**Sensitivity Assessments of the GRNmap Modelling Software Reveals that Certain Manipulations of the Model Input Impacts the Fit of the Model**

**Abstract:**

A gene regulatory network (GRN) depicts regulatory relationships between transcription factors, illustrating the regulation of gene expression. The dynamics of GRNs can offer information about the changes in gene expression over time in response to various environmental stimuli. GRNmap (Gene Regulatory Network modeling and parameter estimation) is a MATLAB application that uses ordinary differential equations to estimate GRN parameters including gene expression thresholds, production rates, and regulatory weights from DNA microarray data. DNA microarray data from wild type *Saccharomyces cerevisiae* and transcription factor deletion strains (∆cin5, ∆gln3, ∆hap4, ∆hmo1, ∆zap1) in which the strains were subjected to 13oC cold-shock for 15, 30, and 60 minutes and 30 and 60 minutes of recovery was used as input to the model. A candidate cold-response GRN with 15 nodes (transcription factors) and 28 edges (regulatory relationships) was derived from the YEASTRACT database. Through this study, a series of experiments were run to determine the sensitivity of the model to manipulations of the input files, including structural changes to the candidate GRN, variable inclusion of strain data, changes to the optimization parameters, and fixing or estimating production rates. Through comparing the fit of the model within these experiments, the sensitivity of the model to these changes was assessed. While changes to the optimization parameters did not impact the fit of the model, structural manipulations of the GRN, including edge and node deletions, along with variable inclusion of strain data and fixing the production rates did have an impact on the model. These results reveal that the model is sensitive to manipulations of the GRN dynamics, while it is not sensitive to changes in optimization parameters.

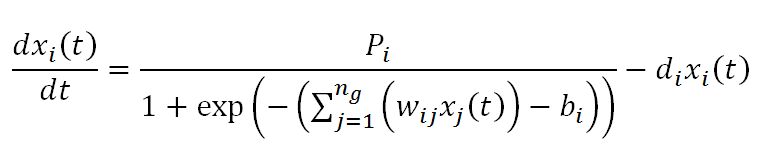
**Introduction**

Microorganisms must be able to adapt to environmental fluctuations in order to survive in natural environments. Environmental stressors, such as temperature, nutrient availability, and acidity, impact the internal homeostasis of the cells, requiring a cellular response to adapt to the changing environment (Schade *et al.,* 2004). In yeast cells, internal responses are necessary after sudden environmental stress in order to maintain growth under the specific circumstances (Gasch *et al.,* 2000). The environmental stress response (ESR) is initiated as the cells react to the changes in the environment causing an alteration in the expression of hundreds to thousands of genes in the yeast genome, specific to the type of environmental stress exposed to the cells. Changes to the expression of certain genes allows for the cells to return to a state of homeostasis, allowing for continued cell growth and division (Gasch *et al.,* 2000).

The optimum growth temperature of budding yeast, *Saccharomyces cerevisiae,* is 30oC, with impaired growth at suboptimal temperatures (Salvadó *et al.,* 2011). Drastic temperature fluctuations affect the cellular functions and characteristics of yeast, impacting the ability for the cells to divide and grow. When grown in environments with suboptimal temperatures, *S. cerevisiae* responds with a change to the transcriptional control of various sets of proteins (Gasch *et al.,* 2002). For example, the response of *S. cerevisiae* to heat shock has been widely studied, showing an effect on the transcriptional control with an induction of a set of heat shock proteins (Mager, 1995). However, the response of yeast to cold shock is less widely studied and the effects on gene expression is not completely understood. Under cold shock conditions, yeast cells are grown at temperatures that hinder, but still allow for cell growth and division (Tai *et al.,* 2007). The cold shock response can be divided into the early cold response (ECR), which constitutes the first twelve hours of growth at low temperature, and the late cold response (LCR), which is the response after twelve hours of growth (Tai *et al.,* 2007). The late cold response has been shown to be regulated by the transcription factors Msn2 and Msn4, while there have been no known main transcription factors that regulate the early cold (Schade et al., 2004). Under cold conditions, decreases in membrane fluidity and membrane-bound enzymes have been shown to cause issues in membrane diffusion (Al-Fageeh *et al.,* 2005). In addition, at low temperatures, protein synthesis has been shown to decrease due to changes in the internal cellular processes (Al-Fageeh *et al.,* 2005). When returned to the optimal growth temperature, yeast cells respond with downregulation of ESR genes, indicating cold shock response recovery (Gasch, 2002).

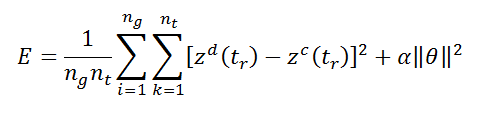
*Saccharomyces cerevisiae* is an ideal model organism for use in systems biology research. Due to the relatively small genome of about 6,000 genes and the availability of molecular genetic tools and datasets, research studies are more accessible and easier to conduct. One such tool is DNA microarray technology, with which the global transcriptional response of a *S. cerevisiae* wild-type strain to cold shock and recovery can be analyzed. DNA microarray technology allows for the determination of the induction or repression of certain genes within a genome (*NIH*, 2015). This technology allows for the visualization of transcriptional changes to the yeast genome after exposure to cold shock through colored fluorescence (Schena, 1996). The 6,189 genes in the yeast genome are included on the DNA microarray slide, allowing for the transcriptional response in each of the genes to be determined. Using this technology, the impact of cold shock on the expression of genes in the yeast genome can be determined.

A gene regulatory network (GRN) is a set of transcription factors which regulate the level of expression of genes encoding other transcription factors. The dynamics of a GRN show how gene expression in the network changes over time. GRNmap (Gene Regulatory Network modeling and parameter estimation) is a MATLAB application that models the change in expression for each gene in the GRN as the production of mRNA minus its degradation using ordinary differential equations with a sigmoidal production function (Equation 1) where *Pi* is the mRNA production rate of the gene, *di* is the mRNA degradation rate of the gene, *b* is the expression threshold, and *w* is the regulatory weight of the gene (Dahlquist et al., 2015)



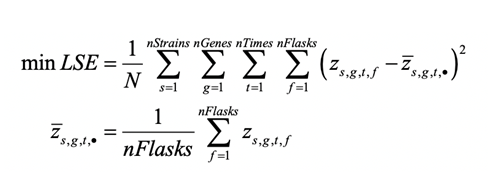
**Equation 1:** Differential equation used in GRNmap which models expression as production minus degradation with a sigmoidal production function.

Given published mRNA degradation rates (Neymotin et al. 2014) and the cold shock microarray data from the Dahlquist lab (t15. t30, t60 for wild type, *Δcin5, Δgln3, Δhap4, Δhmo1,* and *Δzap1* strains), GRNmap then estimates the production rates and expression thresholds for each gene, and regulatory weights for each edge, which denote the direction (activation or repression) and strength of the regulatory relationships. The model uses a penalized least squares approach to estimation (Equation 2). The least squares error assesses the difference between the experimental and simulated model values.



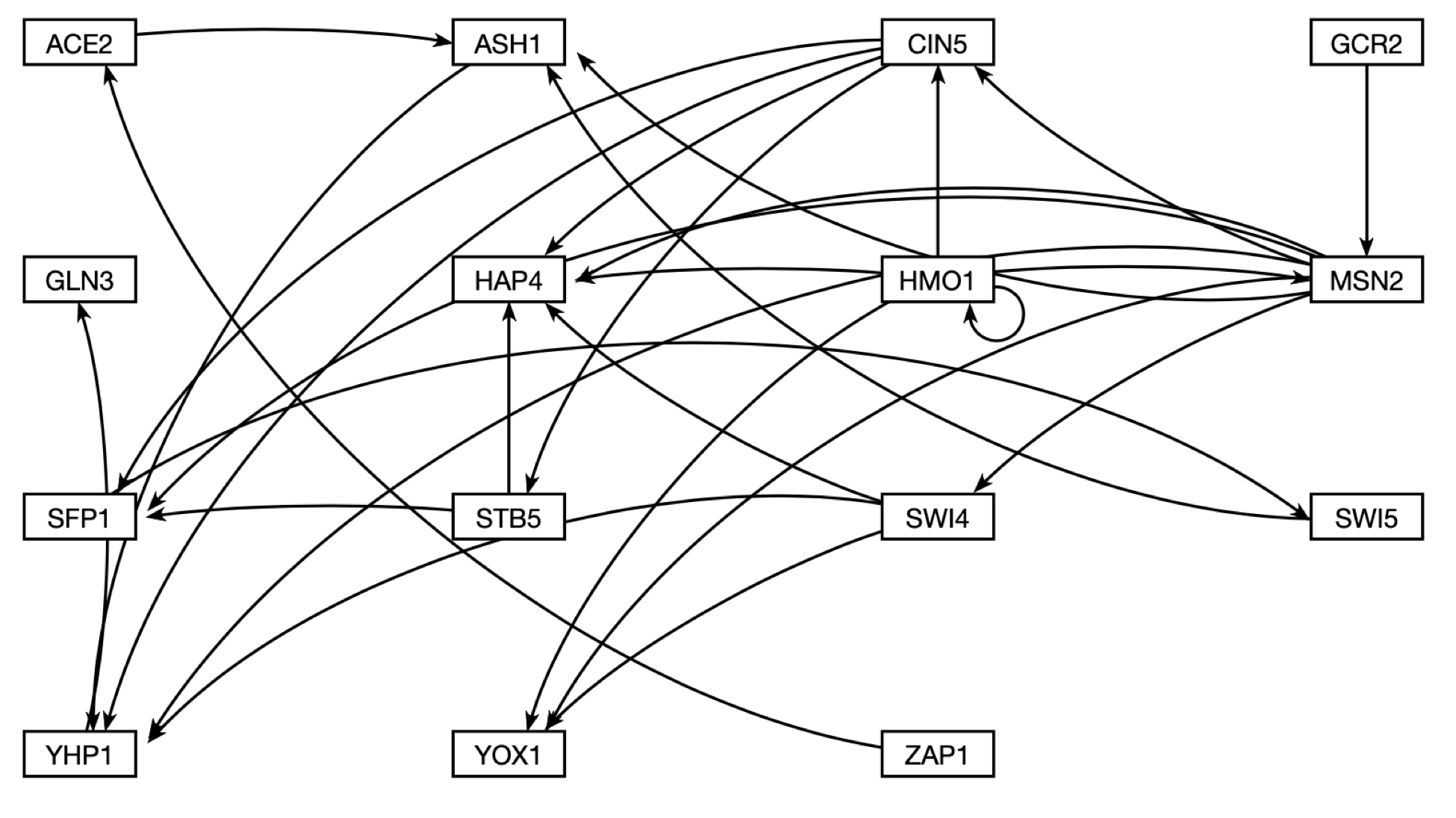
**Equation 2:** Penalized least squares error (LSE) assesses the difference between the experimental and simulated model values.

The fit of the model can be assessed and different models can be compared using the ratio of the least squares error (LSE) to the minimum theoretical least squares error (minLSE) (LSE:minLSE).



**Equation 3:** Minimum theoretical least squares error (minLSE)

The candidate gene regulatory network (GRN) used in this study was db5, which consists of 15 nodes, representing regulatory transcription factors, and 28 edges, representing the regulatory relationships between them (Fig 1). The GRN was generated by inputting the genes that showed a significant change in expression at any time point (Benjamini and Hochberg correct p value < 0.05) in the *cin5* strain microarray data into the YEASTRACT (Yeast Search for Transcriptional Regulators and Consensus Tracking) database, which generates a list of transcription factors that potentially regulate genes in the data, as well as an adjacency matrix denoting their regulatory relationships (Monteiro et al., 2020).



**Fig 1:** Intact db5 gene regulatory network graph visualized using GRNsight. The GRN has 15 nodes (transcription factors) and 28 edges (regulatory relationships).

Through this study, a series of experiments were run to determine the sensitivity of the model to manipulations of the input files, including structural changes to the candidate GRN, variable inclusion of strain data, changes to the optimization parameters, and fixing or estimating production rates. Through comparing the fit of the model within these experiments, the sensitivity of the model to these changes can be assessed.

**Section One: Gene Ontology Enrichment Analysis of Microarray Data Clusters Generated using STEM and k-means Clustering**

**DNA Microarray Data**

Using DNA microarrays, expression data for 6189 genes from the *Saccharomyces cerevisiae* wild type strain and five transcription factor deletion strains (*Δcin5, Δgln3, Δhap4, Δhmo1,* and *Δzap1*) was collected in the Dahlquist lab (NCBI GEO GSE83656). Eachstrain was subjected to cold shock at 13oC for one hour and subsequent recovery at 30oC for another hour. Samples were taken before cold shock, after 15, 30, and 60 minutes of cold shock, and after 60 minutes of cold shock followed by and 30 and 60 minutes of recovery. A within-strain ANOVA was performed on these t15, t30, t60, t90 and t120 samples to determine which genes had a significant change in gene expression at any time point in the experiment. P values were adjusted for the multiple testing problem using the Benjamini and Hochberg correction. Genes with an adjusted p value < 0.05 were then included in the analysis below.

