Science
AAAS

Supplementary Materials for

Title: Pleiotropic effects of *trans*-regulatory mutations on fitness and gene expression

Pétra Vande Zande, Mark Hill, and Patricia J Wittkopp

Correspondence to: wittkopp@umich.edu

This PDF file includes:

Other Supplementary Materials for this manuscript include the following:

Data S1 to S3:

Data S1. *cis*- and *trans*-regulatory mutant identities, growth rates, and effects on *TDH3* expression.

Data S2: Genes significantly differentially expressed between mating type α and mating type **a** reference strains and between reference strains with and without a *URA3* selectable marker.

Data S3: Genes significantly differentially expressed in the *TDH3* null mutant strain and their expression levels in all mutants.

Materials and Methods

Yeast Genotypes

Strains of *S. cerevisiae* bearing *cis*-regulatory mutations used in this study are a subset of the strains used to assay the fitness effects of changing *TDH3* expression constructed and described in Duveau et al *(17).* They are haploid, mating type **a** strains of *S. cerevisiae* derived from S288C and constructed from the progenitor strain YPW1001, which contains a wild type P_{TDH3}-*YFP* construct and a *NatMX4* drug resistance marker at the *HO* locus and alleles of *MKT1*, *SAL1*, *CAT5* and *MIP1* decreasing petite frequency and the alleles of *RME1* and *TAO3* increasing sporulation efficiency, as previously described (*17*). The strains specifically used in this study include: (1) YPW1177: a strain with a deletion of the entire native *TDH3* promoter and coding sequence (0% of wild-type expression), (2) YPW1156: a strain with a C->T point mutation in the native *TDH3* promoter 482 bp upstream of the *TDH3* start codon in a binding site for the GCR1 transcription factor (~20% of wild-type expression), (3) YPW1200: a strain with a C->T point mutation in the native *TDH3* promoter 485 bp upstream of the start codon in the same binding site for GCR1 (~50% of wild-type expression), (4) YPW1188: a strain with a G->A point mutation in the native *TDH3* promoter 510 bp upstream of the start codon in a binding site for the RAP1 transcription factor (~85% of wild-type expression), (5) YPW1189: a strain with the wild type *TDH3* promoter and coding sequence used as a reference strain for the previous four mutant strains, (6) YPW3059: a strain with two copies of the *TDH3* gene separated by a *URA3* selectable marker in which each copy of *TDH3* contained a G->A point mutation in its promoter 505 bp upstream of the start codon in a RAP1 binding site (resulting in a total ~135% of wild-type expression), and (7) YPW2682: a strain with the same *URA3* selectable marker as the strain

with the *TDH3* gene duplication inserted after the native *TDH3* locus used as a reference strain for the overexpression mutant strain.

The 35 *trans*-regulatory mutants analyzed in this study (Data S1) include a subset of those described in Duveau et al (*16*). Briefly, mutants analyzed with mutations in *GCR1* or *RAP1* were constructed by using mutagenic PCR to randomly introduce mutations within each gene and then CRISPR-mediated allele-replacement to substitute the native locus with a mutant allele *(16)*. The 9 *trans*-regulatory mutants with mutations in one of these two genes each contained 1 to 6 mutations (Data S1, Fig. S1). These specific mutant alleles of *RAP1* and *GCR1* were selected from the hundreds of mutant alleles of each gene characterized in Duveau et al. (2021) for their range of effects on *TDH3* expression including both increases and large decreases in expression. The remaining 26 *trans*-regulatory mutants analyzed each contained a single nucleotide change introduced into the genome by site-directed mutagenesis and either *delitto perfetto* or CRISPR, with the specific mutation introduced identified by genetic mapping of mutant genotypes isolated from an EMS mutagenesis screen for altered expression of *TDH3 (16),* with these nucleotide changes located in the coding sequences of genes involved in purine biosynthesis (4 mutants), iron transport (4 mutants), transcriptional regulation (8 mutants), or other processes (10 mutants), all of which are expected to affect *TDH3* expression indirectly (Fig. 1C). The specific mutations in these genes were either a single nucleotide deletion causing a frameshift (n = 1), a nonsense mutation causing an early stop codon (n = 9), or a nonsynonymous mutation that changes one amino acid (n = 16) (Data S1). *Trans*-regulatory mutant strains and the corresponding reference strain are haploid, mating type α strains of *S. cerevisiae* derived from S288C and constructed from the progenitor strain YPW1139, which also contains a wild type P_{TDH3}-*YFP* construct and a

KanMX drug resistance marker at the *HO* locus and alleles of *MKT1*, *SAL1*, *CAT5* and *MIP1* decreasing petite frequency and the alleles of *RME1* and *TAO3* increasing sporulation efficiency, as previously described (*17*). This same progenitor strain YPW1139, re-stocked and renamed as YPW3016, was used as the reference strain for these *trans*-regulatory mutant strains. Thus, the only differences between the reference strains and mutant strain backgrounds for the under-expressing *cis*-regulatory mutants and the *trans*-regulatory mutants are mating type and resistance marker.

Comparing the RNA-seq data for reference strains (methods for which are described below) with a single copy of *TDH3* that were mating type α (YPW3016) and mating type **a** (YPW1189) showed that 29 genes were significantly differentially expressed between these two strains, 23 of which were annotated as mating type genes, dubious, or uncharacterized ORFs, and 6 of which were annotated with different functions (Data S2, Fig. S6A). These genes were not significantly differentially expressed between the reference strain and mutant strains of the same mating type, suggesting that all differentially expressed genes in mutant strains are attributable to the *cis*- or *trans*-regulatory mutations affecting *TDH3* expression. Nevertheless, these genes were excluded from all further analysis.

Comparing the RNA-seq data for the reference strain containing a *URA3* marker, used to control for the presence of the *URA3* marker in the *cis*-regulatory mutant overexpressing *TDH3* overexpression showed differential expression of 138 genes (Data S2, Fig. S6C). These genes were also excluded from all further analyses. Expression of *URA3* was also discovered in 5 of the *trans*-regulatory mutant lines, suggesting that they had not lost the *URA3* containing plasmid

used during construction of the *trans*-regulatory mutant strains (*16*). To ensure that these strains with the unexpected *URA3* expression did not impact the patterns presented in the main text, all analyses were carried out with and without these strains; conclusions were consistent in both cases. Analyses with these 5 mutants excluded are included in the scripts provided at Github (*27*).

