Yeast Beasts

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

Outline

- **A microsoft access database was created with yeast microarray data and gene regulatory network data.**
- ANOVA was run, Bonferroni and Benjamini & Hochberg corrections were calculated to define significance.
- Cluster profiles were visualized to determine the profile with the most genes which was profile 41.
- A GRNmap network was created for profile 41 based on queries conducted on the database which gave new insights

Download processed Sha et al. DNA microarray data from SGD.

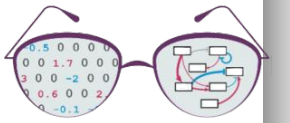
Further process data and run ANOVA analyses.

CHP data was clustered using STEM.

YEASTRACT created a list of RTFs from GO term names.

Access was used to create a GRNmap input workbook and database.

Queries were written to fill the GRNmap input workbook.



A GRNmap was created using GRNsight.

We developed a database to model GRNs

- We analyzed existing microarray data publications and used the GRNmap to construct gene regulatory networks
- Quality assurance was conducted on the existing GRNsight database and the newly created database

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

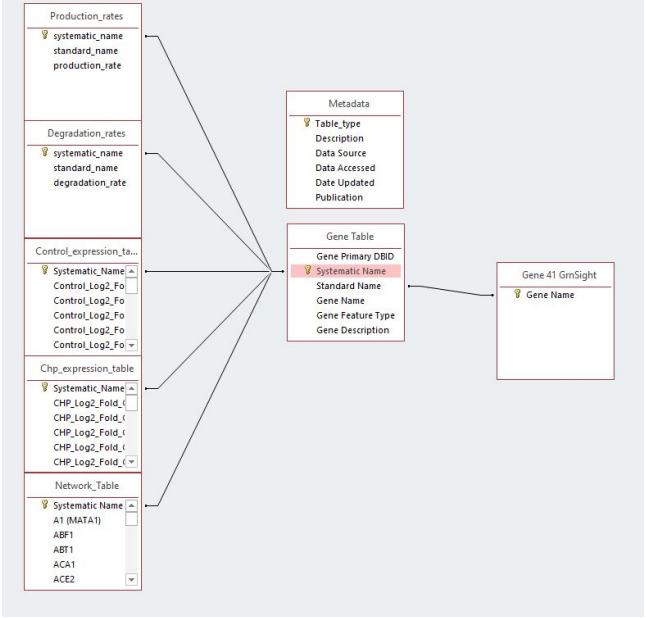
- The microarray dataset was uploaded to the GEO Database as GSE26169.
- The researchers only analyzed data points from times 0-20 minutes.
    - But they uploaded all 120 minutes that they collected.
    - We are using all 120 minutes of their data.

Transcription Factors selectively bind to DNA and influence transcription based on environment

- The paper attempted to predict the binding sites of 203 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).
- A metadata table which includes information about each of the tables.

Outline

- A microsoft access database was created with yeast microarray data and gene regulatory network data.
- **ANOVA was run, Bonferroni and Benjamini & Hochberg corrections were calculated to define significance.**
- Cluster profiles were visualized to determine the profile with the most genes which was profile 41.
- A GRNmap network was created for profile 41 based on queries conducted on the database which gave new insights

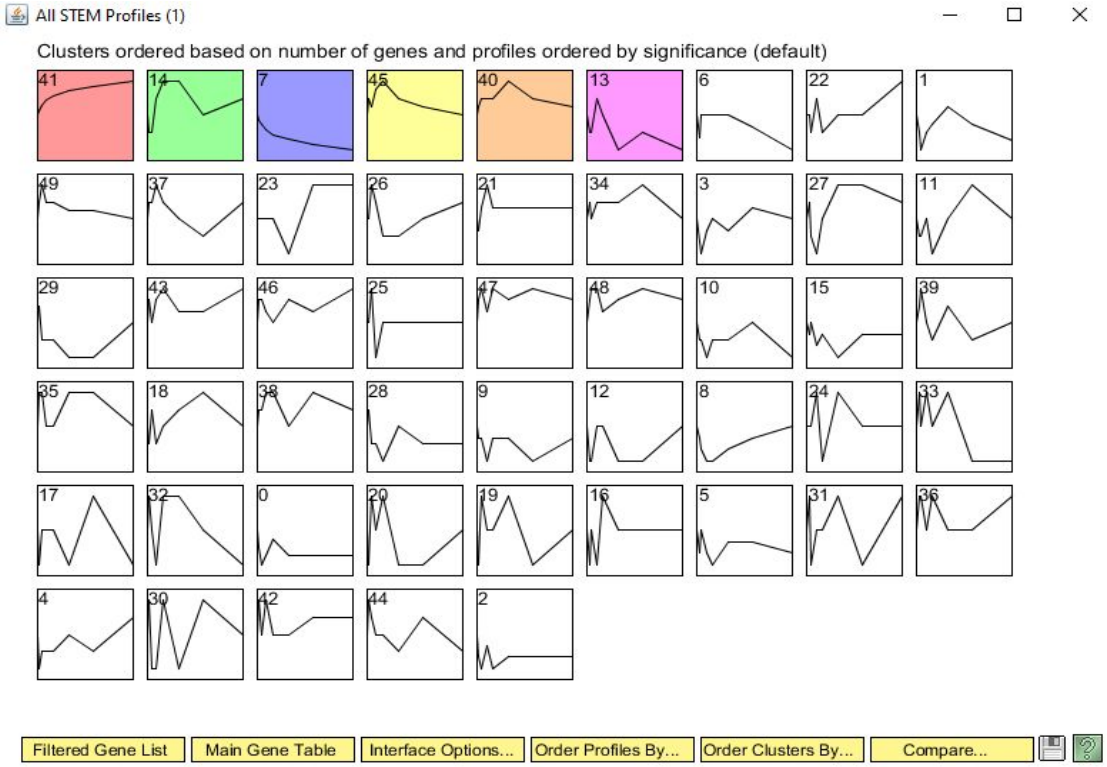
Selecting p<0.00001 for B&H ANOVA yielded ~25% significance for CHP and control datasets.

