

Review

Contributions of mutation and selection to regulatory variation: lessons from the *S. cerevisiae* *TDH3* gene

Patricia J. Wittkopp

Department of Ecology and Evolutionary Biology
Department of Molecular, Cellular, and Developmental Biology
University of Michigan, Ann Arbor, MI 48109
wittkopp@umich.edu
ORCID: 0000-0001-7619-0048

Subject Areas: evolution, genetics, molecular biology

Keywords: *cis*-regulatory, *trans*-regulatory, regulatory network, fitness, pleiotropy, gene-by-environment

Abstract

Heritable variation in gene expression is common within and among species, and contributes to phenotypic diversity. Mutations affecting either *cis*- or *trans*-regulatory sequences controlling gene expression give rise to variation in gene expression, and natural selection acting on this variation causes some regulatory variants to persist in a population longer than others. To understand how mutation and selection interact to produce the patterns of regulatory variation we see within and among species, my colleagues and I have been systematically determining the effects of new mutations on expression of the *TDH3* gene in *Saccharomyces cerevisiae* and comparing them to the effects of polymorphisms segregating within this species. We have also investigated the molecular mechanisms by which regulatory variants act. Over the last decade, this work has revealed properties of *cis*- and *trans*-regulatory mutations including their relative frequency, effects, dominance, pleiotropy, and fitness consequences. Comparing these mutational effects to the effects of polymorphisms in natural populations, we have inferred selection acting on expression level, expression noise, and phenotypic plasticity. Here, I summarize this body of work and synthesize its findings to make inferences not readily discernible from the individual studies alone.

1. Introduction

Gene expression is important for physiology, development, and evolution. It is regulated by complex molecular networks built from interactions among *trans*-acting proteins, RNAs, and *cis*-regulatory DNA sequences (e.g., promoters, enhancers). Mutations in any of these components can contribute to expression differences within and between species, but studies have found that some types of changes contribute to polymorphism and divergence more often than others [1–3]. For example, comparing the relative contributions of *cis*- and *trans*-regulatory variation affecting expression of genes within and among species of *Saccharomyces* yeast [4] or *Drosophila* flies [5–7] showed that *trans*-regulatory variation was primarily responsible for expression differences within a species, with an increasing proportion of regulatory divergence attributable to *cis*-acting differences over evolutionary time. Some studies suggest that this pattern arises from purifying selection preferentially removing *trans*-regulatory variation [8] whereas others support a model of *trans*-regulatory variants evolving more neutrally with positive selection favoring *cis*-regulatory changes [9,10]. Understanding how the process of mutation introduces regulatory variation and how natural selection acts on it are necessary to understand why we see the patterns of regulatory variation we see as well as to predict the most likely paths of future evolutionary change.

Much of what we've learned about how new mutations affect gene expression has come from mutation accumulation studies. In these studies, a single individual (for selfing species) or a single pair of individuals (for sexually reproducing species) is used to produce the next generation, for many generations, allowing all but the most deleterious (e.g., lethal) mutations to accumulate over time [11]. To determine how these new mutations impact gene expression, mRNA transcript abundance was surveyed and compared among mutation accumulation lines. Such studies have been conducted in fruit flies [12,13], nematode worms [14,15], and baker's yeast [16], and comparing the observed mutational variance to the variation in gene expression seen in natural populations has provided evidence of both stabilizing and directional selection acting on different genes. Importantly, this work shows that the increase in a gene's expression variance due to new mutations each generation differs among genes, and, in yeast at least, correlates with the number of predicted *trans*-acting regulators [16]. However, because each mutation accumulation line carries many mutations, the effects of individual mutations cannot be discerned, and it is thus unknown whether these mutation(s) affect expression of any given gene in *cis* or in *trans*. Moreover, because few mutations affecting expression of any one gene are recovered in mutation accumulation studies [17], these data are insufficient for describing gene-specific distributions of mutational effects, which are important for understanding how expression of individual genes evolves.

To address these knowledge gaps, my colleagues and I have studied how hundreds of new mutations affect expression of the same focal gene. After establishing an experimental system suitable for isolating and characterizing many such new regulatory mutations, we quantified and compared properties of these *cis*- and *trans*-regulatory mutations including their relative frequency, effects, dominance, pleiotropy, interactions with the environment, and fitness. We then contrasted the properties of these new regulatory mutations with properties of regulatory polymorphisms segregating in the wild to infer the impacts of selection. Finally, we

examined the molecular mechanisms by which these new mutations affect expression of the focal gene. In this review, I synthesize these data for the first time, describing the system we used for this work and then integrating the major findings from each of these prior studies. Together, these data provide the most comprehensive understanding of regulatory variation available for any gene in any species. They also underscore the power that comes from studying the same system from many complementary angles. Future directions for this work as well as complementary work needed to assess the generalizability of these findings to other genes and other species are also discussed below.

2. Study system for isolating and characterizing new regulatory mutations

Characterizing the properties of new mutations affecting expression of any focal gene is challenging because such mutations are expected to arise rarely and most often have small effects, making them difficult to identify and isolate. Moreover, describing the properties of such mutations requires recovery of hundreds, if not thousands, of individual mutations all affecting expression of the same gene. To overcome these hurdles, we chose to isolate and characterize regulatory mutations in the single-celled baker's yeast *Saccharomyces cerevisiae*. By using *S. cerevisiae*, we were able to work with clonal cultures of either haploid or diploid cells and induce sexual reproduction when needed for genetic mapping. The genomic resources available for *S. cerevisiae*, including extensive information about the regulatory networks controlling gene expression [18], along with its rapid generation time and ability to store strains stably in the freezer, were also key to the success of our experiments.

We used a focal gene for this work consisting of the native *S. cerevisiae* *TDH3* promoter driving expression of a yellow fluorescent protein (YFP) because it provided a read-out of the focal gene's expression that could be easily quantified in individual cells in a high-throughput manner using flow cytometry (Figure 1A). We chose the *TDH3* promoter, which natively drives expression of a glyceraldehyde-3-phosphate dehydrogenase (GAPDH) protein involved in glycolysis, gluconeogenesis, and other processes [19,20] (Figure 1B), because it was one of the best characterized promoters in *S. cerevisiae*, with empirically verified binding sites for the transcription factors *RAP1p* and *GCR1p* (Figure 1C, [21,22]). In addition, deletion of *TDH3* is not lethal [23], presumably because its paralog *TDH2* has overlapping function (Figure 1C), suggesting that we would be able to recover mutations that alter, or even eliminate, *TDH3* expression unless they cause lethality because of their effects on other genes. Finally, the *TDH3* promoter drives high levels of YFP expression throughout the cell cycle [24], making it readily quantifiable using flow cytometry without cell synchronization.

We inserted the *P_{TDH3}-YFP* reporter gene into the *S. cerevisiae* genome rather than using an episomal vector to ensure it was regulated as similarly as possible to native *S. cerevisiae* genes. This reporter gene was originally inserted into a pseudogene on chromosome 1 in a derivative (BY4724) of the commonly used lab strain S288C, but after isolating hundreds of new regulatory mutations in this genetic background [25], we found that the high rate of spontaneous loss of mitochondrial function (i.e., petite formation) and the low rate of sporulation of this strain complicated measuring mutational effects and genetic mapping. We therefore modified this strain by introducing genetic variants identified previously [26,27] that reduce the frequency of petite formation and increase sporulation [28]. We also moved the *P_{TDH3}-YFP*

reporter gene from the pseudogene to the *HO* mating type locus on chromosome IV after discovering that the pseudogene was not present in all wild strains of *S. cerevisiae* [29]. Effects of mutant *TDH3* promoter alleles were highly correlated when the P_{TDH3} -YFP reporter gene was placed in these two different genomic locations as well as when these promoter alleles were assayed for their effects on expression of the native *TDH3* gene using a YFP fusion protein (Figure 1C, [28]). Effects of mutant promoter alleles on reporter gene expression were also highly correlated between different genetic backgrounds (Figure 1C, [30]).

Ideally, we would have used spontaneous mutations to generate our collection of mutants with individual *cis*- or *trans*- regulatory mutations affecting expression of P_{TDH3} -YFP; however, we reasoned (and later found to be true [25]) that the rate at which such mutations would arise spontaneously would be too low to realistically isolate the hundreds of mutations all affecting expression of the same gene needed to describe the properties of such regulatory mutations. We therefore used the chemical mutagen ethyl methanesulfonate (EMS) to elevate the mutation rate while still introducing mutations randomly throughout the genome. We used EMS to elevate mutation rate because (a) we could control the number of mutations introduced (by controlling EMS concentration and/or exposure time), (b) we could introduce mutations in millions of cells all at once, and (c) the mutation rate could be returned to wild-type levels by washing away the mutagen, creating stable mutant strains. Mutations introduced by EMS are almost exclusively G:C->A:T transitions, which happen to be the most common types of point mutations observed in *S. cerevisiae* mutation accumulation lines [31] and the most common type of synonymous mutation segregating within *S. cerevisiae* [32,33]. Haploid cells were used for these experiments so that we could see the effects of all mutations (i.e., no mutations were recessive).