***STEM Profiles***

The expression patterns for the genes that had a significant change in expression at any timepoint for five strains (excluding the *Δhmo1* data) were clustered using the Short Time-series Expression Miner (STEM) individually for each strain (Ernst and Bar-Joseph, 2006). STEM groups genes by their expression patterns into pre-determined profiles based on the number of timepoints and determines which profiles show a statistically significant enrichment of genes that show that expression pattern (Ernst and Bar-Joseph, 2006).

The resulting STEM profiles for each strain were compared to determine commonalities between the strains. An analysis of the STEM profiles revealed that three profiles— profiles 45, 9, and 22— were common in the clustering results of all of the strains. Other profiles were found in several, but not all, of the strains, such as profiles 48, 2, 7, 40, 0, and 28, while other profiles were unique to a single strain, such as profiles 31 and 38 (Fig 2).

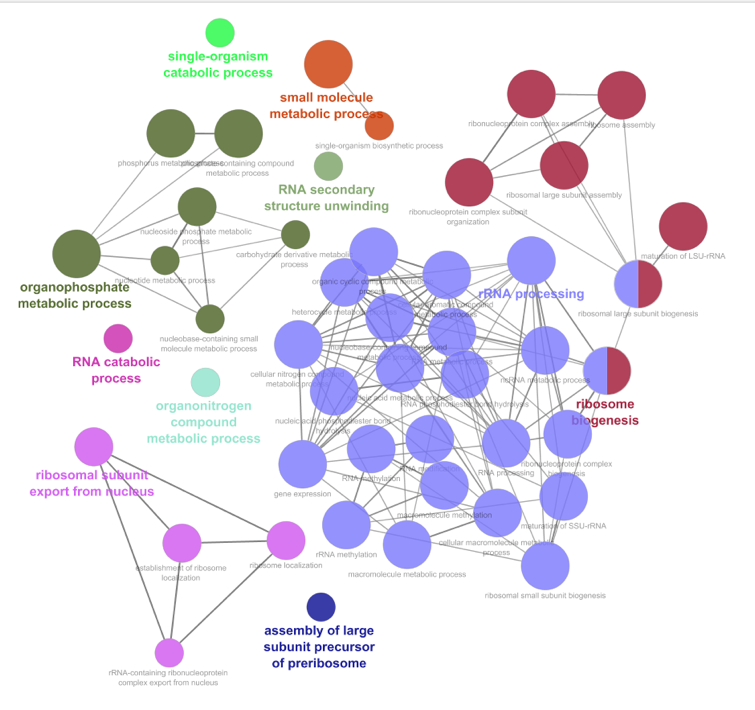
Profile 45 is characterized by an initial increase in gene expression during cold shock followed by a decrease during the recovery period (Fig 2). Profile 9 shows consistent down regulation during cold-shock with a return to the standard expression towards the recovery period. Profile 22 shows no change in expression throughout the cold-shock time points, with upregulation of the genes during the recovery period (Fig 2). These three profiles are likely to contain genes that are important for the cell’s response to cold shock because they have different levels of expression during the cold shock time points versus the recovery time points.

|  |
| --- |
| **Figure 2:** The significant STEM profiles, with trend lines, indicating the pattern of expression common to the genes in the cluster. Profile identifiers listed across the top row and strains are listed to the left. |

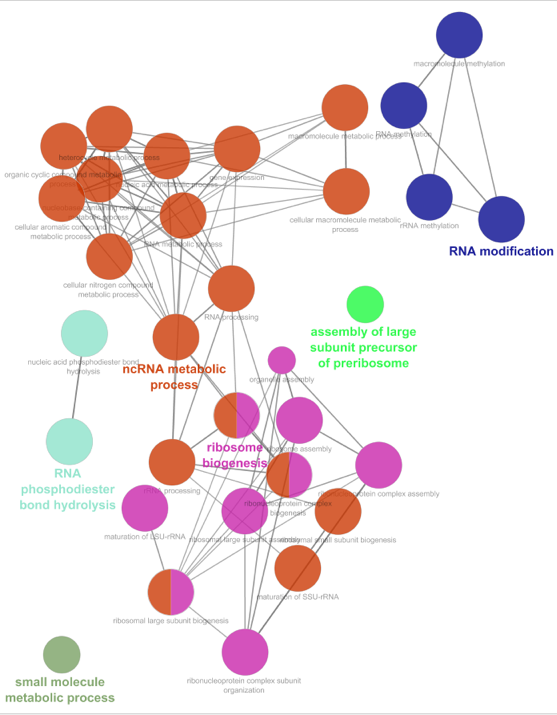
***ClueGO Gene Ontology Analysis of STEM Profiles***

A Gene Ontology (GO) term enrichment analysis for each of the STEM clusters was conducted using ClueGO, a Cytoscape plugin that creates gene ontology networks. ClueGO also facilitates comparisons between two lists of data (Maere et al., 2005). The plugin creates a visualization for groups of GO terms with networks and charts, using kappa statistics to generate links between GO terms. The gene lists and GO terms for STEM profiles 45, 9, and 22 for each strain were input into ClueGO and the resulting GO networks were used to determine the overrepresented functional categories. The full set of methods used to input gene lists or GO terms into ClueGO and examples of properly formatted input files can be found at <https://openwetware.org/wiki/Alice_Finton_Online_Lab_Notebook#How_to_run_ClueGO>.

For profile 45, the overrepresented GO categories included those involved in ribosome biogenesis and RNA processing (Figure 3). Previous studies have shown that ribosome biogenesis and rRNA-processing genes were found to be upregulated during cold shock, indicating their role in the cell’s adaptation to the low temperature (Al-Fageeh et al., 2006). In addition, Sahara et al. found that genes associated with RNA processing, especially RNA polymerase I, which is involved in the production of rRNAs, are upregulated during early cold shock (Sahara et al., 2002). In the present study, it was found that genes associated with ribosome biogenesis and RNA processing were upregulated initially then downregulated during the recovery period, which is consistent with the previous findings.



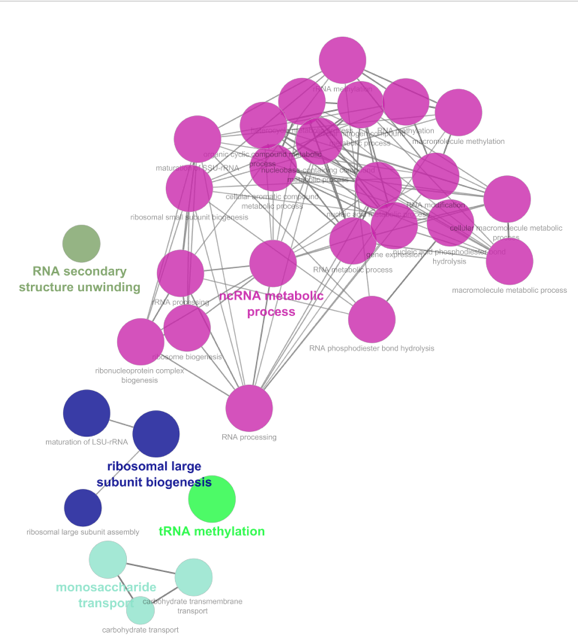
Wild-Type



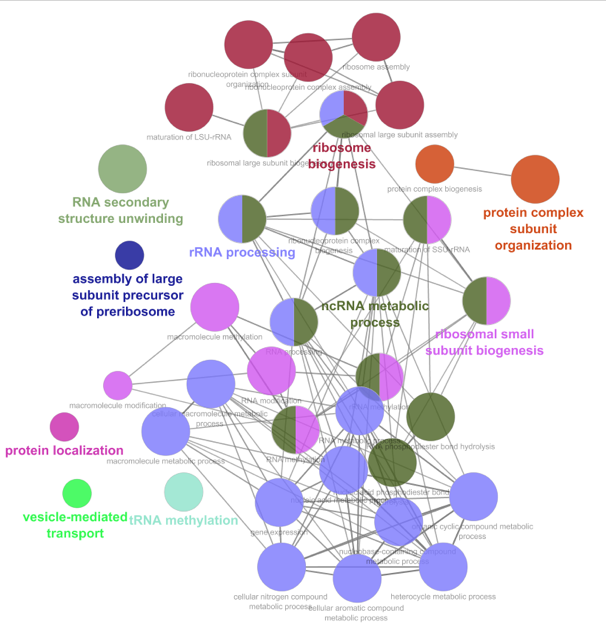
*∆cin5*



*∆gln3*



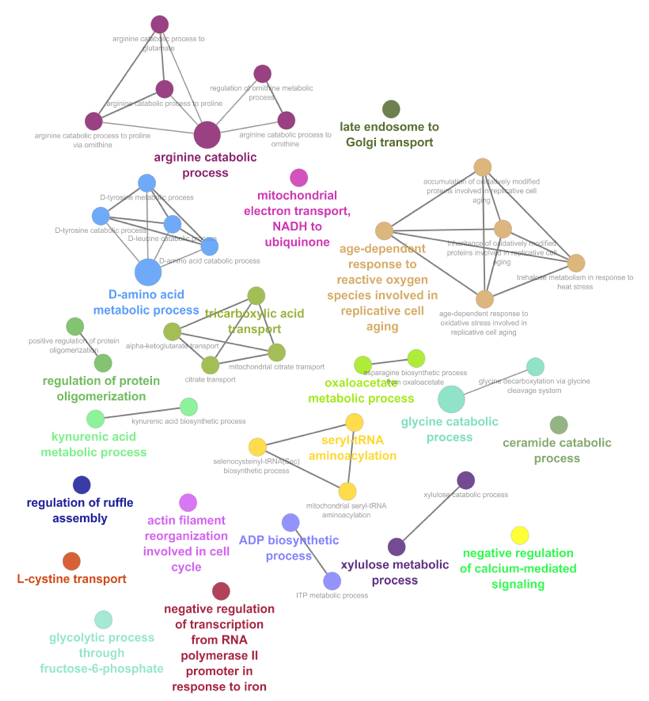
*∆hap4*



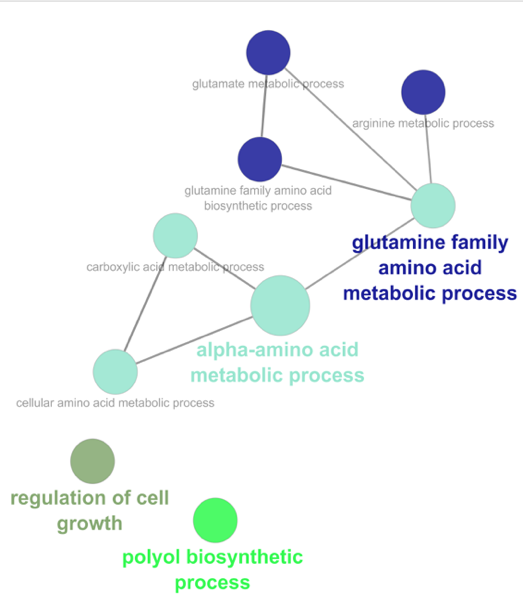
*∆zap1*

**Figure 3:** GO networks representing the overrepresented GO categories in profile 45 for each of the strains. The color of the nodes in the networks reflect a colorized clustering of common terms within each strain. The color of nodes between strain networks do not necessarily reflect common terms.

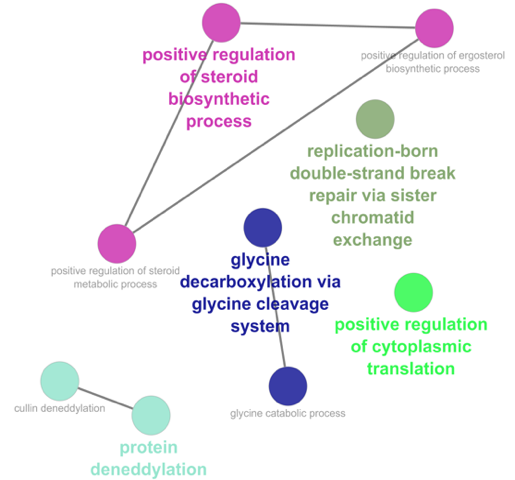
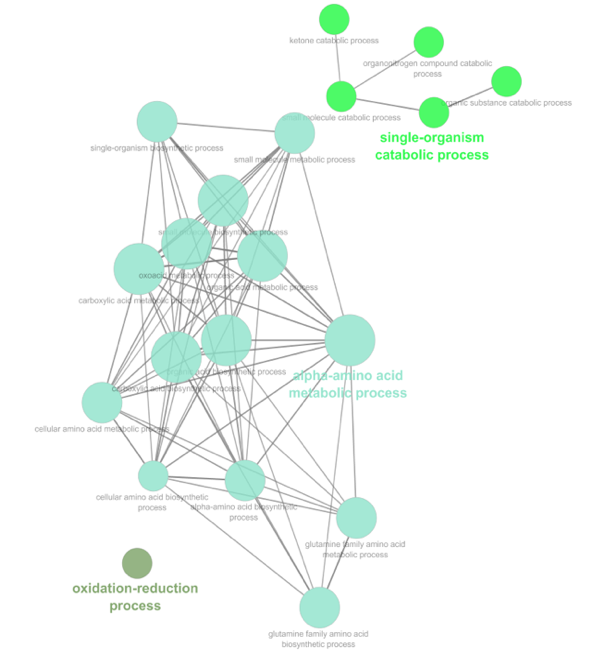
Profile 9 contains genes that were downregulated during the cold shock time points and increased back to pre-shock levels during the recovery period. For this profile, the overrepresented GO categories similarly enriched in each of the strains were involved in catabolic and metabolic processes, such as amino acid and carboxylic acid metabolism, and cell cycle regulation (Figure 4). In the early cold shock response, it has been found that genes associated with lipid and amino acid metabolism are upregulated, which is inconsistent with our findings (Schade et al., 2004). In addition, genes involved in general cellular metabolic processes are upregulated during the late cold response (Al-Fageeh et al., 2006). Further, cell cycle control has also been shown to be induced in the late cold response, regulated by the Msn2/Msn4 complex, which, again, is inconsistent with our findings (Al-Fageeh et al., 2006).