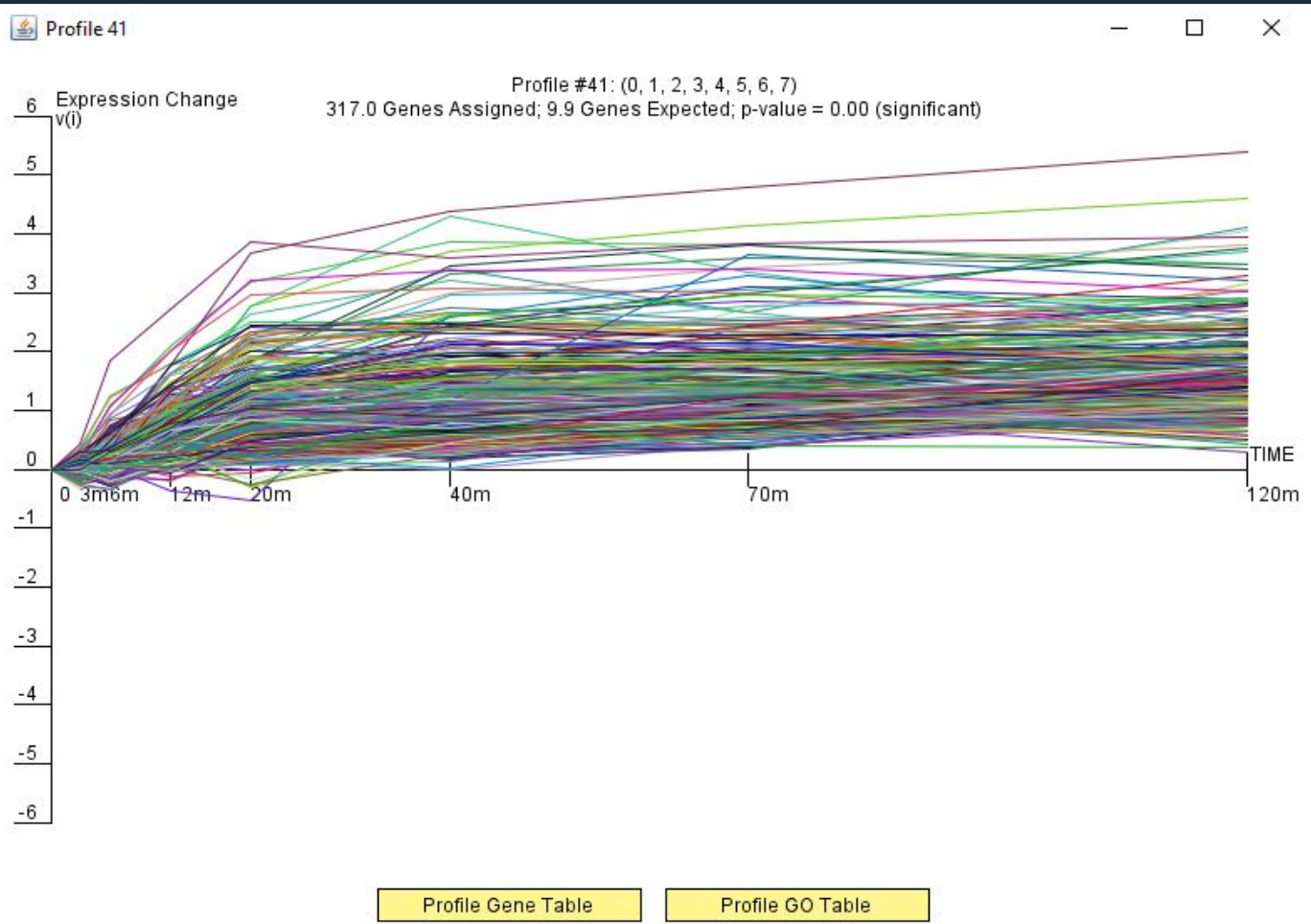
B&H ANOVA	Control Genes	CHP Genes
p<0.05	3699 (78.7%)	2856 (60.8%)
p<0.01	3219 (68.6%)	2390 (50.9%)
p<0.001	2558 (54.4%)	1857 (39.5%)
p<0.0001	1921 (40.9%)	1419 (30.2%)
p<0.00001	1325 (28.2%)	1058 (22.5%)