As described in Gruber et al. [25], we titrated EMS exposure to determine a dose that made recovery of hundreds of mutants with altered P_{TDH3} -YFP expression feasible, but predicted few mutants would harbor more than one mutation affecting P_{TDH3} -YFP expression. This dose was 75% less than commonly used for an EMS mutagenesis screen in *S. cerevisiae*, with exposure time also reduced [25]. Genetic mapping and whole genome sequencing have since shown that mutants isolated under these conditions carry an average of 24 mutations, with only one of these mutations responsible for the effects on P_{TDH3} -YFP expression in 95% of the 43 mutants tested [34,35]. Fluorescent activated cell sorting (FACS) was used to isolate individual cells from mutagenized populations, selecting cells randomly [28] or based on their fluorescence [25]. In both cases, the sorted single cells were deposited onto a solid media, grown into colonies, and then used to inoculate liquid cultures, from each of which thousands of cells were individually measured for their fluorescence per unit cell size (Figure 1D). In all but the earliest study [25], four or more replicate populations were inoculated and analyzed for each mutant genotype. Collecting expression phenotypes in this way allowed us to calculate not only the average expression level of each genotype (corrected for differences in cell size), but also the variability in expression among genetically identical cells in the same environment (Figure 1D). This variability is defined as “expression noise” [36], and is a genetically controlled trait subject to selection [37–40]. This method of quantifying gene expression allowed us to consistently detect changes in the average reporter gene’s expression level as small as 1%, which is much more sensitive than methods such as RT-qPCR or RNA-seq. In Duveau et al.

[39], we compared YFP fluorescence per unit cell size to the abundance of YFP mRNA measured by pyrosequencing [41] and found a nonlinear relationship caused by saturation of the flow cytometry signal at high expression levels. We corrected for this nonlinearity in Duveau et al. [39] and all subsequent papers by using the empirically determined relationship to convert YFP fluorescence to estimated mRNA abundance.

As a complement to the collection of EMS mutants affecting P_{TDH3} -YFP expression, we used targeted mutagenesis to generate hundreds of mutant strains with mutations in the reporter gene's *TDH3* promoter and used the same flow cytometry methods to measure their effects. Initially, we made 236 strains each containing a single point mutation that changed one of the 241 G or C nucleotides in the wildtype *TDH3* promoter to an A or T, respectively, mimicking EMS mutations (Figure 1E, [30]). One strain developed inconsistent expression and was excluded from subsequent studies. We later expanded this set of mutant strains to include other types of promoter mutations [39] and introduced a subset of these *cis*-acting mutations into the native *TDH3* gene to measure their effects on fitness [39,42] and expression of other genes [43]. Since the publication of this work, two deep mutational scanning studies of the *TDH3* promoter have been published [44,45], although many of the alleles in those studies included multiple mutations and their effects were measured with different techniques, making it difficult to integrate the datasets.

3. Properties of *cis*- and *trans*-regulatory mutations

As described in the introduction, the primary motivation for setting up this system was to quantify properties of new regulatory mutations expected to influence their contribution to regulatory variation segregating within and among species. In particular, we sought to compare these properties between mutations that affected a focal gene's expression in *cis* or in *trans* given that our prior work showed that *cis*- and *trans*-regulatory variation make different contributions to regulatory divergence over evolutionary time [4,5,7]. Thus far, we have examined the relative frequency of *cis*- and *trans*-regulatory mutations affecting expression of the same focal gene, their distributions of effects on expression level and expression noise, their relative dominance, their impacts on fitness, and their impact on expression of other genes as a measure of pleiotropy. We have also investigated how the effects of these mutations and their impacts on fitness vary among environments. Primary conclusions from this work are summarized below.

3.1. Frequency

New mutations can alter a gene's activity by affecting its expression or the function of its encoded gene product (RNA and/or protein). To a first approximation, the mutational target size for a gene product can be predicted by the length of the gene's coding sequence, but the mutational target size for regulatory mutations is much less clear. In *S. cerevisiae*, *cis*-regulatory mutations tend to be located in non-coding sequences flanking the gene (Figure 2A), but only a subset of these sequences typically impact expression. *trans*-acting mutations can be located virtually anywhere in the genome, in either coding or non-coding sequences of direct or indirect regulators (Figure 2A). By using EMS to introduce mutations randomly throughout the genome, screening for their effects on expression of the focal gene, and classifying each mutation as

affecting the focal gene's coding, *cis*-regulatory, or *trans*-regulatory sequences, we directly compared the frequencies at which each of these types of mutations arise.

After isolating 231 mutants from the tails of the EMS-treated population that each caused a statistically significant change in expression of the P_{TDH3} -YFP focal gene, we re-sequenced the reporter gene and found that 7% had nonsynonymous mutations in the YFP coding sequence and 2% had putative *cis*-regulatory mutations in the promoter [25]. 10% of the mutants had two copies of the reporter gene [25], which we later found were due to spontaneous whole genome duplications and aneuploidies of the chromosome carrying the reporter gene (unpublished data). The remaining 81% of mutants were classified as *trans*-regulatory because the causative mutation was located somewhere outside of the coding or non-coding sequences of the focal gene [25]. The *cis*- and *trans*-acting effects of mutants classified as *cis*- or *trans*-regulatory were confirmed by testing for allele-specific effects on expression of the focal gene [25,46]. Taken together, these data suggest that new mutations alter a focal gene's expression much more often than they alter the function of its gene product, and that the vast majority of new mutations altering a gene's expression act in *trans*.

This study, Gruber et al. [25], provided the first empirical estimate of the relative frequency of *cis*- and *trans*-regulatory mutations for any eukaryotic gene, but these data were not well suited to robustly estimating the overall rate of *cis*- and *trans*-regulatory mutations. For example, because we only analyzed EMS-treated cells sorted from the extreme tails of the fluorescence distribution, we missed mutations with smaller effects on the focal gene's expression. In addition, we scored each potential mutant genotype in only a single population, and our subsequent mapping efforts suggested a higher false positive rate than we had anticipated [34]. We thus repeated the EMS mutagenesis experiment but sorted cells randomly from the EMS-treated population and analyzed each mutant genotype in at least 4 replicate populations to assess whether they harbored a mutation that affected P_{TDH3} -YFP expression [28]. Based on the results from Gruber et al. [25], we assumed that *cis*-regulatory mutants affecting expression of the focal gene would be negligible in the EMS-treated population and considered all 1,485 mutants isolated to be *trans*-regulatory. We compared these data to the 235 *cis*-regulatory mutations generated by site directed changes in the reporter gene's *TDH3* promoter [30].

After taking into account that the *cis*-regulatory mutations were sampled from 235 bp of the 678 bp *TDH3* promoter and the *trans*-regulatory mutations were sampled from all G:C basepairs in the 12 Mb *S. cerevisiae* genome, we estimated a mutational target size of 306 bp for *cis*-regulatory mutations and ~118,000 bp for *trans*-regulatory mutations affecting expression of the P_{TDH3} -YFP reporter gene using a significance threshold of $p = 0.05$ to identify statistically significant changes in expression [28]. However, because these estimates depend on the power of the experimental design and statistical threshold used to call significant expression changes, we also examined the relative frequency of *cis*- and *trans*-acting mutations with particular effect sizes (Figure 2B). Although *trans*-regulatory mutations were much more common than *cis*-regulatory mutations for most ranges of effect sizes, *cis*-regulatory mutations were more common among mutants that decreased expression of P_{TDH3} -YFP more than ~7.5% (Figure 2B,

[28]. Nonetheless, the overall greater frequency of *trans*-regulatory mutations suggests that they are likely to be a more abundant source of new variation in gene expression than *cis*-regulatory mutations.

3.2. Effects on expression level

New mutations can either increase or decrease expression of a focal gene to varying degrees. Given that *TDH3* is one of the most highly expressed genes in the *S. cerevisiae* genome [47], we anticipated that most new regulatory mutations would decrease expression of P_{TDH3} -YFP. We were thus surprised when our first study of new regulatory mutations recovered more mutants with increased than decreased expression of P_{TDH3} -YFP [25]. When we repeated the EMS mutagenesis experiment, sorting cells randomly from the EMS-treated population rather than from the tails, we again observed more mutations causing large (>7.5%) increases than decreases in expression [28]. However, when taking the full distribution of effects for EMS-treated mutants into account, the distribution was much more symmetrical (Figure 2D, [17,28]. Interestingly, the effects of new regulatory mutations altering expression of reporter genes driven by other *S. cerevisiae* promoters showed more asymmetry, with some distributions biased toward increased expression (e.g., *STM1*) and others (e.g., *RNR2*) biased toward decreased expression (Figure 2C, [17]. These differences in the distribution of effects for new regulatory mutations predict that even in the absence of natural selection, regulatory evolution is expected to vary from gene-to-gene.

