Winter 2021

Mapping Genetic Variants Associated with Dynamic Protein Abundance in Haploid Yeast

Tanner Thuet-Davenport
Western Washington University, thuetdt@wwu.edu

Mapping Genetic Variants Associated with Dynamic Protein Abundance in Haploid Yeast

By

Tanner Thuet-Davenport

Accepted in Partial Fulfillment
Of the Requirements for the Degree
Master of Science

ADVISORY COMMITTEE

Dr. Dan Pollard

Dr. Lynn Pillitteri

Dr. Matthew Zinkgraf

GRADUATE SCHOOL

David L. Patrick, Dean

Master's Thesis

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Western Washington University, I grant to Western Washington University the non-exclusive royalty-free right to archive, reproduce, distribute, and display the thesis in any and all forms, including electronic format, via any digital library mechanisms maintained by WWU.

I represent and warrant this is my original work, and does not infringe or violate any rights of others. I warrant that I have obtained written permissions from the owner of any third party copyrighted material included in these files.

I acknowledge that I retain ownership rights to the copyright of this work, including but not limited to the right to use all or part of this work in future works, such as articles or books.

Library users are granted permission for individual, research and non-commercial reproduction of this work for educational purposes only. Any further digital posting of this document requires specific permission from the author.

Any copying or publication of this thesis for commercial purposes, or for financial gain, is not allowed without my written permission.

Tanner Thuet-Davenport
1-27-2021

...

Mapping Genetic Variants Associated with Dynamic Protein Abundance in Haploid Yeast

A Thesis
Presented to
The Faculty of
Western Washington University

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science

by
Tanner Thuet-Davenport
January 2021

Abstract:

As organisms respond to changes in their environment genetic variation between individuals can directly affect organismal trait phenotypes by altering gene expression. Historically, studies have focused on the effect of genetic variation on mRNA synthesis (transcription) and decay rates. Relatively few studies have probed the relationship between DNA variants and protein-specific regulation of individual genes, despite the plethora of evidence that RNA levels are often poor proxies for protein levels. No study to date has mapped genetic variation associated with dynamic protein levels. In this study we investigated the location and identity of genetic variants acting on protein expression dynamics for the genes Fig1, Fus3, and Tos6 during mating pheromone response in isolates of the budding yeast *Saccharomyces cerevisiae*. We classified protein level variation as either local (driven by variation within the locus of the gene of interest) or as *trans*-acting (driven by variation elsewhere in the genome) by swapping gene of interest alleles between lab and clinical strain isolates. We previously found that Fig1p protein levels are controlled by trans-acting genetic variants and in this study we found evidence of local effects acting on TOS6 protein abundance but could not disentangle local from *trans* effects for the Fus3 gene. To map quantitative trait loci associated with Fig1 protein level variation (pQTL) during mating pheromone response we used a novel time-based Bulk Segregant Analysis (BSA) approach combined with Fluorescence Activated single-Cell Sorting (FACS) and Next-Gen Sequencing (NGS). These findings demonstrate the value of mapping protein expression dynamics and shed light on the complex nature of genotype-phenotype relationships in natural populations.

Table of Contents

Introduction:

Genetic diversity acts as the raw material for natural selection and evolution as populations respond to changing environments. Some individuals in a population possess genetic variants that are better suited for the selective pressures imparted by their environment (Morjan et al. 2004). Thus, genetic variability within a population allows populations to adapt and survive. DNA is used as an instruction manual for producing each of the thousands of proteins that are required for cells to operate properly. This flow of genetic information from DNA to final protein product is commonly referred to as gene expression, and differences in gene expression result in the observable differences between living organisms (Stern and Orgogozo, 2008). DNA instructions differ ever so slightly between individuals in a population, and differ more significantly between divergent species. The disparities observed in DNA sequences can result in unique versions of functional proteins, or in varying amounts of protein being produced by the cell. These differences at the protein level manifest into the observable variation seen in the traits of individuals in a population, or in an individual's susceptibility to a given disease (Schadt et al. 2005). Moreover, natural variation in when, where, and how much genes are expressed has been associated with ecological adaptation and has direct implications in evolutionary biology (Shapiro et al. 2004), (Wittkopp et al. 2004), (Lackner et al. 2012).

Deciphering the genetic code has improved our understanding of how variation at the DNA level results in different versions of proteins, however relatively less is known about how genetic differences affect when or how much a given protein is expressed. Furthermore, studies have suggested that sequence divergence is far more common in DNA regions that control

expression patterns than in orthologous protein coding regions between closely related species (Wittkopp et al. 2008), (Dermitzakis and Clark, 2002), (Borneman et al. 2007), suggesting that phenotypic differences between divergent species are more a result of differential expression patterns than expression of unique proteins (Ronald et al. 2005), (Schmidt et al. 2010). Four central processes control the concentration of proteins in a cell: transcription (mRNA synthesis), mRNA decay, translation (protein synthesis), and protein decay (Figure 1). While in theory, differences at the DNA level should be able to affect the rate of all four of these processes, most research has exclusively focused on the genetics involved in differential mRNA synthesis and differential mRNA decay rates (Albert and Kruglyak, 2015). Affordability and relative ease of mRNA sequencing has made quantitative transcript analysis a more attractive avenue for studying gene expression than quantitative proteomic analysis (Albert et al. 2014), (Foss et al. 2007), (Marguerat et al. 2012). However, evidence suggests that mRNA expression profiles don't always agree with protein expression profiles (Straub et al. 2011), (Ghazalpour et al. 2011), (Foss et al. 2011), (McManus et al. 2012), (Pollard et al. 2016), and recent studies on global mRNA and Protein abundances show only modest correlations ($r \sim 0.4$) in both humans (Gry et al. 2009) and yeast (Brion et al. 2020) suggesting that a major gap exists in our understanding of gene expression regulation. Further, since proteins, not mRNA transcripts are the functional molecular form of a given gene, understanding the mechanisms that govern differential protein expression provides greater insight into the heritable variation involved in cellular physiology.

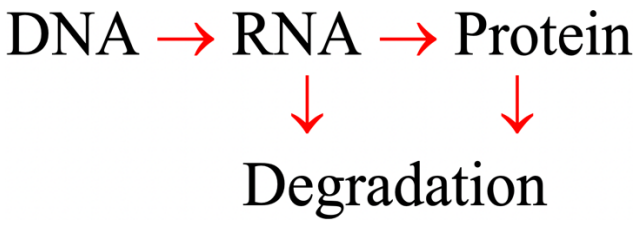


Figure 1: The central dogma of molecular biology, where each red arrow represents a cellular process in which DNA variation can act on gene expression dynamics.

Heritable phenotypic traits can be influenced by a variety of genetic processes which together make up the genetic architecture of a trait. Understanding these underlying mechanisms that contribute to the genetic architecture of phenotypic traits is a major focus in modern day genetics (Timpson et al. 2017) (Pomp et al. 2004). The architecture for most heritable traits is genetically complex meaning these traits are influenced in part by a number of genetic variants acting simultaneously. Making matters more complex, genetic variants that affect expression of individual genes can either be local (near the gene in question) or located in distant regions of the genome. Allele-specific expression assays allow for the genetic mechanisms underlying expression differences to be teased apart by directly comparing the expression of alleles within the same cellular context. Allele-specific expression differences are an indication of *cis*-acting variation (differences driven by variation within the gene of interest allele) while lack of allele-specific effects is usually interpreted as evidence of *trans*-acting variation (driven by variation elsewhere in the genome) (Wittkopp et al. 2004), (Knight, 2004) (Salinas et al. 2016) (Khan et al. 2012) (Figure 2). Fluorescent gene tags enable the quantification of single-cell protein abundance of individual genes (Huh et al. 2003) (Rines et al. 2002) and research employing chimeric gene of interest (GoI) allele swaps have been able to tease apart allele specific differences from genomic background differences on the expression of many phenotypic traits

(Sadhu et al. 2016) (Deutschbauer and Davis, 2005). A similar allele-swap methodology combined with quantitative fluorescent microscopy and a reporter-tagged GoI can be employed in order shed light on the genetic architecture of protein expression phenotypes (Figure 3). Although these experiments cannot distinguish between *cis*-acting variation and local variation acting by a *trans* mechanism, this assay can tell us whether GoI expression is influenced primarily by local polymorphisms or distant *trans*-acting factors. For alleles where local variation drives expression differences individual causal polymorphisms can be fine-mapped directly via subsequent rounds of systematic chimeric allele swaps, but for variants acting on expression in *trans* genome-wide mapping is necessary prior to fine mapping causal variants.

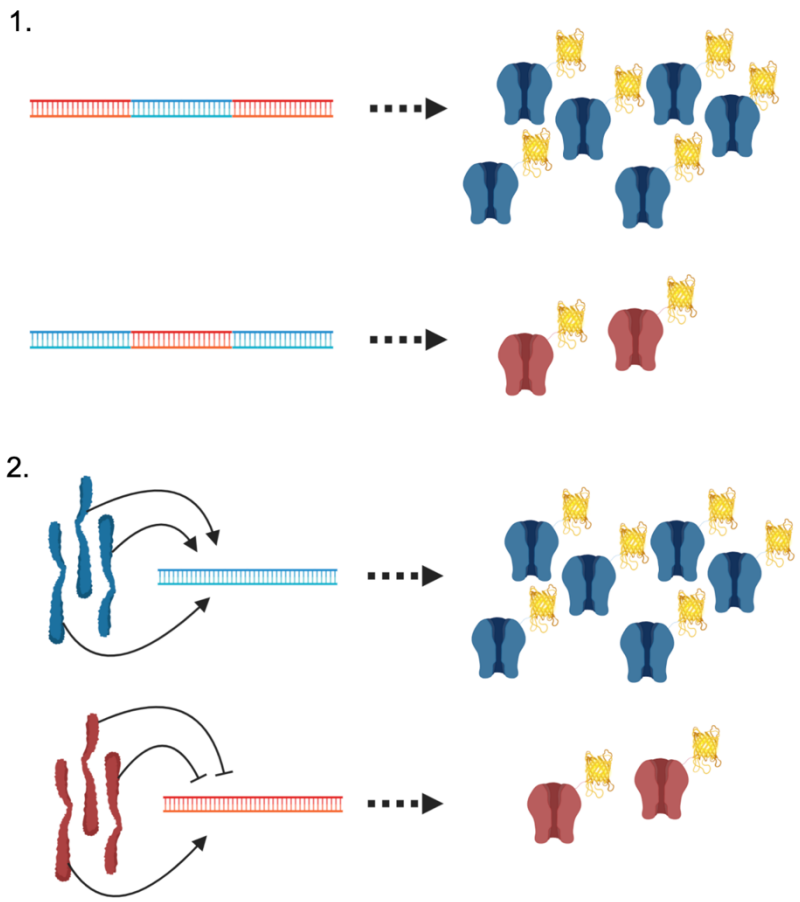


Figure 2: Illustration depicting mechanistic differences between local and distant genetic interactions contributing to variable protein level expression. 1) Local level effects where variation within the GoI allele primarily contributes to variable GoI protein expression. 2) Distant or *trans* level effects where variation in distant regions of the genome contribute to variable GoI protein expression.

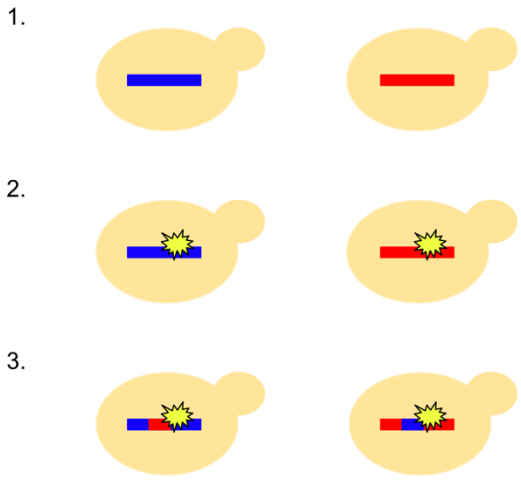


Figure **3**: Example illustration of the systematic allele swap experimental framework where the expression of GoI alleles are compared across genomic backgrounds. 1) Native strain without fluorescence reporter attached to the gene of interest. 2) Native strain background with gene of interest locus replaced with native strain chimeric gene of interest fluorescent reporter allele. 3) Allele swap strains with reciprocal strain's chimeric gene of interest fluorescent reporter inserted into the otherwise native strain genomic background.

While individual *cis*-acting variants often have much larger phenotypic effects than individual *trans*-acting variants, *trans* variants explain the majority of the variance for the expression of most genes (Metzger et al. 2016) (Albert et al. 2018) (Signor and Nuzhdin, 2018). Quantitative Trait Loci (QTL) are genomic locations that harbor polymorphisms associated with a given quantitative (complex) trait. Bulk Segregant Analysis (BSA) is a highly efficient method for mapping QTL in which large numbers of individuals are sorted according to phenotype and subsequently bulk DNA-sequenced (Duveau et al. 2014) (Edwards and Gifford, 2012) (Salunkhe et al. 2011) (de Vries et al. 1998) (Ehrenreich et al. 2010) (Figure 4). Recent BSA studies in yeast have identified pQTL acting on steady-state protein levels of thousands of genes (Albert et al, 2014) (Picotti et al, 2013) (Parts, 2014), and research on mRNA level QTL (eQTL) have found that nearly all steady state *trans*-acting variants cluster into just 102 eQTL hotspots (Albert et al,

2018). Counterintuitively, eQTL often lack corresponding protein level pQTL suggesting that mechanisms controlling cellular protein levels can act independently of transcriptional regulation. Together these findings point towards a class of polymorphisms that influence gene expression specifically at the protein level, motivating further research into the genetic architecture underlying protein abundance of specific genes.

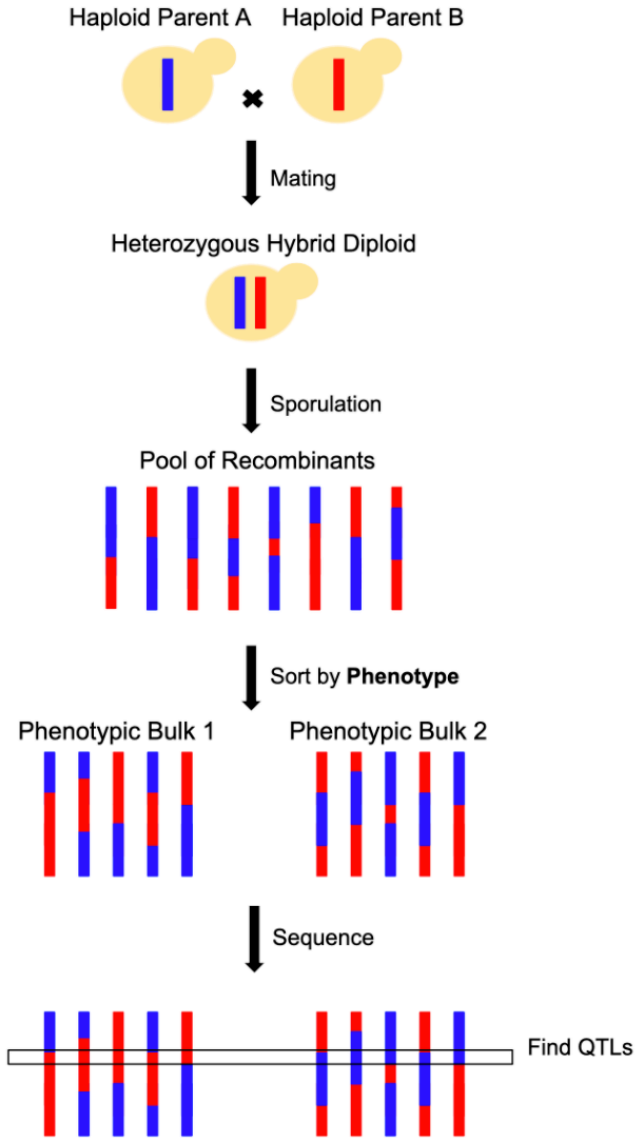


Figure **4**: Schematic depicting the Bulk Segregant Analysis experimental framework. First haploid parents are mated in order to generate diploid hybrids. Diploid hybrids are subsequently sporulated to generate a pool of haploid recombinants which are then sorted by phenotype into separate bulks and independently bulk-sequenced

To this point quantitative gene expression research has focused predominately on gene expression under steady state conditions. Despite the abundance of critical biological processes that rely on dynamic gene expression responses such as development (Verd et al. 2017) (Assoul et al. 2010), aging (Pascual-Ahuir et al. 2019) (Seroude et al. 2002), and responding to environmental perturbation (Hook et al. 2007) (de Nadal et al. 2011) (Strassburg et al. 2010) the extent to which genetic differences affect differences in dynamic gene expression processes is still poorly understood. And while some dynamic states have been relatively well studied, most research has focused exclusively on the transcriptional side of gene expression (Delile et al. 2019) (Gloss et al. 2017) (Hook et al. 2010) (Miller et al. 2011). Moreover, whether the QTL that are known to regulate steady-state mRNA and protein levels extend to non-steady-state conditions has yet to be thoroughly investigated. In this study we employed Allele Specific Expression assays as well as Bulk Segregant Analysis pQTL-mapping in order to probe the genetic architecture underlying observed dynamic gene expression differences between populations of Brewer's yeast, *Saccharomyces cerevisiae.*

Background

Brewer's yeast is an ideal model organism for studying general Eukaryotic cellular processes due to its rapid reproduction rate, unique haploid-diploid reproductive cycle, its relatively small and fully sequenced genome, and the plethora of established experimental protocols available for yeast research, most notably reliable genome editing via homologous recombination. (Liti et al. 2009) (Liu et al. 2017) (Roberts et al. 2004). While yeast are similar to familiar model organisms such as *Escherichia coli* in that they are unicellular, robust, and easy to maintain in a lab setting, they are Eukaryotic and despite ~1.5 billion years of evolutionary divergence, 1/3 of their ~6500 genes have direct homologs to human genes. Furthermore, as a result of geographic and ecological isolation *S. cerevisiae* has a complex population structure with thousands of genetically divergent strains (Schuller et al. 2012) (Peter et al. 2018), making yeast a uniquely appropriate system for studying the heritable nature of complex traits.