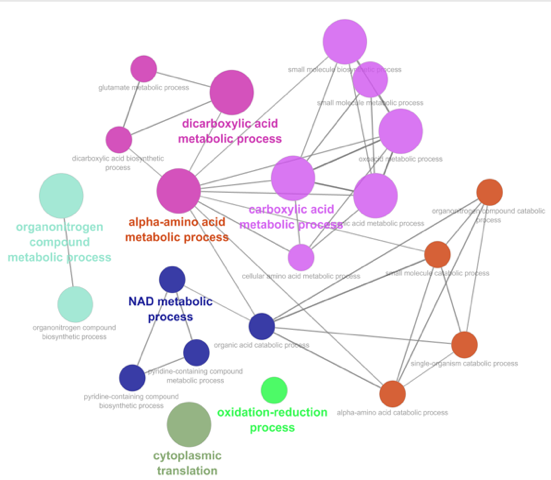
Wild-Type



*∆cin5*



*∆hap4*

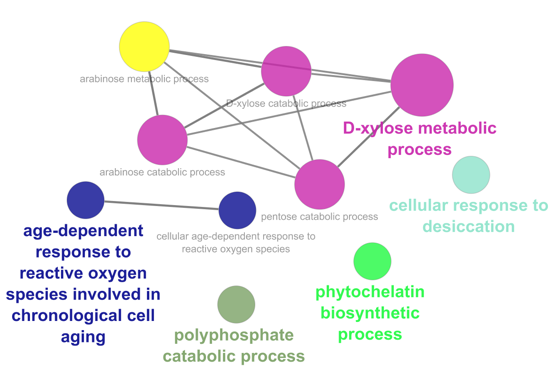
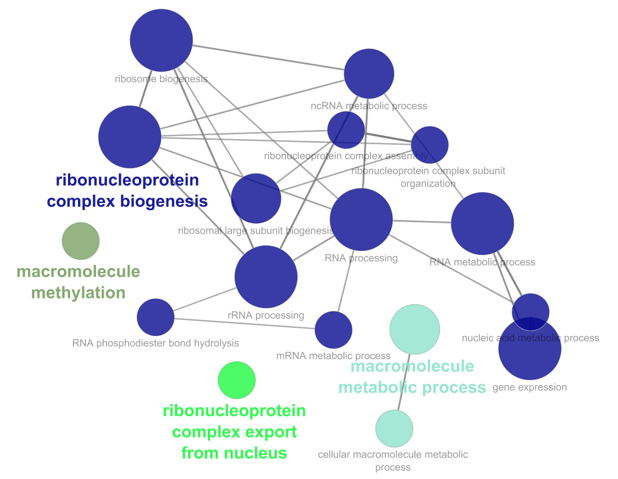
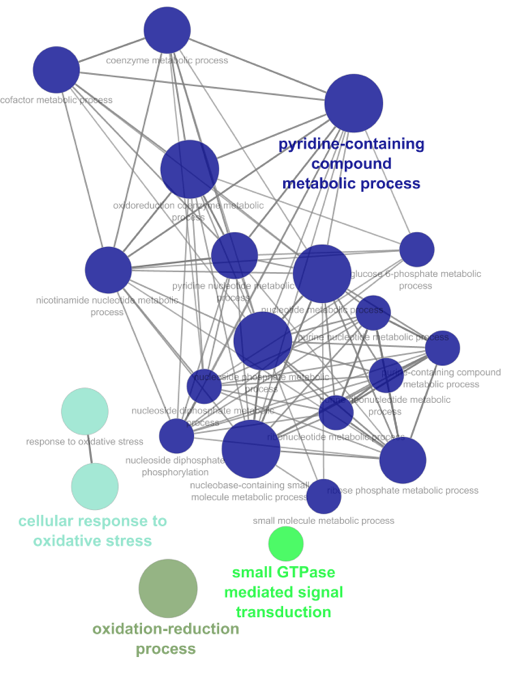
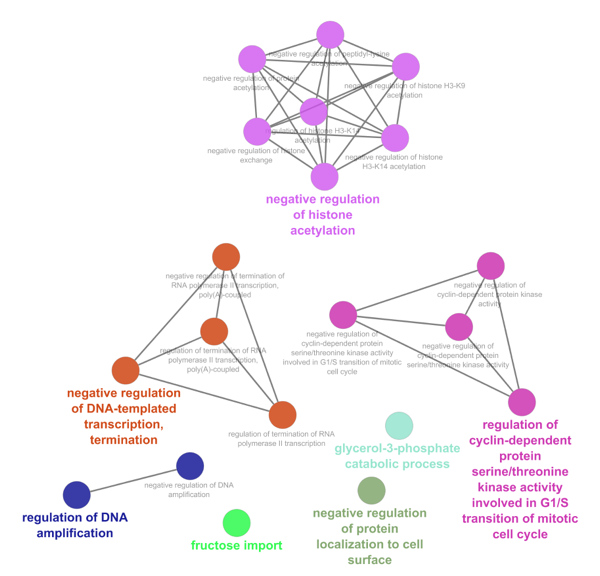
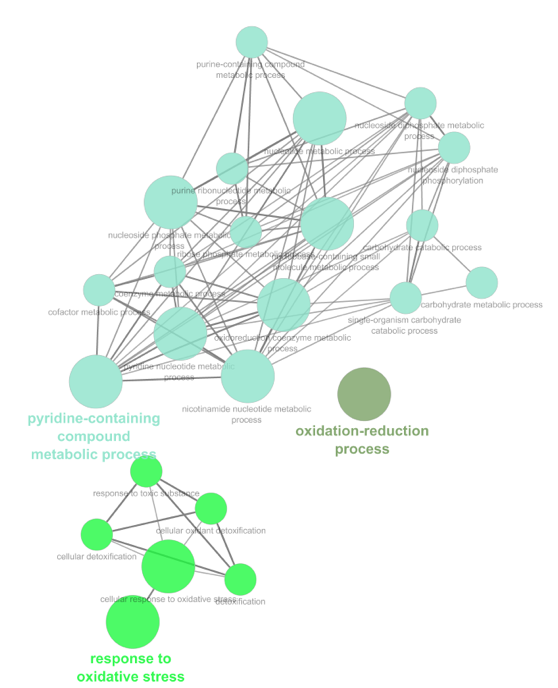


*∆zap1*

*∆gln3*

**Figure 4**: GO networks representing the enriched GO categories in profile 9 for each of the strains. The color of the nodes in the networks reflect a colorized clustering of common terms within each strain. The color of nodes between strain networks do not necessarily reflect common terms.

Profile 22 is defined by an upregulation of genes during the recovery period. Within this profile, the GO categories enriched in all strains included those involved in cell-aging and stress responses, such as desiccation and oxidative stress (Figure 5). In a previous study, genes associated with cell rescue, defense, and virulence were downregulated through the early and late cold shock response (Sahara et al., 2002). However, at near freezing temperatures, these genes were upregulated, indicating more stress at that temperature (Homma et al., 2003). When cells were grown in optimal temperatures, the transcription of these genes were not induced (Homma et al., 2003). Therefore, our findings are not consistent with previous studies.



Wild-Type

*∆cin5*

*∆gln3*

*∆hap4*

*∆zap1*

**Figure 5:** GO networks representing the enriched GO categories in profile 22 for each of the strains. The color of the nodes in the networks reflect a colorized clustering of common terms within each strain. The color of nodes between strain networks do not necessarily reflect common terms.

***k*-means Clustering of the Microarray Data**

In addition to clustering the microarray data through STEM, it was additionally clustered using *k-*means clustering in MATLAB. The STEM clustering method takes data time points into account in the clusters, whereas *k-*means clustering simply clusters the complete dataset, without considering time points (Ernst et al., 2005). The *k-*means clustering method partitions the data into a predetermined number of clusters (*k*) based on similarity (Wagstaff et al., 2001). Therefore, *k-*means clustering was implemented in order cluster the entire microarray data independently from the time points to determine if cold shock and recovery expression patterns differed.

The implementation of k-means clustering in MATLAB could not handle missing values that occurred for some replicates of the microarray data. Thus, before carrying out the clustering, the log fold change values for each gene at each of the timepoints for each strain were averaged and the remaining missing values were set to zero. Genes that showed no significant change in gene expression (Benjamini and Hochberg corrected ANOVA p value >= 0.50 across all of the strain data were filtered out, leaving 2441 genes in the analysis. These genes were then grouped into ten clusters using *k*-means clustering (Figure 6). The genes were then reordered manually by similarity (left to right) and the experiments were ordered by strain and time point (top to bottom). A heatmap was generated with green indicating downregulation and red indicating upregulation (Figure 5). (The code and files for microarray clustering and heatmap clustering are in the appendix.)

|  |
| --- |
| **Figure 6:** Heatmap of microarray clusters with associated cluster number designated below the heatmap. Strain data ordered by strain and time point is listed on the y-axis and the genes which showed significant changes in expression are on the x-axis. Green indicates downregulation and red indicates upregulation. The strains and timepoints on the right hand side are listed top to bottom as follows: ∆*hap4* t120, ∆*zap1* t120*, ∆swi4* t120*, ∆gln3* t120*, ∆cin5* t120*, wt* t120*, ∆hmo1* t120*, ∆hmo1* t90*, ∆hap4* t90*, ∆zap1* t90*, ∆swi4* t90*, ∆gln3* t90*, ∆cin5* t90*, wt* t90*, ∆hmo1* t60*, ∆hmo1* t30*, ∆gln3* t30*, ∆zap1* t60*, ∆zap1* t30*, ∆swi4* t60*, ∆swi4* t30*, ∆hap4* t60*, ∆zap1* t15*, ∆hmo1* t15*, ∆hap4* t30*, ∆hap4* t15*, ∆gln3* t30*, ∆gln3* t15*, ∆cin5* t60*, ∆cin5* t30*, ∆cin5* t15*, wt* t60*, wt* t30*, wt* t15*.* |

The heatmap (Figure 6) shows that there are differences in the expression of genes during cold shock versus the recovery period. Certain clusters showed more extreme changes in expression, such as in clusters 2, 9, 8 and 1, 5, 7 (Fig 6). Other clusters do not display as strong of a change, as shown in clusters 10, 4, and 6 (Fig 6).

There are fifteen regulatory transcription factors in the candidate db5 network modeled in the current study (Ace2, Ash1, Cin5, Gcr2, Gln3, Hap4, Hmo1, Msn2, Sfp1, Stb5, Swi4, Swi5, Yhp1, Yox1, and Zap1). Eight of them (ACE2, GCR2, GLN3, HAP4, STB5, SWI4, SWI5, and ZAP1) were not present in the k-means clustering analysis because they did not change expression significantly in at least one strain (Table 1). The remaining seven were spread out among different clusters. ASH1 is present in cluster 10. This indicates that ASH1 expression change was significant for at least one time point in at least one strain, but there was only a slight downregulation of the gene during cold shock and the recovery period. Ash1 is a transcription factor which negatively regulates mating type switching and the G1/S checkpoint in mitosis (Ash1, SGD). CIN5 and HMO1 are found in cluster 6, which indicates that the expression of these genes was increased during the cold shock time points and returned to baseline during the recovery period. Cin5 is involved in the regulation of DNA binding and the salinity stress response (Cin5, SGD). Hmo1 is associated with DNA structure compaction and the maintenance of the yeast genome (Hmo1, SGD).

|  |  |
| --- | --- |
| **Table 1:** Regulatory transcription factors in the db5 GRN and the k-means cluster number in which they are found. | |
| **Gene** | **Cluster** |
| ACE2 | none |
| ASH1 | 10 |
| CIN5 | 6 |
| GCR2 | none |
| GLN3 | none |
| HAP4 | none |
| HMO1 | 6 |
| MSN2 | 4 |
| SFP1 | 4 |
| STB5 | none |
| SWI4 | none |
| SWI5 | none |
| YHP1 | 4 |
| YOX1 | 2 |
| ZAP1 | none |

The Gene Ontology is a controlled vocabulary used to annotate genes and proteins in three categories, biological process, cellular component, and molecular function (Gene Ontology Consortium, 2015). The list of genes in each of the clusters was loaded into the Gene Ontology enrichment tool on the website (<http://geneontology.org/>) and the list of significant (p<0.05) gene ontology terms was generated. Cluster 4 was not enriched for any GO terms, but the rest of the clusters were enriched for particular GO terms.

