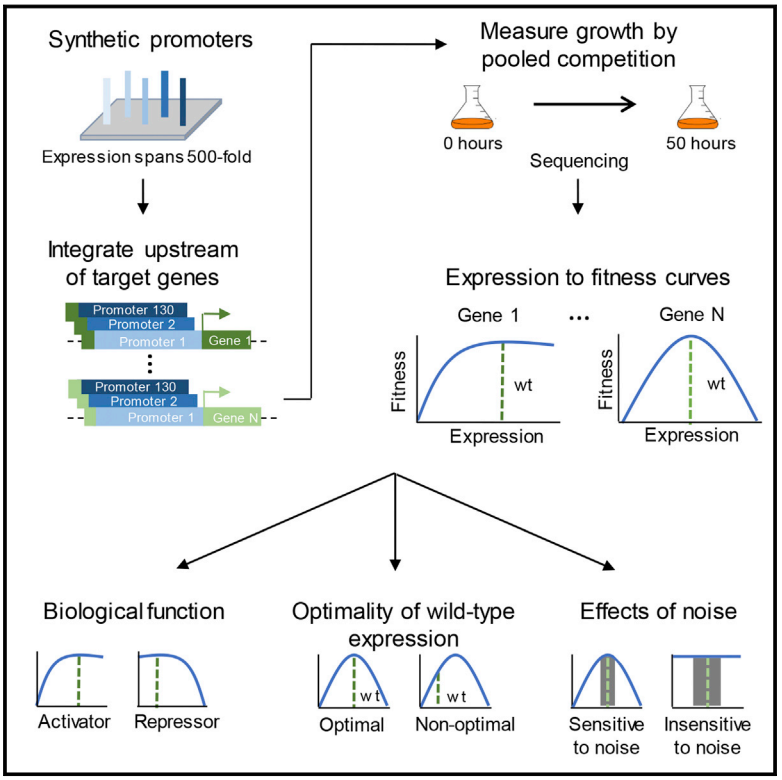
Article

Cell

Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness

Graphical Abstract



Authors

Leeat Keren, Jean Hausser, Maya Lotan-Pompan, ..., Uri Alon, Ron Milo, Eran Segal

Correspondence

eran.segal@weizmann.ac.il

In Brief

How does gene expression variation affect fitness? Systematic probing informs gene function, optimality of wild-type expression, and the impact of noise.

Highlights

- Multiplexed approach unravels causal effects of gene expression levels on fitness
- Expression of many yeast genes altered at high resolution across a 500-fold range
- Condition-specific effect of expression on fitness reveals biological function
- Wild-type expression is optimal for fitness in one condition, but not in another

Data Resources

GSE83936

Keren et al., 2016, Cell *166*, 1282–1294
August 25, 2016
http://dx.doi.org/10.1016/j.cell.2016.07.024

CellPress

Cell

Article

Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness

Leeat Keren,[1,2,4] Jean Hausser,[2] Maya Lotan-Pompan,[1,2] Ilya Vainberg Slutskin,[1,2] Hadas Alisar,[1,2] Sivan Kaminski,[3] Adina Weinberger,[1,2] Uri Alon,[2] Ron Milo,[4] and Eran Segal[1,2,5,*]

[1]Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel
[2]Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel
[3]Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel
[4]Plant and Environmental Sciences, Weizmann Institute of Science, Rehovot 76100, Israel
[5]Lead Contact
*Correspondence: eran.segal@weizmann.ac.il
http://dx.doi.org/10.1016/j.cell.2016.07.024

SUMMARY

Data of gene expression levels across individuals, cell types, and disease states is expanding, yet our understanding of how expression levels impact phenotype is limited. Here, we present a massively parallel system for assaying the effect of gene expression levels on fitness in *Saccharomyces cerevisiae* by systematically altering the expression level of ~100 genes at ~100 distinct levels spanning a 500-fold range at high resolution. We show that the relationship between expression levels and growth is gene and environment specific and provides information on the function, stoichiometry, and interactions of genes. Wild-type expression levels in some conditions are not optimal for growth, and genes whose fitness is greatly affected by small changes in expression level tend to exhibit lower cell-to-cell variability in expression. Our study addresses a fundamental gap in understanding the functional significance of gene expression regulation and offers a framework for evaluating the phenotypic effects of expression variation.

INTRODUCTION

Cells invest vast resources in the regulation of gene expression. Genes are regulated because ultimately their expression levels affect cellular and organismal fitness. Expressing a gene at abnormal levels may be detrimental to proper biological function and lead to malignancies, such as cancer (Emilsson et al., 2008). Remarkable technological advances in high-throughput sequencing are leading to constantly increasing catalogs of gene expression levels in different organisms, cell types, cell states, and individuals, and much research is invested in interrogating the mechanisms leading to observed expression levels (ENCODE Project Consortium, 2004). However, the ability to draw functional conclusions from expression data lags behind because our understanding of how changes in expression levels affect specific traits and overall fitness is limited and mostly qualitative. For example, for most genes it is unknown whether a 2-fold change in expression has functional consequences and if so, what they are.

Genome-wide libraries of knock-outs and overexpression delineate for many genes the effects of completely lacking the gene or expressing it at extremely high levels and have provided valuable information on gene function (Deutschbauer et al., 2005; Gelperin et al., 2005; Gilbert et al., 2014; Sopko et al., 2006; Steinmetz et al., 2002). However, such extreme expression levels are typically far from wild-type expression levels and thus do not reveal the dependence of phenotype on expression in the vicinity of wild-type expression, where natural variation generally occurs. Moreover, measuring a single expression point does not provide information on the relationship between gene expression and phenotype along the entire expression spectrum. Thus, it remains unclear how sensitive or robust are different cellular properties to the expression levels of different genes, whether wild-type expression levels maximize fitness, and how natural variability in expression affects different traits.

To infer the phenotypic effects of changes in expression, it is necessary to vary the expression level of a gene in small increments across a large range and measure the consequential phenotype for each expression level. One common method to achieve this is by replacing the promoter of the gene by an inducible promoter and measuring the phenotype at different concentrations of the inducer (Gossen and Bujard, 1992). For the *Escherichia coli* lac operon (Dekel and Alon, 2005; Dykhuizen et al., 1987; Perfeito et al., 2011; Stoebel et al., 2008) and the yeast gene *LCB2* (Rest et al., 2012), where such measurements were performed, novel insights were provided on the relationship between expression and phenotype—expression levels were shown to affect growth rate in a nontrivial, gene-specific manner, with a linear dependence for one gene (Dykhuizen et al., 1987) and a step-like function for another (Rest et al., 2012). A recent approach used CRISPR-directed activators and inhibitors to modulate expression levels of three mammalian genes and measure resistance to ricin (Gilbert et al., 2014). Although informative, these experiments are laborious, as each expression level requires an independent measure, and they have thus been applied only to a handful of genes, mostly for a single trait (Bauer et al., 2015).

