UNIVERSITY OF CALIFORNIA, SAN DIEGO

Dissecting protein dynamics in *Saccharomyces cerevisiae*

A dissertation submitted in partial satisfaction of the requirements for the degree in Doctor of Philosophy

in

Biology

by

Sidney Kuo

Committee in charge:

Professor Scott A. Rifkin, Chair
Professor Lin Chao, Co-Chair
Professor Nan Hao
Professor Richard Kolodner
Professor Maho Niwa

2017

The Dissertation of Sidney Kuo is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____
Co-Chair

_____
Chair

University of California, San Diego

2017

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

Chapter 1 is currently being prepared for submission for publication of the material. Kuo, Sidney; Egertson, Jarrett; MacCoss, Michael; Chevrier, Nicolas; Pollard, Dan; Rifkin, Scott. The dissertation author was the primary investigator and author of this material.

Chapter 2 is currently being prepared for submission for publication of the material. Kuo, Sidney; Schwartz, Ruth; Rifkin, Scott. The dissertation author was the primary investigator and author of this material.

VITA

2010 Bachelor of Science, Duke University

2017 Doctor of Philosophy, University of California, San Diego

FIELDS OF STUDY

Major Field: Biology

Studies in Evolutionary Genetics
Professor Scott A. Rifkin

ABSTRACT OF THE DISSERTATION

Dissecting protein dynamics in Saccharomyces cerevisiae

by

Sidney Kuo

Doctor of Philosophy in Biology

University of California, San Diego, 2017

Professor Scott A. Rifkin, Chair
Professor Lin Chao, Co-Chair

Protein dynamics during cellular responses to extracellular stimuli has been widely studied to understand how cells respond to change. Here, we employ two distinct techniques to look at two separate aspects of cellular response in *Saccharomyces cerevisiae*. First, we use a transcriptomics and proteomics approach to examine the pheromone response of yeast, using RNA-sequencing and mass spectrometry, respectively. We find that in general, mRNA

abundance is not sufficient as a predictor of protein levels during dynamic responses such as pheromone response, but by using a mass-action kinetics model, we can reconcile these discrepancies. Performing this experiment in both a lab strain and the clinical isolate, YJM145, allows us to further identify possible genes where post-transcriptional regulation is affected by polymorphisms, allowing possible mechanisms of post-transcriptional regulation to be further explored. Second, we used an extreme bulk segregant analysis approach to map alleles in wild yeast strains, L-1374 and YPS606, which cause a difference in molecular response during the yeast's response to hyper-osmotic shock by tracking the protein, STL1p. We find that many alleles have transient effects on STL1p levels during the osmotic stress response, but only a few have an impact on the cellular state after adaptation to the new environment. In addition, we found that alleles in the wild strain affect STL1p levels in both directions, suggesting possible compensatory mutations. Most genes were not found to be in the known osmotic stress response pathway suggesting regulation mechanisms of STL1p that were previously unknown.

INTRODUCTION

The dynamics of protein regulation has been a widely studied area in biology, as it is essential for cells to respond to external environment changes. As cells experience a host of different environments, cells must be able to adapt and react in order to maximize their survivability and growth potential. Protein regulation has been studied in numerous ways, from studying the precursor mRNA regulatory changes to measuring the proteins themselves. Most studies in the past, however, have focused on steady-state mRNA or protein changes, since abundance measurements are noisy, labor-intensive, and/or expensive. With the advancement of new techniques, however, we set out to study protein regulatory dynamics in the model organism, *Saccharomyces cerevisiae*, while also considering the temporal aspects of protein regulatory dynamics. The two pathways we have chosen to focus in *S. cerevisiae* are two mitogen-activated protein kinase (MAPK) pathways. MAPK pathways are found in most eukaryotes, from fungi such as yeast to mammals such as humans. In yeast, MAPK pathways are responsible for certain vital cellular responses. For our studies, we will focus on two of those responses, pheromone response and hyperosmotic stress response.

During the pheromone response of yeast, when a haploid yeast comes into contact with another haploid yeast of the opposite mating type, the corresponding MAPK pathway will activate which causes the yeast to undergo cellular and morphological changes in order to mate with the other cell, forming a single diploid cell[1]. For most observed strains of *S. cerevisiae*, those that can sporulate to form haploid yeast cells have been observed to mate[2]. While mating may not be an essential cellular function, it allows yeast to undergo sexual reproduction, which is thought to enhance fitness by increasing genetic diversity[3]. The response of mating type **a** yeast cells to the mating pheromone, alpha-factor, is widely studied

at both genome-wide level and a more detailed, molecular response level[1,4,5]. Using this system as a model for other regulatory systems, we set out to address two key questions. The first question is whether the mRNA abundances measured correspond to the protein abundances, during the dynamic cellular response to alpha-factor. The second is what kind of impact do natural genetic variants found in wild yeast strains have this mRNA-protein relationship. To address these questions, we combine quantitative transcriptomics, via RNA-sequencing (RNA-seq), with quantitative proteomics, via quantitative mass spectrometry, and perform these experiments on a lab strain and a clinical yeast isolate for a period of 8 hours after pheromone induction.

For our second model system, the response to hyperosmotic stress in yeast, we studied more closely the impact of naturally found genetic variants on that particular pathway. The hyperosmotic stress response pathway in yeast is an essential pathway for the cell. When cells are subject to an environment with high osmolarity, it must undergo changes in order to maintain turgor pressure to prevent the cell wall from bursting. To react to such an environmental change, cells have sensory proteins on the cell surface, and they signal downstream proteins when a change in external osmolarity is detected, which eventually converges on the MAPK HOG1[6]. HOG1 activates a host of downstream cellular responses, including inducing transcription of the protein Stl1p[7,8]. We adopted a quantitative trait locus (QTL) mapping approach to evaluate the effects of genetic variants in two wild strains during the Stl1p response to hyperosmotic stress. Though several other protein-QTL (pQTL) studies have been performed before, we do our experiment across three time points in order to find loci that have temporary as well as persistent effects, and which of those loci overlap. This allows us to also evaluate what proportion of polymorphisms cause protein expression differences to be buffered by downstream components.

Overall, our experiments focus on exploring the regulatory dynamics of protein expression in the context of natural genetic variation. We address some gaps in our understanding of the dynamics of how protein is translated from mRNA and subsequently regulated via protein degradation. In addition, we assess the possible effects of naturally occurring genetic variation and how they can affect those dynamics. Together, these aid in our understanding of protein regulation, and they add a small part to the larger picture of how these regulatory pathways are subject to mutation and selection.

CHAPTER 1: PROTEIN AND MRNA DYNAMICS DURING THE YEAST PHEROMONE RESPONSE

INTRODUCTION

Controlling protein levels is an essential task of cells and takes on particular importance when they need to muster new proteins to respond to external stimuli. The yeast mating response is an example of such a system that has been extensively studied. In *Saccharomyces cerevisiae*, when mating type **a** cells come into contact with alpha-factor, cells activate a host of downstream responses including cell-cycle arrest, inducing or activating a number of proteins, and elongating to form a shmoo, in preparation for mating with another cell of the opposite mating type[1]. During this process, many proteins are activated in order to alter the physiology and morphology of the yeast. Many previous studies have focused on the transcriptional response when yeast has been exposed to alpha-factor[4,5]. However, even though the transcriptional response may be well-studied, it is unclear how protein levels, which ultimately affect cell physiology and morphology, are affected by transcriptional changes during pheromone response and other similar pathways[9–17]. In addition, most studies focus solely on the lab strain, S288c. Natural genetic variation that affect protein regulation is important as it is known to be associated many phenomena such as ecological adaption[18]. Thus, we would like to study how natural genetic variation influences post-transcriptional regulation of cellular pathways such as pheromone response.

Recently, many methods have been developed and refined in order to study mRNA or protein on the transcriptome or proteome level, respectively. For transcriptomics, RNA

sequencing (RNA-Seq) is currently the predominant method. It can detect even low level transcripts within cells, and it also allows for quantification of transcripts within cells[19,20]. Proteomic methods, on the other hand, have had slower development. However, recent developments in quantitative mass spectrometry (MS) have allowed researchers to more reproducibly quantify levels of protein in cellular samples[21,22]. Together, these tools allow us to investigate the genome-wide response of yeast to mating pheromone for both mRNA and protein levels simultaneously.

The relationship between mRNA and protein is not static, on either evolutionary or cellular timescales. On the cellular level, the connection between protein levels and mRNA levels during pheromone response in yeast has been shown to be weak. Post-transcriptionally, protein levels may be impacted by translation rates and degradation rates. These rates are not generally known for most genes which is why it is seemingly difficult to predict protein levels from mRNA[23]. On the evolutionary level, natural genetic variation adds another layer of complexity by affecting protein expression. Previous studies have examined how natural variation impacts mRNA transcription dynamics such as transcription rates and mRNA degradation rates, which all affect the abundance of the final protein product[24,25]. Thus, we would like to characterize the relationship between RNA abundance and protein abundance on a genome-wide level, and survey the differences in that relationship between two strains in order to understand the amount of variation that can exist between strains.

Here, we seek to address these questions using by performing a time course experiment on two strains, S288c and YJM145, during pheromone response, profiling the transcriptome via RNA-Seq and profiling the proteome via quantitative MS. YJM145, also

known as YJM789, has a high quality genome sequence available[26], and some of its phenotypic differences have been studied and mapped to genetic variants[27]. In addition, its transcriptional profile during pheromone response has also been found to have several significant differences compared with the lab strain[4]. Thus, by obtaining quantitative data during pheromone response, we can determine the relationship between protein levels and their corresponding mRNA levels across the proteome. Performing these experiments in two separate strains allows us to survey how natural genetic variation affects protein dynamics. With follow up studies in the future, this may elucidate previously-unknown post-transcriptional regulatory mechanisms, as well as shed insight into how these cellular pathways are subject to the forces of evolution.

METHODS

STRAINS AND EXPERIMENTAL CONDITIONS

The two strains used in this study were S288c and YJM145. Both are mating type **a** with the following genes knocked out via the loxP system[28]: *URA3, HO, AMN1,* and *BAR1***.** In both strains, *FIG1* had been tagged with yECitrine using the Ca.URA3 marker for other studies. Strains were grown in 50 mL YPD overnight as a starter culture, with the experimental culture being in 1L YPD in a 3L Fernbach flask to provide proper aeration. Experimental cultures were diluted to an OD of ~0.1 and incubated for 2 hours before inducing with pheromone. Pheromone was added to achieve a final concentration of 50 nM in the media. Samples taken for the mRNA were at hours 0, 0.5, 1, 1.6, 2.4, 3.2, 4.8, 6.4, 8 where hour 0 is taken immediately before induction. Samples for the protein MS were taken from hours 0, 1.6, 3.2, 4.8, 6.4, 8. Fifty mL of each sample was taken from the flask at each

time point, spun down, and flash frozen in liquid nitrogen. Experiments were performed on three separate days with the RNA and protein handled together afterwards.

Total RNA extraction of all the samples was done using the Ambion RiboPure kit. RNA-Seq preparation was using a custom protocol developed by Nicholas Chevrier and adapted by Sergey Kryazhimskiy. The protocol produces mRNA reads proportional to the number of transcripts independent of transcript length. Sequencing was done on an Illumina HiSeq4000 using single reads 50 base pairs in length. Quantitative MS was performed by the MacCoss laboratory using a data-independent LC-MS/MS method, and processing of raw MS data was done in Skyline[21].

PROCESSING AND STATISTICAL METHODS

Processing of RNA-Seq data was done on Galaxy, using BWA to align reads to the *S. cerevisiae* reference genome (R64-2-1, 2015-01-13) and htseq-count to extract read counts for each transcript[29–31]. Estimation of read counts was performed using DESeq2, with the generalized linear model parametrized as *count ~ timepoint + strain + timepoint:strain*, with *timepoint* being a factor variable[19]. Likelihood ratio tests between the strains use the form *count ~ timepoint + strain* as the reduced model, which would not include differences in the pre-induction time point between the strains.

