H., and Kokubo, T. (2008). *Saccharomyces cerevisiae* HMO1 interacts with TFIID and participates in start site selection by RNA polymerase II. *Nucleic Acids Research*, 36(4), 1343-1357. DOI: 10.1093/nar/gkm1068.
- Koschützki, D., & Schreiber, F. (2008). Centrality analysis methods for biological networks and their application to gene regulatory networks. *Gene Regulation and Systems Biology*, 2, GRSB-S702. DOI: 10.4137/GRSB.S702.
- Larochelle, M., Drouin, S., Robert, F., and Turcotte, B. (2006). Oxidative Stress-Activated Zinc Cluster Protein Stb5 Has Dual Activator/Repressor Functions Required for Pentose Phosphate Pathway Regulation and NADPH Production. *Molecular and Cellular Biology*, 26(17), 6690-6701. DOI: 10.1128/MCB.02450-05.
- Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., ... and Zeitlinger, J. (2002). Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*, 298(5594), 799-804. DOI: 10.1126/science.1075090.
- Lempiäinen, H., Uotila, A., Urban, J., Dohnal, I., Ammerer, G., Loewith, R., and Shore, D. (2009). Sfp1 Interaction with TORC1 and Mrs6 Reveals Feedback Regulation on TOR Signaling. *Molecular Cell*, 33(6), 704-716. DOI: 10.1016/j.molcel.2009.01.034.
- Lillacci, G., & Khammash, M. (2010). Parameter estimation and model selection in computational biology. *PLoS Computational Biology*, 6(3), e1000696. DOI: 10.1371/journal.pcbi.1000696.
- Mangan, S. and Alon, U. (2003). Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences*, 100(21), 11980-11985. DOI: 10.1073/pnas.2133841100.
- Martinez-Pastor, M.T., Marchler, G., Schüller, C., Marchler-Bauer, A., Ruis, H., and Estruch, F. (1996). The *Saccharomyces cerevisiae* zinc finger proteins Msn2p and Msn4p are required for transcriptional induction through the stress response element (STRE). *The EMBO Journal*, 15(9), 2227-2235. DOI: 10.1002/j.1460-2075.1996.tb00576.x.
- Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science*, 298(5594), 824-827. DOI: 10.1126/science.298.5594.824.
- Momose, Y., Matsumoto, R., Maruyama, A., and Yamaoka, M. (2010). Comparative analysis of transcriptional responses to the cryoprotectants, dimethyl sulfoxide and trehalose, which confer

tolerance to freeze–thaw stress in *Saccharomyces cerevisiae*. *Cryobiology*, 60(3), 245-261. DOI: 10.1016/j.cryobiol.2010.01.001.

Morano, K.A., Grant, C.M., and Moye-Rowley, W.S. (2012). The Response to Heat Shock and Oxidative Stress in *Saccharomyces cerevisiae*. *Genetics*, 190(4), 1157-1195. DOI: 10.1534/genetics.111.128033.

Narang, V., Ramli, M.A., Singhal, A., Kumar, P., de Libero, G., Poidinger, M., & Monterola, C. (2015). Automated identification of core regulatory genes in human gene regulatory networks. *PLoS Computational Biology*, 11(9), e1004504. DOI: 10.1371/journal.pcbi.1004504.

Neymotin, B., Athanasiadou, R., and Gresham, D. (2014). Determination of in vivo RNA kinetics using RATE-seq. *RNA*, 20(10), 1645-1652. DOI: 10.1261/rna.045104.114.

Özgür, A., Vu, T., Erkan, G., and Radev, D.R. (2008). Identifying gene-disease associations using centrality on a literature mined gene-interaction network. *Bioinformatics*, 24(13), i277-i285. DOI: 10.1093/bioinformatics/btn182

Panday, A., and Grove, A. (2016). The high mobility group protein HMO1 functions as a linker histone in yeast. *Epigenetics and Chromatin*, 9(1), 13. DOI: 10.1186/s13072-016-0062-8.

Pavlopoulos, G.A., Secrier, M., Moschopoulos, C.N., Soldatos, T.G., Kossida, S., Aerts, J., ... and Bagos, P.G. (2011). Using graph theory to analyze biological networks. *BioData Mining*, 4(1), 10. DOI: 10.1186/1756-0381-4-10.

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

Sahara, T., Goda, T., and Ohgiya, S. (2002). Comprehensive Expression Analysis of Time-dependent Genetic Responses in Yeast Cells to Low Temperature. *Journal of Biological Chemistry*, 277(51), 50015-50021. DOI: 10.1074/jbc.M209258200.

Schade, B., Jansen, G., Whiteway, M., Entian, K.D., and 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.

Simon, I., Barnett, J., Hannett, N., Harbison, C.T., Rinaldi, N.J., Volkert, T.L., ... and Young, R.A. (2001). Serial Regulation of Transcriptional Regulators in the Yeast Cell Cycle. *Cell*, 106(6), 697-708. DOI: 10.1016/S0092-8674(01)00494-9.

Teixeira, M.C., Monteiro, P.T., Palma, M., Costa, C., Godinho, C.P., Pais, P., ... and Sá-Correia, I. (2017). YEASTRACT: an upgraded database for the analysis of transcription regulatory networks in *Saccharomyces cerevisiae*. *Nucleic Acids Research*, 46(D1), D348-D353. DOI: 10.1093/nar/gkx842.

Verghese, J., Abrams, J., Wang, Y., and Morano, K. A. (2012). Biology of the Heat Shock Response and Protein Chaperones: Budding Yeast (*Saccharomyces cerevisiae*) as a Model System. *Microbiology and Molecular Biology Reviews*, 76(2), 115-158. DOI: 10.1128/MMBR.05018-11.

- Vu T.T., and Vohradsky J. (2007). Nonlinear differential equation model for quantification of transcriptional regulation applied to microarray data of *Saccharomyces cerevisiae*. *Nucleic Acids Research*, 35, 279–287. DOI:10.1093/nar/gkl1001.
- Wingfield, J.C. (2013). Ecological processes and the ecology of stress: the impacts of abiotic environmental factors. *Functional Ecology*, 27(1), 37-44. DOI: 10.1111/1365-2435.12039.
- Xiao, L., and Grove, A. (2009). Coordination of Ribosomal Protein and Ribosomal RNA Gene Expression in Response to TOR Signaling. *Current Genomics*, 10(3), 198-205. DOI: 10.2174/138920209788185261.
- Youn, J. Y., Friesen, H., Ba, A. N. N., Liang, W., Messier, V., Cox, M. J., ... and Andrews, B. (2017). Functional Analysis of Kinases and Transcription Factors in *Saccharomyces cerevisiae* Using an Integrated Overexpression Library. *G3: Genes, Genomes, Genetics*, 7(3), 911-921. DOI: 10.1534/g3.116.038471.