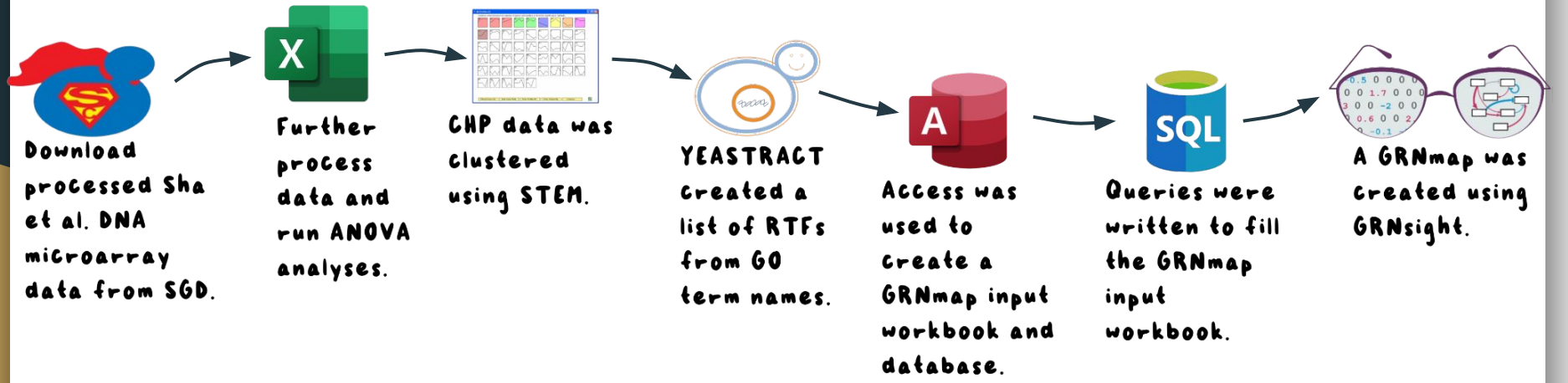
Yeast Beasts

Biol 367
Dr. Dahlquist
Hailey Ivanson, Dean Symonds, Charlotte Kaplan,
Katie Miller, Natalija Stojanovic, Andrew Sandler

Outline

- Data from Sha et al. was downloaded and prepared.
- Cluster profiles were visualized to determine the profile with the most genes which was found to be profile 41.
- A Microsoft Access database was created using the above data and gene regulatory network data.
- A GRNmap network was created for profile 41 based on queries conducted on the database.
- Our findings agreed and disagreed with Sha et al. findings.



Flow chart of data processing and technical workflow of the project. Images from sources cited.

Outline

- **Data from Sha et al. was downloaded and prepared.**
- Cluster profiles were visualized to determine the profile with the most genes which was found to be profile 41.
- A Microsoft Access database was created using the above data and gene regulatory network data.
- A GRNmap network was created for profile 41 based on queries conducted on the database.
- Our findings agreed and disagreed with Sha et al. findings.

The data analysts used microarray data from a 2013 study on cumene hydroperoxide exposure in yeast.

- The microarray dataset was uploaded to the GEO Database as GSE26169.
- The researchers only analyzed data points from times 0-20 minutes.
    - But they uploaded all 120 minutes that they collected.
    - We are using all 120 minutes of their data.

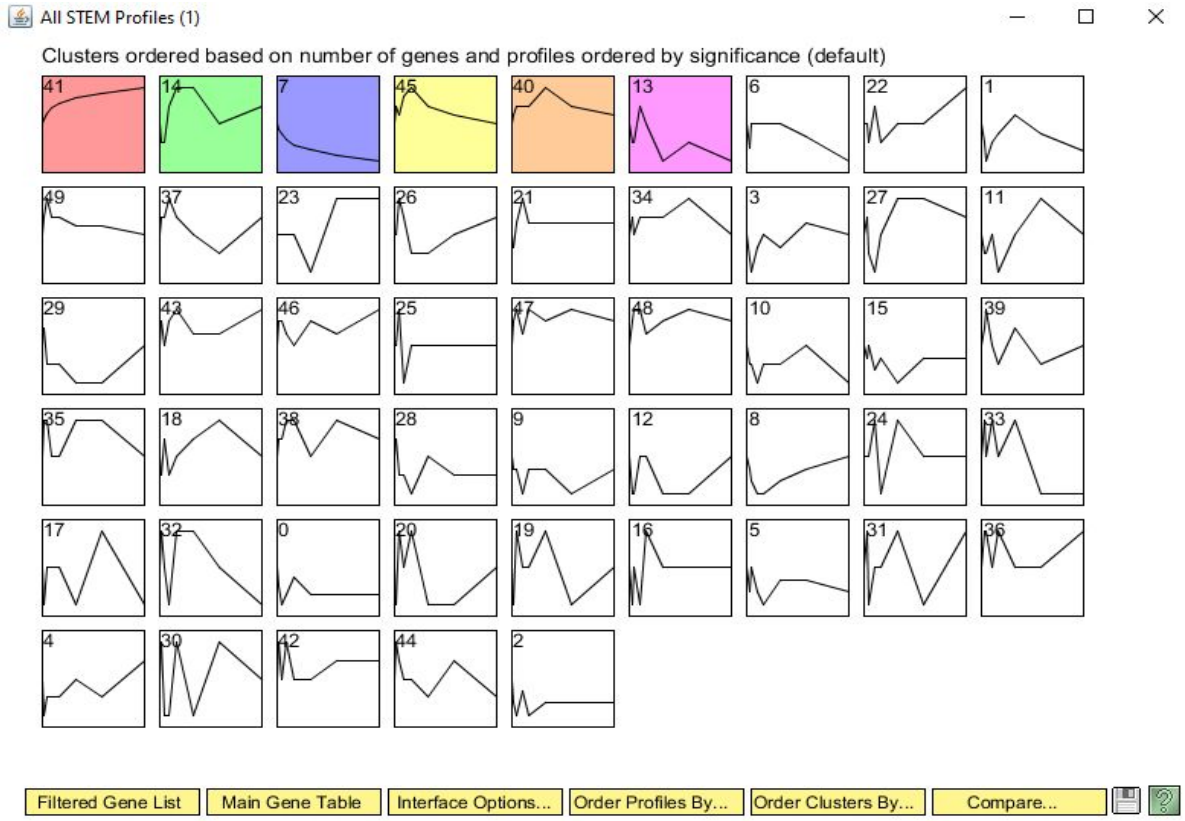
Selecting p<0.00001 for B&H ANOVA yielded ~25% significance for CHP and control datasets.