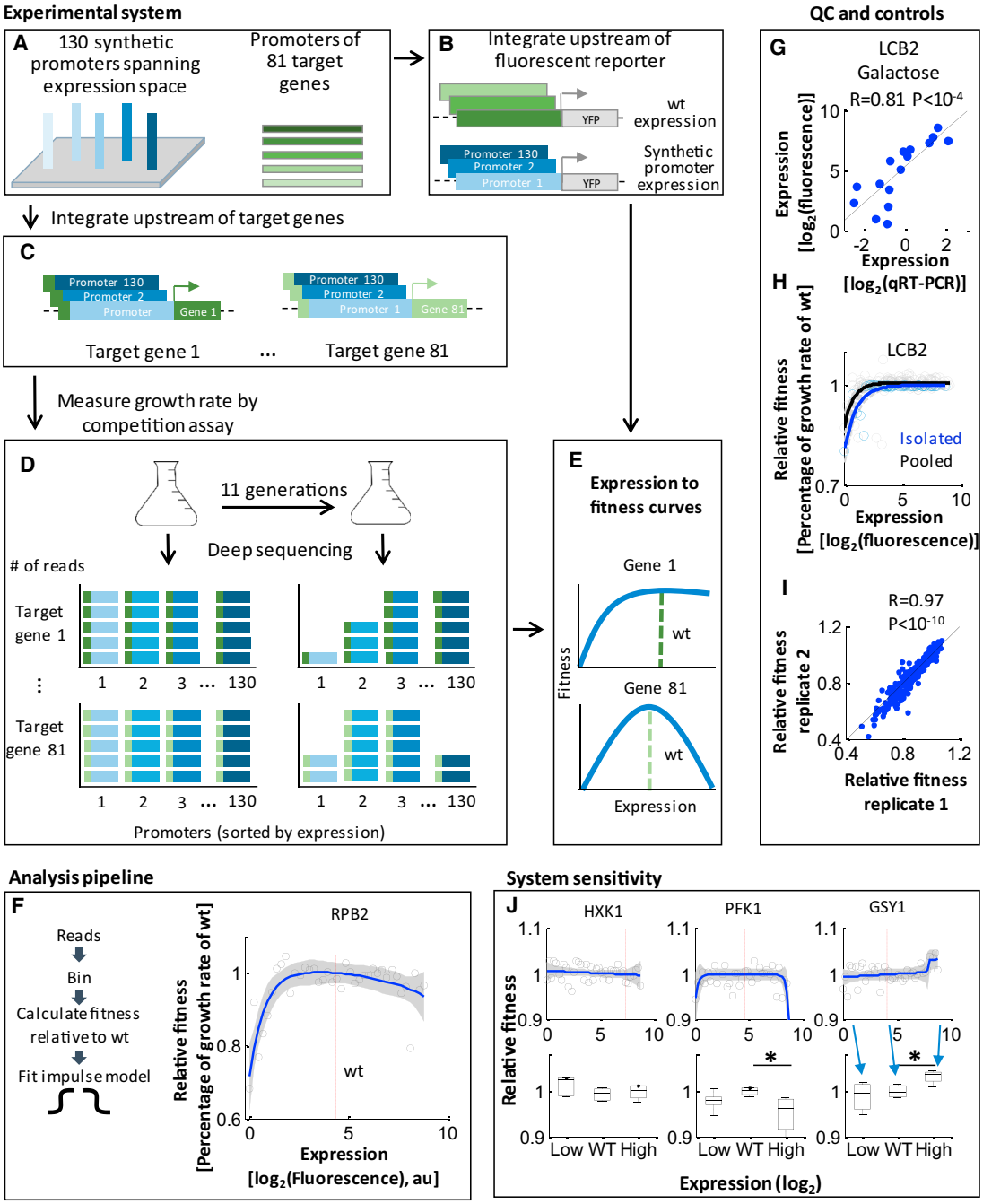


Figure 1. An Experimental System for the Systematic Interrogation of the Effects of Expression Levels of Endogenous Genes on Fitness

(A–E) Illustration of the method. (A) Barcoded, synthetic promoters, spanning an expression range of 500-fold are synthesized on an Agilent array. Endogenous yeast promoters are amplified from the yeast genome by PCR. (B) Promoters are integrated upstream of a fluorescent protein to measure their activity. (C) The set of 130 synthetic promoters is genomically integrated upstream of 81 barcoded target genes, resulting in a library of $81 \times 130 \approx 10^4$ strains. Each strain represents a specific expression level (inferred by a promoter barcode) of a single gene (inferred by a gene barcode). (D) Strains are subjected to a pooled competition assay. Samples are taken to NGS at the beginning and at the end of the assay to extract relative strain abundance and compute growth rates. (E) Example curves for target genes depicting possible dependencies of growth rate on expression level. Wild-type expression is marked by dashed lines.

(F) A schematic of the data analysis pipeline and the output for one target gene. For each strain carrying a synthetic promoter upstream of *RPB2* (gray circles) shown is its relative fitness, calculated as its growth rate relative to wild-type (y axis) as a function of the promoter activity (x axis). Blue line is the best fit of a parametric impulse model to the data, and shaded gray areas mark the 95% confidence intervals. Dashed red line marks wild-type expression levels.

(legend continued on next page)

In this study, we devised a high-throughput approach to systematically study the effect of gene expression levels on different traits, based on the parallel integration of 130 different synthetic promoters upstream of endogenous genes. We applied our method to systematically vary the expression of 81 *Saccharomyces cerevisiae* genes across a wide expression range and at high resolution and measured the resulting growth rate in two environmental conditions by a pooled competition assay. We validated that our method has high sensitivity and can robustly identify changes in growth rate as small as 2%. We found that functional groups of genes differ in their relative sensitivities to low and high expression levels and in their robustness to changes around wild-type expression levels. We found that wild-type expression does not maximize growth in one of the tested conditions for ~20% of the tested genes, despite many changes in expression that occur in this condition. We also found an anti-correlation between expression noise and sensitivity to changes in expression, suggesting that evolutionary selection forces act to shape properties of wild-type expression. We suggest that similar approaches may be employed to study the phenotypic consequences of expression levels for other genes, traits, and organisms.

RESULTS

Construction of a Barcoded Library of Yeast Strains with Different Expression Levels of Endogenous Genes

We chose 81 target genes for the analysis, spanning different cellular functions (metabolism, cytoskeleton, signaling, etc.) and groups (essential and non-essential, singletons and duplicates, high and low expression levels, etc.). We determined the wild-type promoter activities of the genes by integrating their promoters upstream of a yellow fluorescent protein (YFP) in a previously described strain of *S. cerevisiae* (Keren et al., 2013; Sharon et al., 2012) and measuring YFP expression by flow cytometry (Table S2; STAR Methods). Next, we synthesized 130 barcoded synthetic promoters and similarly measured their activities upstream of YFP (Figures 1A and 1B; Table S2; STAR Methods). The synthetic promoters span a range of 500-fold in expression with high resolution (a mean difference of 5% in expression between consecutive promoters), allowing us to explore both downregulation and upregulation for most target genes (Figure S2A). We then mass-transformed our synthetic promoter pool upstream of the target genes in their native genomic location, which resulted in a barcoded library of $81 \times 130 \approx 10^4$ strains (Figure 1C; STAR Methods). Each strain represents a specific expression level (determined by the synthetic promoter) of one of the target genes. Strain identity can be inferred by sequencing a region containing a promoter barcode and gene barcode. Figure S1 depicts the full construction process.