Clusters 1, 3, 5, and 7 contain genes that were upregulated during cold shock and downregulated during the recovery period. GO categories overrepresented in cluster 1 included ribosome biogenesis and RNA metabolic processes. Upregulation of genes associated with ribosome biogenesis indicates that it is an important factor in the adaptation of the yeast cells to the cold environment. In a previous study, it was determined that the induction of these genes is necessary for the maintenance of cell proliferation and growth (Albert et al., 2019). Therefore, the induction of these genes indicates that in order to maintain this homeostasis, the yeast cells had an increase in expression of ribosome-related genes. Cluster 3 is enriched for GO categories involved in RNA metabolic processes, such as transcription, processing, and degradation. It was previously determined that RNA metabolism is increased during the early cold shock response, which is consistent with these findings (Schade et al., 2004). Cluster 5 is enriched for GO terms associated with carbohydrate (pentose, mannose, fructose, and glucose) and proton transmembrane transport. The mobilization of carbohydrates in yeast has been shown to increase during the early cold shock response (Schade et al., 2004). Reserves of carbohydrates, specifically glucose and trehalose, have been associated with cold shock (Schade et al., 2004). Therefore, the transport and metabolism of carbohydrates is likely an adaptation of the cells to the cold in order to maintain cell membrane fluidity and resist freezing (Al-Fageeh et al., 2006). Lastly, cluster 7 is enriched for GO categories that are associated with cellular zinc homeostasis and fatty acid metabolic processes. Lipid metabolism has been shown to be upregulated in cold shock, which is consistent with these findings (Schade et al., 2004).

Clusters 2, 8, and 9 contain genes that were downregulated during cold shock and upregulated during recovery. GO terms enriched in cluster 2 include terms that are associated with oxidation-reduction and phosphorus metabolic processes. Cluster 8 is enriched for GO terms involving protein folding, amino acid metabolism, and RNA processing. These results are inconsistent with previous findings, which report an upregulation of genes associated amino acid metabolism (Schade et al., 2004). However, genes involved with protein folding have been shown to be downregulated during the early cold response, but induced during the late cold response (Al-Fageeh et al., 2006). Cluster 9 was enriched for only two GO terms: glucose 6-phosphate metabolic processes and generation of metabolite precursors. Glucose 6-phosphate is an intermediate in glycolysis and its presence negatively regulates the progression of glucose metabolism (Berg et al., 2002). Therefore, when there are high amounts of glucose 6-phosphate, glycolysis rates decrease. The downregulation of metabolite precursors and glucose 6-phosphate indicates that metabolism increases, which is consistent with Murata et al. (2006), which determined that energy and metabolism functions increase during cold shock.

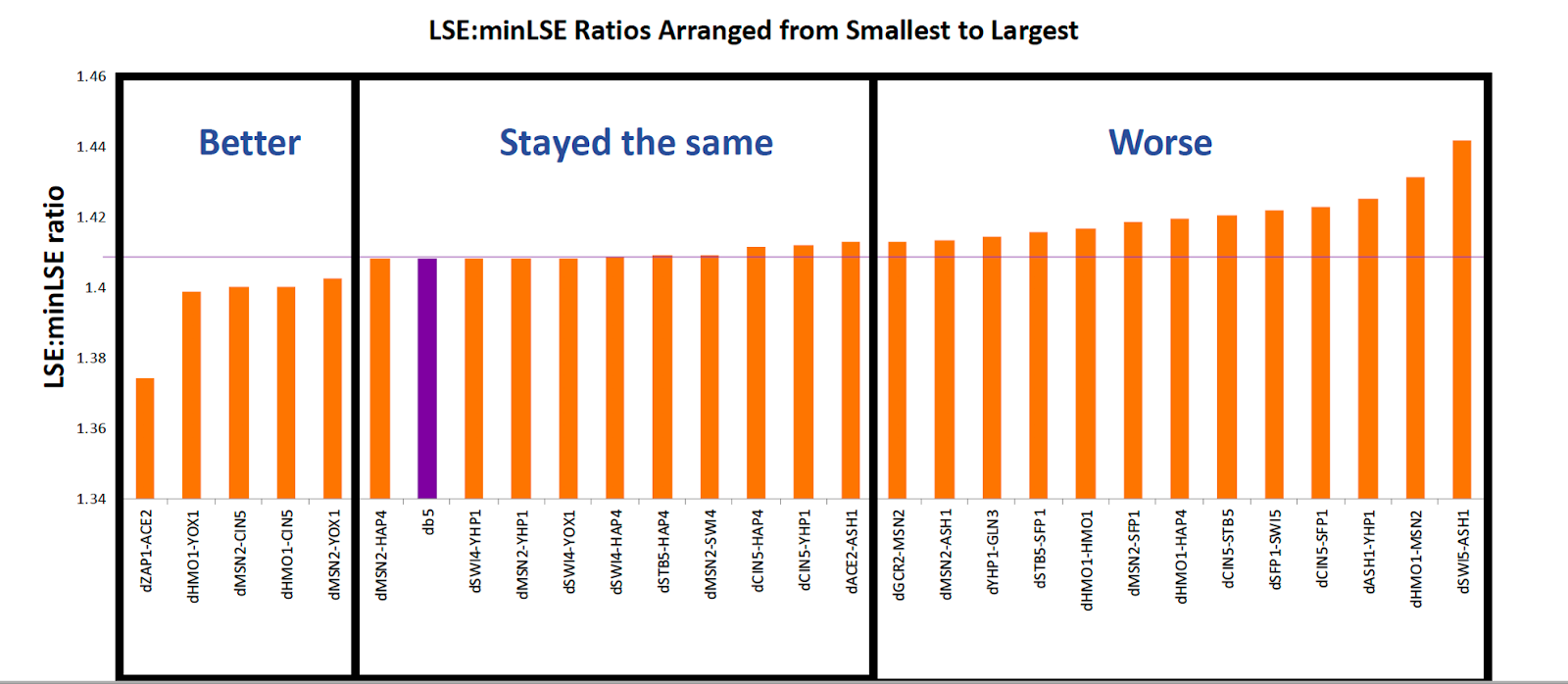
Clusters 6 and 10 contained genes that did not show major changes in expression between the cold shock and recovery time points. There was an overall upregulation for the genes in cluster 6, with more extreme upregulation during the cold shock time points. Only one GO term was enriched in this cluster: mitochondrial translation. There was an overall downregulation for the genes in cluster 10 for both the cold shock and recovery time points. This cluster was enriched for GO categories associated with amino acid metabolism and translational termination. These results are consistent with the findings from profile 9 of the STEM analysis, but are inconsistent with Schade et al. (2004), which indicated amino acid metabolism and protein synthesis are upregulated during cold shock.

**Section Two: Sensitivity of the Model to Network Structure and Strain Data**

Having estimated the parameters for the db5 network, we performed a series of modifications to the model in order to assess the sensitivity to different conditions. First, the structure of the network was manipulated through systematic deletion of the 28-edges of the cold response candidate network (Fig 1) one at a time. Next, model runs were performed with variable inclusion of strain data upon which to base the estimation. Finally, certain nodes were deleted from the network.

**Systematic Deletion of Edges from the db5 Network**

To determine the importance of each edge in the intact network, the edges of the GRN were systematically deleted, one at a time, generating 28 new GRNs, whose parameters were then estimated with GRNmap. To compare the fit of these runs, the LSE:minLSE ratio of each was determined. A ratio higher than that of the intact network indicates that the model is a worse fit, while a ratio lower than the intact network indicates that the model is a better fit. In these 28 edge-deletion networks, LSE:minLSE ratios indicated that five networks performed better than the intact network, while ten networks performed about the same, and thirteen performed worse (Fig 7). The edge-deletions involving the Hmo1, Msn2, and Cin5 transcription factors resulted in a poor performance of the model, indicating that those edges represent important regulatory relationships in the cold shock response (Kelly, 2019).



**Figure 7:** LSE:minLSE ratios for each of the 28 runs, ordered smallest to largest.

Further, *k*-means clustering was performed on the edge weight values from the intact network and edge-deletion networks, resulting in three clusters. Sixteen edge-deletion networks occurred in the same cluster as the intact network, while the other two clusters had four and eight networks, respectively. Compared to the by-eye clustering of edge-deletion networks based on LSE:minLSE ratio, the clusters resulting from *k-*means clustering show that the different clustering methods were generally consistent, though more networks were clustered with the intact network. In addition, the network that performed the best based with regards to the LSE:minLSE ratio was included in cluster 3 of figure 7, which contains networks that performed worse than the intact network. In addition, some of the networks that were included in the “worse” cluster of figure 7 occurred in the same cluster as the intact network in figure 8. Therefore, there are some inconsistencies between the by-eye clusters and the *k-*means clusters.

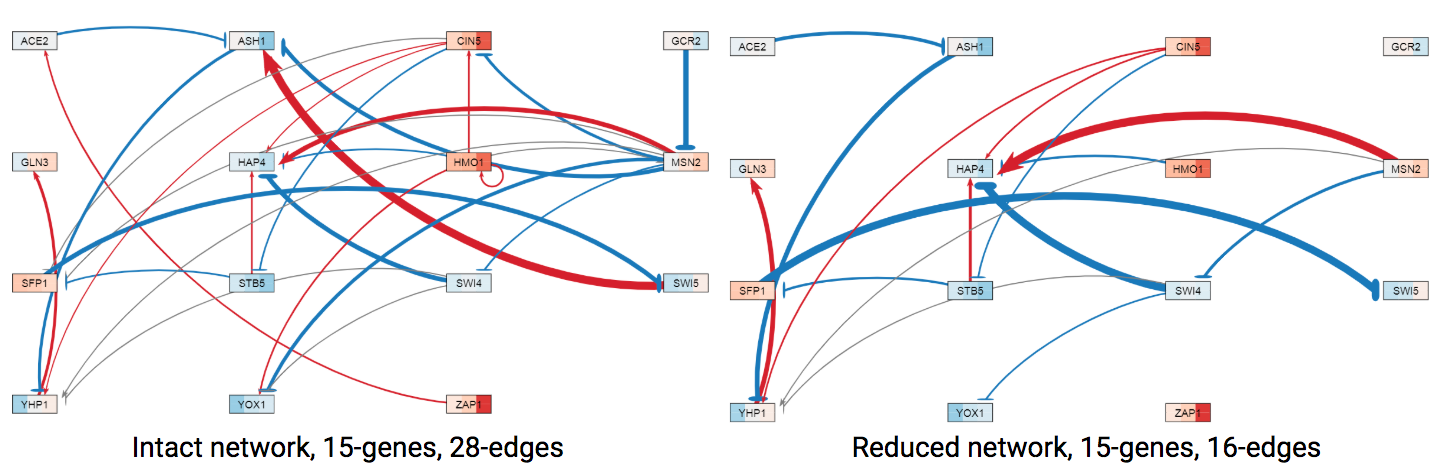
An examination of the clusters showed that certain transcription factors were commonly found in clusters 1 and 3, indicating a deviation from the pattern of weight values for the intact network. It was determined that the weight values deviated from those of the intact network for 6 of 9 edge-deletions involving Msn2, 4 of 5 edge-deletions involving Hmo1, and 3 of 6 edge-deletions involving Cin5 (Fig 8). In addition, the edge-deletions involving Gcr2 and Zap1 caused major deviation in weight values from the intact network (Fig 8). This edge-deletion analysis reveals the importance of these regulatory relationships in the cold shock response.



**Figure 8:** *k*-means clustering of the edge-deletion model runs. The three clusters are marked as cluster 1, 2, and 3. The LSE:minLSE ratio for each of the runs is given by the bar graph. The purple bar and line indicate the ratio value for the intact network. The heat map shows the regulatory weight values for each of the 28 edges in the network in each model run. Red indicates activation and blue indicates repression. The intensity of the color reflects the magnitude of the activation or repression.

These results indicate that the transcription factors Msn2, Hmo1, and Cin5 are important for the cold shock stress response. In a previous study, it was determined that Msn2 and Msn4 are key transcription factors (TFs) in the environmental stress response in yeast (Gasch et al., 2000). These TFs recognize the stress response element (STRE) promoter sequence, which causes the induction of the stress response (Gasch et al., 2000). In addition, Schade et al. (2004) reported that the late cold response is controlled by Msn2/Msn4 transcription factors, which induce a large fraction of the genes associated with the late response, such as glycogen and trehalose synthesis. However, this Msn2/Msn4 complex have not been shown to influence the early cold shock response (Schade, 2004). Further, Berger et al. (2007) reported that the deletion of the HMO1 gene caused a decrease in the transcription of genes indirectly associated with ribosome biogenesis. Therefore, Hmo1 is an important transcription factor in the cold shock response because genes associated with the GO term ribosome biogenesis are overrepresented among genes that are upregulated during cold shock. Furthermore, Cin5 has been found to be induced upon stress conditions and regulates genes associated with ribosome biogenesis and protein synthesis while maintaining chromosome stability (Uniprot, 2019; Fernandes et al., 1997). Disruptions to the pattern of weights when edges involving Msn2, Hmo1, or Cin5 are deleted from the network are supported by literature showing the importance of these TFs in controlling the stress response.

