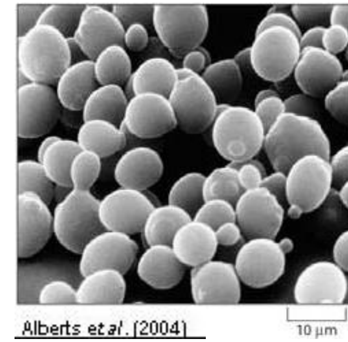
Modeling gene expression regulation using cold shock response in *Saccharomyces cerevisiae*

Elizabeth Polidan
Department of Mathematics
Loyola Marymount University

Studying Response to Temperature Shock is a Good Way to Explore Gene Expression

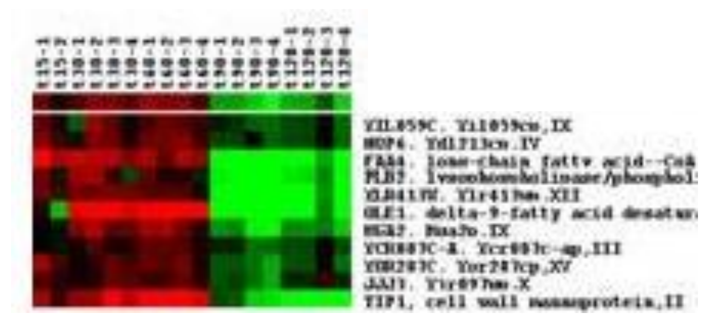


Alberts et al. (2004)

- *Saccharomyces cerevisiae* is subject to temperature extremes in both natural and artificial environments (Aguilera, et al., 2007)
  - While heat shock as been well studied, little is known about cold shock, especially how gene expression is regulated as yeast respond to the sudden extreme change.
- *S. cerevisiae* was cold shocked from 30^{o}C to 13^{o}C. DNA microarray data was taken and analyzed. The gene expression data was used to construct a model to estimate the relative contribution of select transcription factors (TFs) on the up or down regulation of genes in the first hour (early response).
  - This data was taken on five strains. This presentation focuses on the data from a wild type strain.

Using DNA Microarrays to obtain Gene Expression Information

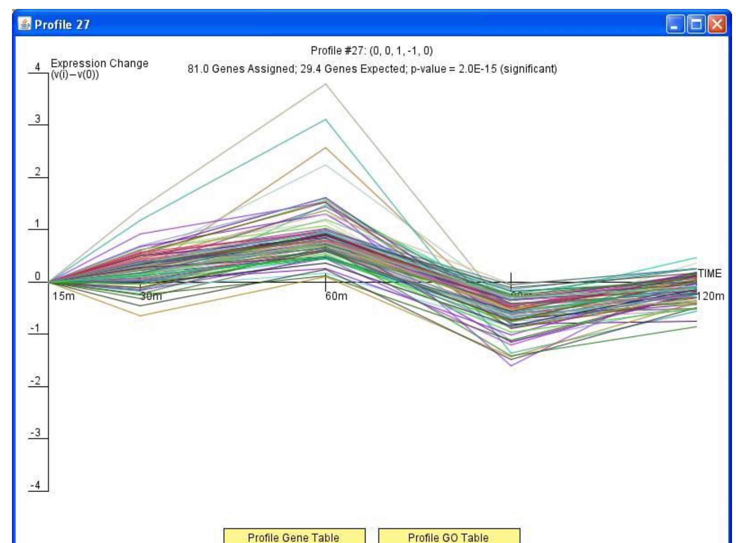
- Baseline DNA microarray data was obtained at 30^{o} C, then replicates of microarray data were obtained at a number of time steps after reducing the temperature to 13^{o} C. This data was processed to determine change in gene expression between time steps.
  - Ratios of red/green fluorescence (expression increase/decrease) were determined, then converted to $\log_2$.



- The data was scaled and centered, then statistical analysis was performed
  - Average Log_2 fold change and standard deviation were determined.
  - T-tests were performed and the p-values were used to understand the likelihood that the expression changes in each gene are real changes.