We performed several experimental procedures to ensure the high quality of our library. First, we confirmed correct genomic integration by using a split selection-marker technique (Shalem et al., 2013), coupled to Sanger sequencing (Figures S1 and S2B; STAR Methods). Second, to ensure minimal effects on the neighboring gene, we inserted our insulated promoter cassette between the endogenous sequence and the target gene and did not perturb the endogenous intergenic region. Third, to ensure sufficient representation of less-fit strains in the final pool, transformants were selected on plates, and at least 700 colonies were collected for each target gene. Adequate representation of all strains was validated by next generation sequencing (NGS) (Figure S2D). Fourth, barcodes were designed to allow good distinction between strains. Error in mapping was $<10^{-5}$, and sequencing the same strains with different barcodes results in a correlation of 0.99 (Figure S2E), indicating little effect of the barcodes on quantifying strain abundance.

Together, these results validate the construction of a pooled library of $\sim 10^4$ strains that spans the expression of 81 different target genes at high resolution and show that our system accurately recovers the identity of these strains by NGS. Our method allows us to pool both expression levels and target genes together and conduct a single experiment for each phenotype.

Measurement of the Effects of Expression Levels on Fitness by a Pooled Competition Assay

To test how expression levels of endogenous genes affect growth in an environment with galactose as a carbon source, we performed a pooled competition assay (Pierce et al., 2007) (STAR Methods). Briefly, we grew the strains in galactose medium and took samples for sequencing at time zero and after 27 hr, corresponding to 11 doublings of the pool (Figures 1D and 1E). This time frame provides sufficient time for small differences in fitness to become visible, but does not allow enough time for genetic alterations to shape the population (Pierce et al., 2007). We mapped the reads to the different strains and binned the data into 40 logarithmically equally spaced expression bins. For each strain we calculate the relative fitness, defined as the number of doublings of this strain in the time that wild-type undergoes one doubling (Rest et al., 2012). Thus, a relative fitness of 1 indicates that a strain is equally fit to the wild-type under the set conditions, and values below or above 1 indicate lower or higher fitness than wild-type,

(G) For 20 synthetic promoters, driving the expression of the gene *LCB2*, shown is the correlation between *LCB2* mRNA levels measured by qRT-PCR (x axis) and expression as measured by fluorescence (y axis).
(H) For 50 synthetic promoters, driving the expression of the gene *LCB2*, shown is the relative fitness (y axis) as a function of expression (x axis) for strains competing in a pool (gray circles) or in a direct head-to-head competition with a fluorescently tagged wild-type (blue circles). Blue and black lines depict loess fits to blue and gray points, respectively.
(I) Shown are relative fitness values measured in two biological replicates of the pooled competition experiment.
(J) Shown is the relative fitness (y axis) as a function of expression (x axis) for three target genes: *HXK1*, *PFK1*, and *GSY1*, as described in (F). Boxplots displaying the distribution of relative fitness at either low (<2), wild-type (wt ± 1), or high (>7) expression are shown for each gene. Asterisks mark significant differences as computed by Student's t test.
See also Figures S1, S2, and S3.

respectively. By including the wild-type in the pooled experiment and measuring its growth rate in isolation, we could account for clonal interference in the pool and obtain individual fitness values from the pooled format. For each target gene, we then fitted to the data a previously described parametric impulse function (Chechik et al., 2008), composed of two sigmoidal functions. To avoid overfitting, we used a 10-fold cross-validation scheme. We calculated the 95% confidence intervals and obtained fitness-to-expression curves for 81 genes (Figure 1F; STAR Methods).

We performed several analyses to gauge the integrity and accuracy of our high-throughput approach. First, we validated that for each target gene, the change in fitness between any two strains with similar expression levels (<20% change in expression) is small (<0.1 for 90% of the strains, $p < 10^{-10}$ compared to randomized data; Figures S2F and S2G). Target genes for which this was not the case (*RPL5*, *ACT1*, *RPP0*, *PGI1*) were removed from further analysis. Second, we confirmed by qRT-PCR measurements of 20 isolated strains that the expression measurements derived from fluorescence are a good indication for the expression levels of an endogenous gene driven by the same promoters ($R = 0.81$, $p < 10^{-4}$; Figure 1G; STAR Methods). Promoters driving similar expression levels as inferred from fluorescence intensity measurements lead to similar fitness values (Figures S2F and S2G), further corroborating that reporter expression reflect target gene expression. Third, we verified that the fitness values measured in a pooled competition format are equivalent to the fitness of isolated strains assayed individually (Figures 1H, S3A, and S3B; STAR Methods). Fourth, we assessed our reproducibility by performing a biological replicate of our experiment. We find that our results are highly reproducible, both at the individual strain ($R = 0.97$ $p < 10^{-10}$; Figure 1I) and at the target gene level ($R = 0.99$ $p < 10^{-10}$; Figure S3C). Finally, we confirmed that our experimental system robustly detects ~2% differences in fitness (Figure 1J). Taken together, our results suggest that we can sensitively measure the dependence of fitness on expression for 81 genes in a single experiment. Data and fits for all genes are presented in Table S2 and Figure S4.

The Effect of Expression Levels on Fitness Is Gene Specific

We examined the relationship between expression and fitness in galactose and found that expression levels affect growth rate in an asymmetric and gene-specific manner (Figure 2A). Specifically, some genes exhibit lower fitness at low expression (e.g., *SPT4*) and others at high expression (*PHD1*). Some genes are sensitive to changes in expression levels around wild-type (*TUB1*) whereas others are more robust (*UGP1*). For some genes, fitness changes gradually with expression (*SEC53*), whereas for others we observe sharper, step-like functions (*GAL10*). Most genes in the library have some effect on fitness, and only a few (*HXK1*) do not exert changes on fitness greater than our detection level for the expression range assayed. Such relationships cannot be captured by traditional one-point overexpression or deletion libraries.