Msn2, Hmo1, or Cin5 regulate or are regulated by multiple other transcription factors, as shown in the candidate network, and while most of the edge deletions involving these transcription factors caused changes to the regulatory weight pattern compared to the intact network, not all of their associated edge-deletions resulted in deviations from the main cluster. Therefore, a reduced network where the edges in clusters 1 and 3 of figure 8 were deleted from the GRN was created and the resulting network had 16 edges. In this GRN, the Gcr2 and Zap1 nodes were left floating because they were no longer connected to the other nodes in the network (Fig 9). Furthermore, the deletion of edges containing Gcr2 or Zap1 transcription factors caused major deviations in the pattern of weight values from the intact network, which indicates that they also play a role in the stress response (Fig 8). Gcr2 is a transcription factor which activates glycolytic genes, thereby increasing carbohydrate metabolism and energy production. Chen et al. (2009) found that Gcr2 was found to be the only consistently upregulated stress-response regulator in all of the environmental stress conditions. Further, Zap1 is a transcription factor that is itself regulated by the presence of zinc in the environment (Yeastgenome.org). In the absence of zinc, Zap1 represses other transcription factors while in the presence of zinc, it induces other transcription factors. Therefore, both Gcr2 and Zap1 regulate other transcription factors when introduced to some environmental stress.



**Figure 9:** Gene regulatory networks for db5 were modified by manipulating the edges clusters 1 and 3. The left GRN is the intact network with 15 genes and 28 edges and the right GRN is the network where the edges not included in cluster 2 were deleted.

Due to the involvement of Gcr2 and Zap1 in the cold shock response and the impact of deleting the edges on the model, three new model estimations were run to determine the fit of the model when the intact network was reduced (Table 2). In one model, the network was left intact, with 15 genes and 28 edges, but the *Δzap1* data sheet was completely deletedfrom the input. Another model was run where the edges in clusters 1 and 3 were deleted from the input so as to create a 15-gene and 16-edge network (Fig4, reduced network). The last model involved the deletion of Gcr2 and Zap1 from the excel sheets, resulting in a 13-gene network, and their edges were also deleted, generating a reduced 26-edge network. To analyze the fit of the models compared to the intact network, the LSE:minLSE ratios of each of the models were compared.

|  |  |  |
| --- | --- | --- |
| **Table 2:** Models run to determine the importance of Gcr2 and Zap1 in the network. The edges included in each of the networks and the data included in the model run are indicated. | | |
| **Genes** | **Edges** | **Data** |
| **15** | **28-Edge** | All data |
| **15** | **16-Edge:** Those within the same k-means cluster as the intact network were included | All data |
| **13** (No Gcr2, Zap1 genes) | **26-Edge:** Edges involving Gcr2 and Zap1 were deleted | All data |
| **15** | **28-Edge** | *Δzap1* data deleted |

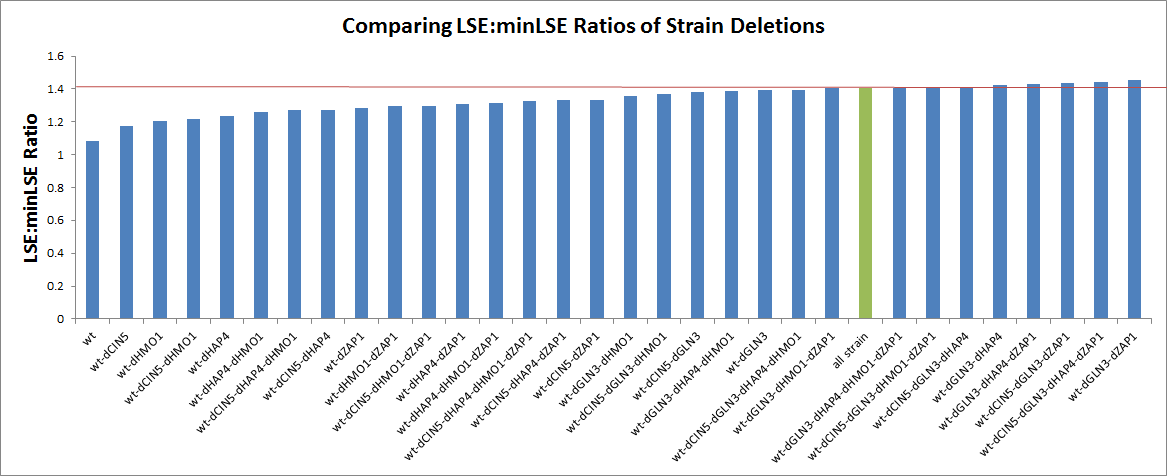
When all edges not included in the main cluster were deleted from the network, the LSE:minLSE ratio was about the same as the intact network, indicating that there was not a major difference in the fit of the model (Fig 10). When only the two edges involving Gcr2 and Zap1 were deleted, the LSE:minLSE ratio decreased even further, indicating a better fit of the model. When *Δzap1* data was deleted from the model, the LSE:minLSE ratio decreased slightly, indicating a slightly better fit (Fig 10). These results indicate that the model is sensitive to the deletion of the Gcr2 and Zap1 edges from the network.

|  |
| --- |
| **Figure 10:** LSE:minLSE ratios for the model runs: the 16-edge reduced network (16-edge), intact network without GCR2 and ZAP1 (26-edge), no dZAP1 data included (28-edge no Zap1), and the intact db5 network. |

**Variable Inclusion of Strain Data**

Up until this point, most of the model runs on the db5 network were performed using all available microarray data from each strain (wild type, *Δcin5, Δgln3, Δhap4, Δhmo1,* and *Δzap1* strains). To test the sensitivity of the model to the inclusion of each strain’s data, 32 model runs were conducted with the wild type data and all possible permutations of inclusion of the other strains’ data. The runs were generated with wild type (*wt*) only, *wt* plus one, *wt* plus two, *wt* plus three, *wt* plus four, and *wt* plus all five deletion strains (A table listing the runs can be found in the appendix). A re-estimation of model parameters, including production rates, expression thresholds, and regulatory weights, was performed in GRNmap.

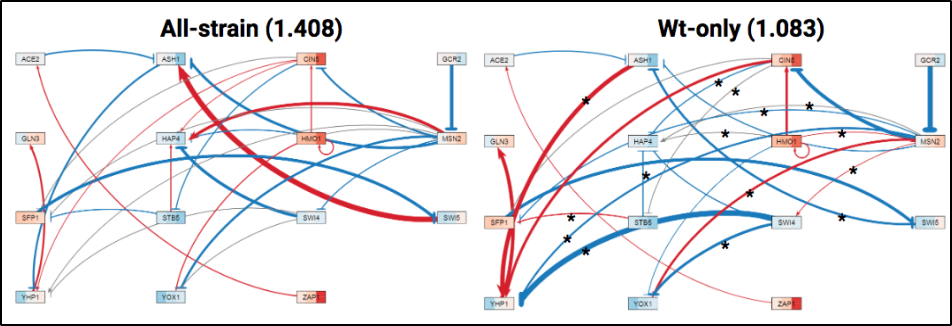
In order to assess the fits of the models, the LSE:minLSE ratio was determined and compared between the 32 model runs (Figure 11). A ratio higher than that of the intact network indicates that the model is a worse fit, while a reduced ratio indicates that the model is a better fit than the intact network. The LSE:minLSE ratio increased for eight of the new models. Of these eight models, all of them included data from the *Δgln3* strain. The LSE:minLSE ratio decreased for 23 of the models, including the wild-type only model and all of the wild-type plus one deletion strain models. Overall, the model run with wild-type only data resulted in the best fit, while the inclusion of the *Δgln3* strain data resulted in a fit that was about the same as the all-strain network or worse. Further research is necessary to determine why the inclusion of *Δgln3* strain data causes a worse fit.

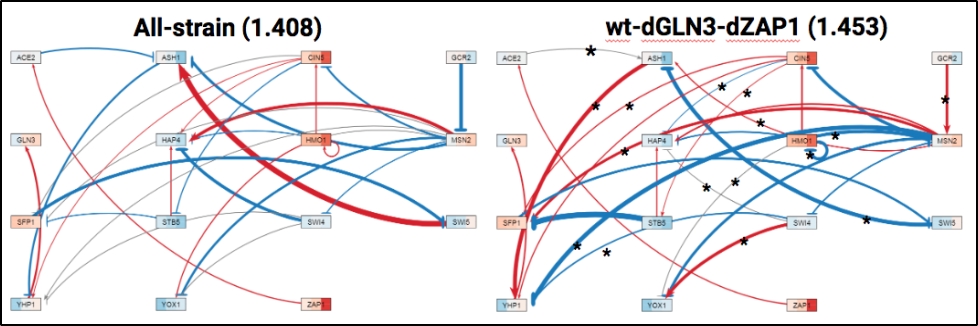


**LSE:minLSE Ratios of Model Runs with Variable Inclusion of Strain Data**

**Figure 11:** LSE:minLSE ratios for each of the model runs. The models are ordered by increasing LSE:minLSE ratio. The green bar represents the all-strain network and the red line indicates the ratio value for the all-strain network.

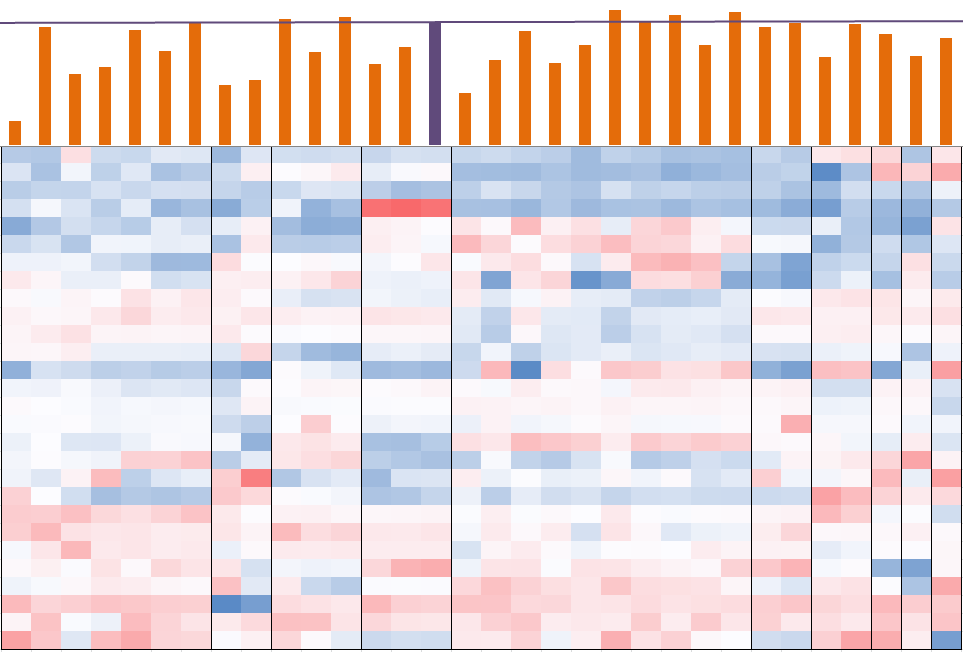
The networks were then visualized with GRNsight. The all-strain intact network was compared to the model with the lowest LSE:minLSE ratio(wt-only), the model highest LSE:minLSE ratio (wt-dGLN3-dZAP1), and the network that did not include *Δgln3* strain data (wt-dCIN5-dHAP4-dHMO1-dZAP1). Compared to the all-strain network, the wt-only GRN had 15 edges that changed the sign of the weight values (Fig 12). The wt-dGLN3-dZAP1 had the highest of the LSE:minLSE ratios. When compared to the all-strain network, 16 of the 28 edges changed the sign of the weight values (Fig 13). Further, when *Δgln3* data was included, the model consistently performed worse. Therefore, a comparison was made between the all-strain data model and the model where only the *Δgln3* data was not included. The network run using the wt-dCIN5-dHAP4-dHMO1-dZAP1 data had four edge weights that changed sign compared to the all-strain network (Fig 14). These results indicate that the model is sensitive to the strain data used to estimate the model parameters.

**Figure 12:** Gene regulatory network (GRN) generated using GRNsight. The red edges indicate activation and the blue indicate repression. The thickness of the edge represents the strength of that relationship. In the GRN to the left, all data was used in estimation, and in the network to the right, wt-only was the data used in the estimation. The LSE:minLSE ratio is given in the parentheses. The asterisks on the edges of the wt-only GRN indicate edges that have changed sign from the all-strain network. 15 of the 28 edges changed.

**Figure 13:** Gene regulatory network (GRN) generated using GRNsight.The red edges indicate activation and the blue indicate repression. The thickness of the edge represents the strength of that relationship. The GRN to the left is the all-strain data network, while the network to the right is the wt-dGLN3-dZAP1 run. The LSE:minLSE ratios are given in the parentheses. The asterisks refer to the edges that changed.