Because nearly all of the mutations isolated from the EMS mutagenesis screen are expected to be *trans*-regulatory [25], we compared the effects of these mutations to the effects of the 235 *cis*-regulatory mutations introduced by site-directed mutagenesis. 45% of these *cis*-regulatory mutations increased expression of the reporter gene, and 55% decreased it; however, the mutations decreasing P_{TDH3} -YFP expression had much larger effects, resulting in an overall skew of *cis*-regulatory mutant effects toward decreased expression (Figure 2D, [28,30]. With the largest effect *trans*-regulatory mutations biased toward increased expression of the P_{TDH3} -YFP reporter gene [25,28], *cis*- and *trans*-regulatory mutations affecting activity of the *TDH3* promoter might be able to often compensate for each other. Ignoring the direction of effects, we found that *cis*-regulatory mutations tended to have larger effects on P_{TDH3} -YFP expression than *trans*-regulatory mutations (Figure 2D inset, [28]). If this tendency holds for most genes, the larger effects of *cis*-regulatory mutations might cause them to be more strongly advantageous or deleterious than *trans*-regulatory mutations (assuming a monotonic relationship between expression and fitness and ignoring possible effects of these mutations on expression of other genes), resulting in quicker fixation or elimination within a population.

3.3 Effects on expression noise

Mutations affecting the average expression level of a genotype also often affect expression noise, which is the variability in expression among genetically identical cells in the same environment [38]. For the P_{TDH3} -YFP reporter gene, we found that *cis*-regulatory mutations almost always increased expression noise whereas *trans*-regulatory mutations had more symmetrical effects, with *trans*-regulatory mutations causing much larger changes in expression noise than *cis*-regulatory mutations [28]. Consistent with prior work [38], *cis*-regulatory (but not

trans-regulatory) mutations showed a strong negative correlation between their effects on expression level and expression noise (Figure 2E, [28]). To try to disrupt this correlation, we created 171 new *cis*-regulatory mutant alleles of P_{TDH3} -YFP in which we (a) added binding sites for a transcription factor previously shown to affect expression noise [48], (b) mutated the TATA box, which also influences expression noise [38], and (c) combined these mutations with mutations in binding sites for the direct regulators RAP1p and GCR1p [39]. We also created strains with duplicated copies of *TDH3* carrying some of these mutations [42]. This set of new *cis*-regulatory mutants showed a weaker negative correlation between average expression level and expression noise than the original 235 point mutations tested [39], suggesting that some genetic variants can alter mean without changing noise and vice versa, potentially allowing natural selection to independently optimize both expression level and expression noise.

3.4 Dominance

Most laboratory strains of *S. cerevisiae* are haploid, yet wild isolates of *S. cerevisiae* are predominantly diploid [49]. In diploid populations, the dominance of new mutations impacts how they are affected by natural selection [50]. Consequently, differences in the dominance of *cis*- and *trans*-regulatory mutations should affect their relative evolutionary fate. For 212 of the 231 regulatory mutants isolated from our first EMS mutagenesis screen of haploid cells, we assessed dominance by crossing each mutant to another haploid cell of the opposite mating type and analyzing YFP fluorescence in the resulting diploid cells using flow cytometry [25]. We found that all 4 *cis*-regulatory mutants had similar effects on YFP fluorescence in haploid and diploid cells, indicating that they were dominant (Figure 2F). By contrast, 88% of the 171 *trans*-regulatory mutants tested showed no statistically significant change in YFP fluorescence in diploid cells, indicating that they were recessive (Figure 2F). This difference in dominance suggests that *cis*-regulatory mutations are more likely to be subject to selection as soon as they arise, whereas *trans*-regulatory mutations are more likely to segregate silently in a population until they are frequent enough to be subject to selection in homozygotes. Consistent with this observation, Lemos et al. [51] also concluded “that *cis*-acting alleles may be preferentially fixed by positive natural selection because of their higher additivity.”

3.5 Fitness

Natural selection favors the fixation of some alleles over others based on their relative fitness, which is measured by their reproductive success. For a single-cell organism like *S. cerevisiae*, the relative fitness of two genotypes can be estimated by comparing their relative growth rates, either by measuring growth of single genotypes individually or by competing two (or more) genotypes against each other by growing them together in a single culture. To measure the fitness effects of *cis*-regulatory mutations shown to alter expression of the P_{TDH3} -YFP reporter gene, we constructed new strains of *S. cerevisiae* in which each mutation of interest was inserted into the *TDH3* promoter of the native *TDH3* gene. The fitness effects of altering *TDH3* expression were then determined for each of these mutations by competing each strain against a common reference strain that had wild type *TDH3* expression and measuring their relative growth rates ([42]. Completely eliminating *TDH3* expression caused a ~7% reduction in fitness in rich media (consistent with prior studies, [23]). Promoter alleles driving *TDH3* expression between 95% to 120% of wildtype levels (reported as fluorescence rather than estimated mRNA

levels) had the highest (optimal) fitness and showed no statistically significant difference among them, suggesting that they might be selectively neutral (Figure 3A, [42]). Interestingly, 92% of 235 G:C → A:T point mutations tested in the *TDH3* promoter had effects in this neutral range (Figure 3A, [42]), suggesting they might be invisible to selection. Mutant promoter alleles driving *TDH3* expression between 0% and 94% of wild-type levels, or at 135% of wild-type levels, showed smaller reductions in fitness than the *TDH3* deletion (Figure 3A, [42]).

Impacts of *cis*-regulatory mutations on expression noise have also been predicted to impact fitness [37], but the correlated effects of mutations on expression level and expression noise and the anticipated small fitness effects of changing expression noise have made them difficult to measure. By quantifying the relative fitness effects of 43 *cis*-regulatory mutant alleles with a reduced correlation between expression level and expression noise, we were able to isolate and estimate the fitness effects attributable to changing expression noise [39]. We were also able to select five pairs of mutant *TDH3* alleles from these data with similar expression levels but differences in expression noise. All five pairs showed significant differences in fitness, but increased expression noise was not always advantageous or deleterious. For the 2 pairs of mutants with a median expression level close to wildtype, the genotype with higher noise had fitness reduced by ~0.1% (Figure 3B). By contrast, for the 3 pairs of genotypes with an average *TDH3* expression level further from wildtype, the genotype with higher noise increased fitness by ~0.2 - 0.4% (Figure 3B). Theoretical work [52] and simulation studies [39] also showed that higher levels of expression noise are deleterious when the average expression level is close to the fitness optimum and advantageous when it is not.

To measure the fitness effects of *trans*-regulatory mutations from the EMS mutagenesis screen, we first needed to map individual mutations responsible for changing *P_{TDH3}-YFP* expression because each EMS mutant carried an average of 24 new mutations [35]. We then tested the effects of the mapped mutation on reporter gene expression in the absence of other EMS-induced mutations. Using a bulk-segregant mapping strategy predicted to be able to map changes in expression as small as 1% [34], we identified the single point mutation responsible for altered reporter gene expression in 29 EMS-induced mutants, introduced that mutation into the un-mutagenized reference strain, and measured its effect on gene expression and fitness (Figure 3C, [43]). We also measured the fitness effects of 6 *RAP1* or *GCR1* mutant alleles that caused significant changes in expression of *P_{TDH3}-YFP* and were isolated from a deep mutational scanning experiment [35]. After excluding 2 *trans*-regulatory mutants for flocculation that affected estimates of growth rate, we found that 6 mutants grew 1% or more faster than wildtype and 23 mutants grew 1% or more slower (Figure 3D, [43]). Increases in growth rate ranged from 1.3% to 7% whereas decreases ranged from 1.2% to 70% (Figure 3D), with the four largest effects seen for mutant alleles of *GCR1* and *RAP1* [43]. Nine of these *trans*-regulatory mutations reduced fitness more than a complete deletion of *TDH3* (Figure 3D), suggesting that these *trans*-acting mutations are also affecting fitness in ways not mediated by their impact on *TDH3* expression.

3.6 Pleiotropy

trans-regulatory mutations are hypothesized to be more pleiotropic and thus more deleterious than *cis*-regulatory mutations affecting expression of the same focal gene because they should have effects on expression of the focal gene as well as effects on expression of other genes in the genome (Figure 4A). To determine whether *trans*-regulatory mutations were indeed more deleterious than *cis*-regulatory mutations, we used 5 *cis*-regulatory mutations and the 35 *trans*-regulatory mutations described above to cause *TDH3* expression to vary from 0% to 135% of wildtype levels and compared the relative fitness of each mutation conditioned on its effects on *TDH3*. (Effects of these mutations on gene expression are discussed in the molecular mechanism section below.) After excluding the two flocculant *trans*-regulatory mutants, we found that 17 of the 33 *trans*-regulatory mutations were more deleterious than *cis*-acting mutations with similar effects on *TDH3* expression, 15 had fitness indistinguishable from such *cis*-regulatory mutations, and 1 was significantly more beneficial (Figure 4B, [43]). The impact of each *trans*-regulatory mutation not attributable to its effects on *TDH3* (i.e., its pleiotropic fitness effect) was estimated by calculating the difference between the relative fitness of the *trans*-regulatory mutation and the predicted fitness of a *cis*-regulatory mutation with the same effect on *TDH3* expression (Figure 4B). Together, these differences were used to estimate a distribution of pleiotropic fitness effects (Figure 4C, top panel). Repeating this analysis with an additional 1106 *trans*-regulatory mutations using independent data from a gene deletion collection [29,53] showed a similar distribution, with the frequency of pleiotropic effects increasing or decreasing fitness by 10% largely symmetrical, but with a significant tail of pleiotropic fitness effects decreasing fitness more than 10% (Figure 4C, bottom panel) [43]. To the best of my knowledge, these data provide the first empirical estimates of the distribution of pleiotropic fitness effects for any trait. Such distributions are important for modeling evolutionary change because they predict how fitness effects not mediated by a focal trait of interest can impact the evolutionary fate of a mutation.