Hierarchical clustering of our data according to the shapes of the fitness curves (Figure 2B; STAR Methods) revealed that functionally related genes, belonging to the same metabolic pathway or complex, generally tend to display similar shapes of curves. For example, proteasome subunits were mostly sensitive to downregulation, whereas topoisomerases were equally sensitive to both up and downregulation (Figure 2C). To systematically identify genes important for growth in galactose, we ranked our strains by their relative fitness in the most extreme points of the curves—strong downregulation or strong upregulation (Figures 2D and 2F). Our results agree well with previous studies, as we find that genes previously shown to be essential on galactose (Giaever et al., 2002) tend to have lower fitness if downregulated in our library (Wilcoxon rank-sum test, $p < 10^{-3}$). Accordingly, low fitness values were obtained for genes that have been previously identified to be toxic for growth in galactose when strongly upregulated (Gelperin et al., 2005; Sopko et al., 2006) (Figures S6A and S6B).

More broadly, we found that most genes tested (83%) had an effect on fitness, either at low or high expression levels. Most of these (67%) had an effect at both low and high levels (Figure S6E). In general, for most genes it appears that cells are more sensitive to low expression levels than to high expression levels (Figures 2D and 2F). For example, 30% of the genes exhibit relative fitness values below 0.9 at high expression, compared to 53% for low expression. Next, we subjected the ranked lists of genes to GO enrichment analysis (Figures S6F and S6G). We found that genes of galactose metabolism are particularly sensitive to low expression ($p < 10^{-3}$). The only exception was *GAL80*, a repressor of galactose catabolism (Sellick et al., 2008) (Figure 2E). We found enrichment for two distinct groups of genes sensitive to high expression. The first was genes that are related to cytoskeleton and more specifically chromosome organization (*TOP1*, *TOP2*, *TUB1*, *TUB2*, *TUB3*, *RAP1*) (Figure S6G). The second were transcription factors (TFs) (Figure 2G) in agreement with previous overexpression studies (Sopko et al., 2006). Increased sensitivity to high expression of TFs may be the result of the accumulated effects of subsequent activation or repression of downstream targets (Gill and Ptashne, 1988).

The Relationship between Expression Levels and Fitness Is Indicative of Gene Function

We next examined specific gene groups, starting with the Leloir pathway, used in yeast for galactose metabolism (Holden et al., 2003). Galactose is transported into the cell via hexose transporters, primarily *GAL2*, and metabolized by *GAL1*, *GAL7*, and *GAL10* (Figure 3A), which are repressed in glucose and induced ~1,000-fold in galactose (Sellick et al., 2008). For the enzymes, we found that fitness is maximal at wild-type expression and sharply drops as expression levels go down (Figure 3B). Most of the expression space results in inviability. The shape of these curves probably reflects the high flux carried by these enzymes when growing on galactose (Fendt and Sauer, 2010). In contrast, the transporter, *GAL2*, displayed a more gradual decrease in fitness from wild-type levels and its fitness leveled off at a higher value. This behavior, distinct from the enzymes in the pathway, may result from redundancy in transport, as other transporters may compensate for the reduction in *GAL2* (Holden et al., 2003).

The regulatory core of the pathway exhibited distinct relationships between expression and fitness, matching their function.

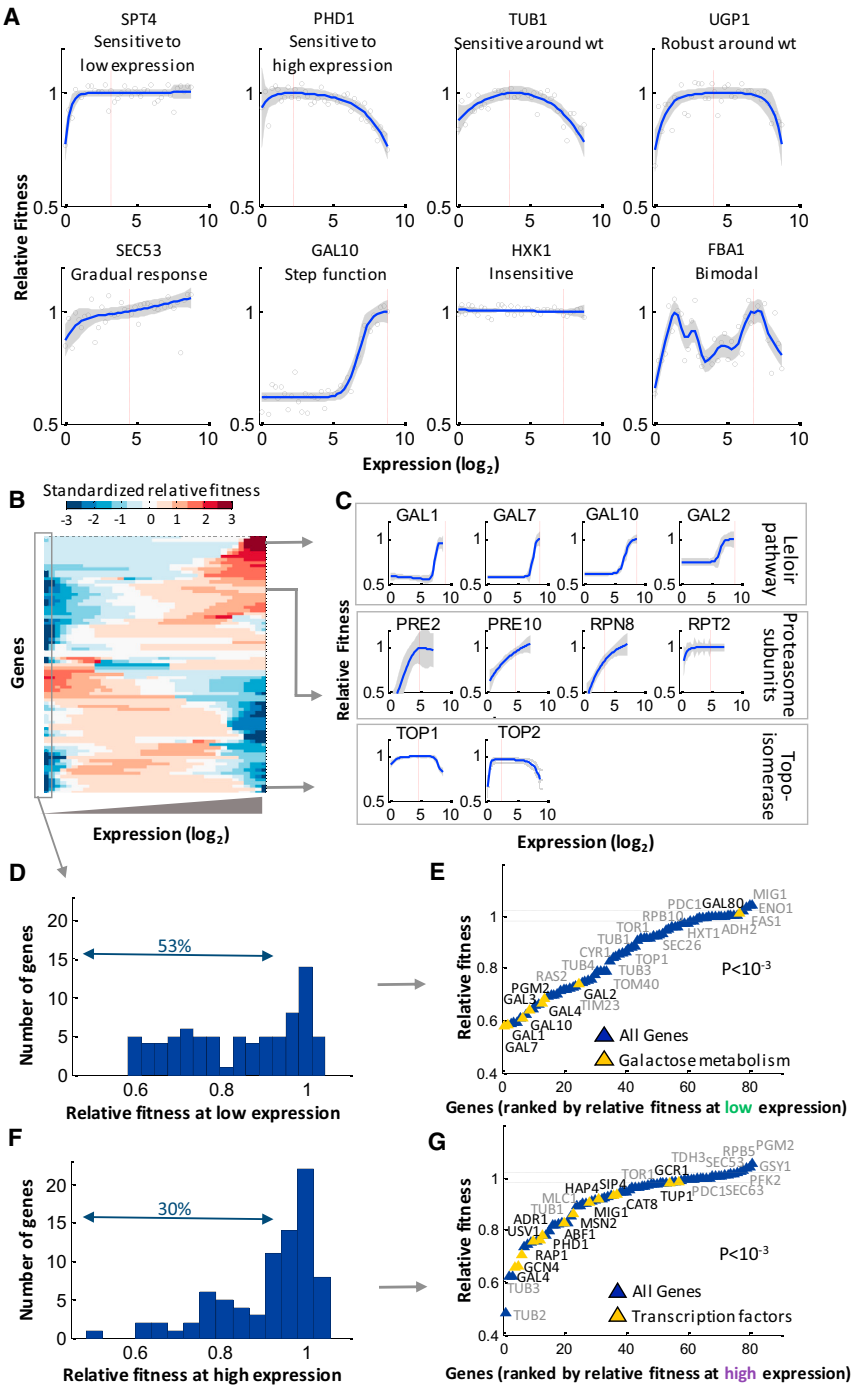


Figure 2. Measurements of the Dependence of Fitness on Expression for 81 Genes in Galactose Reveal Distinct, Gene-Specific Shapes

(A) Shown is the relative fitness of eight representative genes (y axis) as a function of expression (x axis) when in galactose, as described in Figure 1F. For *FBA1*, blue line represents smoothed local regression.