Yeast culturing conditions

These 43 strains (5 *cis*-regulatory mutants, 35 *trans*-regulatory mutants, and 3 reference strains), as well as 3 deletion mutants not used in this study, were randomly arrayed into a 96-well plate containing YPD media (2% dextrose monohydrate, 2% peptone, 1% yeast extract, weight to volume in milliQ purified water and sterilized by autoclave), including 3 replicates of the reference for *trans*-regulatory mutants to provide an estimate of within-replicate variance. The outer rows and columns of this 96-well plate were filled only with sterile media because slight differences in yeast growth at the outer wells of the 96-well plate had been previously observed. 4 unique random plate arrays, each containing all mutants and reference strains (with 3 replicates of the *trans*-regulatory reference strain) in different positions on each plate were designed and assembled. Each plate was then grown to saturation in YPD media with glass beads while shaking at 250 rpm to keep cells in suspension. 100 uL of each culture from each well in each of these four plates was mixed with 23 uL of 80% glycerol and stored at -80 C.

Estimating relative fitness

Relative fitness was estimated based on quadruplicate measures of growth rate for each genotype. A pin tool was used to transfer cells from the four replicate 96-well plates containing

glycerol stocks to a solid YPG (5% glycerol by volume, 2% peptone weight to volume, 2% agar weight to volume, 1% yeast extract weight to volume in milliQ purified water and sterilized by autoclave) agar plates to prevent the formation of the "petite" phenotype which can be common upon thawing. These plates were then incubated at 30 C for ~3 days to allow colony growth. Cells from each genotype were then transferred using a pin tool from the agar plate into 500 uL of liquid YPD media in a 1 mL plate with glass beads and grown with shaking at 30 C for ~3.5 days. 5 uL of these saturated cultures were transferred into 100 uL of fresh YPD liquid media in a Costar 96 well plate with lid, which was then inserted into a BioTek Synergy (Agilent). Cells were grown for 24 hrs at 30 C while being continually shaken to maintain suspension, pausing to take optical density measurements of each well, including blank control wells, every hour. Two strains, containing a mutation in either *CYC8* or *SSN2* showed visible evidence of flocculation (a known phenotype for deletions of both of these genes (*28*)), were excluded from growth rate analysis and subsequent analyses using growth rate measurements. Optical density curves from 0 to 18 hrs (when the diauxic shift occurred), were then plotted and fit to a sigmoidal growth curve using the R program 'growthcurver' *(29)*. Maximal growth rates, calculated as the maximal slope of each fitted curve, were used to calculate growth rate relative to the appropriate control strain to yield average relative growth rates and standard errors (scripts used for analysis and raw data available at Github, *27*).

RNA extraction

Cells used for RNA-seq were sampled from the 4 replicate plate glycerol stocks described above. A pin tool was used to transfer cells from each glycerol stock plate to solid YPG agar plates and cells were grown for ~3 days at 30 C to prevent the formation of the "petite" phenotype as

described above. Cells were then transferred by pin tool into 500 uL liquid YPD media in 1 mL 96-well plates with glass beads and grown at 30 C while shaking at 250 rpm for 2 days until all strains were once again at saturation. 100 uL of the saturated culture was then transferred to a Costar 96 well plate and OD measured using a Tecan Sunrise (Tecan). The OD was then used to calculate cell density. A separate 1 mL 96 well plate was then filled with 500 uL of YPD in each well, and each well inoculated with a volume of the saturated culture calculated to grow to a cell density of $5x10^6$ cell per mL (an OD of about 0.4) after 12.5 hr. These cultures were then grown rotating on a wheel at 30 C for 12.5 hrs, after which 100uL was removed and used to measure OD to ensure all cultures were between an OD of 0.26 to 0.48. Plates were then centrifuged for 5 min at 3000 g, and liquid media pipetted off. Remaining cell pellets were frozen by plunging the entire plate containing cell pellets into liquid nitrogen. Plates were sealed with foil and stored at -80 C until RNA extractions were performed. For any very slow growing strains that did not reach an OD of between 0.26 and 0.48, this process was repeated and multiple cell pellets pooled at the RNA extraction stage to achieve uniform cell numbers across all strains and replicates.

Frozen cell pellets were resuspended in 700 uL of lysis buffer (100mM Tris-HCl, pH 7.5, 500 mM LiCl, 10mM EDTA, pH 8, 1% LiDS, 5mM DTT) containing beta-mercaptoethanol and transferred to a plate containing ~250 uL of acid-washed 425-600 um beads. These plates were vortexed 10x for 1 min each with 1 min on ice in between. Plates were centrifuged at 3000 rpm at 4 C for 4 min, and 400 uL of lysis supernatant were removed and transferred into a new 96 well plate containing 50 uL oligodT magnetic beads resuspended in lysis buffer (Dynabeads mRNA DIRECT Kit, Ambion, cat# 61011). Beads and lysate were incubated at room temperature with agitation for 5 minutes, and then placed on a magnetic stand. Supernatant was

pipetted off and beads were washed 2x with Wash Buffer A (10mM Tris-HCl, pH 7.5, 0.15 M LiCl, 1 mM EDTA, 0.1% LiDS) and 2x with Wash Buffer B (10mM Tris-HCl, pH 7.5, 0.15 M LiCl, and 1mM EDTA) and then eluted in 10 uL of elution buffer (10mM Tris-HCl, pH 7.5) for 2 min at 72 C. After incubation, each plate was placed immediately back on the magnetic stand and the supernatant containing eluted RNA was transferred to an RNAse-free plate. After extraction, several random samples were run on Agilent Bioanalyzer to check RNA quality and concentration before moving on to library preparation. Selected wells were tested for the presence of *NatMX* or *KanMX* resistance markers present in distinct strains and control samples by RT-PCR products visualized on a gel to determine whether well-cross contamination had taken place at any step previous to library prep; we found no evidence of such cross contamination.

