Potassium Starvation in Yeast Triggers Changes in the Expression of Genes Related to Different Metabolic and Biosynthetic Pathways

Aby Mesfin, Jonar Cowan, David Ramirez, Mihir Samdarshi, Christina Dominguez
Loyola Marymount University
BIOL 367-01
December 13, 2019

Introduction

Potassium has been proven to be important to normal cell function. As the most abundant cation in the cell, it is involved in numerous different pathways that regulate the cell. Potassium is necessary for the maintenance of certain metabolic properties and metabolic pathways within cells. The cation, for instance, is responsible for generating the membrane potential that exists within cells, along with other cations such as sodium (Adrian, 1956). Without potassium balance, threatening conditions occur both to the cell. Potassium in the cell is also involved with other cations, such as ammonium. The effects of potassium have been shown to limit growth when ammonium is used as a nitrogen source as well as to induce ammonium toxicity under high potassium conditions due to over-expression of ammonium transports (Hess et al 2006). Due to potassium's importance in the cell and relation to other cations, understanding the regulation of transcriptional responses in yeast in relationship to potassium levels is crucial.

This study analyzes the data presented in "The Short-Term Response of Yeast to Potassium Starvation" by Barreto L., Canadell, D., Valverde-Saubí, D., Casamayor, A., and Ariño, J., who examined how potassium starvation induces changes in gene expression and how it relates to cellular pathways. They studied potassium starvation in yeast cells for 10, 20, 40, 60, and 120 minutes and analyzed induction and repression of genes through microarray analsyis. The results of Barreto, et al. found that potassium starvation is linked to changes in the expression of genes associated with oxidative stress, methionine/cysteine biosynthesis, cyclin levels, septin rings, retrograde pathway, methylglyoxal production, and trehalose metabolism. The roles of these metabolic pathways in the yeast cell are essential to cellular function.

These findings of Barreto, et al. expanded on results from other studies. A study on halotolerant yeast *Debaryomyces hasenii* found that potassium stress regulated amino acid synthesis, which could be involved in pathways of methionine/cysteine biosynthesis found in the Baretto results (Martinez et al 2012). Similarly, another study on the relationship between potassium limitation and its effects on the production of ammonium toxicity also links potassium starvation with changes in gene expression as seen in the Baretto study (Hess et al 2006). Recently, septin has been documented to play an important role in the cell cycle and cell shape in yeast. Well-assembled septin cortex have been found to be necessary in developing yeast shapes. When septin is disorganized by the absence or lack of a critical nutrients like potassium, then the septin cortex is unable to grow a healthy bud (Gladfelter et al 2005). Therefore, as stated in the Baretto, et al., a lack of potassium inhibits the growth of yeast by causing changes in the expression of septin rings.

This study reanalyzes the data presented by Barreto microarray results in order necessary in order to better understand the transcriptional responses in yeast during potassium starvation. Understanding how transcription factors are involved in a gene regulatory network related to potassium concentrations in the cell will aid in expanding knowledge about potassium's role in cellular function. GRNmap and GRNsight were used to create this regulatory network to demonstrate the relationships among genes during potassium starvation from the microarray data. A MS Access database was also used in order to run GRNmap. Doing so facilitates its current and future usage to better understand and explore new information from microarray data. This paper aims to discover the gene regulatory networks induced by potassium starvation from

the Barreto, et al. microarray data in order to expand upon the existing knowledge regarding the function of potassium in the cell.

Methods/Results/Discussion

Figure 1: A flow chart of the project illustrating sequential steps to creating the gene regulatory network and problems encountered throughout.

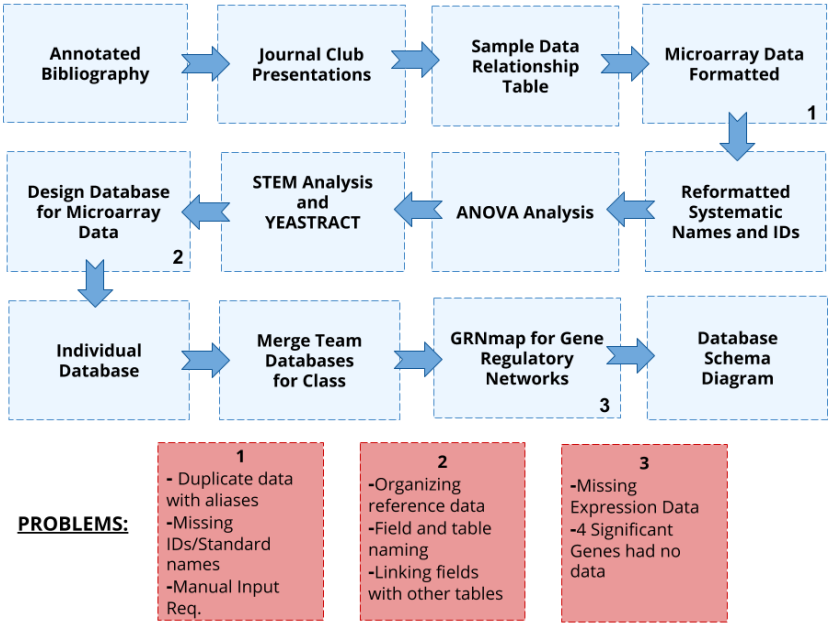


Table 1: Sample-relationship table formatted to regularize data into sample, strain, treatment, timepoint, and replicate.