3.7 Gene-by-environment interactions

All of the analyses described above were performed in the same environment: a rich YPD media with glucose as the primary carbon source. However, it is well known that the regulation of gene expression and the effects of mutations can vary among environments. To test for environment-specific effects of mutations altering *TDH3* expression, we compared expression of the *P_{TDH3}-YFP* reporter gene in the 235 *cis*-regulatory mutant strains with individual G:C->A:T mutations in media containing glucose, galactose, or glycerol as a primary carbon source. We found that the wild type *TDH3* promoter exhibits plasticity in expression among these three types of media, with expression decreasing on galactose and increasing on glycerol relative to expression in cells grown on glucose (Figure 4D, [54]). The effects of some *cis*-regulatory mutations also varied among environments, providing evidence of gene-by-environment interactions. Specifically, mutations with the largest effects when cells were grown on the fermentable carbon source glucose had much smaller effects when cells were grown on the non-fermentable carbon source glycerol (Figure 4E, [54]). Mutational effects were much more similar (i.e., showed less gene-by-environment interactions) between cells grown on glucose and galactose, both of which are fermentable carbon sources (Figure 4F). These findings are consistent with *TDH3* expression being regulated by different transcription factor binding sites in

different environments [55] and show how single nucleotide changes can give rise to gene-by-environment interactions.

4. Inferring selection by comparing mutational effects to polymorphisms

In the absence of natural selection, genetic variants are fixed randomly by genetic drift and new mutations thus determine the pattern of variation expected to be seen within a population [56]. Understanding how new mutations introduce variation in gene expression is thus critical for developing neutral models of regulatory evolution that serve as null hypotheses for understanding whether and how natural selection has impacted the regulatory variation seen in the wild [12,14,16,57]. Using this general framework, we have examined the effects of selection acting on *cis*- and *trans*-regulatory variation contributing to differences in *TDH3* expression among strains of *S. cerevisiae* recently isolated from the wild.

4.1 *cis*-regulatory variation

Among 86 strains of *S. cerevisiae*, we identified 44 polymorphisms in the *TDH3* promoter, which is less than expected if this sequence were evolving neutrally [30] and indicates the *TDH3* promoter has been subject to purifying (negative) selection in the wild. As described above, mutations in the *TDH3* promoter can have deleterious effects because of their impacts on expression level or expression noise. To determine whether selection for either of these expression traits might be responsible for this signature of purifying selection, we determined the effect of each polymorphism on P_{TDH3} -YFP expression in the same genetic background as the original set of 235 *cis*-regulatory mutations and compared the distributions of effects for these mutations and polymorphisms for both expression level and expression noise. We found no significant difference in the distributions of effects for mutations and polymorphisms for expression level (Figure 5A), but did see a significant difference in their effects on expression noise (Figure 5B, [30]). Specifically, we found that mutations increased expression noise more often than polymorphisms, suggesting that selection has acted to minimize expression noise in natural populations [30].

These observations are consistent with our lab-based fitness assays showing that increased expression noise is deleterious when expression level is near the fitness optimum [39]. They are also consistent with the *TDH3* expression level being robust to most new mutations; less than 10% of mutations in the *TDH3* promoter caused changes in expression with measurable deleterious effects and the polymorphisms had small effects on *TDH3* expression predicted to be in the neutral range [42]. We expect that natural selection is acting to remove mutations at the small number of sites with large effects on *TDH3* expression, but that elimination of these mutations is not enough to cause a statistically significant difference in the distribution of effects for mutations and polymorphisms. Natural selection might also be eliminating variants in the *TDH3* promoter because of environment-specific effects that were not captured in our lab-based fitness assays. Indeed, a comparison of effects for mutations and polymorphisms on plasticity between glucose- and galactose-based media showed less plasticity among the polymorphisms than mutations, suggesting that *TDH3* promoter alleles with a particular degree of plasticity are more fit than others (Figure 5C). Taken together, these

studies identify multiple ways that promoter mutations can have deleterious effects on gene expression and thus influence promoter diversity.

4.2 *trans*-regulatory variation

Comparing the effects of individual mutations and polymorphisms to infer the impacts of selection on *trans*-regulatory variation is much more challenging than for *cis*-regulatory variation because *trans*-acting changes can be located virtually anywhere in the genome and it is difficult to isolate their individual effects [2]. To overcome these hurdles, we (a) used the EMS mutants described above to infer a distribution of *trans*-regulatory effects for individual new mutations (Figure 5D), (b) repeatedly sampled from this distribution of mutational effects to simulate multiple trajectories of neutral evolution (Figure 5E), and (c) compared these simulations of neutral evolution to the observed net *trans*-regulatory effects of polymorphisms in each of 56 strains of *S. cerevisiae* (Figure 5F, [58]). The net effect of *trans*-regulatory variation in each strain was determined empirically by introducing a wildtype P_{TDH3} -YFP reporter gene into each strain and measuring its expression [58]. We found that these strains showed less variability in *TDH3* expression than predicted to evolve neutrally from sampling the mutation distribution (Figure 5F), suggesting that *TDH3* expression is evolving under stabilizing selection [58]. Such stabilizing selection can cause the elimination of genetic variants that alter *TDH3* expression or the maintenance of sets of variants with off-setting, compensatory effects. Mapping quantitative trait loci affecting expression of the reporter gene (eQTLs) in 3 of these 56 strains relative to a common lab strain (Figure 5G) supported the latter model: ~100 eQTL with similar frequencies of eQTL increasing and decreasing expression in *trans* were identified in each comparison (Figure 5H, [58]). These observations provide rare empirical support for theoretical models showing that stabilizing selection acting on quantitative traits can maintain many alleles with compensatory effects [59].

5. Molecular mechanisms of *cis*- and *trans*-regulatory variants

Connecting genetic changes to their effects on gene expression and fitness can help us predict how gene expression might evolve, but this information alone provides little insight into *why* these genetic changes have the fitness effects that they do. We aim to fill this knowledge gap by investigating the molecular mechanisms by which new regulatory mutations alter gene expression. I anticipate that understanding these mechanisms more completely will not only help us make sense of current patterns of regulatory variation, but will also help us predict future evolutionary changes in gene expression and the phenotypes they impact based on the structure of regulatory networks.

5.1 *cis*-regulatory mechanisms

Mutations in *cis*-regulatory sequences such as promoters (and enhancers in multicellular organisms) are most often thought to alter gene expression by affecting binding sites for transcription factors. The Rap1p and Gcr1p transcription factors directly regulate expression of *TDH3* as well as other glycolytic genes [21,22,60]. We found that mutations in these binding sites had the largest effects on expression of the P_{TDH3} -YFP reporter gene of the 236 G:C->A:T promoter mutations tested (Figure 6A, [30]). A few other clusters of mutations with similar effects on expression were also observed suggesting that they might be binding sites for other

transcription factors (Figure 6A), but which transcription factors (if any) bind to these sites remains unknown. In addition to Rap1p and Gcr1p, the YEASTRACT database suggests *TDH3* might be directly regulated by Fhl1p, Yap1p, Gcn4p, Hsf1p, Skn7p, Tye7p, Gcr2p, and/or Cin5p [18,35], and a recent genomic study of transcription factor binding in *S. cerevisiae* and *S. paradoxus* adds Msn1p, Swi5p, Fkh1p, and Fkh2p to that list of candidate direct regulators [61]. Promoter mutations can also affect gene expression by altering nucleosome occupancy (with or without changes in transcription factor binding), which has also been shown to impact the mutability of promoters in *S. cerevisiae* [62]. Because sets of transcription factors regulating a gene's expression can differ among environments, the same *cis*-regulatory mutation can have different environment-specific effects, as we have observed for *TDH3* [54].

cis-regulatory mutations that impact expression of their focal gene can also impact the expression of other ("downstream") genes. For *TDH3*, we found that its deletion caused statistically significant changes in expression of 140 other genes [43]. When we measured expression of these genes in *cis*-regulatory mutants with *TDH3* expression varying from 20% to 135% of the wild-type levels, we found a strong correlation between the expression level of genes in this group and the level of *TDH3* expression (Figure 6B, [43]). Although there is some evidence that *TDH3* itself might act as a transcriptional regulator [20], we find it more likely that homeostatic feedback mechanisms responding to changes in metabolite levels caused by altering *TDH3* expression are responsible for these effects on downstream genes. For example, *TDH2*, a gene paralogous to *TDH3*, was upregulated in response to reductions in *TDH3* expression [43]. Sequences of Tdh2p and Tdh3p are highly conserved, suggesting that this upregulation of *TDH2* in response to reduced *TDH3* expression might ameliorate the mutation's fitness effects. *GCR1*, which regulates expression of *TDH2* as well as *TDH3*, was also upregulated upon deletion of *TDH3*, suggesting that it might be involved in the molecular mechanism responsible for compensatory expression of *TDH2* (Vande Zande et al. in prep). These data underscore the complexity of regulatory networks and highlight the fact that *cis*-regulatory mutations can have far-reaching impacts beyond expression of the gene they most immediately regulate.