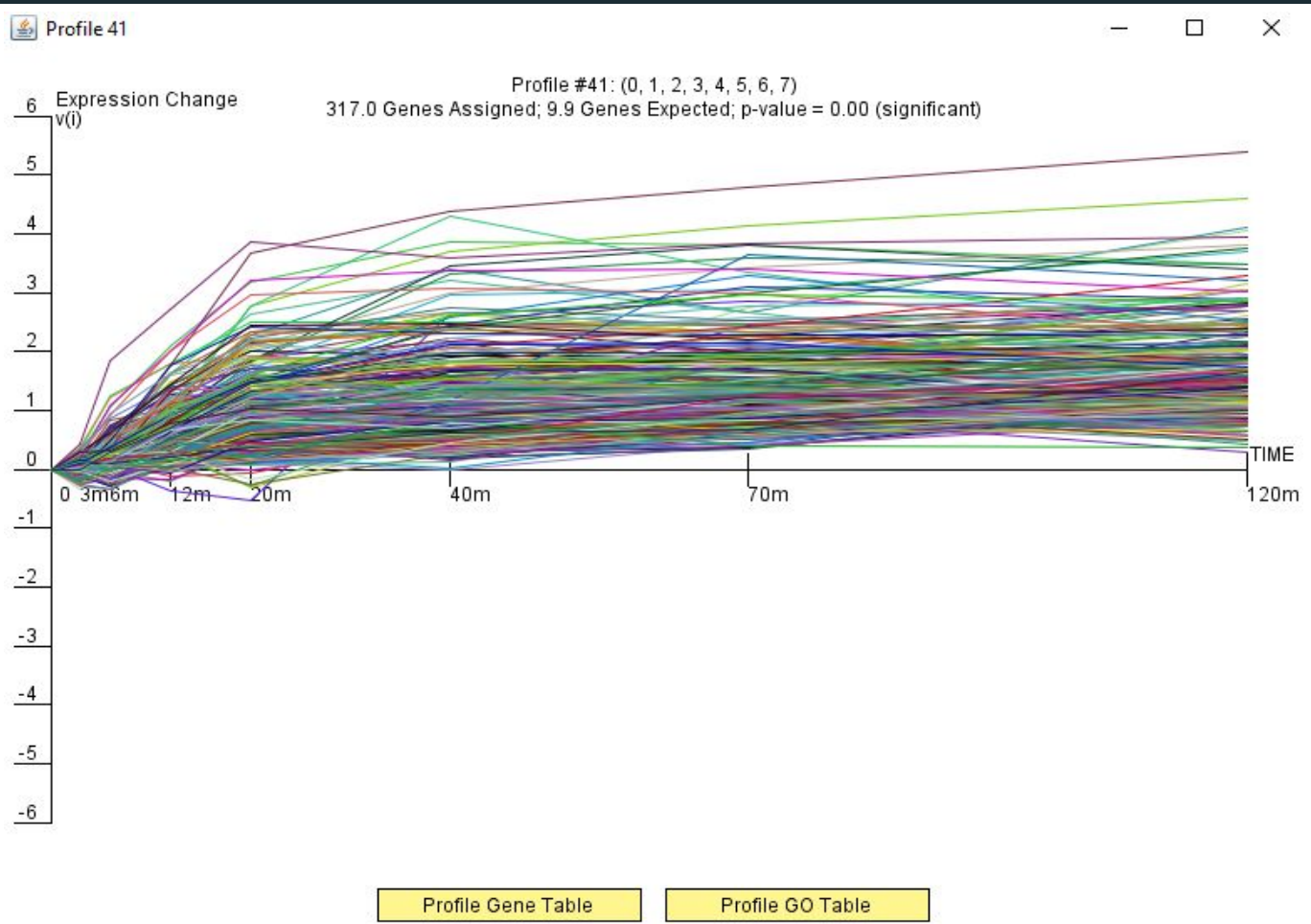
B&H ANOVA	Control Genes	CHP Genes
p<0.05	3699 (78.7%)	2856 (60.8%)
p<0.01	3219 (68.6%)	2390 (50.9%)
p<0.001	2558 (54.4%)	1857 (39.5%)
p<0.0001	1921 (40.9%)	1419 (30.2%)
p<0.00001	1325 (28.2%)	1058 (22.5%)