Sample	Strain	Treatment	Timepoint (min)	Replicates
wt, K+ free media, 10min	wt	K+ Free Media	10	2
wt, K+ free media, 20min	wt	K+ Free Media	20	4
wt, K+ free media, 40min	wt	K+ Free Media	40	4
wt, K+ free media, 60min	wt	K+ Free Media	60	4
wt, K+ free media, 120min	wt	K+ Free Media	120	4

Table 2: ANOVA results showing number and percent of genes that had relative p-values from the microarray dataset.

ANOVA	Percent of Genes
P < 0.05	2,985 (71.9 %)
P < 0.01	2,403 (57.9 %)
p < 0.001	1,703 (41 %)
p < 0.0001	1,198 (28.9 %)
B & H p < 0.05	2,839 (68.4 %)
Bonferroni p < 0.05	776 (18.7 %)

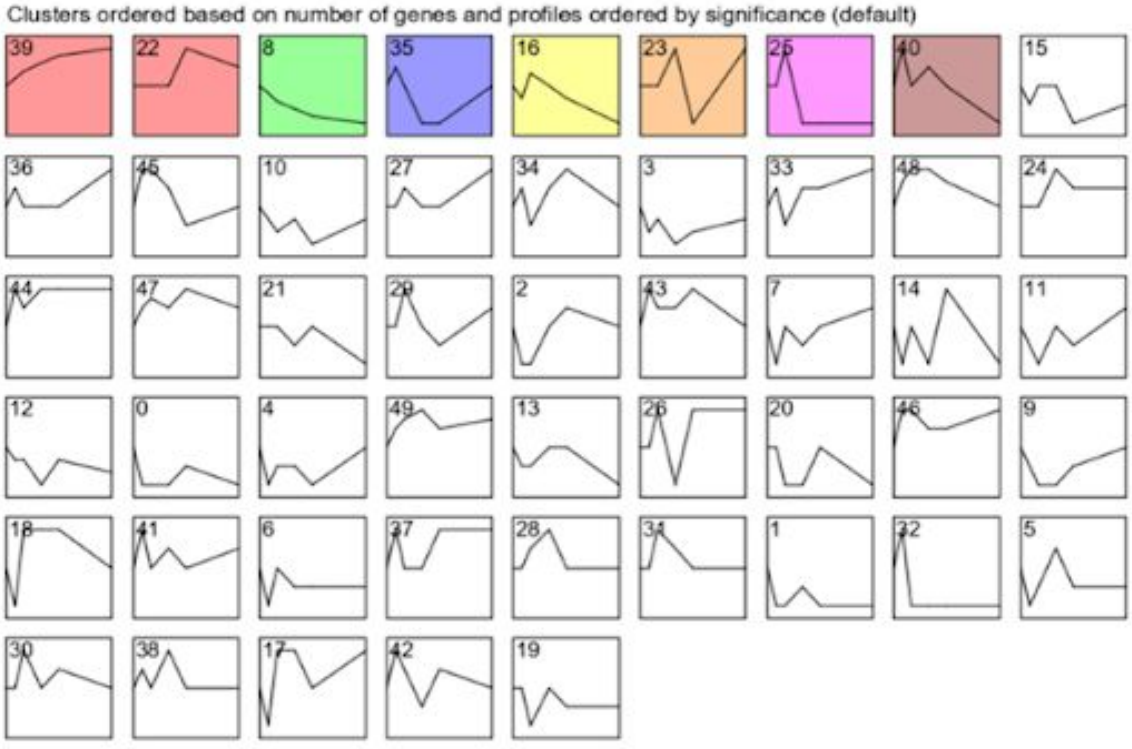


Figure 2: Eight significant gene clusters were observed. Profile 39 was the most significant cluster and was used for the rest of the study.