5.2 *trans*-regulatory mechanisms

trans-regulatory mutations are generally assumed to have more wide-spread effects on gene expression than *cis*-regulatory mutations altering expression of the same focal gene because they should have effects similar to *cis*-regulatory mutations on expression of downstream genes as well as affect expression of other genes in parallel (Figure 4A). To test this assumption, we compared the impact of 5 *cis*- and 35 *trans*-regulatory mutations affecting expression of *TDH3* on expression of all other genes in the *S. cerevisiae* genome. For *cis*- and *trans*-regulatory mutations with similar impacts on expression of the focal gene, we found that the *trans*-regulatory mutation nearly always affected expression of more genes (Figure 6C, [43]). Using expression data from a gene deletion collection to expand this analysis to 748 other focal genes, we found that this relationship generally holds and can be explained by the structure of regulatory networks controlling gene expression [63]. Gene expression changes that are due to a *trans*-regulatory mutation could be independent of the mutation's effect on the focal gene or mediated by the change in the focal gene's expression. We used the effects of *cis*-regulatory

mutations to predict how each *trans*-regulatory mutation should alter the expression of downstream genes via the change in *TDH3* expression. Surprisingly, in many cases, we found that a *trans*-regulatory mutation had impacts that differed from this prediction (Figure 6D, [43]), indicating that the *trans*-regulatory mutation was altering expression of genes downstream of the focal gene in a way independent of its impacts on *TDH3* (Figure 6E).

The 35 *trans*-regulatory mutants used to compare the effects of *cis*- and *trans*-regulatory mutations on gene expression all harbored changes in the coding sequence of proteins that presumably affected their function. These proteins encoded the transcription factors Rap1p and Gcr1p known to directly regulate *TDH3* expression as well as 8 other transcriptional regulators (Mrn1p, Tup1p, Bre2p, Caf40p, Cyc8p, Ssn2p, Tra1p, and Tye7p) that could regulate *TDH3* expression directly or indirectly [43]. The remaining *trans*-regulatory mutations presumably affected *TDH3* expression indirectly and altered 4 proteins involved in purine biosynthesis, 4 involved in iron transport, and 10 involved in other processes. Targeted mutagenesis was used to obtain the 6 mutant alleles of *RAP1* and *GCR1* that altered *TDH3* expression [35], but the other 29 *trans*-regulatory mutants were isolated from the random EMS mutagenesis screen [35]. Only 1 of these 66 mutations was located in a non-coding region, suggesting that new *trans*-regulatory mutations (at least within the range of effects we could map) primarily affect coding sequences. Of the 42 genes affected by these 66 *trans*-regulatory mutations, transcription factors predicted by YEASTRACT to regulate *TDH3* expression were affected more often than expected by chance (Figure 6F, [35]), yet more than 90% of all *trans*-regulators identified were not transcription factors. Interestingly, the 66 *trans*-regulatory mutations were found more often in regions identified as having eQTL affecting *TDH3* expression in our prior work [58] than expected by chance [35], suggesting that our knowledge of how *trans*-regulatory mutations are distributed within the *TDH3* regulatory network might have predictive power for the sources of *trans*-regulatory variation affecting *TDH3* expression segregating in natural populations. While this might seem expected, prior work suggests that knowledge of regulatory networks has had limited success predicting quantitative variation in genome-wide association studies [64].

6. Concluding remarks and future directions

Taken together, the data presented here provide an unparalleled look at the causes and consequences of regulatory variation affecting expression of a gene and how natural selection acts on it. With many properties found to differ between *cis*- and *trans*-regulatory mutations that can impact their contributions to regulatory evolution -- some in apparently contradicting ways (Table 1) -- modeling work is now needed to better understand how these factors interact with each other and with natural selection under various conditions. Importantly, the empirical data described above provide the information needed to parameterize such models with realistic values. These models can be used, for example, to determine the conditions under which *cis*-regulatory changes accumulate preferentially relative to *trans*, as has been reported for both flies and yeast [4,7]. Work in progress to determine how epistasis alters distributions of mutational effects, how the relationship between expression and fitness varies among environments, and how regulation of *TDH3* has diverged between species and among its paralogs, will allow us to use this system to further explore regulatory evolution in greater depth.

But important questions remain about how the properties of regulatory mutations affecting expression of the *S. cerevisiae* *TDH3* gene relate to properties of regulatory mutations affecting expression of other genes in *S. cerevisiae* and other species. Based on what is known about the structure of regulatory networks, I expect that, as seen for *TDH3*, *trans*-regulatory mutations affecting a focal gene's expression will arise more often than *cis*-regulatory mutations for all genes and species because the differences in mutational target size are assumed to result from features of regulatory network structure shared by all eukaryotes. I also expect that the larger average effects of *cis*- than *trans*-regulatory mutations on the focal gene's expression as well as the greater dominance of *cis*-regulatory mutations that we observed for *TDH3* will also hold for other genes and species because these properties likely result from general molecular mechanisms of gene regulation. In addition, as described above, our recent work suggests that the greater average pleiotropy we observed for *trans*- relative to *cis*-regulatory mutations affecting *TDH3* expression is generalizable and explainable by the structure of regulatory networks having few highly-connected and many lowly-connected nodes. Properties that I expect to vary more significantly from gene-to-gene in *S. cerevisiae*, as well as among genes in other species, include the balance of mutations that increase or decrease expression of the focal gene, which we have observed among reporter genes driven by nine other *S. cerevisiae* promoters [17]; because most new mutations are expected to disrupt molecular function, I hypothesize that this difference in the relative frequency of mutations that increase or decrease expression of the focal gene results from the balance of activators and repressors in each gene's regulatory network. The relationship between gene expression levels and fitness is also expected to vary among genes and species. In thinking about how our findings from a single-celled organism might relate to a multicellular species, I predict that we'll see similar properties if analyzing effects of new regulatory mutations within a single cell type, but that the effects of those mutations will vary among cell types, similar to the differences we observed when comparing effects of individual regulatory mutations in *S. cerevisiae* among environments.

Despite the need for comparable data for other genes and species, studies of *cis*- and *trans*-regulatory mutations and polymorphisms affecting expression of the *S. cerevisiae* *TDH3* gene provide an important multi-dimensional look at the interplay of mutation and selection in the evolution of gene expression, considering both the arrival and the survival of the fittest. As such, this work provides a much needed bridge between the too often disconnected fields of molecular and evolutionary biology. Such bridges are needed to achieve a "functional synthesis" [65] that integrates our current understanding of molecular, cellular, and developmental biology with the more theoretical models of population genetics, quantitative genetics, and evolutionary biology. Working at the interface of these fields is essential for understanding biology given that developmental and molecular systems are the product of evolution as well as the context in which new traits evolve.

Acknowledgements

The body of work synthesized in this review would not have been possible without the input, creativity, dedication, and energy of former and current members of the Wittkopp lab, including (chronologically) Gizem Kalay, Jonathan Gruber, David Yuan, Brian Metzger, Kara Vogel,

Elizabeth Walker, Bing Yang, Hussein Al-Asadi, Fabien Duveau, William Toubiana, Stephen Tryban, Katya Mack, Natasha Sood, Tiffany Brooks, Andrea Hodgins-Davis, Crisandra Diaz, Petra Vande Zande, Lisa Kim, Mark Hill, Anna Redgrave, Mohammad Siddiq, Nicholas Brown, Eden McQueen, and Hannah Kania. An additional thanks to Petra Vande Zande, Brian Metzger, Fabien Duveau, and Mo Siddiq for feedback on the manuscript. Figures from [28] and [42] reproduced by permission of Oxford University Press and the Society for Molecular Biology and Evolution. Figures reprinted from [66] with permission from Elsevier.

Funding statement

Financial support for this work came from the John Simon Guggenheim Memorial Foundation, Alexander von Humboldt Foundation, National Science Foundation (DEB-1911322 and MCB-1929737), National Institutes of Health (R35GM118073 and R01GM108826), March of Dimes (5-FY07-181), and Alfred P. Sloan Foundation.

