

# **JASPAR information is part of a new feature of GRNsight that increases the application's ability to interpret the results of gene profiles**

**Antonio Porras - Department of Biology**

**Eddie Bachoura - Department of Computer Science**

**Quinn Lanners - Department of Mathematics**

**Simon Wroblewski - Department of Computer Science**



**Biological Databases CMSI/BIOL 367-01**

**December 12, 2017**



# Outline

- **GRNsight maps the relationships between transcription factors of gene profiles constructed from microarray data**
- ***JASPAR the Friendly Ghost* team members worked simultaneously on separate but related tasks**
- **Relationships were visualized between the most significant transcription factors in Profile 22 of dZAP1**
- **JASPAR provided integral information regarding transcription factors for the added feature to GRNsight**
- **Future work includes ensuring robust functionality and adding to the information available for the genes on GRNsight**

# Outline

- **GRNsight maps the relationships between transcription factors of gene profiles constructed from microarray data**
- *JASPAR the Friendly Ghost* team members worked simultaneously on separate but related tasks
- Relationships were visualized between the most significant transcription factors in Profile 22 of dZAP1
- JASPAR provided integral information regarding transcription factors for the added feature to GRNsight
- Future work includes ensuring robust functionality and adding to the information available for the genes on GRNsight

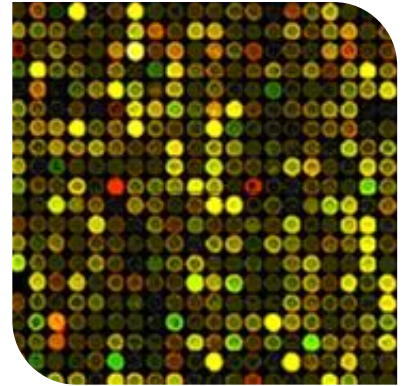
# **Cold shock has been found to have an impact on the expression of the yeast genome**

- ***Saccharomyces cerevisiae* is an ideal model organism for the studying of gene expression.**
- **The effect of environmental stressors on gene expression is a common theme to study in genomics.**
- **Extreme cold shock (30°C to 4°C) has been found to significantly impact the expression of over 100 genes**

**(Becerra, et. al., 2003).**

# Studying gene profiles provides information on how cold shock effects related genes

- Gene expression profiles can be created from microarray data.
- Gene profiles create a snapshot of all of the genes significantly impacted by a particular stressor.
  - Allowing for the clumping of genes that perform related functions and/or are controlled by related regulators
- It is difficult to determine which transcription factors belong in, and play an important role in, the gene regulatory network.



# GRNsight allows for the visualization of the relationships between transcription factors

- The application GRNmap can be run on MATLAB to mathematically model and simulate gene regulatory networks (About GRNsight).
  - However, outputted data is in the form of tabulated data
- GRNsight allows for visualization of this data in an interactive map (About GRNsight).
  - Nodes: genes
  - Edges: regulatory relationships



# **JASPAR Database offers crucial information regarding transcription factors**

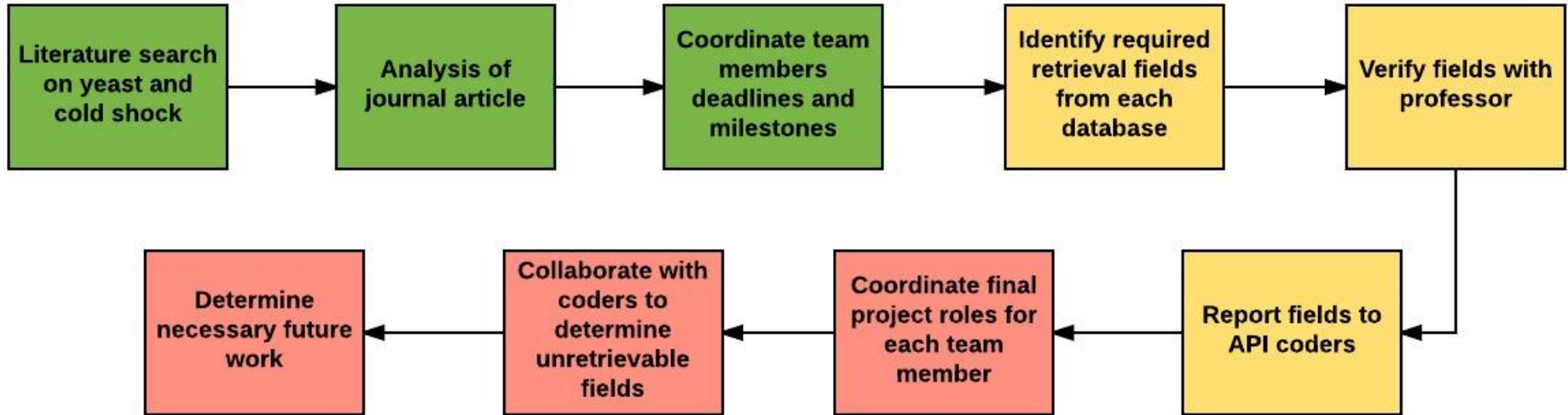
- **General information regarding each gene would improve the functionality of GRNsight.**
- **Important information regarding genes can be generally obtained from five databases:**
  - **Ensembl**
  - **NCBI**
  - **SGD**
  - **UniProt**
  - **JASPAR**
- **JASPAR offers information tailored specifically for transcription factors.**