Brewer's yeast have a life cycle in which individuals can alternate between haploid and diploid cell stages (Figure 5). Both diploid and haploid cells are stable and capable of reproducing via mitosis. Haploid cells belong to one of two mating types, mat-a and mat- **α** (alpha), where cells respond to and mate with haploid cells of the opposite mating type (Haber, 2012). When haploid cells detect the opposite mating type's mating pheromone, they respond by growing a shmoo in the direction corresponding to the highest concentration of detected pheromone. The mat-a and mat- **α** haploid cells join at the site of the shmoo and form a diploid cell. While diploid cells are stable, under stressful environmental conditions diploid cells can undergo meiosis and produce 4 recombinant haploid spores contained within a single ascus structure.

Mating pheromone regulated genes are dynamically expressed (Erdman et al. 1998) (Paliwal et al. 2007), and genetic variation between closely related yeast populations contribute to observed transcriptional differences between pheromone treated cells (Zheng et al. 2010). The deeply conserved MAPK pathway is critical in regulating yeast cellular physiology and plays a critical role in the switch between vegetative growth and mating physiological states (Herskowitz, 1995). Research has found that the MAPK pathway shows differential regulation between closely related yeast strains (Treusch et al. 2015) (Chen and Thorner, 2007), and has uncovered causal polymorphisms in common lab strains with respect to ancestral yeast populations that effect downstream MAPK signaling and mating physiology (Lang et al. 2009).

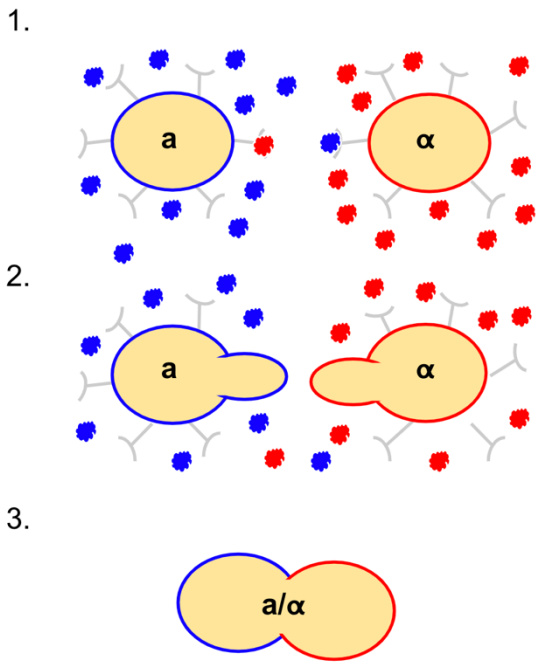


Figure 5: Simplified illustration of yeast mating process. 1) Haploid yeast belong to one of two mating types, mat-a and mat-alpha. Haploid cells produce and release mating pheromone into the environment. 2) Cells respond to the opposite mating types' mating pheromone by growing a schmoo in the direction of the highest concentration of detected environmental pheromone. 3) Cells join at the site of the schmoo and fuse to become diploid hybrid cells.

The mating pheromone response pathway in yeast is an ideal model system for studying differential gene expression dynamics because cells are responding to their environment by coordinating the expression of a cascade of genes. Under a dynamic gene expression model, the expression of specific genes is both quantitatively and temporally regulated in response to their environment (Hecker et al. 2009). Prior studies focusing on environmental stress response in yeast have found evidence of differential transcription of distinct groups of genes at different timepoints following various forms of environmental perturbation (Dong et al. 2017) (Sethiya et al. 2019). As described above, a growing body of research focused on natural variation in steady state gene expression has suggested that mRNA abundance is often an insufficient proxy for protein abundance for many genes (Parts et al. 2014) (Vogel and Marcotte, 2012). Meanwhile, whether mRNA abundance is the key player in explaining protein abundance under non-steady state conditions - such as mating pheromone response pathways – has only recently begun to be explored (Pollard et al. 2016). Moreover, the degree to which the genetic architecture of steady-state and dynamically expressed genes overlap is so far yet to be determined.

Two strains of *S. cerevisiae,* S288c a common lab strain, and YJM145 a clinical strain originally isolated from the lungs of an immuno-suppressed AIDS patient, are known to vary in their response to mating pheromone. These strains differ by ~60,000 genomic polymorphisms, show differences in dynamic expression of pheromone responsive genes, and therefore present an ideal system for probing the genetics underlying differential dynamic gene expression phenotypes.

Prior research in the Pollard Lab has identified a group of genes with protein expression dynamics that differ between lab and clinical strains during pheromone response (Pollard et al. 2016) (Figure 6)**.** This research focuses on elucidating the genetic architecture and molecular mechanisms underlying these observed differences in dynamically expressed pheromone response genes between strains. Fig1 is a membrane bound protein involved in calcium influx that is locally expressed at the site of the shmoo (Muller et al. 2003) (Cavinder and Trail, 2012), and is essential for yeast mating (Erdman et al. 1998). Fus3 is a protein kinase that is a component of the mitogen active kinase pathway that transduces the pheromone signal (Elion et al. 1993). Tos6 a glycosylphosphatidylinositol (GPI)-dependent cell wall protein (Hamada et al. 1998). Fig1, Fus3, and Tos6 each express more protein molecules per RNA molecule in the S288c strain compared to the YJM145 strain (Pollard et al. 2016) (Figure 6). These protein-level differences suggest either higher protein synthesis rates, lower protein decay rates, or both in S288c compared to YJM145. Cycloheximide decay rate assays confirm that decay rates are indeed significantly lower in S288c compared to YJM145 and differential equation modeling supports the hypothesis that protein synthesis rates are significantly higher in S288c compared to YJM145. Allele-specific expression assays on the Fig1 gene did not find evidence of allele specific expression, which implies that Fig1 protein expression differences between strains is due to genetic variants that are located outside of the Fig1 locus and act in *trans*. This thesis project is an effort toward further elucidating the genetic architecture and molecular mechanisms underlying the differences in protein expression dynamics for these three genes.

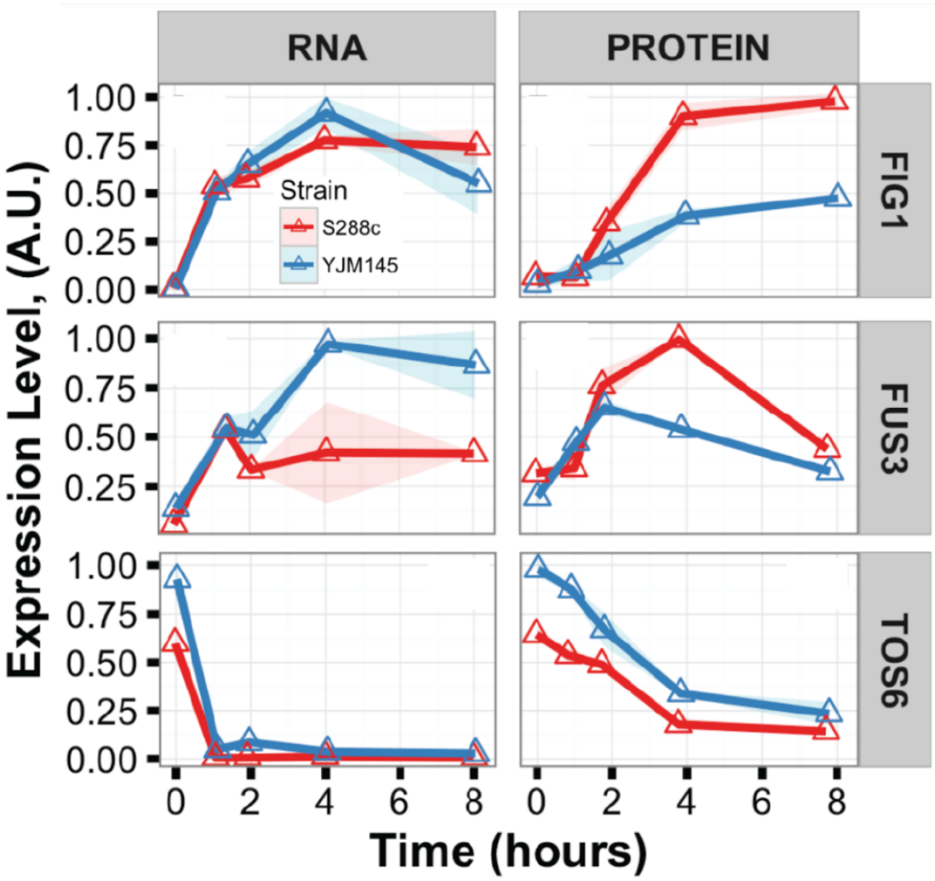


Figure 6: Example RNA and Protein expression time courses illustrating protein per RNA differences for the Fig1, Fus3, and Tos6 genes following addition of **α** -factor mating pheromone at timepoint zero in both S288c and YJM145 native strain backgrounds. Red lines represent S288c while blue lines represent YJM145. Triangles represent mean expression while shadows represent 95% confidence intervals. Plots are rescaled to arbitrary units with a maximum value of one for visualization purposes.

For the first step towards illuminating the genetic architecture underlying divergent protein synthesis and/or decay rates for these genes, we characterized protein expression variation as either being subject to locally-acting or distantly-acting genetic variants. As stated above, Fig1 had already been identified as having distant trans-acting genetic variation (Pollard et al. 2016). Here, we performed allele swaps combined with fluorescence microscopy for Fus3 and Tos6 to tease apart whether protein level differences are driven by local or *trans*-acting genetic

variation. For the Fus3 gene we failed to uncover consistent expression patterns, preventing categorization of the location of regulatory genetic variants. However, for Tos6 we found evidence of local variation affecting protein abundance.

For the second step towards understanding the genetic architecture and ultimately the molecular mechanisms underlying divergence in protein expression dynamics, we used a bulk segregant approach to map the genomic locations of genetic variants acting in trans on Fig1 protein expression. We generated a recombinant segregant population through multiple generations of random mating between S288c and YJM145 isolates, selected high and low expressing bulks of cells at multiple time points during pheromone response using fluorescence active single-cell sorting, and identified alleles associated with Fig1 expression levels from next generation sequencing of the selected bulks. We identified a collection of novel *trans*-acting QTL affecting FIG1 protein abundance and intriguingly found evidence of QTL that differentially contribute to FIG1 protein levels across different pheromone response stages.

Methods:

Characterizing Expression Variation as Local or Trans-acting

Strain Engineering

Allele swap strains were constructed using a targeted mutagenesis approach (Gietz and Schiestl, 2007) with selection for specific genetic markers. Alleles were swapped in two steps by first deleting the GoI locus in each genomic background before inserting the reciprocal strain allele. First, the native allele was knocked out using a KanMX knockout cassette by targeting the 5' and 3' sequences immediately flanking the GoI locus (Table 1). After knocking out the native allele, a second transformation using targeted mutagenesis inserted a strain specific GoI allele cassette with a 3' tethered yECitrine fluorescent protein domain along with an independently transcribed Ura3 selectable marker. This effectively replaced the native GoI locus with the reciprocal strain's GoI locus attached in an open reading frame to a yECitrine fluorescent protein domain. Transformants were first selected by plating on either KanMx or -Ura minimal media, passaged on rich YPD media, and then selected a second round on the appropriate selective media. Following the second round of selection transformant colonies were confirmed by PCR before being frozen as -80 degree C glycerol stocks.

Target Loci	Upstream Homology Tag	Downstream Homology Tag	Selectable Marker (plasmid ID)
Tos6	CGGATCTTTTGTGTTGCTTGGAAGTGTGATCAATACCCAT	GAAGCGGAATATCCTTCCTAGTTTAAGTTGTCCATGCAAC	loxp.KanMX.loxp (pUG6)
Tos6	CGCACGCAACGATACTAAGA	TAACAAAGCAAAGGCAGCAG	pTEF-CaUra3 (pDAP4)
Fus3	GTAAGGCCCAAAGAGAATAGACAAAATGAAGTAATATCAT	TACATTGTTCTTCGGGTTGATATTTTAATGATAATGATGG	loxp.KanMX.loxp (pUG6)
Fus3	GGGTGAATTCTTCGGCATTA	ACTAAATATTTCGTTCCAAA	pTEF-CaUra3 (pDAP4)

Table 1: Homology tags and selectable markers that were used in each allele swap genomic transformation. GoI loci were first knocked out using a KanMX selectable marker, and the reciprocal strain specific GoI allele was knocked in using a Ura3 selectable marker. Plasmid IDs of each selectable marker are given when applicable.

Microscopy Prep

Strains were streak plated from their respective -80 degree C glycerol stocks and grown for 2-4 days. Single colonies were then transferred into low fluorescence media, and grown for 6-8 hours in a 30 degree C shaking incubator. Cultures were then diluted back and grown overnight into early log-phase at 30 degrees C. The following day, cultures were diluted to an OD of ~0.2, and subsequently treated with alpha-factor mating pheromone to a final concentration of 50 nM. Treated cell cultures were placed back into the 30 degree C shaking incubator for ~3.75 hours, until they were loaded into a glass bottom 96-well plate. Plates were treated with

Concanavalin A to adhere cells to glass bottom wells. Following addition of cells, plates were transferred to the Leica Fluorescence Microscope for imaging.

Image Acquisition

Four hours after initial pheromone treatment, Concanavalin A adhered cell cultures were placed into a custom temperature-controlled microscope housing so that the Leica Fluorescence Microscope environment maintained a constant temperature of 30 degrees C. Cells were imaged at 63x magnification using a brightfield 21-image z-stack paired with a single YFP fluorescent filter image acquired at 150 milliseconds and gain set at 6.

Microscopy Analysis Script

Each image was analyzed using a custom microscopy image analysis pipeline. First, a segmentation script used a 21-image z-stack series to define the boundaries of each cell in an image. Next, the YFP filter image matching a given z-stack was used to quantify the per cell fluorescence for each cell in the image. Finally, fluorescence was normalized against the autofluorescence of each respective native strain. The normalized per-cell fluorescence was averaged across each strain to determine strain-specific mean fluorescence strength for each experimental replicate.

Allele Swap Statistics

Normalized per-cell fluorescence microscopy results were used to compare expression levels between native allele and allele swap strains. To assess the effects that strain background and GoI allele have on fluorescent reporter expression levels two-way ANOVA models including interaction effects were constructed using the aov function in R version 3.5.3 (R Core Team, 2019). Native strain and allele swap strain RNA and protein level ratios were compared via Welch's t-test using the t.test function in R version 3.5.3.

Mapping Trans Acting Variation

Strain Construction

Fig1 QTL mapping strains were constructed using a combination of targeted mutagenesis and plasmid-programmed genomic recombination. To select for Mat-a cells capable of responding to **α**-factor each strain was first modified to procure a histidine auxotrophy. Both lab and clinical backgrounds had their His3 loci knocked-out using a G418-resistance (KanMX) cassette targeting the 5' and 3' sequences immediately flanking the His3 locus (Table 2). The KanMX knockout cassette was then excised from the His3 locus of the lab background (S288c) using Cre-Lox recombination. Next a modified synthetic genetic array methodology (Yan Tong and Boone, 2006) (Costanzo and Boone, 2009) was employed in order to select for Mat-a individuals by inserting a chimeric ste2-His5_Sp cassette into the Can1 locus of the lab background. The ste2 promoter is only transcriptionally active in Mat-a cells and the His5 allele from *Schizosaccharomyces pombe* has analogous functionality to the *S. cerevisiae* His3 gene,

permitting selection of Mat-a individuals on -histidine growth media. The intact KanMx allele in the clinical background, if combined with a selectable marker in the lab background, enables selection of mated diploids on double selection media. To take advantage of this feature, a Hygromycin-resistance (Hph) cassette was inserted into the yercΔ8 locus of the lab background via a targeted mutagenesis approach using homology tags directed at the sequences immediately 5' and 3' of the yercΔ8 locus.

Target Loci	Upstream Homology Tag	Downstream Homology Tag	Selectable Marker (plasmid ID)
His3	TATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCA	TTCATAGGTATACATATATACACATGTATATATATCGTAT	loxp.KanMX.loxp (pUG6)
Fig1	CAGTAATGGCTTGGTTTAGCTTTG	CAGACGGTAATGATTAGAGTTTAGGT	pTEF-CaUra3 (pDAP4)
Can1	AGAGAATGATACGAGATAAAGCACA	TTGTCAATTCAAACTCCGTTCTAAG	Ste2-His5_Sp
yercΔ8	CCCAGTTGTTTGTAGCTGGTTCATATTTAGCGGCAAT	TTGTTGGCATTCCATTGTTGGGAGAGGCTATTATATC	pTEF-Hph (pDAP6)

Table 2: Homology tags and selectable markers that were used in each BSA strain genomic transformation. Plasmid IDs of each selectable marker are given when applicable.

