

Compensatory *trans*-regulatory alleles minimizing variation in *TDH3* expression are common within

*Saccharomyces cerevisiae*

Brian P. H. Metzger<sup>1,2</sup>, Patricia J. Wittkopp<sup>1,3</sup>

<sup>1</sup> Department of Ecology and Evolutionary Biology  
University of Michigan, Ann Arbor, MI 48109 USA

<sup>2</sup> Department of Ecology and Evolution  
University of Chicago, Chicago, IL 60637 USA

<sup>3</sup> Department of Molecular, Cellular, and Developmental Biology  
University of Michigan, Ann Arbor, MI 48109 USA

Corresponding author:

Patricia J. Wittkopp

Department of Ecology and Evolutionary Biology

Department of Molecular, Cellular, and Developmental Biology

University of Michigan, Ann Arbor, MI 48109-1048 USA

wittkopp@umich.edu

Tel: +1.734.763.1548

Fax: +1.734.763.0544

**Article Type:** Letter

**Running Title:** Complex *trans*-regulatory variation among yeast strains

**Keywords:** gene regulation, compensation, stabilizing selection, mapping, eQTL

### Abstract

Heritable variation in gene expression is common within species. Much of this variation is due to genetic differences outside of the gene with altered expression and is *trans*-acting. This *trans*-

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/evl3.137](https://doi.org/10.1002/evl3.137).

This article is protected by copyright. All rights reserved.

regulatory variation is often polygenic, with individual variants typically having small effects, making the genetic architecture and evolution of *trans*-regulatory variation challenging to study. Consequently, key questions about *trans*-regulatory variation remain, including the variability of *trans*-regulatory variation within a species, how selection affects *trans*-regulatory variation, and how *trans*-regulatory variants are distributed throughout the genome and within a species. To address these questions, we isolated and measured *trans*-regulatory differences affecting *TDH3* promoter activity among 56 strains of *Saccharomyces cerevisiae*, finding that *trans*-regulatory backgrounds varied approximately two-fold in their effects on *TDH3* promoter activity. Comparing this variation to neutral models of *trans*-regulatory evolution based on empirical measures of mutational effects revealed that despite this variability in the effects of *trans*-regulatory backgrounds, stabilizing selection has constrained *trans*-regulatory differences within this species. Using a powerful quantitative trait locus mapping method, we identified ~100 *trans*-acting eQTL in each of three crosses to a common reference strain, indicating that regulatory variation is more polygenic than previous studies have suggested. Loci altering expression were located throughout the genome, and many loci were strain specific. This distribution and prevalence of alleles is consistent with recent theories about the genetic architecture of complex traits. In all mapping experiments, the non-reference strain alleles increased and decreased *TDH3* promoter activity with similar frequencies, suggesting that stabilizing selection maintained many *trans*-acting variants with opposing effects. This variation may provide the raw material for compensatory evolution and larger scale regulatory rewiring observed in developmental systems drift between species.

### **Impact Summary**

Gene expression varies among individuals in a population due to genetic differences in regulatory components. To determine how this variation is distributed within genomes and species, we used a

powerful genetic mapping approach to examine multiple strains of *Saccharomyces cerevisiae*.

Despite evidence of stabilizing selection maintaining gene expression levels among strains, we find hundreds of loci that affect expression of a single gene. These loci vary among strains and include similar frequencies of alleles that increase and decrease expression. As a result, each strain contains a unique set of compensatory alleles that lead to similar levels of gene expression among strains. This regulatory variation might form the basis for large scale regulatory rewiring observed between distantly related species.

## Introduction

Heritable variation in gene expression results from genetic variation affecting *cis*-regulatory elements (e.g., promoters, enhancers) and *trans*-acting factors (e.g., proteins, RNAs). These *trans*-regulatory changes are located throughout the genome and are the major source of regulatory variation within species (Wittkopp *et al.* 2004; Wang *et al.* 2007; Sung *et al.* 2009; Zhang and Borevitz 2009; Emerson *et al.* 2010; Bell *et al.* 2013; Schaefer *et al.* 2013; Suvorov *et al.* 2013; Coolon *et al.* 2014; Chen *et al.* 2015). The number, identity, and effects of individual loci contributing to variation in gene expression has been determined in a variety of species using expression quantitative trait locus (eQTL) mapping (Gilad *et al.* 2008; Hansen *et al.* 2008; Majewski and Pastinen 2011; Cubillos *et al.* 2012; Nica and Dermitzakis 2013; Westra and Franke 2014; Albert and Kruglyak 2015; Pai *et al.* 2015), with the most extensive dissection of eQTL coming from studies of two strains of the baker's yeast *Saccharomyces cerevisiae* (Brem *et al.* 2002, 2005; Schadt *et al.* 2003; Yvert *et al.* 2003; Brem and Kruglyak 2005; Ronald *et al.* 2005; Smith and Kruglyak 2008; Albert *et al.* 2014, 2018; Parts *et al.* 2014). These studies have found that (i) expression differences are typically associated with ~10 or fewer eQTL, (ii) most eQTL have individually small effects on expression, and

(iii) most eQTL do not overlap the gene whose expression they affect and thus likely contribute to *trans*-regulatory differences.

Traditional eQTL mapping approaches require genotype and expression data for many individuals to detect significant effects. Consequently, studies mapping the genetic basis of regulatory differences have largely been limited to two strains within any given species. In cases where the extent and variability of regulatory variation has been studied within a species, experiments have focused on *cis*-regulatory variation for technical reasons (de Meaux 2005; Gruber and Long 2009; Kang *et al.* 2016; Moyerbrailean *et al.* 2016; Salinas *et al.* 2016; Kita *et al.* 2017). As a result, key questions about the extent, variability, and genetic basis of *trans*-regulatory variation segregating within a species remain unanswered. For example, do multiple *trans*-regulatory variants affecting a gene's expression often segregate at the same locus within a species? How different are the suites of *trans*-acting eQTL affecting a gene's expression among individuals or strains? Are the effects of *trans*-regulatory variants at different loci often in the same direction, or do they typically have opposing effects, canceling one another out? Addressing these questions requires identifying *trans*-acting eQTL and their effects on expression among multiple individuals or strains of the same species.

While genome-wide association studies use population level variation to identify eQTL, they do not meet this need because they can only detect eQTL alleles that are shared by many individuals.

In addition to these questions about the variability in genetic architecture of *trans*-regulatory variation, questions also remain about the impact of selection on this variation. Prior work has shown that gene expression levels are broadly constrained by stabilizing selection (Denver *et al.* 2005; Gilad *et al.* 2006) and variation in *cis*-regulatory eQTL appears to be limited by purifying selection (Josephs *et al.* 2015; Kita *et al.* 2017). But the impact of natural selection on the number, identity, or genomic distribution of *trans*-acting eQTL is less clear, and there are reasons to suspect

that it might be different than for *cis*-acting eQTL. For example, prior work suggests that *trans*-regulatory mutations arise more frequently than *cis*-regulatory mutations, but tend to have smaller effects on the focal gene's expression (Metzger *et al.* 2016). In addition, *trans*-regulatory mutations are more likely to be recessive and have greater pleiotropic effects than *cis*-regulatory mutations (Stern 2000; Landry *et al.* 2007; Fay and Wittkopp 2008; Lemos *et al.* 2008; Gruber *et al.* 2012). Any or all of these factors might cause selection for the level of gene expression to have different impacts on *cis*- and *trans*-regulatory variation.

Here, we examine *trans*-regulatory variation segregating among genetically distinct strains of *S. cerevisiae*. We focus on the extent and variability of, evolutionary forces acting on, and genetic basis for, *trans*-regulatory variation affecting expression of the *TDH3* gene, which encodes a glyceraldehyde-3-phosphate dehydrogenase. This gene was chosen because prior work has estimated the effects of new *trans*-regulatory mutations on its expression (Gruber *et al.* 2012; Metzger *et al.* 2016) as well as the fitness consequences of changing its expression (Duveau *et al.* 2017, 2018), allowing us to compare the *trans*-regulatory variation segregating in *S. cerevisiae* to empirically-informed models of neutral evolution. We find that although differences in *trans*-regulation affecting *TDH3* promoter activity are common among strains, they generate less variation in *TDH3* promoter activity than predicted by neutral models, suggesting that stabilizing selection has acted on *trans*-regulatory variation affecting *TDH3* promoter activity in the wild. We then use a powerful genetic mapping approach to determine differences in the genetic architecture of this *trans*-regulatory variation by identifying eQTL between each of three strains of *S. cerevisiae* and a common reference strain. In each of these three eQTL mapping experiments, we find an order of magnitude more eQTL affecting activity of the *TDH3* promoter in *trans* than previously known. These loci are often different among strains, have opposing effects on expression, and are spread throughout the genome, indicating diverse sources of *trans*-regulatory variation segregating within

*S. cerevisiae*. These results agree with theoretical predictions that stabilizing selection can maintain genetic variation for polygenic traits (Lande 1976; Dover and Flavell 1984; Turelli 1984; Barton 1986, 1989). They also suggest that natural populations harbor greater regulatory variation than suggested by differences between a single pair of strains and that this variation can impact the evolution of regulatory systems.