# Outline

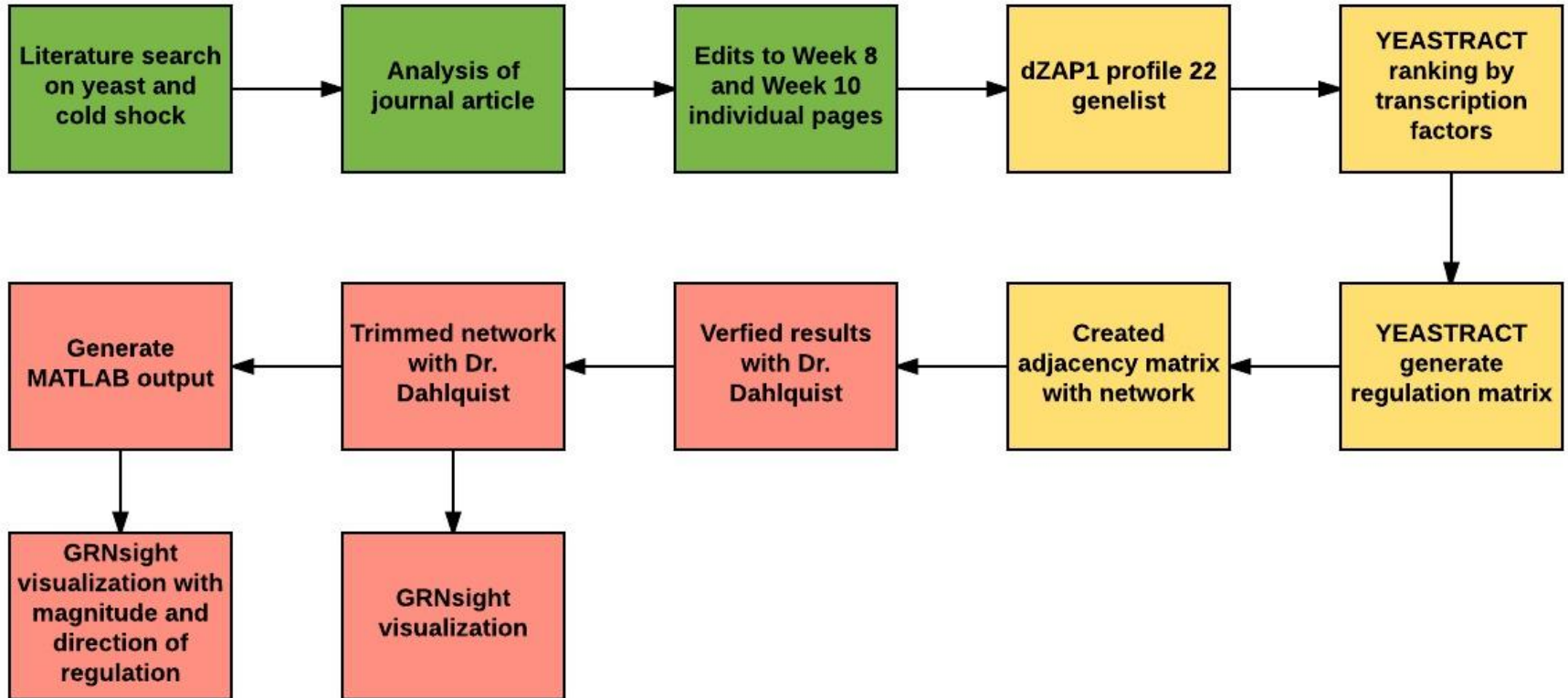
- GRNsight maps the relationships between transcription factors of gene profiles constructed from microarray data
- ***JASPAR the Friendly Ghost* team members worked simultaneously on separate but related tasks**
- Relationships were visualized between the most significant transcription factors in Profile 22 of dZAP1
- JASPAR provided integral information regarding transcription factors for the added feature to GRNsight
- Future work includes ensuring robust functionality and adding to the information available for the genes on GRNsight