The clinical background (YJM145) had its mating type switched from Mat-a to Mat-**α** via transfecting a plasmid containing a galactose driven HO endonuclease. Mat-**α** colonies were identified by Halo-assay, a pheromone production assay where pheromone-sensitive mutant test strains DBY7442 and DBY7730 experience mitotic growth arrest upon exposure to the

opposite mating type's mating pheromone, permitting inference of individual cell's mating types via presence or absence of test-strain zones of inhibition on YPD growth media. Finally, both lab and clinical backgrounds had their native Fig1 locus replaced with strain specific Fig1 reporter cassettes. These reporter cassettes preserved the native Fig1 coding sequence of each strain, while also adding a tethered yECitrine fluorescent protein domain and an independently transcribed Ura3 sequence in order to select for transformants on -Ura media. After each genomic modification, transformant cells were confirmed via colony-PCR genotyping and were then frozen as -80 degree C glycerol stocks.

Mating and Sporulation

S288c and YJM145 strains were subjected to two and five rounds of mating and sporulation in order to increase the frequency of genomic recombination events and thus increase mapping resolution (Parts et al. 2011) (Magwene et al. 2011). The lab and clinical (mat-a and mat-**α** respectively) strain backgrounds were mated in a droplet of water on YPD media plates. After mating, cells were streaked onto G418 + Hygromycin B double selection media in order to select for mated diploid colonies. Diploid colonies were then transferred and grown at ~23 degrees C for 5-7 days in nutrient poor sporulation media to induce sporulation of diploid individuals. A modified sporulation procedure was adapted from Goddard et al. 2005. After 5-7 days in nutrient poor sporulation media cells were rinsed, resuspended and incubated at 37 degrees C for 10 minutes in 1 unit/25uL Zymolyase, followed by a two-hour 37 degree C incubation in 1% SDS in order to digest spore ascii. Following digestion, Eppendorf tubes containing sporulated cells were submerged in water and sonicated using a Torbéo 36810

Series ultrasonicator on low setting for 3 minutes in order to dissociate spores from ascii. Dissociated spores were then resuspended in rich YPD growth media and grown at 30 degrees C overnight without shaking to allow cells to settle and proceed through an additional round of mating. For one iteration of the mapping experiment the mating and sporulation protocol was repeated 4 times for a total of 5 rounds of mating and sporulation, while for the second iteration of the mapping experiment the mating and sporulation procedure was repeated once for a total of 2 rounds of mating and sporulation.

FACS Preparation

Following the final round of mating and sporulation, dissociated spores were suspended in -His + Canavanine double selection media and grown overnight in order to select for Mat-a segregants. After selecting for mat-a individuals, cells were grown to early log-phase in low-fluorescence media, before being treated with **α**-factor mating pheromone to a final concentration of 50nM. Pheromone treated cultures were grown for either 2 hours or 5 hours prior to sorting.

FACS sorting

Cells were sorted on a Nanocellect WOLF Fluorescence Activated Single-Cell Sorter. Gates were carefully drawn to exclude doublets, extremely large and small cells, and non-fluorescing cells (Figure 7). Doublets were excluded by gating out cells with a large FSC-width to FSC-height ratio. Cells near the tails of the FSC-height distribution were excluded in order to only retain the ~68% of intermediately sized cells. Finally, non-fluorescing cells were gated out by normalizing

against S288c cells lacking a yECitrine fluorescent reporter domain. Selection-gates were drawn to reflect the known relationship between cell size and cell fluorescence (Duveau et al. 2014). 3-4% cell fractions were collected from each high and low fluorescent tails, while omitting the highest 2% and lowest 2% of individuals to ensure that artifacts such as doublets and unhealthy cells were excluded. Cells were sorted and collected in 30-minute time intervals, corresponding to the 15 minutes before and after the target 2 hour and 5 hour timepoints. Between 5,000 and 6,000 cells were collected into each bulk from the 5 rounds of random mating and sporulation population and between 10,000 and 11,000 cells were collected into each bulk from the 2 rounds of mating and sporulation populations.

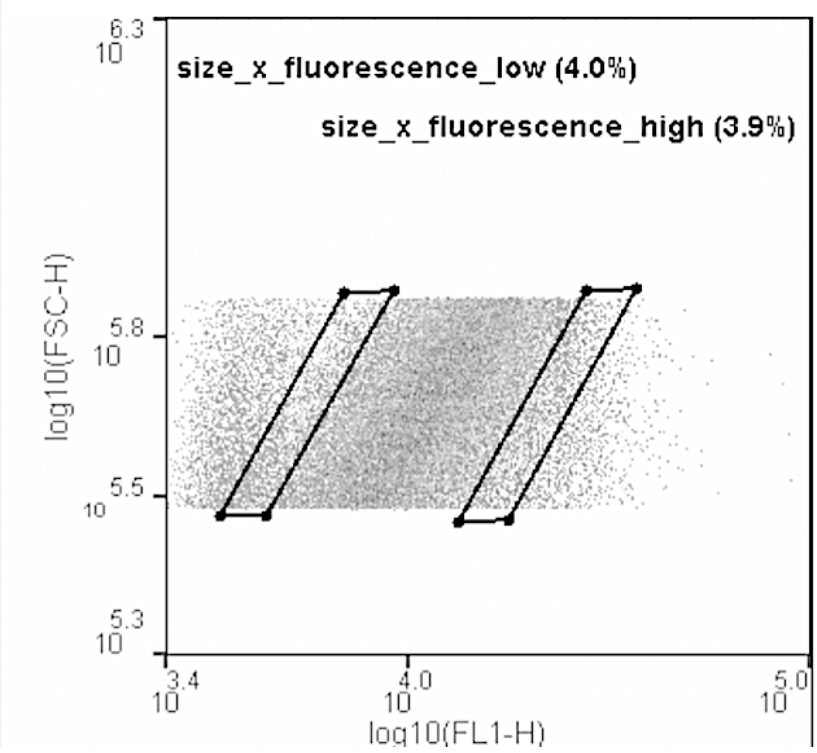


Figure 7. Collection gates were carefully designed to reflect the known relationship between cell-size and protein expression (represented here by FSC-H and FL1 respectively). Special attention was placed to exclude the most extreme cells to avoid collecting experimental artifacts.

Bulk DNA Prep

Following sorting, pooled cells were spun down and rinsed to remove residual mating pheromone, before being resuspended in YPD and grown for 18-30 hours until cultures reached mid to late-log phase. Cells were then pelleted and subjected to phenol:choloroform:isoamyl DNA extraction (CSHL, 2015).

Sequencing

Bulk-pooled DNA library preparation and sequencing was performed by Psomagen, INC. sequencing services. Libraries were prepared using an Illumina Truseq PCR-Free Library Prep kit and bulk sequencing was performed using a NovaSeq6000 S4 150-bp Paired End sequencing platform with ~3Gb production target. Per-bulk DNA Integrity Number values (a measure of DNA quality) ranged from 6.5 to 9.8 with a mean value of 7.7 providing relatively intact, high quality genomic DNA fragments for sequencing.

Read Processing

Bulk-pooled raw sequence reads with genome wide sequencing depths ranging from 339x to 1800x coverage per-pool were processed prior to QTL-mapping analysis. Raw sequence read adapters were trimmed using Scythe (Buffalo, 2013) and reads were quality trimmed using Sickle (Joshi, 2011). Trimmed reads were processed as described in GATK variant calling best practices (DePristo et al., 2011). First, reads were aligned to the S288c R64 reference genome using BWA (Li, 2013) to create aligned bam files for each pool. Unaligned trimmed reads were converted to unmapped bam format and then merged with the aligned bam reads using Picard

(Broad Institute, 2019) to conserve the per-read metadata required for downstream variant calling. The merged reads were indexed using Samtools (Li, 2009) and then compared to the S288c reference genome using GATK HaplotypeCaller (Van der Auwera et al., 2013) to generate pool specific VCF files. Per-sample VCFs were then indexed using GATK IndexFeatureFile and High and Low FACS-sorted bulk VCF files were merged using GATK CombineGVCFs to generate whole-experiment VCF files. The whole-experiment VCF files were then joint-called using GATK GenotypeGVCFs in order to genotype high and low pools together and create joint VCF files. The joint-genotyped VCF files were then exported into tab delimited tables using GATK VariantsToTable.

QTL Identification

The joint-genotyped VCF table was then further processed in the R-package QTLseqr (Mansfeld, 2018) to filter SNPs and develop whole-genome QTL maps. First, low confidence SNPs, SNPs with low depth in both bulks, and SNPs where the reference (lab strain-S288c) allele was overrepresented were filtered so that SNPs better fit a null-G' distribution using the QTLseqr filterSNPs function to produce VCF files with 150x to 2500x coverage across each SNP in the genome. Outliers were identified using the deltaSNP method at a cutoff of 0.05. Using the QTLseqr runGprimeAnalysis function a tricubed-smoothed G' statistic was calculated for each SNP using a sliding window sizes of 30 kb and 50 kb for segregant populations derived from five-rounds and two-rounds of random mating respectively. QTL summary statistics were called for QTL with an FDR cutoff rate of 5% using QTLseqr getQTLTable. Sliding window sizes were derived by first estimating the recombinant and the non-recombinant fractions of the

genome at each additional round of random mating, and then estimating the average physical distance between recombination events throughout the genome. QTL were considered overlapping if any portion of the significant QTL regions overlapped and the entire union of significant intervals was considered for comparison of replicated QTL between experiments.

Comparison to Existing QTL

Prior gene expression-based BSA studies in yeast have uncovered pQTL hotspots (loci implicated in effecting protein abundance of many genes) (Albert, 2014), eQTL hotspots (loci implicated in effecting transcript abundance of many genes) (Albert, 2018), as well as pQTL lacking cognate eQTL. In order to compare these existing QTL sets to the Fig1p QTL a bash script was used to find each existing steady-state QTL peak's nearest SNP across each Fig1 experimental data set and the G' statistic for each QTL's nearest SNP was recorded for each Fig1 BSA replicate. In order to test whether the Fig1 pQTL were enriched for existing QTL, I counted the number of existing QTL peak's nearest SNPs that were significant in each Fig1 experiment as well as the total number of SNPs significant in each Fig1 experiment. A hypergeometric test was then performed for each Fig1 SNPset independently using the R phyper function (R Core Team, 2019) to test for enrichment of significant SNPs. In order to determine whether the G' values for the existing steady-state QTL peak's nearest SNPs were higher than the G' values of randomly selected SNPs, an R script was used to randomly sample SNPs from each Fig1p SNPset and sum their G' values over 10,000 iterations to construct an empirical G' sum distribution. An Empirical Cumulative Distribution Function was then applied using the R function ecdf to estimate the fraction of simulated observations that were less than

or equal to the steady-state QTL's nearest SNP sum and determine the probability of the existing QTL peak's nearest SNPs G' sum occurring given the randomly empirical G' sum distribution. Three pQTL without cognate eQTL were clustered within 8000 basepairs of one another, well within the G' sliding window size, so these steady-state QTL's nearest SNPs were collapsed into a single SNP for statistical testing.

Results:

Characterizing Expression Variation as Local or Distant

Tos6 Allele Swap Results

We performed allele swap protein expression assays to determine if the previously observed differences in TOS6 protein expression between brewer's yeast strains S288c and YJM145 are due to local (within the locus of Tos6) genetic variants, distant (outside the locus of Tos6) genetic variants, or some combination of local and distant genetic variants. We measured protein expression with fluorescence microscopy for two "native" strains with the native allele of Tos6 tagged with YFP and two "allele swap" strains where the native Tos6 allele was replaced with a YFP-tagged Tos6 allele from the opposing strain. We first analyzed the protein levels of native and allele-swap strains via a two-way ANOVA model including allele and genomic background terms as well as interaction effects. We found significant allele-driven effects ($F=21.3$, $df=1$, $p=0.000217$), rejecting the null hypothesis that Tos6 allele has no effect on TOS6 protein levels. The S288c Tos6 allele showed 17% greater mean expression across genomic backgrounds and showed the highest expression in the YJM145 background (Figure 8), suggesting that local genetic variation is acting on TOS6 protein expression. However, this does not rule out the possibility that small-effect *trans*-acting factors also play a role in regulating TOS6 protein levels.

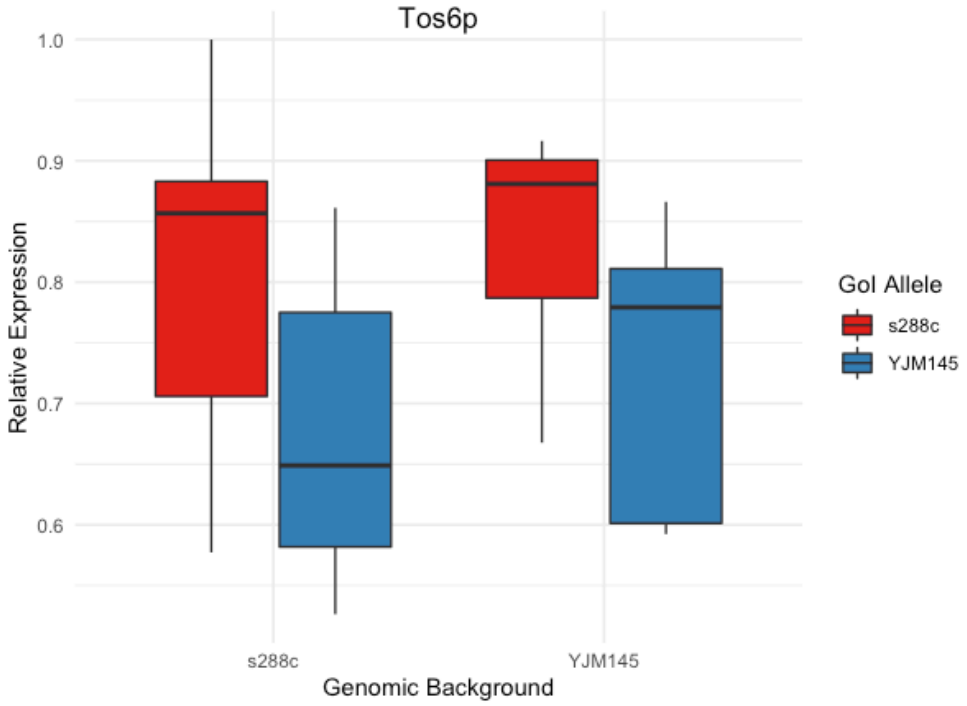


Figure 8: Comparison of both S288c (lab strain) and YJM145 (clinical strain) Tos6 allele relative protein expression in its native vs reciprocal genomic backgrounds. Cells were grown to log phase, treated with mating pheromone for 4 hours, imaged via fluorescent microscopy, and had their fluorescence normalized against non-fluorescent tagged native strains. An ANOVA model assessing genomic background and GoI allele-specific effects found significant differences in TOS6 protein abundance between S288c and YJM145 Tos6 alleles ($F=21.3$, $df=1$, $p=0.000217$).

We next examined if these local effects observed in the Tos6 allele swap strains were driven by expression differences acting at the RNA level, protein level, or both. Previous work on Tos6 expression differences between these strains identified that YJM145 expresses 3.5-7x more Tos6 RNA than S288c, while expressing roughly similar amounts of Tos6 protein between the two strains (Pollard et al. 2016). Additionally, previous work attributed the higher protein per RNA in S288c to higher protein synthesis rate and lower protein decay rate. We reasoned that if the local effects are acting at the RNA level alone, then the previously measured ratio of Tos6

RNA between the strains should be comparable to the ratio of Tos6 protein between native and allele swap strains. We used prior RT-qPCR results to compare the ratio of S288c to YJM145 native strain RNA levels to the allele swap strain Tos6 protein level ratios via Welch's t-test and found that native strain RNA level ratios were significantly different than allele swap strain protein level ratios (t=-10.75, df=6.11, p-value=3.41e-05) (Figure 9). These results suggest that Tos6 RNA level variation cannot explain the observed Tos6 protein abundance expression patterns and provide evidence of local, protein level specific effects on Tos6 protein abundance.

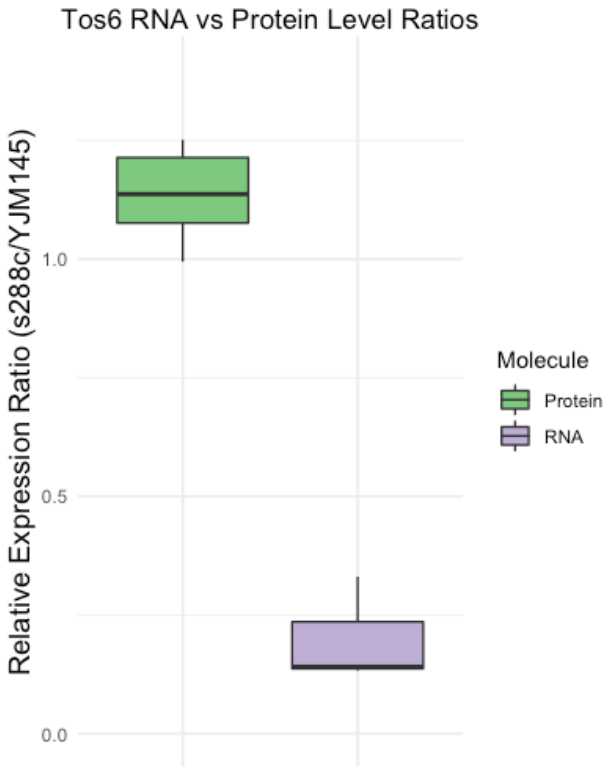


Figure 9: Comparison of Tos6 expression ratio between strains (S288c expression over YJM145 expression) at both the RNA and Protein levels show mean differences in RNA vs Protein level differences between strains at 4 hours post-pheromone treatment. Welch's t-test found significant differences between RNA level and Protein level ratios between strains (t=-10.75, df=6.11, p-value=3.41e-05).