References

1. Signor SA, Nuzhdin SV. 2018 The Evolution of Gene Expression in cis and trans. *Trends Genet.* **34**, 532–544.
2. Hill MS, Vande Zande P, Wittkopp PJ. 2021 Molecular and evolutionary processes generating variation in gene expression. *Nat. Rev. Genet.* **22**, 203–215.
3. Brown KE, Kelly JK. 2022 Genome-wide association mapping of transcriptome variation in *Mimulus guttatus* indicates differing patterns of selection on cis- versus trans-acting mutations. *Genetics* **220**. (doi:10.1093/genetics/iyab189)
4. Metzger BPH, Wittkopp PJ, Coolon JD. 2017 Evolutionary Dynamics of Regulatory Changes Underlying Gene Expression Divergence among *Saccharomyces* Species. *Genome Biol. Evol.* **9**, 843–854.
5. Wittkopp PJ, Haerum BK, Clark AG. 2008 Regulatory changes underlying expression differences within and between *Drosophila* species. *Nat. Genet.* **40**, 346–350.
6. McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, Wittkopp PJ. 2010 Regulatory divergence in *Drosophila* revealed by mRNA-seq. *Genome Res.* **20**, 816–825.
7. Coolon JD, McManus CJ, Stevenson KR, Graveley BR, Wittkopp PJ. 2014 Tempo and mode of regulatory evolution in *Drosophila*. *Genome Res.* **24**, 797–808.
8. Josephs EB, Lee YW, Wood CW, Schoen DJ, Wright SI, Stinchcombe JR. 2020 The Evolutionary Forces Shaping Cis- and Trans-Regulation of Gene Expression within a Population of Outcrossing Plants. *Mol. Biol. Evol.* **37**, 2386–2393.
9. Emerson JJ, Hsieh L-C, Sung H-M, Wang T-Y, Huang C-J, Lu HH-S, Lu M-YJ, Wu S-H, Li W-H. 2010 Natural selection on cis and trans regulation in yeasts. *Genome Res.* **20**, 826–836.
10. Coolon JD, Stevenson KR, McManus CJ, Yang B, Graveley BR, Wittkopp PJ. 2015 Molecular mechanisms and evolutionary processes contributing to accelerated divergence of gene expression on the *Drosophila* X chromosome. *Mol. Biol. Evol.* **32**, 2605–2615.
11. Halligan DL, Keightley PD. 2009 Spontaneous Mutation Accumulation Studies in Evolutionary Genetics. *Annu. Rev. Ecol. Evol. Syst.* **40**, 151–172.
12. Rifkin SA, Kim J, White KP. 2003 Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nat. Genet.* **33**, 138–144.
13. McGuigan K, Collet JM, McGraw EA, Ye YH, Allen SL, Chenoweth SF, Blows MW. 2014 The Nature and Extent of Mutational Pleiotropy in Gene Expression of Male *Drosophila serrata*. *Genetics*. **196**, 911–921. (doi:10.1534/genetics.114.161232)
14. Denver DR, Morris K, Streelman JT, Kim SK, Lynch M, Thomas WK. 2005 The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. *Nat. Genet.* **37**, 544–548.
15. Deiss TC, Konrad A, Bergthorsson U, Katju V. 2021 Global Gene Expression Divergence in Spontaneous Mutation Accumulation Lines of *Caenorhabditis elegans* under Varying

Efficiency of Selection. *bioRxiv.* , 2021.05.03.442498. (doi:10.1101/2021.05.03.442498)

16. Landry CR, Lemos B, Rifkin SA, Dickinson WJ, Hartl DL. 2007 Genetic properties influencing the evolvability of gene expression. *Science* **317**, 118–121.
17. Hodgins-Davis A, Duveau F, Walker EA, Wittkopp PJ. 2019 Empirical measures of mutational effects define neutral models of regulatory evolution in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 21085–21093.
18. Teixeira MC *et al.* 2018 YEASTRACT: an upgraded database for the analysis of transcription regulatory networks in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **46**, D348–D353.
19. McAlister L, Holland MJ. 1985 Isolation and characterization of yeast strains carrying mutations in the glyceraldehyde-3-phosphate dehydrogenase genes. *J. Biol. Chem.* **260**, 15013–15018.
20. Ringel AE, Ryznar R, Picariello H, Huang K-L, Lazarus AG, Holmes SG. 2013 Yeast Tdh3 (glyceraldehyde 3-phosphate dehydrogenase) is a Sir2-interacting factor that regulates transcriptional silencing and rDNA recombination. *PLoS Genet.* **9**, e1003871.
21. Yagi S, Yagi K, Fukuoka J, Suzuki M. 1994 The UAS of the yeast GAPDH promoter consists of multiple general functional elements including RAP1 and GRF2 binding sites. *J. Vet. Med. Sci.* **56**, 235–244.
22. Huie MA, Scott EW, Drazinic CM, Lopez MC, Hornstra IK, Yang TP, Baker HV. 1992 Characterization of the DNA-binding activity of GCR1: in vivo evidence for two GCR1-binding sites in the upstream activating sequence of TPI of *Saccharomyces cerevisiae*. *Molecular and Cellular Biology.* **12**, 2690–2700. (doi:10.1128/mcb.12.6.2690)
23. Pierce SE, Davis RW, Nislow C, Giaever G. 2007 Genome-wide analysis of barcoded *Saccharomyces cerevisiae* gene-deletion mutants in pooled cultures. *Nat. Protoc.* **2**, 2958–2974.
24. Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B. 1998 Comprehensive Identification of Cell Cycle-regulated Genes of the Yeast *Saccharomyces cerevisiae* by Microarray Hybridization. *MBoC* **9**, 3273–3297.
25. Gruber JD, Vogel K, Kalay G, Wittkopp PJ. 2012 Contrasting properties of gene-specific regulatory, coding, and copy number mutations in *Saccharomyces cerevisiae*: frequency, effects, and dominance. *PLoS Genet.* **8**, e1002497.
26. Deutschbauer AM, Davis RW. 2005 Quantitative trait loci mapped to single-nucleotide resolution in yeast. *Nat. Genet.* **37**, 1333–1340.
27. Dimitrov LN, Brem RB, Kruglyak L, Gottschling DE. 2009 Polymorphisms in Multiple Genes Contribute to the Spontaneous Mitochondrial Genome Instability of *Saccharomyces cerevisiae* S288C Strains. *Genetics.* **183**, 365–383. (doi:10.1534/genetics.109.104497)
28. Metzger BPH, Duveau F, Yuan DC, Tryban S, Yang B, Wittkopp PJ. 2016 Contrasting Frequencies and Effects of cis- and trans-Regulatory Mutations Affecting Gene Expression. *Mol. Biol. Evol.* **33**, 1131–1146.

29. Maclean CJ, Metzger BPH, Yang J-R, Ho W-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.
30. Metzger BPH, Yuan DC, Gruber JD, Duveau F, Wittkopp PJ. 2015 Selection on noise constrains variation in a eukaryotic promoter. *Nature* **521**, 344–347.
31. Lynch M *et al.* 2008 A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 9272–9277.
32. Liti G *et al.* 2009 Population genomics of domestic and wild yeasts. *Nature* **458**, 337–341.
33. Schacherer J, Shapiro JA, Ruderfer DM, Kruglyak L. 2009 Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* **458**, 342–345.
34. Duveau F, Metzger BPH, Gruber JD, Mack K, Sood N, Brooks TE, Wittkopp PJ. 2014 Mapping small effect mutations in *Saccharomyces cerevisiae*: impacts of experimental design and mutational properties. *G3* **4**, 1205–1216.
35. Duveau F, Vande Zande P, Metzger BPH, Diaz CJ, Walker EA, Tryban S, Siddiq MA, Yang B, Wittkopp PJ. 2021 Mutational sources of trans-regulatory variation affecting gene expression in *Saccharomyces cerevisiae*. *Elife* **10**, e67806.
36. Raser JM, O'Shea EK. 2004 Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811–1814.
37. Lehner B. 2008 Selection to minimise noise in living systems and its implications for the evolution of gene expression. *Mol. Syst. Biol.* **4**, 170.
38. Hornung G, Bar-Ziv R, Rosin D, Tokuriki N, Tawfik DS, Oren M, Barkai N. 2012 Noise-mean relationship in mutated promoters. *Genome Res.* **22**, 2409–2417.
39. Duveau F, Hodgins-Davis A, Metzger BP, Yang B, Tryban S, Walker EA, Lybrook T, Wittkopp PJ. 2018 Fitness effects of altering gene expression noise in *Saccharomyces cerevisiae*. *Elife* **7**, e37272.
40. Schmiedel JM, Carey LB, Lehner B. 2019 Empirical mean-noise fitness landscapes reveal the fitness impact of gene expression noise. *Nat. Commun.* **10**, 3180.
41. Wittkopp PJ. 2011 Using pyrosequencing to measure allele-specific mRNA abundance and infer the effects of cis- and trans-regulatory differences. *Methods Mol. Biol.* **772**, 297–317.
42. Duveau F, Toubiana W, Wittkopp PJ. 2017 Fitness Effects of Cis-Regulatory Variants in the *Saccharomyces cerevisiae* TDH3 Promoter. *Mol. Biol. Evol.* **34**, 2908–2912.
43. Vande Zande P, Hill MS, Wittkopp PJ. 2022 Pleiotropic effects of trans-regulatory mutations on fitness and gene expression. *Science* **377**, 105–109.
44. Kotopka BJ, Smolke CD. 2020 Model-driven generation of artificial yeast promoters. *Nat. Commun.* **11**, 2113.
45. Li S, Ma L, Fu W, Su R, Zhao Y, Deng Y. 2022 Programmable Synthetic Upstream Activating Sequence Library for Fine-Tuning Gene Expression Levels in *Saccharomyces*