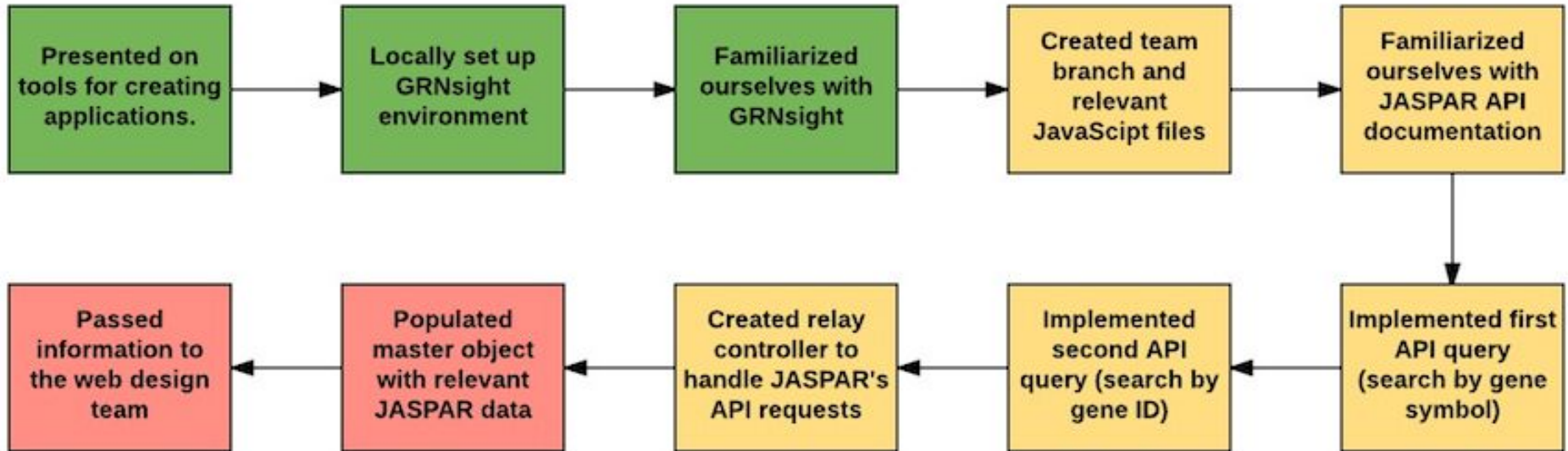
# Quality Assurance/Project Manager coordinated team activity and outlined necessary deliverables