Fus3 Allele Swap Results

Similar to the above analysis of Tos6, we performed an allele swap experiment to determine if the previously observed differences in Fus3 expression between yeast strains S288c and YJM145 are due to local genetic variants, distant genetic variants, or some combination of the two. Unlike Tos6, we did not observe consistent effects of Fus3 alleles (Figure 9) but we did observe a trend toward higher Fus3 expression for the YJM145 strain background (Figure 9). A two-way ANOVA model with GoI allele and genomic background terms failed to uncover significant differences in Fus3 expression between alleles ($F=0.057$, $df=1$, $p=0.81$) but found a nearly significant difference between genomic backgrounds ($F=4.22$, $df=1$, $p=0.055$). Previous work done in the Pollard lab found significantly higher Fus3 expression in the native S288c strain background (Pollard et al. 2016), not YJM145 as observed here. Other studies have found that Fus3 expression is highly sensitive to environmental conditions such as cell-cycle stage and pheromone concentration even within clonal populations (Conlon et al. 2016) (Li et al. 2017) which may explain why these results are inconsistent with previously observed FUS3 protein expression patterns. Regardless, we did not find evidence for local or distant variants acting on FUS3 protein expression dynamics.

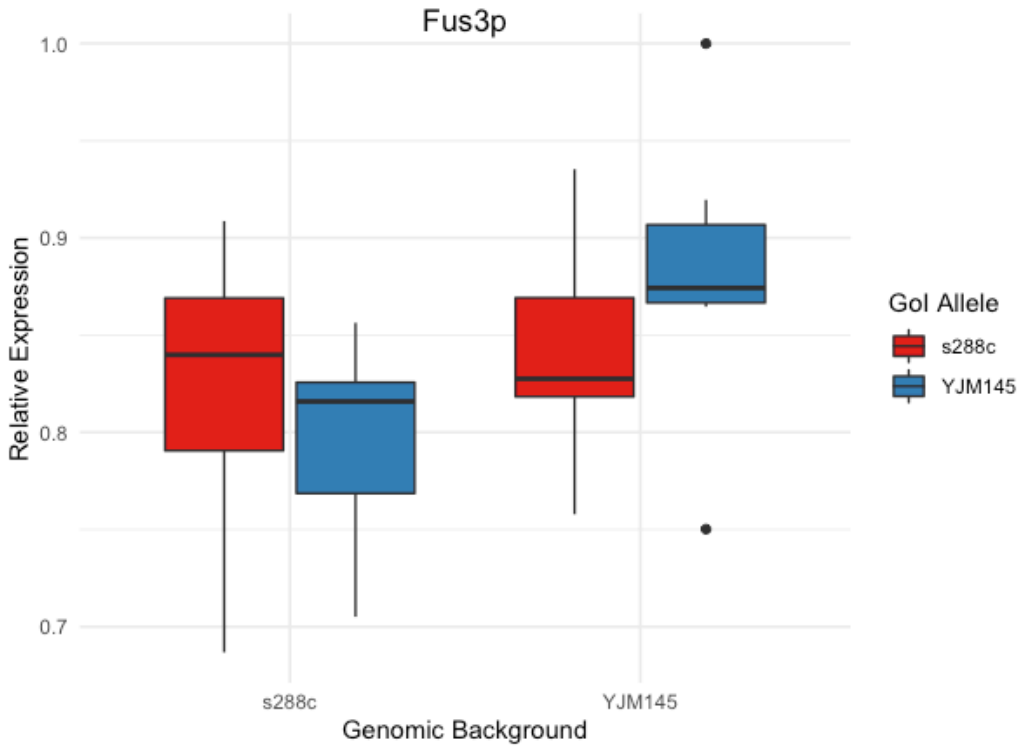


Figure 9: Comparison of both S288c (lab strain) and YJM145 (clinical strain) Fus3 allele relative protein expression in its native vs reciprocal genomic backgrounds. Cells were grown to log phase, treated with mating pheromone for 4 hours, imaged via fluorescent microscopy, and had their fluorescence normalized against non-fluorescent tagged native strains. An ANOVA model comparing effects of genomic background and Fus3 allele found no significant differences in Fus3 expression across background ($F=4.22$, $df=1$, $p=0.055$) or GoI allele ($F=0.057$, $df=1$, $p=0.81$).

Mapping Trans Acting Variation

FIG1 Whole Genome pQTL Maps

We crossed together yeast strains S288c and YJM145, remated offspring for two or five generations, and performed BSA experiments on the segregants from the cross in order to find genomic locations associated with differences in Fig1 protein expression between strains. In order to evaluate the dynamics of the genetic network acting on Fig1 protein expression we performed BSA experiments at two hours and five hours after the addition of mating pheromone. We repeated the experiments using segregants generated from either two or five rounds of random mating in order to better understand the methodological trade-offs between more or less mating. Across the four BSA experiments, 17 unique significant FIG1 pQTL were detected at an FDR of 5% (Figure 10). Of these 17 unique QTL, two QTL (Chromosomes 13 and 14) were significant in both replicates of both timepoints. Meanwhile four QTL were specific to the two-hour timepoint experiments, one QTL was only replicated at the five-hour timepoint and 12 of the significant QTL were detected in a single experimental replicate but were not found in other replicates (Figure 11). Interestingly, twice as many distinct QTL were detected between the two-hour timepoint replicates as the five-hour timepoint replicates. Further, comparison of maximum $\log_{10}$ p-values across shared QTL intervals suggest that the early pheromone response FIG1 level phenotype might have more genetic complexity than that of the late pheromone response phenotype (Figure 12). Two verified expression level causal polymorphisms were located within significant FIG1 pQTL peaks. One such variant located within the Mkt1 locus was significant across all four experimental replicates, and has been implicated in regulating both transcript levels and protein levels for a wide variety of other

genes (Albert et al. 2014) (Albert et al. 2018) (Brion et al. 2020). Another polymorphism within the Gpa1 locus known to affect both mating and growth-related phenotypes (Lang et al. 2009) was significant in both two-hour timepoint experiments but not significant in either five-hour timepoint experiments.

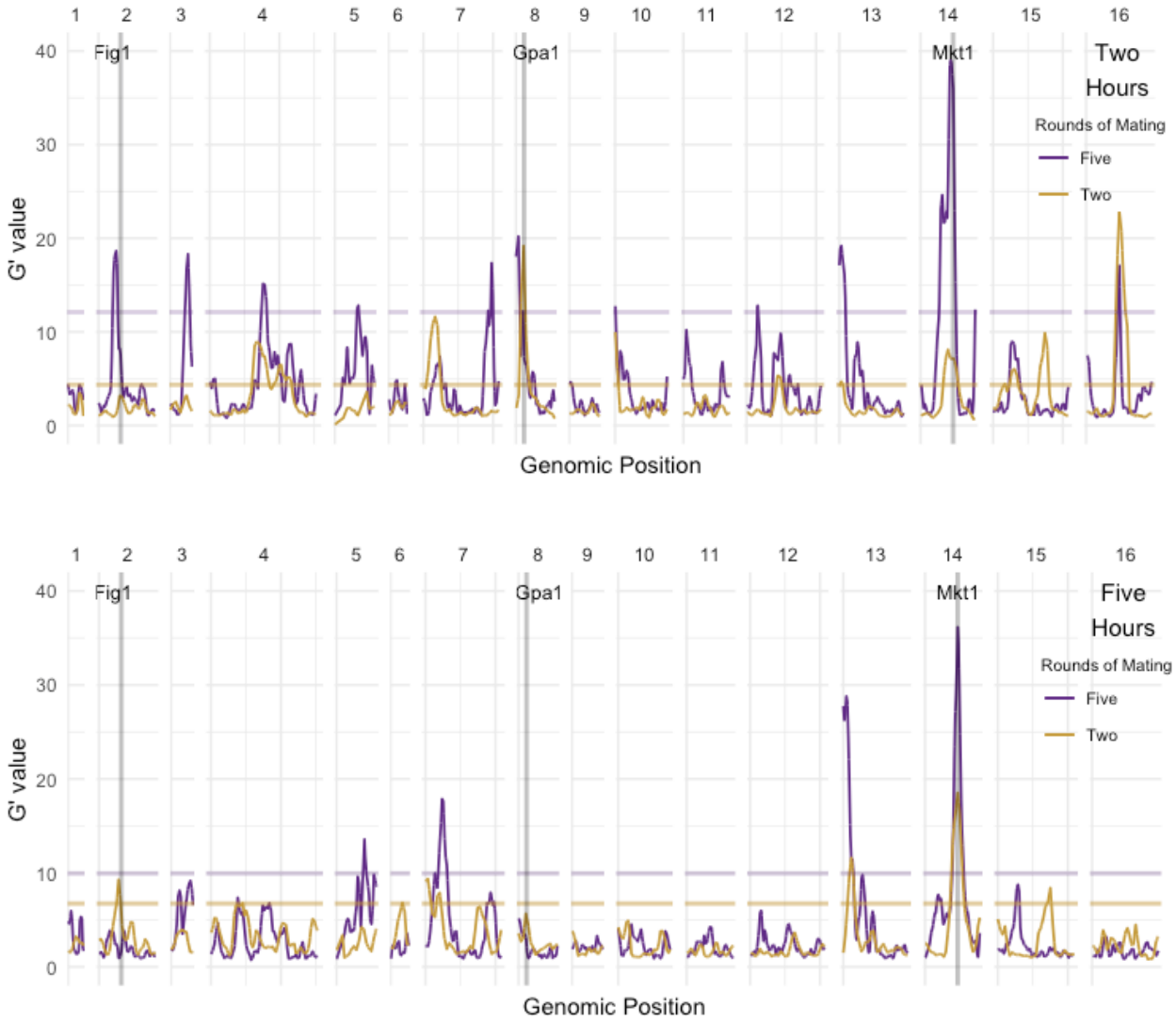


Figure 10**.** Whole-genome FIG1 QTL-maps. Peaks represent genomic loci associated with FIG1 protein expression 2 hours (top) after exposure to mating pheromone, and 5 hours (bottom) after exposure to mating pheromone. Horizontal lines represent an FDR cutoff of 0.05. Purple and gold maps correspond to BSA experiments with segregant populations derived from five rounds and two rounds of random mating respectively.

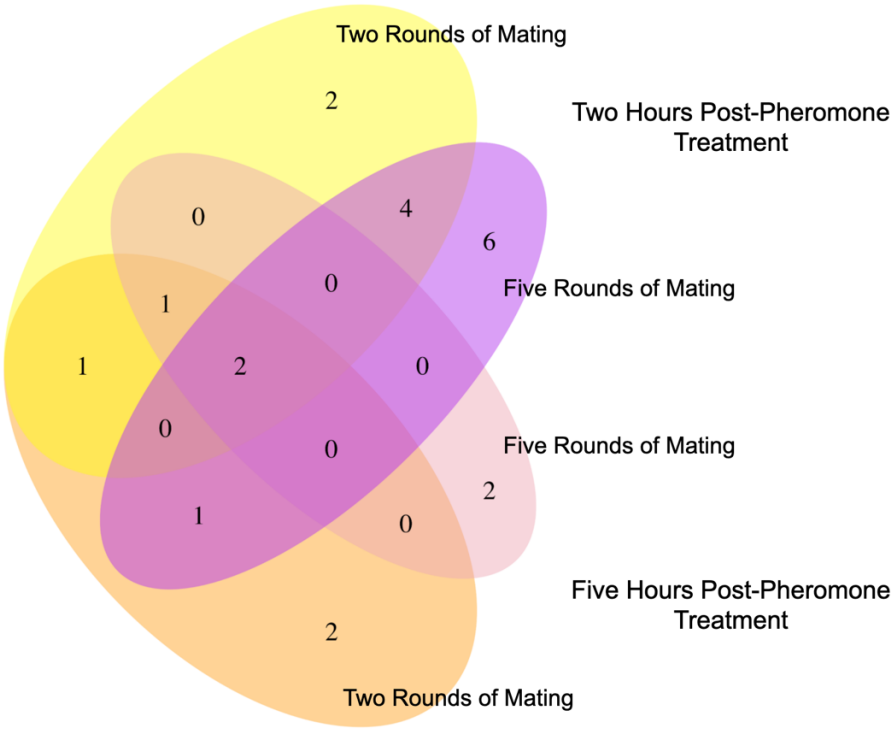


Figure 11. Four-way Venn Diagram of shared (overlapping) significant QTL intervals between experiments. Intervals were classified as shared if any portion of significant regions overlapped between experiments. Yellow oval represents two rounds of mating, two hours post mating pheromone treatment QTL. Orange oval represents two rounds of mating, five hours post mating pheromone treatment QTL. Purple oval represents five rounds of mating, two ours post mating pheromone treatment QTL. Pink oval represents five rounds of mating, five hours post pheromone treatment QTL.

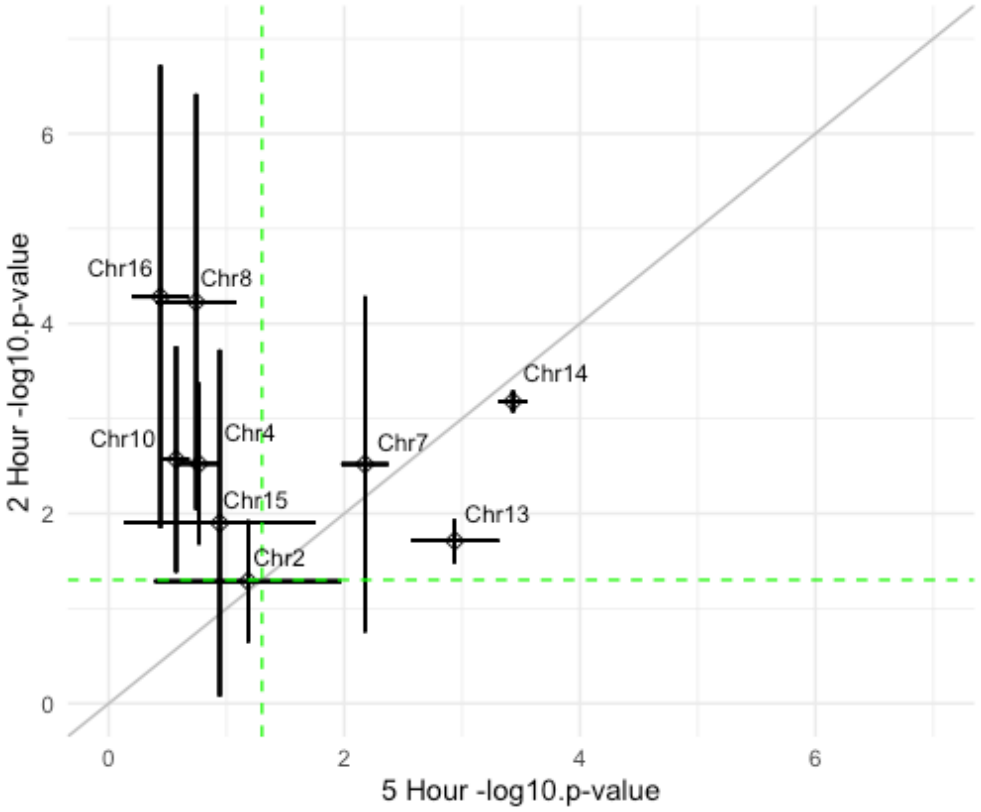


Figure 12. Relationship between two-hour and five-hour timepoint overlapping QTL peak -log10 p-values. Overlapping QTL regions were defined as the union of significant QTL regions and -log10 p-values were calculated from the local maxima within each overlapping region. Black boxes represent mean -log10 p-value, and whiskers represent the range of -log10 p-values for each timepoint. Dotted green lines represent significance cut-offs at an FDR=0.05.

18 total QTL were detected in the two-rounds of random mating experiments while 15 total QTL were detected in the five-rounds of random mating experiments. The slightly lower number of total QTL in the five-rounds of random mating experiments may be partly attributable to increased selection for mating-related alleles. YJM145 and S288c have known differences in mating efficiency, and in theory alleles that improve mating efficiency should increase in frequency in a population undergoing subsequent rounds of random mating and selection. As a result, increased rounds of random mating may unintentionally select for said

mating related alleles, decreasing the allelic diversity near these implicated loci through linkage and decreasing power to detect certain QTL.