Pre-processing of MS intensity values was done using Skyline, followed by sample normalization and log transformation via MSStats[21,32]. A general linear model using log intensities for each protein that follows the same form as the RNA-seq model was fit to the data, and the likelihood ratio test was performed the same way.

All analysis and modeling was performed in R. Model fitting to the differential equation model was performed using residuals on a log scale as errors are thought to be symmetrical for protein levels in a log scale.

RESULTS

We profiled the transcriptome and proteome for 8 hours after pheromone induction in two strains, S288c and YJM145. Transcriptome profiling was done via RNA-Seq, with data being taken at 9 separate time points in 3 biological replicates for each strain, while proteome profile was done via quantitative MS, with data being taken at 6 separate time points in the same 3 biological replicates. We used DESeq2 to analyze the quantitative RNA-Seq data, which fits a generalized linear model to estimate mRNA counts at each time point[19]. Similarly, for the MS data, we modeled the log-abundances using a general linear model[33]. For RNA-Seq, 6247 different genes had transcripts detected at one or more time points in at least one strain, and 2345 different genes had peptides detected at one or more time points by quantitative mass spectrometry. Using the pre-induction time point as a control, in S288c, we detected 1261 (22.5%) genes to be differentially expressed by RNA-Seq during at least one point in the time course, and for YJM145, we detected 1767 (31.6%) differentially expressed genes. Similarly, we detected 816 (34.8%) and 848 (36.1%) differentially expressed proteins via mass-spec during at least one time point for S288c and YJM145, respectively.

For mRNA expression, out of the set of 1261 and 1767 differentially expressed genes for each strain, we detected a common set of 921 (16.5%) genes that were differentially expressed in both strains. Differential expression patterns are similar, but not identical. After filtering for only genes that had measurable mRNA transcripts at every time point, we calculate the correlation between mRNA expression trajectories between the two strains

across the remaining 5595 genes at 1 hour post-induction to be 0.49 (Figure 1). The median Pearson correlation across these 5595 genes up to the first 2.4 hours is 0.35, suggesting that for many of the genes, the direction of gene expression change is congruent but perhaps not the timing and magnitude for the duration of the time course (Figure 1). Using the likelihood ratio test with our generalized linear models, we can directly test for differences between the two strains. Given the way we have parametrized our linear model, the likelihood ratio test only tests for dynamic differences after the pre-induction time point. We find 219 (4%) genes with dynamic differences with $p<0.05$ using this test. One example gene, SED1 is shown here (Figure 1).

For protein expression, out of the 816 (S288c) and 848 (YJM145) genes that were detected to be significantly different, 519 (22.1%) of them overlap between the two strains. The likelihood ratio test detected 85 (3.6%) genes that have dynamic differences in protein expression. The Pearson correlation between all proteins at 1.6 hours post induction is 0.48 (Figure 2). These figures are in line with what was discovered in the mRNA. Out of 85 genes that have dynamic differences, we found that three pathways are over-represented using gene ontology enrichment analysis. In particular, the chromisate biosynthetic pathway (p=4.2e-2), the arginine biosynthetic pathway (p=3.74e-2), and oxidation reduction process (p=3.98e-3) were found to be significantly over-represented in this list of genes that are dynamically different (Figure 3).

CORRELATIONS BETWEEN MRNA AND PROTEIN LEVELS ARE POOR

In general, the correlations between mRNA levels and protein levels are poor. The median Pearson correlation between mRNA and protein for S288c is 0.0154, and the median Spearman rank correlation was not much higher at 0.05 (Figure 4). The clinical isolate,

YJM145 had median Pearson and Spearman correlations of 0.19 and 0.2, respectively, which is only slightly higher than S288c and still does not suggest a strong correlation. Both sets of correlations are lower than values reported in the literature[17,34] However, most of the literature focuses on steady state correlations between mRNA and protein, and few studies have been conducted on dynamic time courses such as ours. For dynamic time series data, we expect that for genes with active translation, the protein product should integrate the mRNA curve, and, depending on the protein's degradation rate, the protein curve may or may not follow the decline in mRNA. Alterations to cell growth conditions may also lead to active degradation of other proteins.

We evaluated the qualitative changes in mRNA and protein by asking whether, for each differentially expressed mRNA, the directions of change are incongruous with the protein (i.e. whether one is induced and the other repressed or vice versa). To do that, we calculated the area under the log-fold-change curve for both mRNA and protein, normalizing the pre-induction time point to zero and comparing whether the areas are positive or negative; a positive sign represents induction, and a negative sign represents repression. We find that in S288c, 61.9% of proteins follow the change in direction of the mRNA, and for YJM145, 70.3% of proteins follow the change in direction. This suggests that even qualitative conclusions drawn from mRNA abundance levels are not always representative of behavior in protein levels during pheromone response. In a previous study, Zheng, et al. identified several proteins to be targets of STE12, which is known to activate several pheromone response genes[4]. In S288c, among the 122 such target genes that we detected as differentially expressed, congruence between mRNA and protein only occurs 56.5% of the time and is not improved, as suggested by a previous paper[17]. Thus, for dynamic systems such as pheromone response, a single time point is inadequate to predict protein behavior.

A MASS ACTION KINETICS MODEL BRIDGES THE GAP BETWEEN PROTEIN LEVELS AND MRNA MEASUREMENTS

In order to bridge the disconnect we observe between mRNA levels and protein levels, we investigate a mass action model of the form $\frac{dp}{dt} = k_s R(t) - k_D P(t)$, where k_s is the translation rate, *R(t)* is the amount of mRNA at time *t*, k_D is the degradation rate, and *P(t)* is the amount of protein at time *t*. We used our mRNA measurements for the curve *R(t)*, fitting the parameters k_s and k_D by minimizing the sum of squares differences between the predicted log-transformed *P(t)* and our data measured by quantitative mass spectrometry. The log-transformation is appropriate here as the errors in the protein quantification via MS are thought to be normally distributed after log-transformation[33].

In general, our mass action model fit most proteins well. We characterized the fit of our models via the normalized root mean square error and the cosine distance. The median normalized root mean square error for S288c is 0.010 (Figure 5), with 96% percent of genes having a normalized root mean square error below 0.05. This means that, on average, the prediction deviates less than 5% from the mean value of the measured values. The cosine distance, which is a metric similar to the Pearson correlation but less susceptible to noise, can also be used as a metric to estimate the goodness of fit[35]. The median cosine distance, which ranges from 0 for identical curves to 1 for completely anti-correlated curves, is 0.0028 between the predicted protein curves and the experimental ones Figure 5). The fits for YJM145 were slightly better, with a median normalized root mean square error of 0.009, with 96% of genes having a normalized root mean square error below 0.05, and a median cosine distance of 0.0026.

While our model fits well for most genes, we noticed for several of the fits, the value of one of the parameters is zero, suggesting negligible contribution from either the degradation or the production term. For S288c, 7.0% of genes had a k_D of 0, suggesting that degradation is negligible compared to the rate of protein production from mRNA. Conversely, 9.1% had a k_s of 0, which suggests that protein degradation is the dominant mechanism determining the change in protein level. One possible explanation for this might be sub-compartmentalization of mRNA in the cell, so even though the mRNA is present and can be measured by RNA-seq, it is not contributing to protein production via translation. Similarly, for YJM145, 9.4% of genes only need the production term while 9.7% of genes only need the degradation term. The slight increase in number of genes for which the degradation term has a negligible impact on protein level may partially account for the higher RNA-protein correlation reported above.

We estimated the median protein half-life in S288c to be 1.19 hours for S288c and 1.65 hours for YJM145 (Figure 6). One example of our model is shown here for the gene YER155C (

Figure 7). Given our experimental conditions, in S288c, 12.3% of all proteins that we have parameter estimates for seem to be rapidly degraded, having a half-life of under 0.2 hours. On the other hand, 18.1% of proteins are quite stable and have an estimated half-life of over 5 hours. For YJM145, 10.7% experience rapid turnover and 21.5% of are long-lived, respectively. These results share some similarities with what was reported in the literature. Although our predicted protein half-lives were approximately log-normally distributed36, we found many more proteins to be short-lived and a much smaller proportion of long-lived proteins34. This could be the result of MS bias, which tends to only pick up the most

abundant proteins, as well as the fact that pheromone response induces drastic physiological changes, which may lead to active protein degradation or certain proteins compared with steady-state growth conditions.

The rate of protein synthesis and degradation may differ between strains even for proteins of similar abundance and dynamic patterns, perhaps due to altered mRNA dynamics. The mean Pearson correlation between protein trajectories from the two strains is 0.64 with a standard deviation of 0.46 (Figure 8). One question that also arises is whether the translation rates and degradation rates are similar between the two strains. To do that, we look at the genes that have over 90% average difference in parameters, quantified by the absolute value of the difference in each parameter divided by the sum that parameter across both strains. Surprisingly, the mean Pearson correlation only drops modestly to 0.51, with a similar standard deviation of 0.48 (Figure 8). Furthermore, if we look at genes where protein abundances are very similar across both strains, defined by having a Pearson correlation of 0.85 or greater, the average difference in parameters remains similar to the average difference across all genes, only dropping modestly from 73.4%, with a standard deviation of 30%, across all genes to 62% (Figure 8). This suggests that polymorphisms that exist in wild strains may alter mRNA abundance, protein synthesis rates and degradation rates even while protein abundances are relatively similar, which is in line with what was reported previously[34].

FITTING A GLOBAL MODEL DOES NOT GIVE US BETTER PARAMETER ESTIMATES

Since our experimental setup uses measurements of mRNA and protein taken from the same flask throughout the entire time course, done in three replicates, we wanted to

evaluate whether the paired structure gives us extra information with regards to model fitting. Specifically, are the flask-specific mRNA curves able to predict the protein measurements from the corresponding flasks better than the averaged mRNA levels for those same protein trajectories? By considering the experimental design, this question is essentially asking whether the variability in mRNA measurements via RNA-seq overwhelm the biological variability across different flasks on different days.

In order to evaluate this, we fit a global mass action model, identical in form to the two-parameter model mentioned above. Instead of utilizing the average mRNA measurements to form our mRNA curve and generating one predicted protein curve and calculating the residuals to that predicted curve from our three measured protein curves, we generate three separate predicted protein curves from the three mRNA curves, and we calculate the residuals from the corresponding protein measurements which were extracted from cells grown in the same flask. By comparing the residuals obtained from each method, we can evaluate which method produces better results (Figure 9). In S288c, we found that for 27.6% of genes, the global method generated better fits than using the average mRNA, and for YJM145, we found 29.9% of genes had better global fits, quantified by having a lower residual sum of squared errors. This suggests that using the averaged levels of mRNA generally gives more consistent results for most genes. It seems that the technical variability introduced during the sample preparation and sequencing process for RNA-seq obscures any biological variability.

DISCUSSION

Many studies have been performed at the mRNA level looking at dynamic responses of yeast to pheromone response on the transcriptome level. However, it is uncertain how

transcript levels affect protein levels, which, in turn, affect the overall cellular phenotype. In our study, we took a quantitative time course data to characterize cellular change in two different strains in order to investigate to what degree protein levels reflect mRNA levels during the pheromone response process, and how natural genetic variation can affect this process. We show that on the proteome level, there is no correlation between mRNA levels and protein levels, and that even qualitative differences are not good predictors for differences in protein levels. However, with a simple mass-action model, we can reconcile many of these differences and find physiologically plausible parameters that represent the relationship between mRNA and protein. This indicates that time course data is important in understanding how mRNA translates into a phenotype.

This opens up several questions such as what are the factors that affect these parameters. By examining the differences between the two strains, we can conclude that there are likely many genetic factors that may lead to a difference in translation or degradation rates of proteins. However, the mechanistic details of how that occurs, whether genetic polymorphisms affect mRNA localization or half-lives, or whether they affect part of the translation or degradation machinery, remains to be uncovered. The mechanistic details would likely have to be studied on an individual gene basis with other techniques in order to obtain the full molecular picture.