# Data Analyst interpreted STEM results of dZAP1 and generated a network for GRNsight visualization



# Coders wrote code to retrieve data from JASPAR and delivered results to the design team



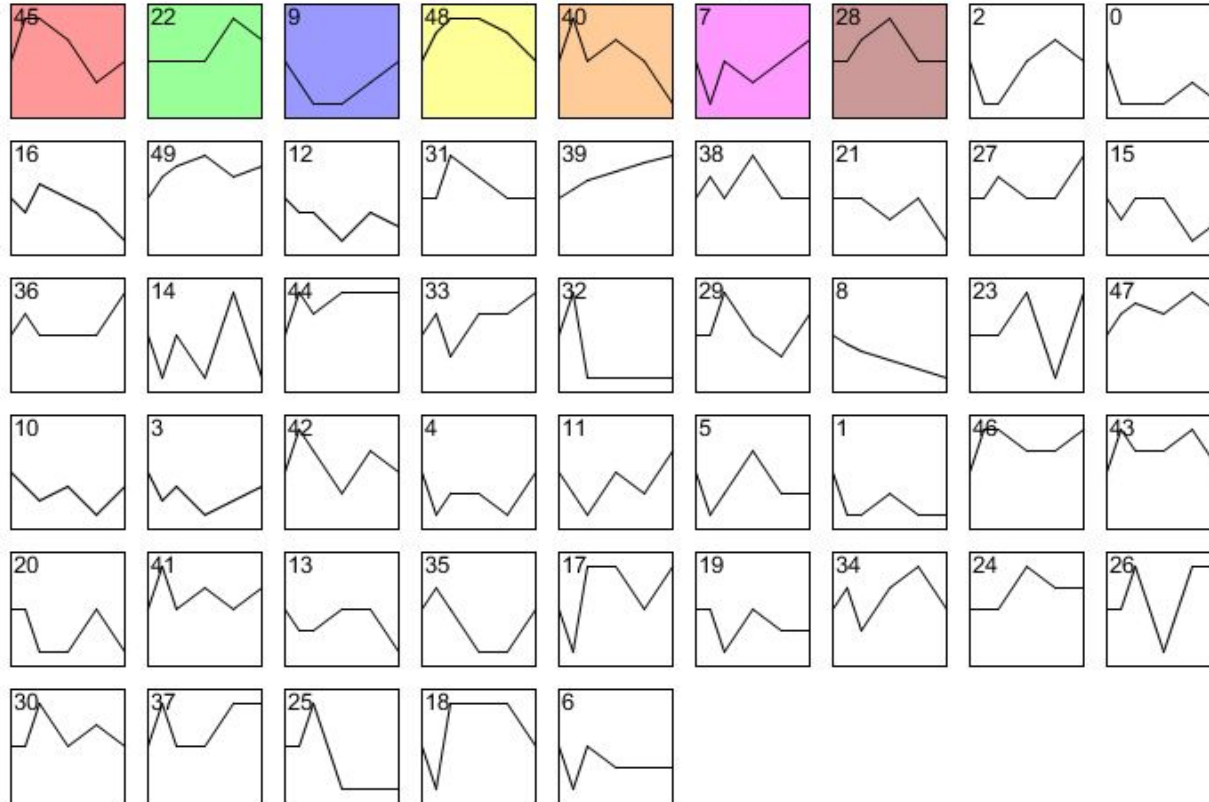
# Outline

- GRNsight maps the relationships between transcription factors of gene profiles constructed from microarray data
- *JASPAR the Friendly Ghost* team members worked simultaneously on separate but related tasks
- **Relationships were visualized between the most significant transcription factors in Profile 22 of dZAP1**
- JASPAR provided integral information regarding transcription factors for the added feature to GRNsight
- Future work includes ensuring robust functionality and adding to the information available for the genes on GRNsight

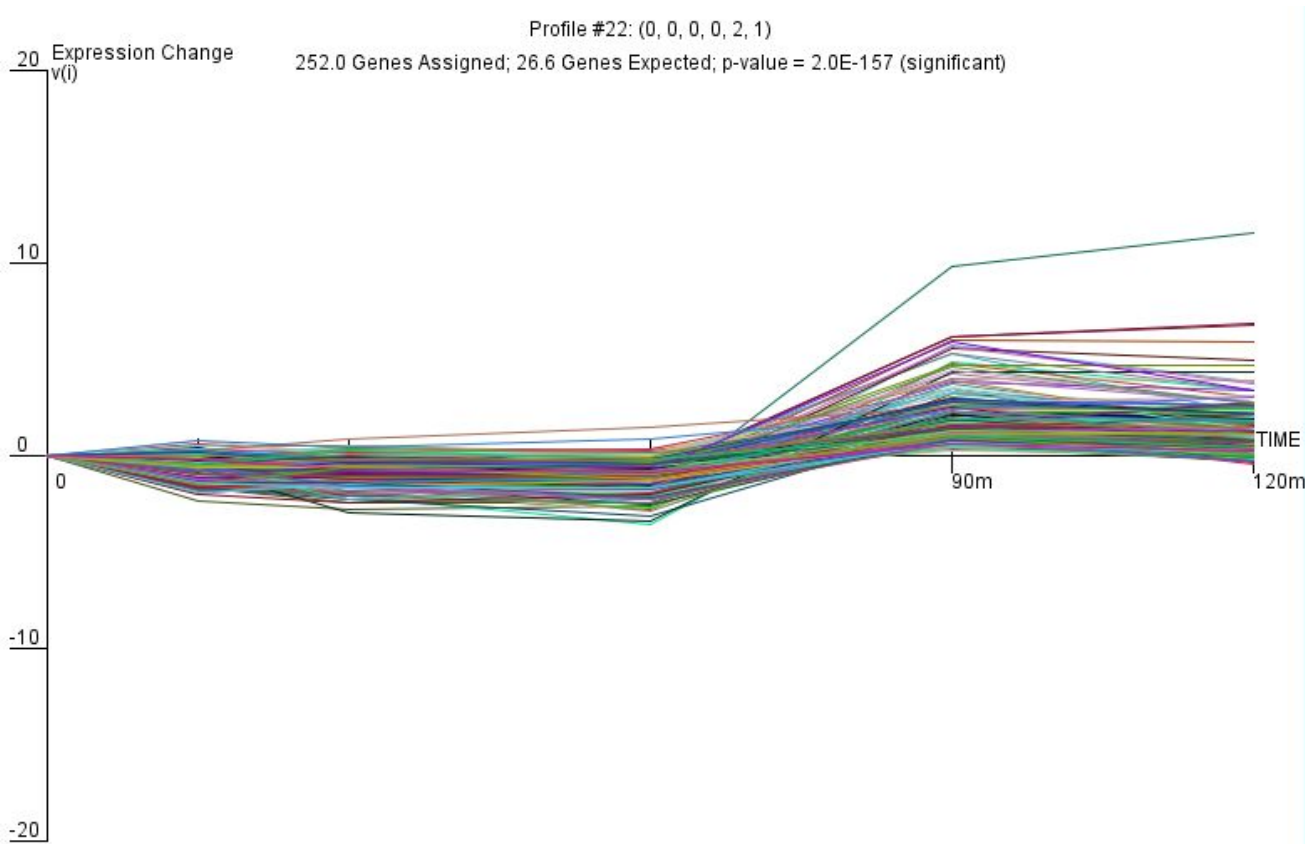
# Genes had expression change significantly different than zero at any timepoint