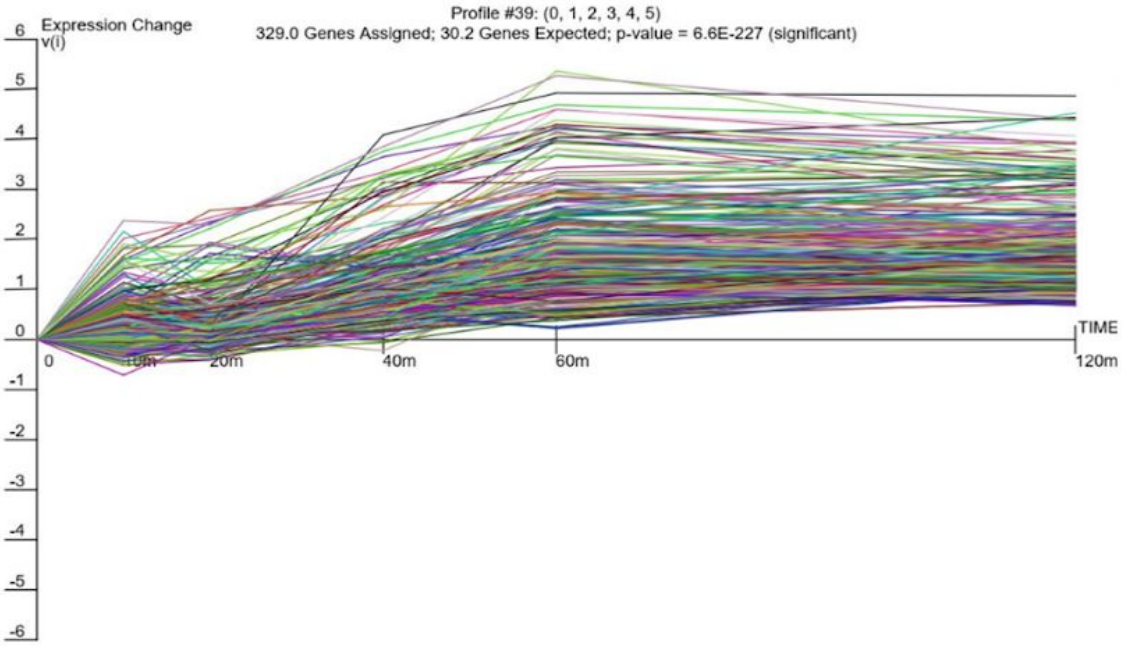


Figure 3: Gene plot of Profile 39 displaying significant changes of all 329 assigned genes at every time-point

Table 3: Gene ontology results from GO terms show the number of genes in the yeast cells from Profile 39 with significant p values categorized according to different cellular function.

Category Name	**#Genes Category**	**#Genes Assigned**	**#Genes Expected**	**#Genes Enriched**	**p-value**	**Corrected p-value**	**Fold**
Cysteine biosynthetic process	10	9	1.2	7.8	3.1E-08	<0.001	7.8
Cellular amino acid biosynthetic process	67	24	7.8	16.2	1.4E-07	<0.001	3.1
Methionine biosynthetic process	26	13	3	10	1.4E-06	<0.001	4.3
Hydrogen sulfide biosynthetic process	6	6	0.7	5.3	2.3E-06	<0.001	8.6
Molecular function	694	114	80.4	33.6	6.2E-06	<0.001	1.4
Oxidation-reduction process	211	45	24.5	20.5	1.8E-05	0.004	1.8

Table 4: Top 15 Significant Transcription Factor genes from Yeastract. Fold change data for RPN4, PDR1, PDR3, and HSF1 genes were removed from the list because the data was not measured in the study. Therefore, the query results for these four genes were empty

Transcription Factor	% in user set	% in s. cerevisiae	p-value	Transcription Factor	% in user set	% in s. cerevisiae	p-value
Gcn4p	98.78 %	5.37%	0	Hsf1p	50.15 %	9.17%	0
Gat1p	60.79 %	8.19%	0	Sok2p	59.27 %	8.51%	0
Rpn4p	96.35 %	5.43%	0	Yrr1p	53.50 %	9.20%	0
Pdr1p	91.49 %	5.77%	0	Gis1p	24.62 %	13.78%	0
Pdr3p	97.26 %	5.34%	0	Mga2p	35.87 %	10.65%	0
Bas1p	79.64 %	8.74%	0	Rlm1p	27.66 %	12.87%	0
Fhl1p	51.37 %	8.68%	0	Met4p	28.27 %	14.40%	0
Hac1p	39.51 %	8.92%	0				

