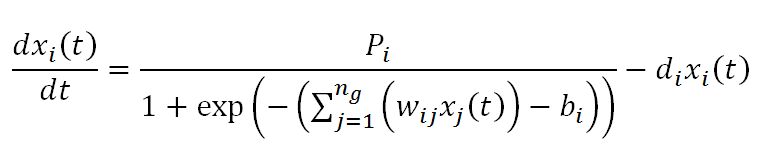
Alice Finton

BIOL 588

March 27, 2020

**Section 3: Data Deletion Sensitivity Tests**

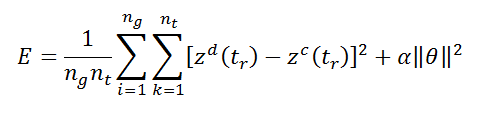
A gene regulatory network (GRN) depicts the regulatory relationship between transcription factors illustrating gene expression mechanism control. The dynamics of GRNs can offer information about the changes in gene expression over time in response to various environmental stimuli. GRNmap (Gene Regulatory 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. GRNmap models gene expression change as the production of mRNA minus the degradation using a sigmoidal production function,



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

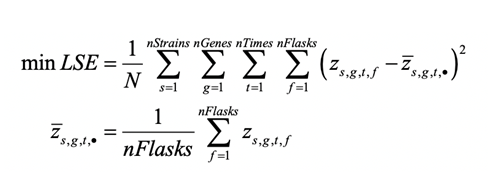
where *Pi* is the mRNA production rate of the gene, *di* is the mRNA degradation rate of the gene, *b* is the expression threshold, and *w* is the regulatory weight of the gene (Dahlquist et al., 2016).

The model uses a penalized least squares approach to estimation. 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) (Dahlquist et al., 2016).



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

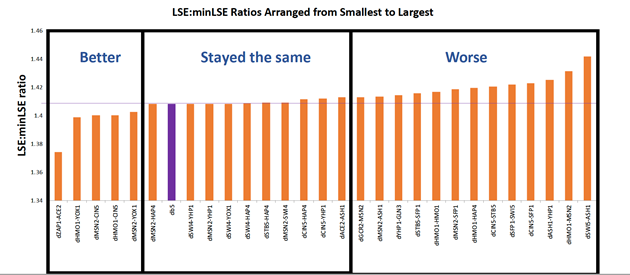
The candidate gene regulatory network (GRN) used in data deletion sensitivity tests was db5, which consists of 15 nodes, representing transcription factors, and 28 edges, representing a regulatory relationship (Fig 1). The GRN was derived through inputting the DNA microarray data into the YEASTRACT (Yeast Search for Transcriptional Regulators and Consensus Tracking) database, which generates a list of transcription factors that potentially regulate genes in the data (P.T. Monteiro et al., 2020). Using genes that showed significant expression changes (p-value<0.05) in the microarray data, the candidate GRN was created.

|  |
| --- |
| **Fig 1:** Intact (all-strain) db5 gene regulatory network generated using GRNsight. The GRN has 15 nodes (transcription factors) and 28 edges (regulatory relationships). |

In order to assess the sensitivity of the model, the db5 candidate network was manipulated through systematic edge deletions and variable inclusion of strain data. In addition, certain nodes were deleted from the network to determine the fit of the model in response to the manipulation.

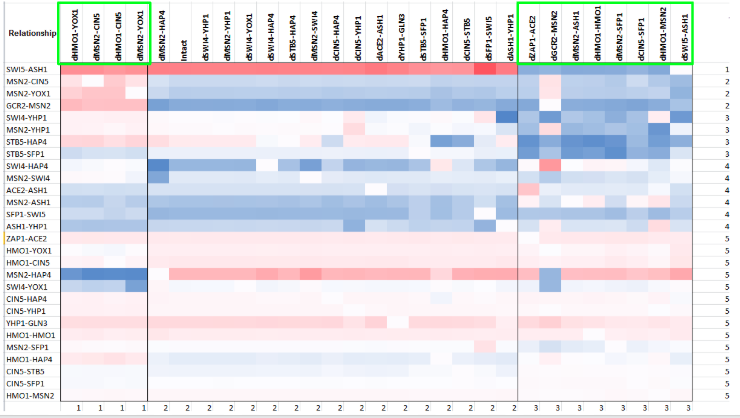
**Systematic Deletion of Edges in the db5 Network**

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 by GRNmap. In order 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 reduced ratio indicates that the model is a better fit than the intact network. 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 2). 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 2:** LSE:minLSE ratios for each of the 28 runs, which are ordered by increasing ratio.