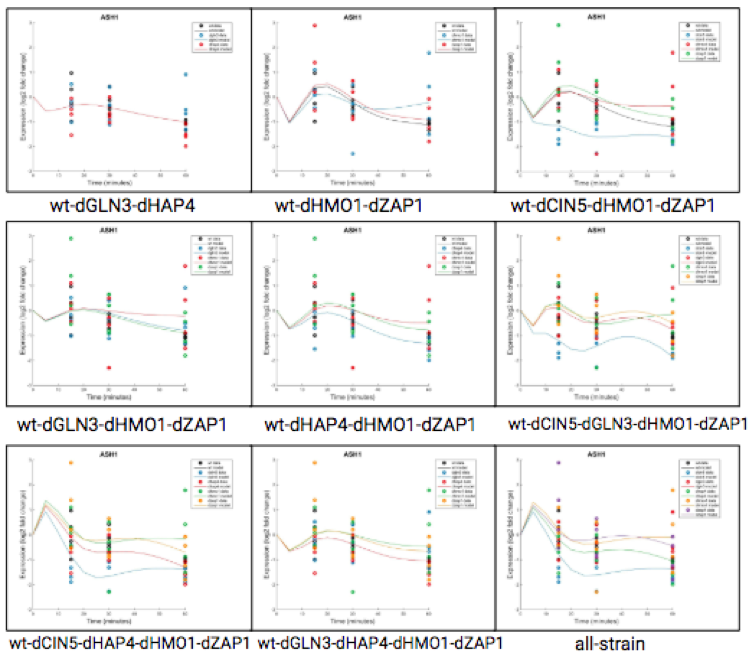
|  |
| --- |
| **Figure 14:** Gene regulatory network (GRN) generated using GRNsight. The red edges indicate activation and the blue indicate repression. The thickness of the edge represents the strength of that relationship. Comparison of all-strain and wt-dCIN5-dHAP4-dHMO1-dZAP1 GRNs. Four of the 28 edges changed when dGLN3 was not included in the model. The asterisks indicate edges that have changed in the new model. |

To discern patterns in the weight values, it was necessary to use *k*-means clustering in MATLAB and generate a heat map. In clustering the heat map in both dimensions, a cluster containing the all-strain, cluster 4, showed a similar pattern of weight values, which was not reflected in the other clusters (Fig 15). The two models that were most similar to the all-strain’s regulatory weights were the models that either excluded *Δgln3* strain data or both the *Δgln3* and *Δzap1* data (Fig 15). Again, the runs that included *Δgln3* data had the highest LSE:minLSE ratio, indicating the worst fit. This trend is reflected in the comparison of the GRNs, where the regulatory weights values did not deviate from the all-strain network as severely as other models when *Δgln3* and *Δzap1* data were not included in the model (Fig 14). In addition, when looking at the expression plots, it was determined that the inclusion of *Δcin5*, *Δhmo1*, and *Δzap1* data did not cause the simulated model data to diverge, indicating that their expression is similar to that of the wild type strain. However, when *Δ*cin5, *Δhmo1*, and *Δzap1* data were not included, the simulated model data did diverge, indicating that their expression dissimilar to the other strains (Fig 16).



**1 2 3 4 5 6 7 8 9 10**

**Figure 15:** Clustered heatmap of regulatory weights with corresponding LSE:minLSE ratio. The heatmap was clustered in the x and y axes, using ten clusters. The purple bar indicates the intact network. Clusters are demarcated by the black lines. Red indicates activation and blue indicates repression. The intensity of the color reflects the magnitude of the activation or repression.



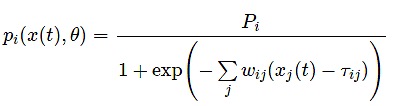
**Figure 16:** Expression plots for candidate model runs for the ASH1 gene. The inclusion of dCIN5, dHMO1, or dZAP1 data caused the simulated expression to diverge for each strain. Divergence is indicated by multiple expression lines, while no divergence results in a single expression line. (the rest of the expression plots can be found in the appendix)

The variable inclusion of strain data impacted the fit of the model, with models including *Δgln3* data performing worse than the intact network. This trend is reflected in the LSE:minSLE ratios, GRNs, *k-*means clustering, and expression plots from the model runs. Gln3 is a transcription factor that regulates glutamine metabolism and has been found to be associated with the nitrogen catabolite repression system (Tate et al., 2007, SGD). In a previous study, it was determined that upregulated functional categories during the early cold shock response included those that are associated with amino acid metabolism (Schade et al., 2004). Therefore, Gln3 plays a role in the cold shock response in *S. cerevisiae.*

**Section Three: Sensitivity Assessment of the Model to Changing Production Rates**

GRNmap performs estimations of regulatory weight, expression threshold, and production rate parameters. In a standard model, each of these parameters is estimated through the program, but the threshold and production rate parameters can be input as fixed values.  When fixed, the initial guesses for the production rate and/or threshold will remain throughout the model estimation, which could affect the fit of the model. Therefore, in order to determine the impact of fixing or estimating these parameters, models were run where production rate and threshold were either estimated or fixed.

When estimated, these parameters are optimized using ordinary differential equations in GRNmap. The production rate is estimated using the sigmoidal function:

, where *Pi* is the initial guess production rates and *pi*is the estimated rate (Dahlquist et al., 2015). The expression over time is then modeled by production minus degradation rate. The initial guess for the production rate was derived through doubling the degradation rates from the Neymotin et al. (2014) paper.

The threshold indicates the point at which activators either induce production, when expression is above threshold, or stop production, when expression is below threshold. For repressors, production is induced when the expression is below the threshold level, while production is stopped when the expression is above threshold (Dahlquist et al., 2015). In the model, threshold is estimated using:

https://lh4.googleusercontent.com/AGQtm2MrzQPC6JV_VCxIj2n_ydkK4ZMGBl7TnlnEc4CnSFxR8eQ5JAQdPwm7oBKojKbhlkR3Pqu5dWJPurGWnuuxf7uDea-pXg-GTWRNylR2X8Uyf5BegqQgBgPwvjzqLdZjZlZP, where *wi j* is the weight of a regulatory relationship and τ*i j* is the threshold at which the production turns off or on (Dahlquist et al., 2015). In standard model runs, the initial guess for the threshold was set to zero.

**Comparison of Estimated Production Rates with Published Experimentally-derived Rates**

In GRNmap, the change in expression over time for each transcription factor in the network is modeled as production of mRNA minus degradation (Section One, Equation 1). The model takes as input degradation rates and the user has the option to either provide production rates or estimate them from an initial guess. For the model runs described previously, the production rates were always estimated and the initial guesses were set as two times the degradation rate for each gene, which were derived from a study by Neymotin et al. (2014) that experimentally determined mRNA half-life data for all genes in yeast using a method called RATEseq. Assuming that the system is in steady state with the relative expression of all genes equal to 1, (P/2) - lambda = 0, where lambda is the degradation rate, is a reasonable initial guess. We wondered how closely the production rates estimated by GRNmap matched the experimentally-derived rates measured by Neymotin et al. (2014). To perform this comparison, production rates were derived from Neymotin et al. (2014). The production rates were calculated using k = α[RNA], where k is the rate constant for synthesis and α[RNA] is the RNA abundance (Neymotin et al., 2014). This RNA abundance was calculated as α = αRNA + αgrowth, where α is the RNA concentration and αRNA is the degradation rate constant and αgrowth is the cell’s division rate constant (Neymotin et al., 2014). Using this equation, the Neymotin et al. (2014) production rates were derived from their dataset (Table 3).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table 3:** Comparison of production rates from Neymotin et al. (2014), estimated production rates from GRNmap, and initial guesses. The estimated production rates are those from the db5 all-strain model run. The initial guess rates were calculated as two times the Neymotin et al. (2014) degradation rate | | | | | |
| **Gene** | **Initial Guess Production Rate** | **Estimated Production Rate** | **Neymotin Production Rate** | **Estimated - Neymotin** | **Percent Change (%)** |
| **ACE2** | 0.224 | 0.202 | 0.180 | 0.022 | 12.22 |
| **ASH1** | 0.433 | 1.677 | 1.037 | 0.640 | 61.72 |
| **CIN5** | 0.201 | 0.656 | 0.063 | 0.593 | 941.27 |
| **GCR2** | 0.193 | 0.232 | 0.327 | -0.095 | -29.05 |
| **GLN3** | 0.322 | 0.302 | 0.365 | -0.063 | -17.26 |
| **HAP4** | 0.272 | 1.302 | 1.827 | -0.525 | -28.74 |
| **HMO1** | 0.099 | 0.307 | 1.406 | -1.099 | -78.17 |
| **MSN2** | 0.408 | 2.557 | 0.487 | 2.070 | 425.05 |
| **SFP1** | 0.693 | 1.555 | 1.199 | 0.356 | 29.69 |
| **STB5** | 0.140 | 0.120 | 0.080 | 0.040 | 50.00 |
| **SWI4** | 0.283 | 0.316 | 0.157 | 0.159 | 101.27 |
| **SWI5** | 0.322 | 1.921 | 0.340 | 1.581 | 465.00 |
| **YHP1** | 0.173 | 0.208 | 0.283 | -0.075 | -26.50 |
| **YOX1** | 0.730 | 1.391 | 1.028 | 0.363 | 35.31 |
| **ZAP1** | 0.104 | 0.128 | 0.082 | 0.046 | 56.10 |

In comparing the estimated production rates and the Neymotin rates, there are differences between the values for each of the genes. For ten of the fifteen genes, the estimated production rate was greater than the Neymotin rate, while five of the genes had a lower estimated rate (Table 3). Certain genes, such as CIN5, MSN2 and SWI5 showed a major difference between the Neymotin et al. and estimated production rates, revealing inconsistency in the rates between them (Table 1). However, other genes, like GLN3 and ACE2, had relatively consistent production rates between the estimated and Neymotin production rates (Table 3).

Further, using a scatter plot, the similarities between the estimated production rates and the Neymotin et al. rates could be visualized. A slight linear relationship could be seen, indicating there is some commonality between the production rates, but this relationship is small with R2 values of 0.1631 for the estimated vs. Neymotin et al. production rates (Fig 17). Therefore, there is not a major consistency in the Neymotin et al. production rates and estimated rates.

|  |
| --- |
| **Fig 17:** Neymotin production rates compared to with the estimated rates. The associated gene is labeled to the right of the data points and the solid blue line is the linear trendline for the data points. The dotted gray line is a reference line with a slope of 1. |

These results indicate that, although there are slight commonalities between the Neymotin et al. (2014) production rates and estimated rates, the similarities are weak, indicating more inconsistencies between the three groups.

**Variably Fixed or Estimated Weights, Thresholds, and Production Rates**

Using db1-db7 all-strain data input files, six new modified models per network were created where thethresholds *(b)* and production rates *(P)* were either estimated or fixed in the model and the weights *(w)* were estimated in each new model (Table 4). For the models where the production rates were fixed, additional models were run where the production rates used as initial guesses were derived from the Neymotin et al. (2014) paper. For the models that were run with fixed threshold values, the initial guess was set to 0. The models were run in GRNmap and from the output, the LSE:minLSE ratio was calculated for each model in order to assess and compare the fit of the models.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 4:** parameter model runs with indication of which parameters were fixed or estimated. Weights were estimated for all model runs, while the threshold and production rate were variably estimated or fixed between the runs. The initial guess production rates were derived by doubling the Neymotin et al. (2014) degradation rates. Neymotin production rates were derived from the Neymotin et al. (2014) paper and used in the models where indicated.  The threshold initial guess was set to 0. | | | |
| **Model Run** | **Weight *(w)*** | **Threshold *(b)*** | **Production Rate *(P)*** |
| **Est *P, b, w*** | Estimated | Estimated | Estimated |
| **Est *P, w;* Fix *b*** | Estimated | Fixed | Estimated |
| **Est *b, w;* Fix *P*** | Estimated | Estimated | Fixed |
| **Est *b, w;* Fix *P*** (using Neymotin Production Rates) | Estimated | Estimated | Fixed |
| **Est *w;* Fix *P, b*** | Estimated | Fixed | Fixed |
| **Est *w;* Fix *P, b*** (using Neymotin Production Rates) | Estimated | Fixed | Fixed |

In order to determine the effect of estimating or fixing parameters on the fit of the model, the LSE:minLSE ratios from each model run were compared. The lower the LSE:minLSE ratio, the better the fit of the model. Db5 was chosen as the candidate network for analyzing the effect on model sensitivity, as db1-db7 model runs resulted in the same trends for increased or decreased LSE:minLSE ratio among the estimate/fix model runs.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 5:** Production rates that were fixed (first two columns) or estimated (last two columns). | | | | |
| **Gene** | **Fix *P* (2x D)** | **Fix *P* (Neymotin)** | **Est *P, b, w*** | **Est *P, w*, Fix *b*** |
| **ACE2** | 0.224 | 0.180 | 0.202 | 0.124 |
| **ASH1** | 0.433 | 1.027 | 1.677 | 0.785 |
| **CIN5** | 0.201 | 0.063 | 0.656 | 1.247 |
| **GCR2** | 0.193 | 0.327 | 0.232 | 0.236 |
| **GLN3** | 0.322 | 0.365 | 0.302 | 0.263 |
| **HAP4** | 0.272 | 1.827 | 1.302 | 0.884 |
| **HMO1** | 0.099 | 1.406 | 0.307 | 0.160 |
| **MSN2** | 0.408 | 0.487 | 2.557 | 3.316 |
| **SFP1** | 0.693 | 1.199 | 1.555 | 1.598 |
| **STB5** | 0.140 | 0.080 | 0.120 | 0.205 |
| **SWI4** | 0.283 | 0.157 | 0.316 | 0.528 |
| **SWI5** | 0.322 | 0.340 | 1.921 | 3.806 |
| **YHP1** | 0.173 | 0.283 | 0.208 | 0.257 |
| **YOX1** | 0.730 | 1.028 | 1.391 | 1.664 |
| **ZAP1** | 0.104 | 0.082 | 0.128 | 0.129 |