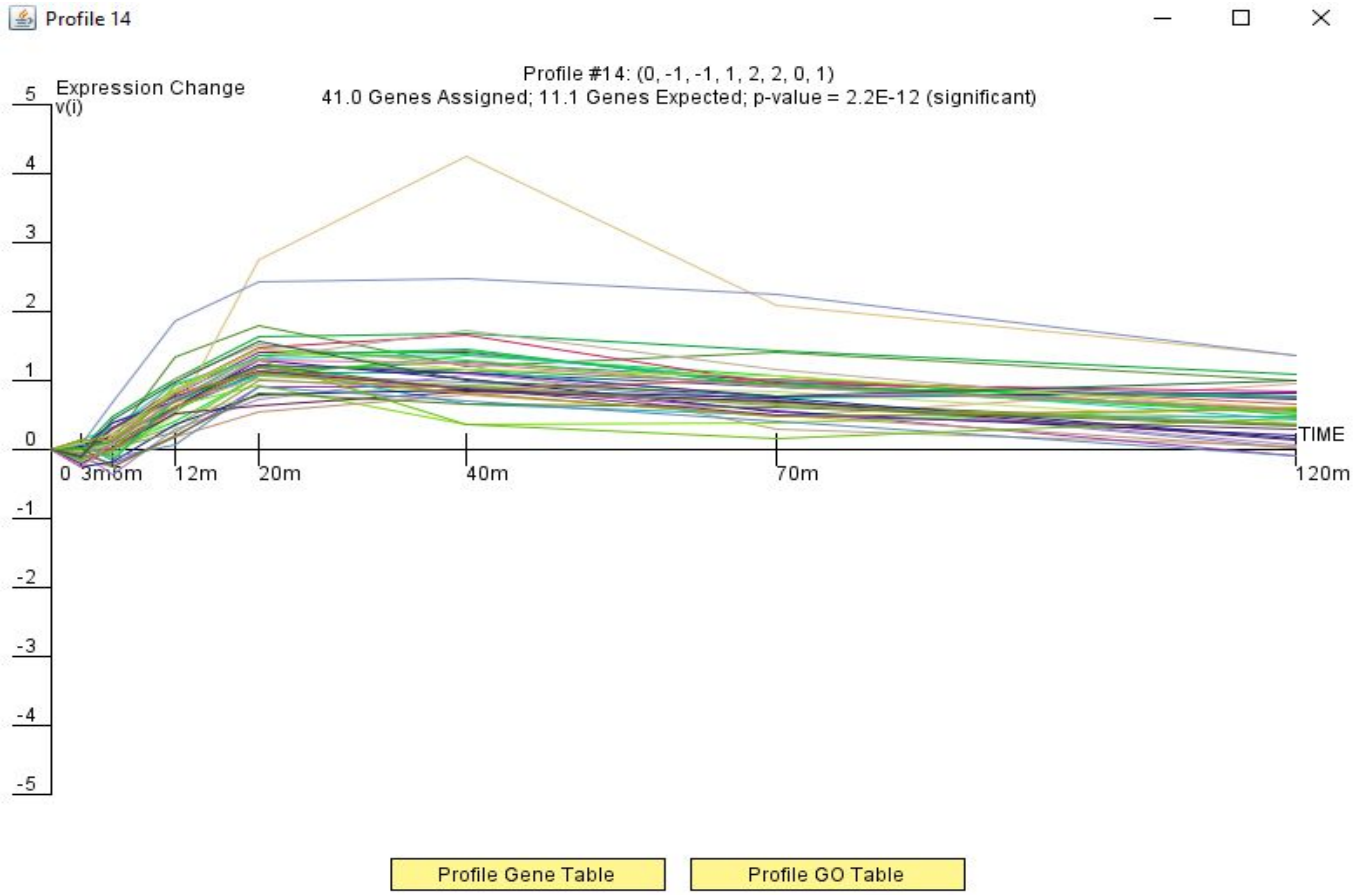
Outline

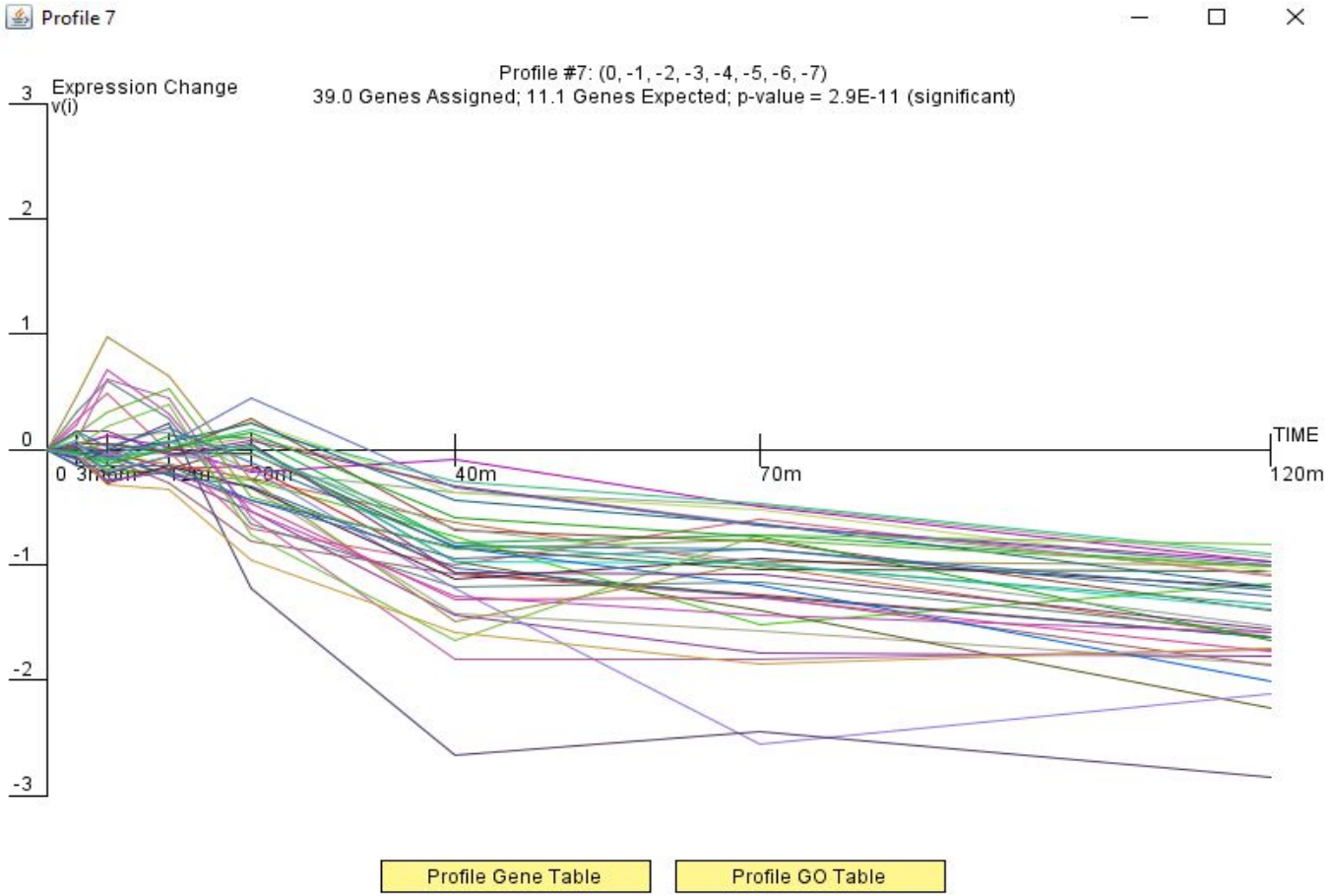
- Data from Sha et al. was downloaded and prepared.
- **Cluster profiles were visualized to determine the profile with the most genes which was found to be profile 41.**
- A Microsoft Access database was created using the above data and gene regulatory network data.
- A GRNmap network was created for profile 41 based on queries conducted on the database.
- Our findings agreed and disagreed with Sha et al. findings.

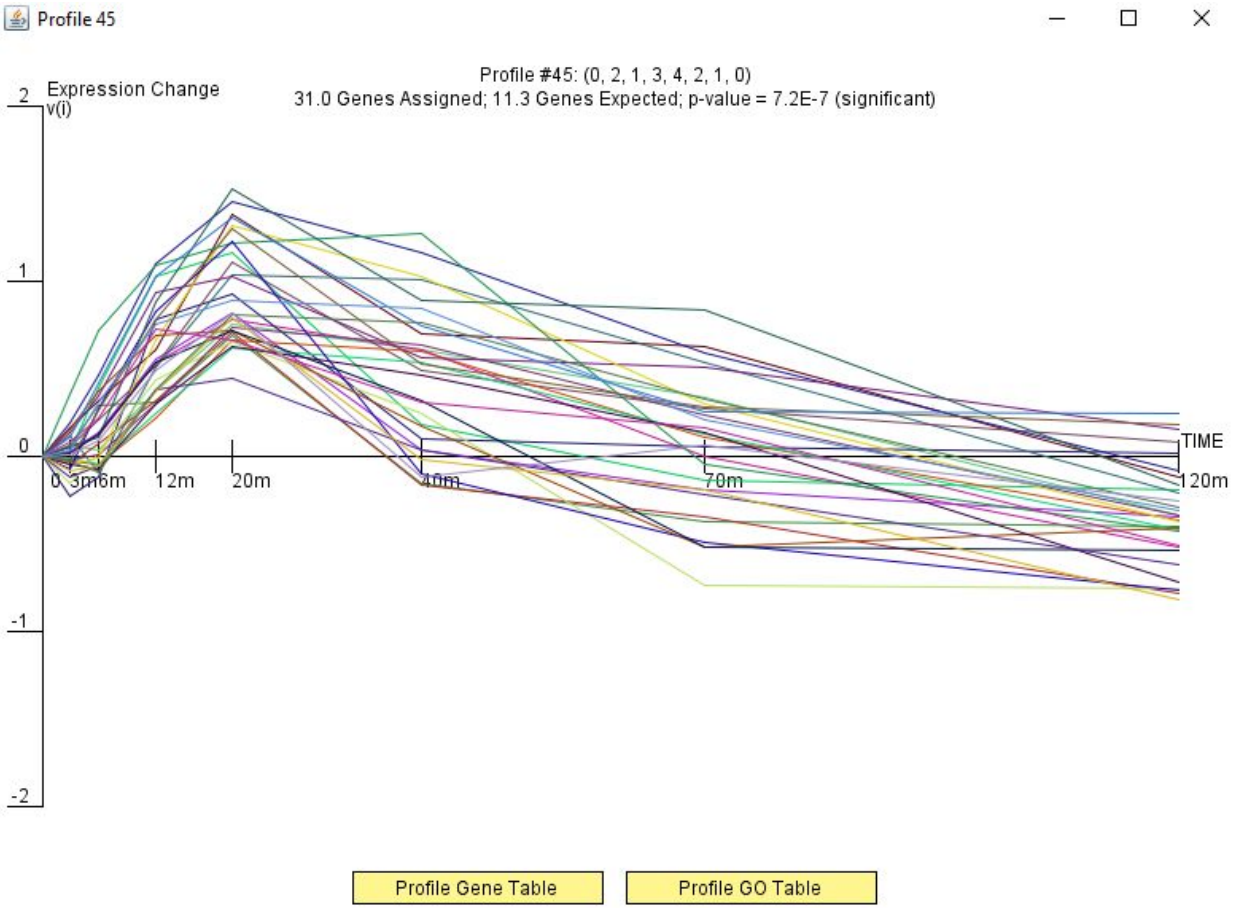
Stem clustering was run on only the CHP data.

