Alice Finton

BIOL 588

February 28, 2020

**Gene Ontology Analysis of STEM Profiles and *k*-Means Clustering of Microarray Data**

**DNA Microarray Data**

Using DNA microarrays, the gene expression data of *Saccharomyces cerevisiae* wild type and deletion strains (*Δcin5, Δgln3, Δhap4, Δhmo1,* and *Δzap1*) was collected from the Dahlquist lab. Thestrains were subjected to cold shock at 13oC for one hour and subsequent recovery at 30oC for another hour. Samples were taken after 15, 30, and 60 minutes of cold shock and 30 and 60 minutes of recovery. These t15, t30, t60, t90 and t120 samples were then used for analysis.

**STEM Profile and Gene Ontology Analysis**

*STEM Profile*

The expression patterns of the 6189 genes on the DNA microarray chip can be clustered and similarities between the genes can be analyzed using STEM. Short-Time Expression Minor (STEM) is a software that can be used to analyze gene expression patterns from microarray data (Ernst and Bar-Joseph, 2006). STEM uses large gene sets and short time point data to determine the statistically significant expression from the data and groups genes by their expression patterns (Ernst and Bar-Joseph, 2006). These unique expression patterns are shown through profiles which depict the time course expression.

The microarray data for the five strains, excluding *Δhmo1* data, was input into STEM and a comparison of the resulting profiles was conducted to determine commonalities between the strains. An analysis of the STEM profiles revealed that three profiles— profile 45, 9, and 22— were common to all of the deletion strains. Other profiles had overlap between the deletion strains; such as profiles 48, 2, 7, 40, 0, and 28; while others were unique to a single strain, like profiles 31 and 38 (Fig 1).

Profile 45 is characterized by an initial increase in gene expression during cold shock followed by a decrease during the recovery period (Fig 1). 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 1).

|  |
| --- |
| **Fig 1:** The significant STEM results for the gene clustering profiles, with gene expression trend lines, indicating commonality between the genes in the cluster. Profiles included are listed in the top row and strains are listed in the first column. |

*ClueGO Gene Ontology Analysis*

The gene ontology (GO) analysis for the strains was conducted using ClueGO, a Cytoscape plugin that creates gene ontology networks from the overrepresented GO terms from each strain and allows for a comparison between two lists of data (Maere et al., 2005).The plugin interprets biological information and visualizes the functional groups of terms through networks and charts, using kappa statistics to generate these network pathways and links between GO terms. The gene lists and ontology terms for each strain within STEM profiles 45, 9, and 22 were input into ClueGO and the resulting GO networks were used to determine the overrepresented functional categories of each strain. (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 functional categories included those involved in ribosome biogenesis and RNA processing. In a previous study, ribosome biogenesis and transcription-related 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 the present study, it was found that these were upregulated initially then downregulated during the recovery period, which is consistent with the previous findings. However, protein synthesis and processing have been found to be downregulated during this period (Murata et al., 2006). The upregulation of transcription-related genes is inconsistent with these findings.

|  |
| --- |
| **Fig 2:** GO networks representing the overrepresented functional categories for each of the strains in profile 45. |

Profile 9 contains genes that were downregulated during the cold shock time points and increased back to normal expression during the recovery period. For this profile, the overrepresented functional categories common to the strains were involved in catabolic and metabolic processes, such as amino acid and carboxylic acid metabolism, and cell cycle regulation. In a previous study, consistently upregulated genes during the time course of cold shock included those involved in energy and metabolism functions, which is inconsistent with these findings (Murata et al., 2006). In the early cold shock response, it has been found that genes associated with lipid and amino acid metabolism and transcription are upregulated, which is, again, inconsistent with our findings (Schade et al., 2004).

|  |
| --- |
| **Fig 3:** GO networks representing the overrepresented functional categories for each of the strains in profile 9. |

Profile 22 is defined by an upregulation of genes during the recovery period. Within this profile, the common overrepresented functional categories included those involved in cell-aging and stress responses, such as desiccation and oxidative stress. Therefore, during the recovery period, the expression of these genes increased. In a previous study, genes associated with cell rescue, defense, and virulence were consistently upregulated during the cold shock response (Murata et al., 2006). These results support the findings of this study.

|  |
| --- |
| **Fig 4:** GO networks representing the overrepresented functional categories for each of the strains in profile 22. |