Figure 4

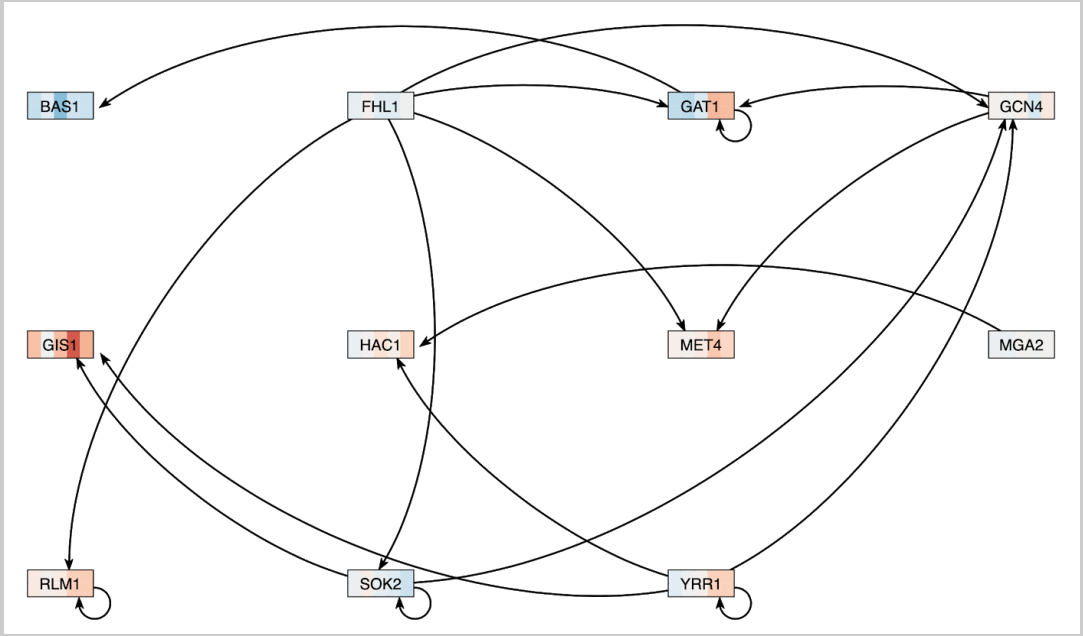


Figure 5

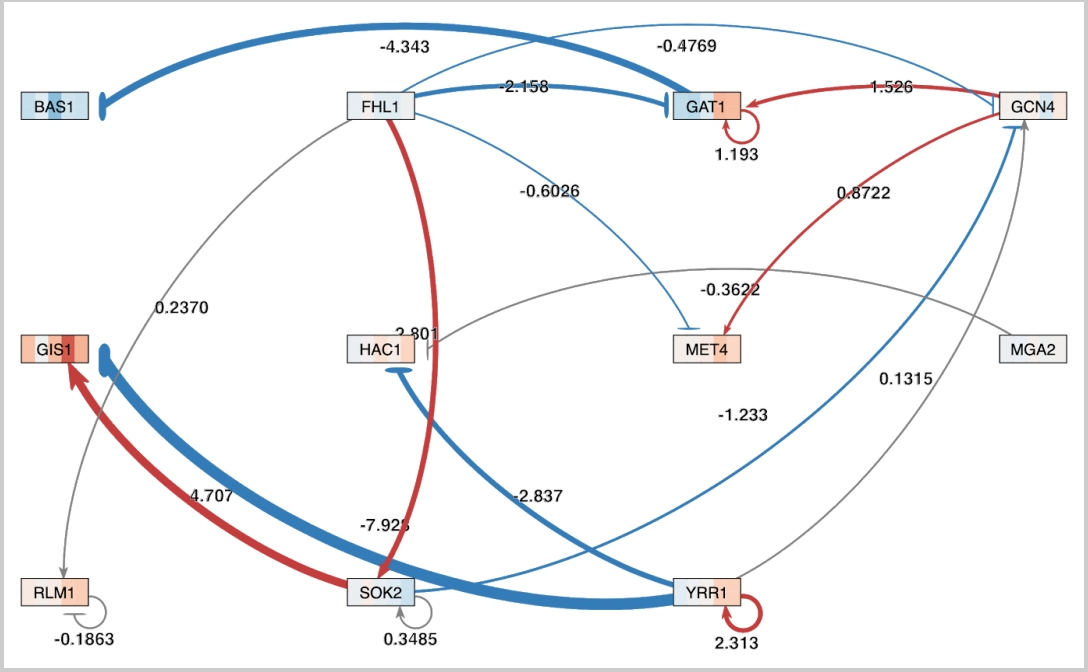


Figure 4,5: Regulatory network established the relationship between eleven genes using GRNmap.

Table 6. Weights for each relationship between transcription factor involved in the network.

rows gene affected/cols genes controlling	BAS1	FHL1	GAT1	GCN4	GIS1	HAC1	MET4	MGA2	RLM1	SOK2	YRR1
BAS1	0	0	-4.343	0	0	0	0	0	0	0	0
FHL1	0	0	0	0	0	0	0	0	0	0	0
GAT1	0	-2.158	1.193	1.526	0	0	0	0	0	0	0
GCN4	0	-0.477	0	0	0	0	0	0	0	-1.233	0.131
GIS1	0	0	0	0	0	0	0	0	0	4.707	-7.928
HAC1	0	0	0	0	0	0	0	-0.362	0	0	-2.837
MET4	0	-0.603	0	0.872	0	0	0	0	0	0	0
MGA2	0	0	0	0	0	0	0	0	0	0	0
RLM1	0	0.237	0	0	0	0	0	0	-0.186	0	0
SOK2	0	2.801	0	0	0	0	0	0	0	0.348	0
YRR1	0	0	0	0	0	0	0	0	0	0	2.313