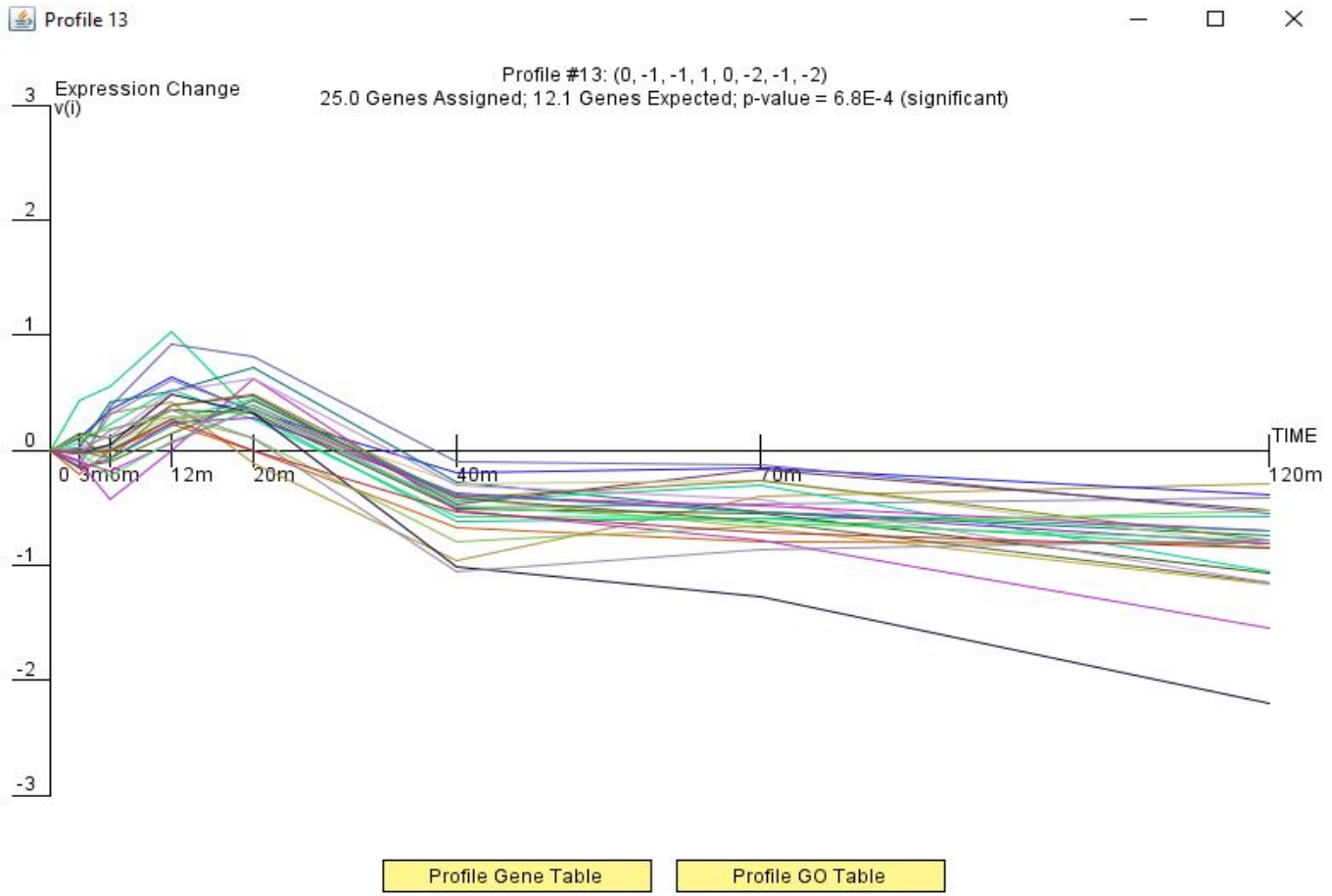



Profile 41
Profile #41: (0, 1, 2, 3, 4, 5, 6, 7)
317.0 Genes Assigned; 9.9 Genes Expected; p-value = 0.00 (significant)
Expression Change
v(i)
6
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
0
3m
6m
12m
20m
40m
70m
120m
TIME
Profile Gene Table
Profile GO Table


Profile 14
Profile #14: (0, -1, -1, 1, 2, 2, 0, 1)
41.0 Genes Assigned; 11.1 Genes Expected; p-value = 2.2E-12 (significant)
Expression Change
v(i)
5
4
3
2
1
0
-1
-2
-3
-4
-5
0
3m
6m
12m
20m
40m
70m
120m
TIME
Profile Gene Table
Profile GO Table

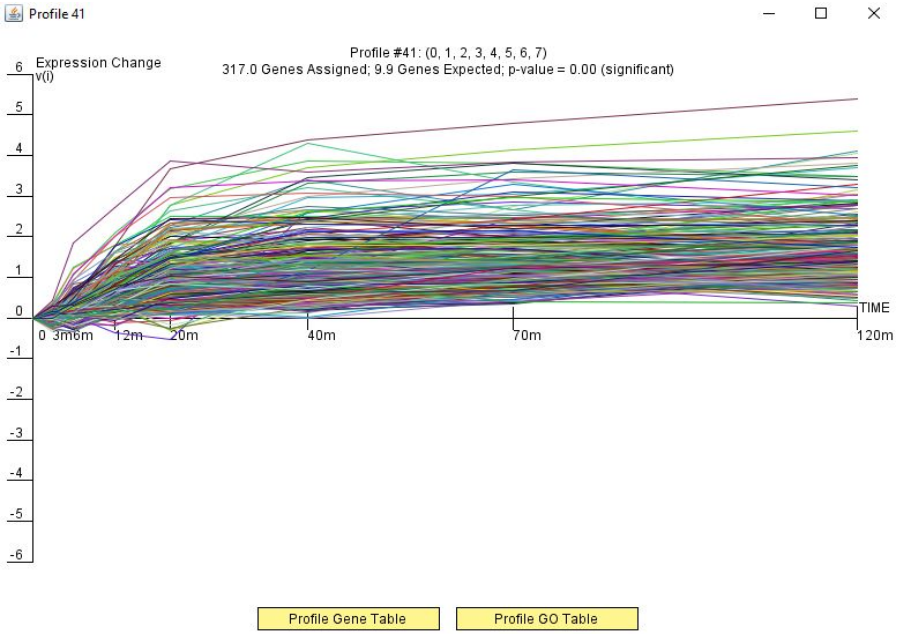

Profile 7
Profile #7: (0, -1, -2, -3, -4, -5, -6, -7)
39.0 Genes Assigned; 11.1 Genes Expected; p-value = 2.9E-11 (significant)
Expression Change
v(i)
3
2
1
0
-1
-2
-3
TIME
0
40m
70m
120m
Profile Gene Table
Profile GO Table


Profile 45
Profile #45: (0, 2, 1, 3, 4, 2, 1, 0)
31.0 Genes Assigned; 11.3 Genes Expected; p-value = 7.2E-7 (significant)
Expression Change
v(i)
2
1
0
-1
-2
0
3m
6m
12m
20m
40m
70m
120m
TIME
Profile Gene Table
Profile GO Table


Profile 13
Profile #13: (0, -1, -1, 1, 0, -2, -1, -2)
25.0 Genes Assigned; 12.1 Genes Expected; p-value = 6.8E-4 (significant)
Expression Change
v(i)
3
2
1
0
-1
-2
-3
TIME
0
3m
6m
12m
20m
40m
70m
120m
Profile Gene Table
Profile GO Table

Profile 41 was selected because it was assigned the most genes and the highest significance.

- Unlike the Sha et al. paper, we used all timepoints 0-120 mins.
  - Different patterns would be observed if only minutes 0-20 were focused on.
  - More on this later.



The Most Significant GO Results are for Cell Wall Organization

GO Biological Process	Place in Saccharomyces cerevisiae Ref. List	FDR Corrected p-value
glutamate catabolic process (GO:0006538)	3	1.38E-02
protein localization to endoplasmic reticulum exit site (GO:0070973)	4	3.93E-02
regulation of fungal-type cell wall organization (GO:0060237)	20	4.82E-04
regulation of cell wall organization or biogenesis (GO:1903338)	21	5.28E-04
nucleotide transport (GO:0006862)	39	3.25E-02

Even at all time points, cell wall transcriptional changes remain significant.

- The 2013 paper focused on the fact that CHP damages peripheral cell structures like the cell wall due to its large size.

GO Results Show Genes Belonging to Sorting Pathways

GO Biological Process	Place in Saccharomyces cerevisiae Ref. List	FDR Corrected p-value
cellular lipid catabolic process (GO:0044242)	50	1.01E-02
multivesicular body sorting pathway (GO:0071985)	58	7.52E-03
late endosome to vacuole transport via multivesicular body sorting pathway (GO:0032511)	51	4.29E-02
endosome transport via multivesicular body sorting pathway (GO:0032509)	57	2.50E-02
small molecule catabolic process (GO:0044282)	141	2.43E-02

Using Yeastract, we made a list of candidate regulatory transcription factors.