Cluster Genes Based on Similar Expression Profiles

- Genes with similar expression profiles are likely to be encoded for similar functions
- Short Time Series Expression Miner (STEM) software performs analysis and assigns genes to expression profiles (Ernst and Bar-Joseph, 2006).
  - p-values give likelihood that the genes belong in the cluster



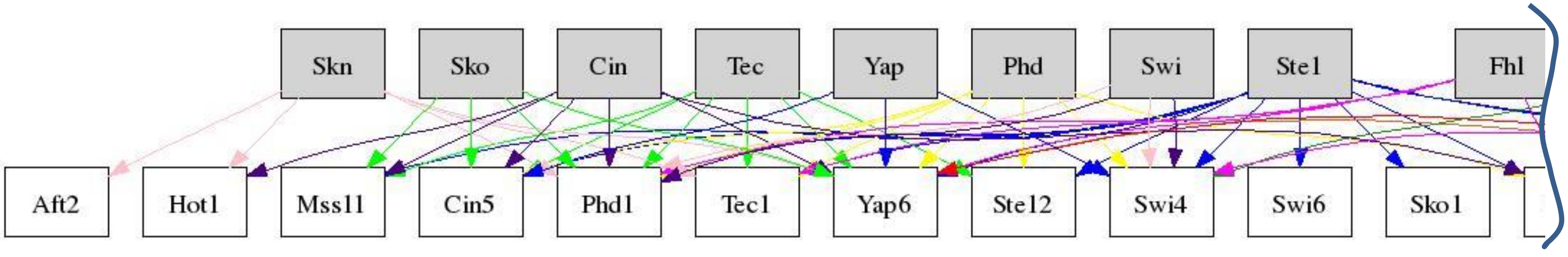
Gene Ontology Analysis is Also Performed Using STEM

- Gene ontology clusters provides a picture of the cellular components, molecular functions, and biological processes genes in these clusters are involved in. (http://www.geneontology.org/GO.doc.shtml)
- p-values are assigned to determine the likelihood that these genes are in these ontological groups by chance vs. really belonging there.

GO Results for Profile 27 based on the actual number of genes assigned to the profile (0.0,0.0,1.0,-1.0,0.0)								
Category ID	**Category Name**	**#Genes Category**	**#Genes Assigned**	**#Genes Expected**	**#Genes Enriched**	**p-value**	**Corrected p-value**	**Fold**
GO:0000469	cleavage involved in rRNA processing	65	5	0.9	4.1	1.50E-03	0.288	5.9
GO:0090501	RNA phosphodiester bond hydrolysis	66	5	0.9	4.1	1.60E-03	0.298	5.8
GO:0016071	mRNA metabolic process	277	10	3.6	6.4	3.10E-03	0.452	2.8
GO:0005730	nucleolus	278	10	3.6	6.4	3.20E-03	0.466	2.7
GO:0000466	maturation of 5.8S rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	77	5	1	4	3.20E-03	0.466	5
GO:0000460	maturation of 5.8S rRNA	78	5	1	4	3.40E-03	0.476	4.9
GO:0006417	regulation of translation	113	6	1.5	4.5	3.50E-03	0.478	4.1
GO:0030686	90S preribosome	89	5	1.2	3.8	6.00E-03	0.606	4.3
GO:0000956	nuclear-transcribed mRNA catabolic process	91	5	1.2	3.8	6.60E-03	0.634	4.2
GO:0010608	posttranscriptional regulation of gene expression	131	6	1.7	4.3	7.10E-03	0.652	3.5
GO:0006402	mRNA catabolic process	93	5	1.2	3.8	7.20E-03	0.66	4.1
GO:0006396	RNA processing	522	14	6.8	7.2	7.20E-03	0.662	2
GO:0030684	preribosome	145	6	1.9	4.1	0.01	0.756	3.2
GO:0006401	RNA catabolic process	111	5	1.5	3.5	0.01	0.822	3.4
GO:0016072	rRNA metabolic process	256	8	3.4	4.6	0.02	0.868	2.4
GO:0006820	anion transport	129	5	1.7	3.3	0.03	0.92	3
GO:0090305	nucleic acid phosphodiester bond hydrolysis	180	6	2.4	3.6	0.03	0.944	2.5
GO:0034470	ncRNA processing	347	9	4.5	4.5	0.04	0.962	2
GO:0034660	ncRNA metabolic process	404	10	5.3	4.7	0.04	0.962	1.9

YEASTRACT Provides Transcription Factors for the Cluster of Genes

- YEASTRACT produced a list of transcription factors. Of the top 10 factors, I chose three to add to the network of 21 genes Drs. Dahlquist and Fitpatrick have developed. They were chosen because they regulate the highest percentage of genes in the network.
  - RAP1
  - STE12
  - TEC1
- YEASTRACT also produced a regulation network. Upon examination, this network contained essentially the same information as the network Dr. Dahlquist created.