(B) Genes were clustered by the shape of their fitness-to-expression curves. Data was standardized such that genes will have a mean of zero and SD one. The standardized fitness for the ranked expression values (columns) for each gene (rows) is shown.

(C) Shown are genes in the library belonging to different functional categories. The different groups display different shapes of curves, but the shapes are similar for different genes within the group. Arrows denote the location of these genes in the clustered data.

(D) Histogram of the relative fitness of all genes in the lowest expression level measured. Double-edged arrow depicts the percent of genes with fitness <0.9.

(E) Genes were ranked according to their relative fitness in low expression and plotted in ranked order. Genes participating in galactose metabolism are highlighted in yellow and are enriched at low fitness values.

(F) Histogram of the relative fitness of all genes in the highest expression level measured. Double-edged arrow depicts the percent of genes with fitness <0.9.

(G) Genes were ranked according to their relative fitness in high expression and plotted in ranked order. Transcription factors are highlighted in yellow and are enriched at low fitness values.

See also Figures S4 and S6.

For the main transcription factor of the pathway, *GAL4*, we found high sensitivity to both down- and upregulation from wild-type levels. *GAL3* and *GAL80* exhibit mirror-like curves, reminiscent of their activity as an activator and repressor, respectively: high fitness is achieved at high *GAL3* expression and low *GAL80* expression (Figure 3C). Expression of *GAL80* below wild-type levels results in higher fitness than wild-type in agreement with flux measurements for a Δ*gal80* strain (Ostergaard et al., 2000) and evolution experiments in galactose (Hong et al., 2011).

While the general dependence of fitness on expression matches the known function of these genes, more intricate aspects of the relationship are not readily comprehensible. For example, it is unclear why upregulation of *GAL4* is detrimental for fitness, as it leads to increased galactose uptake rates (Bro et al., 2005). It is also unclear how cells are robust to moderate upregulation of *GAL4* (up to 4-fold). These adverse effects may result from accumulation of metabolic intermediates (de Jongh et al., 2008), squelching (Gill and Ptashne, 1988), or non-specific activation of undesirable targets (Kodadek, 1993).

Next, we focused on genes that belong to central carbon metabolism, specifically, glycolysis, gluconeogenesis, and fermentation. Yeast ferment glucose to ethanol by flowing flux through glycolysis. Yeast can also reverse this pathway to generate glucose by gluconeogenesis. Whereas some of the

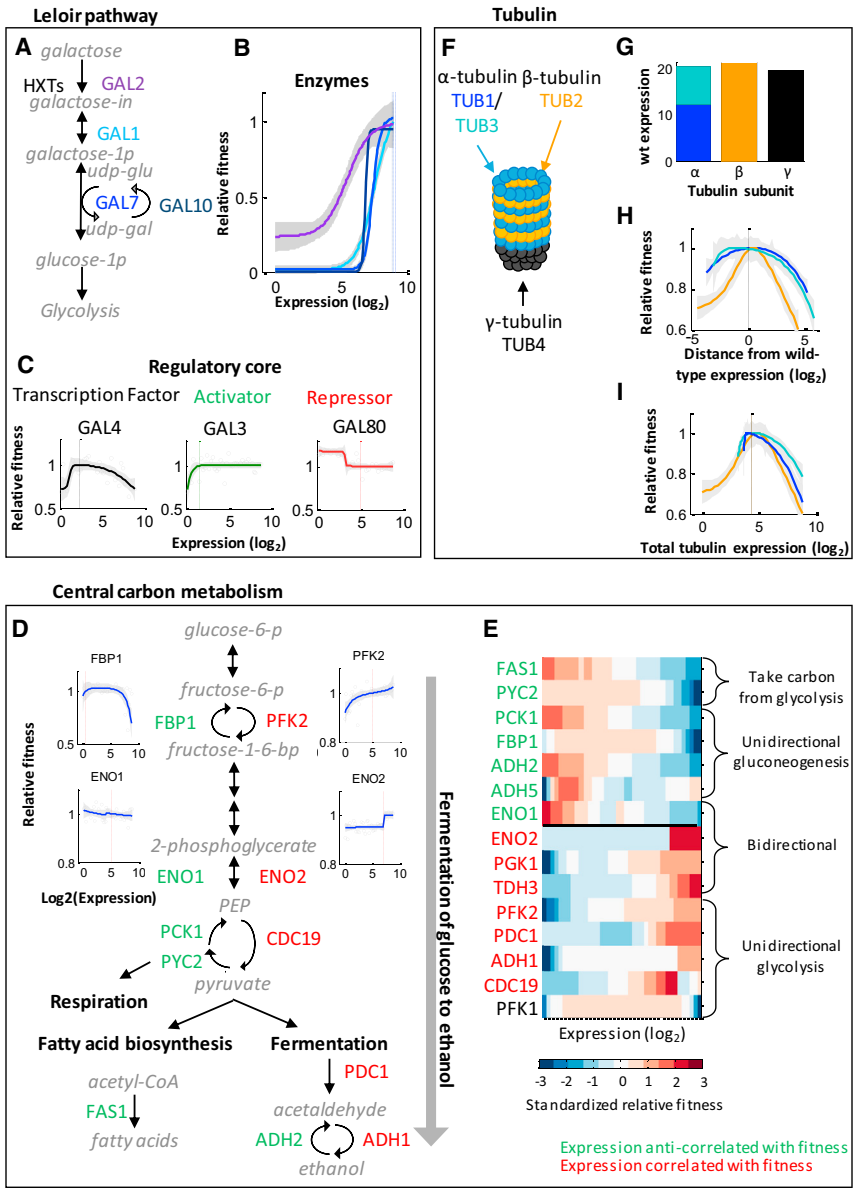


Figure 3. The Relationship between Expression Levels and Fitness Indicates Gene Function

(A) Schematic illustration of the structural components of the Leloir pathway to metabolize galactose.

(B) The relative fitness of the enzymes of the Leloir pathway (y axis) as a function of expression (x axis) in galactose media is shown as in Figure 1F. Genes are colored as in (A). Data after 13 hr of growth is plotted.

(C) The relative fitness as a function of expression for regulatory core of galactose metabolism is shown as in Figure 1F.

(D) Genes participating in central carbon metabolism are depicted next to the enzymatic reactions they catalyze. Red and green font depict genes for which there is a positive or negative correlation between fitness and expression, respectively. The relative fitness as a function of expression for four genes is shown as in Figure 1F.

(E) The fitness for the ranked expression values (columns) for the genes in (D) (rows) is shown. Genes are sorted by their enzymatic activity as either: (1) enzymes that withdraw carbon from glycolysis and fermentation into other metabolic processes, (2) enzymes that catalyze reactions supporting gluconeogenesis, (3) bidirectional enzymes, or (4) enzymes that catalyze reactions supporting glycolysis. Data was standardized such that genes will have a mean of zero and SD one.