- We used the 23 most significant transcription factors in profile 41 to move on with.
- We recognized some from the Sha et al. paper.
  - More on this later.

Outline

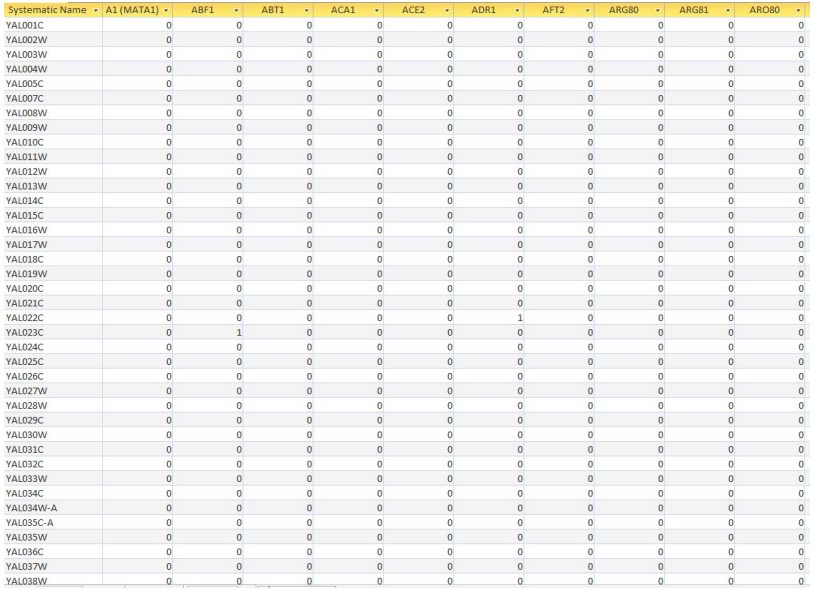
- Data from Sha et al. was downloaded and prepared.
- Cluster profiles were visualized to determine the profile with the most genes which was found to be profile 41.
- **A Microsoft Access database was created using the above data and gene regulatory network data.**
- A GRNmap network was created for profile 41 based on queries conducted on the database.
- Our findings agreed and disagreed with Sha et al. findings.

We developed a database to model GRNs.

- This allowed us to:
  - Learn new things about yeast genes under oxidative stress.
  - Check the quality of existing publications.
    - Check the quality of our own new database.

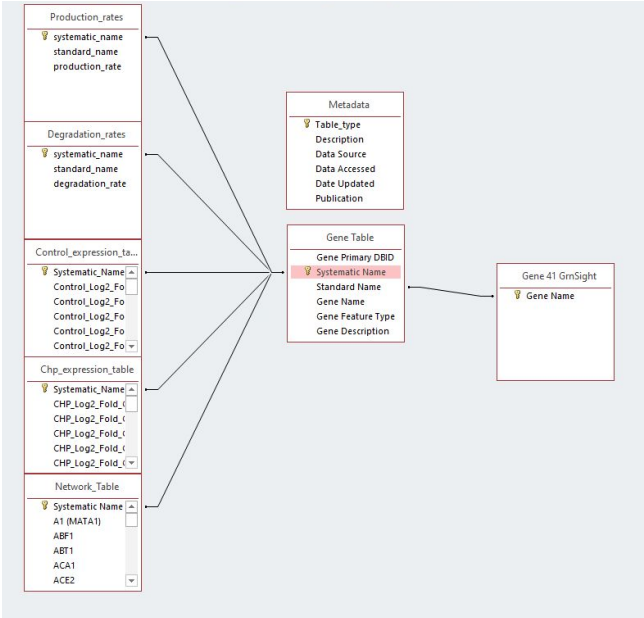
The coder/designers used data from a 2004 study that predicted DNA-binding transcriptional regulators.

- Data obtained from the paper was used to determine if there was a significant binding result between the genes and regulators.



Systematic Name	A1 (MATA1)	ABF1	ABT1	ACA1	ACE2	ADR1	AFT2	ARG80	ARG81	ARO80
YAL001C	0	0	0	0	0	0	0	0	0	0
YAL002W	0	0	0	0	0	0	0	0	0	0
YAL003W	0	0	0	0	0	0	0	0	0	0
YAL004W	0	0	0	0	0	0	0	0	0	0
YAL005C	0	0	0	0	0	0	0	0	0	0
YAL007C	0	0	0	0	0	0	0	0	0	0
YAL008W	0	0	0	0	0	0	0	0	0	0
YAL009W	0	0	0	0	0	0	0	0	0	0
YAL010C	0	0	0	0	0	0	0	0	0	0
YAL011W	0	0	0	0	0	0	0	0	0	0
YAL012W	0	0	0	0	0	0	0	0	0	0
YAL013W	0	0	0	0	0	0	0	0	0	0
YAL014C	0	0	0	0	0	0	0	0	0	0
YAL015C	0	0	0	0	0	0	0	0	0	0
YAL016W	0	0	0	0	0	0	0	0	0	0
YAL017W	0	0	0	0	0	0	0	0	0	0
YAL018C	0	0	0	0	0	0	0	0	0	0
YAL019W	0	0	0	0	0	0	0	0	0	0
YAL020C	0	0	0	0	0	0	0	0	0	0
YAL021C	0	0	0	0	0	0	0	0	0	0
YAL022C	0	0	0	0	0	1	0	0	0	0
YAL023C	0	1	0	0	0	0	0	0	0	0
YAL024C	0	0	0	0	0	0	0	0	0	0
YAL025C	0	0	0	0	0	0	0	0	0	0
YAL026C	0	0	0	0	0	0	0	0	0	0
YAL027W	0	0	0	0	0	0	0	0	0	0
YAL028W	0	0	0	0	0	0	0	0	0	0
YAL029C	0	0	0	0	0	0	0	0	0	0
YAL030W	0	0	0	0	0	0	0	0	0	0
YAL031C	0	0	0	0	0	0	0	0	0	0
YAL032C	0	0	0	0	0	0	0	0	0	0
YAL033W	0	0	0	0	0	0	0	0	0	0
YAL034C	0	0	0	0	0	0	0	0	0	0
YAL034W-A	0	0	0	0	0	0	0	0	0	0
YAL035C-A	0	0	0	0	0	0	0	0	0	0
YAL035W	0	0	0	0	0	0	0	0	0	0
YAL036C	0	0	0	0	0	0	0	0	0	0
YAL037W	0	0	0	0	0	0	0	0	0	0
YAL038W	0	0	0	0	0	0	0	0	0	0

An Access database was created with 6 tables to assist in data analysis.

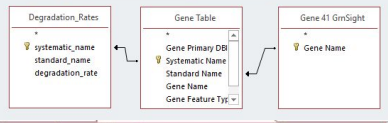


- Degradation rates table contains the degradation rates from Neyomotin et al. (2014).
- Production rates contains the initial guesses for the production rates of each gene.
- 2 expression tables to assess the how the genes were expressed when they were treated with CHP.
- Network table using data obtained from Harbison et al. (2004).

Outline

- Data from Sha et al. was downloaded and prepared.
- Cluster profiles were visualized to determine the profile with the most genes which was found to be profile 41.
- A Microsoft Access database was created using the above data and gene regulatory network data.
- **A GRNmap network was created for profile 41 based on queries conducted on the database.**
- Our findings agreed and disagreed with Sha et al. findings.

The Queries from the database layed a foundation for the GRNmap network.