- cerevisiae*. *ACS Synth. Biol.* **11**, 1228–1239.
46. Wittkopp PJ, Haerum BK, Clark AG. 2004 Evolutionary changes in cis and trans gene regulation. *Nature* **430**, 85–88.
 47. Newman JRS, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, Weissman JS. 2006 Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* **441**, 840–846.
 48. Sharon E, van Dijk D, Kalma Y, Keren L, Manor O, Yakhini Z, Segal E. 2014 Probing the effect of promoters on noise in gene expression using thousands of designed sequences. *Genome Res.* **24**, 1698–1706.
 49. Hanson SJ, Wolfe KH. 2017 An Evolutionary Perspective on Yeast Mating-Type Switching. *Genetics* **206**, 9–32.
 50. Fisher RA. 1923 XXI.—On the Dominance Ratio. *Proceedings of the Royal Society of Edinburgh* **42**, 321–341.
 51. Lemos B, Araripe LO, Fontanillas P, Hartl DL. 2008 Dominance and the evolutionary accumulation of cis- and trans-effects on gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 14471–14476.
 52. Tănase-Nicola S, ten Wolde PR. 2008 Regulatory control and the costs and benefits of biochemical noise. *PLoS Comput. Biol.* **4**, e1000125.
 53. Kemmeren P *et al.* 2014 Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. *Cell* **157**, 740–752.
 54. Dubeau F, Yuan DC, Metzger BPH, Hodgins-Davis A, Wittkopp PJ. 2017 Effects of mutation and selection on plasticity of a promoter activity in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E11218–E11227.
 55. Kuroda S, Otaka S, Fujisawa Y. 1994 Fermentable and nonfermentable carbon sources sustain constitutive levels of expression of yeast triosephosphate dehydrogenase 3 gene from distinct promoter elements. *J. Biol. Chem.* **269**, 6153–6162.
 56. Kimura M. 1968 Evolutionary rate at the molecular level. *Nature* **217**, 624–626.
 57. Smith JD, McManus KF, Fraser HB. 2013 A novel test for selection on cis-regulatory elements reveals positive and negative selection acting on mammalian transcriptional enhancers. *Mol. Biol. Evol.* **30**, 2509–2518.
 58. Metzger BPH, Wittkopp PJ. 2019 Compensatory trans-regulatory alleles minimizing variation in TDH3 expression are common within *Saccharomyces cerevisiae*. *Evol Lett* **3**, 448–461.
 59. Barton NH. 1986 The maintenance of polygenic variation through a balance between mutation and stabilizing selection. *Genet. Res.* **47**, 209–216.
 60. Deminoff SJ, Santangelo GM. 2001 Rap1p requires Gcr1p and Gcr2p homodimers to activate ribosomal protein and glycolytic genes, respectively. *Genetics* **158**, 133–143.

61. Krieger G, Lupo O, Wittkopp P, Barkai N. 2022 Evolution of transcription factor binding through sequence variations and turnover of binding sites. *Genome Res.* **32**, 1099–1111.
62. Hornung G, Oren M, Barkai N. 2012 Nucleosome organization affects the sensitivity of gene expression to promoter mutations. *Mol. Cell* **46**, 362–368.
63. Vande Zande P, Wittkopp PW. 2022 Network topology can explain differences in pleiotropy between cis and trans-regulatory mutations. *Mol. Biol. Evol.* **in press**.
64. Flint J, Ideker T. 2019 The great hairball gambit. *PLoS Genet.* **15**, e1008519.
65. Dean AM, Thornton JW. 2007 Mechanistic approaches to the study of evolution: the functional synthesis. *Nat. Rev. Genet.* **8**, 675–688.
66. Siddiq MA, Wittkopp PJ. 2022 Mechanisms of regulatory evolution in yeast. *Curr. Opin. Genet. Dev.* **77**, 101998.

Figures Legends and Table

Figure 1. Isolating and characterizing *cis*- and *trans*-regulatory mutations

(A) Image shows expression of YFP produced by the reporter gene in living cells, which allows quantification of reporter gene expression by flow cytometry. Photo by Fabien Duveau. (B) The proteins encoded by *TDH3* and its paralog *TDH2* encode GAPDH proteins that function in both glycolysis (blue) and gluconeogenesis (red). (C) Schematic shows the P_{TDH3} -YFP reporter gene, containing the 678 bp *TDH3* promoter regulated by a TATA box and binding sites for the experimentally confirmed RAP1p and GCR1p transcription factors, the coding sequence for a yellow fluorescent protein (YFP, yellow), and a *CYC1* terminator sequence (gray), that was integrated into the *S. cerevisiae* genome at a pseudogene on chromosome I (shown) or the *HO* locus on chromosome IV (not shown). (D) Plots show highly correlated effects of mutated *TDH3* promoter alleles in the P_{TDH3} -YFP reporter gene integrated into the pseudogene and the *HO* locus (left panel), in the reporter gene at the pseudogene and at the native *TDH3* gene on chromosome VII fused to a YFP fusion protein (middle panel), and in the reporter gene at the pseudogene in different genetic backgrounds (right panel). “Exp.” = expression of the reporter gene. Top and middle panels reproduced from Figure S3A and Figure S3B in [28], respectively, and bottom panel reproduced from Extended Data Figure 7A in [30]. (E) Experimental design for sampling *trans*-regulatory mutations is shown. A population of cells carrying the P_{TDH3} -YFP reporter gene was treated with a low dose of ethyl methanesulfonate (EMS), and then single cells were sorted either from the tails of the fluorescence distribution [25] or randomly [28] onto solid media. Each colony grown from a sorted single cell was used to inoculate one or more liquid cultures, with fluorescence and cell size measured for thousands of cells analyzed from each of these cultures using flow cytometry. Average expression level (μ) and expression noise (related to σ) were then calculated for each population. (E) A schematic of the wildtype *TDH3* promoter is shown with binding sites for RAP1p and GCR1p indicated. The 236 mutant promoter alleles, each with a single G:C → A:T mutation (red X), are also shown.

Figure 2. Frequency, effects, and dominance of new regulatory mutations

(A) Schematic shows the mutational target sizes for *cis*- and *trans*-regulatory mutations. *cis*-regulatory mutations are limited to *cis*-acting DNA sequences, typically located near the focal gene. *trans*-regulatory mutations can be located in the coding or non-coding sequences of direct or indirect regulators located anywhere in the genome. Reproduced from Figure 1 in [66]. (B) The relative frequency of *cis*-regulatory (red) and *trans*-regulatory (blue) mutations with various effect sizes relative to wildtype (WT) are shown. Dotted lines indicate the maximum potential target size for *cis*-regulatory (678 bp) and *trans*-regulatory (12 Mb) mutations. Modified from Figure 2D in [28]. (C) Distribution of mutational effects determined by EMS mutagenesis for P_{TDH3} -YFP (*TDH3*, maroon) and nine other reporter genes driven by promoters from other *S. cerevisiae* genes are shown. X-axis shows effects of mutations measured as Z-scores relative to the un-mutagenized, wild type strain. Reproduced from Figure 3A in [17]. (D) Histograms show the effects of 235 *cis*-regulatory mutations in the *TDH* promoter (red) and 1485 putatively *trans*-regulatory mutations introduced by EMS (blue) on expression of P_{TDH3} -YFP reporter gene. Inset boxplots show the relative magnitude of effects for these mutations, with the 1 *cis*-regulatory and 4 *trans*-regulatory mutants causing changes in expression greater than 11% not shown. Modified from Figures 1A, 1C, and 2A in [28]. (E) *cis*-regulatory (red), but not *trans*-regulatory (blue), mutations have

negatively correlated effects on expression level and expression noise. In each case, oval contains 95% of the mutants, and dashed lines show principal components for each set of mutations. Reproduced from Figure 5 in [28]. (F) Effects of *cis*-regulatory (red) and *trans*-regulatory (blue) mutations in haploid (X-axis) and diploid (Y-axis) cells show that *cis*-acting mutations tend to be dominant and *trans*-acting mutations tend to be recessive. Modified from Figure 4A in [25].

