Contrasting Frequencies and Effects of cis- and trans-Regulatory Mutations Affecting Gene Expression

Brian P. H. Metzger,¹,¹ Fabien Duveau,¹,¹ David C. Yuan,¹,²,³ Stephen Tryban,¹ Bing Yang,² and Patricia J. Wittkopp*¹,²

¹Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor
²Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor
³Department of Biology, Stanford University

*These authors contributed equally to this work.

Corresponding author: E-mail: wittkopp@umich.edu.

Associate editor: Juliette de Meaux

Abstract

Heritable differences in gene expression are caused by mutations in DNA sequences encoding cis-regulatory elements and trans-regulatory factors. These two classes of regulatory change differ in their relative contributions to expression differences in natural populations because of the combined effects of mutation and natural selection. Here, we investigate how new mutations create the regulatory variation upon which natural selection acts by quantifying the frequencies and effects of hundreds of new cis- and trans-acting mutations altering activity of the TDH3 promoter in the yeast Saccharomyces cerevisiae in the absence of natural selection. We find that cis-regulatory mutations have larger effects on expression than trans-regulatory mutations and that while trans-regulatory mutations are more common overall, cis- and trans-regulatory changes in expression are equally abundant when only the largest changes in expression are considered. In addition, we find that cis-regulatory mutations are skewed toward decreased expression while trans-regulatory mutations are skewed toward increased expression. We also measure the effects of cis- and trans-regulatory mutations on the variability in gene expression among genetically identical cells, a property of gene expression known as expression noise, finding that trans-regulatory mutations are much more likely to decrease expression noise than cis-regulatory mutations. Because new mutations are the raw material upon which natural selection acts, these differences in the frequencies and effects of cis- and trans-regulatory mutations should be considered in models of regulatory evolution.

Key words: expression noise, distribution of mutational effects, neutral evolution, mutational target size, TDH3.

Introduction

Variation in gene expression is a common source of phenotypic diversity within and between species (Zheng et al. 2011). Much of this variation is heritable, arising from mutations in DNA sequences encoding cis-regulatory elements (e.g., promoters and enhancers) and trans-regulatory factors (e.g., transcription factors, noncoding RNAs, and signaling molecules) (Carroll 2008; Stern and Orgogozo 2008). Studies investigating the genetic basis of intra- and interspecific expression differences have shown that both cis- and trans-acting changes contribute to differences in gene expression, but the contributions of cis- and trans-acting loci are rarely equal (Yvert et al. 2003; Gibson and Weir 2005; Rockman and Kruglyak 2006; Gilad et al. 2008; Tiros et al. 2009; Goncalves et al. 2012; Schaeke et al. 2013; Coolon et al. 2014). In particular, identification of the specific genetic loci responsible for variation in gene expression has revealed differences in both the relative frequency and average effects of cis- and trans-acting loci (Gibson and Weir 2005; Rockman and Kruglyak 2006; Gilad et al. 2008).

Differences in the frequency and effects of cis- and trans-regulatory alleles result from the combined action of mutation and natural selection, with new mutations generating the genetic variation upon which natural selection acts. Mutation accumulation experiments, in which mutations are allowed to accrue in the near absence of selection, have shown how gene expression levels change on a genomic scale in response to new mutations (Denver et al. 2005; Rifkin et al. 2005; Landry et al. 2007; Simola et al. 2010; McGuigan et al. 2014; Hodgins-Davis et al. 2015), but many questions remain about the impact of new mutations on cis- and trans-regulation of a particular gene. For example, what are the relative frequencies of such cis- and trans-regulatory mutations? Do cis- and trans-regulatory mutations cause similar changes in a gene’s expression? Are they equally likely to increase or decrease expression? Empirical answers to these questions are needed to develop realistic neutral models of regulatory evolution that can be used to infer the impact of selection on regulatory variation observed in natural populations.

For any specific gene, the mutational target size for trans-regulatory changes is expected to be larger than the mutational target size for cis-regulatory changes because cis-regulatory mutations are typically limited to sequences close to the gene, whereas trans-regulatory mutations can...
be located anywhere in the genome. However, not all mutations in these potential cis- and trans-regulatory sequences will affect expression of the focal gene. The subsets of mutations that actually do affect expression of the focal gene define realized target sizes for cis- and trans-regulatory mutations (Wittkopp 2005; Lang and Murray 2008), and it is these realized target sizes that are important for determining the relative contributions of cis- and trans-regulatory mutations to expression variation. Prior studies suggest that the realized target size for trans-regulatory mutations remains larger than the realized target size for cis-regulatory mutations (Denver et al. 2005; Landry et al. 2007; Gruber et al. 2012); however, the magnitude of this difference remains uncertain.

In addition to differences in frequency, differences in the distributions of effects on gene expression for cis- and trans-regulatory mutations are also important for determining the relative contributions of cis- and trans-regulatory changes to expression variation. These differences include how much mutations alter gene expression level (magnitude of effects), and whether mutations increase or decrease expression (direction of effects). For example, studies of quantitative trait loci affecting gene expression (eQTL) have shown that cis-acting eQTLs tend to have larger magnitudes of effect on gene expression than trans-acting eQTL (Schadt et al. 2003). This pattern could arise if new cis-regulatory mutations have, on average, larger effects on expression than new trans-regulatory mutations. Differences in the direction of effects for cis- and trans-regulatory mutations are harder to predict and more likely to vary among genes, but could result from a difference in the relative frequency of activators and repressors between direct trans-regulatory factors (those that bind a gene’s cis-regulatory sequences) and indirect trans-regulatory factors (those that regulate a gene’s expression by altering the abundance or activity of direct regulators).

To address these questions, we compared the frequency and effects of cis- and trans-regulatory mutations altering expression of a focal gene using a reporter gene previously constructed for studying the effects of cis-regulatory sequences in the baker’s yeast Saccharomyces cerevisiae. We isolated 1,485 new mutants with the potential to affect expression of the reporter gene (Metzger et al. 2015). We also collected and analyzed 202 mutants enriched for rare trans-regulatory mutations with large effects on reporter gene expression. Finally, because prior work suggests that heritable variation affecting the variability in gene expression among genetically identical cells (i.e., expression noise) exists within natural populations and can be subject to selection (Zhang et al. 2009; Wang and Zhang 2011; Metzger et al. 2015), we also measured the effects of each of these 1,922 mutants on expression noise. We found that cis- and trans-regulatory mutations altering reporter gene expression differed in their frequency, magnitude of effect, and direction of effect for both average expression level and expression noise. These differences in the frequencies and effects of cis- and trans-regulatory mutations should be considered in null models of regulatory evolution.