Further, *k*-means clustering was performed on the edge weight values from the intact network and edge-deletion networks. An examination of the clusters showed that 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 3). In addition, the edge-deletions involving Gcr2 and Zap1 caused major deviation in weight values from the intact network (Fig 3). This edge-deletion analysis reveals the importance of these regulatory relationships in the cold shock response.



**Figure 3:** *k*-means clustering resulted in three distinct clusters. The intact network is included in the second cluster. Edge deletions including Gcr2, Msn2, and Zap1 were consistently found in clusters one and three, indicating deviation from the intact network.

These results indicate that the transcription factors Msn2, Hmo1, andCin5 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 another study, the deletion of the Hmo1 gene caused a decrease in the transcription of genes indirectly associated with ribosome biogenesis (Berger et al., 2007). Therefore, Hmo1 is an important transcription factor in the cold shock response, as ribosome biogenesis has been shown to be an overrepresented upregulated functional category. Further, Cin5 has been found to be activated when there is a change in the environment, whether that be temperature, carbohydrate availability, etc.. It has also been associated with increasing ribosome biogenesis and protein synthesis (Uniprot, 2019). Therefore, when the edge deletions included Msn2, Hmo1, or Cin5 they deviated from the all-strain due to the importance of these TFs in controlling the stress response.

Further, when the transcription factors Gcr2 and Zap1 were not included in the model, the major weight deviations indicate that they play a role in the stress response. Gcr2 is a transcription factor which activates glycolytic genes, thereby increasing carbohydrate metabolism and energy production. In a previous study, Gcr2 was found to be the only consistently upregulated stress-response regulator in all of the environmental stress conditions (Chen et al., 2009). 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.

|  |
| --- |
|  |

**Deletion of Gcr2 and Zap1 Transcription Factors from db5 Network**

The deletion of edges containing Gcr2 or Zap1 transcription factors caused major deviations in the weight values for the db5 network. When the edges not included in the cluster containing the intact network were deleted from the GRN, the resulting network consisted of 16 edges. In this GRN, Gcr2 and Zap1 nodes were left floating in the network, indicating that all of their regulatory relationships deviated from the intact network (Fig 4). Therefore, in order to determine the impact of these TFs on the fit of the model, two runs were conducted with a reduced network, where the edges that were not in the clusters were deleted from the input. One model was run where Gcr2 and Zap1 were completely deleted from all of the sheets in the excel workbook and one where the *Δzap1* data sheet was completely deleted. In the last model run, the network was left intact, but Gcr2 and Zap1 were deleted from the matrix, making a 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

|  |
| --- |
| **Fig 4:** Modified gene regulatory networks where: intact network (1), weights in clusters 1 and 3 were put close to zero (2), and edges not in the main cluster deleted (4). |

|  |  |  |
| --- | --- | --- |
| **Table 1:** 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. | | |
| **Model** | **Edges** | **Data** |
| **16-edge** | Those within the same k-means cluster as the intact network were included | All data |
| **26-edge** | Edges involving Gcr2 and Zap1 were deleted | All data |
| **28-edge no Zap1** | All edges | *Δzap1* data deleted |
| **Intact** | All edges | All data |

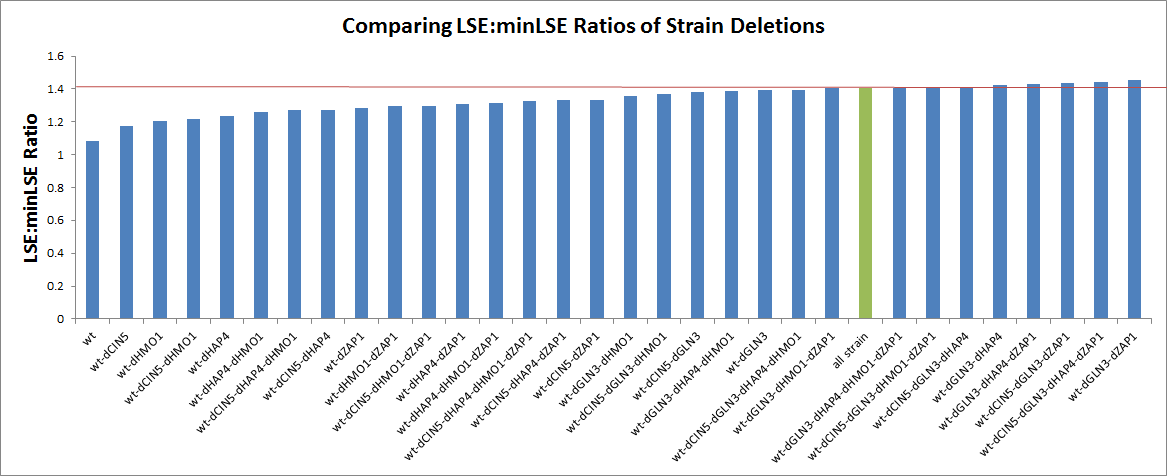
When all edges that were not included in the main cluster were deleted from the network, the LSE:minLSE ratio was about the same as the intact network, indicating that there was not a major difference in the fit of the model. When only the two edges involving Gcr2 and Zap1 were deleted, the LSE:minLSE ratio decreased, 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 5).