The lowest ratio occurred for the models where *w, b,* and *P* were all estimated, therefore, it was determined that fixing any parameter decreased the fit of the model (Fig 18).  Estimating or fixing the threshold did not change the LSE:minLSE ratio value greatly, which indicates that threshold does not have a major impact on the fit of the model. Conversely, fixing the production rates made the LSE:minLSE ratio increase more, indicating their importance on the fit of the model (Fig 1). In addition, the use of the Neymotin et al. (2014) production rates as the fixed values caused the LSE:minLSE ratio to increase even more, indicating that the initial guesses for the  production rates are a better fit for the model than the Neymotin et al. (2014) values.

|  |
| --- |
| https://lh3.googleusercontent.com/d8SLZTq85iL4uoIm7OTqLabfKN1OpAW27T-EiOJlj1ViVtQ15vKM4mhAQ61MwtgGNFOK4Xz0IYwzK6jNuu1exG9otXA4jfWqD4kPxRUoTlGJEtioyCpyxjSiQMTcRLcAMKqwxP8g  **Fig 18:** The candidate db5 network LSE:minLSE ratios for the estimate/fix production rate (*P*), weight (*w)*, and threshold (*b)* model runs. Production rates used were the initial guess rate or the Neymotin et al. (2014) production rate (designated by “Neymotin”). For db1-db7, fixing production rate caused a greater ratio, the greatest of which were fixing the Neymotin production rates. |

From these results, we found that fixing the production rate as the initial guesses caused the model to perform worse, but the worst fit came from the models that were run with the Neymotin et al. (2014) production rates. This indicates that our initial guess production rates are a better fit for the model than the Neymotin values. Overall, it was determined that the model is not too sensitive to changes in the optimization parameter, but is sensitive to fixing the production rates

**Section Four: Sensitivity Assessment of the Model to Manipulation of Optimization Parameters**

GRNmap uses the optimization parameters kk\_max, MaxIter, TolFun, MaxFunEval, and TolX to optimize the fit of the model. The parameter kk\_max is the number of times that the optimization loop is rerun, which can improve the model performance. The MaxIter is the maximum iterations that the model will run. This parameter should be high enough that the model does not stop before completing the estimation. Therefore, if the MaxIter is too low, the model may not complete the optimization. The TolFun is the difference between the least squares of each rerun that signifies when the model is no longer making improvements. The MaxFunEval is the maximum number of times that the function will evaluate the least squares cost. Lastly, the TolX indicates the maximum difference between the least errors cost before the program signifies there is no more improvement of the model (Dahlquist et al., 2018).

The standard optimization parameter values used in the model are 1 for kk\_max, 1.00x108 for MaxIter, 1.00x10-6 for TolFun, 1.00x108 for MaxFunEval, and 1.00x10-6 for TolX.  In order to assess the sensitivity of the model to changes in these optimization parameters, each parameter was changed by an order of magnitude in either direction and then run. The LSE:minLSE ratios of the models were compared to that of the standard model run with the midpoint values. Deviation from the standard model indicates that either the model ran better, as in the case of a decrease in the LSE:minLSE ratio, or worse, indicated by an increase in the ratio.  In addition, the iteration counts of the models were compared to determine whether the changes to the optimization parameters affected the number of times the model evaluated the least squares before stopping the model run.

Using db5 all-strain data, the optimization parameters in the input Excel workbook were changed by one degree of magnitude in each direction for kk\_max, MaxIter, TolFun, MaxFunEval, and TolX (Table 6).  One optimization parameter was changed per new model, with the other parameters set with their standard values.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 6:** Optimization parameter changes. | | | |
| **Optimization Parameter** | **Low** | **Mid** | **High** |
| **kk\_max** | 0.1 | 1 | 10 |
| **MaxFunEval** | 1.00x107 | 1.00x108 | 1.00x109 |
| **MaxIter** | 1.00x107 | 1.00x108 | 1.00x109 |
| **TolFun** | 1.00x10-7 | 1.00x10-6 | 1.00x10-5 |
| **TolX** | 1.00x10-7 | 1.00x10-6 | 1.00x10-5 |

The resulting ten new model input files were run through GRNmap and the LSE:minLSE ratios and iteration counts of each model were then compared to determine if the changes to the optimization parameters affected the fit of the model. The standard model, where none of the optimization parameters were changed, was the db5 all-strain model run.  The model run where kk\_max was set to 0.1 failed to complete the estimation, therefore, it was not included in the comparison.

The LSE:minLSE ratio for the run under standard conditions (mid-point value for each parameter) was 1.4081816 (Table 7). When the parameters were changed by one magnitude in either direction, the ratio changed only for kk\_max at 10 and TolX at 1.00E-07, but the change was minimal with a 2.56x10-4% decrease for kk\_max at 10 and a 3.18x10-4% decrease for TolX at 1.00E-07.   The other runs caused no difference in the ratio. Therefore, changing the value for the optimization parameters made very minimal difference to the fit of the model.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 7:** Iteration count and LSE:minLSE ratio for the low, mid, and high values for each of the optimization parameters. The asterisks indicate the values that differed from the mid-point values. | | | |
| Optimization Parameter | Value Magnitude | Iteration Count | LSE:minLSE |
| kk\_max | 0.1 | - | - |
| 1 | 111242 | 1.4082 |
| 10 | 645513\* | 1.40817\* |
| MaxFunEval | 1.00x107 | 111242 | 1.40818 |
| 1.00x108 | 111242 | 1.40818 |
| 1.00x109 | 111242 | 1.40818 |
| MaxIter | 1.00x107 | 111242 | 1.40818 |
| 1.00x108 | 111242 | 1.40818 |
| 1.00x109 | 111242 | 1.40818 |
| TolFun | 1.00x10-7 | 111242 | 1.40818 |
| 1.00x10-6 | 111242 | 1.40818 |
| 1.00x10-5 | 111242 | 1.40818 |
| TolX | 1.00x10-7 | 115171\* | 1.40817\* |
| 1.00x10-6 | 111242 | 1.40818 |
| 1.00x10-5 | 111242 | 1.40818 |

The iteration count of the model indicates the number of times the least squares is evaluated by the program before the model run stops (Dahlquist et al., 2018). The iteration count for the run under standard conditions (mid-point value for each parameter) is 111,242. Again, changing the kk\_max to 10 and the TolX to 1.00x10-7 caused a change in the iteration count. The change in iteration count for these parameters was greater than the change that occurred for the LSE:minLSE ratio, especially for kk\_max which showed a 480% increase, while the change in TolX caused a 3.5% increase (Fig 19).  This increase could be expected however, because changing kk\_max to 10 increases the number of times the optimization loop is rerun, which in turn can increase the number of iterations the model runs through. Likewise, decreasing the TolX by one magnitude decreases the maximum allowed difference in the least squares cost before the model stops, which may cause the model to run through more iterations.

 From these sensitivity assessments of the model, it can be determined that changes in the optimization parameters of the model do not impact the fit of the model, but an increase in the kk\_max value does increase the iteration count. Similarly, it was determined that although estimating the weights, thresholds, and production rates resulted in the best fit of the model, fixing the threshold parameter did not impact the model dramatically.

|  |
| --- |
| https://lh6.googleusercontent.com/jTrZ70fbhrew2MmHQ9xkX4jeBRfpqCiG1iXSUm4o4EUhVKmdakgEr7fT8EJ0_erT-ymnNUSdZz6U3qsSzk_XiX95gVr_RadSFhMMSA0_RA9ElPJsyrhVJ2jFyAJtFteE_Qa7rbqGhttps://lh4.googleusercontent.com/88CW7IvbGN64n1Vb4l2028uWsRpynDrtU_AvhjajKK2lDQTf1hxVZ9B49_oF4yRsDULnLEnPeYS3mYHwo9BRL05iEjtw9UCt-8I-iPrXlUFokhZ5hz1pyjtr20vBptnLTQ7dJr5k  **Fig 19:** The LSE:min LSE ratio remained the same for all of the optimization parameter tests except for when kk\_Max was set to 10 and TolX was set to 1.00x10-7 in which case, the ratio decreased. Likewise, the iteration count remained the same except for these same optimization parameter changes. The iteration count increased in both of these cases. The mid-value model is the one run with the standard values for each optimization parameter. |

**Section Five: Revisiting the Dahlquist et al. (2015) GRN model**

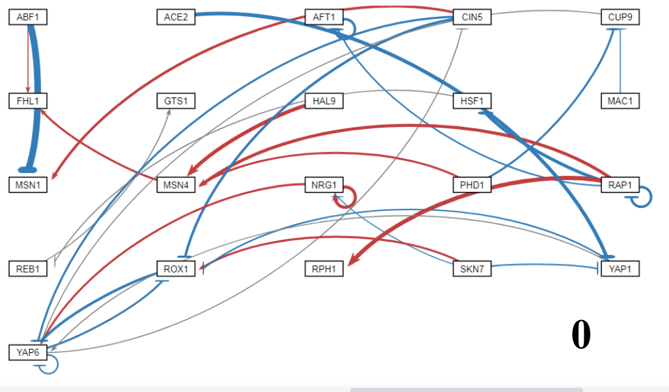
An early version of the GRNmap software was reported in Dahlquist et al (2015). A 21-gene, 31-edge network was derived from the ChIP-chip data reported in Lee et al. (2002) and Harbison et al. (2004). Production rates, weights, and threshold b values were estimated using microarray data from Schade et al. (2004) and protein degradation rates from Belle et al. (2006). Schade et al. (2004) subjected *S. cerevisiae* cells to a cold shock environment (10℃) for 10, 30, and 120 minutes and subsequently harvested the cultures. Results from Dahlquist et al. (2015) can be visualized in GRNsight (Dahlquist et al., 2016; https://dondi.github.io/GRNsight/ demo #4). Since Dahlquist et al. (2015) was published, the GRNmap software evolved to have the following additional features. It can now handle replicate input microarray data instead of just average log2 fold change values, including missing values in the data. It can also handle data from both the wild type strain and multiple deletion strains. Because of the differences in both the data and the GRNmap software, we wanted to revisit the Dahlquist et al. (2015) model to see how changes to the model and data affect the parameter estimation. A series of model runs was undertaken in a stepwise fashion to examine this (Table 8).

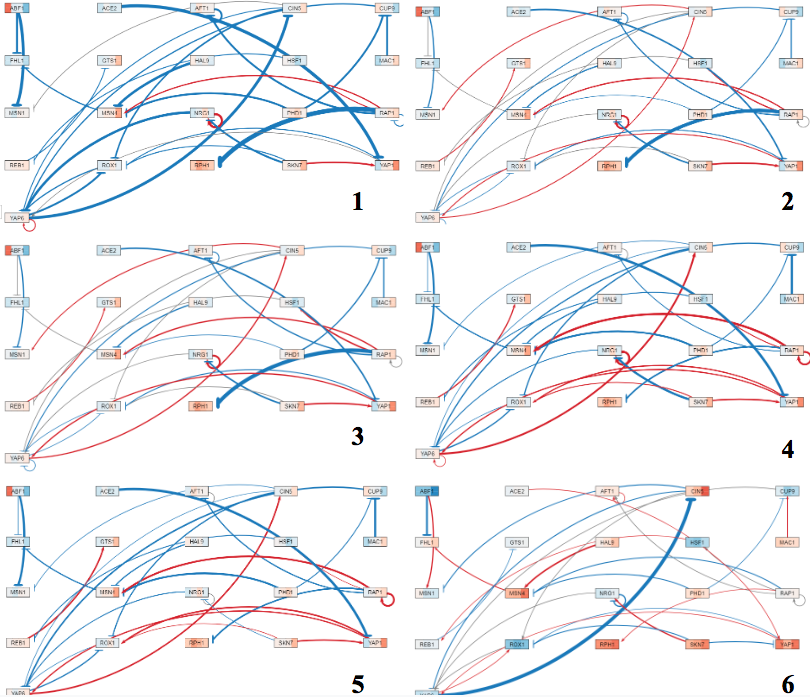
The first model run took the original input file used in Dahlquist et al. (2015) and formatted it for GRNmap v1.10. The second run was performed with the optimization parameters used for the current db5-based models: the alpha to 0.02, MaxIter to 1.00E+08, TolFun to 1.00E-06, MaxFunEval to 1.00E+08, and TolX 1.00E-06. This modification was carried throughout subsequent model runs. The third model run included the replicate expression data from Schade et al. (2004). The fourth and fifth runs had the Belle et al. (2006) degradation rates, which were replaced with the Neymotin et al. (2014) degradation rates. The last model was run with the Schade et al (2004) expression data replaced with the Dahlquist wild type expression data. The Dahlquist lab data was compiled using the MS Access Database found at: <https://github.com/kdahlquist/DahlquistLab/blob/master/data/Spring2019/Expression-and-Degradation-rate-database_2019.accdb>.