Results
Effects of cis- and trans-Regulatory Mutations on Gene Expression Level
To quantify the effects of new cis- and trans-regulatory mutations on gene expression, we used a reporter gene (PTDH3-YFP) containing the S. cerevisiae TDH3 promoter (PTDH3) and yellow fluorescent protein (YFP) coding sequence incorporated into the S. cerevisiae genome (Gruber et al. 2012). Effects of cis-regulatory mutations were determined by reanalyzing a published data set describing the effects of 235 individual point mutations in the 678-bp TDH3 promoter on YFP fluorescence (Metzger et al. 2015). As described in Metzger et al. (2015), these 235 genotypes were constructed by using site-directed mutagenesis to systematically mutate Gs to As and Cs to Ts within the TDH3 promoter. Overall, there was no significant difference in the number of cis-regulatory mutations that increased (n = 105) or decreased (n = 130) expression (P = 0.12, binomial test; fig. 1A). However, cis-regulatory mutations that decreased expression had significantly larger effects on expression (measured as percent change in expression level) than cis-regulatory mutations that increased expression (P = 0.0001, Wilcoxon, fig. 1B). This difference resulted in a significant skew in the distribution of cis-regulatory effects toward decreased expression (skewness = –7.9, P < 1 × 10^{-6}, Bootstrap). Mutations in known transcription factor binding sites (TFBS) for RAP1 and GCR1 (supplementary fig. S1, Supplementary Material online) contributed to this skew: 16 of the 18 mutations in previously identified TFBS decreased expression, including all mutations that decreased expression more than 6% (fig. 1A). After excluding all mutations in known TFBS, cis-regulatory mutations that decreased expression still had significantly larger effects on gene expression than cis-regulatory mutations that increased expression (P = 0.008, Wilcoxon, supplementary fig. S2, Supplementary Material online), indicating that the skew toward decreased expression is a general property of cis-regulatory mutations affecting activity of the TDH3 promoter (skewness = –1.3, P = 2 × 10^{-5}, Bootstrap). The relative effects of cis-regulatory mutations were robust to changes in the reporter gene insertion site, genetic background, and even fusion of YFP to the TDH3 coding sequence (supplementary fig. S3, Supplementary Material online).

To determine the effects of trans-regulatory mutations on PTDH3-YFP expression, we generated mutations throughout the genome using a low dose of ethyl methanesulfonate (EMS) that introduced ~32 (21–43, 95% percentiles) mutations per cell. We then randomly collected and analyzed 1,485 mutant cells from the EMS-treated population, irrespective of their effects on fluorescence. Because the potential target size for trans-regulatory mutations is ~18,000 times larger than the potential target size for cis-regulatory mutations (12.1 Mb S. cerevisiae genome vs. 678 bp TDH3 promoter), we expect that cis-regulatory mutations make a negligible contribution to changes in expression among these mutants and...
attributed all effects to trans-regulatory mutations. This is a conservative assumption when testing for significant differences between the effects of cis- and trans-regulatory mutations. For each mutant isolated, we used flow cytometry to measure YFP fluorescence in four biological replicate populations containing ~5,000 cells each and compared this measure of gene expression with YFP fluorescence of non-EMS-treated controls. This procedure is similar to that used for the cis-regulatory mutants described above and both data sets were analyzed using identical methods (Metzger et al. 2015). We found no significant difference in the number of trans-regulatory mutants that increased \((n = 747)\) or decreased \((n = 738)\) \(P_{TDH3}-YFP\) expression \((P = 0.84, \text{binomial test; fig. 1C})\), nor any significant difference in the magnitude of effects between trans-regulatory mutants that increased or decreased expression \((P = 0.68, \text{Wilcoxon, fig. 1D})\).

To determine the relative effects of cis- and trans-regulatory mutations on YFP fluorescence, we compared the relative effect sizes of the 235 cis-regulatory mutants and the 1,485 trans-regulatory mutants by examining the absolute value of effects on gene expression. We found that the cis-regulatory mutants had significantly larger effects on gene expression than trans-regulatory mutants \((P = \frac{3}{10^{11}}, \text{Wilcoxon, fig. 2A})\). This result was robust to removing the large effects of cis-regulatory mutants with mutations in the known TFBS \((P = 1 \times 10^{-7}, \text{Wilcoxon, supplementary fig. S28, Supplementary Material online})\). Interestingly, for mutations that increased expression, the effects of cis-regulatory
Mutants were only marginally larger than the effects of trans-regulatory mutants \( (P = 0.013, \text{Wilcoxon, fig. 2B}) \) and the larger overall effect of cis-regulatory mutants than trans-regulatory mutants was caused primarily by the significantly larger effects of cis-regulatory mutants than trans-regulatory mutants for mutants that decreased expression \( (P = 4 \times 10^{-11}, \text{Wilcoxon, fig. 2C}) \). These results were also robust to removal of the effects of cis-regulatory mutations in known TFBS (supplementary fig. S2C and D).

Differences in the percentage of cis- and trans-regulatory mutations sampled in our collections might affect our conclusions: 34.6\% of the 678 sites in the TDH3 promoter were mutated in our collection of cis-regulatory mutants, but only \(~39\%\) of sites \((32 \text{ mutations/mutant} \times 1,485 \text{ mutants} = 47,520 \text{ mutations})\) in the \(~12.1\text{ Mb genome}\) were mutated in our collection of trans-regulatory mutants. It is therefore possible that increased sampling of trans-regulatory mutations might recover more large-effect mutations, which would decrease the difference in effect size observed between cis- and trans-regulatory mutations. To better characterize trans-regulatory mutations with large effects on expression, we used fluorescence-activated cell sorting (FACS) to isolate 103 mutants enriched for increased YFP fluorescence and 99 mutants enriched for decreased YFP fluorescence levels from a new EMS-treated population. When each of these 202 mutants was grown clonally and reanalyzed in four replicate populations, the average absolute effect on \( P_{TDH3} \) expression was significantly higher than the average absolute effect of the trans-regulatory mutants chosen randomly from the population \( (P = 1 \times 10^{-5}, \text{Wilcoxon, fig. 2B and C}) \), consistent with the selection of these mutants from the tails of the fluorescence distribution. Trans-regulatory mutants enriched for large increases in expression had an average effect on reporter gene expression that was similar to that seen for the cis-regulatory mutants \( (P = 0.91, \text{Wilcoxon}) \), whereas the average effect of trans-regulatory mutants enriched for large decreases in expression remained significantly smaller than the average effect of the cis-regulatory mutants \( (P = 0.035, \text{Wilcoxon, fig. 2B and C}) \). Twelve of the 202 mutants enriched for large effect mutations showed changes in \( P_{TDH3} \) expression greater than 7.5\%, 11 of which increased expression (fig. 1E and supplementary fig. S4, Supplementary
Frequencies and Effects of cis- and trans-Regulatory Mutations

To determine whether these large increases in expression were caused by duplications of the reporter gene, we used pyrosequencing to compare copy number of the P_{TDH3}-YFP reporter gene in each of these mutants with that of the ancestral wild-type strain. We found no evidence of increased copy number in the mutants, indicating that the large increases in expression were not caused by duplication of the P_{TDH3}-YFP construct. Taken together, these data indicate that the effects of trans-regulatory mutants are significantly skewed toward increased expression whether or not the large effects trans-regulatory mutants are (skewness = 3.0, P = 0.0007, Bootstrap) or are not (skewness = 2.3, P = 0.047, Bootstrap) included. This skew toward increased expression for trans-regulatory mutants affecting TDH3 expression is consistent with earlier reports (Gruber et al. 2012) and contrasts with the skew of cis-regulatory mutations toward decreased expression (fig. 1A and B).