RNA-seq pipeline and DESeq2 analysis

RNA-seq libraries were prepared using ⅓ volume reactions from TruSeq RNA Sample Preparation v2 Kit and multiplexable adapters from the TruSeq RNA CD Index Plate (cat# 20019792). Ten of the 198 library preps used in this study failed, resulting in those strains being analyzed in triplicate rather than quadruplicate. All samples from the same replicate plate were pooled and run on one lane on the Illumina HiSeq 4000 by the University of Michigan Advanced Genomics Core, for a total of 4 replicate plates run on 4 sequencing lanes. Raw reads were run through the FastQC (version 0.11.5) read quality software and passed the program's benchmarks for read quality. Reads were then trimmed using Cutadapt ((*30*), version 1.10) and pseudo-mapped to the a transcriptome index generated using Salmon ((*31*), version 0.9.1) on *S. cerevisiae* cDNA (Ensemble, release 38, retrieved from

ftp://ftp.ensemblgenomes.org/pub/release-38/fungi/fasta/saccharomyces_cerevisiae/cdna/). Read counts from Salmon were imported into R using TxDb (version 3.2.2). Read counts for the entire dataset were supplied to DESeq2 (*32*) to model gene expression levels, with different strain backgrounds being analyzed separately. The DESeq2 'contrast' wrapper was then used to estimate $\log_2$ fold changes for each strain relative to its appropriate reference and identify differentially expressed genes and estimate $\log_2$ fold changes in expression relative to the reference. We used PCA analysis to identify any outliers amongst the replicates sequenced and identified one *trans*-regulatory mutant control replicate as well as the replicate of *TDH3 cis*-regulatory mutants with expression of 0%, 20%, 50%, and 85% from the fourth replicate plate as outliers (Fig. S7). These samples were excluded from the analyses.

To assess the reliability of DESeq2 estimates of expression levels from our RNA-seq data, we compared the RNA-seq expression measures for the *cis*-regulatory mutants to previously published expression driven by the same set of *cis*-regulatory mutant alleles controlling expression of a fluorescent reporter gene (*17*). The reporter gene was located at the *HO* locus in these strains, so the fluorescence measures were first compared to fluorescence measures in set of strains with *cis*-regulatory mutations at the native *TDH3* locus driving expression of a TDH3-YFP fusion protein *(18).* This relationship was used to predict expression values at the native locus for strains that contained *cis*-regulatory mutations that were not present in a fusion protein strain themselves. These predictions were then compared to the expression values estimated directly from the RNA-seq data. RNA-seq and fluorescence estimates were strongly correlated with an r^2 value of 0.97 (Fig. S8). Strains containing the fluorescent reporter gene had the same genetic background (YPW1139) as the *trans*-regulatory mutants analyzed in this study, further

reconfirming that mating type and resistance marker do not affect the output of the *TDH3* promoter.

Statistical analysis of differences between cis and trans pleiotropy

In addition to comparing the number of differentially expressed genes in *cis* and *trans*-regulatory mutant strains, we compared the average absolute $\log_2$ fold change or the Euclidean distance among $\log_2$ fold change estimates, which capture the magnitude of expression changes estimated regardless of statistical significance (Fig. S9). These analyses showed similar results to numbers of differentially expressed genes. Permutation tests for each of these comparisons were conducted to assess differences in the number of differentially expressed genes or the Euclidean distance between $\log_2$-fold changes for *cis*- and *trans*-regulatory mutants. These permutation tests were used because they take into account the differences in sample size for *cis*- and *trans*-regulatory mutants. Because there are 5 *cis*-regulatory mutants, 5 *trans*-regulatory mutants were drawn from the total of 35 for each permutation (without replacement), and the median and variance of each sample was calculated. This sampling was repeated 1000 times to create a distribution of the medians and variances measured for 5 *trans*-regulatory mutants, which was then compared to the observed median and variance values for the 5 *cis*-regulatory mutants. P-values were calculated as the proportion of the 1000 random samples of *trans*-regulatory mutants with a median less than or equal to the median or variance of the *cis*-regulatory mutants (Fig. S3).

Gene ontology analysis

Gene ontology analysis was performed using the *Saccharomyces* genome database (SGD) 'GO Term Finder' (Version 0.86) tool (*28*). Both gene sets were tested for enrichment against a background set of all genes included for analysis in our RNA-seq experiment. GO process terms significant at an FDR of 0.01 were examined, and those terms with direct gene associations with the most significant enrichments (rather than parent terms) were reported.

Comparing the effects of trans*-regulatory mutations on genes 'downstream' of TDH3*

To compare the downstream effects of perturbing *TDH3* expression via *cis*- and *trans*-regulatory mutations, for each gene significantly differentially expressed in the *TDH3* null mutant (Data S3), the $\log_2$ fold change estimates across all *cis*-regulatory mutants were converted to percent wild-type estimates. A linear model was then fit to each gene's expression level regressed on the percent wild-type expression level of *TDH3* across all *cis*-regulatory mutants using the base R 'lm' function. We then applied a Benjamini-Hochberg multiple testing correction to the p-values generated from the F-test of the linear regression to identify the 123 genes that were significantly linearly related to *TDH3* expression at an FDR of 10%. Repeating this analysis with more stringent FDR cutoffs of 1% and 5% reduced the number of genes considered significantly linearly related as expected (Fig. S5A) but maintained a similar distribution of regression coefficients among the genes called significant (Fig. S5B). These linear models were then used to predict each of the 122 genes' expression levels in *trans*-regulatory mutants using *TDH3* expression level in each mutant as the predictor. The absolute difference between the predicted gene expression level and the actual observed gene expression level (i.e., the residual) was calculated for each mutant for each of the 122 genes. Whether this residual from predicted expression level was considered significant was determined by calculating the 95% prediction

intervals for the linear model generated from the *cis*-regulatory mutant data using the base R 'predict.lm' function with the interval = "prediction" designation. If the *trans*-regulatory mutant expression level for that gene was greater than one standard error of the mean outside of these 95% prediction intervals, the *trans*-regulatory mutation was said to have a significant pleiotropic effect on expression of that gene. All scripts used to perform these analyses are available on Github (*27*).

Gene deletion collection microarray data analysis

The microarray expression data used in this study to estimate the pleiotropic effects of *trans*-acting gene deletions in *S. cerevisiae* was first published in Kemmeren *et al*. (2014). The processed data file from that publication named "deleteome_all_mutants_ex_wt_var_controls.txt", was downloaded from http://deleteome.holstegelab.nl/ on 06/19/2020. This file included all M values and p-values for expression changes of each gene on the microarray (n = 6,123) for each gene deletion strain assayed relative to the wild type control, where the M value is the $\log_2$ fold-change in expression and the p-value is obtained after Benjamini-Hochberg FDR correction for a statistically significant change in expression relative to a wild-type strain as calculated using the *limma* R package. *Limma* uses linear expression models and an empirical Bayes model to moderate standard errors and calculate a moderated t-statistic and log-odds of differential expression *(33)*. The file also included experiments for strains grown in different media types, which were removed from the dataset for the analyses conducted in this paper. As described in their methods, Kemmeren et al. were aware of the aneuploid strains that have been identified in yeast gene deletion mutants and analyzed all expression profiles for evidence of aneuploidy. Any strains

showing evidence of aneuploidy were remade and re-assayed or excluded from the dataset. This downloaded file was read into R (version 3.5.2) where all statistical analysis was performed (scripts available at Github). A gene was considered significantly differentially expressed in a deletion mutant if it showed a 1.7 or greater $\log_2$-fold change in expression (M value) and a *P*-value less than or equal to 0.05 after Benjamini-Hochberg FDR adjustment, in accordance with the cutoffs used by the authors of the original publication. We excluded 127 gene deletion assays in which the deletion of a gene did not cause a statistically significant decrease in that gene's expression level.