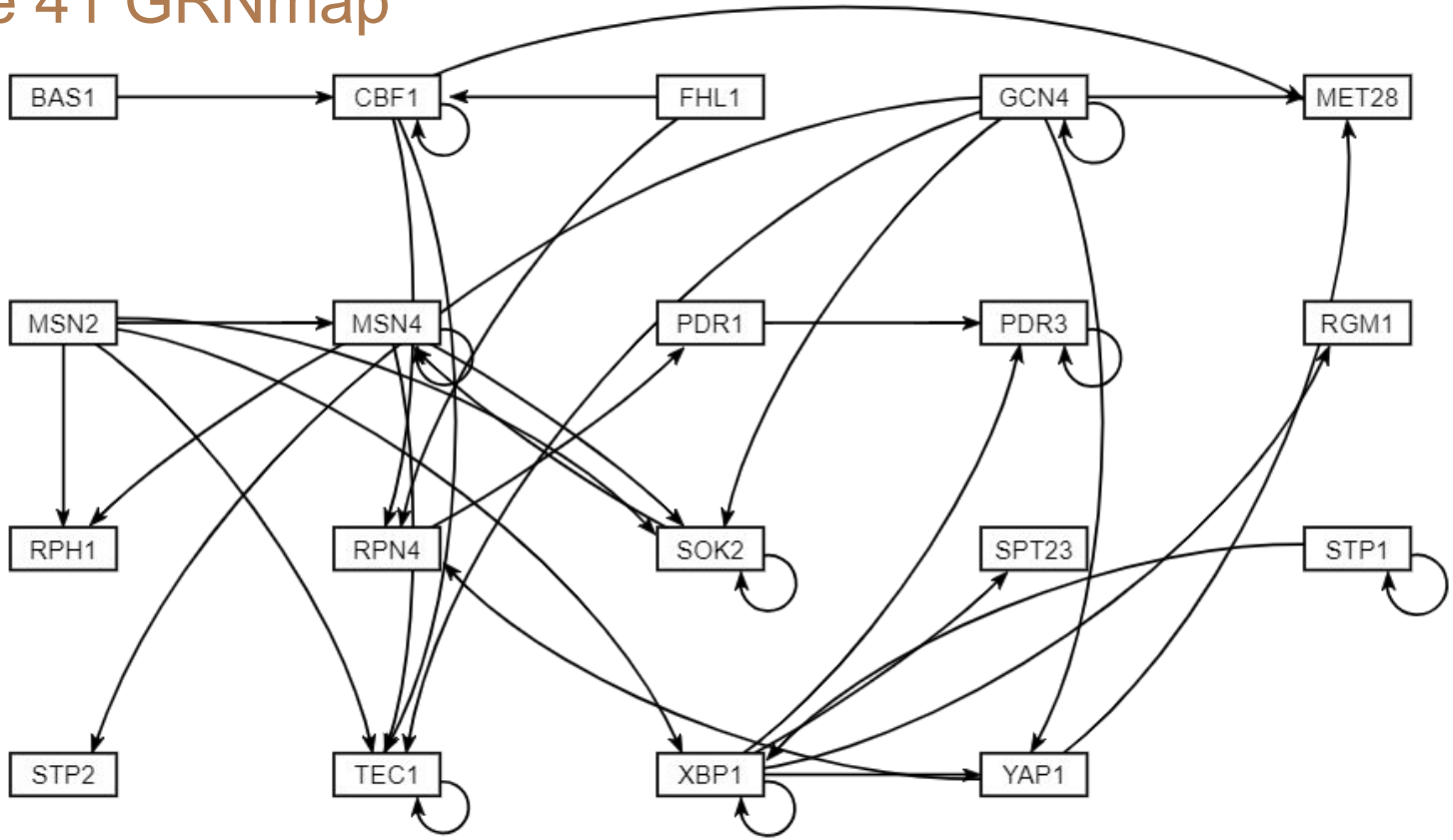
systematic_name	standard_name	degradation_rate
YKR099W	BAS1	0.1066
YJR060W	CBF1	0.0835
YNL027W	CRZ1	0.1238
YPR104C	FHL1	0.105
YEL009C	GCN4	0.0513
YIR017C	MET28	0.0542
YPL038W	MET31	0.055
YMR037C	MSN2	0.2039
YKL062W	MSN4	0.1386
YGL013C	PDR1	0.1083
YBL005W	PDR3	0.1359
YMR182C	RGM1	0.2666
YPL089C	RLM1	0.2236
YER169W	RPH1	0.105
YDL020C	RPN4	0.1136
YMR016C	SOK2	0.4332
YKL020C	SPT23	0.1284
YDR463W	STP1	0.0845
YHR006W	STP2	0.0912
YBR083W	TEC1	0.2773
YIL101C	XBP1	0.0912
YML007W	YAP1	0.0835

- Using queries in design view, the tables were connected based on their primary keys.
- The queries revealed the production rates, degradation rates, and the networks for the different genes.
- These query results allowed for a GRN map to be created.
- This GRNmap should allow a network of profile 41 to be created.

Instead of using the Harbison et al. data, 2024 data and GRNsight were used to create the GRNmap.

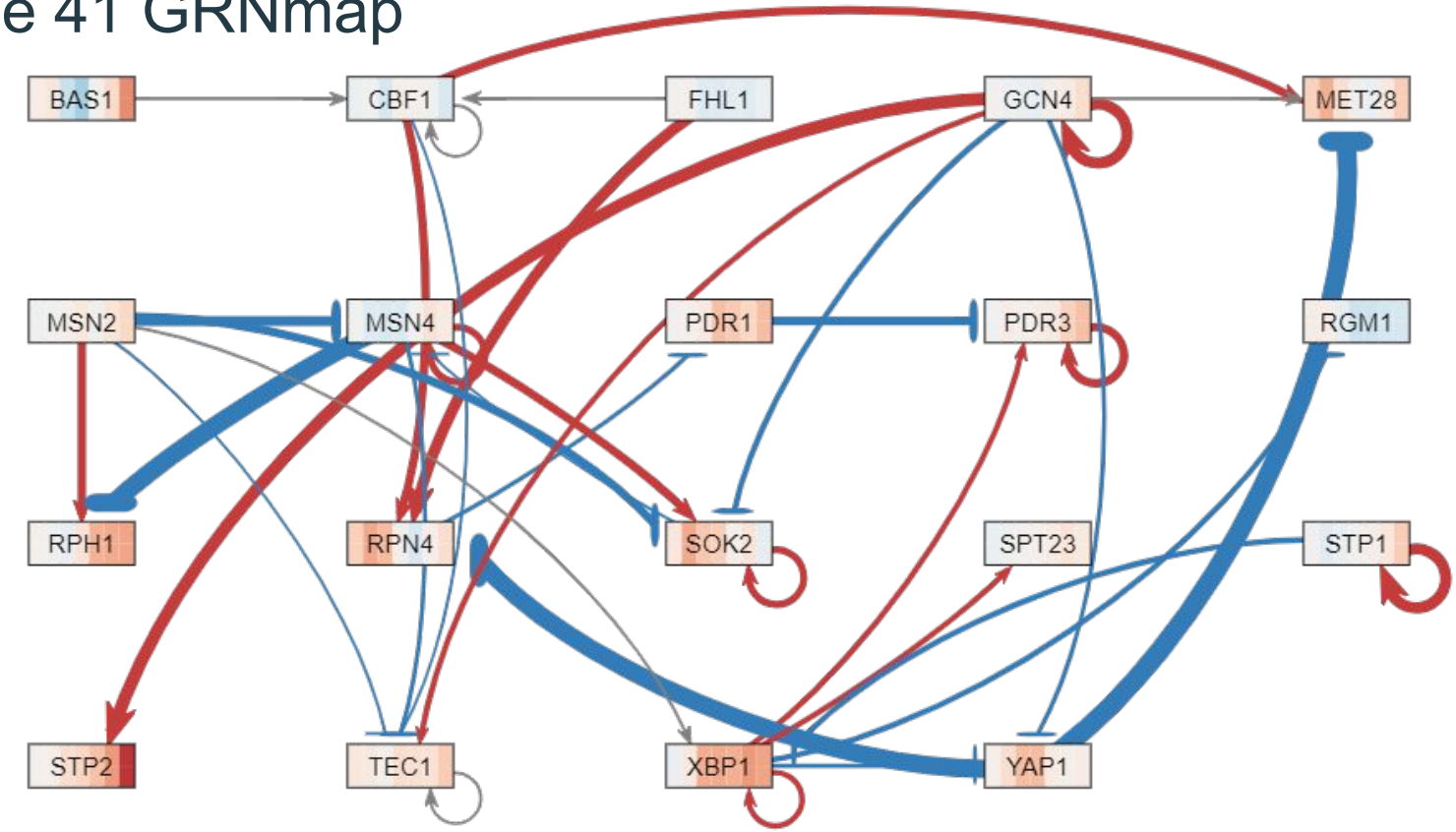
- There were not enough edges to construct a GRNmap model using the Harbison data.
- 2024-03-19 version of the Saccharomyces Genome Database regulation data from YeastMine was used instead.
- CRZ1, MET31, MGA2, and RLM1 were removed as they lacked connections to other transcription factors.