There are inherent limitations to our transcriptome and proteome-wide approach, one of which is the technical variability in quantitative estimates with the two techniques. These can be eventually refined with more data, and genes of interest may be individually examined for more precise measurements. In addition, one inherent bias in the significance calling pipeline is that because of the nature of –omics experiments, while the false discovery rate is

usually controlled for by various methods, the false negative rate, and therefore power, is not known and difficult to estimate with our or other existing data sets. Thus, any conclusions drawn from genes not called as differentially expressed must be made with caution, and in our study, we tried to utilize the estimates of the log-fold changes instead.

Other aspects that we cannot address are changing parameters during the time course and the heterogeneity of cells during a response such as mating response[1,23,37]. Since yeast cells are no longer in steady state upon exposure to pheromone, many of the transcription, translation or degradation rates may change throughout the time course. For a longer time course such as ours, our parameter estimates are sufficient for coarse-grained estimates of the average, but for focusing more precise physiological values, other methods such as ribosome profiling would be necessary to provide estimates of the true rates especially during the immediate response. In addition, new technologies such as single-cell RNA-seq may give allow researchers to discover the extent of variability between individual cells during this type of process.

Regulatory mechanisms are a key part of producing phenotypic differences. With the advent of technology, we are now able to quantitatively characterize an organism at both its transcriptome and proteome level, allowing us to infer regulatory effects for individual genes. The results we present here contribute to our knowledge of how mRNA levels and protein levels are related, and emphasize the importance of taking post-transcriptional regulation into consideration both on a cellular physiology level and on an evolutionary level.

FIGURES

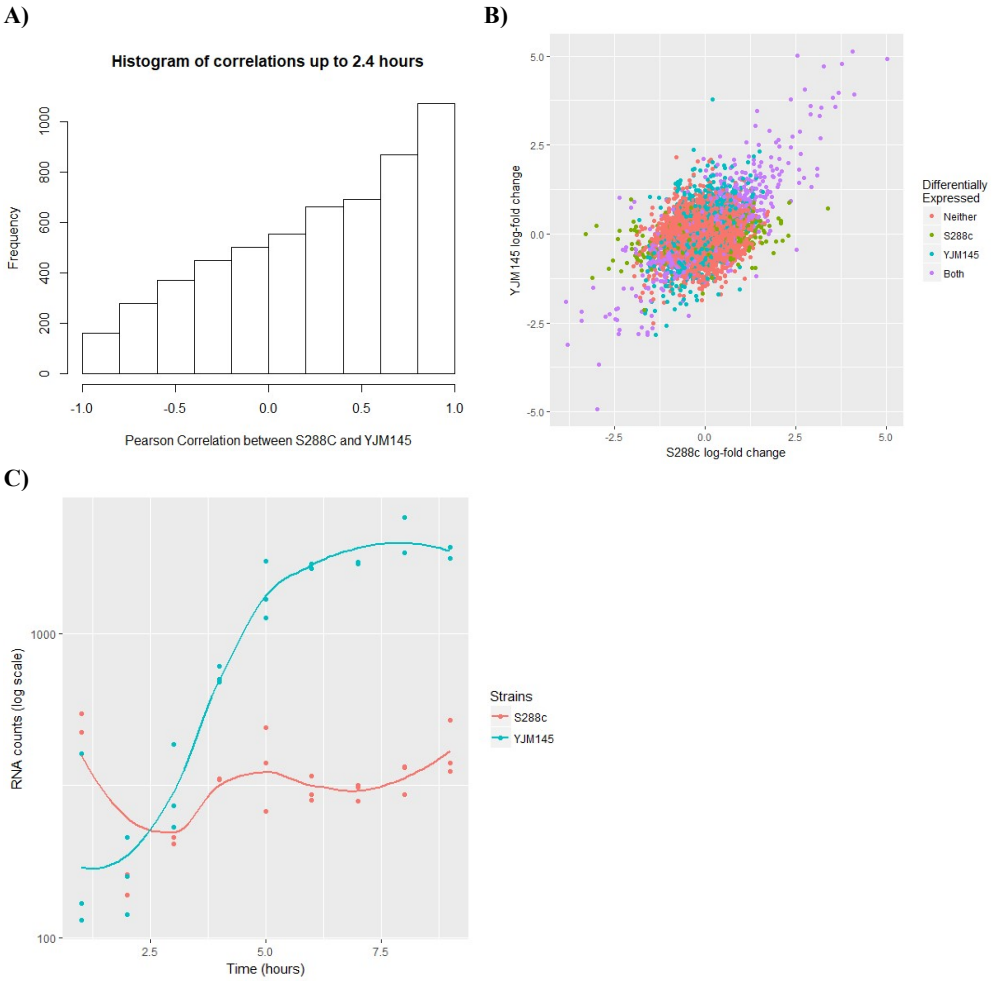


Figure 1: Transcript abundance between the two strains. A) A histogram of Pearson correlations between mRNA expression trajectories in S288c and YJM145 up to the first 2.4 hours. The median correlation is 0.35, suggesting that mRNA behaviors are likely to be directionally similar. B) A scatter plot of the log-fold change value 1 hour post pheromone induction for each strain. Each dot represents an individual gene, and the position reflects its level compared with the pre-induction time point. The color reflects whether it was found to be significantly different from the pre-induction time point. Pearson correlation is 0.49 between the two strains across all genes. C) One example of a highly dissimilar gene in terms of mRNA expression between the two strains. This gene, YDR077W (SED1) was identified by the likelihood ratio test performed using our generalized linear models. YDR077W is a stress induced cell wall glycoprotein.

Figure 2: Scatter plot of protein expression changes at 1.6 hours post pheromone induction. Similar to the mRNA expression plot, each dot represents a single protein and its measured level compared to the pre-induction time point. Colors represent whether each protein was detected to be significantly different from pre-induction. Pearson correlation was calculated to be 0.48 across all proteins between the two strains.

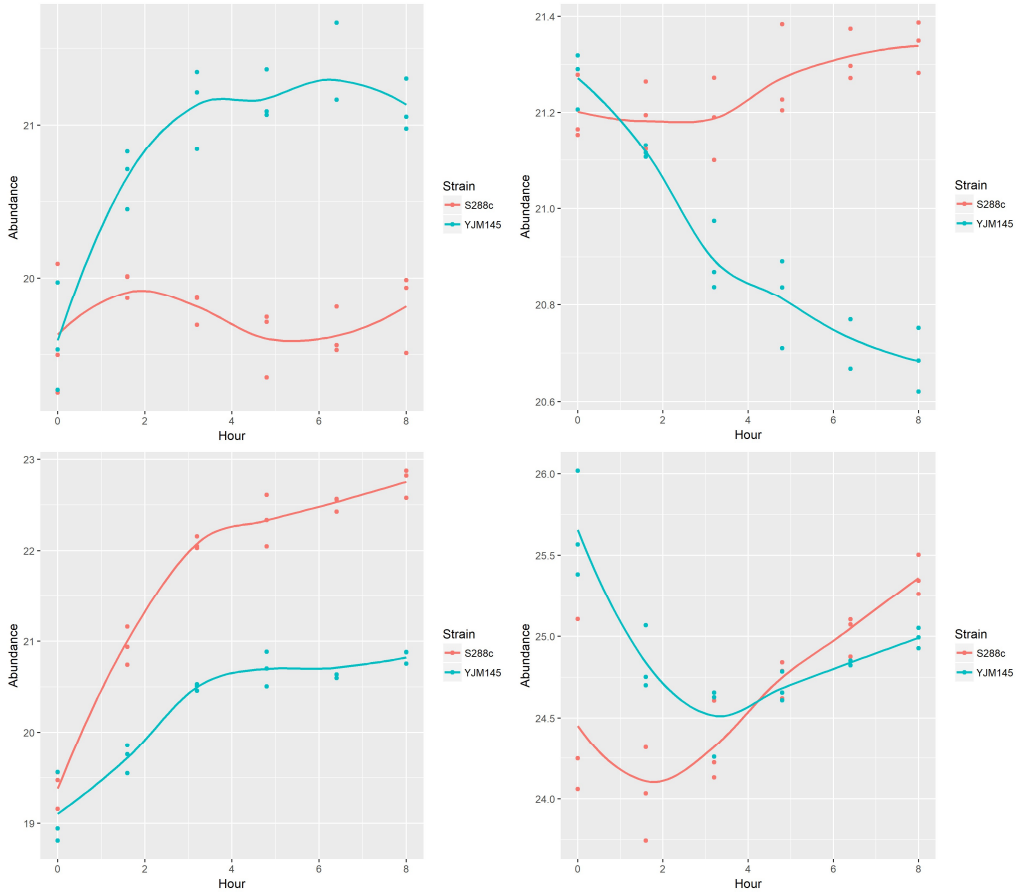
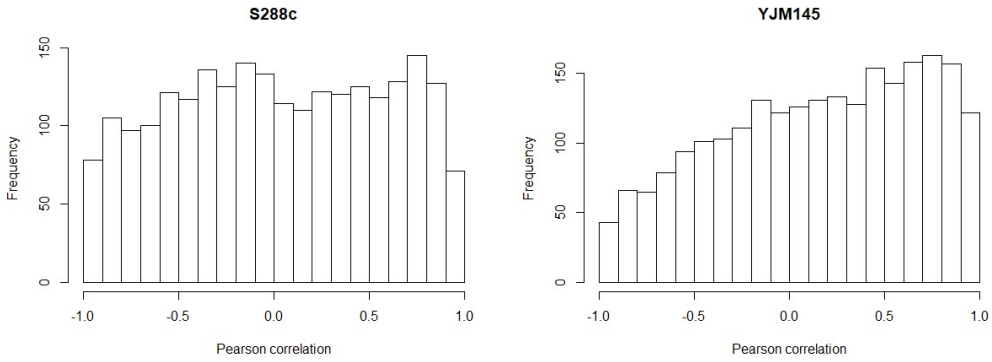


Figure 3: Four proteins that were detected to be significantly different according to the likelihood ratio test. Top-left: YDR044W (HEM13) is involved in the heme biosynthetic pathway. Top-right: YDR127W (ARO1) is involved in chromisate biosynthesis, which is the precursor to aromatic amino acids. Bottom-left: YKL127W (PGM1) codes for phosphoglucomutase, which converts glucose-1-phosphate to glucose-6-phosphate, a process necessary for hexose metabolism. Bottom-right: YOL058W (ARG1) is an enzyme in the arginine biosynthesis pathway.

A)



B)

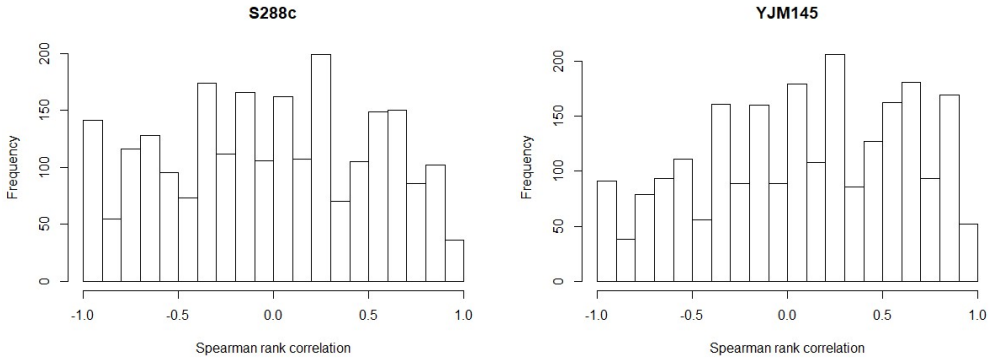


Figure 4: Histogram of mRNA to protein correlations within each strain. A) For the total of 2337 genes detected in both strains for both mRNA and protein, median Pearson correlation for S288c is 0.0154 and median correlation for YJM145 is 0.11. B) Median Spearman rank correlations are 0.05 and 0.2 for S288c and YJM145, respectively.