(F) The microtubules are polymerized from heterodimers of α- and β-tubulin, nucleated by γ-tubulin (Winsor and Schiebel, 1997).

(G) Bars depict wild-type expression levels of the genes encoding α- (*TUB1* in blue, *TUB3* in cyan), β- (*TUB2*, orange), and γ-tubulin (*TUB4*, black).

(H) The relative fitness (y axis) as a function of expression (x axis), for *TUB1* (blue), *TUB3* (cyan), and *TUB2* (orange) is shown as in Figure 1F. Data is aligned by wild-type expression.

(I) The relative fitness (y axis) as a function of the total expression of the relevant tubulin subunit (x axis) for *TUB1* (blue), *TUB2* (orange), and *TUB3* (cyan) is shown. For example, the fitness of *TUB1* is plotted as the function of its expression level plus the wild-type expression of *TUB3*. Dashed lines mark wild-type expression.

enzymes are shared between both pathways and function both in the forward (glycolysis) and reverse (gluconeogenesis) directions, others are unidirectional (Figure 3D). We find that genes catalyzing both bidirectional (e.g., *PGK1*, *TDH3*) and unidirectional reactions in the forward direction (e.g., *PFK2*, *CDC19*, *PDC1*, *ADH1*) generally show a positive correlation between fitness and expression, with high fitness at high expression levels (Figure 3E, red genes). Correspondingly, we find that genes catalyzing unidirectional reactions in the reverse direction (e.g., *FBP1*, *PCK1*, *ADH2*) and genes that divert flux away from fermentation to other processes (*PYC2*, *FAS1*) display a negative correlation between fitness and expression: fitness is lower at high expression levels (Figure 3E, green genes). Our results suggest that increased flux through glycolysis and fermentation facilitates faster growth in galactose. This increase in flux can be achieved by downregulation of reactions that divert carbon away from glycolysis, by downregulation of the reverse reactions, or by upregulation of the forward reactions.

The data also provides information on directionality in metabolic reactions and divergence of paralogs. For example, for *ENO2* fitness was correlated with expression, whereas for its paralog, *ENO1*, it was anti-correlated (Figure 3E). Thus, we propose that *ENO2* specializes in catalyzing the forward reaction (2PG → PEP) whereas *ENO1* specializes in the reverse reaction (PEP → 2PG) by means of compartmentalization, differential binding partners, etc. Our suggestion is supported by expression data, showing that *ENO2* is highly expressed in glycolytic carbon sources, whereas *ENO1* is upregulated in gluconeogenic carbon sources (DeRisi, 1997; Keren et al., 2013).

Next, we analyzed genes of the tubulin family—the structural elements of microtubules (Figure 3F). We found that the sum of wild-type expression values of the two genes encoding for α-tubulin, *TUB1* and *TUB3*, is similar to the expression of *TUB2*, encoding for β-tubulin (Figure 3G). The microtubules are polymerized from heterodimers of α- and β-tubulin and this result suggests that in wild-type yeast 1:1 stoichiometry of α and β subunits is facilitated by the relative expression levels of *TUB1*, *TUB2*, and *TUB3*.

We also found that the fitness data provides information on the stoichiometry of the microtubule complex and redundancy between the paralogs. *TUB2* is highly sensitive to changes in expression, whereas *TUB1* and *TUB3* are more buffered. (Figure 3H). Plotting the curves as a function of the total amount of the relevant tubulin subunit in the cell, either α (for *TUB1* and *TUB3*) or β (for *TUB2*) results in a high correlation between the three genes (R = 0.99, Figure 3I), indicative of the 1:1 stoichiometry of α and β subunits and compensation of the two paralogs (Papp et al., 2003). While the curves are highly correlated, they still do not overlap entirely, which may suggest additional, differential functions of these genes.

Together, we show that extracting high resolution fitness data for functionally related genes can be used to gain insights on enzyme activity, pathway directionality, parallel or divergent function of paralogs, and stoichiometry of complexes.

The Effect of Expression on Fitness Depends on the Environment

To examine how environmental conditions impact the dependence of fitness on expression, we repeated our measurements of expression and fitness in glucose medium (Figure 4B) and performed validations for their integrity (Figures S2A, S6H, and S6I). We find very good agreement between our results in glucose to published data of deletion strains (Giaever et al., 2002; Kemmeren et al., 2014) (Saccharomyces Genome Database [SGD] pathway database [http://pathway.yeastgenome.org]) (Figures S6C, S6D, S6L, and S6M). Data for all genes is presented in Table S3 and Figure S5.

We compared how fitness curves differed between the two conditions. For each gene, we calculated the Euclidean distance and Pearson correlation between its fitness curves in both conditions. Histograms of both distances (Figure 4A) and correlations (Figure S6J) reveal that most genes exhibit similar dependence of fitness on expression in both conditions (distance <1 for 86% of the genes). The group of genes with similar dependence of fitness on expression included most housekeeping genes (e.g., DNA replication and transcription). The tail of the distribution shows genes that display changes in their fitness curves between the two conditions and is enriched with genes of glucose ($p < 10^{-4}$) and galactose metabolism ($p < 10^{-5}$). The three genes with the largest distance were *GAL1*, *GAL7*, and *GAL10* (Figure 4C). The *GAL* genes display high sensitivity to expression levels in galactose (mean fitness across the expression range = 0.71 ± 0.14), but no sensitivity in glucose (1 ± 0.03), consistent with the large metabolic flux carried by the pathway in galactose, but not in glucose. Genes participating in glycolysis and fermentation exhibited greater sensitivity to their expression levels in glucose (0.75 ± 0.17) than in galactose (0.96 ± 0.04, Figure 4C). These results indicate that in media containing glucose as the carbon source, the glycolytic genes exert high control over growth, whereas in galactose, reactions upstream of glycolysis are more limiting for growth.

TFs exhibited different degrees of correlation between the conditions (Figure 4D). Approximately half (e.g., *SPT15*, *GCN4*, and *HAP4*) had similar fitness curves in both conditions, suggesting a common function in both environments. The other half (8 out of 15 genes) displayed large differences in their fitness-to-expression curves between the two conditions, in agreement with their canonical function. For example, we found a strong dependence of fitness on expression levels of *GCR1*, one of the main transcriptional activators of glycolysis (Chambers et al., 1995), in glucose, but not in galactose. Correspondingly, in galactose, but not in glucose, we found strong sensitivity to the expression levels of *GAL4*. For some of the TFs, the differential sensitivities of fitness to expression levels across conditions is less clear and may hint at new functions or regulatory interactions. For example, we found high sensitivity to high expression levels of *PHD1*, a transcriptional activator of pseudohyphal growth (Borneman et al., 2006), in galactose, but not in glucose. Overall, we find that environment plays an important role in determining the fitness outcome of changes in expression and that the same expression level of a gene may exert different effects on fitness, depending on the environmental conditions.