ANOVA	WT	<i>dZAP1</i>
$p < 0.05$	2528 (40.85%)	2485 (40.2%)
$p < 0.01$	1652 (26.70%)	1609 (26.0%)
$p < 0.001$	919 (14.85%)	885 (14.3%)
$p < 0.0001$	496 (8.01%)	457 (7.4%)
Benjamini & Hochberg-corrected $p < 0.05$	1822 (29.44%)	1766 (28.5%)
Bonferroni-corrected $p < 0.05$	248 (4.01%)	209 (3.4%)

# Seven significant expression profiles were generated using high-level analysis



# Profile #22 displayed no significant change in expression until timepoint 60m



# GOlist provided the nature and expression of significant genes in Profile #22

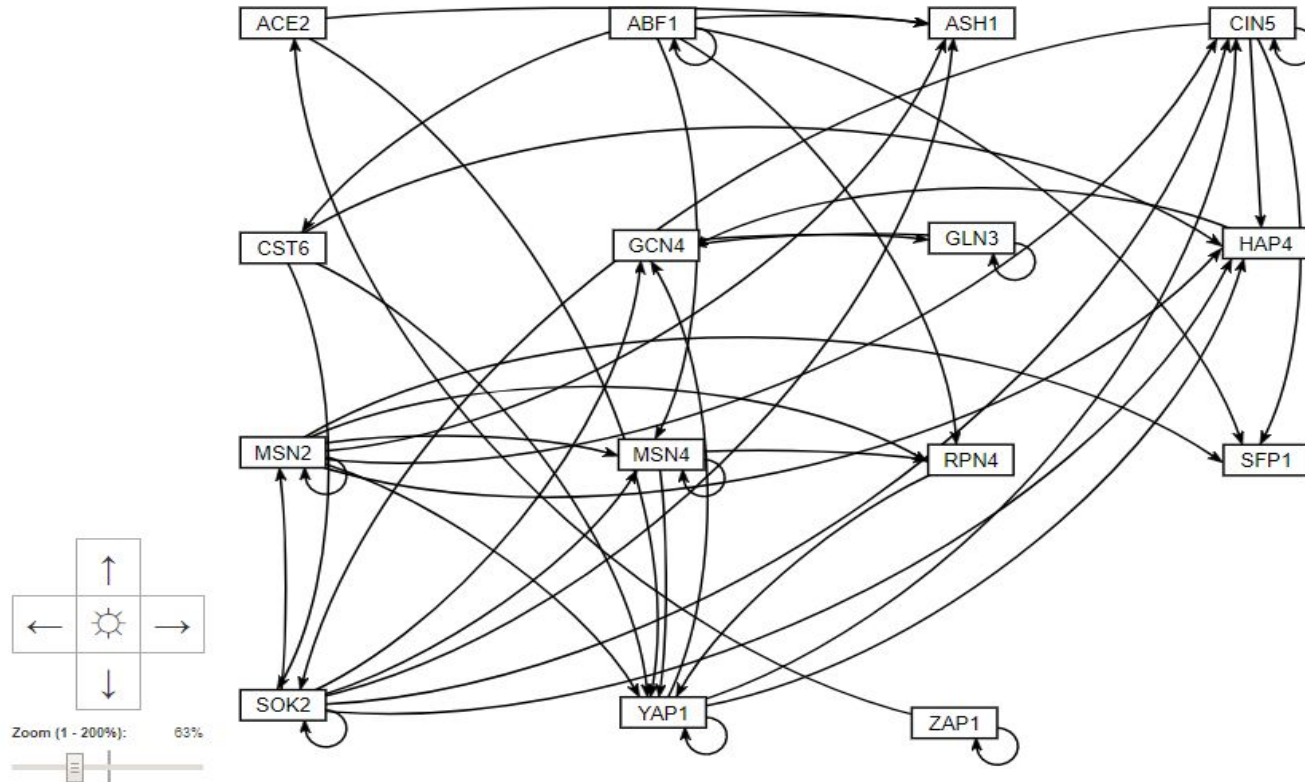
Category Name	#Genes Category	#Genes Assigned	#Genes Expected	#Genes Enriched	p-value	Corrected p-value	Fold
cytoplasm	1238	215	174.8	40.2	3.00E-10	<0.001	1.2
oxidoreductase activity	131	45	18.5	26.5	8.10E-10	<0.001	2.4
cellular response to oxidative stress	51	25	7.2	17.8	1.50E-09	<0.001	3.5