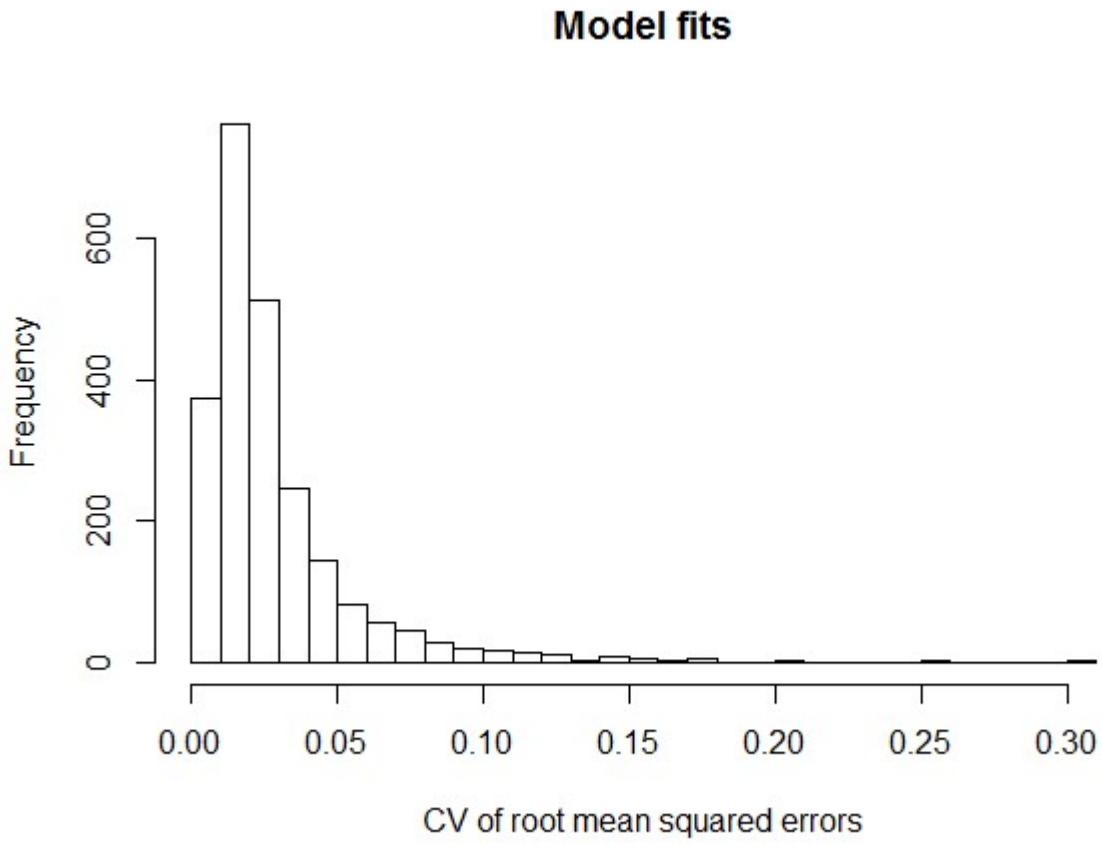


Figure 5: Histogram of the coefficient of variation of the root mean squared errors (or normalized RMS) for model fits using the simple mass-action model. The simple model is as follows

$$\frac{dP}{dt} = k_r R(t) - k_p P(t)$$

where R(t) is the amount of mRNA at time t and P(t) is the amount of protein at time t. This suggests that even a simple model can more accurately predict protein amounts than the Pearson correlations suggest.

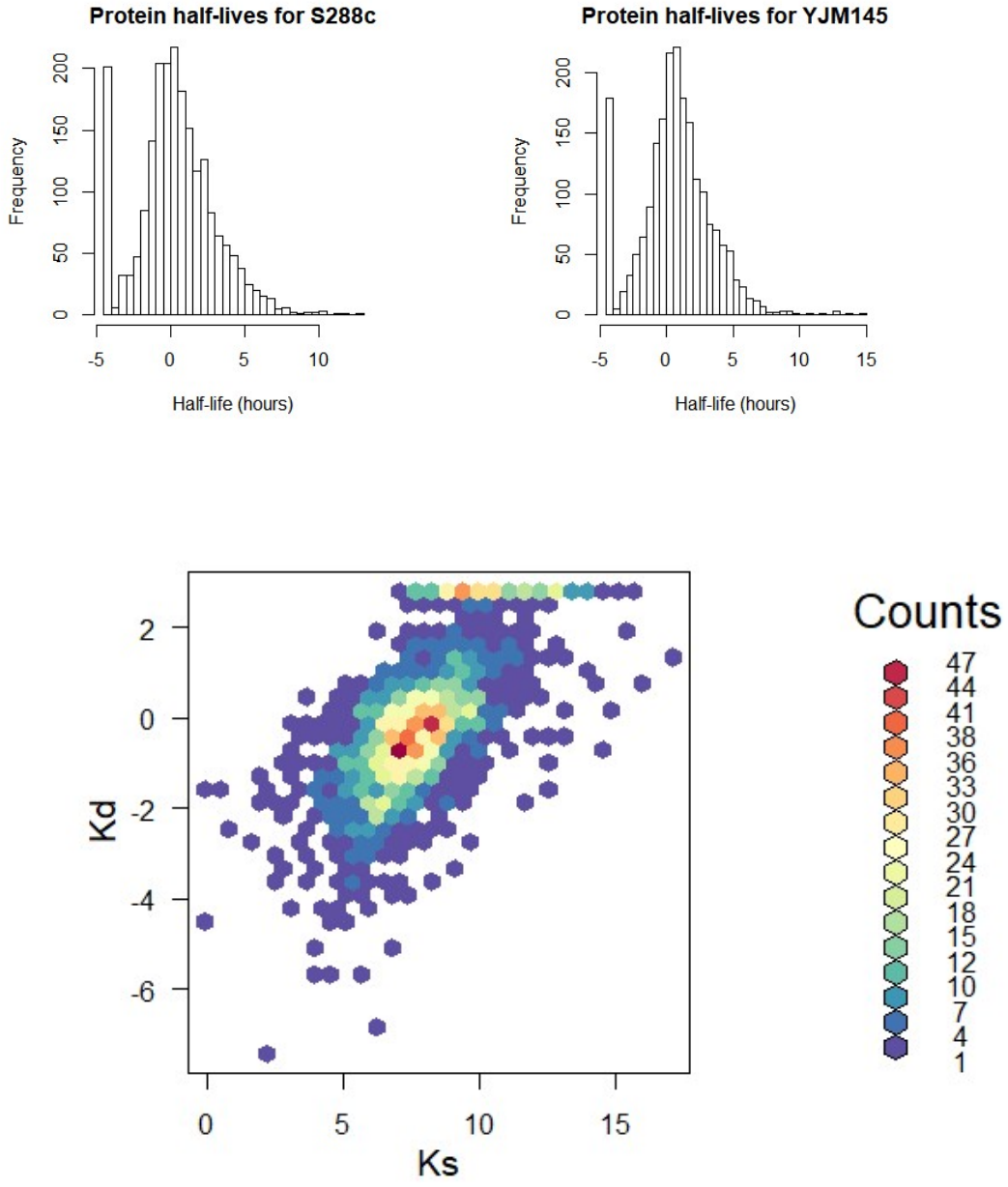


Figure 6: 2-D histogram of protein half-lives, and a density plot of translation rate-degradation rate pairs. Median half-life in S288c is 1.16 hours and 1.73 hours for YJM145. The density plot shows parameter-pair values plotted on a log-scale.

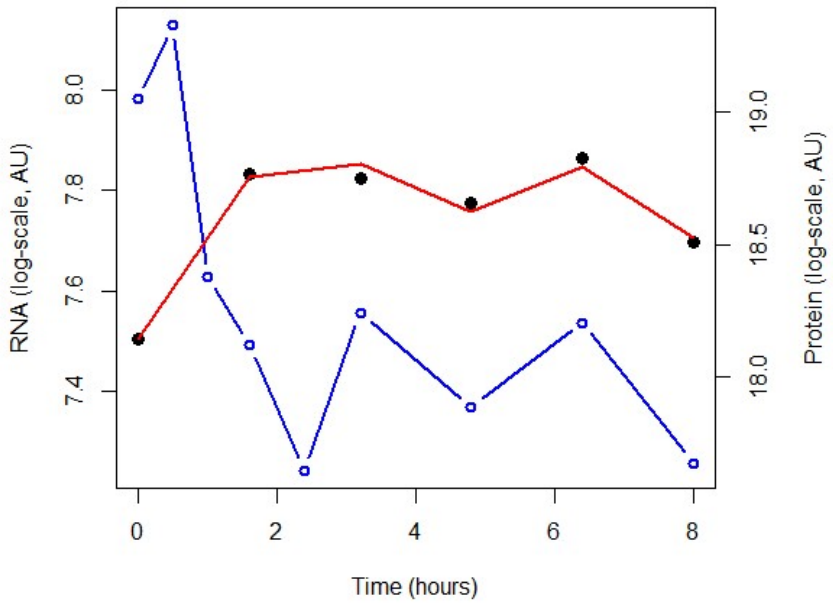


Figure 7: Example model fit for gene YER155C. The mRNA trajectory is depicted in blue while the predicted protein trajectory is depicted in red. The black points are the actual experimental values. All concentrations are on a log scale.

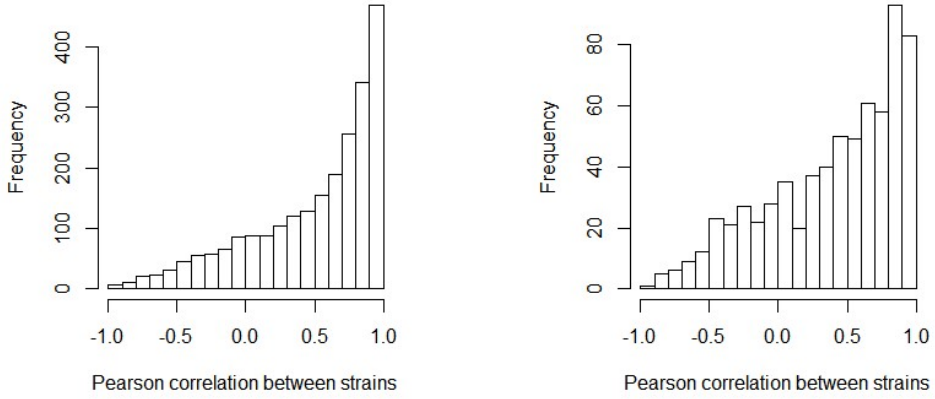


Figure 8: Histogram of protein trajectory correlations between the two strains among all genes (left) and among genes with greater than 90% parameter differences (right). Pearson correlations for protein levels remain relatively unchanged even when fitted parameter values differ between the two strains suggest that while genotypic variability may affect translation and degradation rates, the protein abundances remain similar.