Two approaches were used to construct the models

- Sigmoidal Model (based on Vu and Vohradsky, 2006)

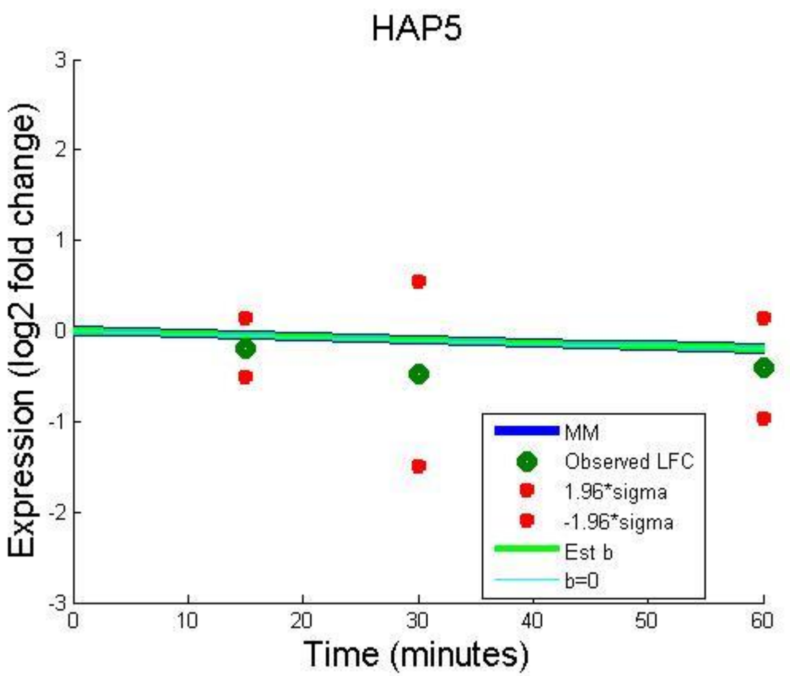
$$\frac{dx_i}{dt} = \frac{P_i}{1 + e^{-\sum\limits_{\mathrm{j}} w_{ij} x_j \, b_i}} - \lambda_i x_i$$

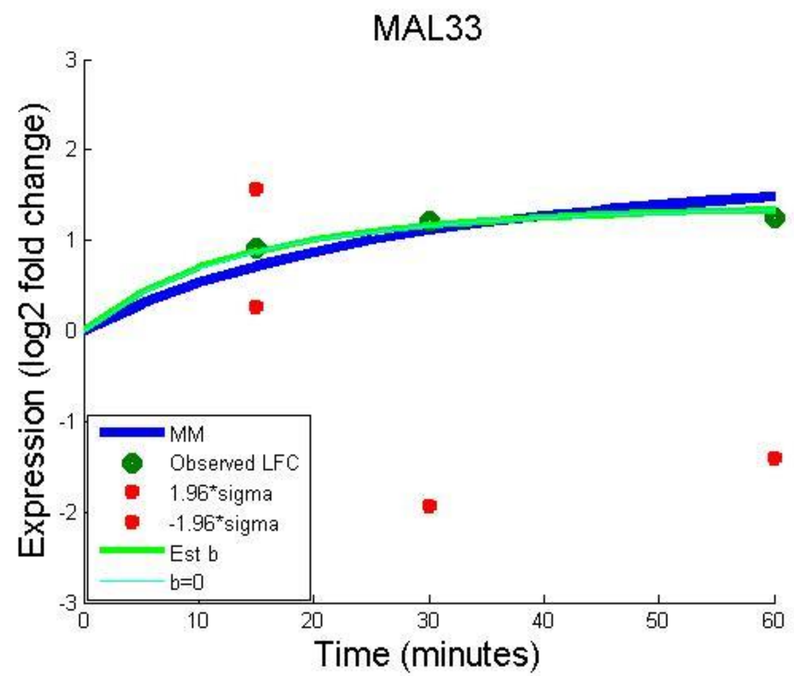
  - where
    P is production rate, w is weight of other genes, b is activation threshold, and λ is degradation rate
    The state variable, x, is concentration
  - "Second" Model: hold b=0 to compare with Michaelis-Menten equation
- Michaelis-Menten

$$\frac{dx_i}{dt} = P_i \left[\sum_j \frac{w_{ij}}{w} \left(\frac{|w_{ij}| x_{ij}}{1 + |w_{ij}| x_j} \right) \right] I_i - \lambda_i x_i, \qquad w = \sum |w_{ij}|$$