Wild-Type Expression Levels Are Optimized for Growth in Glucose, but Not in Galactose

Next, we evaluated the extent to which wild-type expression levels maximize fitness in a given environment. For each gene in both conditions, we computed the difference between their relative fitness in low or high expression and their wild-type fitness. A threshold of 2% increase over wild-type was used to determine that up/downregulation of a gene had improved fitness over wild-type. Several procedures were employed to ensure a minimal number of false positives (STAR Methods).

In glucose, we found that none of the genes had higher fitness than wild-type if downregulated, and only one gene, *HAP4*, a global regulator of respiration (Zampar et al., 2013), conferred higher fitness upon upregulation. A global upregulation of the TCA cycle may be beneficial in this condition for the generation of amino acid precursors, as the medium did not contain amino acids except for histidine. In galactose, we identified 13 and 2 genes for which high or low expression levels, respectively, increase fitness over wild-type expression levels (Figures 5A and S6K). These results suggest that *S. cerevisiae* is highly adapted to maximizing growth on glucose, but it is possible to tune its gene expression to increase its growth on galactose. By changing the expression of only one gene, it is possible to increase growth in galactose by up to 13%.

The 15 genes whose wild-type expression levels are not optimal in galactose belong to three cellular categories: carbon metabolism, RNA polymerase, and the protein secretion pathway, which may hint at the major processes that are bottlenecks for fast growth (metabolism, transcription, and membrane biogenesis, Figure 5A). Overexpression of *PGM2*, a phosphoglucomutase, which links galactose metabolism to glycolysis, conferred the most dramatic increase in fitness (13%) compared

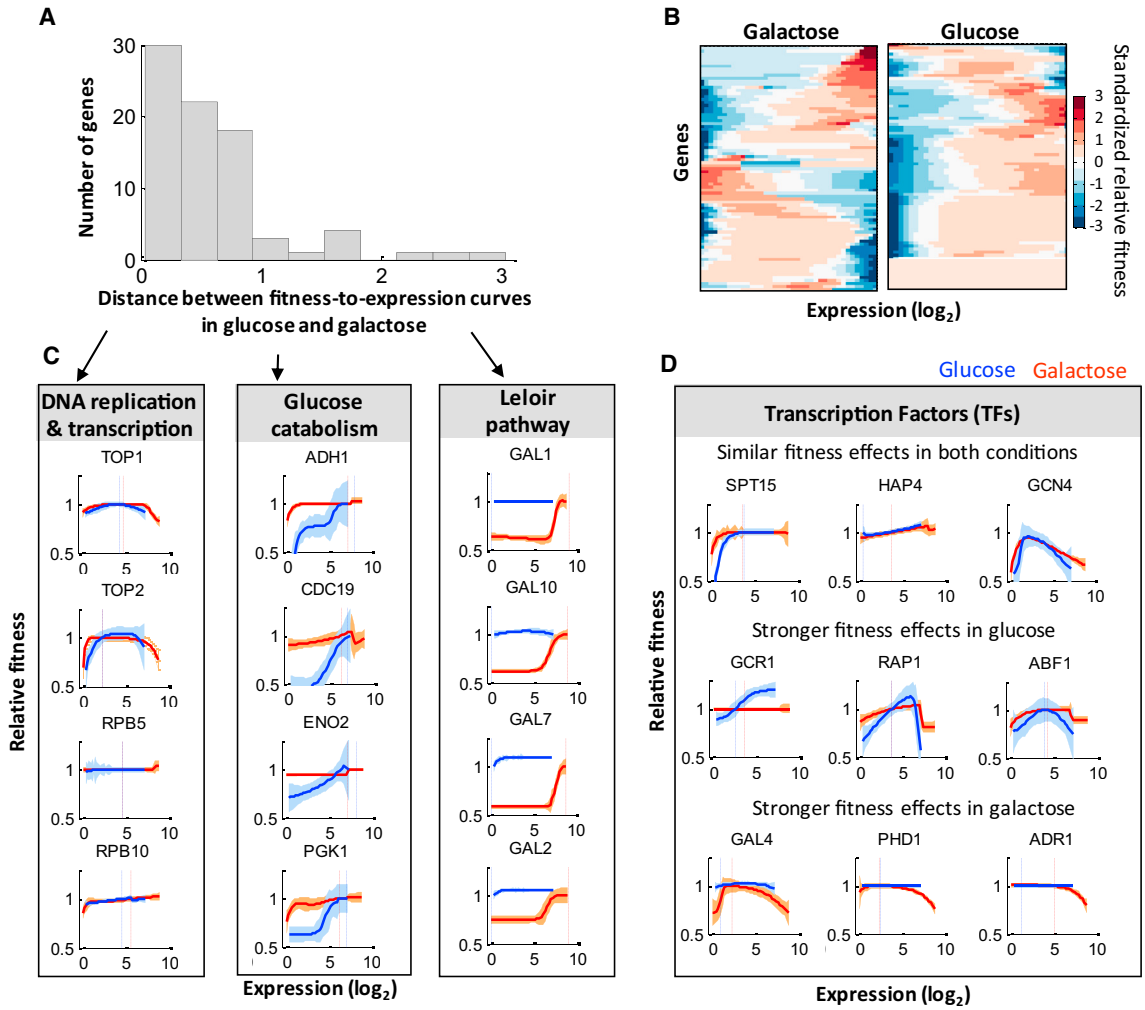


Figure 4. Comparison of the Dependence of Fitness on Expression across Conditions Is Instructive of Gene Function

(A) The Euclidean distance between the fitness-to-expression curves in glucose and galactose was computed for all genes in the library. Shown is a histogram of the distance values.

(B) Fitness-to-expression data for each gene was clustered and plotted as in Figure 2B for growth in galactose (left) and glucose (right).

(C) Genes were ranked by the distance between their fitness-to-expression curves in glucose and galactose and analyzed for GO enrichment. Shown are groups of genes that were enriched at the tail (Leloir pathway, Glucose catabolism) or body of the distribution (DNA replication and transcription). For each group, the relative fitness (y axis) as a function of expression (x axis) in both galactose (red) or glucose (blue) is depicted for representative genes.

(D) Genes were classified by the standard deviation of their fitness into groups that have stronger fitness variations in glucose, galactose, or similar variation in both. For nine transcription factors, three from each group, shown is their relative fitness (y axis) as a function of expression (x axis) in both galactose (red) or glucose (blue).