**Clustering Microarray Data and Gene Ontology Analysis**

*k-means Clustering of the Microarray Data*

The raw microarray data contained missing values, therefore, 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 across all of the strain data were filtered out, resulting in 2441 genes left in the data. The leftover significant genes were loaded into MATLAB and clustered into ten clusters using *k*-means clustering. The gene clusters were then reordered by similarity and the strains were ordered by time point and strain. The resulting data was input into MATLAB and a heatmap was generated, with green indicating downregulation and red indicating upregulation.

|  |
| --- |
| **Fig 5:** Clustered heatmap of microarray data with associated cluster number designated below the heatmap. Strain data ordered by time point and strain 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. |

From the clustered heatmap, it can be determined that there are differences in the expression of genes during cold shock versus the recovery period. Certain clusters showed more severe change in expression, such as in clusters 2, 9, 8 and 1, 5, 7 (Fig 5). Other clusters did not have as strong of a change, as shown in clusters 10, 4, and 6 (Fig 5).

|  |  |  |
| --- | --- | --- |
| **Table 1:** Trends in upregulation or downregulation for each of the clusters in the cold shock and recovery period. | | |
| **Cluster** | **Cold Shock** | **Recovery** |
| **1** | Up | Down |
| **2** | Down | Up |
| **3** | Up | Down |
| **4** | Some Up | Some Up |
| **5** | Up | Down |
| **6** | Up | Some Up |
| **7** | Up | Down |
| **8** | Down | Up |
| **9** | Down | Up |
| **10** | Some Down | Some Down |

The presence of transcription factors included in the candidate db5 network (Ace2, Ash1, Cin5, Gcr2, Gln3, Hap4, Hmo1, Msn2, Sfp1, Stb5, Swi4, Swi5, Yhp1, Yox1, and Zap1) was checked in order to determine their changes in expression. ACE2, GCR2, GLN3, HAP4, STB5, SWI4, SWI5, and ZAP1 were not left in the heatmap because the expression change was not significant in at least one strain. ASH1 was included and present in cluster 10. This indicates that ASH1 expression change was significant, 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 were included in cluster 6, which indicates that the expression of these genes was increased during the cold shock timepoints and a slight upregulation during the recovery period. Cin5 is a transcription factor that 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). MSN2, SFP1, and YHP1 were included in cluster 4, which indicates that there was no major change in expression during the cold shock and recovery time points, but there was some downregulation seen in this cluster. Msn2 is associated with the cold shock response through its induction of stress response genes by binding to stress response elements (Schade et al., 2004). Sfp1 regulates ribosome biogenesis genes and has been shown to be associated with the regulation of the stress response (Sfp1, SGD). Yhp1 negatively regulates transcription during the G1/S stage of mitosis (Yhp1, SGD). Lastly, YOX1 was included in cluster 2, which indicates that the expression of this gene was decreased during cold shock and increased during the recovery period. Yox1, like Yhp1, is associated with the negative regulation of transcription during the G1/S stage of mitosis (Yox1, SGD).

|  |  |  |
| --- | --- | --- |
| **Table 2:** The presence or absence of transcription factors included in the db5 GRN. The cluster that the present genes were found is included. | | |
| **Gene** | **Presence** | **Cluster** |
| ACE2 | No | - |
| ASH1 | Yes | 10 |
| CIN5 | Yes | 6 |
| GCR2 | No | - |
| GLN3 | No | - |
| HAP4 | No | - |
| HMO1 | Yes | 6 |
| MSN2 | Yes | 4 |
| SFP1 | Yes | 4 |
| STB5 | No | - |
| SWI4 | No | - |
| SWI5 | No | - |
| YHP1 | Yes | 4 |
| YOX1 | Yes | 2 |
| ZAP1 | No | - |

The Gene Ontology Consortium is a database that utilizes past research to annotate genes and proteins with their biological roles through gene ontologies (Gene Ontology Consortium, 2015). The list of genes in each of the clusters was loaded into the Gene Ontology Consortium website (<http://geneontology.org/>) and the list of significant (p<0.05) gene ontology terms was generated. Cluster 4 did not enrich any significant GO terms, but the rest of the clusters had overrepresented functional categories.