Estimating the pleiotropic fitness effects of gene deletions with an effect on TDH3 expression

To further sample the distribution of pleiotropic fitness effects for *trans*-acting mutations affecting expression of *TDH3*, we also examined previously published data (*22,23*) including the effects of 1106 single gene deletions on *TDH3* expression and fitness (Fig. S2). Despite differences in genetic background and experimental conditions used to collect these data, the distribution of pleiotropic effects observed for these *trans*-acting gene deletions was similar to that observed for the 33 better-controlled *trans*-regulatory mutations (Fig. 2C). To estimate this distribution, we used the same dataset described above (from *(22)*) to identify the impact of each gene deletion on the level of *TDH3* expression to estimate a larger distribution of pleiotropic fitness effects of gene deletions on *TDH3*. We used the $\log_2$ fold change in *TDH3* expression reported in the dataset to calculate the expected fitness effect of the gene deletion attributable to its impact on *TDH3* alone, as predicted by the fitness-expression relationship we previously modeled using our *cis*-regulatory mutants (Fig. 2A).

We then turned to another dataset *(23)* to calculate a pleiotropic, or residual, fitness effect for each gene deletion mutant by subtracting the expected fitness effect due to change in *TDH3* expression (described above) from the total fitness effect of that gene deletion strain. Total fitness measurements for gene deletion strains were taken from Maclean et al. (2017) as reported in "Supplementary Data 6" of that publication. The reported fitness measures of gene deletions relative to the reference strain in YPD media were used directly, without any significance cutoff restricting the analysis to deletions that resulted in a statistically significant decrease in fitness. The resulting distribution of residuals is shown in Fig. S2B.

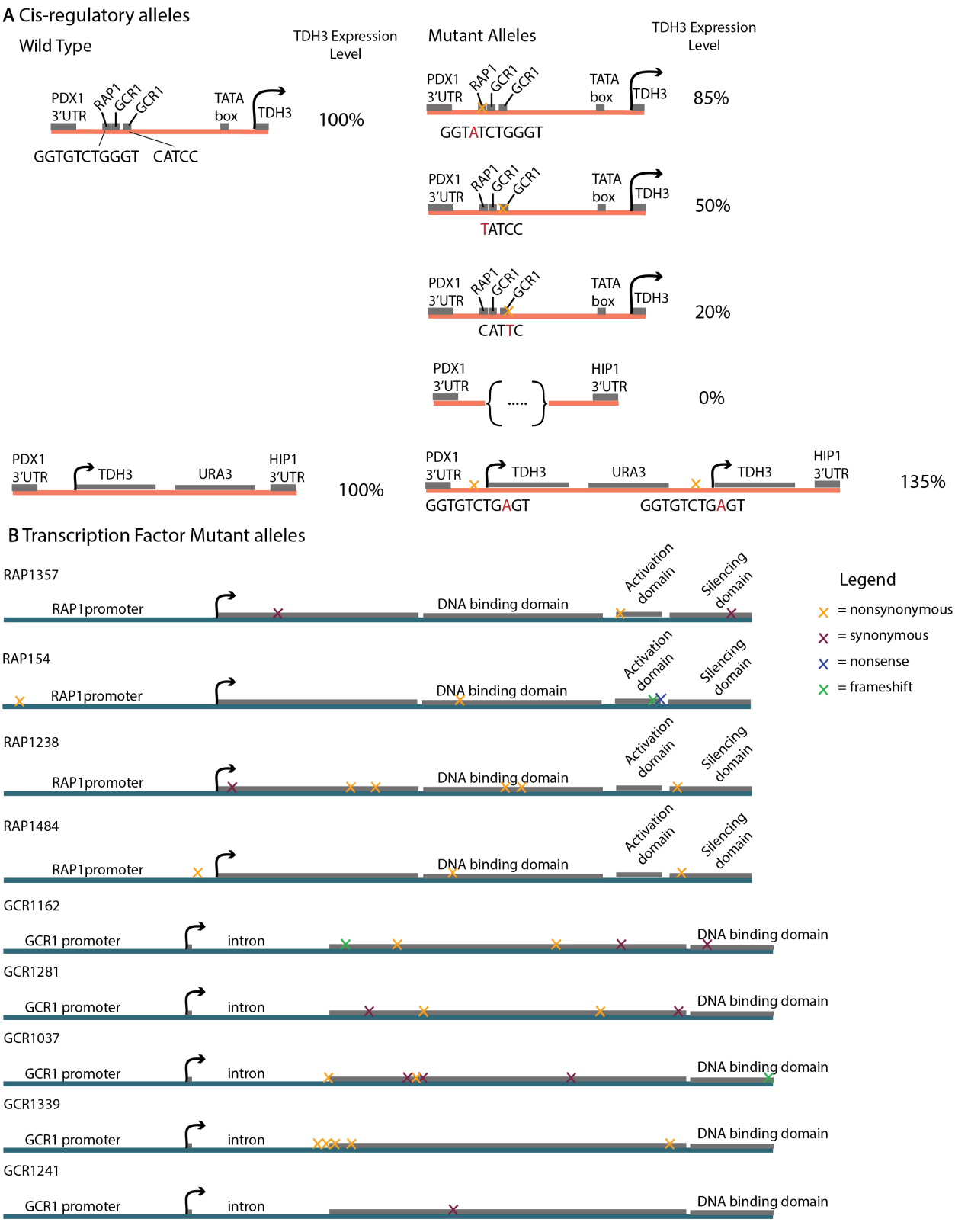
A Cis-regulatory alleles
Wild Type
TDH3 Expression Level
PDX1 3'UTR
RAP1
GCR1
GCR1
TATA box
TDH3
GGTGTCTGGGT
CATCC
100%
Mutant Alleles
TDH3 Expression Level
GGTATCTGGGT
85%
TATCC
50%
CATTC
20%
HIP1 3'UTR
0%
URA3
GGTGTCTGAGT
135%
B Transcription Factor Mutant alleles
RAP1357
RAP1154
RAP1238
RAP1484
RAP1promoter
DNA binding domain
Activation domain
Silencing domain
Legend
= nonsynonymous
= synonymous
= nonsense
= frameshift
GCR1162
GCR1281
GCR1037
GCR1339
GCR1241
GCR1 promoter
intron