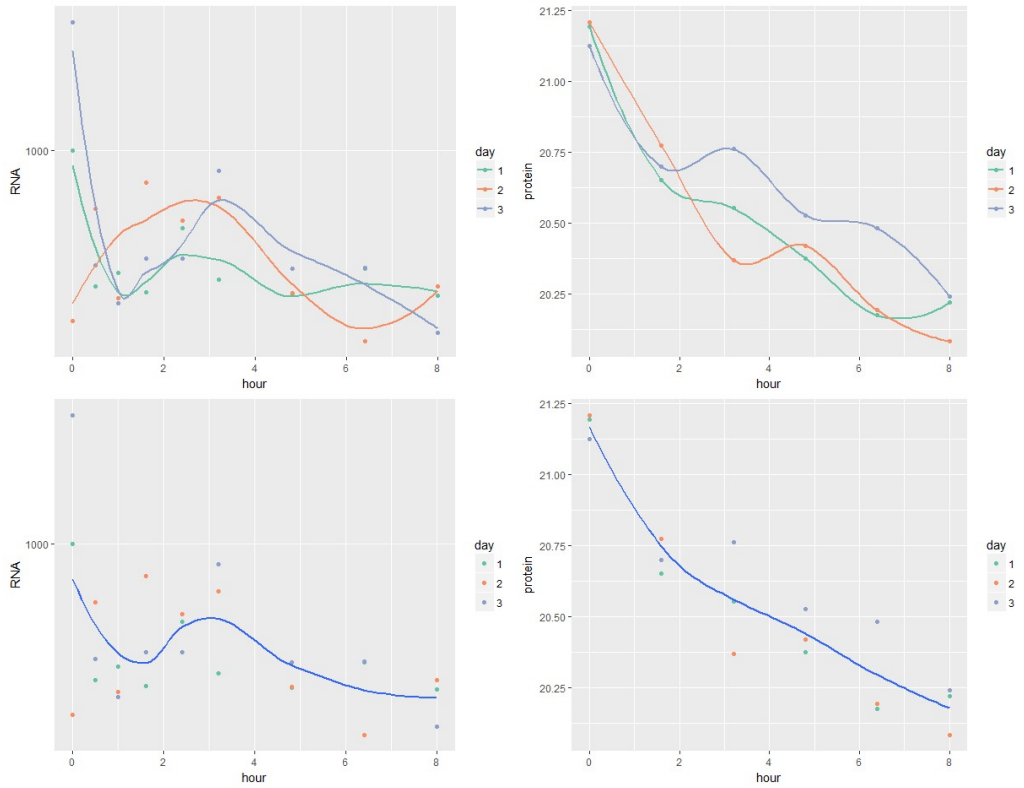


Figure 9: Global versus average models. We fit each mRNA trajectory on 3 different days (encoded by color) to three different protein trajectories and sum the residuals. The corresponding colors on the RNA and protein graphs on top represent samples taken from the same flask used for the time course. For our regular model, we take the estimated mRNA levels and predict a single protein curve, calculating the residuals similarly by summing the residuals of the individual points.

ACKNOWLEDGEMENTS

Chapter 1 is currently being prepared for submission for publication of the material. Kuo, Sidney; Egertson, Jarrett; MacCoss, Michael; Chevrier, Nicolas; Pollard, Dan; Rifkin, Scott. The dissertation author was the primary investigator and author of this material.

CHAPTER 2: MAPPING NATURAL GENETIC VARIANTS WITH TEMPORAL AND PERSISTENT EFFECTS ON THE YEAST HYPEROSMOTIC STRESS RESPONSE

INTRODUCTION

Understanding variation in dynamic cellular responses to external stimuli has been of central importance to fields as diverse as personalized medicine, climate change biology, and experimental evolution. Much work has been conducted in model systems on responses at a variety of levels ranging from the immediate transcriptional response to the overall cellular growth phenotypes. One important question that follows is how genetic variation can impact both transient and persistent cellular responses when cells are subject to stimuli. In addition, it is unknown whether the impact of genetic variation comes from genes identified in core pathways identified by previous studies, or in peripheral or maybe previously unknown loci, that exert its effects in other mechanisms. The response of *Saccharomyces cerevisiae* to high external osmolarity is a great model system to address these questions, because its molecular response to osmotic stress is well characterized, and it has a tractable mating system to facilitate mapping of genetic variants which alter those responses.

Response to hyper-osmotic stress in yeast is channeled through a well-studied MAPK cascade, of which Hog1p, the MAPK, is a homolog to the mammalian p38 known to play key roles in cancer[6,38]. In yeast, the cell signaling pathway eventually induces a cellular response by expressing proteins such as Gpd1, Gpp1/Gpp2, and Stl1, as well as affecting the behavior of other proteins such as Fps1[39–41]. Together these genes help cells accumulate glyercol to

counteract the increased external turgor pressure[42]. Studies have shown that, while the core MAPK pathway is evolutionarily conserved, many of the peripheral genes vary genetically, which may alter molecular response dynamics[43,44]. In addition, effects of genetic variation on the cellular growth adaptation to high external osmolarity environments has been studied and polymorphisms have been found that have an effect on growth in high salt[45]. However, the effects of genetic variants on the more immediate molecular response are not well understood. Examining variation in the molecular responses allow us to clarify how and what sorts of genetic variation impacts dynamic protein behavior at a molecular level, and allows us to evaluate whether transient variations lead to altered steady-state behaviors. This, in turn, may give us a better picture on which types of mutations persist during the organism's evolutionary history in signaling pathways.

We will take advantage of a promising method that has been developed to dissect multi-allelic traits, called extreme-quantitative trait locus (X-QTL) mapping, to identify casual loci for molecular phenotypes tracking protein expression[45,46]. This method utilizes a fluorescence activated cell sorter (FACS) to sort pools into extreme populations, followed by deep sequencing to genotype the selected pools. Several advantages of this method include its high statistical power, allowing even small phenotypic differences to be separated, as well as its rapidity, allowing us to perform sorting on intermediate time points during the cellular response which captures transient and persistent effects on protein levels. With these improvements over traditional QTL mapping approaches, we can elucidate dynamic effects of genetic variants during the hyperosmotic stress response in yeast.

Here, we used a Stl1p-GFP fusion protein as a molecular phenotype of the yeast response to osmotic shock, and perform X-QTL mapping at three different time points during

cellular adaption across the lab strain, S288c, and two wild strains, L-1374 and YPS606. We mapped casual alleles to several loci, especially in L-1374, across the time course. We found that L-1374 contains alleles altering its Stl1p levels in both directions, and that most alleles were found to affect Stl1p levels transiently but not alter its levels at steady state. In addition, we developed a method using the same sequenced pools to find loci involved in the variability of Stl1p, which was used to identify a locus that is involved in both Stl1p level and variance. These discoveries show how temporal dynamics of proteins in response to stimuli may be altered by genetic variation.

METHODS

STRAINS AND EXPERIMENTAL CONDITIONS

The strains used in this experiment were S288c (MAT **a**), YPS606 (MAT **a**) and L-1374 (MAT alpha). The following genes were knocked out with the loxP system[47]: URA3, HO, AMN1. STL1 was tagged with GFP-HphMX and selected on hygromycin. HIS3 was knocked out with a NatMX and selected on nourseothricin. CAN1 was replaced in S288c and YPS606 with a Prom-STE2 driving Sp.HIS5 in order to facilitate selection of selection of MAT **a** haploids[48,49]. Cells were mated in YPD and sporulated for up to a week by observing spore formation percentage. Sporulated cells were selected on synthetic minimal media supplemented with uracil with the drugs canavanine, hygromycin and nourseothricin added for 24 hours. Then cells were scrapped off plates with a cell spreader and frozen down. Each plate was frozen into two cryo-tubes.

For FACS, cells were thawed from the cryo-tubes into 50 mL YPD for 10-12 hours, diluted to an OD of 0.7, and incubated for another 1.5 hours before sorting. Three replicates were performed for each cross, with each replicate being taken from a different cryo-tube.

Before adding salt, 1 mL of cells for each sample was retrieved and suspended in 1x PBS and put on ice. Cells were then induced in NaCl by adding YPD+4M NaCl to the culture flask to a final concentration of 0.8M NaCl. Cell sorting was performed immediately after retrieving the 45 minute sample, re-suspending in 1x PBS, and placing on ice. The 90 minute sample was prepared in a similar fashion. Low and high (lowPool and hiPool) populations were taken from the top and bottom 2% of cells, for a total of 50,000 cells each, and the null pool was sorted without any selection on fluorescence with also 50,000 cells. Sorted cells were recovered in 1 mL YPD and frozen. DNA preparation was performed on all samples simultaneously after growing the sorted pools in 3 mL YPD overnight using the mini-prep protocol from *Methods in Yeast Genetics*[50].

SEQUENCING AND PROCESSING

Sequencing was performed by BGI on an Illumina HiSeq4000, using two lanes for our 54 samples. The DNA yield for four samples were too low and ultimately not sequenced, so the remaining two replicates were used in the downstream analysis. Alignments were performed using BWA to the reference genome (R64-2-1, 2015-01-13) using the Galaxy servers[29,31]. Variant calling was performed locally using Samtools *mpileup* command[51,52]. Subsequent analyses and plots were done in Python using scripts adapted from Lee, et al., 2017[53]. LOD peaks were called using MULTIPOOL[54] using options according to previous studies[37,45].

Level QTLs were detected by contrasting the high pools and low pools, using all three replicates pooled together to minimize flask-specific effects. Variance QTL were detected by adding the pooled allele frequencies of the low pool and high pool at each time point, multiplying by the total read count at each locus, and contrasting that with null pool.

MICROSCOPY

Microscopy was performed on a Nikon TiE using Hog1p-YFP and with 100x magnification and photos taken every minute. Hog1p-YFP nuclear localization was calculated using the. The top ten percentile of fluorescent pixels, which represents to nuclear fluorescent intensity, was divided by the cellular mean fluorescent intensity to generate localization curves. Automated cell segmentation software was adapted from Hansen lab software[55,56] in MATLAB, and subsequent fluorescent calculations were also performed automatically in MATLAB[57].

RESULTS

USING A MOLECULAR PHENOTYPE DURING THE YEAST OSMOTIC SHOCK RESPONSE

In order to study the osmotic shock pathway in yeast, we selected the Stl1p protein to use as a readout for the cellular response to osmotic shock. Stl1p is known to be expressed at a low basal level and induced significantly upon exposure to osmotic stress[58]. We measured Stl1p expression by generating a Stl1p-GFP fusion protein and monitoring its activity during adaption to 0.8 M NaCl using a flow cytometer. Surveying several of the strains from *Saccharomyces* Genome Resequencing Project collection[59], we found two in particular that show dynamic differences compared to the lab strain, S288c (Figure 10). We adopted a QTL mapping approach developed by Kruglyak, dubbed X-QTL, to identify casual loci for these observed differences[48].

We generated two pools of haploid offspring using our three strains. The first pool was a cross between S288c and L-1374. Flow cytometry data shows the segregant pool to have expected levels of Stl1p-GFP throughout the time course: distributed between the two extreme parental phenotypes (Figure 10). However, in our second pool, which is a cross between L-1374 and YPS606, the distribution of fluorescence levels in the offspring is not simply a mixture of the parental phenotypes, which may be indicative of interactive alleles at the underlying loci[60]. Using a fluorescence-activated cell sorter (FACS), we isolated the high and low Stl1p-GFP extreme populations and a null population with no selection on fluorescence at three different time points in each cross. We performed three biological replicates for each sample. The first time point represents a glucose steady-state growth condition, before the cells have been exposed to NaCl. The second time point, at 45 minutes

after exposure to osmotic stress, represents the intermediate adaptation state. The third time point represents the steady-state adapted levels of Stl1p-GFP to 0.8 M NaCl. In both crosses, the L-1374 parent had higher expression at the first and last time points but had lower expression at the intermediate time point. With our mapping approach, we can see whether the same alleles cause this difference at all three times.

MAPPING STL1P-GFP LEVELS AS A DYNAMIC TRAIT

After isolating the extreme segregants in three biological replicates using FACS, we genotyped the pools via Illumina sequencing. We pooled the genotype data from the replicates to calculate allele frequencies at each SNP. The median sequencing coverage at each SNP was ~300x after pooling across most samples, giving us good statistical power to detect smaller effect QTL. After generating the allele frequencies at each SNP, we used MULTIPOOL to calculate LOD scores across the genome[54]. We used a LOD threshold cutoff of 5, as previously reported[45,48], to identify the casual loci throughout the time course.

Examining the allele frequencies and LOD scores for the intermediate time point shows that Stl1p fluorescence level is indeed a multi-locus trait (Figure 11). Comparing the data from the two segregant pools, we find that the two strongest peaks are at the same loci, on chromosomes 9 and 15. This may not initially be surprising because the allele frequency graph shows that the L-1374 allele is over-represented at these two loci, and L-1374 is the common strain in these two crosses. However, the phenotype data (Figure 10) at 45 minutes shows that the L-1374 parental strain exhibits *lower* Stl1p-GFP levels, and, therefore, its alleles should be under-represented. Thus, the strongest LOD peaks correspond to polymorphisms that bring Stl1p-GFP levels more in line with what is observed in S288c and YPS606, counteracting several smaller peaks that lower Stl1p levels. Together, the smaller

peaks must have a stronger effect since that is what was observed in the L-1374 parental strain. These intervals harbor two candidate genes that might have an effect on Stl1p levels: SLN1 and HAL9. SLN1, in particular, is known to be one of the sensor proteins for osmotic shock response in yeast[61]. L-1374 has 23 non-synonymous mutations in SLN1 compared to S288c, while S288c and YPS606 have the same protein product.