Comparison of Three Models Reveals Small Differences

- For the most part there was agreement between the three models
- For some data the sigmoidal fit was better
- There was little difference if estimating b instead of holding b=0.
  - Therefore I chose to use data with estimated b.





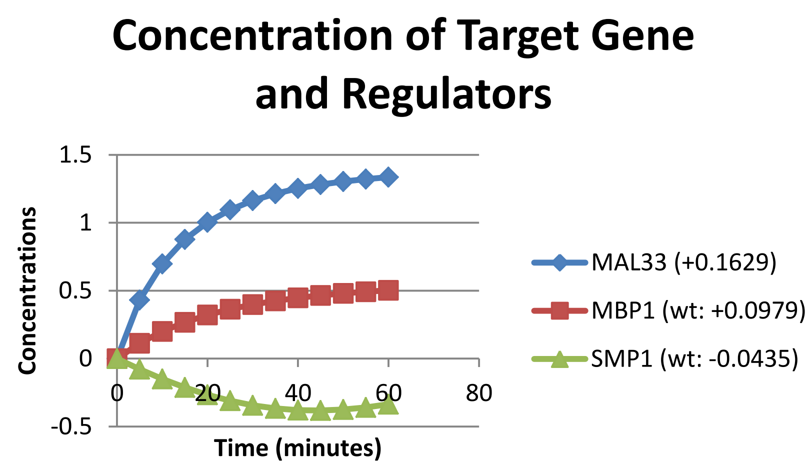
The Addition of RAP1, STE12, TEC1 to Main Set of Genes Affected Weights

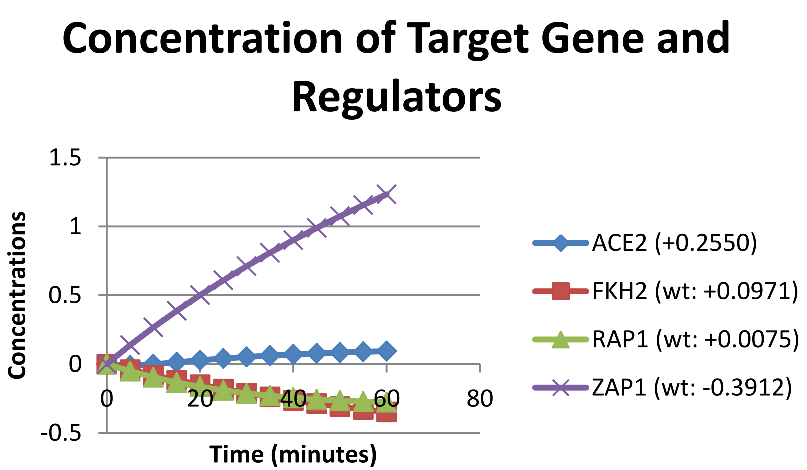
Controller	Target	Weights Main Set	Weights Main +3	Controller	Target	Weights Main Set	Weights Main +3
CIN5	HOT1	+7.98E-06	+0.1817	PHD1	SMP1	-0.0011	+0.0058
CIN5	SMP1	-0.1252	+0.3868	SKN7	HOT1	+0.2679	-0.0206
CIN5	YAP6	+1.74E-05	-0.0006	SKN7	PHD1	+0.1307	-0.0347
FHL1	FKH2	+0.0237	-0.1308	SKN7	YAP6	-9.0E-05	+0.8024
FHL1	HMO1	+0.6772	-0.0528	SKO1	PHD1	+0.0816	-0.0146
FLH1	PHD1	-7.0E-05	+0.0543	SMP1	MAL33	-0.1851	+0.0576
FHL1	SMP1	-0.0172	+0.0863	SMP1	MGA2	-0.0014	+0.0626
FKH2	ACE2	+0.0861	+0.6956	SWI4	PHD1	-0.0305	+0.3454
FKH2	FKH1	+0.0172	+0.5142	SWI6	PHD1	+0.0436	-0.8003
MAL33	SWI4	+0.2147	-0.0005	SWI6	SWI4	-0.0162	-0.7681
PHD1	CIN5	+0.0638	-0.0002	YAP6	SWI4	+0.0252	-0.0196
PHD1	PHD1	-0.0045	+0.3668	YAP6	YAP6	-0.0551	+0.1087