Comparing FIG1 pQTL to Established pQTL Hotspots

Where prior experiments have mapped pQTL at steady state there have yet to be pQTL mapped for pheromone responsive genes. In order to better understand whether steady state pQTL extend across environments and play a role in pheromone responsive FIG1 protein levels, FIG1 pQTL were compared to a set of established steady-state pQTL hotspots (Albert et al. 2014) by finding the single-nucleotide polymorphisms (SNPs) in our FIG1 BSA datasets nearest each pQTL hotspot peak (Supplementary Figure 1). These comparisons should in theory provide insight as to whether the pQTL acting on steady state protein levels are the same or different as the pQTL acting on dynamic protein levels during mating pheromone response. A pQTL hotspot at position 465007 on chromosome 14 was significant in all four experimental replicates. Additionally, a hotspot on chromosome 13 was significant in both of the 5-hour timepoint experiments while the hotspot on chromosome 8 was significant in both of the 2-hour timepoint experiments. A hypergeometric test found the mild levels of enrichment for significant SNPs near existing pQTL hotspots were not significant for any of the four experiments (Table 5). An ecdf test found that the pQTL hotspots' nearest SNPs had significantly higher G' values than what would be expected from selecting genomic SNPs at random for the five-hour five-rounds of mating experiment ($p=0.0281$, $alpha=0.05$) while the other experiments showed similar but slightly weaker non-significant trends. Across both hypergeometric and ecdf pQTL hotspot tests the five-rounds of random mating experiments

(which correspond to decreased G' tri-cubed sliding window size) tended to have lower p-values than their two-rounds of random mating counterparts. Further, the five-hour timepoints tended to have lower p-values than their two-hour counterparts for both hypergeometric and ecdf pQTL hotspot tests. These results suggest that while some of the steady-state pQTL hotspots seem to also be contributing to FIG1 protein levels, for example the hotspot on chromosome 14, many of these detected FIG1 pQTL are distinct with regards to existing steady-state pQTL.

Time Point		Two Hours		Five Hours	
Rounds of Mating		Two	Five	Two	Five
pQTL Hotspots	enrichment value	1.43x	1.9x	1.97x	2.36x
	hypergeometric test p-value	0.138	0.08465	0.07861	0.05032
	ecdf test p-value	0.1539	0.0527	0.0543	*0.0281
eQTL Hotspots	enrichment value	1.19x	1.49x	1.16x	1.39x
	hypergeometric test p-value	0.1786	0.0879	0.2626	0.2629
	ecdf test p-value	0.1207	0.1173	0.3423	0.2056
pQTLs w/o eQTLs	enrichment value	1.58x	2.11x	4.37x	5.25x
	hypergeometric test p-value	0.1205	0.07756	*0.008766	*0.005236
	ecdf test p-value	0.4846	*0.0404	*0.018	*0.0114

Table 3: Enrichment values and p-values for established steady-state QTL' nearest SNP hypergeometric and ecdf tests. *indicates tests that were significant at alpha = 0.05

Comparing FIG1 pQTL to Established eQTL Hotspots

Although prior experiments suggested that FIG1 protein level differences could not be explained by RNA level differences, this does not necessarily imply that mRNA level QTL (eQTL) do not contribute to FIG1 protein level differences. In order to take advantage of a more comprehensive gene expression QTL dataset and to determine whether steady-state eQTL were impacting FIG1 protein levels, FIG1 pQTL were also compared to established eQTL hotspots (Albert, 2018) by finding the SNPs in our FIG1 BSA datasets closest to each eQTL hotspot peak (Supplementary Figure 2). The eQTL hotspot at position 466588 on chromosome 14's nearest SNP was significant in all four experiments. Additionally, each experimental replicate had at least one significant eQTL hotspot SNP within the first 90,000 bp of chromosome 13. The eQTL hotspots at position 745016 of chromosome 4 and the hotspot at position 46353 of chromosome 16s' nearest SNPs were significant in both two-hour timepoint replicates but significant in neither of the five-hour replicates. While mild enrichment for established eQTL hotspots within the FIG1 QTL was observed across each experiment, a hypergeometric test found the enrichment to be non-significant. An ecdf test found that the pQTL hotspots' nearest SNPs had slightly higher G' values than what would be expected from selecting genomic SNPs at random, yet none of the experiments' results reached significance.

For both hypergeometric and ecdf eQTL hotspot tests the two-hour timepoint replicates tended to have both lower p-values as well as greater overall enrichment than their five-hour timepoint counterparts, while the five-rounds of random mating experiments (which also correspond to decreased G' tri-cubed sliding window size) tended to have lower p-values than their two-

rounds of random mating counterparts. Together these results suggest that while some of the established steady-state eQTL hotspots appear to play a role in Fig1 protein abundance, most of the FIG1 pQTL do not coincide with established eQTL hotspots. This general trend showing only modest overlap between FIG1 pQTL and existing steady-state eQTL hotspots suggest that FIG1 protein levels are affected by different eQTL than observed in steady-state studies, or possibly that FIG1 protein levels are regulated by transcriptionally independent mechanisms (Figure 6). Further, the observation that established steady-state eQTL hotspots show slightly greater association with the two-hour FIG1 pQTL than the five-hour pQTL suggest that *trans*-variants that affect transcript abundance may influence early pheromone response FIG1 protein levels more than they affect late pheromone response FIG1 levels.

Comparing FIG1 pQTL to Established pQTL without eQTL

Upon observing that steady-state eQTL hotspots seemed to be more predictive of early pheromone response FIG1 pQTL than late pheromone response FIG1 pQTL (Table 5), and because FIG1 protein levels appear to be regulated by different genetic elements than those that effect steady-state RNA levels, FIG1 pQTL were also compared to an established set of steady-state pQTL that lack cognate eQTL (Albert, 2018) (Supplementary Figure 3). Three pQTL without eQTL clustered around position 460000 of chromosome 14 were significant in each FIG1 pQTL experimental replicate. Another pQTL lacking a cognate eQTL at position 112600 of chromosome 13 was significant in both five-hour timepoint replicates but neither of the two-hour timepoint experiments. The five-hour timepoint Fig1 pQTL showed 4.37 and 5.25 times enrichment for steady-state pQTL without cognate eQTL respectively for the two and five-

rounds of mating replicates which hypergeometric tests found to be significant, while neither two-hour timepoint replicates showed significant enrichment. An ecdf test found that the pQTL without eQTL's nearest SNPs had significantly higher G' values than what would be expected from selecting genomic SNPs at random for all experimental replicates besides the two-rounds of mating two-hour timepoint experiment. Across both hypergeometric and ecdf pQTL without eQTL tests the five-rounds of random mating experiments tended to have lower p-values than their two-rounds of random mating counterparts. Moreover, the five-hour timepoints tended to have lower p-values than their two-hour counterparts for both hypergeometric and ecdf pQTL without eQTL tests. The observation that the five-hour timepoint FIG1 pQTL were better associated with the steady-state pQTL without eQTL than the two-hour timepoint pQTL suggests that the late pheromone response FIG1 protein level phenotype might be more influenced by known genetic variants that act on protein levels in a transcriptionally independent mechanism.

Discussion:

This project set out to elucidate how genetic variation contributes to dynamic protein level differences between two closely related strains of yeast. For the Tos6 and Fus3 genes we performed GoI allele swaps combined with fluorescence microscopy to determine whether pheromone responsive protein level differences were driven by genetic variation local genetic variation or by genetic variation in distant regions of the genome. We found evidence that local variation within the Tos6 allele drives the observed differences in TOS6 protein levels. Results for the Fus3 gene were inconclusive. Prior experiments had determined that FIG1 protein abundance was regulated by distant genetic variation, so here we employ Bulk Segregant Analysis to map FIG1 pQTL during early and late pheromone response. Further, since S288c and YJM145 share nearly identical Fig1 RNA expression profiles (Figure 6) pQTL detected via BSA are likely acting directly at the protein level. Across BSA experiments we successfully mapped 21 distinct FIG1 pQTL, many of which appear to be novel with respect to prior known steady-state pQTL. Interestingly many of the FIG1 pQTL appear to be dynamic, in that they differentially affect FIG1 protein levels between time points during mating pheromone response.

Allele Swap Expression Experiments

These experiments provide strong evidence that local genetic variation within the Tos6 allele contributes to the observed differences in TOS6 protein expression between strains. Further we are able to infer from previous RNA level measurements that TOS6 protein level differences are driven by protein level specific mechanisms. A prior cycloheximide protein decay rate assay where translation is suspended and cellular protein levels are measured over time found a

significant difference in YJM145 and S288c TOS6 protein decay rates providing evidence that at least one protein-level specific mechanism is at play, however these results do not exclude the possibility of a complex network of variants contributing to TOS6 levels by a variety of mechanisms acting simultaneously. Experimental designs utilizing RNA level and protein level quantification in tandem will be able to disentangle the RNA vs protein level effects, verify our inferences about protein level specific effects, and provide further insight on the cellular mechanisms acting on dynamic TOS6 protein levels. As a next step, the precise location of causal level polymorphisms can be fine-mapped using a systematic allele-swap divide-and-conquer approach, in which various combinations of a chimeric allele are compared until the relative contributions of individual polymorphisms can be distinguished (Sadhu et al. 2016) (Lutz et al. 2019). The FUS3 allele swap results are difficult to discern, and as a result we will likely suspend our interrogation of Fus3 genetic architecture until we can incorporate more sensitive and/or complex assays, such as assays that can simultaneously consider individual cell-cycle stage along with single-cell GoI protein abundance.

Dynamic FIG1 pQTL

Interestingly, we find evidence that some *trans*-acting factors make major contributions to early pheromone response FIG1 protein abundance but not to late pheromone response FIG1 abundance (Figure 12).

One mechanism that could explain these apparent differences is that FIG1 may have higher sensitivity to QTL that act on the transcription rates of other genes early during pheromone response. In this scenario cells that are able to coordinate a stronger initial transcriptional response upon pheromone exposure may receive a jump-start in their mating-related gene expression cascades as mating-specific transcription factors and other regulatory proteins get produced more rapidly. Under this framework broad-acting eQTL would influence early mating response FIG1 levels more so than late mating FIG1 levels. Although the trend is subtle, we did observe that our late pheromone response FIG1 QTL had greater overlap with previously identified steady-state pQTL that regulate protein levels in a transcriptionally independent manner (Table 5).

We are also interested in the possibility is that the apparent differences in genetic complexity between mating response stages are due to differences in broad-sense heritability. Current work has set out to estimate the broad-sense heritability of FIG1 protein expression during pheromone response. Not only will these experiments allow for direct comparison of heritability measurements between mating stages, these endeavors will also help to contextualize the overall effect pQTL have on FiG1 abundance.

Potential Molecular Mechanisms for *Trans*-acting pQTL

To begin addressing questions about the underlying molecular mechanisms contributing to FIG1 protein abundance we first compared the FIG1 pQTL peak locations to genetic variants that had contributed to steady-state protein levels for genes in previous studies. Two such documented expression level polymorphisms were present within our significant QTL peaks, harbored within the genes GPA1 on chromosome 8 and MKT1 on chromosome 14. GPA1, inhibitory alpha subunit of the G-protein coupled receptor that binds mating pheromone peptides and initiates the mating MAPK pathway (Miyajima et al. 1987) (Bardwell, 2005). Prior studies have found that GPA1 harbors a Quantitative Trait Nucleotide (a single nucleotide polymorphism associated with a trait of interest) affecting pheromone response physiology (Yvert et al. 2003). Further, the S288c Gpa1 open reading frame is known to harbor a non-conservative missense mutation with respect to the ancestral yeast Gpa1 coding sequence which leads to loss-of-function and thus decreased capacity to downregulate the pheromone response signaling cascade. The native S288c genomic background shows greater FIG1 expression than the YJM145 genomic background which is consistent with prior characterization of the Gpa1 allele, however counterintuitively, QTL mapping reveals that the S288c GPA1 allele appears to be associated with the low FIG1 protein abundance phenotype (Supplementary Figure 4). Making matters more complex, variation in the Ste20 (a gene coding for a kinase protein involved in mating response pathway signal transduction) gene locus, ~20 kb away from the Gpa1 gene locus and well within the sliding window range for both experiments (30 kb and 50 kb for five-rounds and two-rounds of random mating respectively), has also been implicated in regulating protein abundance of many steady-state genes (Grossbach et al. 2019). Therefore, either or

both of these genes may harbor genetic variants that act on FIG1 protein abundance. Current experiments swapping the alleles of both Gpa1 and Ste20 seek to resolve how each gene is contributing to the QTL on chromosome 8. It is also worth noting that the chromosome 8 QTL appears to have stronger effects early in pheromone response compared to late pheromone response, consistent with the role of Gpa1 and Ste20 in transducing pheromone signal.

The pQTL on chromosome 14 harbors known quantitative trait polymorphisms within the Mkt1 locus, a gene that codes for a nuclease like-protein involved in post-transcriptional regulation (Tadauchi et al. 2004) (Wickner, 1987). Recently, steady-state pQTL and eQTL mapping experiments implicated Mkt1 as a trans-acting regulator for 10 different genes (Brion et al. 2020). The S288c Mkt1 allele contains mutations with respect to ancestral yeast populations which have been shown to influence a wide variety of phenotypic traits ranging from growth rate to drug susceptibility (Deutschbauer and Davis, 2005) (Fay, 2013). The observation that variation within the Mkt1 allele impacts cellular growth rates suggest that FIG1 protein abundance may be influenced by a growth rate – protein dilution mediated mechanism. Cellular growth rates are able to act on cellular protein abundance because every time a cell divides its cellular protein constituents are divided into the two daughter cells, and most protein molecules are relatively stable compared to RNA molecules which have rapid turnover rates. Thus, cells with otherwise equal protein production and degradation rates will have protein abundances inversely proportional to the rate of cellular division. This explanation is also consistent with the observation that the YJM145 strain background has a faster mitotic growth rate than S288c and therefore the decrease in FIG1 protein abundance in the YJM145

background may be a result of a mitotic growth rate - protein dilution effect. A growth rate based - protein dilution mechanism acting directly at the protein level might also explain the enrichment for protein level specific steady-state pQTL that was observed for the late pheromone response FIG1 phenotype.

In an attempt to further elucidate the genetic architecture and molecular mechanisms underlying dynamic FIG1 protein expression we can begin to look for candidate genes within significant pQTL regions. By leveraging existing gene ontology and functional annotation data combined with our genome-wide G' scores for individual SNPs we can find polymorphic genes with large mean G' scores that have protein functions which might influence protein abundance of other genes in *trans*. These candidate genes will provide context and inform our search for causal variants that act on FIG1 protein levels as we begin allele swap experiments to verify and fine-map *trans*-acting polymorphisms. And since *trans*-variants may have small individual effects on FIG1 protein abundance phenotypes, higher powered assays such as flow cytometry may be necessary to evaluate such *trans*-allele specific expression differences.

An ongoing goal of this research is to further our understanding of how cellular protein abundance is dynamically regulated in response to environmental perturbation such as exposure to mating pheromone, and further, the ways in which natural genetic variation influences these processes. To this end, we are currently working to map QTL for 10 more pheromone-responsive genes that show dynamic protein expression differences between lab and clinical strains. This research will disentangle broad-acting QTL that affect protein

abundances of many pheromone-responsive genes from QTL that act specifically on the expression of a single gene. These future studies will illuminate the genetic mechanisms that underly dynamic protein abundance regulation, as well as determine the degree of overlap between dynamic protein level QTL and steady-state protein level QTL. Furthermore, expanding this research to a broader set of pheromone responsive genes will help us to determine which protein abundance regulatory mechanisms are shared across genes, and determine which molecular level these mechanisms tend to act on.

Outstanding Questions and Future Considerations

Several QTL were detected in sub-telomeric regions near the ends of chromosomes. Yeast chromosome ends contain complex structural variation where in many cases S288c differs from ancestral strains by entire blocks of subtelomeric genes (Cubillos et al. 2011) (Albert et al. 2018), which in turn influences a wide-swath of phenotypic traits. Further, the precise location of causal genes or polymorphisms within these regions cannot be determined from our current segregant panel because each segregant either contains all or none of the genes in the region. Therefore, fine mapping causal polymorphisms in these regions may require systematic allele swaps of large chromosomal regions in order to directly compare the effects of copy number variation and moreover allow us to progressively narrow in on the location of causal variants. If any genes within these regions are essential, systematic allele swaps may necessitate a CRISPR-based approach in order to knock out and replace a given genomic loci in a single step (Albert et al. 2018) (Sadhu et al. 2016) (Sadhu et al. 2018).

While further rounds of mating and sporulation are beneficial in that they allow for an increase in meiotic recombination events, further shuffling the genome and justifying the use of a smaller sliding window which in turn provides increased mapping resolution, it appears to come at the cost of decreased power to detect certain QTL. As segregants are subjected to further rounds of mating and sporulation several unintended selective pressures can begin to creep in. For example, S288c and JYM145 are known to have differences in mating efficiency, which means that certain mating-related alleles may increase in frequency in the segregant population through subsequent rounds of mating effectively decreasing the segregant population allelic diversity for certain loci. The sporulation process in which cells are grown at suboptimal temperatures in nutrient deplete media for a week in order to provide enough environmental stress to initiate sporulation during an otherwise stable diploid state may also provide a selective pressure for sporulation efficiency or survivability related alleles. Together, these selective pressures may be contributing to loss in population-wide allelic diversity, making real QTL nearby these implicated alleles more difficult to detect. This observed trade-off between mapping resolution, and power to detect certain QTL will influence future BSA experimental designs.

One challenge presented by BSA, is that allele frequency estimations can be skewed by mapping bias, the tendency for sequencing reads corresponding to the reference strain to map with higher accuracy than reads belonging to the non-reference allele. Mapping bias is especially prevalent in highly polymorphic regions of the genome as reads containing too many polymorphisms can either map to the wrong part of the genome or fail to map altogether.