In order to elucidate the potential mechanism, we also observed the behavior of Hog1p, which is the MAPK that is responsible for activation of Stl1p[58,61,62]. During the response to osmotic stress, Hog1p translocates to the nucleus, acting as a transcription factor for several genes, one of which is Stl1[7]. Measuring Hog1p nuclear localization levels with microscopy during osmotic stress response, we see that the three strains have distinct Hog1p nuclear localization patterns upon exposure to salt shock (Figure 12). The immediate localization pattern is similar across the three strains in timing, but not in localization amount. The amount of Hog1p localized could have an impact on Stl1p-GFP levels, as Hog1p one of the two direct transcriptional activators of STL1[7,8]. On the other hand, the timing in Hog1p localization does not start to diverge until close to 40 minutes after salt shock, which makes it unlikely to have an impact on Stl1p levels we measured at 45 minutes. Thus, given the Hog1p localization pattern, it may be the case that Hog1p localization amount has an impact on the immediate Stl1p levels, in addition to host of other osmo-stress response proteins, which causes slower adaptation in L-1374, which is reflected by its slower Hog1p de-localization and slower increase of Stl1p.

The other LOD peaks which affect Stl1p levels at 45 minutes have effects in the expected direction (i.e. lowering Stl1p levels in L-1374). The genes MLP2 and KAP122 fall into two of the intervals. They are involved in nuclear transport, and they may contribute to

Stl1p-GFP level differences. Since both Hog1p and Hot1p need to be nuclear localized and bound to the STL1 promoter in order to induce Stl1p[7], these are strong candidates responsible for the observed differences. For MLP2, in particular, there are 14 non-synonymous mutations between L-1374 and S288c.

Next, we compared the LOD plots across the three time points for both crosses. For the YPS606 and L-1374 cross, we detected no peaks at either the beginning or the end time point that meet our threshold (Figure 10). Even despite clear phenotypic differences, stringent selection via FACS, and sequencing at high coverage, no alleles of either parent were found to be over-represented between the two extreme pools. This somewhat surprising result reinforces the idea that higher order gene interactions in this particular cross a (Figure 10) resulted in no peaks showing up with our experimental set up. This would be consistent with the phenotypic data. On the other hand, two peaks showed up at both time points for the cross between S288c and L-1374 (Figure 13). In particular, the two peaks in chromosomes 8 and 12 at 90 minutes seem to also be present, covering the same loci, at 45 minutes, and to a lesser degree, be present pre-induction. This suggests that perhaps these loci have a persistent effect on Stl1p during salt shock, and perhaps even have roles in basal Stl1p levels. Chromosome 8 is a fairly narrow interval, and only encompasses two known genes if we take the LOD support interval to be 10% of its maximal value. However, GSY2, which is a glycogen synthase, and HSP60 have no known connections to the osmotic stress pathway. On the other hand, the interval at chromosome 12 encompasses GPA1, which has a known role in feedback of the MAPK pathway. Given that its effect seems stronger at the later time point than initially, it is a strong candidate QTL for affecting Stl1p levels.

NO VARIANCE QTLS WERE FOUND DESPITE VARIABILITY DIFFERENCES

Our experimental setup also allowed us to determine whether QTL contribute to the variance of Stl1p-GFP levels. Even though variance in fluorescence in the parental strains are similar, the segregant pool exhibits increased variance suggesting transgressive segregation (Figure 10). In order to identify QTL affecting variance, we pooled the high and low selected pools into one pool by averaging the frequencies and multiplying by total coverage, and then we contrasted this with our null pool, which is not selected for fluorescence. We generated similar allele frequency graphs and calculated corresponding LOD scores. Surprisingly, despite differences in variability of Stl1p (Figure 10), we did not detect any other peaks that met our LOD threshold. There are a few possibilities as to why we could not detect other loci. Perhaps our study is underpowered to detect such a trait, especially if variability in this particular protein depends on small effects from many loci. Also, there could exist complex genetic interactions as mentioned above, which would not necessarily show up with our design. And finally, it is possible that variability in induced by purely environmental factors.

DISCUSSION

Discovering the molecular dynamics of the osmostress response in yeast has been a longstanding objective for cell biologists. Even though most of the key proteins and their respective interactions are known, many effectors of this canonical pathway are not. Leveraging natural genetic variation and a powerful quantitative trait mapping approach, we uncovered more genes that affect this cell signaling pathway. While QTL studies have been performed on salt tolerance as a growth phenotype in yeast[37], by observing a much more

immediate molecular phenotype such as STL1, we can identified alleles that have an effect on the immediate signal transduction process.

Surprisingly, none of the known upstream activators of STL1, nor the cis-regulatory region were identified as QTL for its level. Two proteins that are responsible for activation of STL1, Hog1p and Hot1p[8] were not detected as significant peaks in our crosses, despite having non-synonymous difference between strains. In addition, two known proteins that complex with Hot1p, Cyc8p and Tup1p[7], were also not detected in this mapping experiment. SKO1, another gene that is known to modulate levels of STL1 induction after salt shock[8], was also not detected as a QTL. This is evidence that natural variation in the temporal dynamics of STL1 induction is likely caused by many perhaps previously unknown effectors of this pathway.

Indeed, most of the loci detected are not part of the core osmotic stress pathway as determined by classical genetics and systems biology. Only GPA1 and SLN1 were from the known salt tolerance pathway, and while GPA1 was detected also as a QTL for growth on salt, SLN1 was not[45]. Interestingly, SLN1 was detected as one of the largest effect loci, but did not seem to have a persistent effect, showing up only in the intermediate time point but not the steady state time point. Perhaps the feedback mechanisms in the Hog pathway or its dual branch nature with SHO1 compensating for any differences. Given that the L-1374 allele of SLN1 appears in many other wild strains, perhaps the functional difference in SLN1 has phenotypic consequences in other conditions. Another possibility is that it evolved as a compensatory mutation because the L-1374 allele makes its Stl1p level more similar to S288c.

A few potential candidate genes sit under the other peaks. HAL9, which is a known transcription factor that activates ENA1, a sodium ion pump[63], may be involved in Stl1p levels, although it is unclear why it would only affect the intermediate time point. In addition, three genes that are involved in nuclear transport, MSN5, MLP1, and KAP122 may have an effect on Stl1p levels, especially given the Hog1p nuclear localization behavior differences between S288c and L-1374 (Figure 12). MSN5, in particular, is known to be involved in the pheromone response pathway[64], another MAPK pathway that shares some pathway components with the Hog pathway, and could perhaps alter Stl1p levels via limited crosstalk. Null mutants of KAP122 exhibit lower osmotic stress tolerance among a variety of other phenotypes[65] could be a strong candidate for further investigation. However, other intervals, such as the peaks at chromosomes 1, 4, and 5, at the intermediate time point, which represent the third, fourth and fifth most significant LOD scores, do not have clear candidates (Figure 11B). Since the LOD scores correlate to the effect size, the intervals on chromosomes 1 and 4 represent the largest effect alleles that lower Stl1p levels in L-1374, and uncovering the mechanistic details would contribute to a fuller understanding of the possible genetic perturbations to functional pathway.

FIGURES

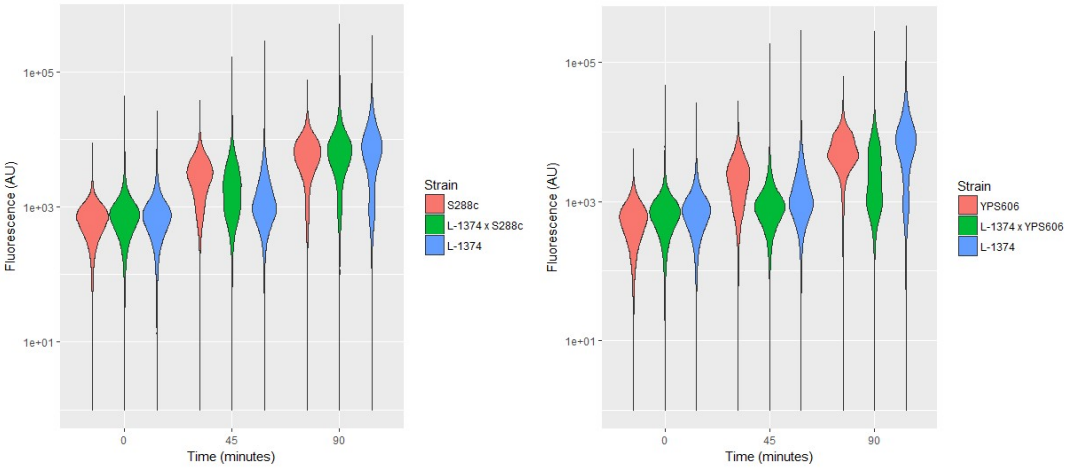


Figure 10: Violin plots of STL1 fluorescence levels at 3 time points during hyperosmotic stress response. For each time point, STL1 levels were measured via a flow cytometer. The area represents a histogram of the single-cell fluorescence levels during that time point. The green in each graph represents the heterogeneous, haploid sporulated pools from the two parental strains.