## Results and Discussion

To isolate the effects of *trans*-regulatory variants segregating among *S. cerevisiae* strains on *TDH3* promoter activity, we inserted a Yellow Fluorescent Protein (YFP) coding sequence under control of the 678bp *TDH3* promoter allele from the BY lab strain into 56 distinct *S. cerevisiae* strains (Figure 1A). These strains (i) were isolated from a range of environments, (ii) differ at more than 100,000 SNPs and small indels, many larger copy number variants chromosomal rearrangements, and (iii) encompass much of the genetic and phenotypic diversity observed within the species (MacLean *et al.* 2017; Peter *et al.* 2018). For each strain, we measured YFP fluorescence in twelve biological replicate populations grown in rich media and used the measured YFP fluorescence to estimate changes in *TDH3* mRNA levels due to differences in *trans*-regulation among strains (Duveau *et al.* 2018). We observed that *trans*-regulatory variation caused differences in expression that ranged from 71% to 147% of the reference strain (Figure 1B, C). This variability in *trans*-regulation was nearly double the variability of *cis*-regulation described in a previous study among a similar set of strains (Figure S1A) (Metzger *et al.* 2015). We detected significant phylogenetic structure for *trans*-regulatory differences among strains, with more closely related strains having on average more similar *TDH3* promoter activity than more distantly related strains ( $\lambda = 0.59$ ,  $p = 0.013$ ;  $K = 0.49$ ,  $p = 0.012$ ; Figure 1D).

In the absence of natural selection, the effects on expression of naturally occurring regulatory variation should be similar to the effects on expression of new mutations. Differences between the effects of naturally occurring variation and new mutations thus provide evidence of natural selection (Metzger *et al.* 2015). To determine whether natural selection has impacted *trans*-regulatory variation affecting *TDH3* promoter activity, we constructed models of neutral evolution using the effects of new mutations defined in prior work (Metzger *et al.* 2016). These mutations were generated with the mutagen EMS (Metzger *et al.* 2016), which mimics the most common type of point mutation in yeast (Zhu *et al.* 2014) and the most common type of polymorphism found among natural yeast strains (G→A and C→T) (MacLean *et al.* 2017). We simulated the neutral evolution of *trans*-regulatory variation affecting *TDH3* promoter activity by sampling these *trans*-regulatory mutational effects on *TDH3* promoter activity (Figure 2A) and tracking how expression changed with the addition of each new mutation, assuming additivity (Figure 2B). We repeated this sampling process 10,000 times and used the observed distributions of expression levels after the addition of each new mutation to define the probability with which we expect to see a given expression level evolve neutrally from the common ancestor after a particular number of genetic changes. Epistasis was ignored in this simulation because its prevalence and effects are unknown for these mutations; it could either increase or decrease the range of expression differences arising from mutation alone. (For a more complete discussion of modeling assumptions and rationale, see the Supplementary Methods section.)

We compared our neutral projections to the observed differences in *TDH3* promoter activity among strains, using the genetic relationships among strains to infer how *TDH3* promoter activity changed along each branch of the phylogeny (Figure 2C). We found that there was significantly less *trans*-regulatory variability among strains than predicted to arise from mutation alone, suggesting that natural selection has constrained *TDH3* promoter activity ( $p < 0.0001$ , Figure S1B, C). Indel mutations

were not included in our distribution of mutational effects but are likely contributing to expression differences among strains; however, these mutations are expected to cause even larger deviations in gene expression than point mutations, further increasing the variation in gene expression expected under neutrality and making our conclusion conservative. To determine whether our inference of stabilizing selection was robust to uncertainty in the inferred phylogenetic relationships among strains and the inferences of changes in *TDH3* promoter activity on the phylogeny, we repeated this analysis using the total genetic distance between pairs of strains instead of the phylogenetic relationships among strains. We again found less *trans*-regulatory variability in *TDH3* promoter activity among strains than predicted by the neutral model, further supporting the hypothesis that *trans*-regulatory variation affecting *TDH3* promoter activity has evolved under stabilizing selection (Figure 2D, Figure S1D).

We also tested for evidence of natural selection acting on *TDH3* *trans*-regulatory variation using a more traditional approach that does not rely on empirical estimates of the effects of new *trans*-regulatory mutations. Specifically, we fit the  $P_{TDH3}$ -*YFP* reporter activity and phylogenetic relationships among strains to two models of quantitative trait evolution: a Brownian Motion model of neutral quantitative trait evolution and an Ornstein-Uhlenbeck model that incorporates stabilizing selection (Bedford and Hartl 2009). We found that the Ornstein-Uhlenbeck model fit the data significantly better than the neutral Brownian motion model ( $p = 0.00007$ , chi-square test, Figure S1E, F), again suggesting that *trans*-regulatory variation affecting *TDH3* promoter activity in *S. cerevisiae* has been shaped by stabilizing selection. The magnitude of effects of *trans*-regulatory variation on *TDH3* promoter activity are expected to decrease fitness by less than 0.1% for 80% of strains based on prior work describing the relationship between *TDH3* expression level and fitness in rich media (Duveau *et al.* 2018), with the largest deviation in *TDH3* promoter activity (71% of wild type) expected to decrease fitness by 0.5% (Figure S1G). Given the large effective population size of



*S. cerevisiae*, we conclude that weak stabilizing selection has constrained the *trans*-regulatory evolution of *TDH3*, although we cannot rule out that selection acting on pleiotropic effects of these regulatory variants might also have contributed to this signal.

In the presence of stabilizing selection, gene expression can be kept similar among strains by purging mutations that alter expression or by maintaining sets of variants with off-setting, or compensatory, effects on expression in the population. To determine which of these mechanisms is more likely to have minimized differences among strains in the *trans*-regulatory effects on *TDH3* promoter activity, we used eQTL mapping to examine the genetic architecture of *trans*-regulatory variation affecting *TDH3* promoter activity in three strains (M22, YPS1000, SK1) relative to a common reference strain (BY). These strains differ in the effects of their *trans*-regulatory background on *TDH3* promoter activity (from 101% to 147%) as well as their phylogenetic relatedness (Figure S2A), making them ideal for determining how the genetic architecture of *TDH3 trans*-regulation varies within the species (Figure S2A). For each focal strain, we used extreme QTL mapping with three rounds of crossing followed by three rounds of selection on  $P_{TDH3}$ -YFP expression using fluorescence assisted cell sorting (FACS) (Figure 3A) (Ehrenreich *et al.* 2010, 2012; Kofler *et al.* 2011; Parts *et al.* 2011; Cubillos *et al.* 2013; Albert *et al.* 2014; Dubeau *et al.* 2014; Schlötterer *et al.* 2014).

In all three crosses we identified ~100 eQTL at a false discovery rate of 10% (n=113 for M22 vs BY, n=101 for YPS1000 vs BY, n=99 for SK1 vs BY) (Figure 3B). This is approximately ten-fold greater than the number of eQTL identified for most genes in a prior eQTL mapping study (Albert *et al.* 2018) between BY and another strain of *S. cerevisiae*, RM11, that is closely related to M22 (Figure S2A), indicating considerable power to detect individual eQTL. The non-BY alleles were evenly split between those that increased and decreased expression for all three mapping experiments (55 of 113 M22 alleles increase expression, 53 of 101 YPS1000 alleles increase expression, and 51 of 99 SK1

alleles increase expression,  $p > 0.6$  for all, binomial test) (Figure 3D). This similarity in the frequency of eQTL increasing and decreasing expression could result from neutral evolution (assuming mutations increasing and decreasing *TDH3* expression arise with similar frequency) or stabilizing selection; however, the evidence of stabilizing selection described above suggests that similar *trans*-regulatory effects on *TDH3* promoter activity are observed among strains of *S. cerevisiae* because compensatory alleles are maintained in the population. Consistent with this conclusion, we found that repeating these eQTL mapping experiments with crossing limited to a single round resulted in considerably fewer eQTL identified, regardless of the number of rounds of selection (black points in Figure 3C, Figure S3). By contrast, reducing the rounds of selection resulted in decreased statistical significance for many eQTL, but did not change the location or direction of effects for most eQTL inferred (red points in Figure 3C, Figure S3). These observations, which were robust to changing the statistical threshold used to call eQTLs (Table S5 and S6), suggest that although additional rounds of selection allowed eQTL with smaller effects to be identified, the high number of eQTL detected results primarily from increased recombination during multiple rounds of meiosis breaking apart physically close eQTL with opposite effects on expression.

To determine the similarity in loci harboring *trans*-regulatory variation affecting *TDH3* promoter activity among strains, we compared the genomic locations of eQTL identified in each pair of strains. If *trans*-regulatory variation is caused by the same loci in all strains, the ~100 eQTLs identified in each comparison should map to similar genomic regions. However, if the sources of *trans*-regulatory variation affecting *TDH3* promoter activity segregating in *S. cerevisiae* are more diverse, eQTL identified in each comparison should map to different genomic regions. We found that the 313 eQTL identified mapped to 180 non-overlapping regions of the genome, with 27% (49 of 180) of these regions containing eQTL in only two of the comparisons and 22% (40 of 180) of these regions containing eQTL in all three comparisons (Figure 3E). Such shared eQTL regions may contain genes

that contribute to variation in *trans*-regulation of *TDH3* promoter activity in multiple strains; however, the 49% of loci overlapping in at least two strains is not greater than expected by chance given the number and width of eQTL observed ( $p = 0.08$ , permutation test, 95% CI: 41-50%). Furthermore, in these shared genomic regions, only 18% of non-BY eQTL alleles had the same direction of effect on *TDH3* promoter activity in two comparisons (26 of 119 for increases, 18 of 120 for decreases), and only 6% of non-BY eQTL alleles had the same direction of effect in all three comparisons (7 of 119 for increases, 8 of 120 for decreases;  $p > 0.53$ , permutation test, Figure 3F-G). These results were robust to the FDR used for identifying peaks (Tables S7 and S8). This lack of consistency in the direction of eQTL effects suggests that even if the same underlying loci contribute to *trans*-regulatory variation in multiple strains, the exact polymorphisms and their effects on *TDH3* promoter activity are likely to differ among strains.