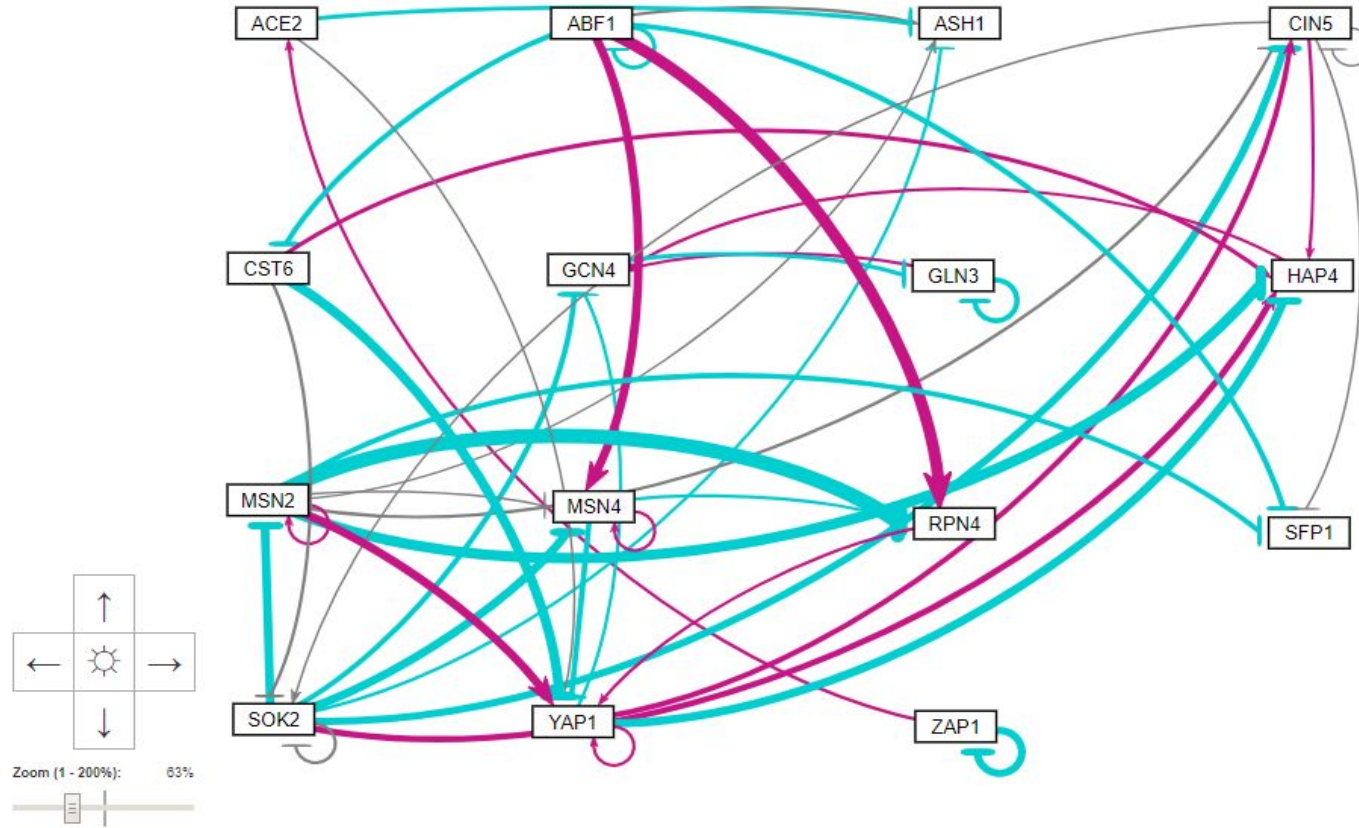
# YEASTRACT generated and ranked 49 transcription factors