Profile 41 GRNmap



https://dondi.github.io/GRNsight/

Profile 41 GRNmap



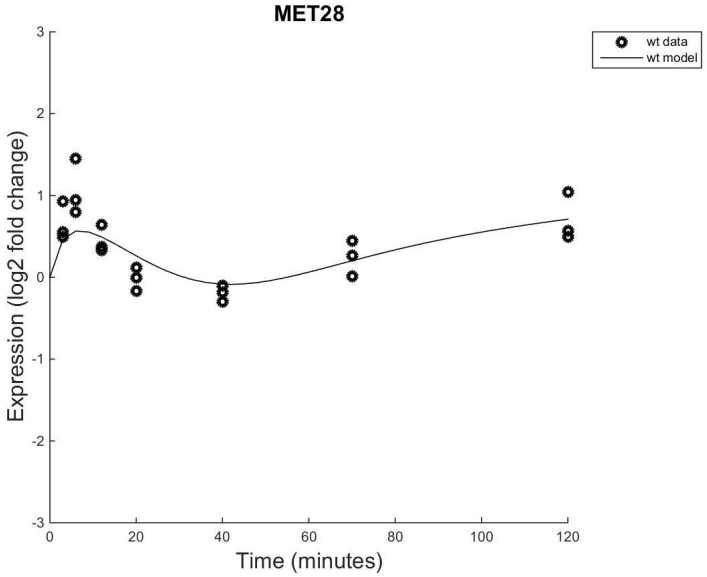
https://dondi.github.io/GRNsight/

Outline

- Data from Sha et al. was downloaded and prepared.
- Cluster profiles were visualized to determine the profile with the most genes which was found to be profile 41.
- A Microsoft Access database was created using the above data and gene regulatory network data.
- A GRNmap network was created for profile 41 based on queries conducted on the database.
- **Our findings agreed and disagreed with Sha et al. findings.**

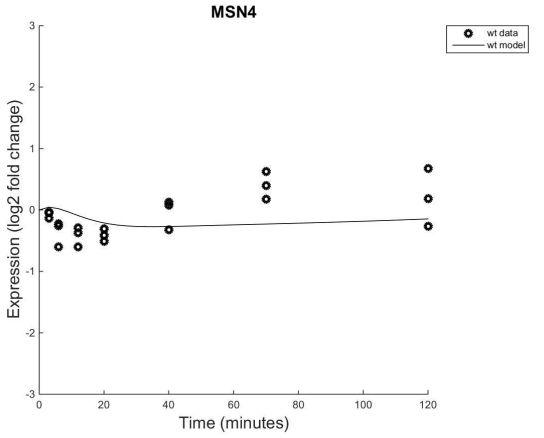
Our findings were in common with the Sha et al. paper regarding transcription factor MET28.

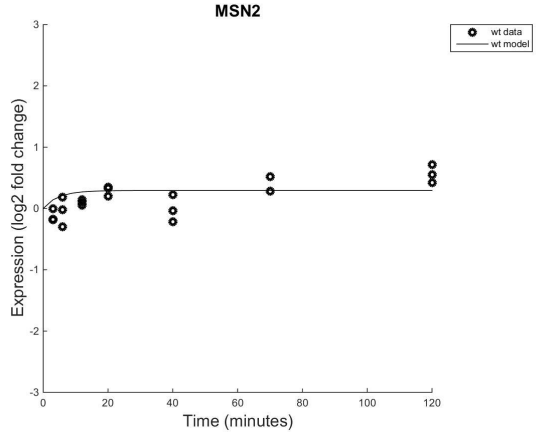
- *MET28 is “an early up-regulated gene.”*
  - Met28 is activated at the 3 minute time point by CBF1 and GCN4 (not discussed by Sha et al.).
  - This is mirrored in the expression plot.



Our findings were *somewhat* in common with the Sha et al. paper regarding transcription factors MSN2 and MSN4.

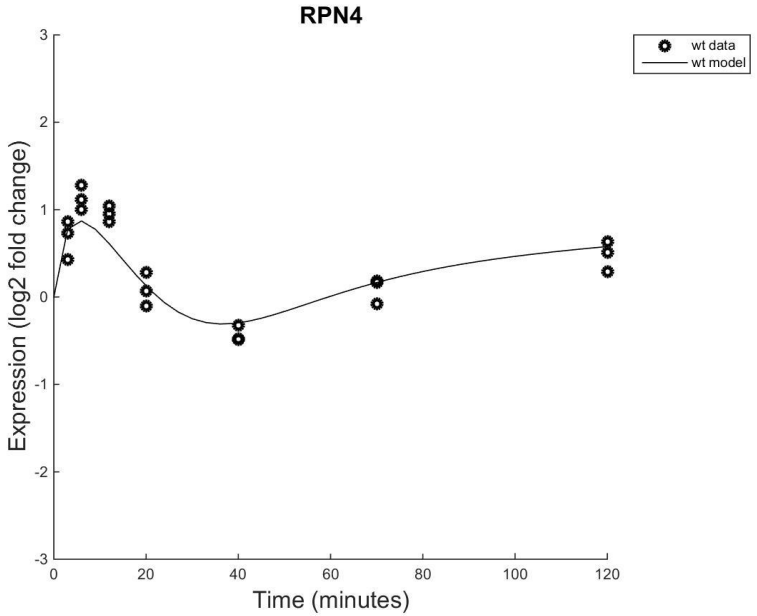
- *MSN2/4 did not show expression changes.*
  - MSN2 repressed MSN4 at the 3 minute mark.
  - MSN4 is repressed by SOK2 at some later time point, to a lesser extent.
  - MSN4 is shown light blue in most of its GRNmap space.
  - The expression plots agree with Sha et al., showing no significant changes.
    - This contradicts the supposed trend of profile 41.





Our findings were *somewhat* in common with the Sha et al. paper regarding transcription factor RPN4.

- *RPN4 is “induced very early.”*
  - RPN4 is shown red in the first half of its GRNmap space.
  - RPN4 is activated by CBF1 at 20 minutes.
  - RPN4 is activated by FHL1 at 40 minutes.
  - The expression plots agree with Sha et al., with an early peak in expression.

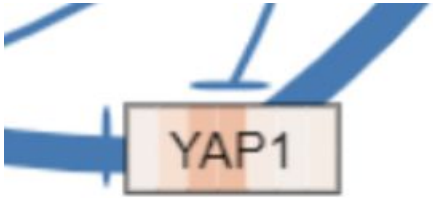


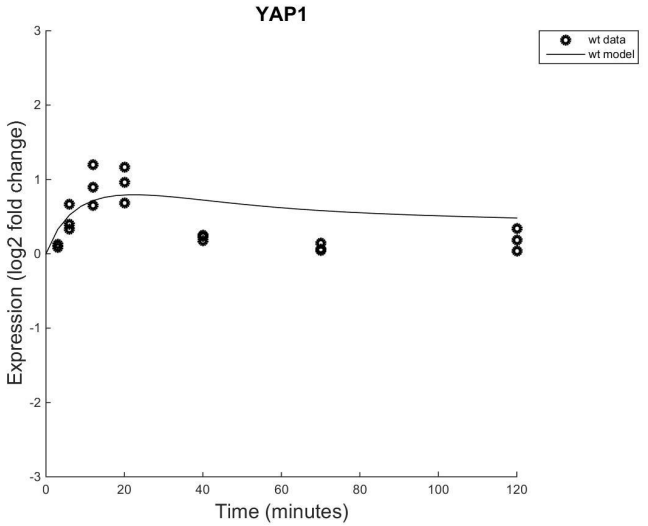
We made new discoveries about Tec1 and XBP1, which were not discussed in the Sha et al. paper.