Frequency of cis- and trans-Regulatory Mutations Affecting Gene Expression Level

To estimate differences in the frequency and target size of cis- and trans-regulatory mutations, we first used t-tests to compare expression between each mutant and a nonmutagenized control genotype. We found that 106 of 235 (45%) potential cis-regulatory mutants and 463 of 1,485 (31%) potential trans-regulatory mutants had significant changes in P_{TDH3}-YFP expression at a threshold of P = 0.05. When a more conservative significance threshold of P = 0.01 was used, 77 of 235 (33%) of cis-regulatory mutants and 232 of 1,485 (16%) of trans-regulatory mutants had significant changes in reporter gene expression. For cis-regulatory mutations, these data suggest a realized target size of 306 bp when using the P = 0.05 threshold or 222 bp when using the P = 0.01 threshold. For trans-regulatory mutants, if we assume that 1) genotypes with wild-type expression do not harbor mutations affecting P_{TDH3}-YFP expression (i.e., no compensatory mutations); 2) the number of mutations per cell significantly affecting P_{TDH3}-YFP expression follows a Poisson distribution (Gruber et al. 2012); 3) epistatic effects on P_{TDH3}-YFP expression are rare among the mutations, and 4) if each mutant has an average 32 mutations, then these data suggest a realized target size of ~118,000 bp (86,000–179,000 bp, 95% percentiles) when using a significance threshold of P = 0.05 and 59,000 bp (43,000–90,000 bp, 95% percentiles) when using a significance threshold of P = 0.01. This represents 0.97% (0.73–1.48%, 95% percentiles) of the S. cerevisiae genome when using the P = 0.05 threshold and 0.49% (0.36–0.74%, 95% percentiles) of the genome when using the P = 0.01 threshold. In other words, using statistical significance to define functional cis- and trans-regulatory mutations, our data suggest that functional trans-regulatory mutations should arise 265 times (based on P = 0.01 threshold) to 362 times (based on P = 0.05 threshold) more often than functional cis-regulatory mutations.

The impact of a mutation on gene expression is expected to be related to its impact on fitness (Rest et al. 2013), thus we also examined how estimates of the realized target size for cis- and trans-regulatory mutations changed when different minimum effect sizes were used to define functional cis and trans-regulatory mutations. To do this, we calculated the number of mutations with effects on P_{TDH3}-YFP expression equal to or larger than a particular value for the range of effects observed. We did this separately for the cis-regulatory mutants, the original collection of trans-regulatory mutants, and the trans-regulatory mutants enriched for mutations of large effect, taking into account that each trans-regulatory mutant has ~32 mutations and that the second set of trans-regulatory mutations was artificially enriched for mutations of large effect (see Materials and Methods). We found that the realized target size for both cis- and trans-regulatory mutations dropped rapidly as the minimum effect size cutoff increased, and that the choice of the specific cutoff used had a drastic effect on the estimated target sizes. In particular, we found that while trans-regulatory mutations remained more common than cis-regulatory mutations for nearly all magnitudes and directions of effect, the relative frequencies of cis- and trans-regulatory mutations varied considerably over the range of effect sizes examined (fig. 2D). For example, trans-regulatory mutations were inferred to be ~10,000 times more common than cis-regulatory mutations among mutations that alter expression less than 1%, but only 10 times more common among mutations that alter expression more than 2.5%. The relative frequency of cis- and trans-regulatory mutations also depended strongly on the direction of effect. For example, although over 100 trans-regulatory mutations that increase expression more than 2.5% are predicted to exist, cis-regulatory mutations resulting in more than a 2.5% increase in P_{TDH3}-YFP expression appear to be either incredibly rare or nonexistent. In contrast, decreases in expression by more than ~7.5% were more likely caused by cis-regulatory mutations than trans-regulatory mutations. These relationships were robust to error in the estimated number of mutations in each trans-regulatory mutant (supplementary fig. S5, Supplementary Material online).

Frequency and Effects of cis- and trans-Regulatory Mutations on Gene Expression Noise

In addition to the effects on the average level of expression, mutations can also alter gene expression noise. To compare the effects of new cis- and trans-regulatory mutations on gene expression noise, we calculated the coefficient of variation (σ/μ) in YFP fluorescence for each cis- and trans-regulatory mutant. For the 235 cis-regulatory mutants, we found that significantly more mutants showed increased expression noise (n = 208) than decreased expression noise (n = 27, P = 8 × 10^{-36}, binomial test) (fig. 3A), consistent with a prior analysis of this data (Metzger et al. 2015). This difference remained after excluding cis-regulatory mutations in the known TFBS (increase = 190, decrease = 27, P = 2 × 10^{-31}, binomial test). We also found that cis-regulatory mutants with increased expression noise had larger effects than cis-regulatory mutants that decreased expression noise, regardless of whether mutations in known TFBS were included (P = 2 × 10^{-5}, Wilcoxon; fig. 3B) or not (P = 5 × 10^{-5}, Wilcoxon; supplementary fig. S6, Supplementary Material online), resulting in a significant skew toward increased gene expression noise.
expression noise for mutations in the TDH3 promoter (all mutations: skewness = 11.0, $P < 1 \times 10^{-6}$, Bootstrap; mutations in known TFBS excluded: skewness = 0.74, $P = 0.001$). In contrast, the randomly collected set of 1,485 trans-regulatory mutants contained significantly more mutants with decreased expression noise ($n = 797$) than increased expression noise ($n = 688$, $P = 0.005$, binomial test; fig. 1C). There was not a significant difference in the effects of trans-regulatory mutants that increased or decreased expression noise ($P = 0.71$, Wilcoxon; fig. 3D). Unlike the effects of mutations on average expression levels, cis-regulatory mutants had smaller effects on gene expression noise than trans-regulatory mutants ($P = 6 \times 10^{-8}$, Wilcoxon; fig. 4A). This was true for both mutations that increased expression noise ($P = 2 \times 10^{-4}$, Wilcoxon; fig. 4B) and decreased expression noise ($P = 9 \times 10^{-8}$, Wilcoxon; fig. 4C).

Using t-tests to compare expression noise between each mutant and a nonmutant control genotype showed that 74 of 235 (31%) potential cis-regulatory mutants and 293 of 1,485 (20%) potential trans-regulatory mutants had significant changes in $P_{\text{TDH3-YFP}}$ expression noise at a significance threshold of $P = 0.05$. Forty of 235 (17%) of cis-regulatory mutants and 110 of 1,485 (7%) of trans-regulatory mutants showed a significant change in expression noise at a significance threshold of $P = 0.01$. From these values, we estimate a realized target size for cis-regulatory mutations affecting gene expression noise of 213 bp when a $P = 0.05$ threshold is used and 115 bp when a $P = 0.01$ threshold is used and a realized target size for trans-regulatory mutations affecting gene expression noise of 74,000 bp (55,000–113,000 bp, 95% percentiles) when a $P = 0.05$ threshold is used and 28,000 bp (21,000–43,000 bp, 95% percentiles) when a $P = 0.01$ threshold is used. For both cis- and trans-regulatory mutations, the realized target sizes predicted for expression noise are smaller than the realized target sizes predicted for average expression using a statistical cutoff based on $P$ values; however, differences in statistical power to detect significant changes in gene expression noise and average expression level make it difficult to meaningfully compare these estimates.

To compare the frequency of cis- and trans-regulatory mutations affecting TDH3 expression noise without relying on statistical cutoffs, we used the same approach described above for average expression level to calculate the realized target size for cis and trans-regulatory mutations affecting gene expression noise over a range of effect sizes. We found that for all magnitudes and directions of effects, trans-regulatory mutations affecting expression noise were more common than cis-regulatory mutations affecting expression noise (fig. 4D). However, as with the average level of expression, the relative frequency of cis and trans-regulatory mutations varied considerably. For example, among mutations that increased expression noise, trans-regulatory mutations were about 10–100 times more common than cis-regulatory mutations for most effect sizes; however, for even moderate decreases in
gene expression noise, trans-regulatory mutations were 1,000–10,000 times more frequent than cis-regulatory mutations.