Fig. S1. Schematics of *cis*-regulatory mutant alleles and mutant alleles of direct regulators *RAP1* and *GCR1*

Schematics show the alleles in each strain bearing either *cis*-regulatory mutations (orange, panel A) or *trans*-regulatory mutations in the transcription factors *RAP1* or *GCR1* (blue, panel B). (**A**) *Cis*-regulatory mutant strains with reduced *TDH3* expression contain a *TDH3* promoter at the native *TDH3* locus, defined as the 678bp intergenic sequence between the gene *PDX1* and the translation start site of *TDH3,* which bears a single mutation in either a RAP1p or GCR1p transcription factor binding site shown with an 'x'. These mutant strains are contrasted with a strain containing a non-mutated *TDH3* promoter. The *cis*-regulatory mutant strain bearing a null allele of *TDH3* consists of a complete deletion of the *TDH3* coding sequence, shown with an ellipses. The *cis*-regulatory mutant strain with increased *TDH3* expression contains a duplication of the entire *TDH3* locus separated by a *URA3* selectable marker. This strain is contrasted with a strain containing a *URA3* selectable marker downstream of the native *TDH3* locus. (**B**) The *trans*-regulatory mutant strains bearing mutations in the transcription factors RAP1p or GCR1p contain alleles of *RAP1* or *GCR1*, at their native loci, bearing from 1 to 6 point mutations. These mutations include nonsynonymous, synonymous, nonsense, and frameshift mutations that can occur in either the promoters or coding sequences of the genes. Notable domains of the transcription factors, including DNA binding domains and activation and silencing domains of the RAP1 protein are shown in grey.

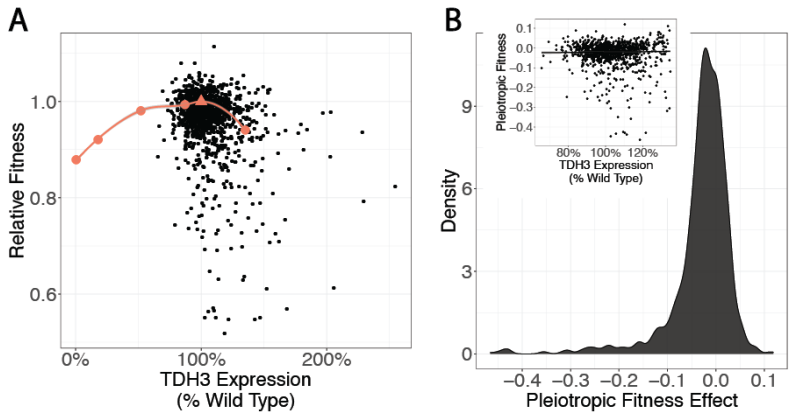

Fig. S2. Pleiotropic fitness effects for deletion strains as *trans*-regulatory mutations to *TDH3*

(**A**) The relative fitness values for 1106 gene deletion strains (as measured in (*23*)) are plotted relative to the expression of *TDH3* in the same deletion strain (as measured in (*22*)). The *cis*-regulatory mutants and fitted LOESS curve from Fig. 2A is included for comparison in orange. (**B**) A smoothed density distribution derived from the pleiotropic fitness measurements (defined the same way as in Fig. 2B) for the deletion data shown in (A) shows a similar shape and negative skew as the smaller dataset shown in Fig. 2C. Inset shows the pleiotropic fitness effects plotted vs the effect on *TDH3* expression including a robust linear regression.