|  |
| --- |
| **Fig 5:** 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**

A systematic deletion of mutant strain data resulting in 32 model runs was conducted to determine the importance of the strain on the fit of the model. 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 of the runs can be found in the appendix). A re-estimation of model parameters, including production rates, expression thresholds, and regulatory weights, was performed using differential equations in GRNmap.

In order to assess the fit of the model, the LSE:minLSE ratio was determined and compared between the 32 model runs. 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 the Gln3 deletion strain. The LSE:minLSE ratio decreased for 23 new 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 caused the best fit, while the inclusion of Gln3 deletion strain data caused the model to run about the same as the all-strain network or worse. This indicates that Gln3 is an important transcription factor for the cold-shock response.



**Fig 6:** LSE:minLSE ratios for each of the model runs is given in the graph. 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.

Further, the output excel sheets were input into GRNsight, which generates a gene regulatory network with the optimized regulatory weights (Dahlquist et al., 2016). The all-strain intact network was compared to the model that ran the best, wt-only; the model that ran the worst, wt-dGLN3-dZAP1; and the network that did not include Gln3 deletion strain data, wt-dCIN5-dHAP4-dHMO1-dZAP1. Compared to the all-strain network, the wt-only GRN had 15 edges that changed (Fig 7). The wt-dGLN3-dZAP1 model performed the worst, with the highest of the LSE:minLSE ratios. When compared to the all-strain network, 16 of the 28 edges changed (Fig 8). When dGLN3 data was included in the input, the model consistently performed worse. Therefore, a comparison was made between the all-strain and model where only *Δgln3* data was not included. The wt-dCIN5-dHAP4-dHMO1-dZAP1 network had four edges that changed compared to the all-strain network (Fig 9). These results indicate that when deletion strain data was not included in the model, all networks showed deviations in the edges.





|  |
| --- |
| **Fig 9:** 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. |

The initial heat map generated with models organized randomly showed no major clusters. Therefore, it was necessary to cluster the heat map using *k*-means clustering in MATLAB. In clustering the heat map in both dimensions, a cluster containing the all-strain, wt-dCIN5-dHAP4-dHMO1-dZAP1, and wt-dCIN5- dHAP4-dHMO1 model runs was generated. The two models that were most similar to the all-strain’s regulatory weights were the models that either excluded *Δgln3* strain data or *Δgln3* and *Δzap1* data (Fig 10). The runs that included *Δgln3* showed the highest LSE:minLSE ratio, indicating the worst fit. This trend is reflected in the comparison of the GRNs, where the regulatory weights 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 9). 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 a better fit. However, when *Δ*cin5, *Δhmo1*, and *Δzap1* data were not included, the simulated model data did diverge, indicating a worse fit (Fig 11). Therefore, again, when *Δgln3* data was included, the model did not perform well.

|  |
| --- |
| **Fig 10:** 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 indicated by the black line. |

|  |
| --- |
| **Fig 11:** Expression plots for candidate model runs. The inclusion of dCIN5, dHMO1, or dZAP1 data caused the simulated model to diverge. 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, YeastGenome.com). 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.*

**Works Cited:**

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.

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.

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.

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.

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.

GLN3. (n.d.). Retrieved from <https://www.yeastgenome.org/locus/S000000842>

HMO1 (n.d.). Retrieved from <https://www.yeastgenome.org/locus/S000002581>

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.

P.T. Monteiro, J. Oliveira, P. Pais, M. Antunes, M. Palma, M. Cavalheiro, M. Galocha, C.P.

Godinho, L.C. Martins, N. Bourbon, M.N. Mota, R.A. Ribeiro, R.Viana, I. Sá-Correia, M.C. Teixeira (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)

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.

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 ConsortiumEuropean Bioinformatics InstituteProtein Information ResourceSIB Swiss

Institute of Bioinformatics. (2019, December 11). AP-1-like transcription factor YAP4. Retrieved from https://www.uniprot.org/uniprot/P40917

ZAP1 (n.d). Retrieved from https://www.yeastgenome.org/locus/S000003592

**Appendix:**

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