Transcription Factor	ABF1	ACE2	ASH1	CIN5	CST6	GCN4	GLN3	HAP4
p-value	1.58E-07	1.69E-08	1.13E-09	4.62E-05	2.14E-05	5.48E-08	0.513787	2.12E-08
Transcription Factor	MSN2	MSN4	RPN4	SFP1	SOK2	YAP1	ZAP1	
p-value	1.48E-11	4.73E-12	2.95E-13	9.66E-10	1.91E-09	1.39E-09	2E-15	

# GRNsight generated a visual representation of the regulatory network



# Weighted regulation map aided in visualizing direction and magnitude of regulation



# Outline

- GRNsight maps the relationships between transcription factors of gene profiles constructed from microarray data
- *JASPAR the Friendly Ghost* team members worked simultaneously on separate but related tasks
- Relationships were visualized between the most significant transcription factors in Profile 22 of dZAP1
- **JASPAR provided integral information regarding transcription factors for the added feature to GRNsight**
- Future work includes ensuring robust functionality and adding to the information available for the genes on GRNsight

# Setup GRNsight environment locally

- Coders cloned the files that make up the GRNsight application to our own machines.
- Ensured local GRNsight is functioning properly with npm install and npm start
- Became familiar with GRNsight and created new files to mark where we would implement the new feature
- Studied JASPAR's API documentation to understand which calls we would be using



**getGeneInformation ()** function searches each database and compiles the data from each into an object

- Integrated knowledge of the JASPAR API documentation into the new function `getGeneInformation ()`
- Created the function `getJasparInfo ()` which would contain the relevant API calls
- Identified the gene information that needed to be called and returned from the JASPAR API
- Wrote and integrated the `getJasparInfo ()` function

```
var getJasparInfo = function (geneSymbol) {
  ...return $.get({
    .....url: "/jaspar/api/v1/matrix/?tax_id=4932&format=json&search=" + geneSymbol,
    .....dataType: "json",
    .....beforeSend: function (xhr) {
      .....xhr.setRequestHeader("content-type", "application/json");
    .....},
  .....}).then(function (data) {
    .....return (data.count === 0 ? {} :
    .....$.get({
      .....url: "/jaspar/api/v1/matrix/" + data.results[0].matrix_id,
      .....dataType: "json",
      .....beforeSend: function (xhr) {
        .....xhr.setRequestHeader("content-type", "application/json");
      .....},
    .....}))
  .....});
  .....});
};
```

**Second API call returns the query's gene object, using the matrix ID retrieved from the first call**

# After finishing the API call function, we were getting an “Access-Control-Allow-Origin” error

- On JASPAR’s documentation page for their API, they say that they support cross-origin.
- This led us to think that we were calling their API wrong.

## CORS requests

JASPAR API also support [Cross-Origin Resource Sharing \(CORS\)](#), which enable users to make cross-origin API requests directly from their web application.

- The easiest solution was to implement a relay controller that would run with GRNsight and control all of the calls to the JASPAR database.



**This function takes the JASPAR response and pulls from it the relevant information that we would like to use for GRNsight.**

```
var parseJaspar = function (data) {  
  ...return {  
    .....jasparID : data.matrix_id, // string  
    .....class: data.class, // string  
    .....family: data.family, // array  
    .....sequenceLogo: data.sequence_logo, // string: URL to image  
    .....frequencyMatrix: data.pfm, // object with keys ACIG, each key mapping to an array of ints  
    .....};  
};
```

```
▼ {jaspar: {...}, ncbi: {...}, emsembl: {...}, uniprot: {...}, sgd: {...}} ⓘ  
  ► emsembl: {ensemblID: "YGL073W", description: "Trimeric heat shock transcr:  
  ▼ jaspar:  
    ► class: ["Heat shock factors"]  
    ► family: ["HSF factors"]  
    ► frequencyMatrix: {A: Array(8), C: Array(8), T: Array(8), G: Array(8)}  
      jasparID: "MA0319.1"  
      sequenceLogo: "http://jaspar.genereg.net/static/logos/svg/MA0319.1.svg"  
    ► __proto__: Object  
  ► ncbi: {ncbiID: "etc.", locusTag: "<OtherAliases>YGL073W", alsoKnownAs: Ari  
  ► sgd: {sgdID: "S000003041", standardName: "HSF1", systematicName: "YGL073W"  
  ► uniprot: {uniprotID: "<name xmlns='http://uniprot.org/uniprot'>HSF_YEAST<
```

**Our team, along with the *Gene hAPI* team, wrote the function that returns *this object*, which is then passed on to the page design and integration teams.**