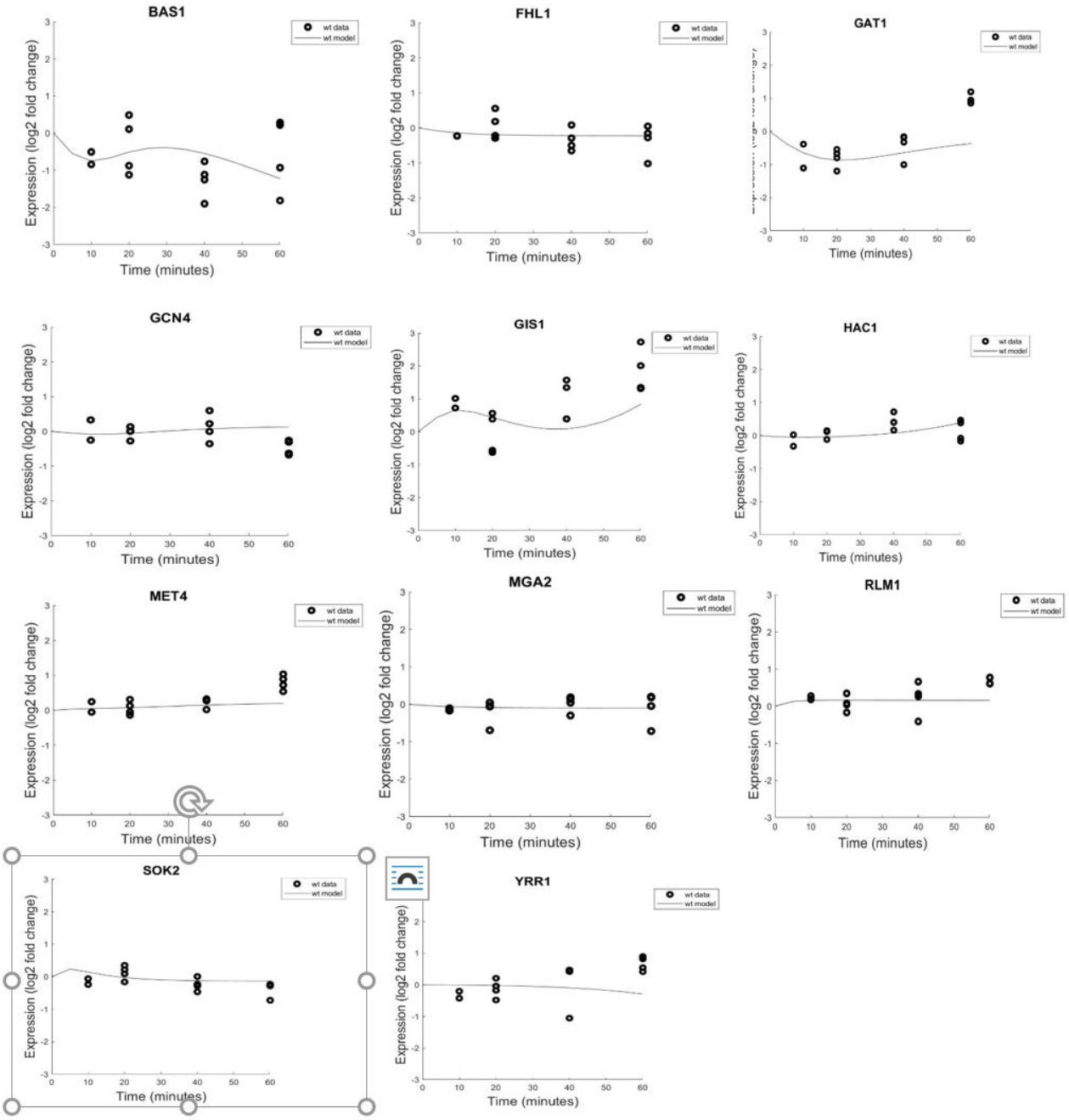


Figure 6. Gene plots produced for each transcription factor by MATLAB

Quality Assurance

Upon downloading of the Barreto et al. data, a sample-data relationship data was created as referred to in the flow chart (Table 1). The microarray data was regularized according to

treatment, time point, and replicate number. All strains were wildtypes, and each was grown in potassium-free media for either 10, 20, 40, 60, or 120 minutes. Each treatment had four replicates, except for the 10 minute time point, which only had two replicates. The format of the sample-data relationship data was used for the column headers of the microarray dataset. The first three column headers included the master index, ID, and standard name. The following columns listed the fold change data for each gene at every timepoint. These columns were standardized according to the following format

```
[yeaststrain]_[treatment]_LogFC_[time]-[replicate number]
```

For example K+ media treatment for ten minutes for replicate 1 had a column header of "BY4741_potassium_50mM_LogFC_t10m-1". Since there were four replicates for all time points except for the 10 minute time point, in which there are only two replicates, eighteen total column headers were formatted to organize the data according to time point and replicate number.

There were problems that occurred in formatting ID and Standard Name for the dataset as referred to in the flowchart as problem 1 (Table 1). Multiple genes contained duplicate data in which ID and Standard Names were changed to aliases. There presence of multiple genes in the data was not mentioned in the research article. Genes that were repeated under different aliases were removed in order to delete any repetition of data that could alter the results. Multiple genes were also also missing IDs or Standard Names. The respective names of the genes was manually inputted in order to complete the dataset using the "OFR List → Gene List" tool on YEASTRACT. Multiple genes were also named under a "SGD" naming format, and so were changed to parallel the ID and Standard Name format for the completion of the dataset. After the

formatting the dataset properly, all the yeast gene IDs were able to be imported into the database successfully.

Data Analysts

In order to identify which genes have a significant fold change in gene expression, an ANOVA test was conducted from the raw data of the Barreto, et al. study. The data was filtered to list only the genes with a p-value of less than 0.05 to only list genes with a significant change in expression. Given that 2,985 (71.9% of the total) genes had a significant change, it is likely that some of the genes' significance is by chance (Table 2). The Bonferroni p-value correction made an adjustment to the p-values to correct for multiple testing problems that might have occurred in the ANOVA test. Once the Bonferroni test was complete, the total amount of significant genes was lowered to 776 (18.7% of the total) genes (Table 2). This new number of significant genes is more accurate to the actual number of genes that have a significant expression change.