To determine whether differences in the eQTL inferred among strains were more likely to be due to biological differences or poor reproducibility among independent experiments, we repeated the mapping experiment between M22 and BY and compared the eQTL identified in the two replicate mapping experiments (Table S9). Of the 74 eQTL found in the second M22/BY eQTL mapping experiment, 73% (54 eQTL) overlapped with eQTL from the initial M22/BY mapping experiment, which is significantly more than expected by chance (95% CI: 27%-40%,  $p < 0.001$ , permutation test, Figure S2C,D). This degree of overlap between the two M22/BY mapping experiments is also significantly greater than the degree of overlap between the second M22/BY experiment and the YPS1000/BY experiment (54%, 40 of 74 eQTL,  $p = 0.03$ , Fisher's exact test) but not the SK1/BY experiment (58%, 43 of 74 eQTL,  $p = 0.08$ , Fisher's exact test). Using a more stringent false discovery rate of 3% to identify eQTL reduced the number of eQTL called in the second M22/BY eQTL mapping experiment to 48, but there was still more overlap between the two M22/BY experiments than expected by chance (36 overlapping eQTL, 75%, 95% CI: 26%-42%,  $p < 0.001$ , permutation test)

(Table S10). Overlap between the two M22xBY crosses was similar to the overlap between the second M22xBY eQTL mapping experiment and either of the other two crosses, however, suggesting that reducing the false discovery rate enriched for eQTL shared among all strains (Table S10). eQTL identified between M22 and BY were also compared to eQTL identified previously between strains RM11 and BY. Consistent with the close phylogenetic relationship between M22 and RM11 (Figure S2A), the 113 eQTL identified in the initial mapping between M22 and BY strains included all 8 regions of the genome identified as affecting *TDH3* promoter activity in a prior cross between the BY and RM11 strains, 7 of which had eQTL alleles with effects in the same direction (Albert *et al.* 2014) (Figure S2E). 6 of these 7 eQTL with effects mapped in the same direction of the first M22 x BY cross and the previously published RM11 x BY cross were also identified with the more stringent FDR cutoff of 3% ( $|G'| > 10$ ) (Figure 2E). In the second M22 x BY mapping experiment, we observed eQTL between M22 and BY with the same direction of effect as eQTL identified between RM11 and BY for 4 of the 8 previously identified eQTL (Figure S2D). In all, more than 100 eQTL have been identified for these closely related crosses, with the majority of eQTL identified in two independent crosses. These results suggest that the differences in eQTLs mapped among strains are unlikely to be explained by low reproducibility of the mapping procedure, but rather reflect real differences in the genetic architecture of *trans*-regulatory variation affecting *TDH3* promoter activity among strains. Taken together, our data suggest that despite stabilizing selection limiting variation in *TDH3* expression, there are hundreds of genetic variants segregating within *S. cerevisiae* that impact *TDH3* promoter activity in *trans*. This number of variants is substantially more than suggested by prior work, indicating that *trans*-regulatory variation is more polygenic than typically appreciated. These variants (i) differ among strains, (ii) cause increases and decreases in *TDH3* promoter activity with similar frequencies, and (iii) are located in hundreds of distinct regions in the genome. In the wild, even more loci are expected to harbor variation affecting *TDH3* expression because the shared

reporter gene used in our experimental design eliminated variation in the *TDH3* promoter and was blind to variation affecting post-transcriptional regulation of *TDH3*. Our experimental design was also unable to determine interactions between *cis*- and *trans*-regulatory variation, epistatic interactions among regulatory variants, the effect size of individual eQTL, or the specific SNPs responsible for each eQTL's effects. Nonetheless, the prevalence and distribution of loci we identified are similar to recent mapping experiments in *S. cerevisiae* that also had high power to detect loci with opposing effects on quantitative traits (Jakobson and Jarosz 2019). They are also consistent with models of complex traits, such as the 'omnigenic model' (Boyle *et al.* 2017; Wray *et al.* 2018; Liu *et al.* 2019); our work specifically supports the idea that many casual variants have *trans*-regulatory effects on expression.

Although it might seem counterintuitive to find such extensive genetic variation affecting a trait whose variance appears to have been limited by stabilizing selection, theoretical work has previously shown that stabilizing selection acting on quantitative traits can maintain abundant cryptic genetic variation with off-setting effects (Lande 1976; Dover and Flavell 1984; Turelli 1984; Barton 1986, 1989). The pervasiveness of genetic variants with opposing effects on expression is consistent with the recurrent observation of compensatory evolution in genomic comparisons of gene expression within and between species (Goncalves *et al.* 2012; Schaefer *et al.* 2013; Coolon *et al.* 2014; Mack *et al.* 2016; Verta *et al.* 2016; Metzger *et al.* 2017). This variation may form the basis for developmental systems drift in which phenotypes stay stable over evolutionary time, but the molecular components responsible for the phenotype change (True and Haag 2001; Brachi *et al.* 2010; Gordon and Ruvinsky 2012; Pavlicev and Wagner 2012): with many combinations of alleles available that can produce the same trait value, changes in the regulation of a trait that do not alter the trait value might be common. Additional genetic mapping experiments that have similar power to resolve closely linked loci and detect alleles with very small effects are needed to determine whether the complex genetic

architecture observed for the *trans*-regulation of the *TDH3* gene in *S. cerevisiae* is common to other genes, traits, and organisms.

## Methods

### *Yeast strains and growth conditions*

Strains used in this work are listed in Table S1. To determine variability in *TDH3* expression segregating within *S. cerevisiae*, we used haploid, MAT $\alpha$  versions of 85 natural *S. cerevisiae* strains created in previous work (MacLean *et al.* 2017). For each strain, we inserted a P<sub>*TDH3*</sub>-YFP reporter at the *HO* locus using a standard lithium acetate transformation approach with minor alterations (Cubillos *et al.* 2009; MacLean *et al.* 2017). The inserted reporter contained a copy of the 678 bp *TDH3* promoter from the BY strain (entire intergenic region 5' of the *TDH3* coding sequence), a Yellow Fluorescent Protein (YFP) coding sequence, a CYC1 terminator, and a NatMX4 drug resistance marker. For 60 strains, we obtained successful integration and correct sequence of the reporter. Prior work indicates that the location of the reporter, or the use of a fusion of YFP to native TDH3, has consistent effects on expression and we thus expect that this experiment primarily is reporting on *trans*-regulatory variation among strains (Metzger *et al.* 2016). Unless noted, all yeast growth was performed at 30°C in YPD (1% Difco yeast extract, 2% peptone, 2% glucose).

### *Measurement of YFP expression*

The *trans*-regulatory effects on *TDH3* promoter activity for each strain were estimated by measuring YFP expression from the P<sub>*TDH3*</sub>-YFP reporter. Strains were first revived from glycerol stocks on YPG (1% Difco yeast extract, 2% peptone, 2% glycerol) at 30°C. After 24 hours, each strain was inoculated into liquid YPD in a 96 well plate. For each plate, YFP positive (PJW1139) and YFP negative (PJW880) strains were included at specific locations

in the 96 well plate as controls. This structure was replicated to solid YPD using a pin tool. To generate replicates, colonies were pin-tool replicated after 24 hours into twelve 96 well plates containing 500  $\mu$ l of liquid YPD and grown for 24 hours. Cultures were then diluted 1/20 into fresh 500  $\mu$ l of YPD and grown for an additional four hours. Samples were diluted 1/10 into 500  $\mu$ l PBS and analyzed on an Accuri C6 flow cytometer connected to an Intellicyt autosampler.

Data was processed using the same procedure as described in Duveau et al. (2018) (Duveau *et al.* 2018). Briefly, hard gates were used to remove flow cytometry artifacts and instances where multiple cells entered the flow cytometry detector at the same time based on estimates of cell size. For each sample, the most abundant monomorphic population was identified and the effect of cell size on fluorescence removed. For each event in a sample, the YFP levels were converted to estimates of mRNA expression using the formula  $E(\text{mRNA}) = \exp(-7.820027 * E(\text{YFP}))$ , which was based on a direct comparison of YFP fluorescence and mRNA abundance first reported in Duveau et al. (2018) (Duveau *et al.* 2018). From these estimates, the population median was calculated. Using the control strains, linear models were used to remove batch effects such as differences among plates and variation due to the position of a sample (row and column) within a plate. Twelve replicate samples from each strain were combined to estimate strain averages. Four strains (NCYC110 (P JW1041), EM93 (P JW1055), Yllc17\_E5 (P JW1038), and DBVPG3591 (P JW1053)) were excluded from analysis due to inconsistent measurements between replicates caused by flocculation and cell settling.