Mapping bias is of concern in these experiments because the *S. cerevisiae* reference genome was constructed using our Lab strain S288c's genome sequence. While in theory, since the G-statistic null-hypothesis assumes equal allele frequencies between bulks rather than assuming an allele frequency of 0.5 in both bulks, the G prime analysis should be fairly robust to biases presented by reference genome choice. However, as experiment-wide mean allele frequency departs from 0.5, the possible allele frequency differences that can exist between bulks decreases, making QTL throughout the genome more difficult to detect. Furthermore, several highly polymorphic regions exist between the S288c and YJM145 backgrounds (Supplementary Figure 5). And while these variant-rich regions are prime candidates for contributing to phenotypic differences between strains, reads in these regions, especially reads harboring one of the many YJM145-specific ORFs, will inevitably show mapping discrepancies between reference and non-reference alleles making QTL in such regions particularly difficult to detect. We indeed find evidence of mapping bias in our experiments with global mean reference allele frequencies of ~0.55 (favoring the S288c allele) for each of our experiments (supplemental figure 6). An important next step for this research is to develop a read processing method that accounts for mapping discrepancies between reference and non-reference reads. One proposed solution to account for such mapping bias involves mapping all reads to both reference and non-reference (S288c and YJM145) genomes and only proceeding with reads that map unambiguously to both strain backgrounds for downstream statistical analysis (Albert et al. 2014).

One overarching problem that continues to evade statistical geneticists in the field of BSA-mediated QTL-mapping is the relative lack of consistency observed between BSA statistical methods. For example, a variety of BSA statistical frameworks ranging from chi-squared oriented tests like G, G', and CMH, to linear, binomial general linear, and quasibinomial general linear based models exist, each including their own inherent trade-offs (Wiberg et al. 2017). Frustratingly, direct comparison between approaches reveal that different statistical techniques can often produce drastically different QTL-mapping results. Most notably each of these established tests suffer from low true positive and high false positive rates, except in designs where BSA experiments are sufficiently replicated, which can quickly become cost-prohibitive. Many tests also include rather ambiguous significance thresholding strategies making QTL classification somewhat arbitrary (Huang et al. 2020). Making matters worse, many user-friendly software packages make these statistical tests and their underlying assumptions less transparent. To this point, there remains the need for a cost-effective, robust, and sensitive BSA statistical framework within the QTL-mapping community, and future studies aiming to address these concerns could provide immense value to the field.

Conclusion

This study provides proof of concept for time-based phenotypic trait mapping and is the first study to date to map pQTL in a dynamic system such as yeast mating pheromone response. In addition to laying the groundwork for future time-based pheromone-response pQTL mapping experiments this work also establishes techniques for generating segregant yeast populations, analyzing BSA Next-Gen sequencing data, and making comparisons between dynamic pQTL and publicly available steady-state gene expression QTL. Notably, we find evidence for *trans*-acting

genetic factors that regulate FIG1 protein abundance, including some loci that affect FIG1 protein abundance in a time-dependent manner, a phenomenon which had been theoretically proposed but had yet to be observed in a natural system. These findings demonstrate the necessity for dynamic trait mapping in gene expression-related phenotypic trait mapping studies. Together this research moves us towards an improved understanding of the genetic architecture underlying dynamic protein abundance phenotypes and sheds light on the complex nature of genotype-phenotype relationships in natural populations.

Works Cited:

Albert, Frank W., & Kruglyak, L. (2015). The role of regulatory variation in complex traits and disease. *Nature Reviews. Genetics*, *16*(4), 197–212. https://doi.org/10.1038/nrg3891

Albert, Frank W., Treusch, S., Shockley, A. H., Bloom, J. S., & Kruglyak, L. (2014). Genetics of single-cell protein abundance variation in large yeast populations. *Nature*, *506*(7489), 494–497. https://doi.org/10.1038/nature12904

Albert, Frank Wolfgang, Bloom, J. S., Siegel, J., Day, L., & Kruglyak, L. (2018). Genetics of trans-regulatory variation in gene expression. *ELife*, *7*, e35471. https://doi.org/10.7554/eLife.35471

Bardwell, L. (2005). A walk-through of the yeast mating pheromone response pathway. *Peptides*, *26*(2), 339–350. https://doi.org/10.1016/j.peptides.2004.10.002

Borneman, A. R., Gianoulis, T. A., Zhang, Z. D., Yu, H., Rozowsky, J., Seringhaus, M. R., Wang, L. Y., Gerstein, M., & Snyder, M. (2007). Divergence of Transcription Factor Binding Sites Across Related Yeast Species. *Science*, *317*(5839), 815–819. https://doi.org/10.1126/science.1140748

Brem, R. B., Yvert, G., Clinton, R., & Kruglyak, L. (2002). Genetic Dissection of Transcriptional Regulation in Budding Yeast. *Science*, *296*(5568), 752–755. https://doi.org/10.1126/science.1069516

Brion, C., Lutz, S., & Albert, F. W. (2020). *Simultaneous quantification of mRNA and protein in single cells reveals post-transcriptional effects of genetic variation* [Preprint]. Genetics. https://doi.org/10.1101/2020.07.02.185413

Broad Institute. (2016). Picard tools.

Buffalo, V. (2011). Scythe-a Bayesian adapter trimmer. *Website: http://github. com/vsbuffalo/scythe*.

Cavinder, B., & Trail, F. (2012). Role of Fig1, a Component of the Low-Affinity Calcium Uptake System, in Growth and Sexual Development of Filamentous Fungi. *Eukaryotic Cell*, *11*(8), 978–988. https://doi.org/10.1128/EC.00007-12

Chen, R. E., & Thorner, J. (2007). Function and regulation in MAPK signaling pathways: Lessons learned from the yeast Saccharomyces cerevisiae. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1773*(8), 1311–1340. https://doi.org/10.1016/j.bbamcr.2007.05.003

Conlon, P., Gelin-Licht, R., Ganesan, A., Zhang, J., & Levchenko, A. (2016). Single-cell dynamics and variability of MAPK activity in a yeast differentiation pathway. *Proceedings of the National Academy of Sciences*, *113*(40), E5896–E5905. https://doi.org/10.1073/pnas.1610081113

Costanzo, M., & Boone, C. (2009). SGAM: An Array-Based Approach for High-Resolution Genetic Mapping in Saccharomyces cerevisiae. In I. Stagljar (Ed.), *Yeast Functional Genomics and Proteomics: Methods and Protocols* (pp. 37–53). Humana Press. https://doi.org/10.1007/978-1-59745-540-4_3

Cubillos, F. A., Billi, E., Zörgö, E., Parts, L., Fargier, P., Omholt, S., Blomberg, A., Warringer, J., Louis, E. J., & Liti, G. (2011). Assessing the complex architecture of polygenic traits in diverged yeast populations. *Molecular Ecology*, *20*(7), 1401–1413. https://doi.org/10.1111/j.1365-294X.2011.05005.x

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498. https://doi.org/10.1038/ng.806

de Nadal, E., Ammerer, G., & Posas, F. (2011). Controlling gene expression in response to stress. *Nature Reviews Genetics*, *12*(12), 833–845. https://doi.org/10.1038/nrg3055

de Vries, A. G., Sosnicki, A., Garnier, J. P., & Plastow, G. S. (1998). The role of major genes and DNA technology in selection for meat quality in pigs. *Meat Science*, *49*, S245–S255. https://doi.org/10.1016/S0309-1740(98)90052-3

Delile, J., Rayon, T., Melchionda, M., Edwards, A., Briscoe, J., & Sagner, A. (2019). Single cell transcriptomics reveals spatial and temporal dynamics of gene expression in the developing mouse spinal cord. *Development*, *146*(12). https://doi.org/10.1242/dev.173807

Dermitzakis, E. T., & Clark, A. G. (2002). Evolution of Transcription Factor Binding Sites in Mammalian Gene Regulatory Regions: Conservation and Turnover. *Molecular Biology and Evolution*, *19*(7), 1114–1121. https://doi.org/10.1093/oxfordjournals.molbev.a004169

Deutschbauer, A. M., & Davis, R. W. (2005). Quantitative trait loci mapped to single-nucleotide resolution in yeast. *Nature Genetics*, *37*(12), 1333–1340. https://doi.org/10.1038/ng1674

Dong, Y., Hu, J., Fan, L., & Chen, Q. (2017). RNA-Seq-based transcriptomic and metabolomic analysis reveal stress responses and programmed cell death induced by acetic acid in Saccharomyces cerevisiae. *Scientific Reports*, *7*(1), 42659. https://doi.org/10.1038/srep42659

Duveau, F., Metzger, B. P. H., Gruber, J. D., Mack, K., Sood, N., Brooks, T. E., & Wittkopp, P. J. (2014). Mapping Small Effect Mutations in Saccharomyces cerevisiae: Impacts of Experimental Design and Mutational Properties. *G3: Genes, Genomes, Genetics*, g3.114.011783. https://doi.org/10.1534/g3.114.011783

Ehrenreich, I. M., Torabi, N., Jia, Y., Kent, J., Martis, S., Shapiro, J. A., Gresham, D., Caudy, A. A., & Kruglyak, L. (2010). Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature*, *464*(7291), 1039–1042. https://doi.org/10.1038/nature08923

Elion, E. A., Satterberg, B., & Kranz, J. E. (1993). FUS3 phosphorylates multiple components of the mating signal transduction cascade: Evidence for STE12 and FAR1. *Molecular Biology of the Cell*, *4*(5), 495–510. https://doi.org/10.1091/mbc.4.5.495

Erdman, S., Lin, L., Malczynski, M., & Snyder, M. (1998). Pheromone-regulated Genes Required for Yeast Mating Differentiation. *Journal of Cell Biology*, *140*(3), 461–483. https://doi.org/10.1083/jcb.140.3.461

Fay, J. C. (2013). The molecular basis of phenotypic variation in yeast. *Current Opinion in Genetics & Development*, *23*(6), 672–677. https://doi.org/10.1016/j.gde.2013.10.005

Foss, E. J., Radulovic, D., Shaffer, S. A., Goodlett, D. R., Kruglyak, L., & Bedalov, A. (2011). Genetic Variation Shapes Protein Networks Mainly through Non-transcriptional Mechanisms. *PLOS Biology*, *9*(9), e1001144. https://doi.org/10.1371/journal.pbio.1001144

Foss, E. J., Radulovic, D., Shaffer, S. A., Ruderfer, D. M., Bedalov, A., Goodlett, D. R., & Kruglyak, L. (2007). Genetic basis of proteome variation in yeast. *Nature Genetics*, *39*(11), 1369–1375. https://doi.org/10.1038/ng.2007.22

Ghazalpour, A., Bennett, B., Petyuk, V. A., Orozco, L., Hagopian, R., Mungrue, I. N., Farber, C. R., Sinsheimer, J., Kang, H. M., Furlotte, N., Park, C. C., Wen, P.-Z., Brewer, H., Weitz, K., Ii, D. G. C., Pan, C., Yordanova, R., Neuhaus, I., Tilford, C., … Lusis, A. J. (2011). Comparative Analysis of Proteome and Transcriptome Variation in Mouse. *PLOS Genetics*, *7*(6), e1001393. https://doi.org/10.1371/journal.pgen.1001393

Gietz, R. D., & Schiestl, R. H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature Protocols*, *2*(1), 31–34. https://doi.org/10.1038/nprot.2007.13

Gloss, B. S., Signal, B., Cheetham, S. W., Gruhl, F., Kaczorowski, D. C., Perkins, A. C., & Dinger, M. E. (2017). High resolution temporal transcriptomics of mouse embryoid body development reveals complex expression dynamics of coding and noncoding loci. *Scientific Reports*, *7*(1), 6731. https://doi.org/10.1038/s41598-017-06110-5

Goddard, M. R., Godfray, H. C. J., & Burt, A. (2005). Sex increases the efficacy of natural selection in experimental yeast populations. *Nature*, *434*(7033), 636–640. https://doi.org/10.1038/nature03405

Großbach, J., Gillet, L., Clément-Ziza, M., Schmalohr, C. L., Schubert, O. T., Barnes, C. A., Bludau, I., Aebersold, R., & Beyer, A. (2019). Integration of transcriptome, proteome and phosphoproteome data elucidates the genetic control of molecular networks. *BioRxiv*, 703140. https://doi.org/10.1101/703140

Gry, M., Rimini, R., Strömberg, S., Asplund, A., Pontén, F., Uhlén, M., & Nilsson, P. (2009). Correlations between RNA and protein expression profiles in 23 human cell lines. *BMC Genomics*, *10*(1), 365. https://doi.org/10.1186/1471-2164-10-365

Haber, J. E. (2012). Mating-Type Genes and MAT Switching in Saccharomyces cerevisiae. *Genetics*, *191*(1), 33–64. https://doi.org/10.1534/genetics.111.134577

Hamada, K., Fukuchi, S., Arisawa, M., Baba, M., & Kitada, K. (1998). Screening for glycosylphosphatidylinositol (GPI)-dependent cell wall proteins in Saccharomyces cerevisiae. *Molecular and General Genetics MGG*, *258*(1), 53–59. https://doi.org/10.1007/s004380050706

Hecker, M., Lambeck, S., Toepfer, S., van Someren, E., & Guthke, R. (2009). Gene regulatory network inference: Data integration in dynamic models—A review. *Biosystems*, *96*(1), 86–103. https://doi.org/10.1016/j.biosystems.2008.12.004

Herskowitz, I. (1995). MAP kinase pathways in yeast: For mating and more. *Cell*, *80*(2), 187–197. https://doi.org/10.1016/0092-8674(95)90402-6

Hook, S. E., Lampi, M. A., Febbo, E. J., Ward, J. A., & Parkerton, T. F. (2010). Temporal patterns in the transcriptomic response of rainbow trout, Oncorhynchus mykiss, to crude oil. *Aquatic Toxicology*, *99*(3), 320–329. https://doi.org/10.1016/j.aquatox.2010.05.011

Hook, S. E., Skillman, A. D., Small, J. A., & Schultz, I. R. (2007). Temporal changes in gene expression in rainbow trout exposed to ethynyl estradiol. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, *145*(1), 73–85. https://doi.org/10.1016/j.cbpc.2006.10.011

Huang, L., Tang, W., Bu, S., & Wu, W. (2020). BRM: A statistical method for QTL mapping based on bulked segregant analysis by deep sequencing. *Bioinformatics*, *36*(7), 2150-2156. https://doi.org/10.1093/bioinformatics/btz861

Huh, W. K., Falvo, J. V., Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S., & O'Shea, E. K. (2003). Global analysis of protein localization in budding yeast. *Nature*, *425*(6959), 686-691. https://www.nature.com/articles/nature02026

Joshi NA, Fass JN. (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.21) [Software]. Available at https://github.com/najoshi/sickle.

Khan, Z., Bloom, J. S., Amini, S., Singh, M., Perlman, D. H., Caudy, A. A., & Kruglyak, L. (2012). Quantitative measurement of allele-specific protein expression in a diploid yeast hybrid by LC-MS. *Molecular Systems Biology*, *8*(1), 602. https://doi.org/10.1038/msb.2012.34

Knight, J. C. (2004). Allele-specific gene expression uncovered. *Trends in Genetics*, *20*(3), 113–116. https://doi.org/10.1016/j.tig.2004.01.001

Lackner, D. H., Schmidt, M. W., Wu, S., Wolf, D. A., & Bähler, J. (2012). Regulation of transcriptome, translation, and proteome in response to environmental stress in fission yeast. *Genome Biology*, *13*(4), R25. https://doi.org/10.1186/gb-2012-13-4-r25

Lang, G. I., Murray, A. W., & Botstein, D. (2009). The cost of gene expression underlies a fitness trade-off in yeast. *Proceedings of the National Academy of Sciences*, *106*(14), 5755–5760. https://doi.org/10.1073/pnas.0901620106

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, Heng. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv:1303.3997 [q-Bio]*. http://arxiv.org/abs/1303.3997

Li, Y., Roberts, J., AkhavanAghdam, Z., & Hao, N. (2017). Mitogen-activated protein kinase (MAPK) dynamics determine cell fate in the yeast mating response. *Journal of Biological Chemistry*, *292*(50), 20354–20361. https://doi.org/10.1074/jbc.AC117.000548

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., … Louis, E. J. (2009). Population genomics of domestic and wild yeasts. *Nature*, *458*(7236), 337–341. https://doi.org/10.1038/nature07743

Liu, G., Lanham, C., Buchan, J. R., & Kaplan, M. E. (2017). High-throughput transformation of Saccharomyces cerevisiae using liquid handling robots. *PLoS ONE*, *12*(3). https://doi.org/10.1371/journal.pone.0174128

Lutz, S., Brion, C., Kliebhan, M., & Albert, F. W. (2019). DNA variants affecting the expression of numerous genes in trans have diverse mechanisms of action and evolutionary histories. *PLOS Genetics*, *15*(11), e1008375. https://doi.org/10.1371/journal.pgen.1008375

Magwene, P. M., Willis, J. H., & Kelly, J. K. (2011). The Statistics of Bulk Segregant Analysis Using Next Generation Sequencing. *PLOS Computational Biology*, *7*(11), e1002255. https://doi.org/10.1371/journal.pcbi.1002255

Mansfeld, B. N., & Grumet, R. (2018). QTLseqr: An R Package for Bulk Segregant Analysis with Next-Generation Sequencing. *The Plant Genome*, *11*(2). https://doi.org/10.3835/plantgenome2018.01.0006