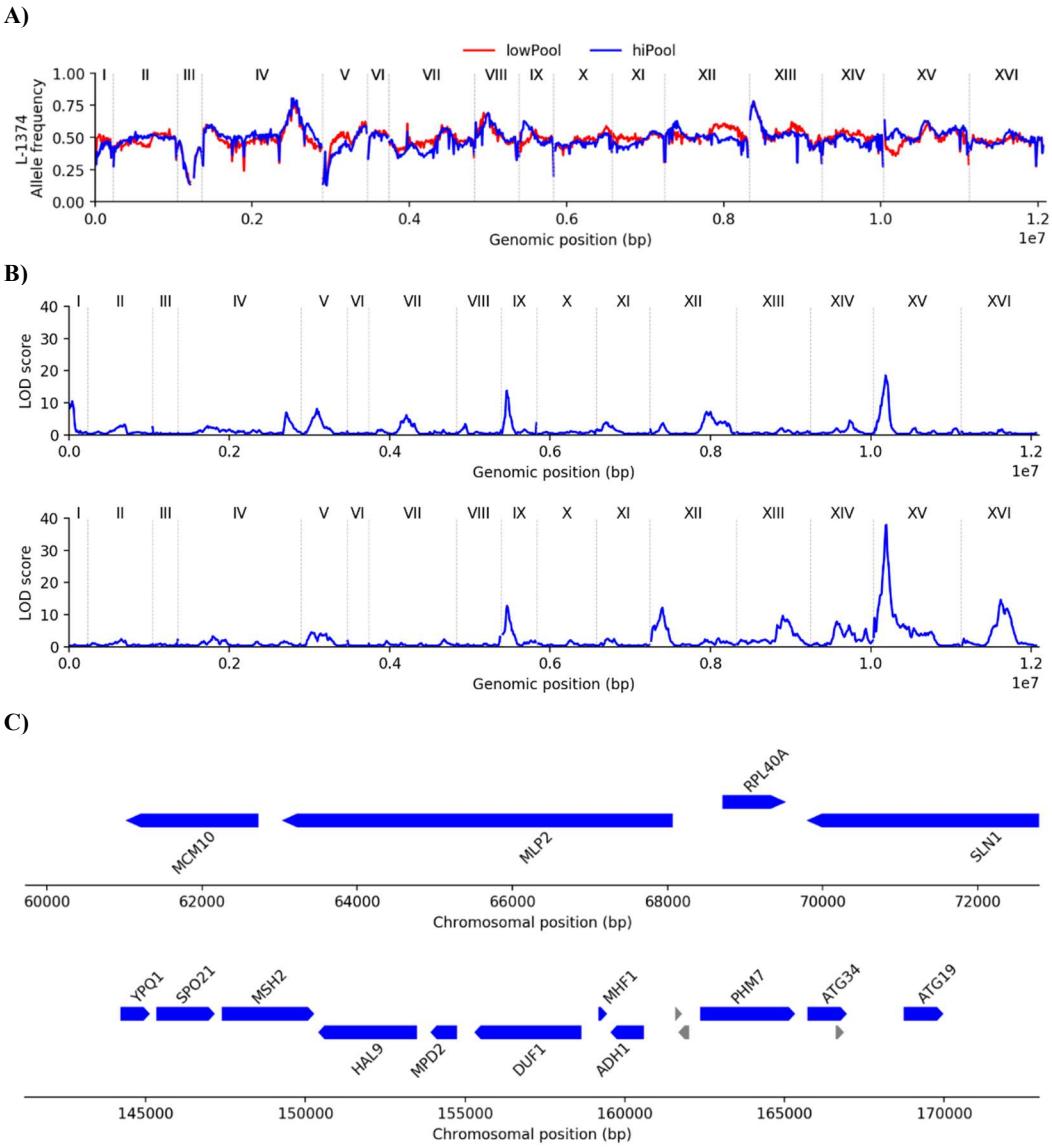


Figure 11: Identification of casual loci that contribute to STL1 differences at 45 minutes post salt shock. A) Allele frequencies were plotted across the genome with the two selected pools in different colors for S288c x L-1374. Based on parental phenotypes, we expect the low pool to higher in L-1374 allele frequency, so any loci where the opposite occurs is an example of a compensatory allele. B) LOD plots were calculated using MULTIPOOL using allele frequencies determined by sequencing. The top LOD plot shows S288c x L-1374, while the bottom one shows YPS606 x L-1374. Two prominent peaks in chromosome 9 and chromosome 15 appear in both pools. C) Gene annotations are shown for chromosomes 9 (top) and 15 (bottom). SLN1 is one of the transmembrane proteins that is involved in osmolarity sensing in yeast, and HAL9 is known to be involved in salt tolerance. Note that according to the allele frequency plot, both of these loci have compensatory alleles in L-1374.

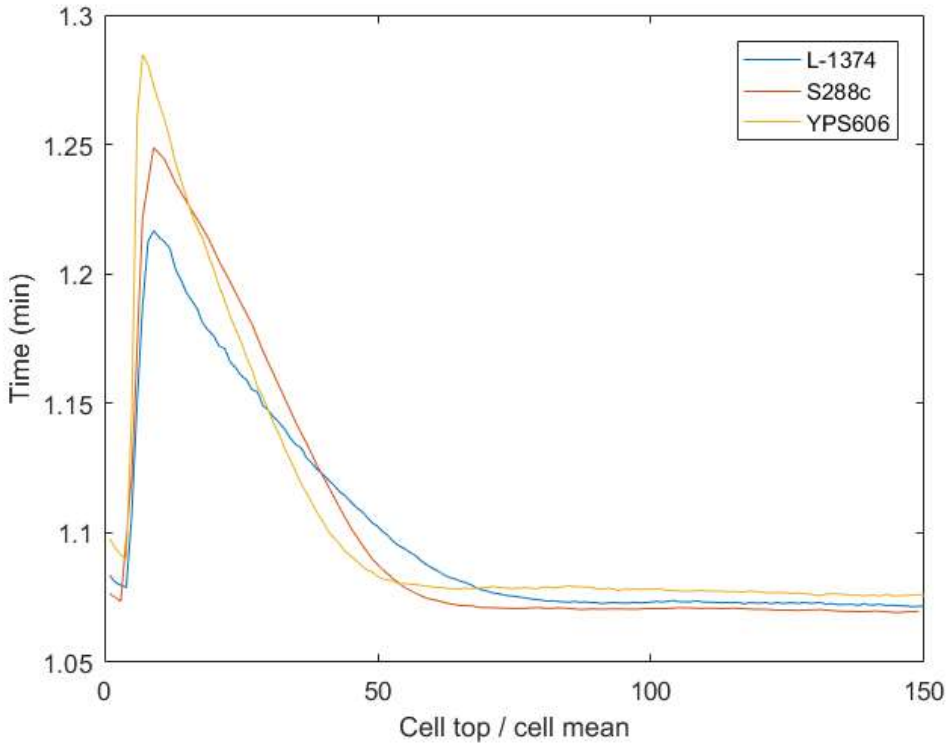


Figure 12: Comparison of Hog1 nuclear localization upon exposure to 0.8 M NaCl. Hog1p was tagged with YFP and nuclear localization was measured via microcopy. To calculate nuclear enrichment, we used the ratio of the top 10 percentile fluorescent pixels, representative of the nuclear Hog1p level, to the cellular mean fluorescence level. Images were taken every two minutes for S288c, and every minute for L-1374 and YPS606.

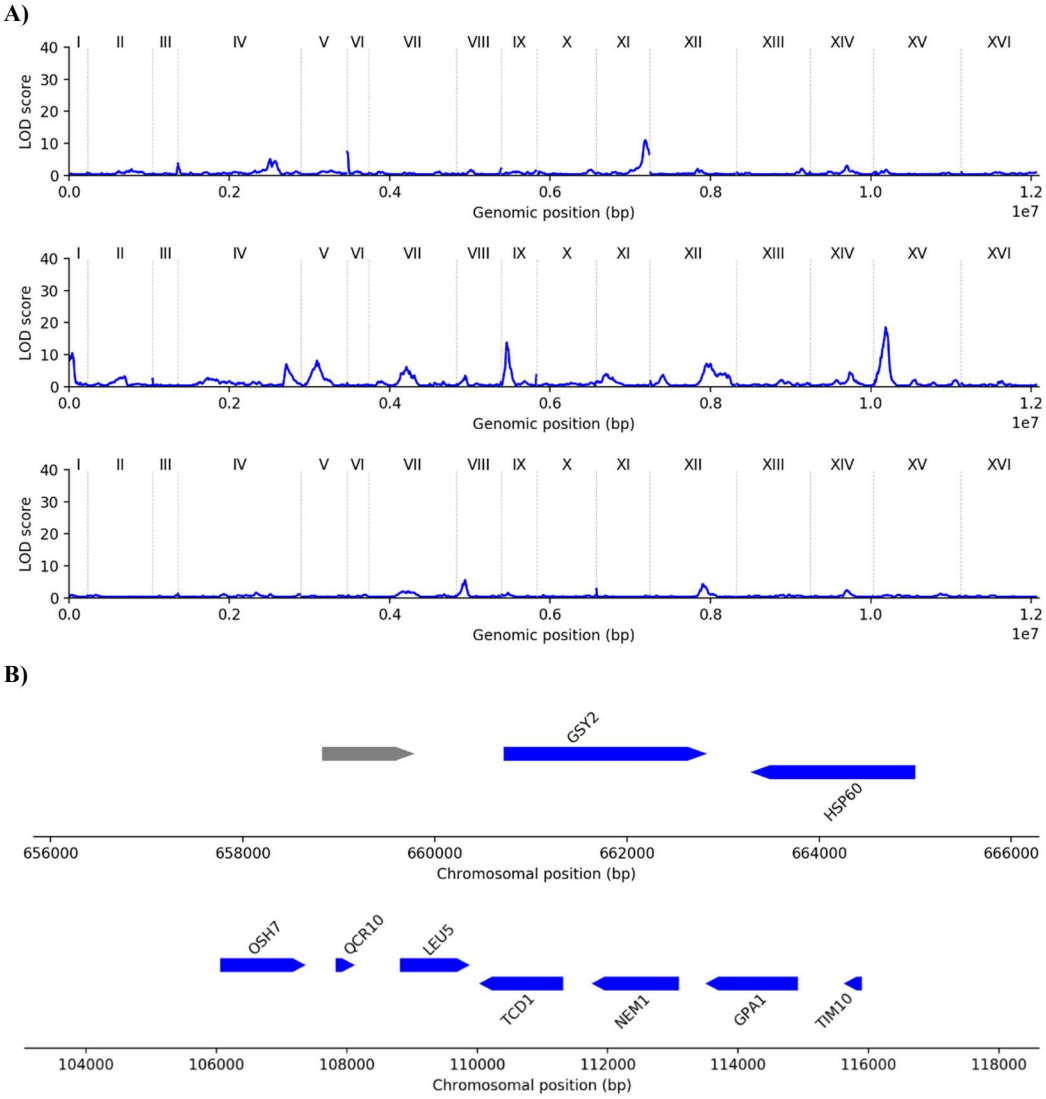


Figure 13: Potential causative loci involved throughout the time course in L-1374 x S288c. A) The top, middle and bottom LOD plots are the 0 (pre-induction) minute, 45 minute and 90 minute time points for the time course. From the bottom two LOD plots, we see that the loci in chromosomes 8 and 12 are involved in the entire osmotic shock response. B) Gene annotations were plotted for the peaks at chromosome 8 (top) and chromosome 12 (bottom).

ACKNOWLEDGEMENTS

Chapter 2 is currently being prepared for submission for publication of the material. Kuo, Sidney; Schwartz, Ruth; Rifkin, Scott. The dissertation author was the primary investigator and author of this material.

BIBLIOGRAPHY

1. Bardwell, L. A walk-through of the yeast mating pheromone response pathway. *Peptides* **26,** 339–350 (2005).

2. Gallone, B., Steensels, J., Prahl, T., Soriaga, L., Saels, V., Herrera-Malaver, B., Merlevede, A., Roncoroni, M., Voordeckers, K., Miraglia, L., Teiling, C., Steffy, B., Taylor, M., Schwartz, A., Richardson, T., White, C., Baele, G., Maere, S. & Verstrepen, K. J. Domestication and Divergence of Saccharomyces cerevisiae Beer Yeasts. *Cell* **166,** 1397–1410.e16 (2016).

3. Gray, J. C. & Goddard, M. R. Sex enhances adaptation by unlinking beneficial from detrimental mutations in experimental yeast populations. *BMC Evol. Biol.* **12,** 43 (2012).

4. Zheng, W., Zhao, H., Mancera, E., Steinmetz, L. M. & Snyder, M. Genetic analysis of variation in transcription factor binding in yeast. *Nature* **464,** 1187–1191 (2010).

5. Roberts, C. J., Nelson, B., Marton, M. J., Stoughton, R., Meyer, M. R., Bennett, H. A., He, Y. D., Dai, H., Walker, W. L., Hughes, T. R., Tyers, M., Boone, C. & Friend, † Stephen H. Signaling and Circuitry of Multiple MAPK Pathways Revealed by a Matrix of Global Gene Expression Profiles. *Science (80-.).* **287,** (2000).

6. Brewster, J., de Valoir, T., Dwyer, N., Winter, E. & Gustin, M. An osmosensing signal transduction pathway in yeast. *Science (80-.).* **259,** (1993).

7. Alepuz, P. M., de Nadal, E., Zapater, M., Ammerer, G. & Posas, F. Osmostress-induced transcription by Hot1 depends on a Hog1-mediated recruitment of the RNA Pol II. *EMBO J.* **22,** 2433–42 (2003).

8. Bai, C., Tesker, M. & Engelberg, D. The yeast Hot1 transcription factor is critical for activating a single target gene, STL1. *Mol. Biol. Cell* **26,** 2357–74 (2015).

9. Lu, P., Vogel, C., Wang, R., Yao, X. & Marcotte, E. M. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat. Biotechnol.* **25,** 117–124 (2007).

10. Wu, G., Nie, L. & Zhang, W. Integrative Analyses of Posttranscriptional Regulation in the Yeast Saccharomyces cerevisiae Using Transcriptomic and Proteomic Data. *Curr. Microbiol.* **57,** 18–22 (2008).

11. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science (80-.).* **324,** 218–223 (2009).