STEM was performed from the ANOVA results to cluster the genes according to their fold changes over each timepoint. STEM results suggested that profiles 39, 22, 8, 35, 16, 23, 25, 40 were significant (Fig. 2). A red color was designated to both profiles 39 and 22, indicating that both profiles belonged to the same gene cluster. Given that profile 39 was listed first in the results, suggesting that it was the most statistically significant cluster, and also had the simplest correlation over the timepoints, it was chosen for further analysis. The gene plot for Profile 39 indicates that most of the 329 genes belonging to this cluster increased in expression after potassium starvation (Fig. 30. Furthermore, the gene ontology results produced from the GO terms of Profile 39 categorized the genes according to their cellular functions within the yeast

cells. The most significant categories were genes involved in the biosynthesis of cysteine, methionine, other cellular amino acids, and hydrogen sulfide. The GO terms were also categorized with the oxidation-reduction processes and general molecular functioning of the cells (Table 3). The top fifteen most significant genes in Profile 39 were Gat1, Gcn4, Met4, Rpn4, Yrr1, Pdr1, Pdr3, Hsf1, Bas1, Hac1, Mga2, Rlm1, Sok2, Fhl, and Gis1 (Table 4).

Barreto, et al. classified the genes involved in the yeast cells response to potassium starvation as being involved in oxidative stress, methionine/cysteine synthesis, cyclin levels, septin rings, the retrograde pathway, methylglyoxal production, and trehalose metabolism. However, when compared to results of Barreto, et al., the GO results of this study only substantiate how potassium starvation may induce changes in the expression of genes related to the biosynthesis of methionine, cysteine synthesis, and other amino acids. Bas1, for instance, has been documented to be a regulator of the histidine pathway (Daignan-Pornier, et al., 1920), and recent studies have suggested that Gcn4 regulates the expression of various genes during amino acid starvation in yeast (Natarajan, et al. 2010). These findings may suggest that potassium starvation leads to a shortage of these amino acids within the yeast cells. Consequently, in response to this stressor, the cells may activate genes responsible for synthesizing these amino acids to compensate for the shortage. Furthermore, Rlm1 has been found to be involved in the regulation of the cell's transition from G1 to the S phase of the cell cycle (Piccirillo, et al., 2017). The regulation of Rlm1 as observed in the network suggests that the yeast cells may be regulating their growth during potassium starvation as well. Given that during the S phase, the cells are replicating their DNA, regulating the transition into the S phase may serve to conserve energy during the potassium starvation. Nonetheless, it should be noted that given that this study

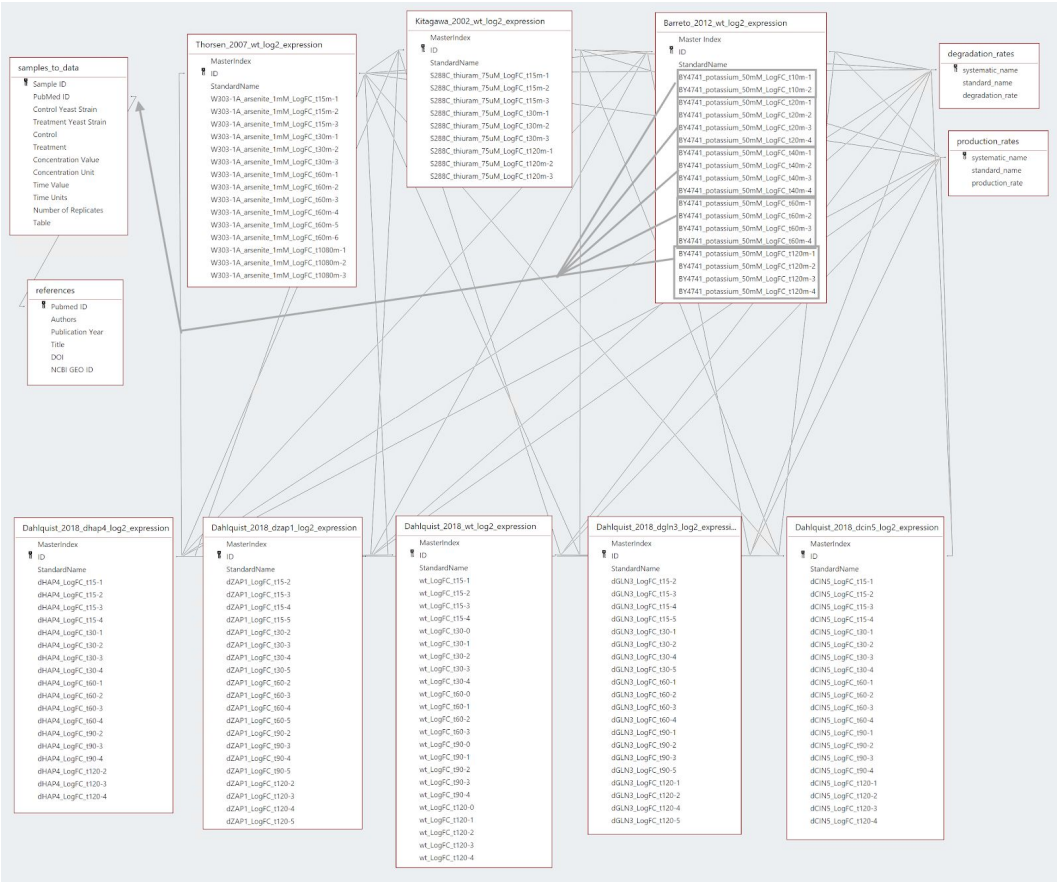
only highlights the genes involved in Profile 39 of the STEM results, it is likely that the genes involved in the other metabolic pathways mentioned by Barreto, et al. may have been clustered in the other gene profiles produced by STEM.