The effects of naturally occurring *cis*-regulatory variants on *TDH3* promoter activity within *S. cerevisiae* used data from Metzger et al. 2015 (Metzger *et al.* 2015) (Flow Repository FR-FCM-ZZBN). The effects of new mutations on *TDH3* promoter activity used data from Metzger et al. 2016 (Metzger *et al.* 2016) (Flow Repository FR-FCM-ZZNR). The original flow cytometry data from these previous studies was reprocessed with the same procedure as used in the current work.

### *Testing for evidence and impacts of selection*

We used the Brownian motion/Ornstein-Uhlenbeck framework to test for the presence of stabilizing selection on *trans*-acting factors affecting *TDH3* promoter activity. We followed the approach of Bedford and Hartl (2009). Briefly, two models of quantitative trait evolution were fit to the observed expression values and phylogenetic relationships among strains. The Brownian motion model allowed for trait values to diverge linearly with time, while the Ornstein-Uhlenbeck model included an additional parameter that reflects the action of stabilizing selection. We tested whether the Ornstein-Uhlenbeck model fit significantly better than the Brownian motion model using a chi-square distribution with a single degree of freedom.

In addition to this standard approach, we developed a method for identifying the action of natural selection on quantitative traits that uses empirically determined effects of new mutations to generate a neutral model of evolution for that specific trait. To inform this model, we used previously collected data on the effects on *TDH3* promoter activity due to new mutations. Briefly, this prior work used ethyl methanesulfonate (EMS) to induce mutations in an isogenic yeast population containing the  $P_{TDH3}$ -YFP reporter and used FACS to isolate ~1500 individual genotypes irrespective of their YFP expression. Each isolated mutant contained ~32 mutations relative to the reference strain, the overwhelming majority of which are expected to be *trans*-acting with respect to *TDH3* promoter activity (only two mutations in the *TDH3* promoter are expected among all



individuals) (Metzger *et al.* 2016). To estimate how *TDH3* promoter activity could evolve in the absence of natural selection, we generated a neutral distribution by sampling effects from this mutational distribution, combining the effects of mutations multiplicatively. We repeated this process 10,000 times to create a distribution of effects on *TDH3* promoter activity expected under neutrality for a given number of mutations.

To test for natural selection, we compared the effects of changes in *TDH3* promoter activity due to *trans*-regulatory differences among *S. cerevisiae* strains to our empirically derived neutral model. To account for the phylogenetic relationships among strains, we used a *S. cerevisiae* phylogeny estimated from whole genome polymorphism data to estimate how many mutations had likely occurred along each branch (MacLean *et al.* 2017). We estimated ancestral *TDH3 trans*-regulation on the *S. cerevisiae* phylogeny using ACE and estimated likelihoods by simulation (Paradis *et al.* 2004). This approach was necessary because the neutral model of *TDH3* promoter activity evolution is based on non-normally distributed data which precludes the assumption of time reversibility that allows for explicit calculation of the effects on expression on an unrooted phylogeny without identifying the direction of expression changes. We then compared the changes in *trans*-regulatory effects along each branch to the corresponding distribution of effects derived from our neutral model. For each branch, we calculated the likelihood of the observed change in expression along that branch given the number of mutations that had occurred. We combined the likelihoods over all observed branches to determine the likelihood of the complete set of observed expression values and changes in expression on the phylogeny. Observed likelihoods less than expected under neutrality are consistent with positive selection for a new phenotypic value, whereas likelihoods greater than expected under neutrality are consistent with phenotypic constraint due to natural selection. While there are biophysical limits on expression, the current modeling framework allows expression to increase or decrease without bounds. What these limits are and how they relate to the

current expression level is unknown. Practically, as long as naturally occurring expression levels are not at a biophysical limit, natural selection can still be detected with this test. Since both increases and decreases in expression were observed among strains, new mutations, and segregants during mapping, this issue is unlikely to have substantially altered our conclusion.

The inference of ancestral states used for this test assumes that the expression value for each ancestral node is the average of the descendent node values weighted by the branch lengths. To test the robustness of our inference to phylogenetic uncertainty and this assumption of ancestral state values, we repeated the analysis using genetic distance between strains instead of phylogenetic branch lengths to estimate the number of mutations that had occurred between strains. To avoid double counting of individual strains, we used each strain exactly once in the comparison. We then sampled which strains were compared 10,000 times to generate a distribution of observed effects.

#### *eQTL Mapping*

Genomic regions responsible for differences in *TDH3* promoter activity were identified by eQTL mapping. We crossed strains YPS1000 (PJW1057), SK1 (PJW1016), and M22 (PJW1072) that were MAT $\alpha$ , nourseothricin resistant, and contained the P<sub>*TDH3*</sub>-YFP reporter to a version of BY (PJW1240, Figure S4) that was MAT $\alpha$ , G418 resistant, and contained the P<sub>*TDH3*</sub>-YFP reporter. This common BY mapping strain also contained a Red Fluorescent Protein (RFP) marker at its mating type locus (Chin *et al.* 2012). Detailed methods for the creation of the common mapping strain are below. For each cross, we selected diploids using a combination of nourseothricin and G418 resistance and choose a single colony to ensure homogeneity in the genetic background. Hybrids for each cross were then sporulated (Figure S2B, P.0). To do so, hybrids were grown on GNA (1% Difco yeast extract, 3% Difco nutrient broth, 5% glucose) media for 12-16 hours, transferred to KAc

plates, and maintained at room temperature until at least 50% of the cells had sporulated.

Cells were then washed twice in 1 ml of water and incubated with 200 ul of 0.3mg/ml 100T zymolyase for one hour with agitation. Next, cells were washed with 1 ml of water and resuspended in 100 ul of water. Cells were vortexed for 2 minutes to stick spores to the tube wall. The supernatant was removed and 1 ml of water was added. Without agitation, this 1 ml was removed and a second 1 ml of H<sub>2</sub>O was added. This 1 ml was also removed and 1ml of triton-X (0.02%) was added. Samples were sonicated on ice for 10 seconds at medium power (3.5 on a Sonic Dismembrator Model 100, Fisher). Spores were confirmed to be separated and diploids absent by visual inspection under a microscope.

After spore isolation, the population was split into thirds. One third was added to 1 ml YPD, grown to saturation overnight, and then frozen at -80°C as a glycerol stock. The second third was sorted for the absence of the RFP marker using fluorescence assisted cell sorting (FACS) on a FACS canto II at the University of Michigan Flow Cytometry Core (Figure S2B, P.1). Because all MAT $\alpha$  and diploid strains express RFP, this sorting captures only MAT $\alpha$  cells. For each cross, we collected > 10<sup>6</sup> individuals lacking RFP fluorescence (Figure S2B, F.1). These were incubated with 1ml YPD and grown for 24-28 hours. The final third was used to initiate additional rounds of crossing by plating onto YPD. After growth overnight, sporulation and spore isolation was repeated. Isolated spores from the second round of sporulation were plated onto YPD and sporulated a final time. After this third round of sporulation, cells lacking RFP fluorescence were again collected (Figure S2B, F.3).

To identify the genetic basis of differences in YFP expression between strains, cells were transferred to 1 ml PBS and cells with the 5% highest (Figure S2B, H.1.1 and H.3.1) or the 5% lowest (Figure S2B, L.1.1 and L.3.1) YFP expression corrected for cell size were sorted from cells within the middle 80% of the cell size distribution based on forward scatter (FSC-

H). For each sample, 100,000 individuals were collected from each tail. Each sorted population was grown in liquid YPD for 20 hours after which one half was frozen. To further enrich genotypes with high and low YFP expression within the sorted populations, the second half of each sample was used to initiate two additional rounds of sorting. For each round, the same sorting procedure as above was followed with one exception; populations originally sorted for high YFP expression were only sorted for high YFP expression and populations sorted for low YFP expression were only sorted for low YFP expression.

After all selection steps were completed, samples were revived from glycerol stocks and grown in 1 ml of YPD for 2 hours. DNA was extracted from each sample using the Puregene Yeast Kit from Qiagen. DNA concentration was determined using a Qubit, and Illumina Nextera XT libraries were prepared following the manufacturers guidelines. Barcodes for each sample are listed in Table S2. Library quality was assessed using the bioanalyzer and all samples were pooled equally using concentration estimates from the Qubit. Sequencing was performed on a HiSeq 2000 using 125 bp paired end sequencing at the University of Michigan Sequencing Core. Sequencing barcodes are listed in Table S2.

#### *QTL Identification*

After sequencing, samples were processed to identify individual eQTL. First, Sickle was used to remove low quality bases from each read using default setting (Joshi, N.A., Fass 2011). Next, Cutadapt was used to remove any adapter sequence from read ends (flags -e 0.2 -O 3 -m 15) (Martin 2011). Samples were aligned to the S228c reference genome using bowtie2 (flags -I 0 -X 1000 --very-sensitive-local) (Langmead and Salzberg 2012) and then sorted and indexed using samtools (Li *et al.* 2009). Overlapping reads were clipped using clipOverlap in bamUtil. SNPs were jointly called within each paired set of samples selected for high and low YFP fluorescence using freebayes (Garrison and

Marth 2012). Identified SNPs were required to reach at least 20% frequency in at least one of the two paired samples and be observed at least 4 times across both samples.