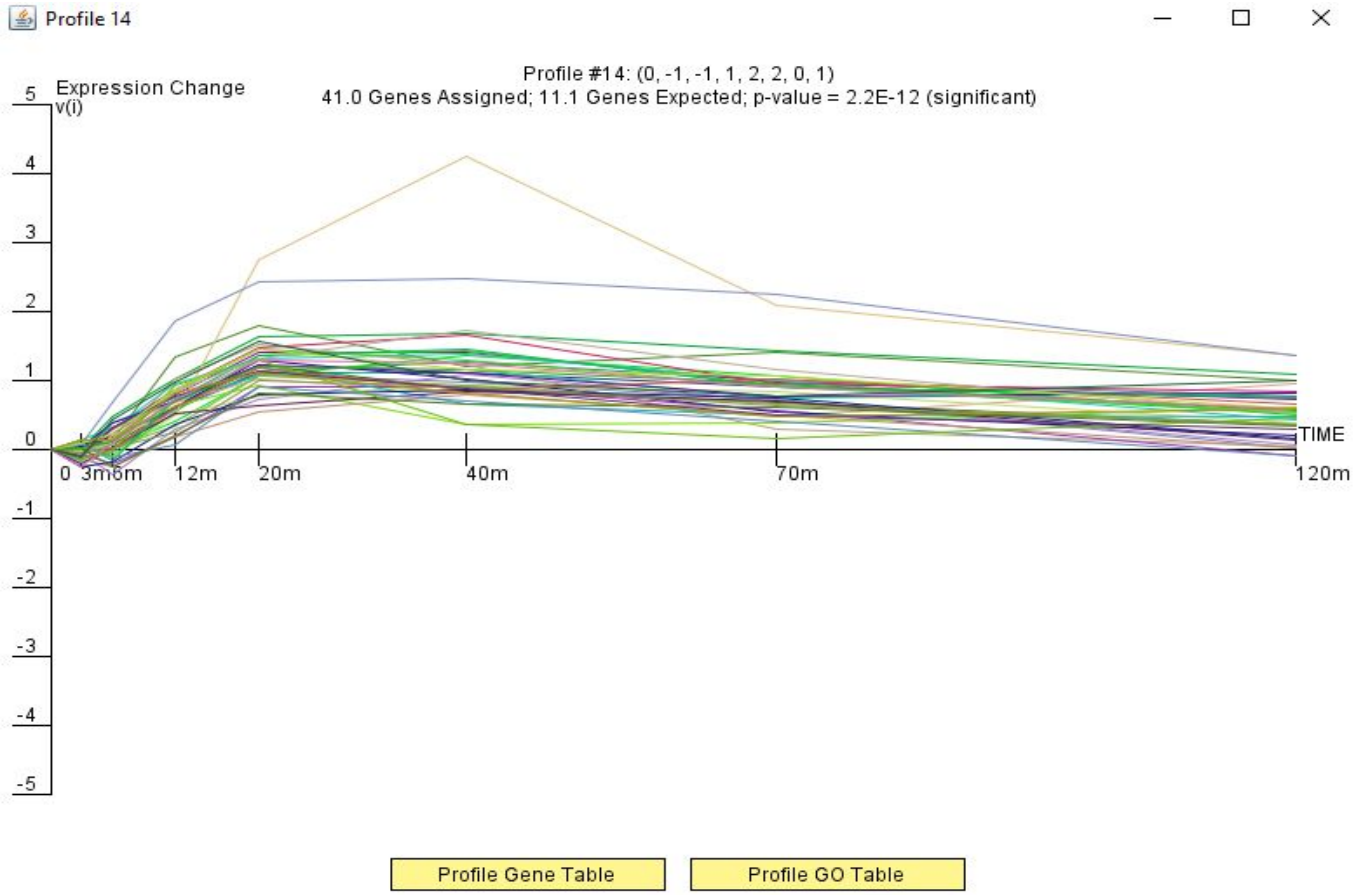
Outline

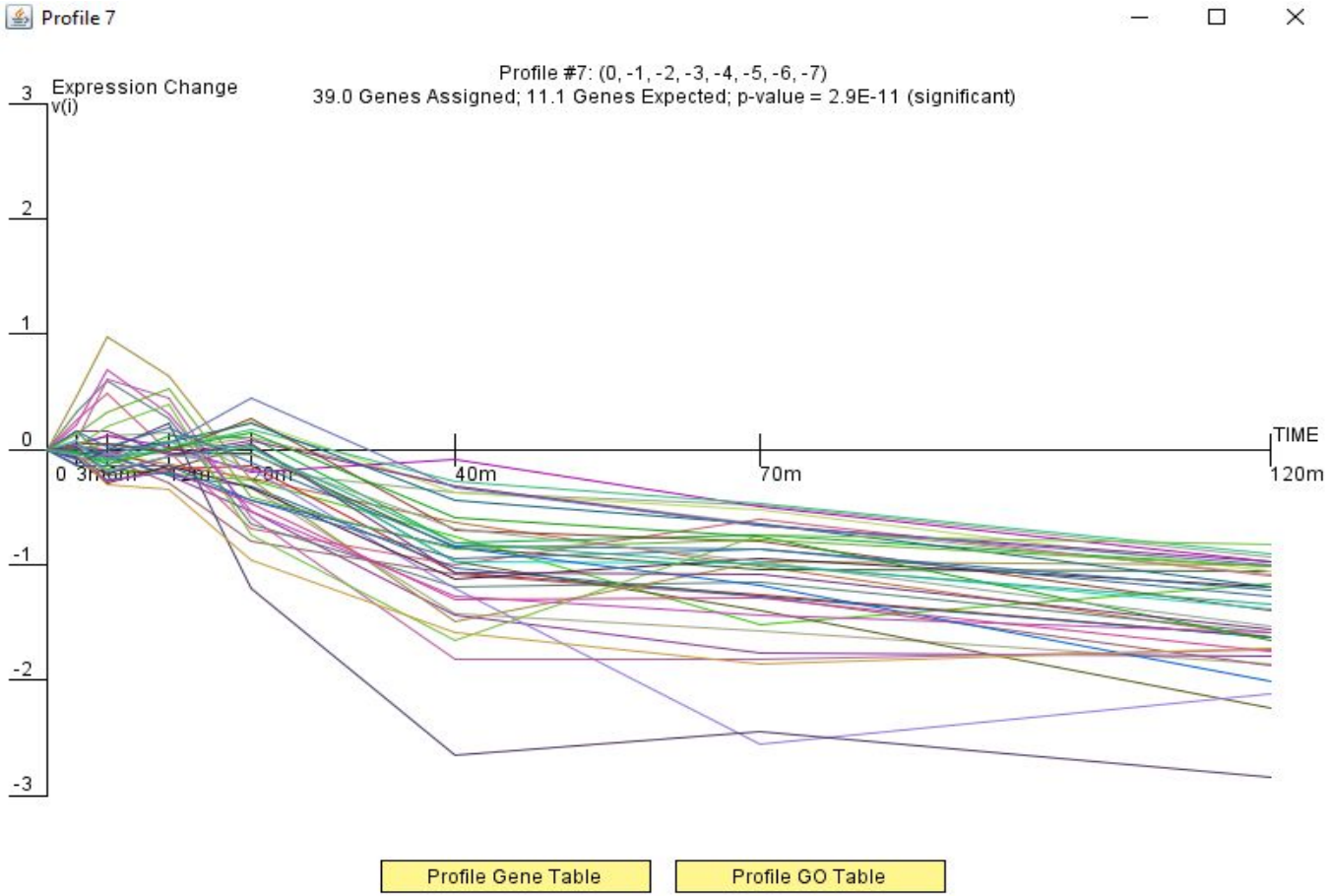
- A microsoft access database was created with yeast microarray data and gene regulatory network data.
- ANOVA was run, Bonferroni and Benjamini & Hochberg corrections were calculated to define significance.
- **Cluster profiles were visualized to determine the profile with the most genes which was profile 41.**
- A GRNmap network was created for profile 41 based on queries conducted on the database which gave new insights