Relationship between Mutational Effects on Average Expression and Expression Noise

New regulatory mutations can alter both the average level of expression and expression noise simultaneously. If the effects of new mutations on average expression level are independent of the effects of mutations on expression noise, however, natural selection can act independently on these two traits. Previous work indicates that there is often a negative correlation between average expression and expression noise across genes (Bar-Even et al. 2006), as well as for individual cis-regulatory mutations (Hornung et al. 2012; Sharon et al. 2014). However, individual cis-regulatory mutations can independently alter average expression and expression noise (Hornung et al. 2012; Sharon et al. 2014; Metzger et al. 2015) and little is known about the relationship between average expression and expression noise for trans-regulatory mutations.

To compare the relationship between effects on average expression level and expression noise for new cis- and trans-regulatory mutations, we used principal components analysis (PCA) to determine the primary axes of variation for the cis- and trans-regulatory mutants separately (fig. 5). Because PCA analysis can be sensitive to extreme outliers, we removed the effects of cis-regulatory mutations in the known TFBS. We found that for cis-regulatory mutants the primary axis of variation had a significant negative slope, indicating a negative correlation between average expression and expression noise (angle of rotation $= 106^\circ$, 99% confidence interval [CI] 100–111$^\circ$, $P < 1 \times 10^{-5}$, Bootstrap). In contrast, for trans-regulatory mutants, the primary axis of variation had a slightly positive slope (angle of rotation $= 91^\circ$, 99% CI 89–93$^\circ$) and this slope was not significantly different from the 90$^\circ$ expected if trans-regulatory mutations had independent effects on average expression level and expression noise ($P = 0.09$, Bootstrap). The angle of rotation was significantly different for cis- and trans-regulatory mutations ($P < 10^{-6}$, permutation tests), indicating that cis- and trans-regulatory mutations can have different relationships between their effects on average expression noise.
expression level and expression noise. These results were robust to inclusion of the cis-regulatory mutations in the known TFBS (supplementary fig. S7A, Supplementary Material online, angle of rotation = 98°, 99% CI 97–108°; difference in angle between cis and trans-regulatory mutants, \( P = 0.041 \)). To better understand how cis- and trans-regulatory mutations might differ in their effects on the relationship between average expression and expression noise, we repeated this analysis using the standard deviation rather than the coefficient of variation (standard deviation/mean). We found that removing the scaling of standard deviation by mean had no effect on the relationship between average expression and expression noise for trans-regulatory mutants (angle of rotation = 90°, 99% CI 89–92°, \( P = 0.52 \), Bootstrap), but reduced the strength of the negative correlation seen for cis-regulatory mutants (angle of rotation = 91°, 99% CI 89–95°, \( P = 0.046 \), Bootstrap) such that there was no longer a statistically significant difference between cis- and trans-regulatory mutations (\( P = 0.11 \), permutation test; supplementary fig. S7B, Supplementary Material online). The different impact on cis- and trans-regulatory mutations of changing the measure of expression noise used suggests that cis- and trans-regulatory mutations might tend to alter expression noise in different ways.

**Discussion**

This study reveals many differences between cis- and trans-regulatory mutations that can impact their likelihood of contributing to variation in gene expression within and between species. For example, we find that cis- and trans-regulatory mutations differ significantly in their effects on both TDH3 average expression level and expression noise, with cis-regulatory mutations skewed toward decreased expression and increased expression noise, while trans-regulatory mutations are skewed toward increased expression and decreased expression noise. The relative frequencies of cis- and trans-regulatory mutations also differ, but the difference in frequencies depends upon the effect size used to define functional cis- and trans-regulatory mutations. For example, if only the largest changes in activity of the TDH3 promoter are considered, then the target size for both cis- and trans-regulatory mutations is in the dozens or hundreds of bases, with trans-regulatory mutations more frequent for increases in expression and cis-regulatory mutations more frequent for decreases in expression. In contrast, if both small and large changes in \( P_{TDH3} \) activity are considered, then the trans-regulatory target size for altering \( P_{TDH3} \) activity is orders of magnitude larger than the cis-regulatory target size, and most new mutations resulting in biologically meaningful changes in TDH3 transcription should be trans-regulatory. To the best of our knowledge, these data provide the first systematic comparison of cis- and trans-regulatory mutations affecting expression of a focal gene in any eukaryote.

**Consequences of Mutation Type on Mutational Distributions**

All the cis-regulatory mutations examined in this study were G→A and C→T transitions. The majority of trans-regulatory mutations examined were expected to be G→A and C→T transitions as well because EMS introduces these types of changes almost exclusively (Flibotte et al. 2010; Duveau et al. 2014). G→A and C→T transitions are also the most common spontaneous point mutations in *S. cerevisiae* (Zhu et al. 2014) and the most common polymorphisms segregating among natural populations of *S. cerevisiae* (Maclean CJ, Metzger BPH, Yang JR, Ho WC, Moyers B, Zhang J, in preparation); however, they are still a minority of all point mutations, comprising ~35% of all spontaneous point mutations (Zhu et al. 2014) and ~37% of single nucleotide polymorphisms segregating in *S. cerevisiae* (Maclean CJ, Metzger BPH, Yang JR, Ho WC, Moyers B, Zhang J, in preparation). So how representative are the effects of G→A and C→T transitions of all point mutations? And how might the effects of point mutations differ from other types of mutations such as insertions or deletions (indels), rearrangements, or copy number variants (CNVs)?

For cis-regulatory sequences, prior work suggests that there are no systematic differences in the effects of different classes of point mutations on gene expression (Patwardhan et al. 2009, 2012; Kwansieski and Mogno 2012; Melnikov et al. 2012; Metzger et al. 2015). Nevertheless, we might have
overestimated the average effect of cis-regulatory mutations in the TDH3 promoter because mutations in known TFBS, which had the largest effects on expression, were overrepresented in our data set (4% of total sequence, but 8% of mutations) due to their higher GC content (18/27 bp, 67%) relative to the rest of the promoter (35%). cis-regulatory mutations had larger effects than trans-regulatory mutations even after excluding mutations in the TFBS, however, suggesting that this overrepresentation does not alter our conclusions. Another consequence of mutating only Gs and Cs is that we failed to mutate functional elements composed of only As and Ts. The canonical TATA box contained within the TDH3 promoter is one clear example of this. This sequence is a key determinant of both average expression level and gene expression noise (Raser and O’Shea 2004; Hornung et al. 2012), suggesting that mutations within it would have had large effects that further increased the magnitude of cis-regulatory mutations we observed.

The impact of using EMS to introduce trans-regulatory mutations differs for coding and noncoding sequences of trans-acting factors. For mutations in noncoding cis-regulatory sequences of trans-acting factors, the consequences of using EMS are expected to be similar to the consequences of mutating only Gs and Cs in the TDH3 promoter: Potentially large effect mutations in functional elements that are AT rich such as TATA boxes might be missed and GC-rich binding sites might be overrepresented, but the distribution of mutational effects should otherwise be unbiased. In contrast, in coding regions, which make up to ~73% of the S. cerevisiae genome (Alexander et al. 2010), using EMS to introduce mutations is expected to result in a biased sampling of amino acid changes because of the genetic code and codon usage (supplementary fig. S8, Supplementary Material online). G→A and C→T transitions are the most common type of spontaneous mutation, thus this bias should be toward the same types of amino acid changes caused most often by spontaneous mutations; however, some amino acids (Alanine, Glycine, and Proline) cannot be created by these types of mutations, whereas other amino acids (Asparagine, Isoleucine, Lysine, and Phenylalanine) and stop codons cannot be mutated. Some amino acid changes are more likely to disrupt protein function than others, resulting in differences in their magnitude of mutational effects (Yampolsky and Stoltzfus 2005). EMS-induced mutations and spontaneous mutations are also expected to differ in the relative frequency of synonymous and nonsynonymous mutations: ~23% of naturally occurring point mutations in coding regions are expected to be synonymous, compared with 31% for EMS-induced mutations (see Materials and Methods). Because synonymous mutations typically have smaller effects than nonsynonymous mutations, the use of EMS might underestimate the effects of new trans-regulatory mutations in coding regions. To determine the full consequence of these biases in the distribution of mutational effects, the impact of other types of point mutations should be examined in future work.