For each pair of samples, SNPs were filtered based on quality and depth. Each SNP was required to have depth of at least 20 to ensure adequate power, a depth below 500 to reduce the number of SNPs called in low complexity sequences, a mapping quality score of greater than 30, and imbalance scores for left/right, center/end, and forward/reverse for SNP position within reads of less than 30. At each position, only the two highest likelihood alleles were retained, with any other alleles observed assumed to result from sequencing errors. For each SNP, we calculated a G statistic using a likelihood ratio test of alternative and reference alleles within the high and low selected populations (Magwene *et al.* 2011). For SNPs where the alternative allele had a higher frequency than the reference allele in the high selected population relative to the low selected population, we maintained the sign of G. For SNPs where the alternative allele had a lower frequency than the reference allele in the high selected population relative to the low selected population, we flipped the sign of G. We then calculated  $G'$  by averaging these estimates over a 40 kb window centered on each individual SNP using a tri-cube kernel function (Magwene *et al.* 2011). To identify QTL peaks, we located all local maxima and minima in  $G'$ . We called significant peaks those with  $G' > 5$  or  $G' < -5$ . The 95% confidence interval on the location of each peak was defined as the distance needed for  $G'$  to drop by 5 from the peak. Local peaks whose confidence intervals overlapped in location were merged into a single peak. The  $G'$  cutoff value was chosen to be conservative based on changes in SNP frequency from prior work (Magwene *et al.* 2011). We also compared allele frequency, G, and  $G'$  for SNPs before and after sorting cells randomly with respect to YFP fluorescence for the M22 x BY cross. This experiment identified 11 eQTL, giving a false discovery rate of approximately 10% for the mapping experiments reported in the main text. To determine how the specific  $G'$  cutoff affects our

analyses, we additionally conducted all analyses using a stricter  $G'$  cutoff of 10, which corresponds to an FDR of 3%. Statistics for both cutoffs are included in tables S5-S9.

### *Creation of mapping strain*

Determining the genetic and molecular mechanisms underlying complex phenotypes often requires identifying the causative genetic loci and nucleotides contributing to these traits (Rausher and Delph 2015). However, in the yeast *Saccharomyces cerevisiae*, the primary laboratory strains, S288c and its descendants, have several phenotypes that limit their usefulness in high throughput mapping approaches.

For example, *S. cerevisiae* isolates from the wild readily undergo meiosis under nutrient starvation and the majority of individual diploids sporulate. By contrast, S288c enters meiosis slowly and only a small proportion of individuals successfully complete meiosis, even under ideal conditions (Deutschbauer and Davis 2005; Gerke *et al.* 2006). Because genetic mapping requires recombination, and thus, meiosis, the limited meiotic abilities of S288c reduces the number and speed at which mapping populations can be created.

In addition to poor sporulation, S288c and its descendants generate petite cells lacking mitochondria with high frequency. As a consequence, these individuals cannot perform aerobic respiration and often have altered phenotypes compared to wild-type individuals (Chen and Clark-Walker 1999). Because linking phenotypes to their genomic location requires high quality phenotyping, additional variation introduced by petite individuals can reduce the accuracy and power of genetic mapping.

Lastly, upon meiosis yeast generate both  $a$  and  $\alpha$  haploids. These haploids will readily reform diploids if not prevented, thus introducing additional variation due to ploidy into a mapping population. Current techniques for limiting the recreation of diploids suffer from a lack of

throughput and poor specificity (Tong *et al.* 2001). As a consequence, the power to map the genetic basis of recessive traits in yeast is reduced.

To overcome these deficiencies, we modified S288c to increase its sporulation rate and density, reduce the frequency at which it generated petites, and to express a fluorescent marker that allowed easy identification of mating type. To accomplish these goals, we obtained several strains derived from S288c. These strains vary in their mating type and auxotrophies, facilitating crossing. In addition, these strains differ at a set of alleles derived from natural *S. cerevisiae* strains that either improve sporulation rate or lower petite frequency. These include versions of *TAO3* and *RME1* that increase sporulation rate (Deutschbauer and Davis 2005) and versions of *SAL1*, *CAT5*, and *MIP1* that decrease petite frequency (Dimitrov *et al.* 2009). An allelic variant at *MKT1* has also been identified that affects both sporulation and petite frequency. However, while the wild-type S288c *MKT1* allele decreases sporulation rate, it also substantially reduces petite frequency compared to the alternative allele and we kept the S288c version (Deutschbauer and Davis 2005; Dimitrov *et al.* 2009).

Through a series of crosses, transformations, and sporulations, we isolated a single individual that contained the desired set of alleles and was free of all auxotrophies except for *ura3Δ0* (Figure S4). We retained the *URA3* auxotrophy to facilitate future genetic manipulation by the *delitto perfetto* method, which requires *5-FOA* counter-selection and therefore a starting strain that is *ura-* (Storici and Resnick 2006). To facilitate the creation of the correct strain, we tracked the allelic identity of each segregating locus using pyrosequencing (Table S3, Table S4). After identification, the isolated individual was turned into a diploid and sporulated to generate isogenic **a** and **α** haploids. To the **a** haploid, we introduced a red fluorescent protein into the *MAT* locus (Chin *et al.* 2012). This marker allows identification of individuals based on their mating type using fluorescence assisted cell sorting

(FACS). Using this marker, populations containing millions of individuals of the same mating type can be collected in minutes. Finally, both  $\mathbf{a}$  and  $\alpha$  strains contain a *TDH3* promoter driving Yellow Fluorescent Protein expression located at the *HO* gene to facilitate mapping of mutations and polymorphisms influencing *TDH3* expression.

Growth was performed using YPD. For crosses involving auxotrophies, synthetic complete media was used, minus the appropriate amino acids (1.7g Yeast nitrogen base, 5 g Ammonium Sulfate, 20g glucose per 1 L water; 20g agar for solid plates). Sporulation was induced by growth on YPD plates for 24 hours at room temperature, followed by plating on KAc plates at room temperature (10g Potassium acetate, 0.5g glucose per 1 L water; 20g agar for solid plates). Ascus walls were dissolved prior to tetrad dissections by incubating spores in 200  $\mu$ l zymolyase (1 mg/ml 20T) for 1 hour without shaking.

To create homozygous diploids from haploids, strains were transformed with plasmid pCM66. pCM66 contains a galactose inducible copy of *HO* and a selective nourseothricin resistance marker. After transformation, nourseothricin resistant cells were grown with galactose as the sole carbon source at 30°C without shaking for 8 hours to induce expression of *HO*. This allowed for mating type switching and subsequent mother-daughter cell mating to produce diploids. Cells were streaked for single colonies on YPD plates and the ploidy of single colonies checked by colony PCR using mating-type specific primers. Diploid colonies were streaked onto fresh, non-selective, YPD plates and assayed for loss of nourseothricin resistance, and thus pCM66.

Pyrosequencing was used to follow sporulation and petite QTN. Methods are as described in (Wittkopp 2012). PCR primers used are Table S3. Dispensation order for pyrosequencing is in Table S4.



## **Acknowledgements**

We would like to thank Fabien Duveau, Andrea Hodgins-Davis, Mark Hill, Mo Siddiq, Petra Vande Zande, and members of the Wittkopp lab for helpful comments on the manuscript. Support and access to flow cytometry equipment was provided by the University of Michigan Center for Chemical Genomics and Flow Cytometry Core. This work was supported by the University of Michigan Rackham Graduate School and National Institutes of Health Genome Sciences training grant (T32 HG000040) to B.P.H.M, and the National Science Foundation (MCB-1021398) and National Institutes of Health (1R35GM118073 and 1 R01 GM108826) to P.J.W. The authors declare no conflicts of interest.

## **Author Contributions**

B.P.H.M. and P.J.W. conceived and designed the study. B.P.H.M. collected and analyzed the data.

B.P.H.M. and P.J.W. jointly drafted the paper.

## **Data Accessibility**

Sequencing data is available from NCBI's dbSNP (<https://www.ncbi.nlm.nih.gov/sra>; PRJNA527754 and PRJNA527772). New flow cytometry data reported in this paper is available for download from Flow Repository (FR-FCM-ZYVQ). Previously published flow cytometry data used for analyses in this paper are also available in the Flow Repository (FR-FCM-ZZBN and FR-FCM-ZZNR). Processed data and R code used for data processing, creating the figures, and performing statistical tests are included in the .zip file available for download from Dryad Digital Repository ([doi:10.5061/dryad.38c01km](https://doi.org/10.5061/dryad.38c01km)) or the University of Michigan's Deep Blue Data Repository (<https://deepblue.lib.umich.edu/data>).

## References

- Albert, F. W., Bloom, J. S., Siegel, J., Day, L., & Kruglyak, L. (2018). Genetics of *trans*-regulatory variation in gene expression. *Elife* 7:1–39.
- Albert, F. W., & Kruglyak, L. (2015). The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.* 16:197–212.
- Albert, F. W., Treusch, S., Shockley, A. H., Bloom, J. S., & Kruglyak, L. (2014). Genetics of single-cell protein abundance variation in large yeast populations. *Nature* 506:494–497.
- Barton, N. (1989). The divergence of a polygenic system subject to stabilizing selection, mutation and drift. *Genet. Res.* 54:59–78.
- Barton, N. H. (1986). The maintenance of polygenic variation through a balance between mutation and stabilizing selection. *Genet. Res.* 47:209–216.
- Bedford, T., & Hartl, D. L. (2009). Optimization of gene expression by natural selection. *Proc. Natl. Acad. Sci.* 106:1133–1138.
- Bell, G. D. M., Kane, N. C., Rieseberg, L. H., & Adams, K. L. (2013). RNA-seq analysis of allele-specific expression, hybrid effects, and regulatory divergence in hybrids compared with their parents from natural populations. *Genome Biol. Evol.* 5:1309–1323.
- Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* 169:1177–1186.
- Brachi, B., Faure, N., Horton, M., Flahauw, E., Vazquez, A., Nordborg, M., *et al.* (2010). Linkage and

Association Mapping of *Arabidopsis thaliana* Flowering Time in Nature. *PLoS Genet.*

6:e1000940.