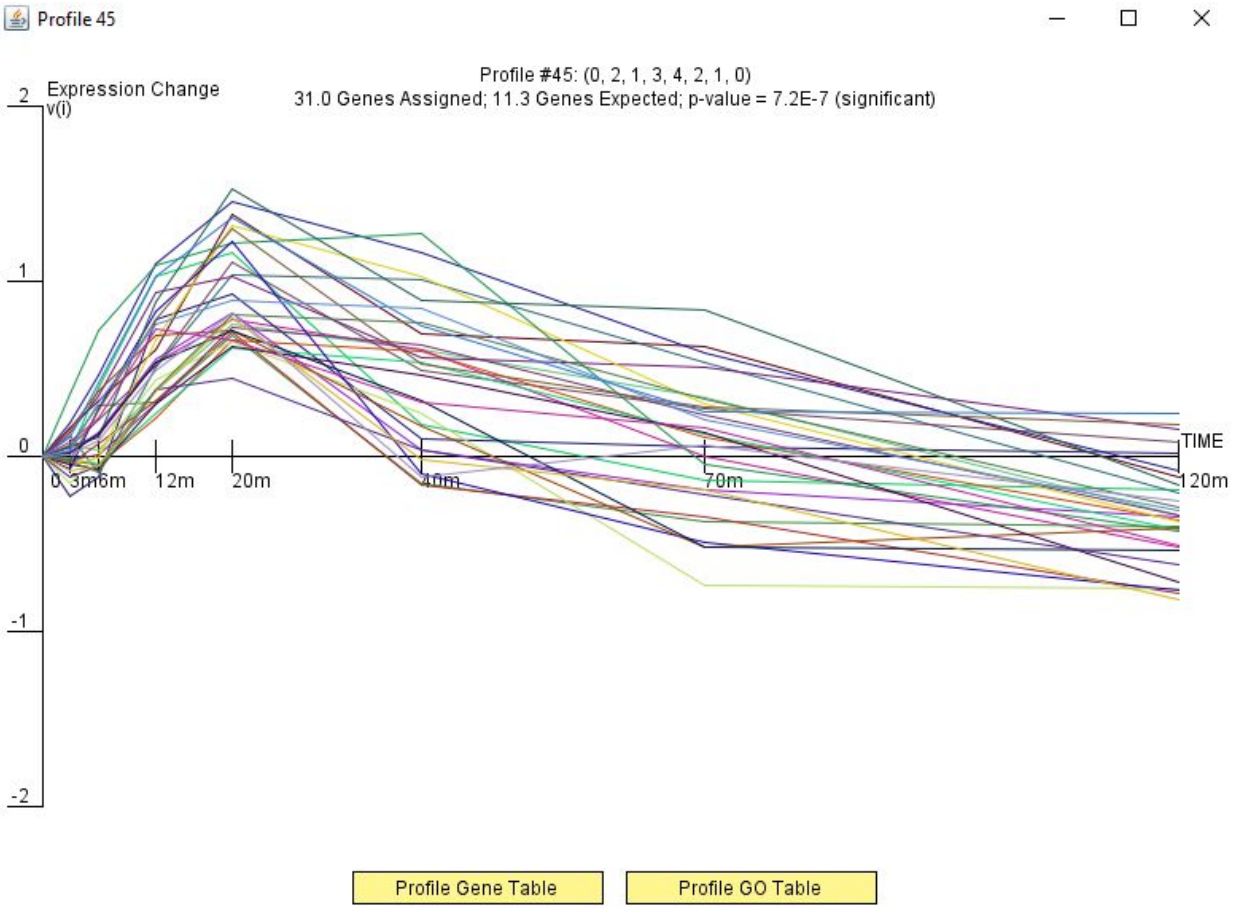
Selecting p<0.00001 for B&H ANOVA yielded ~25% significance for CHP and control datasets.