# Outline

- GRNsight maps the relationships between transcription factors of gene profiles constructed from microarray data
- *JASPAR the Friendly Ghost* team members worked simultaneously on separate but related tasks
- Relationships were visualized between the most significant transcription factors in Profile 22 of dZAP1
- JASPAR provided integral information regarding transcription factors for the added feature to GRNsight
- **Future work includes ensuring robust functionality and adding to the information available for the genes on GRNsight**

# Future work will include additional testing and modification and additions to retrieved data

- **Isolate the code from the server using *Sinon***
  - Mock response
- **Modify the code to check for any servers that are not responding**
  - If one is not responding, the code should mark that database's result as undefined.
- **Identify unretrievable data fields**
  - Find workaround to retrieve these fields
  - Find other database to retrieve this same information



# Conclusion

- **GRNsight visualizes a system of relationships between transcription factors.**
  - **Which can be used to better understand cold shock in yeast**
- ***JASPAR the Friendly Ghost* team members both utilized and enhanced GRNsight functionality to analyze genomic data.**
- **GRNsight visualization of the significant transcription factors in Profile 22 of dZAP1 provided further insight on the effect of cold shock on yeast.**
- **JASPAR information was retrieved as part of a larger `getGeneInformation()` function.**
- **Future work will include further testing and the addition of unretrievable desired information available for genes.**

# Acknowledgements

- Thank you to both Dr. Dahlquist and Dr. Dionisio for teaching this course and providing support throughout the semester.
- The JASPAR The Friendly Ghost team for the collaborative effort
- The Departments of Computer Science and Biology at LMU
- The class for being a wonderful audience

# References

- About GRNsight. (2017, June 2). *GRNsight*. Retrieved on December 7, 2017, from <http://dondi.github.io/GRNsight/about.html>
- About JASPAR. (2017). *JASPAR*. Retrieved on December 7, 2017, from <http://jaspar.genereg.net/about/>
- Anonymous. (2017, February 15). */g/ - Technology*. *Warosu.org*. Retrieved on December 11, 2017, from <https://warosu.org/g/image/VIAIA6TQYmgswrNunM19HQ>
- Becerra, M., Lombardia, L.J., Gonzalez-Siso, M.I., Rodriguez-Belmonte, E., Hauser, N.C., & Cerdan, M.E. (2003). Genome-wide analysis of the yeast transcriptome upon heat and cold shock. *Comparative and Functional Genomics*, 4(4), 366-375. doi: 10.1002/cfg.301
- File:Npm-logo.svg (2017, February 27). *WikiMedia Commons*. Retrieved on December 11, 2017, from <https://commons.wikimedia.org/wiki/File:Npm-logo.svg>

# References continued

- **Gene Ontology Consortium. (2017). The Gene Ontology. Retrieved November 29, 2017, from <http://geneontology.org>**
- **LMU BioDB 2017. (2017). Week 15. Retrieved December 05, 2017, from [https://xmlpipedb.cs.lmu.edu/biodb/fall2017/index.php/Week\\_15](https://xmlpipedb.cs.lmu.edu/biodb/fall2017/index.php/Week_15)**
- **LMU BioDB 2017. (2017). *JASPAR the Friendly Ghost*. Retrieved December 05, 2017, from [https://xmlpipedb.cs.lmu.edu/biodb/fall2017/index.php/JASPAR\\_the\\_Friendly\\_Ghost](https://xmlpipedb.cs.lmu.edu/biodb/fall2017/index.php/JASPAR_the_Friendly_Ghost)**
- **Nature. (2014). Scientists Can Study an Organism's Entire Genome with Microarray Analysis. Retrieved on December 9, 2017, <https://www.nature.com/scitable/topicpage/scientists-can-study-an-organism-s-entire-6526266>**



# References continued

- **Short Time-series Expression Miner (STEM). (2006). Retrieved November 29, 2017, from <http://www.cs.cmu.edu/~jernst/stem/>**
- ***Sinon.JS.* (2017). Retrieved on December 11, 2017 from <http://sinonjs.org/>**
- **Teixeira, M. C., Monteiro, P. T., Guerreiro, J. F., Gonçalves, J. P., Mira, N. P., dos Santos, S. C., ... & Madeira, S. C. (2013). The YEASTRACT database: an upgraded information system for the analysis of gene and genomic transcription regulation in *Saccharomyces cerevisiae*. *Nucleic acids research*, 42(D1), D161-D166.**
- **Tyranuer, E. (2017, October 23). *WT\_p-value\_slide\_ET* [PPTX]. Los Angeles: Emma Trynauer.**