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

Brem, R. B., Storey, J. D., Whittle, J., & Kruglyak, L. (2005). Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature* 436:701–703.

Brem, R. B., Yvert, G., Clinton, R., & Kruglyak, L. (2002). Genetic dissection of transcriptional regulation in budding yeast. *Science* 296:752–755.

Chen, J., Nolte, V., & Schlötterer, C. (2015). Temperature Stress Mediates Decanalization and Dominance of Gene Expression in *Drosophila melanogaster*. *PLoS Genet.* 11:e1004883.

Chen, X. J., & Clark-Walker, G. D. (1999). The petite mutation in yeasts: 50 years on. *Int. Rev. Cytol.* 194:197–238.

Chin, B. L., Frizzell, M. a., Timberlake, W. E., & Fink, G. R. (2012). FASTER MT: Isolation of Pure Populations of  $\alpha$  and  $\beta$  Ascospores from *Saccharomyces cerevisiae*. *G3* 2:449–452.

Coolon, J., McManus, C. J., Stevenson, K. R., Graveley, B. R., & Wittkopp, P. J. (2014). Tempo and mode of regulatory evolution in *Drosophila*. *Genome Res.* 24:797–808.

Cubillos, F. A., Coustham, V., & Loudet, O. (2012). Lessons from eQTL mapping studies: Non-coding regions and their role behind natural phenotypic variation in plants. *Curr. Opin. Plant Biol.* 15:192–198.

Cubillos, F. A., Louis, E. J., & Liti, G. (2009). Generation of a large set of genetically tractable haploid and diploid *Saccharomyces* strains. *FEMS Yeast Res.* 9:1217–1225.

- Cubillos, F. A., Parts, L., Salinas, F., Bergström, A., Scovacricchi, E., Zia, A., *et al.* (2013). High-resolution mapping of complex traits with a four-parent advanced intercross yeast population. *Genetics* 195:1141–1155.
- de Meaux, J. (2005). Allele-Specific Assay Reveals Functional Variation in the Chalcone Synthase Promoter of *Arabidopsis thaliana* That Is Compatible with Neutral Evolution. *Plant Cell* 17:676–690.
- Denver, D. R., Morris, K., Streelman, J. T., Kim, S. K., Lynch, M., & Thomas, W. K. (2005). The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. *Nat. Genet.* 37:544–548.
- Deutschbauer, A. M., & Davis, R. W. (2005). Quantitative trait loci mapped to single-nucleotide resolution in yeast. *Nat. Genet.* 37:1333–1340.
- Dimitrov, L. N., Brem, R. B., Kruglyak, L., & Gottschling, D. E. (2009). Polymorphisms in multiple genes contribute to the spontaneous mitochondrial genome instability of *Saccharomyces cerevisiae* S288C strains. *Genetics* 183:365–383.
- Dover, G. A., & Flavell, R. B. (1984). Molecular coevolution: DNA divergence and the maintenance of function. *Cell* 38:622–623.
- Duveau, F., Hodgins-Davis, A., Metzger, B. P. H., Yang, B., Tryban, S., Walker, E. A., *et al.* (2018). Fitness effects of altering gene expression noise in *Saccharomyces cerevisiae*. *Elife* 7:1–33.
- Duveau, F., Metzger, B. P. H., Gruber, J. D., Mack, K., Sood, N., Brooks, T. E., *et al.* (2014). Mapping Small Effect Mutations in *Saccharomyces cerevisiae*: Impacts of Experimental Design and Mutational Properties. *G3* 4:1205–1216.

- Duveau, F., Toubiana, W., & Wittkopp, P. J. (2017). Fitness Effects of *Cis*-Regulatory Variants in the *Saccharomyces cerevisiae* TDH3 Promoter. *Mol. Biol. Evol.* 8:206–216.
- Ehrenreich, J. M., Bloom, J., Torabi, N., Wang, X., Jia, Y., & Kruglyak, L. (2012). Genetic Architecture of Highly Complex Chemical Resistance Traits across Four Yeast Strains. *PLoS Genet.* 8:e1002570.
- Ehrenreich, J. M., Torabi, N., Jia, Y., Kent, J., Martis, S., Shapiro, J. a., *et al.* (2010). Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature* 464:1039–1042.
- Emerson, J. J., Hsieh, L.-C., Sung, H.-M., Wang, T.-Y., Huang, C.-J., Lu, H. H.-S., *et al.* (2010). Natural selection on *cis* and *trans* regulation in yeasts. *Genome Res.* 20:826–836.
- Fay, J. C., & Wittkopp, P. J. (2008). Evaluating the role of natural selection in the evolution of gene regulation. *Heredity (Edinb).* 100:191–199.
- Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *ArXiv*, doi: 10.1126/science.1241459.
- Gerke, J. P., Chen, C. T. L., & Cohen, B. a. (2006). Natural isolates of *Saccharomyces cerevisiae* display complex genetic variation in sporulation efficiency. *Genetics* 174:985–997.
- Gilad, Y., Oshlack, A., & Rifkin, S. A. (2006). Natural selection on gene expression. *Trends Genet.* 22:456–461.
- Gilad, Y., Rifkin, S. a., & Pritchard, J. K. (2008). Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet.* 24:408–415.
- Goncalves, A., Leigh-Brown, S., Thybert, D., Stefflova, K., Turro, E., Flicek, P., *et al.* (2012). Extensive compensatory *cis-trans* regulation in the evolution of mouse gene expression. *Genome Res.*

22:2376–2384.

Gordon, K. L., & Ruvinsky, I. (2012). Tempo and Mode in Evolution of Transcriptional Regulation.

*PLoS Genet.* 8:e1002432.

Gruber, J. D., & Long, A. D. (2009). *Cis*-regulatory variation is typically polyallelic in *Drosophila*.

*Genetics* 181:661–670.

Gruber, J. D., Vogel, K., Kalay, G., & Wittkopp, P. J. (2012). Contrasting Properties of Gene-specific

Regulatory, Coding, and Copy Number Mutations in *Saccharomyces cerevisiae*: Frequency,

Effects and Dominance. *PLoS Genet.* 8:e1002497.

Hansen, B. G., Halkier, B. a., & Kliebenstein, D. J. (2008). Identifying the molecular basis of QTLs:

eQTLs add a new dimension. *Trends Plant Sci.* 13:72–77.

Jakobson, C. M., & Jarosz, D. F. (2019). Molecular Origins of Complex Heritability in Natural

Genotype-to-Phenotype Relationships. *Cell Syst.* 8:363–379.

Josephs, E. B., Lee, Y. W., Stinchcombe, J. R., & Wright, S. I. (2015). Association mapping reveals the

role of purifying selection in the maintenance of genomic variation in gene expression. *Proc.*

*Natl. Acad. Sci.* 112:15390–15395.

Joshi, N.A., Fass, J. N. (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for

FastQ files.

Kang, E. Y., Martin, L., Mangul, S., Isvilanonda, W., Zou, J., Ben-David, E., *et al.* (2016). Discovering

SNPs Regulating Human Gene Expression Using Allele Specific Expression from RNA-Seq Data.

*Genetics* 204:1057–1064.

Kita, R., Venkataram, S., Zhou, Y., & Fraser, H. B. (2017). High-resolution mapping of *cis*-regulatory

- variation in budding yeast. *Proc. Natl. Acad. Sci.* 114:E10736–E10744.
- Kofler, R., Pandey, R. V., & Schlötterer, C. (2011). PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* 27:3435–3436.
- Lande, R. (1976). The maintenance of genetic variability by mutation in a polygenic character with linked loci. *Genet. Res.* 89:373–387.
- Landry, C. R., Lemos, B., Rifkin, S. A., Dickinson, W. J., & Hartl, D. L. (2007). Genetic properties influencing the evolvability of gene expression. *Science* 317:118–121.
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2.
- Lemos, B., Araripe, L. O., Fontanillas, P., & Hartl, D. L. (2008). Dominance and the evolutionary accumulation of cis- and trans-effects on gene expression. *Proc. Natl. Acad. Sci.* 105:14471–14476.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.* (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.
- Liu, X., Li, Y. I., & Pritchard, J. K. (2019). *Trans* Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell* 177:1022–1034.
- Mack, K. L., Campbell, P., & Nachman, M. W. (2016). Gene regulation and speciation in house mice. *Genome Res.* 26:451–461.
- MacLean, C. J., Metzger, B. P. H., Yang, J.-R. R., Ho, W.-C. C., Moyers, B., & Zhang, J. (2017). Deciphering the Genic Basis of Yeast Fitness Variation by Simultaneous Forward and Reverse Genetics. *Mol. Biol. Evol.* 34:2486–2502.