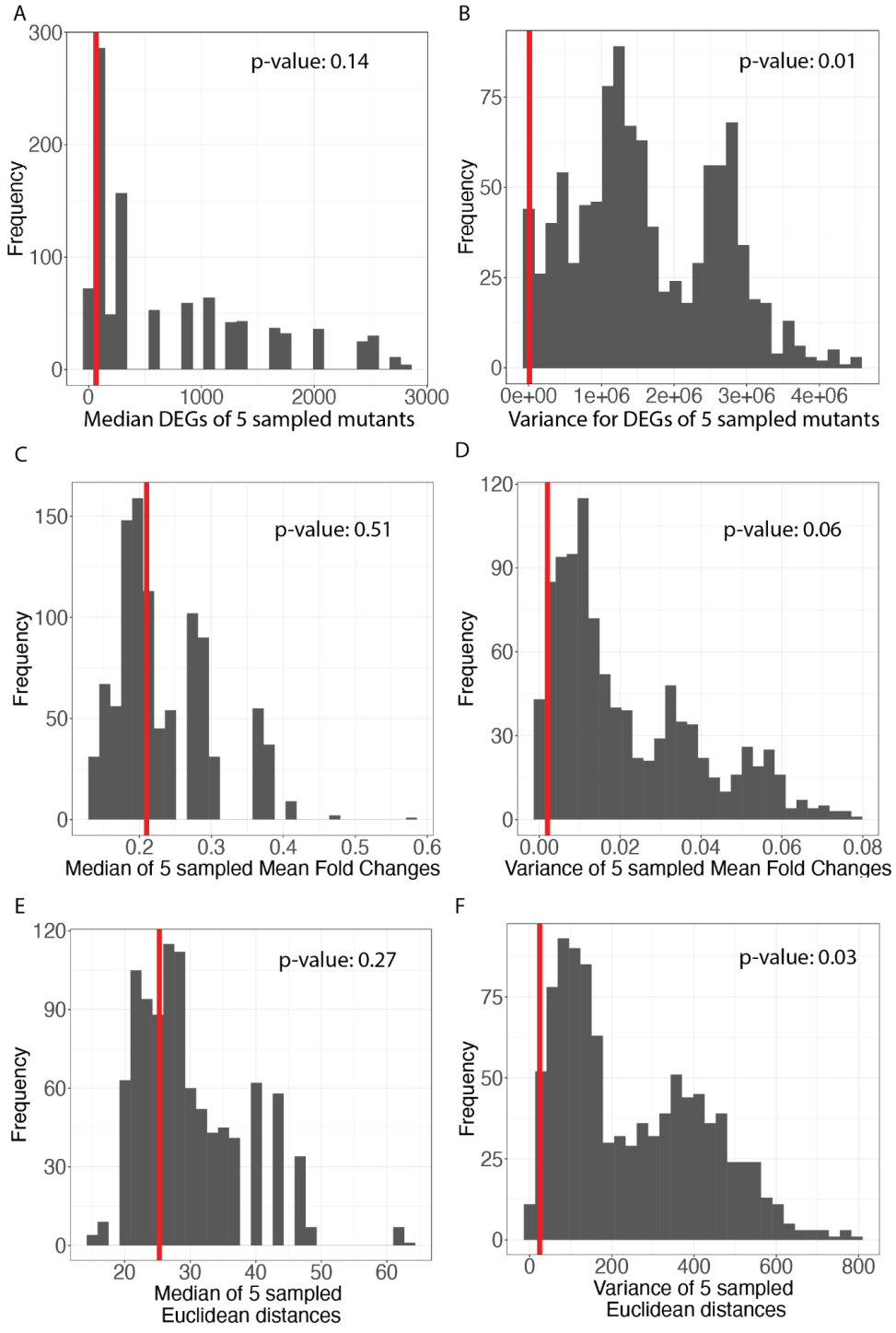

Fig. S3. Permutations tests for comparing effects of *cis*- and *trans*-regulatory mutations on gene expression. (**A**) Histogram (in grey) shows the median number of significantly differentially expressed genes (DEGs) for sets of 5 *trans*-regulatory mutants randomly sampled

from the total set of 35 without replacement (1000 permutations). Red line is the median of the 5 *cis*-regulatory mutants. (**B**) Histogram (in grey) shows the variance in the number of DEGs for sets of 5 *trans*-regulatory mutations randomly sampled from the total set of 35 without replacement (1000 permutations). Red line is the variance of the 5 *cis*-regulatory mutants. (**C,D**) The same information is shown as in A and B, but for average absolute fold change of all genes in the dataset rather than number of DEGs. (**E,F**) The same information is shown as in A and B, but for Euclidean distances among $\log_2$-fold changes rather than the number of DEGs.

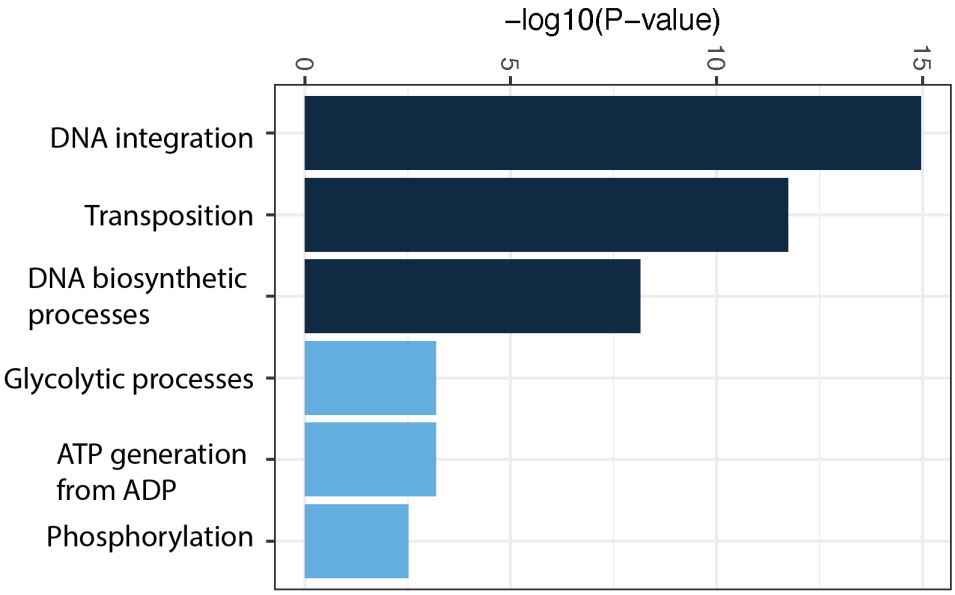

Fig. S4. Gene Ontology (GO) terms enriched in genes differentially expressed in the *TDH3* null mutant. Terms significantly enriched for genes significantly differentially expressed in the *TDH3* null mutant are shown. The three terms associated with transposition are shown in dark blue, and terms associated with glycolysis shown in light blue. Negative $\log_{10}$ P-values were calculated using the SGD online GO term finder tool, using the total set of genes analyzed in this study as the background gene list for enrichment.

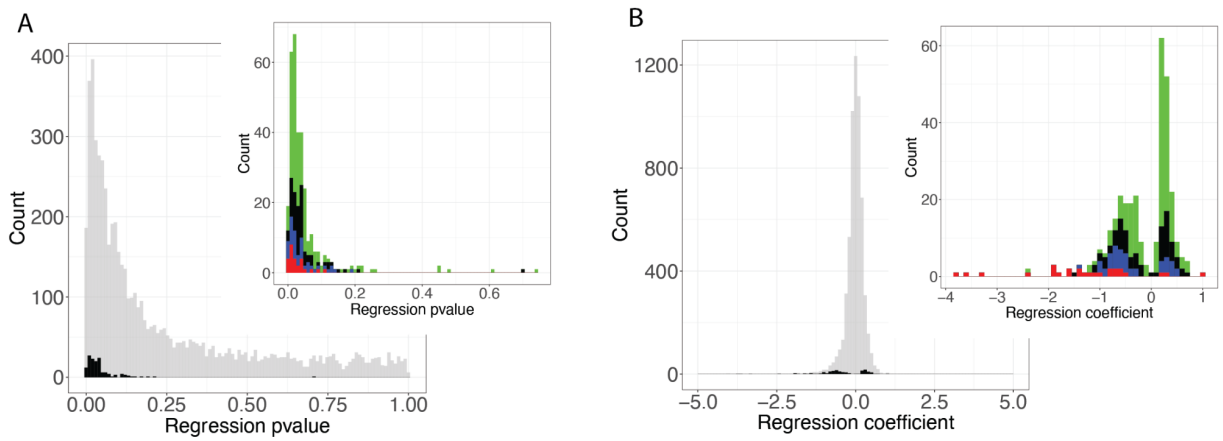

Fig. S5. Linear relationships between *TDH3* expression and expression of other genes in *cis*-regulatory mutants. (**A**) A linear regression was performed for each gene included in the study (n = 6,128) regressed on *TDH3* expression values for *cis*-regulatory mutants. Histograms show the p-value of the F-test calculating *TDH3* level's additional information as a predictor in the linear model. All genes are shown in grey. Genes significantly differentially expressed in the *TDH3* null mutant at an FDR of 10% are shown in black and enlarged in the inset. The inset also shows all genes significantly differentially expressed at a FDR of 20% (green), 5% (blue), and 1% (red). (**B**) The coefficients (slopes) of the linear models in (A) calculated for all genes in the study (grey). As in (A), genes that are significantly differentially expressed in the *TDH3* null mutant at an FDR of 10% are shown in black and enlarged in inset. The inset also shows the coefficients for all genes significantly differentially expressed at a FDR of 20% (green), 5% (blue), and 1% (red).