- TEC1 had the most incoming edges, and is therefore influenced by many transcription factors.
  - Involved in cell stress response.
- XBP1 had the most outgoing edges, and is therefore influences many transcription factor.
  - Involved in glucose deprivation response.

Without all of the transcription factors involved in the cluster, there is still more to be considered before drawing conclusions.

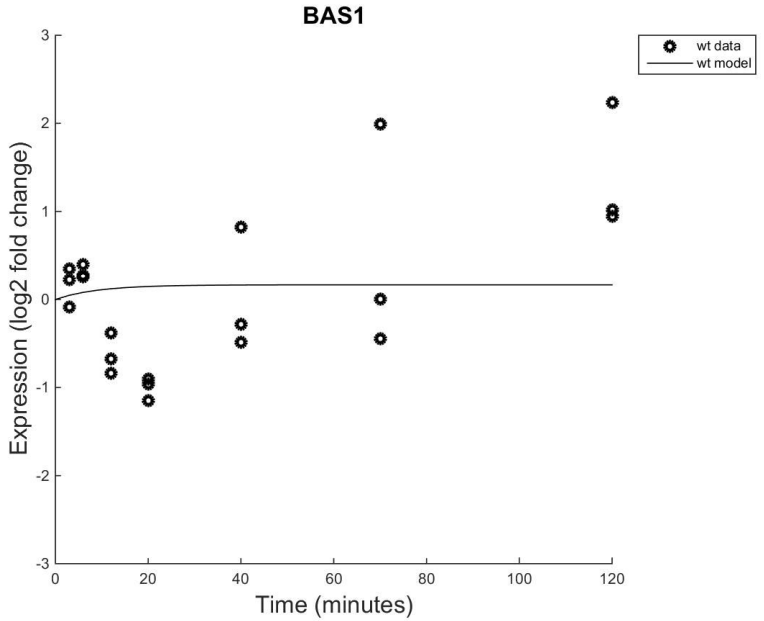
- There are other transcription factors at play, even if not displayed in the GRNmap.
- YAP1 is shown to be repressed by all of its incoming edges.
  - YAP1's expression is still activated, shown by red.
  - Activation of YAP1 is mirrored in its plot.





Because we used all of the timepoints, profile 41 had only a slight trend to track.

- GRNsight results revealed that in profile 41, log2 fold change > 0 on average.
- This is not as interesting as it could have been.
- Using a narrow range of time and comparing to another time range could be more fruitful.



Regulatory transcription factors and corresponding p-values are all significant.

Transcription Factor	p-value	Transcription Factor	p-value
Rpn4	0	Sok2	0
Gcn4	0	Msn2	0
Pdr1	0	Fhl1	0
Xbp1	0	Pdr3	0
Met28	0	Cbf1	1.80E-14
Spt23	0	Rph1	7.60E-14
Bas1	0	Stp1	1.21E-13
Yap1	0	Msn4	1.39E-12
		Tec1	6.80E-12

Future Directions

- Focus on the first different time sets of the CHP/Control Yeast experiment.
  - For example: 3-20 versus 40-120 minutes.
    - To compare the possibility of distinct stress responses occurring in the yeast.
- Include all of the transcription factors when making the GRNmap.
- Include .txt files or access files as data when including data from an experiment instead of Excel.

Conclusion

- Sha et al. data was collected and prepared.
- Cluster/profile 41 was the most significant.
- An Access database was made using the above data and Harbison et al. gene regulatory network data.
- A GRNmap network was made for profile 41 based on queries conducted on the database, but using GRNsight and 2024 data.
- Our findings agreed and disagreed with Sha et al. findings.
  - Our findings allowed for more nuance than the general trends.
- In the future, we could use different time point sets
- In the future, we could use all available transcription factors to make the GRNmap.

Thank You to All That Made This Presentation Possible

Thank you to the Loyola Marymount University Biology department, all of our classmates and fellow group members, as well as our professor, Dr. Dahlquist.

References

Balakrishnan, R., Park, J., Karra, K., Hitz, B. C., Binkley, G., Hong, E. L., Sullivan, J., Micklem, G., & Michael Cherry, J. (2012). YeastMine—An integrated data warehouse for Saccharomyces cerevisiae data as a multipurpose tool-kit. *Database*, *2012*. https://doi.org/10.1093/database/bar062

Ernst, J., Patek, D., & Bar-Joseph, Z. (n.d.). *STEM: Short Time-series Expression Miner* (1.2) [Computer software].

GRNsight v6.0.7. (2022, April 19). GRNsight. http://dondi.github.io/GRNsight/

Harbison, C. T., Gordon, D. B., Lee, T. I., Rinaldi, N. J., Macisaac, K. D., Danford, T. W., Hannett, N. M., Tagne, J.-B., Reynolds, D. B., Yoo, J., Jennings, E. G., Zeitlinger, J., Pokholok, D. K., Kellis, M., Rolfe, P. A., Takusagawa, K. T., Lander, E. S., Gifford, D. K., Fraenkel, E., & Young, R. A. (2004). Transcriptional regulatory code of a eukaryotic genome. *Nature*, *431*(7004), 99–104. https://doi.org/10.1038/nature02800

Laloux, I., Dubois, E., Dewerchin, M., & Jacobs, E. (1990). TEC1, a Gene Involved in the Activation of Ty1 and Ty1-Mediated Gene Expression in Saccharomyces cerevisiae: Cloning and Molecular Analysis. *Molecular and Cellular Biology*, *10*(7), 3541–3550. https://doi.org/10.1128/mcb.10.7.3541-3550.1990

LMU BioDB 2024. (2024). Final Project. Retrieved May 2, 2024, from https://xmlpipedb.cs.lmu.edu/biodb/spring2024/index.php/Final_Project

Mai, B., & Breeden, L. (1997). Xbp1, a Stress-Induced Transcriptional Repressor of the *Saccharomyces cerevisiae* Swi4/Mbp1 Family. *Molecular and Cellular Biology*, *17*(11), 6491–6501. https://doi.org/10.1128/MCB.17.11.6491

Microsoft Access (Access 2016). (2016). [C++]. Microsoft.

Saccharomyces Genome Database | *SGD*. (n.d.). Retrieved May 2, 2024, from https://www.yeastgenome.org/

Sha, W., Martins, A. M., Laubenbacher, R., Mendes, P., & Shulaev, V. (2013). The Genome-Wide Early Temporal Response of Saccharomyces cerevisiae to Oxidative Stress Induced by Cumene Hydroperoxide. *PLoS ONE*, *8*(9), e74939. https://doi.org/10.1371/journal.pone.0074939

Timoshina, N. (2023, February 8). *What is SQL Database: Structure, Types, Examples*. https://www.alphaservesp.com/blog/what-is-sql-database-structure-types-examples

Teixeira, M. C., Viana, R., Palma, M., Oliveira, J., Galocha, M., Mota, M. N., Couceiro, D., Pereira, M. G., Antunes, M., Costa, I. V., Pais, P., Parada, C., Chaouiya, C., Sá-Correia, I., & Monteiro, P. T. (2023). YEASTRACT+: A portal for the exploitation of global transcription regulation and metabolic model data in yeast biotechnology and pathogenesis. *Nucleic Acids Research*, *51*(D1), D785–D791. https://doi.org/10.1093/nar/gkac1041