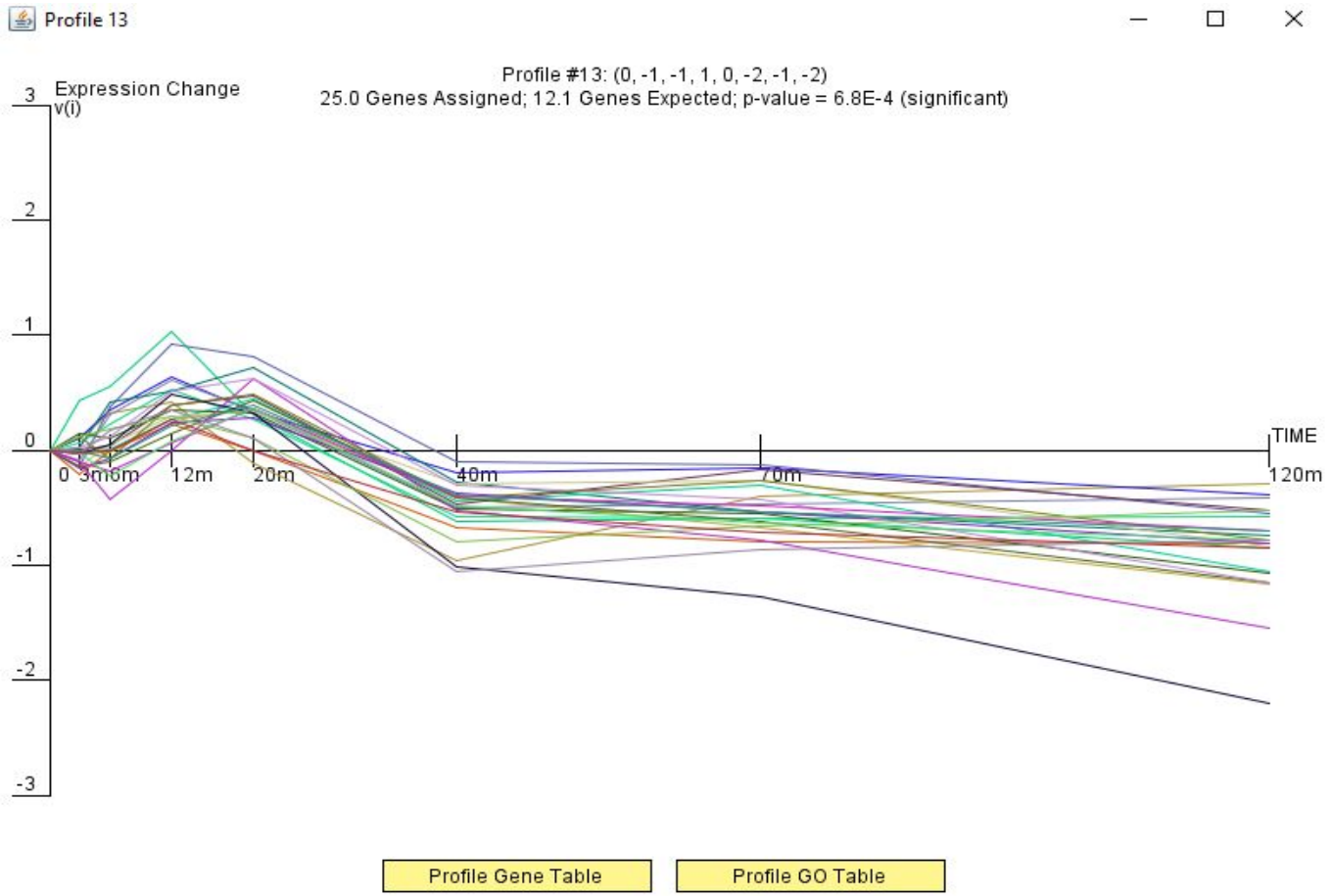



Profile 41
Profile #41: (0, 1, 2, 3, 4, 5, 6, 7)
317.0 Genes Assigned; 9.9 Genes Expected; p-value = 0.00 (significant)
Expression Change
v(i)
6
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
0
3m
6m
12m
20m
40m
70m
120m
TIME
Profile Gene Table
Profile GO Table


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

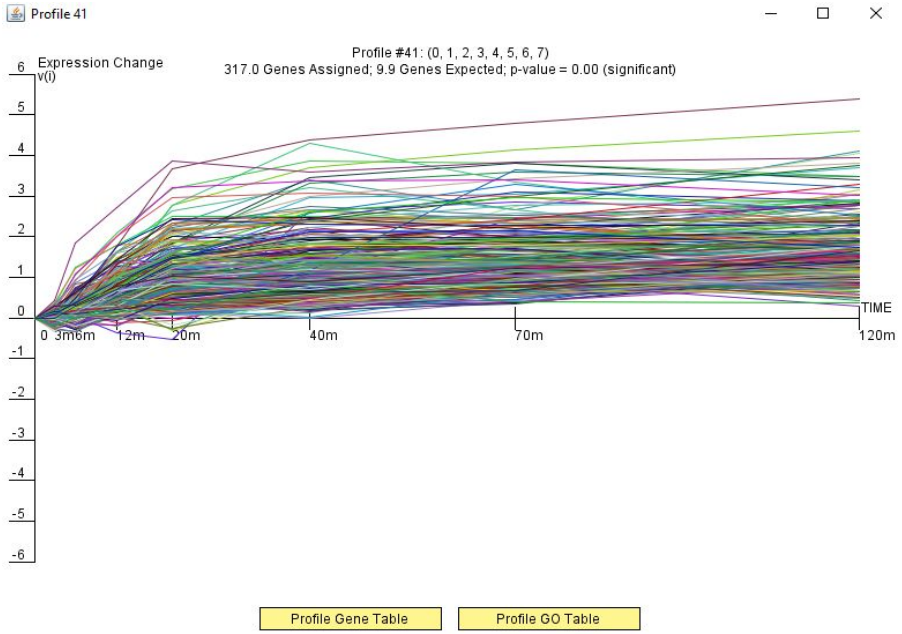

Profile 7
Profile #7: (0, -1, -2, -3, -4, -5, -6, -7)
39.0 Genes Assigned; 11.1 Genes Expected; p-value = 2.9E-11 (significant)
Expression Change
v(i)
3
2
1
0
-1
-2
-3
TIME
0
40m
70m
120m
Profile Gene Table
Profile GO Table


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.



Profile 41 was selected because it was assigned the most genes.

- Unlike the Sha et al. paper, we used all timepoints 0-120 minutes.
  - Different patterns would be observed if minutes 0-20 were focused on.
- Significant genes lack distinct upregulation or downregulation.
- Dr. Dahlquist removed Crz1, Met31, Mga2, Rlm1 because they were not connected to the gene regulatory network.

The Most Significant GO Results are 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

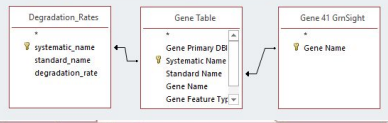
GO Results Show Genes Belonging to Sorting Pathways

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

Outline

- A microsoft access database was created with yeast microarray data and gene regulatory network data.
- ANOVA was run, Bonferroni and Benjamini & Hochberg corrections were calculated to define significance.
- Cluster profiles were visualized to determine the profile with the most genes which was profile 41.
- **A GRNmap network was created for profile 41 based on queries conducted on the database which gave new insights**