Before creating the regulatory network with GRNsight, genes Rpn4, Pdr1, Pdr3, and Hsf1 were removed from the study because fold change data was not collected on these transcription factors by Barreto, et al. A GRNmap input file was created using a database compiled from the Barreto, et al. data. The input worksheet was loaded into MATLAB to produce an output file that would be later loaded into GRNsight for the weighted network. The LSE:minLSE ratio after running MATLAB was 1.176173. The gene regulatory network established a relationship between the eleven transcription factors involved in the study (Fig. 4, 5). Gat1 and Gcn4 were the most regulated genes while Fhl1 regulated the most number of genes. The strongest relationships were observed in the repression of Gis1 by Yrr1 and the induction of Gis1 by Sok2. A relatively strong relationship was also observed in the repression of Bas1 by Gat1. Transcription factors Gat1 and Yrr1 were also found to be self inducing and thus, response for the induction of their own genes. Recent studies have also suggested that a relationship does exist between the genes observed in the compiled regulatory network. Genes Sok2 and Gis1, for example, have been documented to be involved in the regulation of TPK1 in yeast (Pautasso, Rossi, 2014), and similarly, genes Gat1 and Fhl1 have also been found to be involved in similar metabolic pathways (Bandhakavi, 2008). Other studies have also found a linkage between Hac1 and Mga2 (Covina, et al. 2018).

Coder

The physical database containing standardized log-fold change gene expression data presented a number of challenges, as it required the standardization of naming of a number of tables and fields as well as the maintenance of different branches of work for each of the groups working on the project. It was created in Microsoft Access, a database management software. The database was stored in the ".accdb" file format, a container for the Jet Database Engine. Given that the classrooms hosting a central database were not always available, the server was hosted on a Windows 10 virtual machine set up using the Microsoft Azure cloud computing service. The database contained 12 tables - 2 metadata tables, 8 tables containing gene expression data, and 2 tables containing mRNA production and degradation rate information for each gene.

Figure 7: A diagram of all the tables, field names, and linkages between tables and fields created in Microsoft Access.



In order to catalog metadata regarding the gene expression data being stored in the database, two tables were created. The first, the "references" table, contained metadata regarding the article or data source from which all the data came from. Fields include authors, article title, NCBI GEO number, and more. This was linked to the second table, "samples_to_data", via the PubMed ID field contained in both. The "samples_to_data" table contained information regarding the experiment from which the data sample was collected. Data samples were classified with the same unique identifier based on the time point of the treatment. Thus, multiple replicates of a certain time point would be given the same unique Sample ID identifier. Although unable to be able to link the entries in these fields to a given set of fields in a table in Microsoft Access, information for other accessors regarding the table in which the information was contained was encoded in a separate field.

Inside the data tables, a number of changes were made regarding linkages and standardization of table and attribute (column) naming. For all the data tables, including the production and degradation rate tables, the Systematic Name of each table as the primary key, essentially stating that each collection of columns forming a row are a tuple, or collection of related data. Between the primary keys linkages of data from the same gene across multiple datasets were established by linking the primary keys (Fig. 7). Although not possible in Microsoft Access, a PostgreSQL database would be able to link attributes of data to a particular field in a table. This is annotated by the particularly heavy gray lines in Figure 7. Additionally, the naming of each table was standardized across all tables, save for the two accessory data and metadata tables. Table names were formatted to follow the format "[First Author Last Name]_[Year]_[strain (wild-type or mutant)]_log2_expression." For example the table for the

gene expression data for the 2012 Baretto et al. paper would be called "Barreto_2012_wt_log2_expression". The attribute names inside of each of these tables were standardized as well; each of these tables contained attributes for the master index (a number), systematic name of yeast gene, standard name, followed by the gene data, named by the following paradigm: "[Yeast Strain]_LogFC_t[Time Point]-[Replicate Number]". The data for data in a table collected from a *zap1* deletion strain yeast in the 3rd replicate of the 20th time point of an experiment may therefore be named "dzap1_LogFC_t20-3".

The final process required merging the different branches of the database created by each group supervising the correct input in the database. Once all groups had uploaded their individual branch of the database to the shared Azure virtual machine, the database was merged. This was done via the import external data function in Access, where naming conflicts were resolved and redundant tables deleted. Finally, each team reviewed the compiled database, and especially their contributions. Skinny Genes, this group, detected an error in the previous version of our table being used, so the correct version was updated and uploaded to the shared virtual machine, as well as the class website.