The effects of indels, large-scale genome rearrangements, and CNVs must also be considered to fully describe a mutational distribution. Point mutations are more common than these other types of mutations, but indels, rearrangements, and CNVs are known to contribute to variable gene expression in natural yeast populations (Gerstein et al. 2014; Hose et al. 2015) and experimentally evolved yeast populations (Dunham et al. 2002; Kao and Sherlock 2008; Payen et al. 2014; Sunshine et al. 2015). For example, a prior study characterizing trans-regulatory mutations affecting activity of the TDH3 promoter identified 22 spontaneous CNVs, all of which included duplication of the reporter gene (Gruber et al. 2012). In most cases, expression was increased substantially, suggesting that CNVs have distinct effects on expression from cis- and trans-regulatory changes caused by point mutations (Gruber et al. 2012). The effects on gene expression for other types of CNVs (including deletions and aneuploidies of other chromosomes) are less clear, but may often be larger than for point mutations (Sunshine et al. 2015). Indels are also expected to have large effects when they occur in trans-acting coding regions because many will alter the reading frame and create nonfunctional proteins. Together, these observations suggest that the mutational distributions we measured may be missing rare, large, effect mutations. Ultimately, the impact of indels, rearrangements, and CNVs on the distribution of mutational effects will need to be determined empirically.

Generality of the Gene Studied

In eukaryotes, transcription of each gene is controlled by biochemical interactions among many trans-acting factors that culminate in the direct binding of specific transcription factors to cis-regulatory sequences (Wray et al. 2003). Our observation that trans-regulatory mutations are more common overall than cis-regulatory mutations is thus likely to be true for most genes, as is the observation that cis-acting mutations tend to have larger effects than trans-acting mutations (Schadt et al. 2003). Other properties we report, however, such as the relative frequency of mutations that increase or decrease gene expression level or expression noise or the relative frequency of cis- and trans-regulatory mutations with particular effect sizes are expected to vary among genes.

TDH3 encodes a glyceraldehyde-3-phosphate dehydrogenase that is involved in both glycolysis (McAlister and Holland 1985) and chromatin remodeling (Ringel et al. 2013). TDH3 expression is regulated (at least in part) by binding sites for the RAP1 and GCR1 transcription factors in its promoter (Baker et al. 1992; Yagi et al. 1994). Most mutations in these TFBS caused TDH3 promoter activity to decrease, consistent with RAP1 and GCR1 activating TDH3 expression. Other genes involved in glycolysis are also regulated by RAP1 and GCR1 (Chambers et al. 1995; Uemura and Fraenkel 2000; Lieb et al. 2001) and these genes potentially have similar distributions of effects for cis-regulatory mutations. Mutations in TFBS appear to often have the largest effects on gene expression (Patwardhan et al. 2012), suggesting that the density of TFBS within a promoter will strongly influence the distribution of effects for its cis-regulatory mutations. Such mutations are not expected to always decrease expression, however; loss-of-function mutations in TFBS for repressors should
increase promoter activity and can thus potentially skew the effects of cis-regulatory mutations toward increased expression. Outside of the known TFBSs, we found that cis-regulatory mutations tend to have small effects that were equally likely to increase or decrease expression. It remains to be seen if these mutations are disrupting unidentified binding sites for activators and repressors or simply altering chromatin structure more generally with an effect on expression (Voss and Hager 2014).

Because TDH3 is one of the most highly expressed proteins in the yeast genome (Newman et al. 2006), we found the skew toward increased expression of trans-regulatory mutations affecting activity of the TDH3 promoter surprising. This skew reflects the near absence of mutations decreasing reporter gene expression by more than 7.5% despite the presence of multiple mutations causing expression to increase by this magnitude. Because most new trans-regulatory mutations are expected to disrupt activity of a trans-acting factor, we interpret the large frequency and magnitude of effects seen for trans-regulatory mutations that increase $P_{TDH3}$ activity as an indication that repressors play a major role in the regulation of TDH3 expression. Because cis-regulatory mutations are skewed toward decreased expression, this additionally suggests that the trans-regulatory mutants with substantial increase in expression we examined affect regulators that do not bind directly to the TDH3 promoter.

The absence of trans-regulatory mutants with large decreases in $P_{TDH3}$-YFP expression (even after we selected specifically for them) could result from trans-regulatory mutations causing strong decreases in $P_{TDH3}$ activity being nonexistent; however, our understanding of TDH3 cis-regulatory sequences suggests that this is not the case: Mutations in the RAP1 and GCR1 binding sites of $P_{TDH3}$ caused large decreases in $P_{TDH3}$ activity, suggesting that trans-acting mutations eliminating or significantly diminishing the function of RAP1 or GCR1 should also cause large decreases in $P_{TDH3}$ activity. The absence of these mutants might be explained by the low fitness of RAP1 and GCR1 null mutants (Giaever et al. 2002), with cells carrying such mutations either dying or being out-competed by other genotypes during the ~10 generations of growth between the introduction of mutations and the isolation of individual mutant genotypes. Low fitness of such mutants could be due to their effects on TDH3 expression, pleiotropic effects on activity of other genes, or some combination of both. In media containing glucose, null mutations in TDH3 have much smaller fitness consequences than null mutations in RAP1 or GCR1 (Baker et al. 1992; Giaever et al. 2002), suggesting that pleiotropy contributes to the low fitness of these mutants. The inability to recover lethal and nearly lethal mutations in studies of mutational effects such as ours is expected to have minimal impact on the utility of these distributions for making predictions about patterns of evolutionary change, however, because mutations causing very low fitness are also expected to be short lived in natural populations.

Consequences of Mutational Properties for the Evolution of Gene Expression

How does gene expression evolve in natural populations? And what are the forces most often responsible for shaping the patterns of regulatory divergence observed within and between species? These questions are difficult to answer, in part because we currently lack realistic null models of regulatory evolution. Generating the data needed to construct such neutral models was one of our primary goals for characterizing the frequencies and effects of new cis- and trans-regulatory mutations. Comparing regulatory evolution expected in the absence of natural selection with patterns of regulatory variation observed in natural populations can be a powerful approach to detecting natural selection and elucidating the underlying forces responsible for regulatory evolution (Denver et al. 2005; Rice and Townsend 2012; Smith et al. 2013; Metzger et al. 2015). Although it is not yet possible to model all aspects of regulatory evolution, the general properties we observed for new mutations can be qualitatively compared with patterns of cis and trans-regulatory divergence observed in natural populations to begin disentangling the contributions of mutation and selection to the evolution of gene expression.