Clusters 1, 3, 5, and 7 contain genes that were upregulated during cold shock and downregulated during the recovery period. Cluster 1 had overrepresented functional categories that are involved in ribosome biogenesis and RNA metabolic processes. These results are consistent with the previous findings in the STEM profiles. The consistency of 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 contains overrepresented functional categories involved in transcription, namely RNA metabolic processes. 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 contains 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 make more energy and create reserves of carbohydrates for later use. Lastly, cluster 7 had overrepresented functional categories that were 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). However, in another study, it was determined that an upregulation of genes associated with zinc homeostasis has been shown in the heat shock response, but not in the cold shock response, which is inconsistent with these findings (Pirev et al., 2010).

Clusters 2, 8, and 9 contain genes that were downregulated during cold shock and upregulated during recovery. Cluster 2 contained overrepresented functional categories that are associated with oxidation-reduction and phosphorus metabolic processes. Cluster 8 had overrepresented functional categories involving protein folding, amino acid metabolism, and RNA processing. These results are consistent with the STEM profile 9, which showed a downregulation of amino acid metabolism, but are inconsistent with previous findings with report an upregulation if these genes (Schade et al., 2004). In addition, the decrease in RNA processing is consistent with previous studies which found that protein synthesis and processing are downregulated during this period (Murata et al., 2006). Cluster 9 had only two GO terms that were enriched for this cluster: 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 a previous study, which determined that energy and metabolism functions increase during cold shock (Murata et al., 2006).

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 severe upregulation during the cold shock time points. Only one GO term was enriched for this cluster: mitochondrial translation. These results are inconsistent with previous studies, but consistent with our STEM profile, which showed that protein synthesis was upregulated during cold shock (Murata et al., 2006). There was an overall downregulation for the genes in cluster 10 for both the cold shock and recovery time points. This cluster had overrepresented functional 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 previous studies, which indicate amino acid metabolism and protein synthesis are upregulated during cold shock (Scade et al., 2004).

**Conclusion**

Clustering the DNA microarray via STEM profile and *k-*means clustering produced similar findings in terms of gene ontology analysis, with certain profiles and clusters enriching the same or similar functional categories. It was determined that ribosome biogenesis and RNA metabolic processes were commonly upregulated during cold shock and subsequently downregulated during the recovery period. In addition, carbohydrate and lipid metabolism were found to be similarly upregulated in cold shock. There was more variance in the functional categories that were found to be downregulated in cold shock. Amino acid metabolism was consistently downregulated during cold shock, but cell cycle regulation was a term unique to the STEM profile and metabolite precursors, RNA processing, and translational termination were unique terms to the *k-*means clusters. These results reveal the importance of ribosomal processes and carbohydrate metabolism in the cell’s adaptation to cold shock.

**References:**

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.

Al-Fageeh MB, Smales CM. 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

ASH1. (n.d.). Retrieved from <https://www.yeastgenome.org/locus/YKL185W>

Berg, J. M., Tymoczko, J. L., & Stryer, L. (2002). Section 16.2, The Glycolytic Pathway Is Tightly Controlled. *Biochemistry,*.

CIN5 (n.d.). Retrieved from <https://www.yeastgenome.org/locus/S000005554>

Ernst, J., & Bar-Joseph, Z. (2006). STEM: a tool for the analysis of short time series gene

expression data. *BMC Bioinformatics*, *7*(1), 191.

Gene Ontology Consortium. (2015). Gene ontology consortium: going forward. *Nucleic acids research*, *43*(D1), D1049-D1056.

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

Maere S, et al. (2005) ClueGO: a Cytoscape plugin to assess overrepresentation of Gene

Ontology categories in biological networks. *Bioinformatics*. 21:3448–3449.

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.

Pirev, E., Ince, Y., Sies, H., & Kröncke, K. D. (2010). Heat shock but not cold shock leads to disturbed intracellular zinc homeostasis. *Journal of Cellular Physiology*, *223*(1), 103-109.

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.

SFP1 (n.d.). Retrieved from <https://www.yeastgenome.org/locus/S000004395>

Stanford University. (n.d.). Saccharomyces Genome Database. Retrieved from https://www.yeastgenome.org/

YHP1 (n.d.). Retrieved from <https://www.yeastgenome.org/locus/S000002859>

YOX1 (n.d.). Retrieved from <https://www.yeastgenome.org/locus/S000004489>

**Appendix:**

[**https://docs.google.com/document/d/1wLl84bKPZzHhDU8if-JbpDY-ZwwNRsNavYkZaoVv8xw/edit**](https://docs.google.com/document/d/1wLl84bKPZzHhDU8if-JbpDY-ZwwNRsNavYkZaoVv8xw/edit)