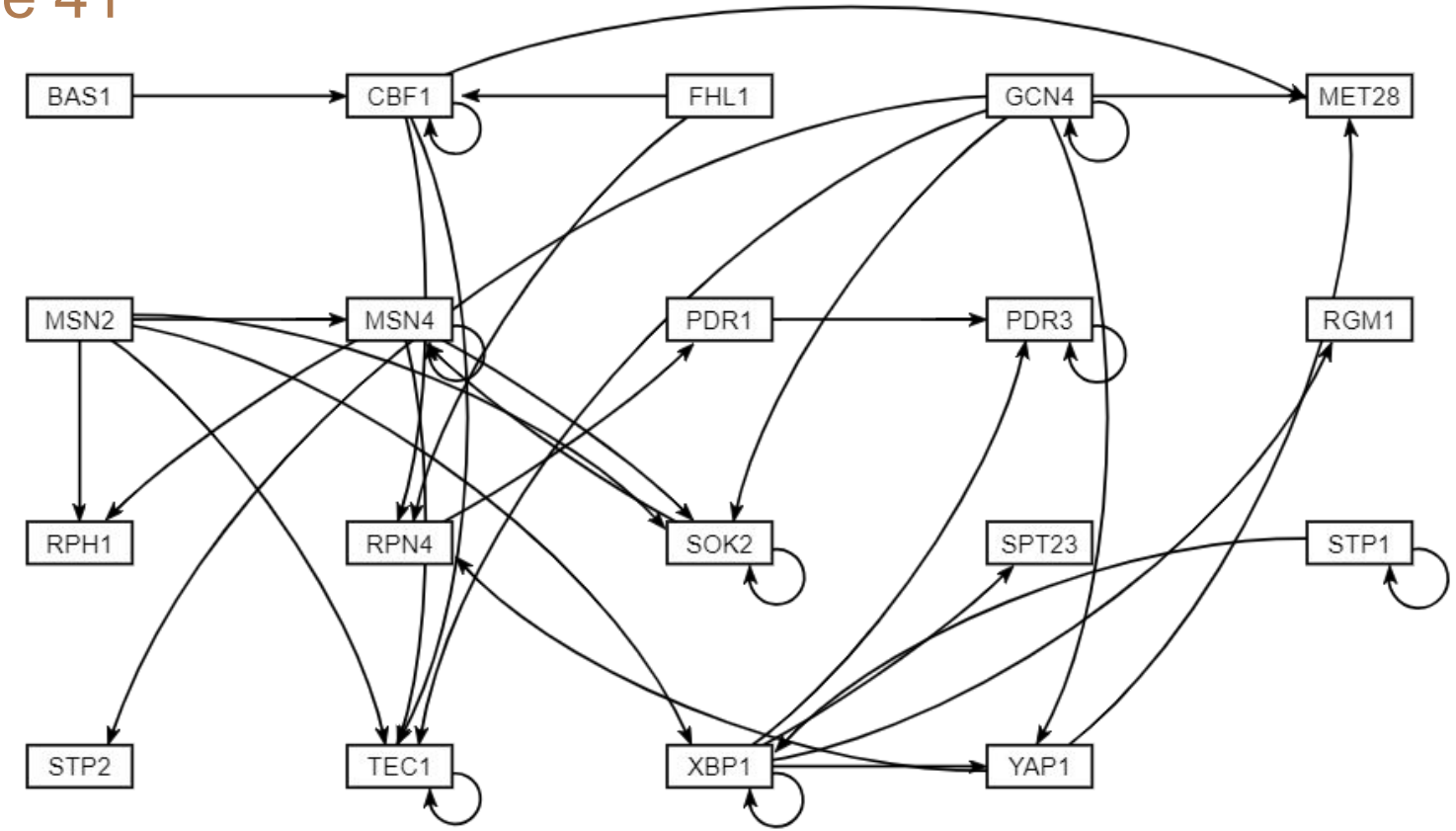
The Queries from the database layed a foundation for the GRNmap network