Conclusion

The aim of this project was to standardize and re-analyze the gene expression data sets of a number of publicly accessible articles regarding environmental stress in *Saccharomyces cerevisae*. This group analyzed the data set of an article by Barreto et al., published in 2012, which conducted an experiment testing the response of yeast to potassium starvation.

Over the course of this project, we utilized the database that we created using the gene expression data. This database allowed the various groups in this project to have centralized

access to a number of curated and easily accessible gene expression data sets. If a group member desired particular information about the data we were analyzing, the group member simply needed to create a simple query. Furthermore, hosting it on a cloud machine increased the accessibility, as anyone with the login credentials could access the database at any time.

The findings of this study substantiate the results of Barreto, et al. since the deprivation of potassium in the yeast cells seemed to trigger changes in gene expression. While the specific transcription factors highlighted in this project were not mentioned by Barreto, et al., the results regarding the biosynthetic pathways involved in potassium starvation parallel those found by Barreto and her colleagues. Both datasets suggest that intracellular potassium concentrations are linked to the changes in the expression of genes associated with amino acid biosynthesis. Thus, it is likely that the deprivation of potassium leads to a shortage of amino acids within the cell, especially methionine and cysteine, leading to an upregulation of the biosynthesis of these amino acids. Furthermore, though the gene classifications of Barreto, et al., only parallel one of the GO classifications found in the study, this project only focused on one of the eight significant gene profiles for analysis. Thus, it is likely that while this project substantiates how potassium starvation may be linked to amino acid biosynthesis, the correlation of genes associated with the other metabolic and biosynthetic pathways observed by Barreto, et al., may have also been observed had other gene profiles been studied from the STEM results.

Upon re-analyzing the data, we decided that if the same experiment were to be performed in the future, we would alter the experiment in a number of ways. Firstly, we would actually measure the expression data for the 4 missing genes. Additionally, we would perform experiments that would describe how the 12 significant genes that we discovered affect other

cations and the effects on the cell, the production and degradation rates of other transcription factors, and cellular function in relation to homeostasis and membrane potential.

Acknowledgments

We would like to thank Dr. Dahlquist, Loyola Marymount University Biology

Department, and the BIOL 367-01 class for their contribution and collaboration on this project.

References

Adrian, R. H. (1956). The effect of internal and external potassium concentration on the membrane potential of frog muscle. *The Journal of Physiology*, *133*(3), 631-658.

Bandhakavi, S., Xie, H., O'Callaghan, B., Sakurai, H., Kim, D. H., & Griffin, T. J. (2008). Hsf1 activation inhibits rapamycin resistance and TOR signaling in yeast revealed by combined proteomic and genetic analysis. PloS one, 3(2), e1598.

Barreto, L., Canadell, D., Valverde‐Saubí, D., Casamayor, A., & Ariño, J. (2012). The short‐term response of yeast to potassium starvation. *Environmental microbiology*, *14*(11), 3026-3042.

Covino, R., Hummer, G., & Ernst, R. (2018). Integrated functions of membrane property sensors and a hidden side of the unfolded protein response. Molecular cell, 71(3), 458-467.

Daignan-Fornier, B., & Fink, G. R. (1992). Coregulation of purine and histidine biosynthesis by the transcriptional activators BAS1 and BAS2. Proceedings of the National Academy of Sciences, 89(15), 6746-6750.

Hess, D. C., Lu, W., Rabinowitz, J. D., & Botstein, D. (2006). Ammonium toxicity and potassium limitation in yeast. PLoS biology, 4(11), e351.

Udensi, U. K., & Tchounwou, P. B. (2017). Potassium Homeostasis, Oxidative Stress, and Human Disease. International journal of clinical and experimental physiology, 4(3), 111–122. doi:10.4103/ijcep.ijcep_43_17

Martínez, J. L., Luna, C., & Ramos, J. (2012). Proteomic changes in response to potassium starvation in the extremophilic yeast Debaryomyces hansenii. FEMS yeast research, 12(6), 651-661. doi: 10.1111/j.1567-1364.2012.00815.x

Natarajan, K., Meyer, M. R., Jackson, B. M., Slade, D., Roberts, C., Hinnebusch, A. G., & Marton, M. J. (2001). Transcriptional profiling shows that Gcn4p is a master regulator of gene expression during amino acid starvation in yeast. Molecular and cellular biology, 21(13), 4347-4368.

Pautasso, C., & Rossi, S. (2014). Transcriptional regulation of the protein kinase A subunits in Saccharomyces cerevisiae: Autoregulatory role of the kinase A activity. Biochimica Et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 1839(4), 275-287.

Piccirillo, S., Neog, D., Spade, D., Van Horn, J. D., Tiede-Lewis, L. M., Dallas, S. L., … Honigberg, S. M. (2017). Shrinking Daughters: Rlm1-Dependent G1/S Checkpoint Maintains Saccharomyces cerevisiae Daughter Cell Size and Viability. Genetics, 206(4), 1923–1938. doi:10.1534/genetics.117.204206