12. Gygi, S. P., Rochon, Y., Franza, B. R. & Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **19,** 1720–30 (1999).

13. Futcher, B., Latter, G. I., Monardo, P., McLaughlin, C. S. & Garrels, J. I. A sampling of the yeast proteome. *Mol. Cell. Biol.* **19,** 7357–68 (1999).

14. Washburn, M. P., Koller, A., Oshiro, G., Ulaszek, R. R., Plouffe, D., Deciu, C., Winzeler, E. & Yates, J. R. Protein pathway and complex clustering of correlated mRNA and protein expression analyses in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 3107–12 (2003).

15. Greenbaum, D., Colangelo, C., Williams, K. & Gerstein, M. Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol.* **4,** 117 (2003).

16. Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K. & Weissman, J. S. Global analysis of protein expression in yeast. *Nature* **425,** 737–741 (2003).

17. Maier, T., Güell, M. & Serrano, L. Correlation of mRNA and protein in complex biological samples. *FEBS Lett.* **583,** 3966–3973 (2009).

18. Shapiro, M. D., Marks, M. E., Peichel, C. L., Blackman, B. K., Nereng, K. S., Jónsson, B., Schluter, D. & Kingsley, D. M. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428,** 717–723 (2004).

19. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).

20. Külahoglu, C. & Bräutigam, A. Quantitative Transcriptome Analysis Using RNA-seq. in *Methods in molecular biology (Clifton, N.J.)* **1158,** 71–91 (2014).

21. Egertson, J. D., MacLean, B., Johnson, R., Xuan, Y. & MacCoss, M. J. Multiplexed peptide analysis using data-independent acquisition and Skyline. *Nat. Protoc.* **10,** 887–903 (2015).

22. MacCoss, M. J. & Matthews, D. E. Quantitative MS for proteomics: teaching a new dog old tricks. *Anal. Chem.* **77,** 294A–302A (2005).

23. MacKay, V. L., Li, X., Flory, M. R., Turcott, E., Law, G. L., Serikawa, K. A., Xu, X. L., Lee, H., Goodlett, D. R., Aebersold, R., Zhao, L. P. & Morris, D. R. Gene expression analyzed by high-resolution state array analysis and quantitative

proteomics: response of yeast to mating pheromone. *Mol. Cell. Proteomics* **3,** 478–89 (2004).

24. Brem, R. B., Yvert, G., Clinton, R. & Kruglyak, L. Genetic Dissection of Transcriptional Regulation in Budding Yeast. *Science (80-.).* **296,** 752–755 (2002).

25. Rockman, M. V. & Kruglyak, L. Genetics of global gene expression. *Nat. Rev. Genet.* **7,** 862–872 (2006).

26. Wei, W., McCusker, J. H., Hyman, R. W., Jones, T., Ning, Y., Cao, Z., Gu, Z., Bruno, D., Miranda, M., Nguyen, M., Wilhelmy, J., Komp, C., Tamse, R., Wang, X., Jia, P., Luedi, P., Oefner, P. J., David, L., Dietrich, F. S., Li, Y., Davis, R. W. & Steinmetz, L. M. Genome sequencing and comparative analysis of Saccharomyces cerevisiae strain YJM789. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 12825–30 (2007).

27. Steinmetz, L. M., Sinha, H., Richards, D. R., Spiegelman, J. I., Oefner, P. J., McCusker, J. H. & Davis, R. W. Dissecting the architecture of a quantitative trait locus in yeast. *Nature* **416,** 326–30 (2002).

28. Gueldener, U., Heinisch, J., Koehler, G. J., Voss, D. & Hegemann, J. H. A second set of loxP marker cassettes for Cre-mediated multiple gene knockouts in budding yeast. *Nucleic Acids Res.* **30,** 23e–23 (2002).

29. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).

30. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31,** 166–9 (2015).

31. Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., Grüning, B., Guerler, A., Hillman-Jackson, J., Von Kuster, G., Rasche, E., Soranzo, N., Turaga, N., Taylor, J., Nekrutenko, A. & Goecks, J. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* **44,** W3–W10 (2016).

32. Choi, M., Chang, C.-Y., Clough, T., Broudy, D., Killeen, T., MacLean, B. & Vitek, O. MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* **30,** 2524–2526 (2014).

33. Herbrich, S. M., Cole, R. N., West, K. P., Schulze, K., Yager, J. D., Groopman, J. D., Christian, P., Wu, L., O'Meally, R. N., May, D. H., McIntosh, M. W. & Ruczinski, I. Statistical Inference from Multiple iTRAQ Experiments without Using Common Reference Standards. *J. Proteome Res.* **12,** 594–604 (2013).

34. Christiano, R., Nagaraj, N., Fröhlich, F. & Walther, T. C. Global Proteome Turnover Analyses of the Yeasts S. cerevisiae and S. pombe. *Cell Rep.* **9,** 1959–1965 (2014).

35. Peshkin, L., Wühr, M., Pearl, E., Haas, W., Freeman, R. M., Gerhart, J. C., Klein, A. M., Horb, M., Gygi, S. P. & Kirschner, M. W. On the Relationship of Protein and mRNA Dynamics in Vertebrate Embryonic Development. *Dev. Cell* **35,** 383–394 (2015).

36. Belle, A., Tanay, A., Bitincka, L., Shamir, R. & O'Shea, E. K. Quantification of protein half-lives in the budding yeast proteome. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 13004–9 (2006).

37. Albert, F. W., Muzzey, D., Weissman, J. S. & Kruglyak, L. Genetic Influences on Translation in Yeast. *PLoS Genet.* **10,** e1004692 (2014).

38. Han, J., Lee, J., Bibbs, L. & Ulevitch, R. A MAP kinase targeted by endotoxin and hyperosmolarity in mammalian cells. *Science (80-.).* **265,** (1994).

39. Ferreira, C., van Voorst, F., Martins, A., Neves, L., Oliveira, R., Kielland-Brandt, M. C., Lucas, C. & Brandt, A. A member of the sugar transporter family, Stl1p is the glycerol/H+ symporter in Saccharomyces cerevisiae. *Mol. Biol. Cell* **16,** 2068–76 (2005).

40. Tamas, M. J., Luyten, K., Sutherland, F. C. W., Hernandez, A., Albertyn, J., Valadi, H., Li, H., Prior, B. A., Kilian, S. G., Ramos, J., Gustafsson, L., Thevelein, J. M. & Hohmann, S. Fps1p controls the accumulation and release of the compatible solute glycerol in yeast osmoregulation. *Mol. Microbiol.* **31,** 1087–1104 (1999).

41. Albertyn, J., Hohmann, S., Thevelein, J. M. & Prior, B. A. GPD1, which encodes glycerol-3-phosphate dehydrogenase, is essential for growth under osmotic stress in Saccharomyces cerevisiae, and its expression is regulated by the high-osmolarity glycerol response pathway. *Mol. Cell. Biol.* **14,** 4135–44 (1994).

42. Klipp, E., Nordlander, B., Krüger, R., Gennemark, P. & Hohmann, S. Integrative model of the response of yeast to osmotic shock. *Nat. Biotechnol.* **23,** 975–982 (2005).

43. Bahn, Y.-S., Xue, C., Idnurm, A., Rutherford, J. C., Heitman, J. & Cardenas, M. E. Sensing the environment: lessons from fungi. *Nat. Rev. Microbiol.* **5,** 57–69 (2007).

44. Connallon, T., Knowles, L. L., Wang, P., Zheng, D., Ding, R., Li, Y., Valentin, F., Wallace, I., Wilm, A., Lopez, R., Thompson, J., Gibson, T. & Higgins, D. Recombination rate and protein evolution in yeast. *BMC Evol. Biol.* **7,** 235 (2007).

45. Treusch, S., Albert, F. W., Bloom, J. S., Kotenko, I. E. & Kruglyak, L. Genetic Mapping of MAPK-Mediated Complex Traits Across S. cerevisiae. *PLoS Genet.* **11,** e1004913 (2015).

46. Ehrenreich, I. M., Torabi, N., Jia, Y., Kent, J., Martis, S., Shapiro, J. A., Gresham, D., Caudy, A. A. & Kruglyak, L. Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature* **464,** 1039–1042 (2010).

47. Gueldener, U. A second set of loxP marker cassettes for Cre-mediated multiple gene knockouts in budding yeast. *Nucleic Acids Res.* **30,** 23e–23 (2002).

48. Ehrenreich, I. M., Torabi, N., Jia, Y., Kent, J., Martis, S., Shapiro, J. A., Gresham, D., Caudy, A. A. & Kruglyak, L. Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature* **464,** 1039–1042 (2010).

49. Tong, A. H. Y. & Boone, C. 16 High-Throughput Strain Construction and Systematic Synthetic Lethal Screening in. in 369–707 (2007). doi:10.1016/S0580-9517(06)36016-3

50. Amberg, D. C., Burke, D., Strathern, J. N., Burke, D. & Cold Spring Harbor Laboratory. *Methods in yeast genetics : a Cold Spring Harbor Laboratory course manual*. (Cold Spring Harbor Laboratory Press, 2005).

51. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. & 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

52. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27,** 2987–2993 (2011).

53. Lee, K. B., Wang, J., Palme, J., Escalante-Chong, R., Hua, B. & Springer, M. Polymorphisms in the yeast galactose sensor underlie a natural continuum of nutrient-decision phenotypes. *PLOS Genet.* **13,** e1006766 (2017).

54. Edwards, M. D. & Gifford, D. K. High-resolution genetic mapping with pooled sequencing. *BMC Bioinformatics* **13 Suppl 6,** S8 (2012).

55. Stockwell, S. R. & Rifkin, S. A. A living vector field reveals constraints on galactose network induction in yeast. *Mol. Syst. Biol.* **13,** 908 (2017).

56. Ricicova, M., Hamidi, M., Quiring, A., Niemistö, A., Emberly, E. & Hansen, C. L. Dissecting genealogy and cell cycle as sources of cell-to-cell variability in

MAPK signaling using high-throughput lineage tracking. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 11403–8 (2013).

57. The MathWorks Inc. MATLAB and Statistics Toolbox Release 2012b. (2012).

58. Rep, M., Krantz, M., Thevelein, J. M. & Hohmann, S. The transcriptional response of Saccharomyces cerevisiae to osmotic shock. Hot1p and Msn2p/Msn4p are required for the induction of subsets of high osmolarity glycerol pathway-dependent genes. *J. Biol. Chem.* **275,** 8290–300 (2000).

59. Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. A., Davey, R. P., Roberts, I. N., Burt, A., Koufopanou, V., Tsai, I. J., Bergman, C. M., Bensasson, D., O'Kelly, M. J. T., van Oudenaarden, A., Barton, D. B. H., Bailes, E., Nguyen, A. N., Jones, M., Quail, M. A., Goodhead, I., Sims, S., Smith, F., Blomberg, A., Durbin, R. & Louis, E. J. Population genomics of domestic and wild yeasts. *Nature* **458,** 337–41 (2009).

60. Brem, R. B. & Kruglyak, L. The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 1572–7 (2005).

61. Saito, H. & Posas, F. Response to hyperosmotic stress. *Genetics* **192,** 289–318 (2012).

62. Wurgler-Murphy, S. M., Maeda, T., Witten, E. A. & Saito, H. Regulation of the Saccharomyces cerevisiae HOG1 mitogen-activated protein kinase by the PTP2 and PTP3 protein tyrosine phosphatases. *Mol. Cell. Biol.* **17,** 1289–97 (1997).

63. Mendizabal, I., Rios, G., Mulet, J. M., Serrano, R. & de Larrinoa, I. F. Yeast putative transcription factors involved in salt tolerance. *FEBS Lett.* **425,** 323–8 (1998).

64. Blondel, M., Alepuz, P. M., Huang, L. S., Shaham, S., Ammerer, G. & Peter, M. Nuclear export of Far1p in response to pheromones requires the export receptor Msn5p/Ste21p. *Genes Dev.* **13,** 2284–300 (1999).

65. Titov, A. A. & Blobel, G. The karyopherin Kap122p/Pdr6p imports both subunits of the transcription factor IIA into the nucleus. *J. Cell Biol.* **147,** 235–46 (1999).