Using statistical significance as a cutoff for defining functional cis- and trans-regulatory mutations affecting $P_{TDH3}$ activity, we found that new trans-regulatory mutations occurred ~250–350 times more often than new cis-regulatory mutations for both average expression and expression noise. When mutations with smaller effects (<1%) were also considered, we found that trans-regulatory mutations were as much as 10,000 times more common than cis-regulatory mutations for both properties of gene expression. These observations suggest that trans-regulatory mutations should be the predominant source of polymorphic expression for species where genetic variation is thought to largely reflect neutral processes. Indeed, many studies of intraspecific expression differences have found that trans-regulatory changes are the primary source of regulatory variation (Lemos et al. 2008; Wittkopp et al. 2008; Emerson et al. 2010; SchaeFke et al. 2013; Coolon et al. 2014, 2015).

Between species, cis-regulatory changes appear to play a larger role (Wittkopp et al. 2008; Tirosh et al. 2009; Coolon et al. 2014). The preferential fixation of cis-regulatory changes over evolutionary time can be caused by reduced purifying selection on cis-regulatory mutations compared with trans-regulatory mutations due to fewer pleiotropic constraints (Wray et al. 2003; Carroll 2005; Stern and Orgogozo 2008, 2009) and/or cis-regulatory mutations more frequently being the target of positive selection (Fay and Wittkopp 2008; Emerson et al. 2010; Coolon et al. 2014, 2015). The larger average effects we observed
for cis-regulatory mutations should make them more likely to be eliminated when deleterious and fixed when advantageous, potentially explaining their greater contribution to divergence than polymorphism. The tendency of cis-regulatory mutations not to be recessive (Lemos et al. 2008; Gruber et al. 2012) might also contribute to their more rapid elimination or fixation within a population.

Taken together, our data are consistent with a model of regulatory evolution in which the neutral process of mutation is primarily responsible for the abundant trans-regulatory variation observed within a species and natural selection is primarily responsible for the excess of cis-regulatory divergence observed between species. Such a model was also suggested and supported by comparisons of cis- and trans-regulatory polymorphisms and divergence using measures of allelespecific expression (Wittkopp et al. 2008; Emerson et al. 2010; Coolon et al. 2015). Future progress in understanding how gene expression evolves in natural populations will require building explicit models of regulatory evolution that account for differences in mutational properties such as pleiotropy, effect size, and dominance between cis and trans-regulatory mutations as well as investigating how these biases are shaped by the action of natural selection. Additional information, such as estimates of pleiotropic effects of individual mutations and measures of their effects on relative fitness, is needed before mature models can be developed.

Materials and Methods

Yeast Strains

Trans-regulatory mutations were created in strain YPW1139 (MATa). This strain is derived from BY4724, BY4722, BY4730, and BY4742 and contains no auxotrophies. In addition, this strain contains five mutations derived from natural yeast strains that fix two defects in the common laboratory strains: High frequency of petites and low sporulation rate. The improved alleles introduced are RME1(ins-308A); TAO3(1493Q) from Deutschbauer and Davis (2005); and SAL1, CAT5-91M, and MIP1-661T from Dimitrov et al. (2009). Finally, this strain contains a copy of the TDH3 promoter, YFP coding sequence, CYC1 terminator, and KanMX4 drug resistance cassette inserted at the HO locus on chromosome IV. cis-regulatory mutations were previously created in strain YPW1 (MATa), as described in Metzger et al. (2015). This strain is derived from BY4724 and is a lys2- ura3- auxotroph. YPW1 contains none of the allele changes made to strain YPW1139, but does carry the same PTDH3-YFP reporter gene. This reporter gene is inserted on chromosome I near the SWH1 pseudogene instead of at the HO locus (Gruber et al. 2012).

To determine whether the differences in genetic background or genomic insertion site of the reporter gene between YPW1139 and YPW1 altered the relative effects of cis-regulatory mutations on reporter gene expression, we compared the effects of 17 different TDH3 promoter haplotypes on PTDH3-YFP expression between the two strains (supplementary table S1, Supplementary Material online). These haplotypes were chosen to capture the entire range of effects on reporter gene expression. We found that the relative effects of these haplotypes were well correlated between the two strains, with an R^2 of >0.96 for both average expression level and expression noise (supplementary fig. S3A and B). This strong correlation shows that differences between these two strains have a negligible effect on the relative effects on reporter expression for individual mutations. To account for absolute differences in expression between the two strains, we used a linear model to estimate the slope of the relationship between YFP fluorescence at the HO locus compared with reporter expression at the SWH1 locus for all 17 TDH3 promoter haplotypes. We used this slope to correct the effects of all 235 cis-regulatory mutants to make their effects comparable with the effects measured for the trans-regulatory mutants.

To further determine whether the relative effects of cis-regulatory mutations measured using the PTDH3-YFP reporter gene were representative of the relative effects of mutations in the native TDH3 promoter, we introduced the same 17 haplotypes into the native TDH3 promoter in a version of YPW1139 that lacked the PTDH3-YFP reporter gene and instead contained the coding sequence for YFP added to the 3’ end of the native TDH3 coding sequence to produce a TDH3::YFP fusion protein at the native TDH3 locus (YPW1452). For the effects of these cis-regulatory mutations on both average expression level and expression noise, we observed a correlation coefficient of R^2 > 0.99 between the reporter gene in YPW1139 and the fusion protein in YPW1452 (supplementary fig. S3C and D, Supplementary Material online). These results are consistent with previous reports indicating that both mRNA (Soboleski et al. 2005) and protein (Kudla et al. 2009) abundance of fluorescent proteins are highly correlated with fluorescence and that the level of YFP fluorescence is a quantitative readout of TDH3 promoter activity (Gruber et al. 2012; Duveau et al. 2014; Metzger et al. 2015).

Mutagenesis

To generate trans-regulatory mutants, a low dose of EMS was used. The specific dose of EMS was chosen to maximize the proportion of cells with a single mutation that significantly altered reporter gene expression while maintaining a low proportion of cells with multiple mutations having significant effects on PTDH3-YFP fluorescence (Gruber et al. 2012). Genome sequencing and genetic mapping of genotypes isolated previously from populations treated similarly with EMS have confirmed that significant changes in PTDH3-YFP expression are typically caused by only a single mutation (Duveau et al. 2014). Because cis-regulatory mutations were more likely to be found in mutants showing low YFP expression, we sequenced the TDH3 promoter in eight EMS-treated mutants that showed an expression level below 95% in this study and found that none carried a cis-regulatory mutation. In addition, sequencing the TDH3 promoter in genotypes isolated from EMS-treated populations that were described in a previous study showed that less than 2% contain mutations in the TDH3 promoter driving YFP fluorescence (Gruber et al. 2012). These genotypes were isolated from the tails of the
fluorescence distribution and were required to have statistically significant effects on YFP fluorescence before being sequenced, thus this 2% represents an upper bound on the frequency of cis-regulatory mutations in the EMS-treated population expected in this study. Indeed, for the 1,485 EMS-treated strains that were randomly isolated, if we assume that each strain carries an average of 32 mutations that are randomly distributed in the genome, we expect only 2 (0–6, 95% percentile) cis-regulatory mutations (0–0.4% of all strains).