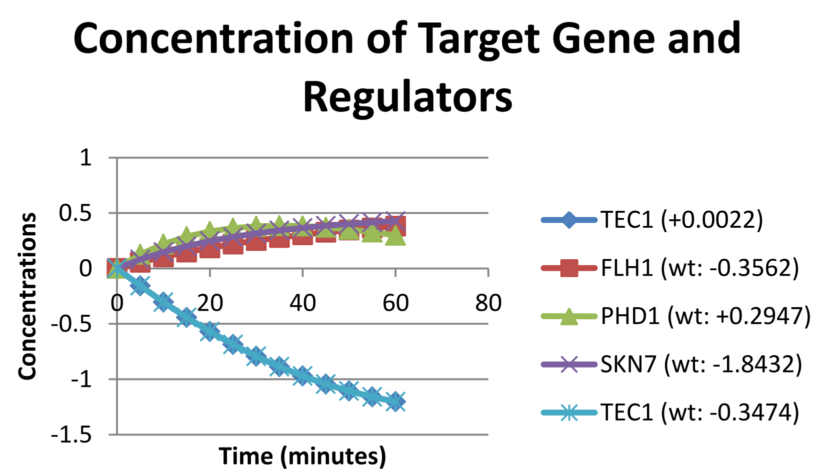
Cannot calculate a total "effective" weight

- Tested to see if the weights can be combined to one "effective" weight
  - Applied weights to model concentrations, per sigmoidal model calculation

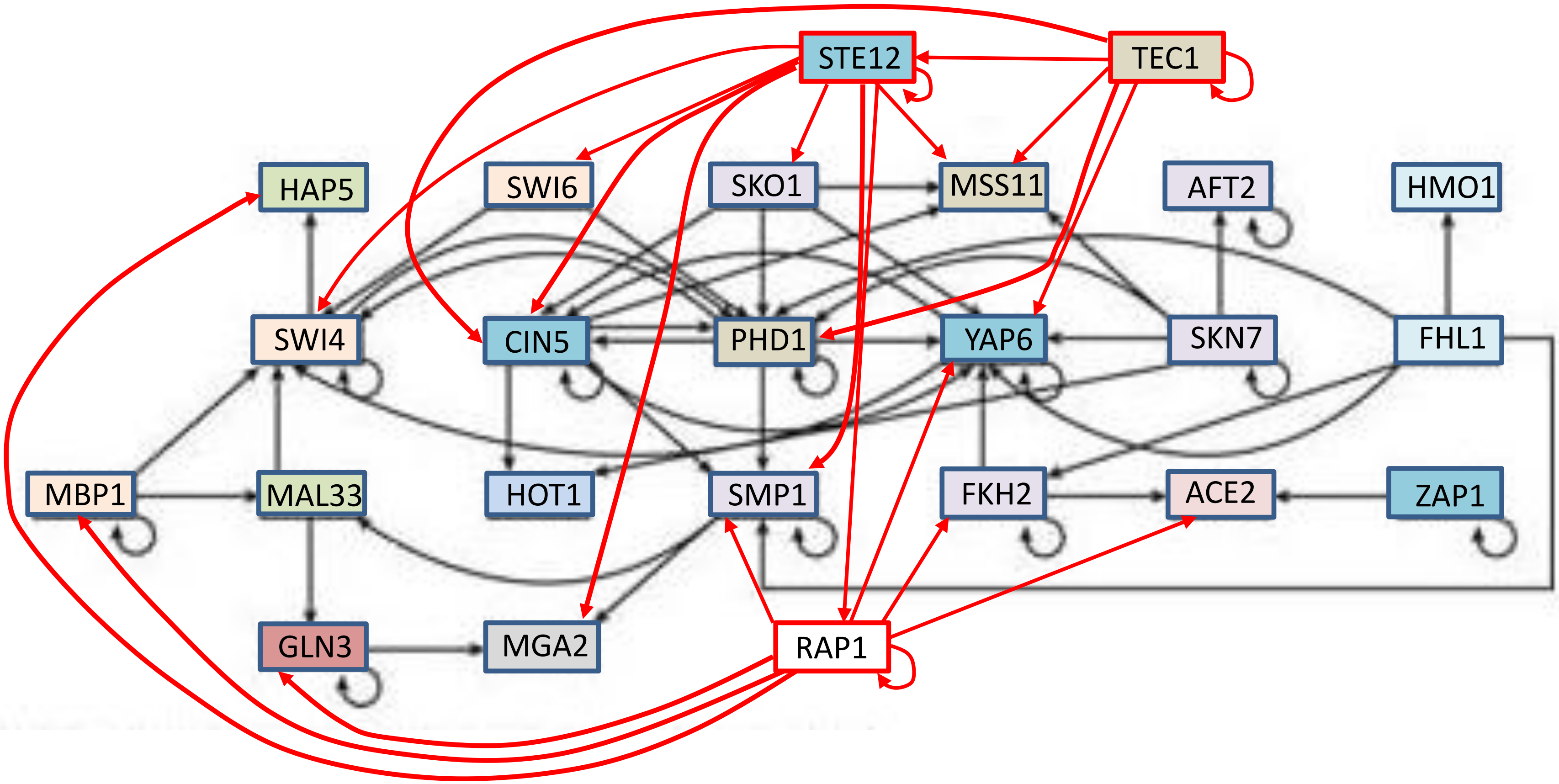
$$\frac{P_i}{1 + e^{-\sum_{j} w_{ij} x_j b_i}}$$