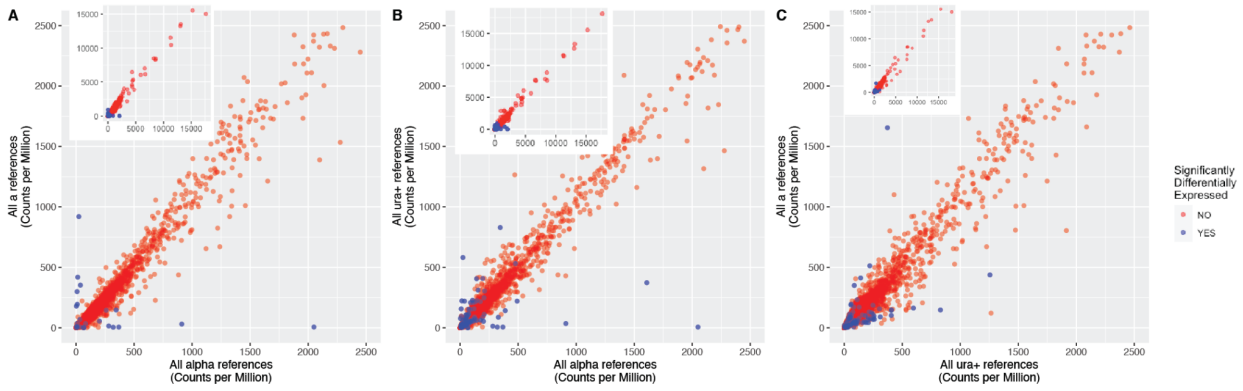

Fig. S6. Comparison of reference genotypes. (**A**) Read counts for each gene, summed across all replicates of the a-type reference strain (YPW1198) and converted to counts per million, are shown on the y-axis. Read counts for each gene summed across all replicates of the alpha-type reference strain (YPW3016) and converted to counts per million are shown on the x-axis. Genes identified as significantly differentially expressed are shown in blue. These genes were removed from all analyses presented in the manuscript. (**B**) Same as in (A), but showing the URA+ reference strain (which is also a-type, YPW2682) on the y-axis and the alpha-type reference (which is ura-, YPW3016) on the x-axis. (**C**) Same as in (A) and (B), but showing the a-type reference (which is ura-, YPW1189) on the y-axis and the URA+ reference (which is also a-type, YPW2682) on the x-axis. These differentially expressed genes, shown in blue, were also removed from the dataset. For all panels, inset panel shows the entire dataset, while the main panel shows an enlargement of the axis from 1-2500 read counts per million.

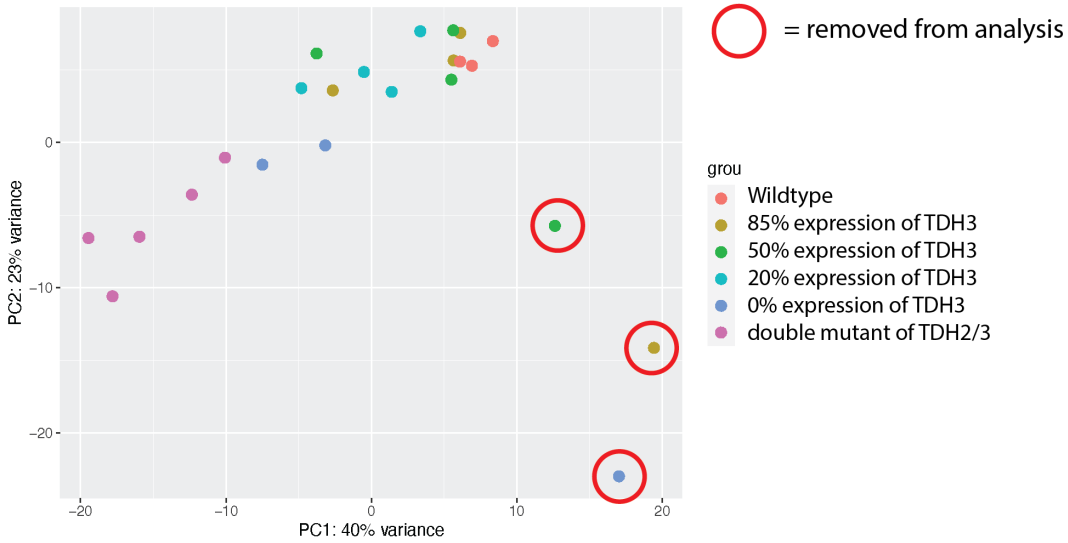

Fig. S7. Principle Components Analysis (PCA) of *cis*-regulatory mutant strains and references showing removal of outlier samples. PCA analysis performed on a variance stabilizing transformed count matrix of all genes with more than 10 reads in all mating type **a** strains, which includes the *cis*-regulatory mutants with *TDH3* expression less than wild-type and the corresponding reference strain. Outliers circled in red were samples excluded from the data set prior to DESeq2 modeling and differential expression analysis.

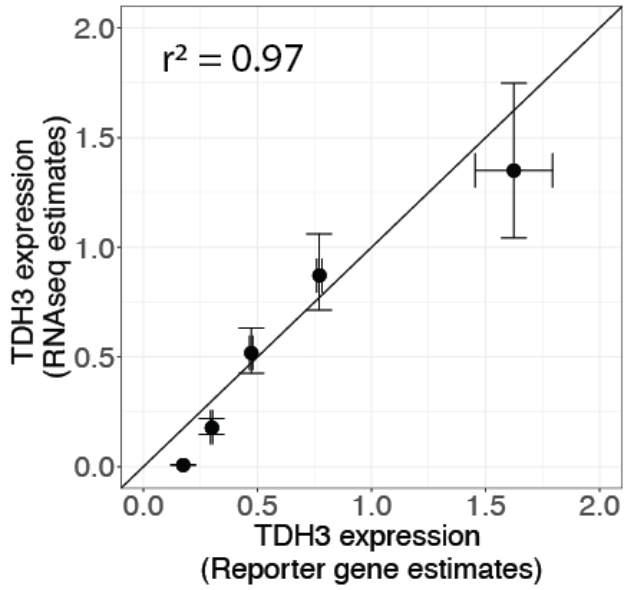

Fig. S8. RNA-seq estimates of *TDH3* expression in *cis*-regulatory mutants are highly correlated with previous fluorescence estimates of the same promoter mutations

RNA-seq estimates of *TDH3* expression correlated highly (Pearson's $r^2 = 0.97$) with expression of a fluorescent reporter gene driven by the same mutations in the *TDH3* promoter measured using flow cytometry (see Methods). Error bars represent one standard error for the $\log_2$ fold change estimates from RNA-seq and flow cytometry.