- Magwene, P. M., Willis, J. H., & Kelly, J. K. (2011). The statistics of bulk segregant analysis using next generation sequencing. *PLoS Comput. Biol.* 7:e1002255.
- Majewski, J., & Pastinen, T. (2011). The study of eQTL variations by RNA-seq : from SNPs to phenotypes. *Trends Genet.* 27:72–79.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17:10–12.
- 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. *Mol. Biol. Evol.* 33:1131–1146.
- Metzger, B. P. H., Wittkopp, P. J., & Coolon, J. D. (2017). Evolutionary dynamics of regulatory changes underlying gene expression divergence among *Saccharomyces* species. *Genome Biol. Evol.* 177:1987–1996.
- Metzger, B. P. H., Yuan, D. C., Gruber, J. D., Duveau, F. D., & Wittkopp, P. J. (2015). Selection on noise constrains variation in a eukaryotic promoter. *Nature* 521:344–347.
- Moyerbrailean, G. A., Richards, A. L., Kurtz, D., Kalita, C. A., Davis, G. O., Harvey, C. T., *et al.* (2016). High-throughput allele-specific expression across 250 environmental conditions. *Genome Res.* 26:1627–1638.
- Nica, A. C., & Dermitzakis, E. T. (2013). Expression quantitative trait loci: present and future. *Philos. Trans. R. Soc. B Biol. Sci.* 368:20120362.
- Pai, A. a., Pritchard, J. K., & Gilad, Y. (2015). The Genetic and Mechanistic Basis for Variation in Gene Regulation. *PLoS Genet.* 11:e1004857.



- Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.
- Parts, L., Cubillos, F. A., Warringer, J., Jain, K., Salinas, F., Bumpstead, S. J., *et al.* (2011). Revealing the genetic structure of a trait by sequencing a population under selection. *Genome Res.* 21:1131–1138.
- Parts, L., Liu, Y.-C., Tekkedil, M. M., Steinmetz, L. M., Caudy, A. a, Fraser, A. G., *et al.* (2014). Heritability and genetic basis of protein level variation in an outbred population. *Genome Res.* 24:1363–1370.
- Pavlicev, M., & Wagner, G. P. (2012). A model of developmental evolution: selection, pleiotropy and compensation. *Trends Ecol. Evol.* 1–7.
- Peter, J., Chiara, M. De, Friedrich, A., Yue, J., Pflieger, D., Bergström, A., *et al.* (2018). Genome evolution across 1011 *Saccharomyces cerevisiae* isolates. *Nature* 556:339–344.
- Rausher, M. D., & Delph, L. F. (2015). Commentary: When does understanding phenotypic evolution require identification of the underlying genes? *Evolution* 69:1655–1664.
- Ronald, J., Brem, R. B., Whittle, J., & Kruglyak, L. (2005). Local regulatory variation in *Saccharomyces cerevisiae*. *PLoS Genet.* 1:0213–0222.
- Salinas, F., de Boer, C. G., Abarca, V., García, V., Cuevas, M., Araos, S., *et al.* (2016). Natural variation in non-coding regions underlying phenotypic diversity in budding yeast. *Sci. Rep.* 6:21849.
- Schadt, E. E., Monks, S., Drake, T., Lusk, A. J., Che, N., Colino, V., *et al.* (2003). Genetics of gene expression surveyed in maize, mouse and man. *Nature* 205:1–6.
- Schaefer, B., Emerson, J. J., Wang, T.-Y., Lu, M.-Y. J., Hsieh, L.-C., & Li, W.-H. (2013). Inheritance of

Gene Expression Level and Selective Constraints on Trans- and Cis- Regulatory Changes in Yeast. *Mol. Biol. Evol.* 30:2121–2133.

Schlötterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014). Sequencing pools of individuals — mining genome-wide polymorphism data without big funding. *Nat. Rev. Genet.* 15:749–763.

Smith, E. N., & Kruglyak, L. (2008). Gene-environment interaction in yeast gene expression. *PLoS Biol.* 6:e83.

Stern, D. L. (2000). Perspective: Evolutionary Developmental Biology and the Problem of Variation. *Evolution* 54:1079–1091.

Storici, F., & Resnick, M. a M. (2006). The delitto perfetto approach to in vivo site-directed mutagenesis and chromosome rearrangements with synthetic oligonucleotides in yeast. *Methods Enzymol.* 409:329–45.

Sung, H.-M., Wang, T.-Y., Wang, D., Huang, Y.-S., Wu, J.-P., Tsai, H.-K., *et al.* (2009). Roles of *Trans* and *Cis* Variation in Yeast Intraspecies Evolution of Gene Expression. *Mol. Biol. Evol.* 26:2533–2538.

Suvorov, A., Nolte, V., Pandey, R. V., Franssen, S. U., Futschik, A., & Schlötterer, C. (2013). Intra-specific regulatory variation in *Drosophila pseudoobscura*. *PLoS One* 8:e83547.

Tong, a H., Evangelista, M., Parsons, A. B., Xu, H., Bader, G. D., Pagé, N., *et al.* (2001). Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* 294:2364–2368.

True, J. R., & Haag, E. S. (2001). Developmental system drift and flexibility in evolutionary trajectories. *Evol. Dev.* 3:109–119.

Turelli, M. (1984). Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the

abdominal bristle. *Theor. Popul. Biol.* 25:138–193.

Verta, J. P., Landry, C. R., & Mackay, J. (2016). Dissection of expression-quantitative trait locus and allele specificity using a haploid/diploid plant system - insights into compensatory evolution of transcriptional regulation within populations. *New Phytol.* 211:159–171.

Wang, D., Sung, H.-M., Wang, T.-Y., Huang, C.-J., Yang, P., Chang, T., *et al.* (2007). Expression evolution in yeast genes of single-input modules is mainly due to changes in trans-acting factors. *Genome Res.* 17:1161–1169.

Westra, H. J., & Franke, L. (2014). From genome to function by studying eQTLs. *Biochim. Biophys. Acta - Mol. Basis Dis.* 1842:1896–1902.

Wittkopp, P. J. (2012). Using Pyrosequencing to Measure Allele-Specific mRNA Abundance and Infer the Effects of *Cis*- and *Trans*-regulatory Differences. P. *in V.* Orgogozo and M. V. Rockman, eds. *Molecular Methods for Evolutionary Genetics.*

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

Wray, N. R., Wijmenga, C., Sullivan, P. F., Yang, J., & Visscher, P. M. (2018). Common Disease Is More Complex Than Implied by the Core Gene Omnigenic Model. *Cell* 173:1573–1580.

Yvert, G., Brem, R. B., Whittle, J., Akey, J. M., Foss, E., Smith, E. N., *et al.* (2003). *Trans*-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat. Genet.* 35:57–64.

Zhang, X., & Borevitz, J. O. (2009). Global analysis of allele-specific expression in *Arabidopsis thaliana*. *Genetics* 182:943–954.

Zhu, Y. O., Siegal, M. L., Hall, D. W., & Petrov, D. a. (2014). Precise estimates of mutation rate and spectrum in yeast. *Proc. Natl. Acad. Sci.* 111:E2310–E2318.

## Figure Legends

Figure 1. Extensive *trans*-regulatory variation affecting *TDH3* expression is segregating among *S. cerevisiae* strains. A. Variation in *TDH3* *trans*-regulatory backgrounds among yeast strains was measured using a reporter gene containing the *TDH3* promoter from the BY strain and a yellow fluorescent protein (YFP). This reporter was integrated into the genome of 56 diverse *S. cerevisiae* strains. Twelve replicate populations were grown in YPD and analyzed by flow cytometry for YFP expression. B. Variation among replicates relative to the BY reference strain was used to calculate the average effect of each strain's *trans*-regulatory background on *TDH3* promoter activity. Darker colors reflect higher *TDH3* reporter activity. Strain names in blue are used in subsequent mapping experiments. C. Frequency of *trans*-regulatory effects relative to reference strain. D. Phylogenetic relationships among strains as estimated from genome wide polymorphism data (MacLean *et al.* 2017). Color of branches corresponds to estimated *trans*-regulatory effect from ancestral character estimation.