Marguerat, S., Schmidt, A., Codlin, S., Chen, W., Aebersold, R., & Bähler, J. (2012). Quantitative Analysis of Fission Yeast Transcriptomes and Proteomes in Proliferating and Quiescent Cells. *Cell*, *151*(3), 671–683. https://doi.org/10.1016/j.cell.2012.09.019

McManus, C. J., May, G. E., Spealman, P., & Shteyman, A. (2014). Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. *Genome Research*, *24*(3), 422–430. https://doi.org/10.1101/gr.164996.113

Metzger, B. P. H., Duveau, F., Yuan, D. C., Tryban, S., Yang, B., & Wittkopp, P. J. (2016). Contrasting Frequencies and Effects of cis- and trans-Regulatory Mutations Affecting Gene Expression. *Molecular Biology and Evolution*, *33*(5), 1131–1146. https://doi.org/10.1093/molbev/msw011

Miller, C., Schwalb, B., Maier, K., Schulz, D., Dümcke, S., Zacher, B., Mayer, A., Sydow, J., Marcinowski, L., Dölken, L., Martin, D. E., Tresch, A., & Cramer, P. (2011). Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Molecular Systems Biology*, *7*(1), 458. https://doi.org/10.1038/msb.2010.112

Miyajima, I., Nakafuku, M., Nakayama, N., Brenner, C., Miyajima, A., Kaibuchi, K., Arai, K., Kaziro, Y., & Matsumoto, K. (1987). GPA1, a haploid-specific essential gene, encodes a yeast homolog of mammalian G protein which may be involved in mating factor signal transduction. *Cell*, *50*(7), 1011–1019. https://doi.org/10.1016/0092-8674(87)90167-X

Morjan, C. L., & Rieseberg, L. H. (2004). How species evolve collectively: Implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology*, *13*(6), 1341–1356. https://doi.org/10.1111/j.1365-294X.2004.02164.x

Muller, E. M., Mackin, N. A., Erdman, S. E., & Cunningham, K. W. (2003). Fig1p Facilitates Ca^{2+} Influx and Cell Fusion during Mating of *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, *278*(40), 38461–38469. https://doi.org/10.1074/jbc.M304089200

Paliwal, S., Iglesias, P. A., Campbell, K., Hilioti, Z., Groisman, A., & Levchenko, A. (2007). MAPK-mediated bimodal gene expression and adaptive gradient sensing in yeast. *Nature*, *446*(7131), 46–51. https://doi.org/10.1038/nature05561

Parts, L., Cubillos, F. A., Warringer, J., Jain, K., Salinas, F., Bumpstead, S. J., Molin, M., Zia, A., Simpson, J. T., Quail, M. A., Moses, A., Louis, E. J., Durbin, R., & Liti, G. (2011). Revealing the genetic structure of a trait by sequencing a population under selection. *Genome Research*, *21*(7), 1131–1138. https://doi.org/10.1101/gr.116731.110

Parts, L., Liu, Y.-C., Tekkedil, M., Steinmetz, L. M., Caudy, A. A., Fraser, A. G., Boone, C., Andrews, B. J., & Rosebrock, A. P. (2014). Heritability and genetic basis of protein level variation in an outbred population. *Genome Research*, gr.170506.113. https://doi.org/10.1101/gr.170506.113

Pascual-Ahuir, A., González-Cantó, E., Juyoux, P., Pable, J., Poveda-Huertes, D., Saiz-Balbastre, S., Squeo, S., Ureña-Marco, A., Vanacloig-Pedros, E., Zaragoza-Infante, L., & Proft, M. (2019). Dose dependent gene expression is dynamically modulated by the history, physiology and age of yeast cells. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1862*(4), 457–471. https://doi.org/10.1016/j.bbagrm.2019.02.009

Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J.-M., Istace, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., … Schacherer, J. (2018). Genome evolution across 1,011 Saccharomyces cerevisiae isolates. *Nature*, *556*(7701), 339–344. https://doi.org/10.1038/s41586-018-0030-5

Picotti, P., Clément-Ziza, M., Lam, H., Campbell, D. S., Schmidt, A., Deutsch, E. W., Röst, H., Sun, Z., Rinner, O., Reiter, L., Shen, Q., Michaelson, J. J., Frei, A., Alberti, S., Kusebauch, U., Wollscheid, B., Moritz, R. L., Beyer, A., & Aebersold, R. (2013). A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature*, *494*(7436), 266–270. https://doi.org/10.1038/nature11835

Pollard, D. A., Asamoto, C. K., Rahnamoun, H., Abendroth, A. S., Lee, S. R., & Rifkin, S. A. (2016). Natural Genetic Variation Modifies Gene Expression Dynamics at the Protein Level During Pheromone Response in Saccharomyces cerevisiae. *BioRxiv*, 090480. https://doi.org/10.1101/090480

Pomp, D., Allan, M. F., & Wesolowski, S. R. (2004). Quantitative genomics: Exploring the genetic architecture of complex trait predisposition,. *Journal of Animal Science*, *82*(suppl_13), E300–E312. https://doi.org/10.2527/2004.8213_supplE300x

R Core Team. (2019). Core R: A Language and Environment for Statistical Computing, Version 3.5. 3. *Vienna: R Foundation for Statistical Computing*. URL https://www.R-project.org/

Rines, D. R., He, X., & Sorger, P. K. (2002). Quantitative microscopy of green fluorescent protein-labeled yeast. In *Methods in Enzymology* (Vol. 351, pp. 16–34). Academic Press. https://doi.org/10.1016/S0076-6879(02)51839-5

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. (2000). Signaling and Circuitry of Multiple MAPK Pathways Revealed by a Matrix of Global Gene Expression Profiles. *Science*, *287*(5454), 873–880. https://doi.org/10.1126/science.287.5454.873

Ronald, J., Brem, R. B., Whittle, J., & Kruglyak, L. (2005). Local Regulatory Variation in Saccharomyces cerevisiae. *PLOS Genetics*, *1*(2), e25. https://doi.org/10.1371/journal.pgen.0010025

Sadhu, M. J., Bloom, J. S., Day, L., & Kruglyak, L. (2016). CRISPR-directed mitotic recombination enables genetic mapping without crosses. *Science*, *352*(6289), 1113–1116. https://doi.org/10.1126/science.aaf5124

Sadhu, M. J., Bloom, J. S., Day, L., Siegel, J. J., Kosuri, S., & Kruglyak, L. (2018). Highly parallel genome variant engineering with CRISPR–Cas9. *Nature Genetics*, *50*(4), 510–514. https://doi.org/10.1038/s41588-018-0087-y

Salinas, F., de Boer, C. G., Abarca, V., García, V., Cuevas, M., Araos, S., Larrondo, L. F., Martínez, C., & Cubillos, F. A. (2016). Natural variation in non-coding regions underlying phenotypic diversity in budding yeast. *Scientific Reports*, *6*(1), 21849. https://doi.org/10.1038/srep21849

Salunkhe, A. S., Poornima, R., Prince, K. S. J., Kanagaraj, P., Sheeba, J. A., Amudha, K., Suji, K. K., Senthil, A., & Babu, R. C. (2011). Fine Mapping QTL for Drought Resistance Traits in Rice (Oryza sativa L.) Using Bulk Segregant Analysis. *Molecular Biotechnology*, *49*(1), 90–95. https://doi.org/10.1007/s12033-011-9382-x

Schadt, E. E., Lamb, J., Yang, X., Zhu, J., Edwards, S., GuhaThakurta, D., Sieberts, S. K., Monks, S., Reitman, M., Zhang, C., Lum, P. Y., Leonardson, A., Thieringer, R., Metzger, J. M., Yang, L., Castle, J., Zhu, H., Kash, S. F., Drake, T. A., … Lusis, A. J. (2005). An integrative genomics approach to infer causal associations between gene expression and disease. *Nature Genetics*, *37*(7), 710–717. https://doi.org/10.1038/ng1589

Schmidt, D., Wilson, M. D., Ballester, B., Schwalie, P. C., Brown, G. D., Marshall, A., Kutter, C., Watt, S., Martinez-Jimenez, C. P., Mackay, S., Talianidis, I., Flicek, P., & Odom, D. T. (2010). Five-Vertebrate ChIP-seq Reveals the Evolutionary Dynamics of Transcription Factor Binding. *Science*, *328*(5981), 1036–1040. https://doi.org/10.1126/science.1186176

Schuller, D., Cardoso, F., Sousa, S., Gomes, P., Gomes, A. C., Santos, M. A. S., & Casal, M. (2012). Genetic Diversity and Population Structure of Saccharomyces cerevisiae Strains Isolated from Different Grape Varieties and Winemaking Regions. *PLOS ONE*, *7*(2), e32507. https://doi.org/10.1371/journal.pone.0032507

Seroude, L., Brummel, T., Kapahi, P., & Benzer, S. (2002). Spatio-temporal analysis of gene expression during aging in Drosophila melanogaster. *Aging Cell*, *1*(1), 47–56. https://doi.org/10.1046/j.1474-9728.2002.00007.x

Sethiya, P., Rai, M. N., Rai, R., Parsania, C., Tan, K., & Wong, K. H. (2020). Transcriptomic analysis reveals global and temporal transcription changes during Candida glabrata adaptation to an oxidative environment. *Fungal Biology*, *124*(5), 427–439. https://doi.org/10.1016/j.funbio.2019.12.005

Shapiro, M. D., Marks, M. E., Peichel, C. L., Blackman, B. K., Nereng, K. S., Jónsson, B., Schluter, D., & Kingsley, D. M. (2004). Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature*, *428*(6984), 717–723. https://doi.org/10.1038/nature02415

Signor, S. A., & Nuzhdin, S. V. (2018). The Evolution of Gene Expression in cis and trans. *Trends in Genetics*, *34*(7), 532–544. https://doi.org/10.1016/j.tig.2018.03.007

Stern, D. L., & Orgogozo, V. (2008). The Loci of Evolution: How Predictable Is Genetic Evolution? *Evolution*, *62*(9), 2155–2177. https://doi.org/10.1111/j.1558-5646.2008.00450.x

Strassburg, K., Walther, D., Takahashi, H., Kanaya, S., & Kopka, J. (2010). Dynamic Transcriptional and Metabolic Responses in Yeast Adapting to Temperature Stress. *OMICS: A Journal of Integrative Biology*, *14*(3), 249–259. https://doi.org/10.1089/omi.2009.0107

Straub, L. (2011). Beyond the Transcripts: What Controls Protein Variation? *PLOS Biology*, *9*(9), e1001146. https://doi.org/10.1371/journal.pbio.1001146

Tadauchi, T., Inada, T., Matsumoto, K., & Irie, K. (2004). Posttranscriptional Regulation of HO Expression by the Mkt1-Pbp1 Complex. *Molecular and Cellular Biology*, *24*(9), 3670–3681. https://doi.org/10.1128/MCB.24.9.3670-3681.2004

Timpson, N. J., Greenwood, C. M. T., Soranzo, N., Lawson, D. J., & Richards, J. B. (2018). Genetic architecture: The shape of the genetic contribution to human traits and disease. *Nature Reviews Genetics*, *19*(2), 110–124. https://doi.org/10.1038/nrg.2017.101

Treusch, S., Albert, F. W., Bloom, J. S., Kotenko, I. E., & Kruglyak, L. (2015). Genetic Mapping of MAPK-Mediated Complex Traits Across S. cerevisiae. *PLoS Genetics*, *11*(1), e1004913. https://doi.org/10.1371/journal.pgen.1004913

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]*, *11*(1110), 11.10.1-11.10.33. https://doi.org/10.1002/0471250953.bi1110s43

Verd, B., Crombach, A., & Jaeger, J. (2017). Dynamic Maternal Gradients Control Timing and Shift-Rates for Drosophila Gap Gene Expression. *PLOS Computational Biology*, *13*(2), e1005285. https://doi.org/10.1371/journal.pcbi.1005285

Vogel, C., & Marcotte, E. M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics*, *13*(4), 227–232. https://doi.org/10.1038/nrg3185

Wiberg, R. A. W., Gaggiotti, O. E., Morrissey, M. B., & Ritchie, M. G. (2017). Identifying consistent allele frequency differences in studies of stratified populations. *Methods in ecology and evolution*, *8*(12), 1899-1909. https://doi.org/10.1111/2041-210X.12810

Wickner, R. B. (1987). MKT1, a nonessential Saccharomyces cerevisiae gene with a temperature-dependent effect on replication of M2 double-stranded RNA. *Journal of Bacteriology*, *169*(11), 4941–4945. https://doi.org/10.1128/jb.169.11.4941-4945.1987

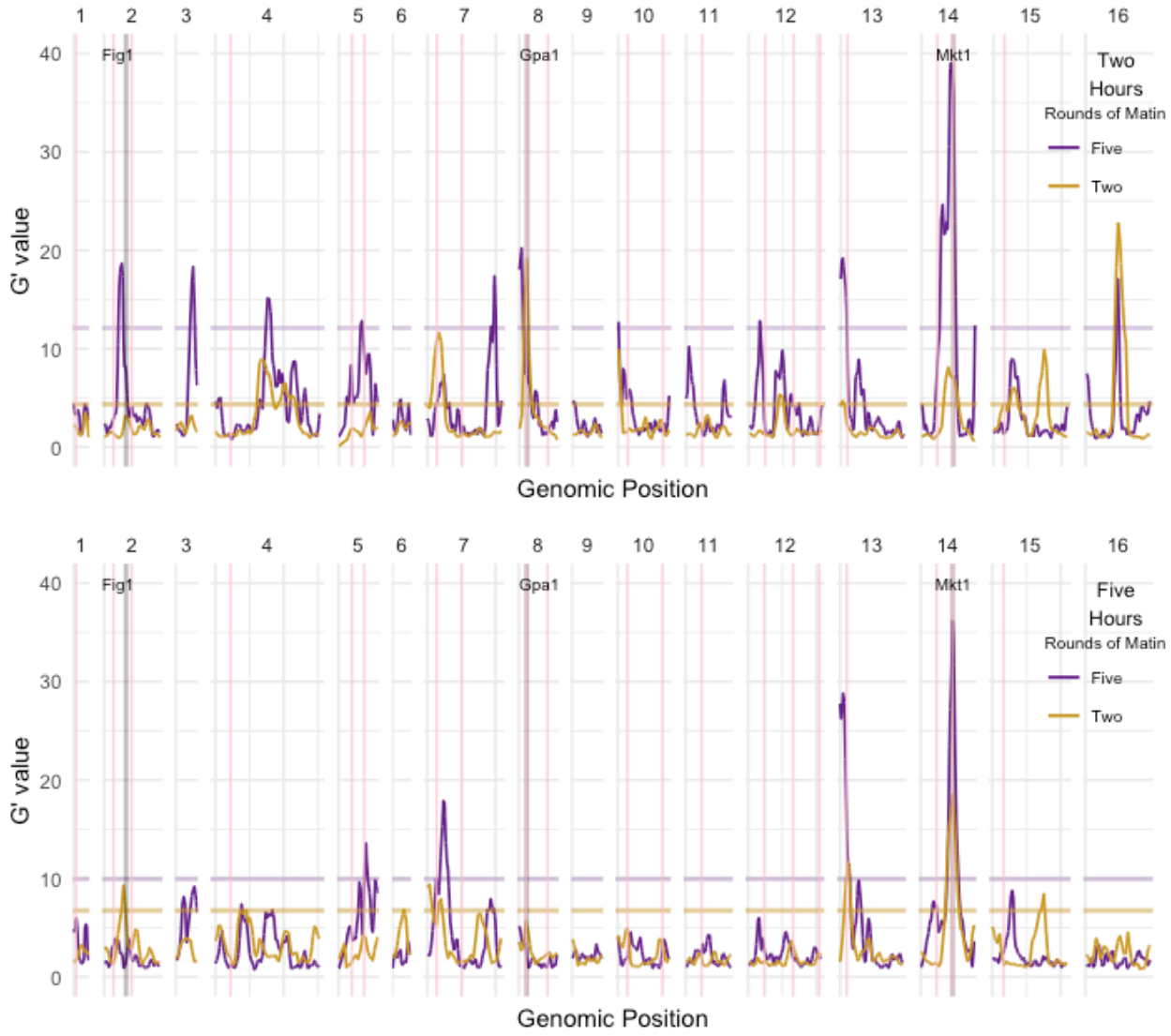
Wittkopp, P. J., Haerum, B. K., & Clark, A. G. (2004). Evolutionary changes in cis and trans gene regulation. *Nature*, *430*(6995), 85–88. https://doi.org/10.1038/nature02698

Wittkopp, P. J., Haerum, B. K., & Clark, A. G. (2008). Independent Effects of cis- and trans-regulatory Variation on Gene Expression in Drosophila melanogaster. *Genetics*, *178*(3), 1831–1835. https://doi.org/10.1534/genetics.107.082032