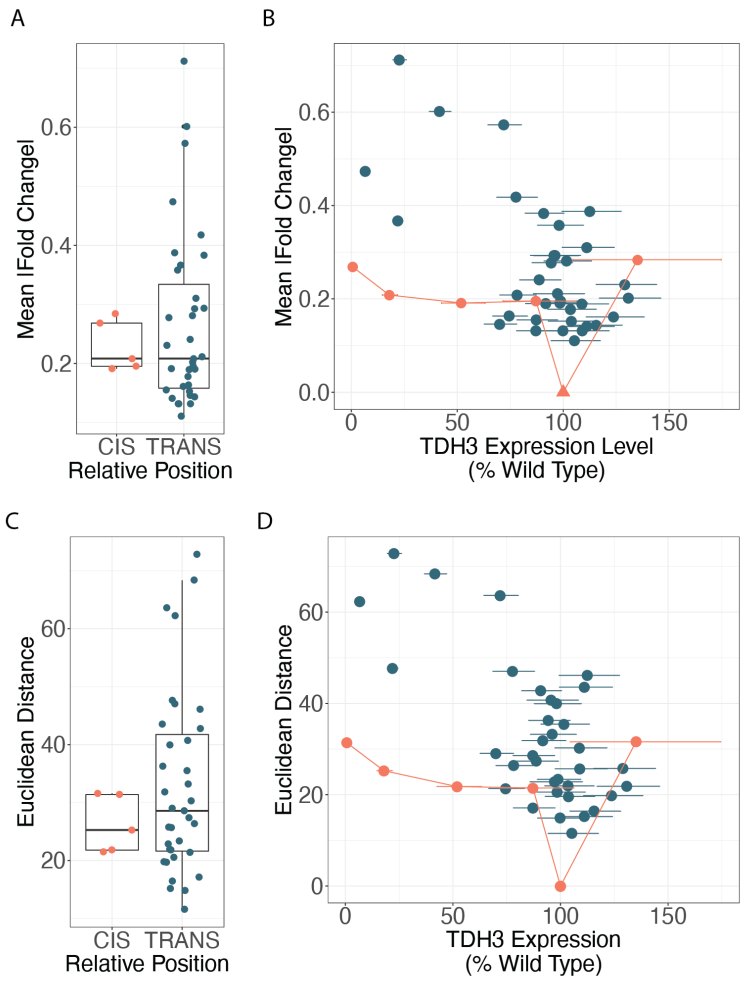

Fig. S9. Comparison of average absolute log2 fold changes and Euclidean distances among all genes between *cis*- and *trans*-regulatory mutants.

(A) Mean absolute $\log_2$ fold changes in expression of all genes in the dataset excluding *TDH3* are shown for *cis*-regulatory (orange) and *trans*-regulatory (blue) mutants. Box plots show median and quartile values. (**B**) Mean absolute $\log_2$ fold changes in expression of all genes are shown for each mutant, plotted according to the mutant's impact on *TDH3* expression. (**C**) Euclidean distances among $\log_2$ fold changes in expression of all genes in the dataset excluding *TDH3* are shown for *cis*-regulatory (orange) and *trans*-regulatory (blue) mutants. Box plots show median and quartile values. (**D**) Euclidean distances among $\log_2$ fold changes in expression are

shown for each mutant, plotted according to the mutant's impact on *TDH3* expression. In both (B) and (D), *cis*-regulatory mutants are shown in orange, with points connected by straight line segments. Error bars for *TDH3* expression are one standard error from RNA-seq data.

Data S1. *cis*- and *trans*-regulatory mutants*:* identities, growth rates, and effects on *TDH3* expression. Each row contains a description of one *cis*- or *trans*-regulatory mutant used in this study. Columns include collection numbers for each strain, the gene in which the mutation is located, the exact position of the nucleotide change for each mutation, the mutation type, the effect on *TDH3* expression relative to the appropriate wild type reference strain, and the growth rate relative to the same reference strain.

Data S2: Genes significantly differentially expressed between mating type α and mating type a **reference strains and between reference strains bearing a *URA3* selectable marker.** Rows are genes significantly differentially expressed between mating type **a** and mating type α reference strains. Columns include gene systematic name, adjusted P-value for differential expression obtained from DESeq2, and functional category or common name of each gene. Un orf = uncharacterized open reading frame, mating = gene associated with mating type differences, dub orf = dubious open reading frame, as annotated in the Saccharomyces Genome Database (*28*). An additional column lists all genes identified as significantly differentially expressed between the reference strain bearing a *URA3* marker and the reference strains without a *URA3* marker. All genes in this data file were removed from the dataset prior to analysing mutant gene expression levels.

Data S3: Genes significantly differentially expressed in the *TDH3* null mutant strain and their expression levels in all mutants. Rows are genes significantly differentially expressed in the *TDH3* null mutant. Columns include the gene's systematic name, the GO terms associated with that gene that were enriched in this gene set, and the expression level of the gene in each of the mutants used in this study.

Supplementary References

27. Code used for data analysis, as well as supporting datasets, are available from GitHub in the repository title "Trans-reg_pleiotropy" at https://github.com/pvz22/Trans-reg_pleiotropy.

28. Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hitz BC, Karra K, Krieger CJ, Miyasato SR, Nash RS, Park J, Skrzypek MS, Simison M, Weng S, Wong ED (2012) Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res. Jan;40(Database issue):D700-5. [PMID: 22110037]

29. K. Sprouffske, A. Wagner, Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics*. **17**, 172 (2016).

30. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*. **17**, 10–12 (2011).

31. R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, C. Kingsford, Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods*. **14**, 417–419 (2017).

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

33. G. K. Smyth, J. Michaud, H. S. Scott, Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics*. **21**, 2067–2075 (2005).