|  |  |
| --- | --- |
| **Table 8:** Model runs for the 21-gene, 31-edge network reported in Dahlquist et al. (2015) using GRNmap v1.10 | |
| **Model Run** | **Manipulation to Input Sheet** |
| **1** | Change format to GRNmap v1.10, Schade et al., 2004 microarray data, Belle et al. 2006 degradation rates |
| **2** | Changes from 1 and Optimization Parameters changed (alpha= 0.02, MaxIter = 1.00x10+8, TolFun = 1.0x10-06, MaxFunEval = 1.00x10+08, and TolX = 1.00x10-06) |
| **3** | Changes from 2 and Schade et al. (2004) replicate log2 fold change expression data instead of averages used |
| **4** | Changes from 2 and Belle et al. (2006) degradation rates and production rate initial guess replaced by those from Neymotin et al. (2014) |
| **5** | Changes from 3 and Belle et al. (2006) degradation rates replaced by those from Neymotin et al. (2014) and Schade et al. (2004) replicate log2 fold change expression data used instead of averages used |
| **6** | Schade et al (2004) expression data replaed with Dahlquist wt-only expression data, Belle et al. (2006) degradation rates replaced by those from Neymotin et al. (2014) |

The resulting six model outputs were visualized in GRNsight to compare changes to the regulatory weights between runs (Figure 20). For each of the visualizations, the edge weight normalization factor in GRNsight was set to 2.971, which was the maximum magnitude weight value from the Dahlquist et al. (2015) model (GRNsight demo #4). Certain edge weights changed sign from activation to repression or vice versa. Of the 31 edges, 14 changed sign when the file was reformatted to GRNmap v1.10, 8 changed sign when the optimization parameters were changed, 9 changed sign when replicate data was added, 10 changed sign when the degradation rates were changed to Neymotin et al. (2014) values with no replicate Schade et al. (2004) expression data was added, 14 changed sign when the rates were changed to Neymotin et al. (2014) values and replicate Schade et al. (2004) expression data was added, and 16 edges changed sign when Dahlquist wt-only expression data was used. Therefore, unexpectedly, reformatting the input file to support GRNmap v1.10 greatly impacted the edge weight signs. Compared to the network from the reformatted model run (1), changing the optimization parameters changed only 4 of the edge weight signs, which was expected based on the previous runs where the optimization parameters were manipulated (Section Four). Compared to the network created when the version of GRNmap was changed, 5 edges changed sign when the Schade et al. (2004) replicate expression data was added, 6 edges changed sign when the replicate expression data was added and the degradation rates were changed to Neymotin et al. (2014), the same 6 edges changed no replicate expression data was added and the degradation rates were changed to Neymotin et al. (2014) values, and 17 edges changed sign when only the Dahlquist wt-only expression data were used.

Although it was not expected that changing the format of the input file to support GRNmap v1.10, the network created from the reformatted file caused almost half of the edges to change signs. However, changing the optimization parameters did not have as much of an effect, which was expected as the previous model runs where the optimization parameters were manipulated did not show a major effect on the fit of the model. Changing the data from Schade et al. (2004) expression data to Dahlquist data and also changing the degradation rates caused more edges to change sign, which was again expected because the data were from experiments with different temperatures, strains, and time points. Therefore, the differences in experimental setup impacts the similarities between the weight values, which was reflected in the edges changing signs.





**Figure 20:** GRNsight visualizations of the model runs from Table 2 Schade model runs where: demo network (0), reformatted to GRNmap v1.10 (1), optimization parameters changed (2), Schade et al. (2004) replicate data added (3), Neymotin et al. (2014) degradation rates with no replicate data (4), Neymotin et al. (2014) degradation rates with Schade et al. (2004) replicate data (5), and Dahlquist wt-only expression data with Neymotin et al. (2014) degradation rates (6).

One edge that consistently changed sign is the regulatory relationship between Rap1 and Rph1, which changed from activation to repression for all of the networks, except for the network using Dahlquist microarray data. Another edge, YAP6 → CIN5, consistently showed activation for four of the five new networks, which is different from the published model results (Dahlquist et al. 2015). For the network using Dahlquist data, however, the edge weight became a strong repression. This edge is involved in a complex feedforward loop with ROX1, so it is not surprising that the weight values would be sensitive to changes in input data (both degradation rates and expression data).

From these analyses, it can be determined that certain regulatory relationships in the 21-gene, 31-edge network are sensitive to changes in the input data. Changing the optimization parameters, adding replicate expression data, or changing the degradation rates greatly impacted the network, changing the sign or intensity of the edges.

**Conclusion:**

*Saccharomyces cerevisiae* must adapt to temperature downshifts through a global transcriptional response in order to maintain homeostasis and continue their proliferation and growth. Using wild-type and five transcription factor deletion strains, this global transcriptional response of yeast to cold shock was assessed in this experiment. These gene expression changes in response to cold shock were determined using DNA microarray technology, which reveals the genes that are either upregulated, downregulated, or show no change in expression. In clustering this microarray data, through both STEM and *k-*means clustering methods, the Gene Ontology of the genes that share similar expression patterns were determined. Further, the regulation of the cold shock response can be modelled using a gene regulatory network and the dynamics of this GRN, such as regulatory weights, production rates, and gene expression thresholds, can be estimated using the MATLAB software, GRNmap. Using ordinary differential equations, GRNmap performs estimations of these parameters and the model runs can be compared to determine the fit of the model.

In this paper, multiple sensitivity experiments were run on the model to determine how manipulations of the model input, such as changing optimization parameters and production rates, deleting edges or nodes of the candidate GRN, or including variable strain data, impact the fit of the model. Through these assessments, we determined that the model is not sensitive to changing the optimization parameters, but is sensitive to fixing the production rate values. However, changing the structure of the GRN or including different strain data did impact the model. Through these assessments, certain edges of the candidate GRN were determined to be important in the cold shock response in yeast, such as those involving Msn2, Cin5, or Hmo1. In addition, through deleting the Gcr2 and Zap1 nodes from the network, the fit of the model increased indicating that the model is sensitive to changes in those nodes. Further, through the variable inclusion of strain data, we determined that the inclusion of ∆*gln3* data decreased the fit of the model, indicating that Gln3 plays an important role in the cold shock response. These results indicate that the model is sensitive to changes in GRN structure, strain data, and fixing production rates; but is not sensitive to changes in the optimization parameters.

**Acknowledgements**

I would like to thank Lauren Kelly for her work in running the models for the edge-deletion experiment, the GRNmap lab alumni (Brandon J. Klein and Lauren Kelly), the current Dahlquist lab researchers (Alyssa Weisblatt and Susanne Carpenter), Dr. Kam Dahlquist, Dr. Ben Fitzpatrick, and the LMU Biology Department.

**References:**

Al-Fageeh MB, Smales CM. (2006) Control and regulation of the cellular responses to cold shock: the responses in yeast and mammalian systems. *Biochem J*. 2006;397(2):247–259. doi:10.1042/BJ20060166

Albert, B., Kos-Braun, I. C., Henras, A. K., Dez, C., Rueda, M. P., Zhang, X., ... & Shore, D. (2019). A ribosome assembly stress response regulates transcription to maintain proteome homeostasis. *Elife*, *8*, e45002. DOI: 10.7554/eLife.45002

ASH1. (n.d.). Retrieved on February 28, 2020 from <https://www.yeastgenome.org/locus/YKL185W>

Belle, A., Tanay, A., Bitincka, L., Shamir, R., & O’Shea, E. K. (2006). Quantification of protein half-lives in the budding yeast proteome. *Proceedings of the National Academy of Sciences*, *103*(35), 13004-13009. DOI: 0.1073/pnas.0605420103

Berg JM, Tymoczko JL, Stryer L. Biochemistry. 5th edition. New York: W H Freeman; 2002. Section 16.2, The Glycolytic Pathway Is Tightly Controlled. Available from: https://www.ncbi.nlm.nih.gov/books/NBK22395/

Berger, A. B., Decourty, L., Badis, G., Nehrbass, U., Jacquier, A., & 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

Chen, T., Li, F., & Chen, B. S. (2009). Cross-talks of sensory transcription networks in response to various environmental stresses. *Interdisciplinary sciences: computational life sciences*, *1*(1), 46-54. DOI: 10.1007/s12539-008-0018-1.

CIN5 (n.d.). Retrieved on February 28, 2020 from <https://www.yeastgenome.org/locus/S000005554>

Dahlquist, K. D., Dionisio, J. D. N., Fitzpatrick, B. G., Anguiano, N. A., Varshneya, A., Southwick, B. J., & 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: 10.7287/peerj.preprints.2068v1

Dahlquist, K. D., Fitzpatrick, B. G., Camacho, E. T., Entzminger, S. D., & 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

Ernst, J., & Bar-Joseph, Z. (2006). STEM: a tool for the analysis of short time series gene expression data*. BMC bioinformatics*, 7, 191. https://doi.org/10.1186/1471-2105-7-191

Fernandes, L., Rodrigues-Pousada, C., & Struhl, K. (1997). Yap, a novel family of eight bZIP proteins in Saccharomyces cerevisiae with distinct biological functions. *Molecular and cellular biology*, *17*(12), 6982-6993. Doi: 10.1128/mcb.17.12.6982

Gasch, A. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., Storz, G., ... & 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

Gasch, A. P., & Werner-Washburne, M. (2002). The genomics of yeast responses to

environmental stress and starvation. Functional & integrative genomics, 2(4-5), 181-192. DOI: 10.1007/s10142-002-0058-2

Gene Ontology Consortium. (2015). Gene ontology consortium: going forward. *Nucleic acids*

*research*, *43*(D1), D1049-D1056. DOI: 10.1093/nar/gku1179

GLN3. (n.d.). Retrieved on February 28, 2020 from

<https://www.yeastgenome.org/locus/S000000842>

HMO1 (n.d.). Retrieved on February 28, 2020 from <https://www.yeastgenome.org/locus/S000002581>

Homma, T., Iwahashi, H., & Komatsu, Y. (2003). Yeast gene expression during growth at low temperature. Cryobiology, 46(3), 230-237. DOI: 10.1016/S0011-2240(03)00028-2

Kelly, L. M. (2019). *Modeling of Gene Regulatory Network Dynamics Predicts which Regulatory Relationships are Important for Controlling the Cold Shock Response in Saccharomyces cerevisiae* (Unpublished thesis). Loyola Marymount University, Los Angeles, CA.

Maere S, et al. (2005) ClueGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in biological networks. *Bioinformatics*. 21:3448–3449. DOI: 10.1093/bioinformatics/btp101

Mager, W. H., & De Kruijff, A. J. (1995). Stress-induced transcriptional activation. *Microbiol.*

*Mol. Biol. Rev.*, *59*(3), 506-531.

Monteiro, P. T., Oliveira, J., Pais, P., Antunes, M., Palma, M., Cavalheiro, M., ... & Mota, M. N. (2020). YEASTRACT+: a portal for cross-species comparative genomics of transcription regulation in yeasts. Nucleic acids research, 48(D1), D642-D649. [doi:10.1093/nar/gkz859](https://doi.org/10.1093/nar/gkz859)

Murata, Y., Homma, T., Kitagawa, E. et al. Extremophiles: Genome-wide expression analysis of yeast response during exposure to 4C (2006) 10: 117. doi:10.1007/s00792-005-0480-1.

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

Salvadó, Z., Arroyo-López, F. N., Guillamón, J. M., Salazar, G., Querol, A., & Barrio, E. (2011).

Temperature adaptation markedly determines evolution within the genus Saccharomyces. Appl. Environ. Microbiol., 77(7), 2292-2302. DOI: 10.1128/AEM.01861-10

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

Schena, M. (1996). Genome analysis with gene expression microarrays. Bioessays, 18(5), 427

431. DOI: 10.1002/bies.950180513

SFP1 (n.d.). Retrieved on February 28, 2020 from <https://www.yeastgenome.org/locus/S000004395>

Stanford University. (n.d.). Saccharomyces Genome Database. Retrieved on February 28, 2020 from <https://www.yeastgenome.org/>

Tai, S. L., Daran-Lapujade, P., Walsh, M. C., Pronk, J. T., & Daran, J. M. (2007). Acclimation of

Saccharomyces cerevisiae to low temperature: a chemostat-based transcriptome analysis. *Molecular Biology of the Cell*, 18(12), 5100-5112. DOI: 10.1091/mbc.E07-02-0131

Tate J. J., Cooper T. G., 2007.  Stress-responsive Gln3 localization in *Saccharomyces cerevisiae* is separable from and can overwhelm nitrogen source regulation. J. Biol. Chem. doi: 18467–18480.

UniProt (2019). AP-1-like transcription factor YAP4. Retrieved on February 28, 2020 from https://www.uniprot.org/uniprot/P40917

Wagstaff, K., Cardie, C., Rogers, S., & Schrödl, S. (2001, June). Constrained k-means clustering with background knowledge. In Icml (Vol. 1, pp. 577-584).

YHP1 (n.d.). Retrieved on February 28, 2020 from <https://www.yeastgenome.org/locus/S000002859>

YOX1 (n.d.). Retrieved on February 28, 2020 from <https://www.yeastgenome.org/locus/S000004489>

ZAP1 (n.d). Retrieved on February 28, 2020 from

<https://www.yeastgenome.org/locus/S000003592>

**Appendix:**

https://docs.google.com/document/d/1wLl84bKPZzHhDU8if-JbpDY-ZwwNRsNavYkZaoVv8xw/edit?usp=sharing