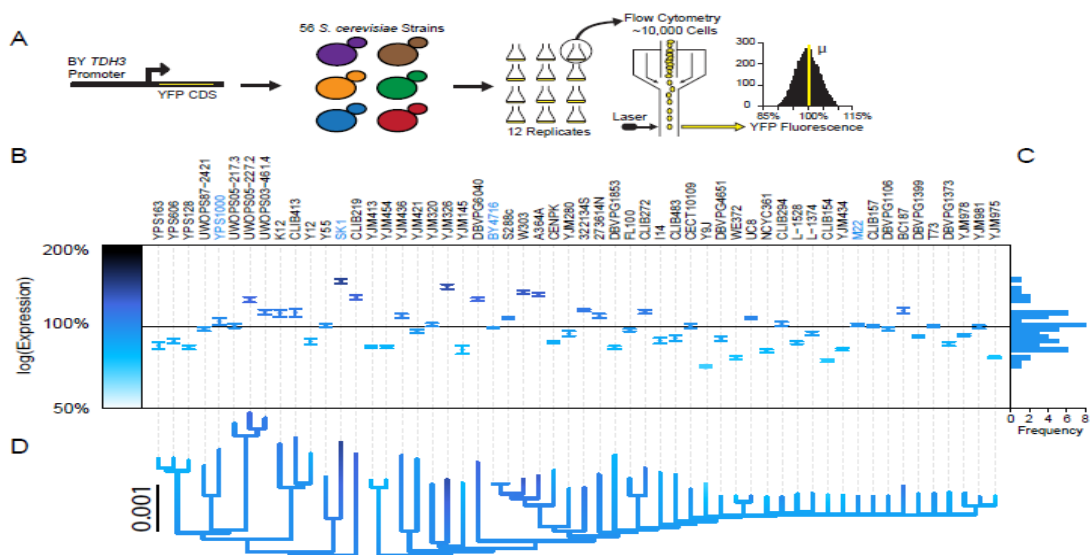
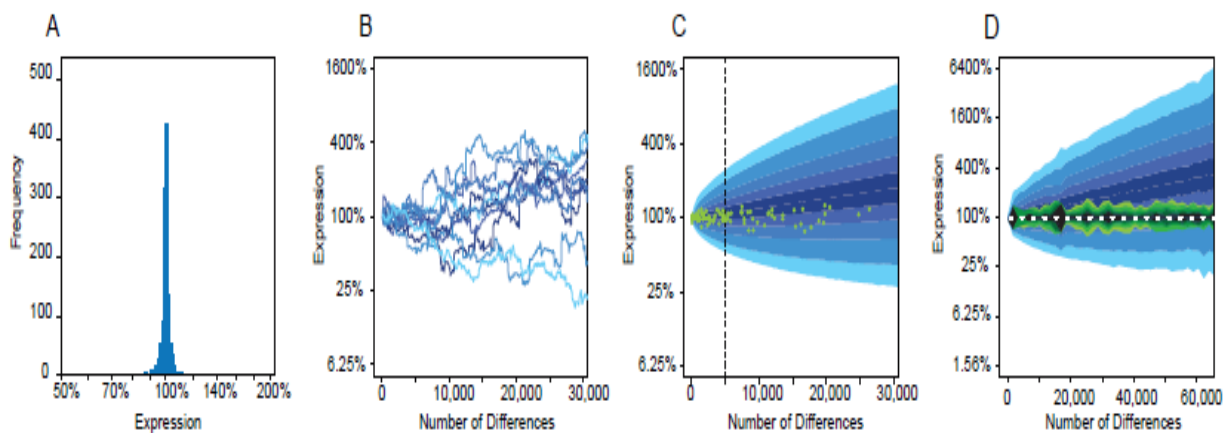


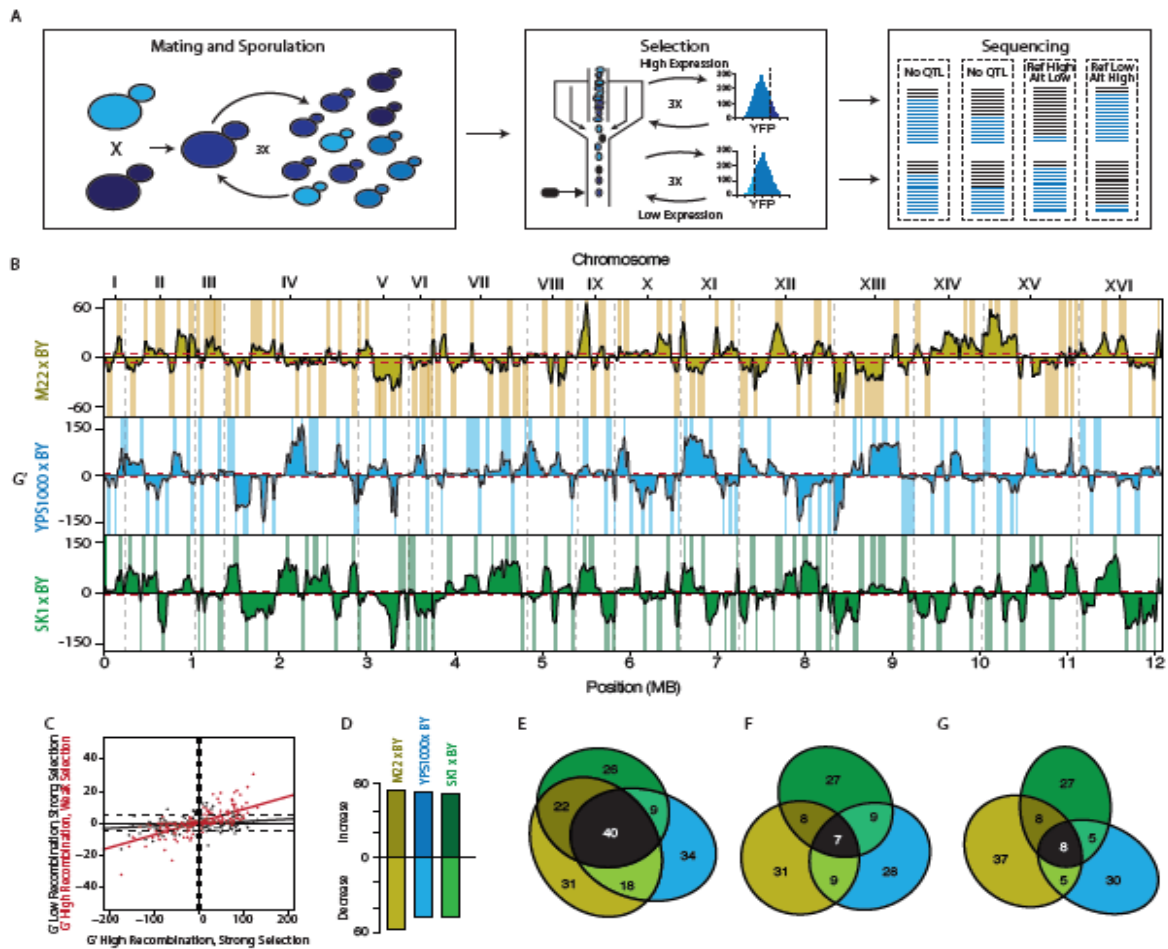
Figure 2. Natural selection has constrained *TDH3* trans-regulatory variation. A. Effects of trans-regulatory mutations on *TDH3* promoter activity. Mutants were collected and analyzed in prior work (Metzger *et al.* 2016). B. Simulated neutral trajectories for *TDH3* promoter activity based on empirically measured effects of new mutations. Lighter colors reflect more extreme values after 30,000 mutations. C. Comparison of observed differences in *TDH3* promoter activity among *S. cerevisiae* strains with neutral expectation. The blue background represents the 95th, 90th, 80th, 70th, and 60<sup>th</sup> percentiles, from light to dark, for the simulated neutral trajectories. Green dots are differences in *TDH3* promoter activity and estimated number of mutations based on the *S. cerevisiae* phylogeny. Dashed line indicates the point where the observed data departs significantly from expectation. D. Same as C, but using genetic distance instead of phylogenetic distance between strains. The green areas represent the 95th, 90th, 80th, 70th, and 60<sup>th</sup> percentiles, from light to dark, for the observed differences from sampling pairs of strains.



A

Figure 3. Compensatory alleles underlie the maintenance of *TDH3* trans-regulatory effects. A. The genomic basis of *TDH3* trans-regulatory variation was mapped using an xQTL approach. Left: Three rounds of mating and sporulation were used to increase mapping resolution. Middle: Three rounds of FACS based selection were used to enrich for alleles increasing and decreasing *TDH3* trans-regulatory activity. In each round, the top or bottom 5% of the population was collected. Right: Comparisons of allele frequency from Illumina sequencing of FACS based pools was used to identify eQTL. Each block (dashed lines) represents a different genomic region. Colored lines represent allele frequencies. Black: Reference strain. Blue: Testing strain. For each block, the top bars are after selection for high YFP fluorescence, while the bottom bars are after selection for low YFP fluorescence. eQTL are identified when allele frequencies among the high and low selected pools differ significantly. B.  $G'$  statistic for evidence of eQTL in each comparison. Effects are relative to the non-BY reference allele. Dashed gray lines indicate chromosome boundaries. Dashed red lines gives threshold for statistical significance. Called eQTL with 95% confidence intervals on the location are highlighted for each strain. Brown: M22 x BY. Blue: YPS1000 x BY. Green: SK1 x BY. C. Relationship between  $G'$  statistic for different mapping procedures. X-axis,  $G'$  statistic for high recombination and strong selection (three rounds of crossing and three rounds of selection). Y-axis, (Black)  $G'$  statistic for low recombination and strong selection (one round of crossing and three rounds of selection). (Red)  $G'$  statistic for high recombination and weak selection (three rounds of crossing and one round of selection). Each point is for an eQTL identified with high recombination and strong selection (three rounds of crossing and three rounds of selection) from the M22 x BY cross. Solid lines show fits from a linear model for high recombination and low selection (red) or low recombination and high selection (black). D. Number of non-BY eQTL increasing or decreasing *TDH3* promoter activity for each cross. E. eQTL shared among the three crosses irrespective of

direction of effect. Areas are proportional to the number of eQTL shared. Brown: eQTL identified only in the M22 x BY cross. Blue: eQTL identified only in the YPS1000 x BY cross. Green: eQTL identified only in the SK1 x BY cross. Black: eQTL identified in all three crosses. F. Same as E, but for non-BY eQTL that increase *TDH3* promoter activity. G. Same as E, but for non-BY eQTL that decrease *TDH3* promoter activity.



Aut