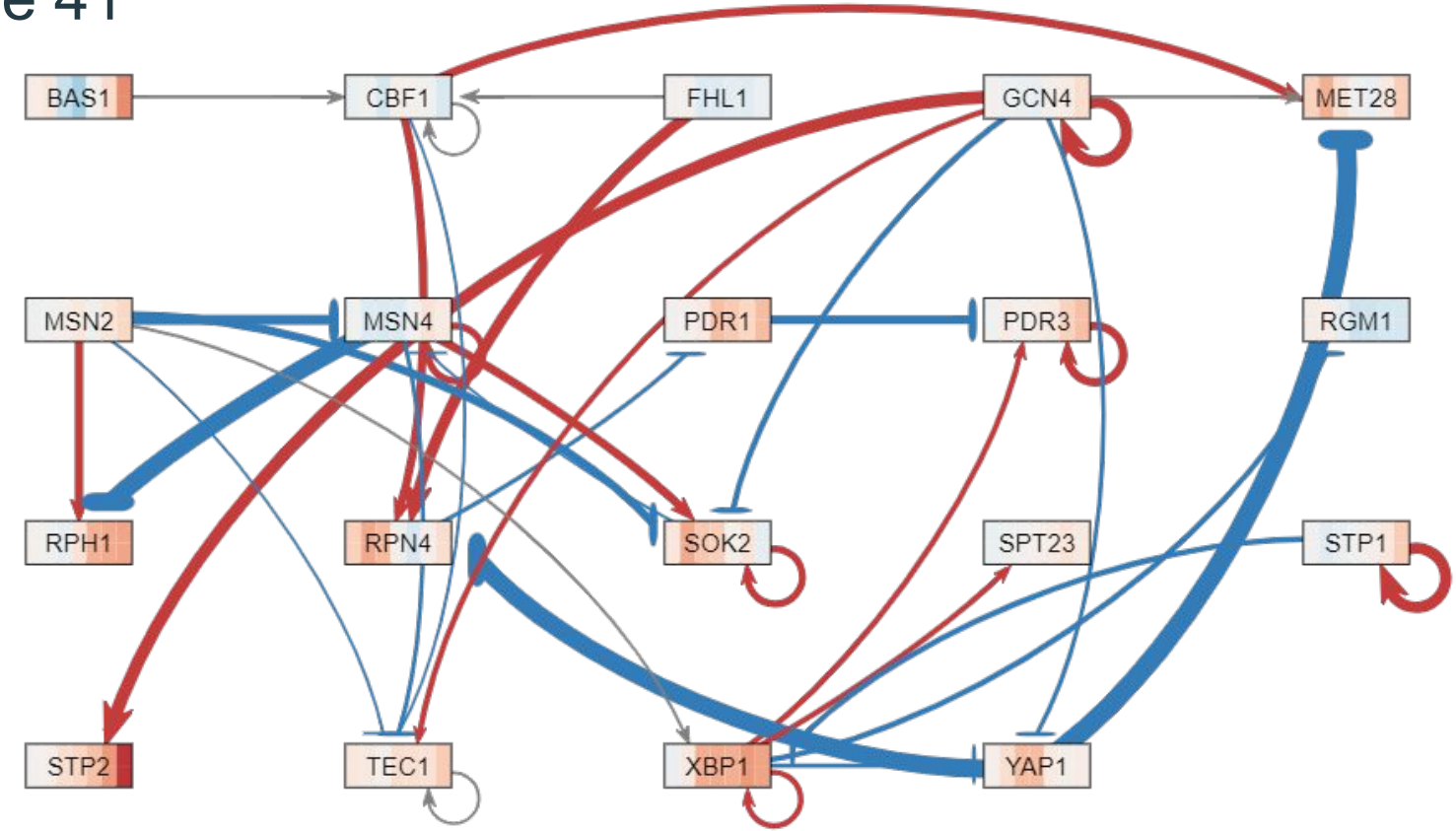
systematic_name	standard_name	degradation_rate
YKR099W	BAS1	0.1066
YJR060W	CBF1	0.0835
YNL027W	CRZ1	0.1238
YPR104C	FHL1	0.105
YEL009C	GCN4	0.0513
YIR017C	MET28	0.0542
YPL038W	MET31	0.055
YMR037C	MSN2	0.2039
YKL062W	MSN4	0.1386
YGL013C	PDR1	0.1083
YBL005W	PDR3	0.1359
YMR182C	RGM1	0.2666
YPL089C	RLM1	0.2236
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 allowed a network of profile 41 to be created.

Profile 41


BAS1
CBF1
FHL1
GCN4
MET28
MSN2
MSN4
PDR1
PDR3
RGM1
RPH1
RPN4
SOK2
SPT23
STP1
STP2
TEC1
XBP1
YAP1

Profile 41


BAS1
CBF1
FHL1
GCN4
MET28
MSN2
MSN4
PDR1
PDR3
RGM1
RPH1
RPN4
SOK2
SPT23
STP1
STP2
TEC1
XBP1
YAP1

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 20 minutes of the CHP/Control Yeast experiment.
- Include txt files or access files as data when including data from an experiment instead of Excel
-

Conclusion

- A microsoft access database was created with yeast microarray data and gene regulatory network data.
- ANOVA was run, Bonferroni and Benjamini & Hochberg corrections were calculated to define significance.
- Cluster profiles were visualized to determine the profile with the most genes which was profile 41.
- A GRNmap network was created for profile 41 based on queries conducted on the database which gave new insights

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

Harbison, C. T., Gordon, D. B., Lee, T. I., Rinaldi, N. J., Macisaac, K. D., Danford, T. W., ... & Young, R. A. (2004). Transcriptional regulatory code of a eukaryotic genome. *Nature, 431*(7004), 99-104. doi: 10.1038/nature02800

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. doi:10.1371/journal.pone.0074939

Short time-series expression miner (stem). STEM: Short Time-series Expression Miner. (n.d.). https://www.cs.cmu.edu/~jernst/stem/

Microsoft Corporation. (2016). *Microsoft Access*. Retrieved from https://www.microsoft.com/en-us/microsoft-365/access

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: 10.7717/peerj-cs.85.

M.C. Teixeira, R. Viana, M. Palma, J. Oliveira, M. Galocha, M.N. Mota, D. Couceiro, M.G. Pereira, M. Antunes, I.V. Costa, P. Pais, C. Parada, C. Chaouiya, I. SáCorreira, P.T. Monteiro (2023)

YEASTRACT+: a portal for the exploitation of global transcription regulation and metabolic model data in yeast biotechnology and pathogenesis

Nucleic Acids Research, advance access (doi:10.1093/nar/gkac1041)
XMLPipeDB. (n.d.). Final Project. Retrieved May 1, 2024, from https://xmlpipedb.cs.lmu.edu/biodb/spring2024/index.php/Final_Project