Mutagenesis was performed on YPW1139 after reviving from a −80 °C glycerol stock. All glycerol stocks were revived on YPG agar medium (10 g/l yeast extract, 20 g/l peptone, 5% vol/vol glycerol, and 20 g/l agar) and grown for 48 h at 30 °C. Approximately 10^6 cells were then transferred to 10 ml of YPD liquid medium (10 g/l yeast extract, 20 g/l peptone, 20 g/l dextrose) and incubated for 24 h at 30 °C with 250 rpm shaking. After growth to a density of ~7 × 10^7 cells/ml, two aliquots of 1 ml were transferred to separate microcentrifuge tubes. Cells were washed twice in 1 ml of H_2O and then suspended in 1 ml of sodium phosphate (0.1 M). In total, 10 µl of EMS (99%; Acros Organics) was added to one sample (EMS treated), but not the other (control), and both samples were incubated for 45 min at room temperature. EMS mutagenesis was quenched by the addition of 1 ml of sodium thiosulfate (5%), a treatment that was also applied to the control sample. Cells were pelleted and suspended twice in 1 ml of sodium thiosulfate (5%), twice in 1 ml of H_2O, and finally suspended in 1 ml of YPD. About 0.125 ml of each sample was then transferred to 3.875 ml YPD in a 15-ml culture tube that was incubated at 30 °C for 24 h. After growth to saturation, 0.125 ml of culture was added for each sample to 3.875 ml YPD and grown for an additional 24 h at 30 °C to allow for ~10 generations of recovery after EMS treatment. A set of mutagen-treated and control cells were collected in each of the two different experiments performed on separate days using the same protocol.

Measuring the Mutation Rate

After allowing cells to recover from EMS mutagenesis, the mutation rate was estimated for control and EMS-treated samples using a canavanine resistance assay (Gruber et al. 2012; Lang and Murray 2008). Briefly, for each culture, 0.1 ml of a 10−1 dilution was plated on arginine dropout medium supplemented with 60 mg/ml L-canavanine sulfate (Sigma-Aldrich) and 0.1 ml of a 2 × 10−4 dilution of cells was plated on arginine dropout medium (6.7 g/l bacto-yeast nitrogen base, 20 g/l dextrose, 2 g/l drop-out mix minus arginine, 20 g/l agar). The number of colony-forming units was counted on each medium after 48 h of growth at 30 °C, and these counts were used to infer the proportion of canavanine-resistant cells in the initial cultures. Previous work has indicated that there are 88 EMS-like point mutations (G→A and C→T transitions) in the CAN1 gene that can result in canavanine resistance (Lang and Murray 2008). We calculated the average number of point mutations per cell as well as a 95% CI for this average using the proportion of resistant colonies and this mutational target size, as described in Gruber et al. (2012).

Isolating Mutants Using FACS

After EMS recovery, individual EMS-treated and control cells were arrayed in a 384-well plate layout on YPD agar plates using FACS (BD FACS Aria III, University of Michigan Flow Cytometry Core). For each sample, 0.5 ml of saturated culture (~7 × 10^7 cells/ml) was mixed with 2 ml of PBS buffer and run on the FACS machine at a flow rate of ~15,000 cells/s. Gating of flow cytometry events was based on width and forward scatter using FACSDiva software to avoid sorting nonyeast events or aggregates. For the first mutagenesis experiment, 1,340 EMS-treated cells and 160 control cells were sorted irrespective of their fluorescence level onto 5 YPD agar plates.

Prior to the second mutagenesis experiment, two gates were set up corresponding to the 2nd and 98th percentiles of the YFP/FSC (fluorescence divided by cell size) distribution obtained from recording 10^6 nonmutant control cells. A total of 550 EMS-treated cells and 550 control cells were then sorted onto 4 YPD agar plates. For both EMS-treated and control samples, 300 cells were sorted irrespective of their fluorescence level, 125 cells were sorted from the 2nd percentile gate (fluorescence lower than 98% of the control cells), and 125 cells were sorted outside of the 98th percentile gate (fluorescence higher than 98% of the control cells). Considering the two mutagenesis experiments together, a total of 1,640 EMS-treated cells and 460 control cells were collected irrespective of fluorescence level. In addition, 125 EMS-treated cells and 125 control cells were sorted from each of the 2% extreme tails of the control fluorescence distributions.

After sorting, cells were grown into colonies by incubating the plates for 48 h at 30 °C. Overall, no growth was observed at 6% of the positions, either due to the presence of a lethal mutation or because no cell was sorted. After growth, 4 quadrants of 96 colonies were transferred to 4 deep 96-well plates containing 0.5 ml YPD in each well using a V&P Scientific pin tool. Fresh YPW1139 cells that went through neither the mutagenesis procedure nor through a single-cell bottleneck were revived from glycerol stocks and inoculated at 20 fixed positions on each plate. These samples were used to correct for position effects in the plate during subsequent flow cytometry experiments (note that these positions were left empty during cell sorting). After 24 h of growth at 30 °C, 100 µl of all cultures was mixed with 23 µl of glycerol (80%) in sterile 96-well plates and kept frozen at −80 °C. In parallel, all samples were transferred to YPG agar plates using the pin tool and grown for an additional 48 h at 30 °C. At this stage, 4.9% of the samples did not grow and were considered petites (cells lacking mitochondria). Ultimately, these procedures resulted in 1,585 EMS-treated colonies sorted irrespective of their YFP fluorescence, 202 EMS-treated colonies sorted from the 2% extreme tails of the control fluorescence distribution (99 from low fluorescence tail and 108 from high fluorescence tail), and 429 control colonies.
Quantifying Fluorescence Using Flow Cytometry
After growth on YPG agar plates, each sample was transferred to 4 replicate 96-well plates containing 0.5 ml of YPD for fluorescence quantification. Two replicates were inoculated 3 h apart on two different days and grown for 22 h at 30 °C to saturation (~7 × 10⁷ cells/ml). Cells were maintained in suspension during growth by the presence of a 3-mm glass bead in each well and by constant shaking at 250 rpm. Immediately prior to flow cytometry, 15 µl of each sample was diluted into 0.5 ml of PBS in a clean 96-well plate. Fluorescence was recorded for ~2 × 10⁴ events per sample using a HyperCyt autosampler (Intellicyt Corp) coupled to a BD Accuri C6 instrument (488-nm laser used for excitation and 533/30 nm optical filter used for acquisition). For cis-regulatory mutants, flow cytometry data are publically available in the FlowRepository under Repository ID FR-FCM-ZZBN. For trans-regulatory mutants, flow cytometry data are publically available in the FlowRepository under Repository ID FR-FCM-ZZNR.

Analysis of Flow Cytometry Data
Flow cytometry data were analyzed with custom R scripts that are similar to those used in Metzger et al. (2015) (supplementary files S1–S4, Supplementary Material online). Samples with less than 1,000 events after removing budding cells and flow artifacts, or with median FSC more than three times the median average deviation, were excluded. Strains with less than three replicates after removing poor samples were also excluded.

For all samples, we measured the median and standard deviation in fluorescence. We then averaged these values across replicates to determine a genotype’s average expression and expression noise. For gene expression noise, we used the coefficient of variation (average standard deviation in expression divided by the average median expression). In previous work, gene expression noise has been divided into intrinsic and extrinsic components. However, our methodology does not allow us to make this distinction and our measure of expression noise captures both components. Finally, we removed samples whose estimates were likely to be unreliable by discarding strains in which the standard deviation in average expression was greater than 0.1, the standard deviation in expression noise was greater than 1, or the standard deviation in FSC (cell size) was greater than 0.1. To confirm the effects of the 12 large effect (>7.5% change in fluorescence) mutants recovered from the tails of the fluorescence distribution of EMS-treated cells, an additional ~5,000 cells in each of the 6 replicate populations for each strain were grown in parallel in a single 96-well plate following the same procedure and analysis steps described above.