Figure 3. Fitness effects of *cis*- and *trans*-regulatory mutations

(A) Fitness effects of changing *TDH3* expression are shown by comparing *TDH3* expression (X-axis) to fitness (Y-axis), both relative to wildtype (WT). Error bars show 95% confidence intervals for expression and fitness, and gray shading shows the 95% confidence interval around a LOESS regression of expression on fitness. Red histogram shows the distribution of effects for 235 *cis*-regulatory mutant strains. Modified from Figure 2C in [42]. (B) Relative fitness (high/low noise genotype) is shown for five pairs of *cis*-regulatory mutant alleles with similar expression levels ("Exp level") but differences in expression noise ("Δ Noise"). Reproduced from Figure 4A in [39]. (C) Individual *trans*-regulatory mutations mapped from EMS mutants had similar effects on P_{TDH3} -YFP expression when tested in isolation. X-axis shows effects on expression in the original EMS mutant, and Y-axis shows effects of the mapped mutation introduced alone. Modified from Figure 2G in [35]. (D) Relative fitness of 33 *trans*-regulatory mutations causing *TDH3* expression to vary from 0% to 135% of wildtype levels (black) are shown overlaid on a violin plot describing these data. Dotted line indicates fitness of the un-mutagenized wild type strain, and the red line indicates the relative fitness of a *TDH3* deletion (i.e., the most severe *cis*-regulatory mutant possible) observed in this study. Produced using data from Figure 2B of [43].

Figure 4. Pleiotropy, plasticity, and gene-by-environment interactions

(A) Schematic shows the anticipated effects of mutations in *cis*-regulatory (red) and *trans*-regulatory (blue) factors affecting expression of the same focal gene. Modified from Figure 1 in [43]. (B) Relative fitness of 5 *cis*-regulatory (red) and 33 *trans*-regulatory (blue) mutations affecting *TDH3* expression are shown (Y-axis) compared to their effects on *TDH3* expression (X-axis). Modified from Figure 2B in [43]. (C) Distribution of pleiotropic effects on fitness, relative to *TDH3*, are shown for 33 *trans*-regulatory mutations (top) and 1106 *trans*-acting gene deletions (bottom). Modified from Figures 2C and S2B in [43]. (D) Expression of the wildtype *TDH3* promoter driving YFP expression was measured in four replicate populations following growth on media containing glucose (Glu), galactose (Gal) or glycerol (Gly) as a primary carbon source. The average expression level for each replicate and the overall mean are shown. Glucose and galactose are fermentable carbon sources, whereas glycerol is non-fermentable. Right panel modified from Figure 1B in [54]. (E) Effects of 235 *cis*-regulatory mutations on expression of P_{TDH3} -YFP are shown for cells grown in media containing glucose (X-axis) or glycerol (Y-axis). Expression is shown as % of wild type. (F) Effects of 235 *cis*-regulatory mutations on expression of P_{TDH3} -YFP are shown for cells grown in media containing glucose (X-axis) or galactose (Y-axis, right panel). Panels E and F reproduced from Figures 3F and 3E in [54], respectively.

Figure 5. Inferring effects of selection on *cis*- and *trans*-regulatory variation.

(A) Distributions of effects are shown for mutations (red) and polymorphisms (black) in the *S. cerevisiae* *TDH3* promoter affecting expression level of the P_{TDH3} -YFP reporter gene. Modified from Figure 3A in [30]. (B) Distributions of effects are shown for mutations (red) and polymorphisms (black) in the *S. cerevisiae* *TDH3* promoter affecting expression noise of the P_{TDH3} -YFP reporter

gene. Modified from Figure 3B in [30]. (C) Distributions of effects are shown for mutations (red) and polymorphisms (black) in the *S. cerevisiae* *TDH3* promoter affecting plasticity of the *P_{TDH3}-YFP* reporter gene in cells grown on media containing glucose or galactose. Modified from Figure 4F in [54]. In panels A, B, and C, effects are shown on the X-axis as percent relative to the un-mutagenized wild type strain. (D) Distribution of effects for individual *trans*-regulatory mutations inferred from EMS mutants impacting expression of *P_{TDH3}-YFP*. Reproduced from Figure 2A in [58]. (E) Changes in expression observed in each of 10 replicate simulations of neutral evolution each produced by drawing 30,000 mutations randomly from the distribution shown in (D). Each color represents a different simulation run. Reproduced from Figure 2B in [58]. (F) Shades of blue show the 95th, 90th, 80th, 70th, and 60th percentiles from light to dark for the range of expression levels predicted by neutral evolution in 10,000 replicate simulations. Green points show expression levels (Y-axis) observed among 56 strains of *S. cerevisiae* with the number of genetic differences from the BY lab strain shown on the X-axis. Reproduced from Figure 2C in [58]. (G) Maps of eQTL affecting expression of *P_{TDH3}-YFP* in three wild strains of *S. cerevisiae* (SK1, green; YPS1000, blue; M22, brown) relative to the BY lab strain. Chromosome numbers (I to XVI) shown along the top and position in Mb (1 to 12) shown along the bottom. The test statistic G' is plotted on the Y-axis, which captures deviations in allele frequency from the null model between the high and low expression pools of cells used for bulk-segregant mapping. Reproduced from Figure 3B in [58]. (H) eQTL increasing and decreasing expression relative to BY were observed at similar frequencies. Reproduced from Figure 3D in [58].

Figure 6. Molecular mechanisms of *cis*- and *trans*-regulatory mutations

(A) Effects of 236 individual point mutations in the *TDH3* promoter on expression of *P_{TDH3}-YFP* in rich media containing glucose are shown. Red bars show statistically significant changes in expression level relative to wildtype. Schematic of the *TDH3* promoter shows locations of RAP1p and GCR1p binding sites (highlighted in gray), in which mutations had the largest effects. Modified from Figures 2A and 2B in [30]. (B) Effects of *cis*-regulatory mutations causing 0%, 20%, 50%, 85%, and 135% of wildtype (WT) *TDH3* expression levels (X-axis) on expression of 140 genes identified as downstream of *TDH3* (Y-axis) are shown. Reproduced from Figure 3B in [43]. (C) Pleiotropy, measured as the number of differentially expressed (“DE”) genes (Y-axis), is shown for *cis*-regulatory (red) and *trans*-regulatory (blue) mutations affecting expression of *TDH3* (X-axis). Reproduced from Figure 2E from [43]. (D) Expression of *GPD2*, a gene downstream of *TDH3*, is shown (Y-axis) in *cis*-regulatory (red) and *trans*-regulatory (blue) mutants affecting *TDH3* expression (X-axis). Effects of *trans*-regulatory mutations outside of the effects predicted by a *cis*-regulatory mutation with similar effects on *TDH3* expression (95% prediction interval shown with outer red lines, 95% confidence interval of regression line in gray) are defined as the pleiotropic effects of *trans*-regulatory mutations on gene expression (dotted line). Reproduced from Figure 3G in [43]. (E) Schematic showing that *trans*-regulatory mutations can impact expression of genes in parallel to and downstream of *TDH3* independent of their effects on expression of the focal gene. Reproduced from Figure 3F in [43]. (F) Network shows transcription factors predicted in YEASTRACT [18] to regulate expression of *TDH3*. Green lines indicate activators; red lines indicate repressors; black lines indicate interactions with unknown directionality. Five mutations impacting *TDH3* expression (yellow stars) were found in these regulators among 66 mutations isolated from EMS mutants, which was more than expected by chance. Modified from Figure 4 in [35].

Table 1. Summary of observations and inferences from studies of *cis*- and *trans*-regulatory mutations affecting expression of TDH3

	Observation	Molecular Explanation	Possible Evolutionary Implications
Frequency	trans mutations arise more often than cis	trans mutations reside in coding or noncoding sequences of direct and indirect regulators; cis mutations tend to be near the focal gene	<i>trans</i> -regulatory mutations should be the predominant source of new regulatory variation
Expression Level	cis mutations tend to have larger effects than trans	cis sequences directly regulate the focal gene's expression; trans mutations have greater opportunity to be buffered by other factors	based on focal gene impact alone, cis mutations should have larger fitness effects (positive or negative) than trans mutations
Expression Noise	mean and noise are more highly correlated for cis than trans mutations	cis sequences alter size or frequency transcriptional bursts; trans factors tend to impact expression noise due to cell-to-cell variation in the trans factor	trans mutations should provide a more effective substrate for independently optimizing expression level and expression noise
Dominance	trans mutations are recessive more often than cis	cis sequences control expression of the linked coding sequence; because trans factors are diffusible molecules, gene products from a wildtype allele can mask effects of a mutant allele	cis mutations should be subject to selection as soon as they arise, but trans mutations may segregate neutrally in heterozygotes
Pleiotropy	trans mutations usually affect expression of more genes than cis	cis mutations affect expression of the focal gene and downstream genes; trans mutations affect the focal gene, downstream genes, and other genes	trans mutations should tend to have more negative impacts on fitness than cis mutations
Fitness	trans mutations more often deleterious than cis	trans mutations changing expression of more genes than cis mutations impact reproduction more than the larger effects of cis mutations on the focal gene	trans mutations should be eliminated by natural selection more often than cis mutations
Environment	cis and trans mutations can have environment-specific effects	cis sequences include TF binding sites used in one environment but not others; trans mutations can affect environment-specific regulators	cis and trans mutations can have different likelihoods of fixation in different environments