Regulation Network with Added Genes



Fatty acid desaturation
TF for Glycerol Biosynthesis Genes
Septum Destruction > Cytokinesis
Regulates Ribosomal Protein Expression
Activator of genes regulated by NCR
Stress Response
Invasive/Pseudohyphal Growth
Osmotic/Oxidative Stress Reponses
Fermentation regulator
Regulates Meiotic Gene Expression

Further Research Must be Done

- Continued exploration of the elements of the network and their relative influences could be ongoing and ongoing….
- The addition of RAP1, STE12, and TEC1 added transcription control to MBP1, SK01, and SWI6.
  - One gene, FLH1, remains without controller.
  - Would like to explore other regulation mechanisms
    - e.g., Zinc regulation of ZAP1 (Wu, et.al., 2008)
    - Where is the “go” button?
- Further exploration of sigmoidal vs. Michaelis Menten models is important.
- Automation of this process would be convenient and not to difficult to achieve in Matlab.

Model Constructed Using Yeast Cold Shock Data Does Provide Relative Contributions from TFs

- The large numbers of genes in a genome provides statistically significant information in a series with few time steps.
- Powerful tools exist to explore hierarchical structures and clustering of expression profiles.
- We are on our way to a model that provides relative regulation contributions of different transcription factors.
- Examination of the weights and concentrations of each individual gene and its regulators is necessary – there is no trick.

Acknowledgements

Thank you to Dr. Fitpatrick and Dr. Dahlquist for all of their guidance.

References

Aguilera, J., Randez-Gil, F., and Prieto, J.A. (2007) Cold response in *Saccharomyces cerevisiae*: new functions for old mechanisms, *FEMS Microbiol. Rev.*, **31**, pp. 327-341.

Ernst, J. and Bar-Joseph, Z. (2006) STEM: a tool for the analysis of short time series gene expression data, *BMC Bioinformatics*, **7**:191.

The Gene Ontology, http://www.geneontology.org... /GO.doc.shtml, accessed 08 May 2013.

Vu, T.T. and Vohradsky, J. (2006) Nonlinear differential equation model for quanitification of transcriptional regulation applied to microarray data of *Saccharomyces cerevisiae*, *Nucleic Acid Research*, **35**:1, pp. 279-287.

Wu, C.-Y., Bird, A.J., Chung, L.M., Newton, M.A., Winge, D.R., and Eide,D.J. (2008) Differential control of ZAP1-regulated genes in response to zinc deficiency in *Saccharomyces cerevisiae*, *BMC Genomics*, **9**:370.