Determining \( P_{TDH3} \)-YFP Copy Number
A prior study (Gruber et al. 2012) found that large increases in fluorescence from \( P_{TDH3} \)-YFP sometimes resulted from duplications of \( P_{TDH3} \)-YFP. To determine whether this was the cause of increased fluorescence in the 16 EMS-treated mutants that showed an increase in fluorescence greater than 7.5% (5 mutants from the unbiased screen and 11 mutants isolated after enrichment for large increases in YFP fluorescence), we used the same quantitative pyrosequencing assay used in the prior study (Gruber et al. 2012) to test for evidence of \( P_{TDH3} \)-YFP duplications. Briefly, the 16 mutant strains were crossed to a common reference strain that carried a single nucleotide substitution in the TDH3 promoter located upstream of the YFP coding sequence. Pyrosequencing was performed in 3 independent replicates for each diploid to quantify the frequency of the TDH3 promoter mutation in DNA, which was expected to be 0.5 in the absence of duplication of the \( P_{TDH3} \)-YFP construct and 0.66 if 2 copies of the construct were present in the mutant strain. All 16 diploids showed an average \( P_{TDH3} \)-YFP mutation frequency between 0.49 and 0.56, suggesting that all 16 strains with expression greater than 7.5% contain only a single copy of the \( P_{TDH3} \)-YFP construct.

Estimating Mutational Target Sizes
Mutational target size was determined in two ways. First, \( t \)-tests were used to identify individual mutants with YFP fluorescence significantly different from nonmutant controls (\( P < 0.01 \)). However, the power of this approach is dependent on sample size and does not necessarily reflect biological significance. As an alternative, we estimated mutational target size for all observed mutational effects. We first assumed that the potential target size was equal to the total number of bases available to be mutated. In addition, we assumed that the realized target size for mutations of no effect is the same as the potential target size. For cis-regulatory mutants, each strain carries only a single mutation and we attributed all differences in fluorescence between a mutant and control to the effect of that mutation. For each mutant we then counted the number of mutants with fluorescence equal to, or more extreme than, that specific mutant’s fluorescence. This was done separately for mutations that increase fluorescence and mutations that decrease fluorescence.

To estimate the target size for trans-regulatory mutants, we used the same procedure with two modifications. First, we accounted for the fact that each trans-regulatory mutant contains multiple mutations. We assumed that epistasis among mutations was rare and that all mutants that did not have effects greater than a specific cutoff did not have mutations with effects greater than that specific cutoff, that is, we assumed that large compensatory mutations were rare. We then assumed that the number of mutations with a specific effect within a mutant follows a Poisson distribution. Under this assumption, the fraction of mutants without an effect beyond a specific cutoff is proportional to the Poisson distribution rate parameter (fraction without effect = \( e^{-\lambda} \)). We used the estimated rate parameters to determine the number of mutants expected to have multiple mutations larger than a specific effect and correct for this bias. This bias is expected to be largest for mutants with small effects and nearly absent for mutants with large effects on YFP fluorescence.

Second, we accounted for the larger effects on expression expected when sorting of large effects. Given the distribution of effects on YFP fluorescence within a sample for nonmutant
control cells and the specific cutoffs used for sorting, we calculated for each effect size the expected enrichment relative to an unsorted sample. We then used this enrichment to correct the target size estimates. We found that using this correction caused estimates of target size to be similar for trans-regulatory mutants sorted irrespective of their effects and trans-regulatory mutants sorted from the tails of the YFP fluorescence distribution, suggesting that it correctly accounted for the expected bias due to sorting from the extremes of the distribution. It also suggests that accurate estimates of target size can be gained by collecting individuals with extreme phenotypes if appropriate corrections can be found.

To determine if the number of mutations expected in each trans-regulatory mutant altered estimates of the mutational target size, we performed the identical calculations but assuming the extremes of the 95% percentiles on the number of mutations within each trans-regulatory mutant (21–43). We also tested our conclusion that cis- and trans-regulatory mutations have different distributions of effects further by comparing the observed effects of trans-regulatory mutants (blue in supplementary fig. S5B and C, Supplementary Material online) with the effects of hypothetical mutants with a similar Poisson distributed number of mutations (λ = 32) drawn from the observed distribution of effects for single cis-regulatory mutations. This “upsampling” of the effects of cis-regulatory mutations was done with replacement, multiplying the effects of cis-regulatory mutations with each other (i.e., assuming no epistasis). We simulated 1,485 hypothetical mutant effects (equivalent to the number of observed trans-regulatory mutants) in this manner using the effects of cis-regulatory mutations with (black in supplementary fig. S5B and C, Supplementary Material online) and without (red in supplementary fig. S5B and C, Supplementary Material online) the effects of mutations in the known TFBS included.

**Estimating Target Size and the Percentage Mutated**

The percentage of the target size mutated was calculated as the number of sites mutated divided by the total number of possible sites that could have been mutated. For cis-regulatory mutants, there were 235 Gs and Cs in the 678-bp TDH3 promoter. For trans-regulatory mutants, we assumed that each mutant contained 32 mutations and that overall there were 32 × 1,485 = 47,520 individual mutations created. Assuming a genome size of 12,071,326, this represents less than 1% of the potential trans-regulatory target and the likelihood of the same mutation occurring twice is thus low.

**Comparing Effects of EMS and Spontaneous Mutations on Amino Acid Sequences**

To determine the expected frequency of amino acid changes in the absence of natural selection, we combined S. cerevisiae codon usage (Gardin et al. 2014) with spontaneous point mutation rates determined from mutation accumulation assays (Zhu et al. 2014). To determine the effects of restricting the type of mutations used to EMS-like GC→AT transitions, we set the mutation rate for all other mutation types to zero and recalculated the expected frequencies of specific amino acid changes. Synonymous mutation rates were calculated as the percentage of mutations expected to result in the same amino acid assuming either natural mutation rates or EMS-like transitions only.

**Statistical Analyses**

All statistical analyses were performed in R (version 3.0.2, R Core Team 2013) using custom code available in supplementary file S1, Supplementary Material online, and the following R packages: flowCore (Hahne et al. 2009), flowClust (Lo et al. 2009), mixtools (Benaglia et al. 2009), moments (Komsta and Novomestky 2015), and lawstat (Hui et al. 2008).

**Supplementary Material**

Supplementary table S1, figures S1–S8, and files S1–S4 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

**Acknowledgments**

We thank the University of Michigan Center for Chemical Genomics, the Flow Cytometry Core, Andrea Hodgins-Davis and Elizabeth Walker for technical assistance with flow cytometry, and Joe Coolon, Andrea Hodgins-Davis, Alisha John, Jennifer Lachowiec, John Massey, and Jonathan Gruber for comments on the manuscript. This work was supported by the March of Dimes (5-FY07-181 to P.J.W.), Alfred P. Sloan Research Foundation (P.J.W.), National Science Foundation (MCB-1021398 to P.J.W.), National Institutes of Health (1 R01 GM108826 to P.J.W.), University of Michigan Rackham Graduate School (B.P.H.M.), National Institutes of Health Genome Sciences training grant (T32 HG00040 to B.P.H.M.), European Molecular Biology Organization postdoctoral fellowship (EMBO ALTF 1114-2012 to F.D.), and National Institutes of Health Genetics training grant (T32 GM007544 to D.C.Y.).

**References**