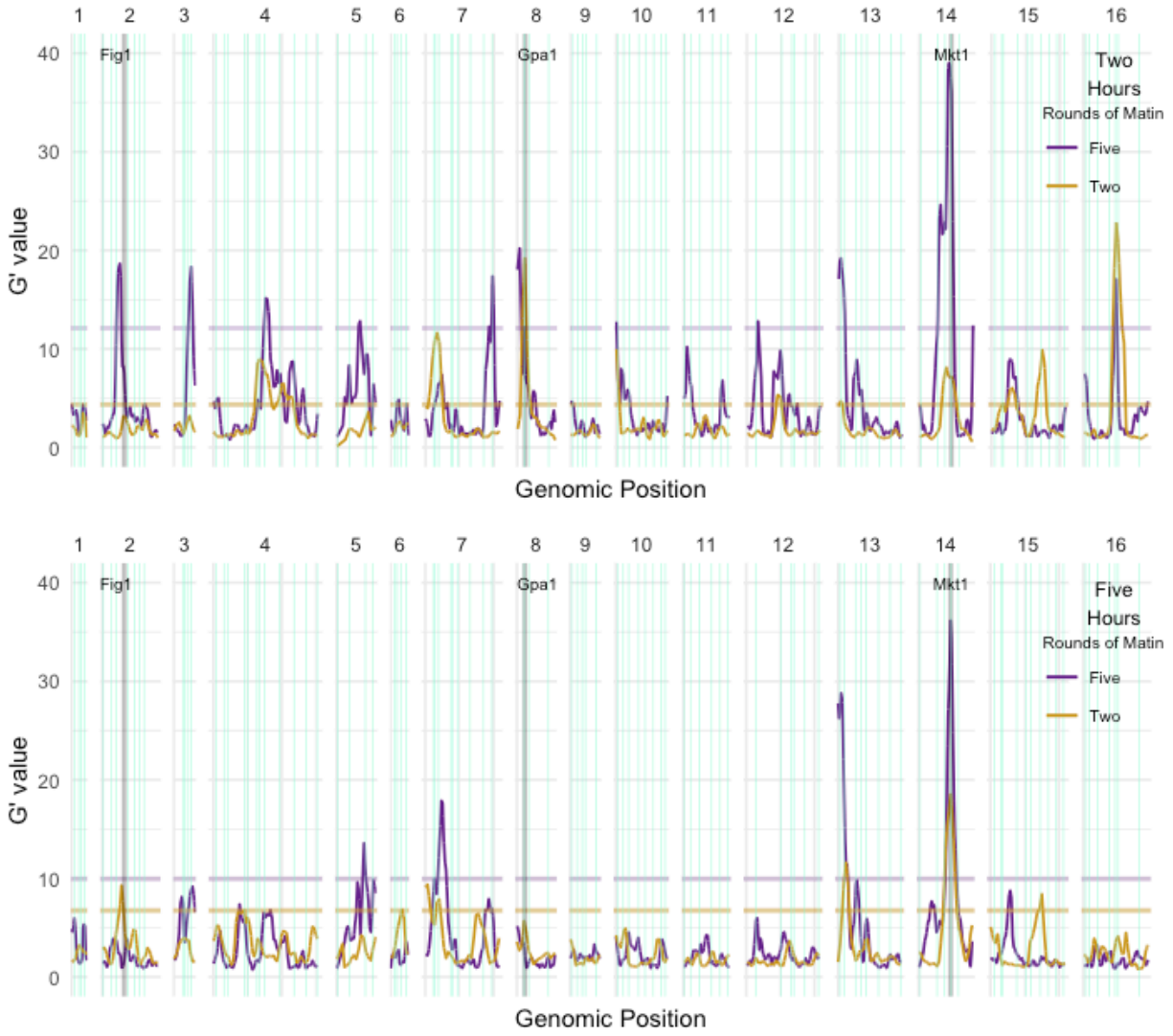
Yan Tong, A. H., & Boone, C. (2006). Synthetic Genetic Array Analysis in Saccharomyces cerevisiae. In W. Xiao (Ed.), *Yeast Protocol* (pp. 171–191). Humana Press. https://doi.org/10.1385/1-59259-958-3:171

Zheng, W., Zhao, H., Mancera, E., Steinmetz, L. M., & Snyder, M. (2010). Genetic analysis of variation in transcription factor binding in yeast. *Nature*, *464*(7292), 1187–1191. https://doi.org/10.1038/nature08934

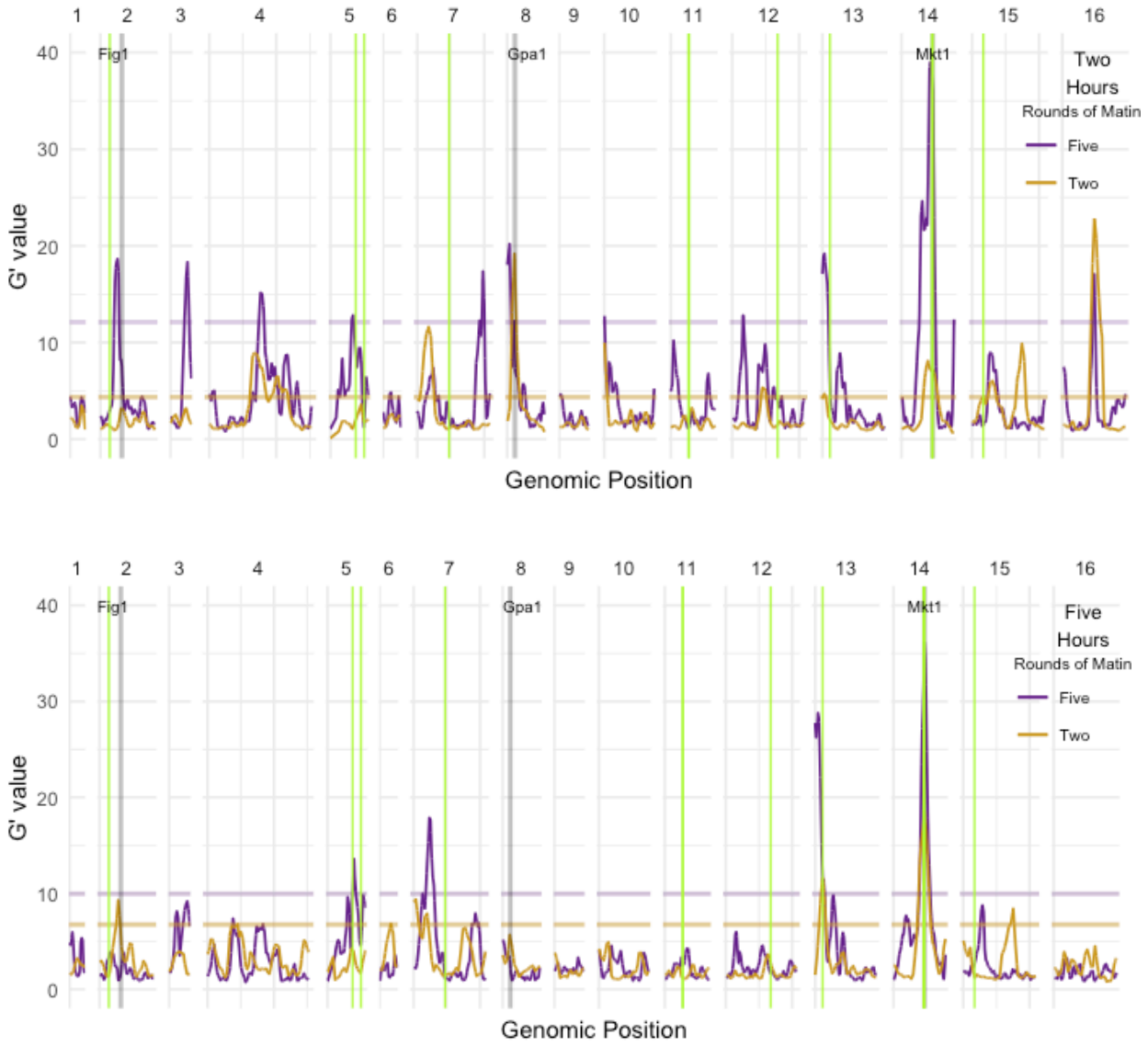
Supplemental Figures:



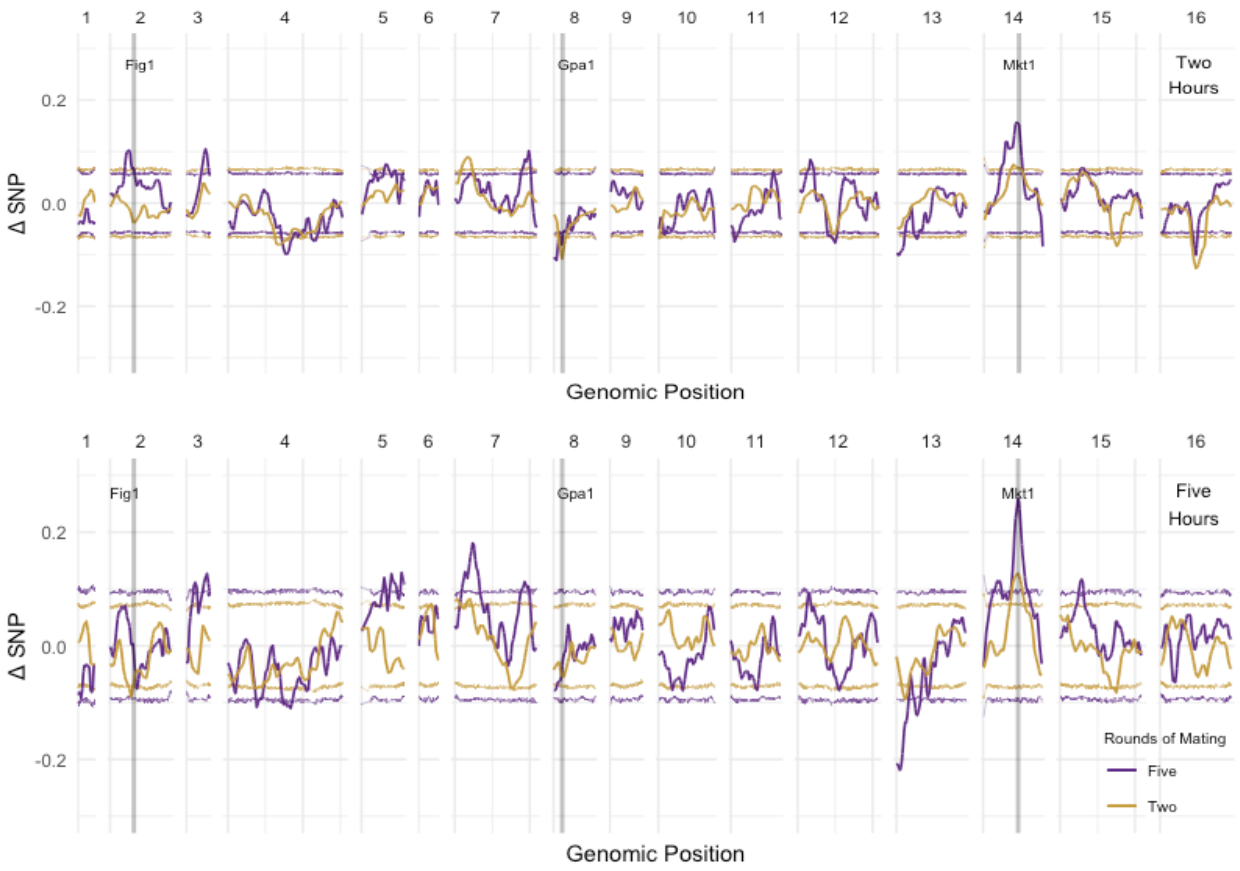
Supplementary Figure 1: Whole-genome FIG1 QTL-maps highlighting the location of steady-state pQTL hotspots (pink bars) (Albert et al. 2014). Peaks represent genomic loci associated with FIG1 protein expression 2 hours (top) after exposure to mating pheromone, and 5 hours (bottom) after exposure to mating pheromone. Horizontal lines represent an FDR cutoff of 0.05. Purple and gold maps correspond to BSA experiments with segregant populations derived from five rounds and two rounds of random mating respectively.



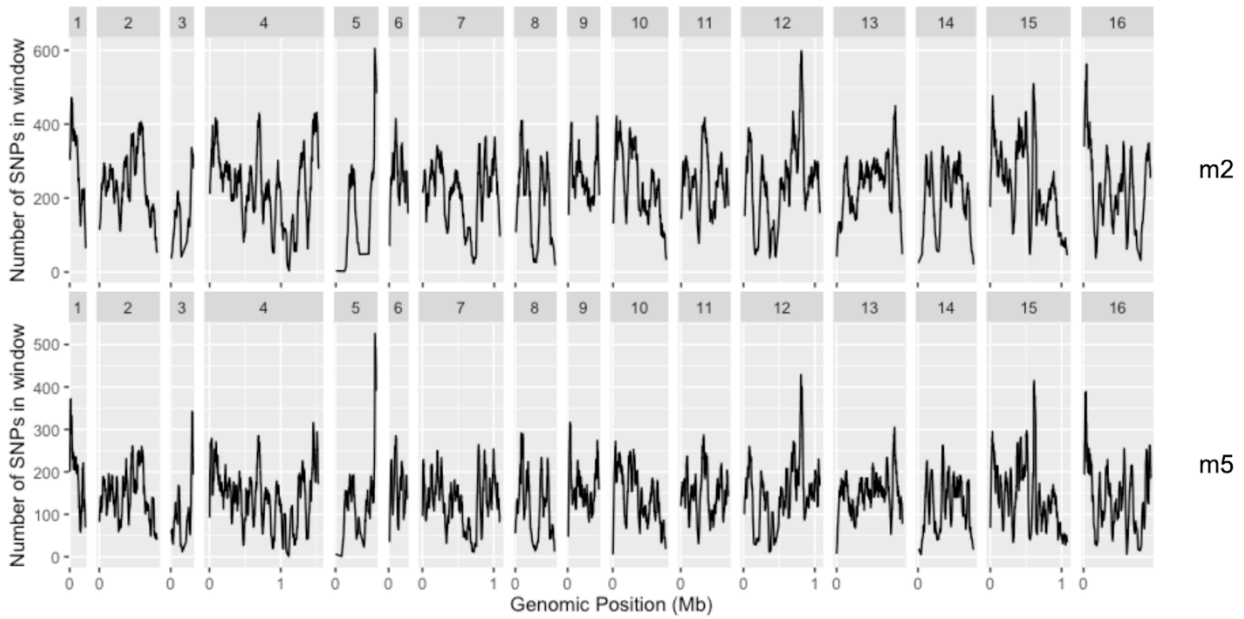
Supplementary Figure 2: Whole-genome FIG1 QTL-maps highlighting the location of steady-state eQTL hotspots (blue bars) (Albert et al. 2018). Peaks represent genomic loci associated with FIG1 protein expression 2 hours (top) after exposure to mating pheromone, and 5 hours (bottom) after exposure to mating pheromone. Horizontal lines represent an FDR cutoff of 0.05. Purple and gold maps correspond to BSA experiments with segregant populations derived from five rounds and two rounds of random mating respectively.



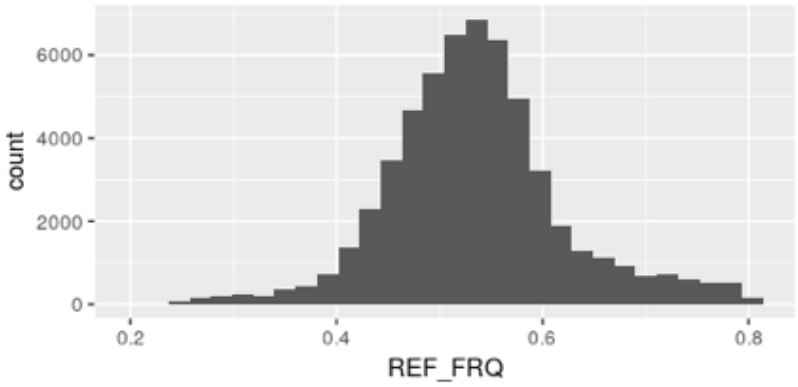
Supplementary Figure 3: Whole-genome FIG1 QTL-maps highlighting the location of steady-state pQTL lacking cognate eQTL (green bars) (Albert et al. 2014). Peaks represent genomic loci associated with FIG1 protein expression 2 hours (top) after exposure to mating pheromone, and 5 hours (bottom) after exposure to mating pheromone. Horizontal lines represent an FDR cutoff of 0.05. Purple and gold maps correspond to BSA experiments with segregant populations derived from five rounds and two rounds of random mating respectively.



Supplementary Figure 4: : Whole-genome FIG1 QTL-maps. Peaks represent genomic loci with allele frequency differences between high and low Fig1 protein expression bulks 2 hours (top) and 5 hours (bottom) after exposure to mating pheromone. Purple and gold correspond to BSA experiments with segregant populations derived from five rounds and two rounds of random mating respectively. Horizontal lines reflect a 95% Confidence Interval for either segregant population. Positive values represent loci where the s288c allele was favored in the high bulk and negative values represent loci where the YJM145 allele was favored in the high bulk.



Supplementary Figure 5: Number of SNPs considered in a sliding window for segregant populations derived from two rounds (top) and five rounds (bottom) of mating and sporulation for each chromosome in the yeast genome. SNP density is not evenly distributed across the genome, and some regions, such the sub-telomeric region on right arm of chromosome 5 or the center of chromosome 12, are especially divergent between strain backgrounds.



Supplementary Figure 6: Distribution of reads whose SNPs mapped to the reference genome for the five rounds of mating, two-hour time-point experiment. This trend shows to be consistent across experiments. Genome wide median SNP reference frequencies deviating from 0.5 suggest either 1) that segregant populations were subjected to selection pressures that favored certain reference strain (S288c) alleles such as selection for mating efficiency related alleles, or 2) that mapping biases exist in our SNPsets which may also bias our QTL mapping statistical approach.