See also Figures S5 and S6.

with wild-type levels. Previous studies found that overexpression of *PGM2* results in improved glycolytic yields on galactose (Bro et al., 2005; Lee et al., 2011). We found that overexpression of *SEC53*, another phosphoglucomutase and a homolog of *PGM2*, also conferred higher fitness than wild-type. Thus, *SEC53* may assume an active role in central carbon metabolism in addition to its canonical function in ER-associated glycosylation (Bernstein et al., 1985). High expression of *GSY1*, which functions in glycogen synthesis, also results in higher fitness than wild-type. Although it is unclear why glycogen synthesis is beneficial for yeast growing on galactose, this result agrees with measurements showing that yeast evolved in galactose have increased glycogen levels (Hong et al., 2011). Increased flux through glycolysis and fermentation also facilitates faster growth. In agreement with previous results (Bro et al., 2005; Ostergaard et al., 2000), we found that low expression of *MIG1*, the regulator of glucose repression, improves fitness over wild-type. *MIG1* represses expression of the GAL genes, and thus reducing its levels may allow upregulation of the entire pathway simultaneously (Figure 5B).

We also found that wild-type expression levels were sub-optimal for growth on galactose for several components of

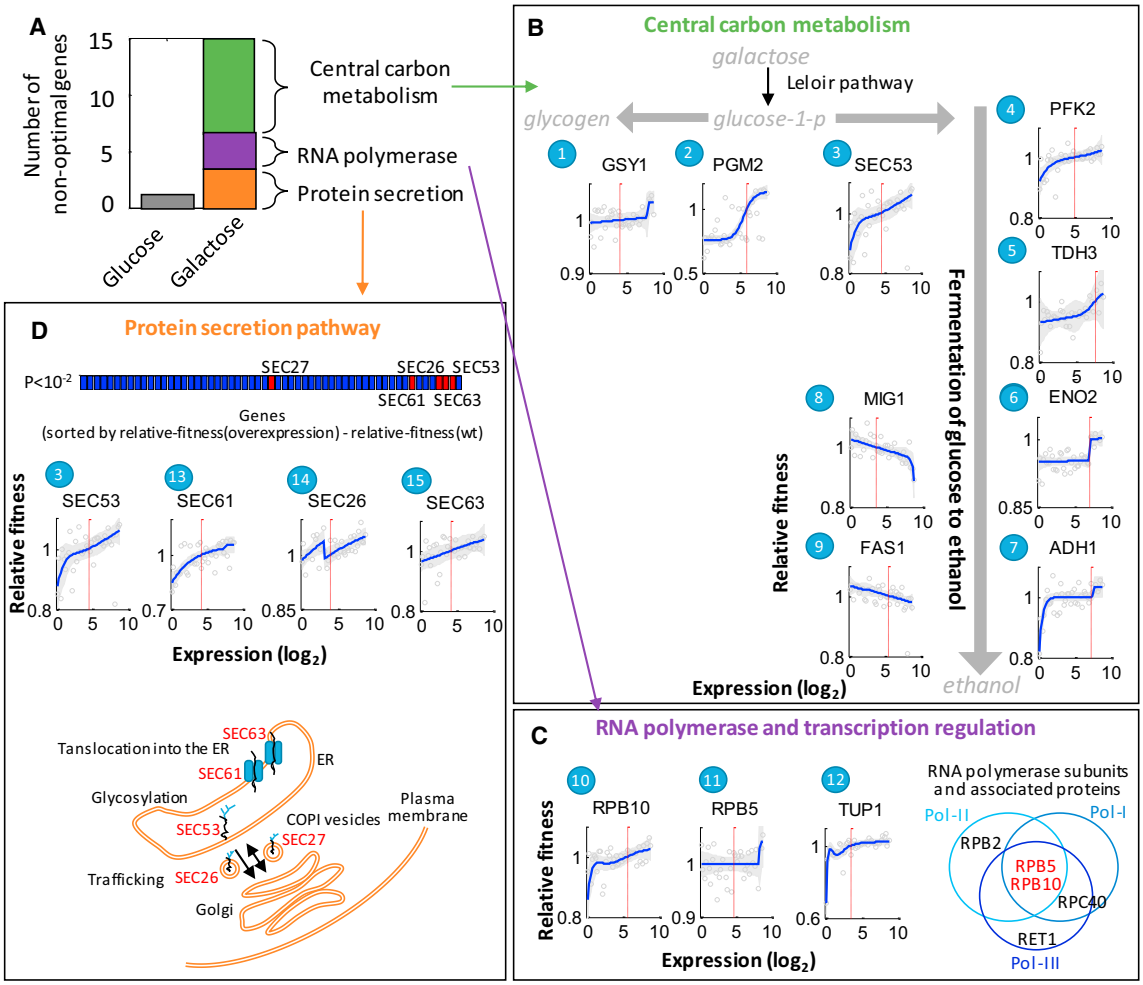


Figure 5. Wild-Type Expression Levels Are Optimal for Growth in Glucose, but Not in Galactose

(A) Bars depict the number of genes whose fitness at either low or high expression is at least 2% higher than their fitness at wild-type expression.

(B–D) For the 15 genes with non-optimal wild-type expression levels in galactose, the functional cellular categories to which they belong and their fitness-to-expression curves in galactose media is shown as in Figure 1F. (B) Central carbon metabolism. (C) Global transcription regulation. Right: Venn diagram depicting which of the RNA polymerase subunits essayed in this study belong to the complexes of polI, polII, and polIII. (D) Protein secretion pathway. Top: genes were sorted by the difference between their relative fitness in high expression and their relative fitness for wild-type expression levels. Components of the protein secretion pathway are colored red. Bottom: illustration of the genes' function in the secretory pathway.

See also Figure S6.

RNA polymerase and the protein secretion pathway (Figures 5C and 5D). Our library included five components of RNA polymerase (*RPC40*, *RPB2*, *RPB5*, *RPB10*, *RET1*) and the global transcriptional regulator, *TUP1*. Three of these (*RPB5*, *RPB10*, *TUP1*) were found to confer higher fitness than wild-type (2%–5%) when upregulated, in agreement with the literature (Lee et al., 2011). Our library also included five components of the secretory pathway (*SEC26*, *SEC27*, *SEC53*, *SEC61*, *SEC63*). Four of these (all but *SEC27*, Wilcoxon rank-sum test, $p < 10^{-2}$) were found to confer higher fitness than wild-type (3%–6%) when upregulated. Our results suggest high control coefficients on growth of these pathways in galactose.

Together, we found that wild-type expression levels are optimized for growth on glucose, but not for growth on galactose, despite the major expression changes that occur in the presence of this carbon source (Keren et al., 2013). Optimizing yeast for growth on galactose also has biotechnological implications, and we identify 15 targets for bioengineering. A prediction of our study is that expression of these genes will likely be subject to change when yeast evolves to grow on galactose. Indeed, *PGM2* and *GSY1* were upregulated in yeast evolved on galactose (Hong et al., 2011), as predicted by our fitness-to-expression curves.

Fitness Curvature at Wild-Type Expression Levels Is Anti-correlated with Expression Noise

We investigated the relationship between our fitness-to-expression curves and the degree of wild-type noise levels. Wild-type

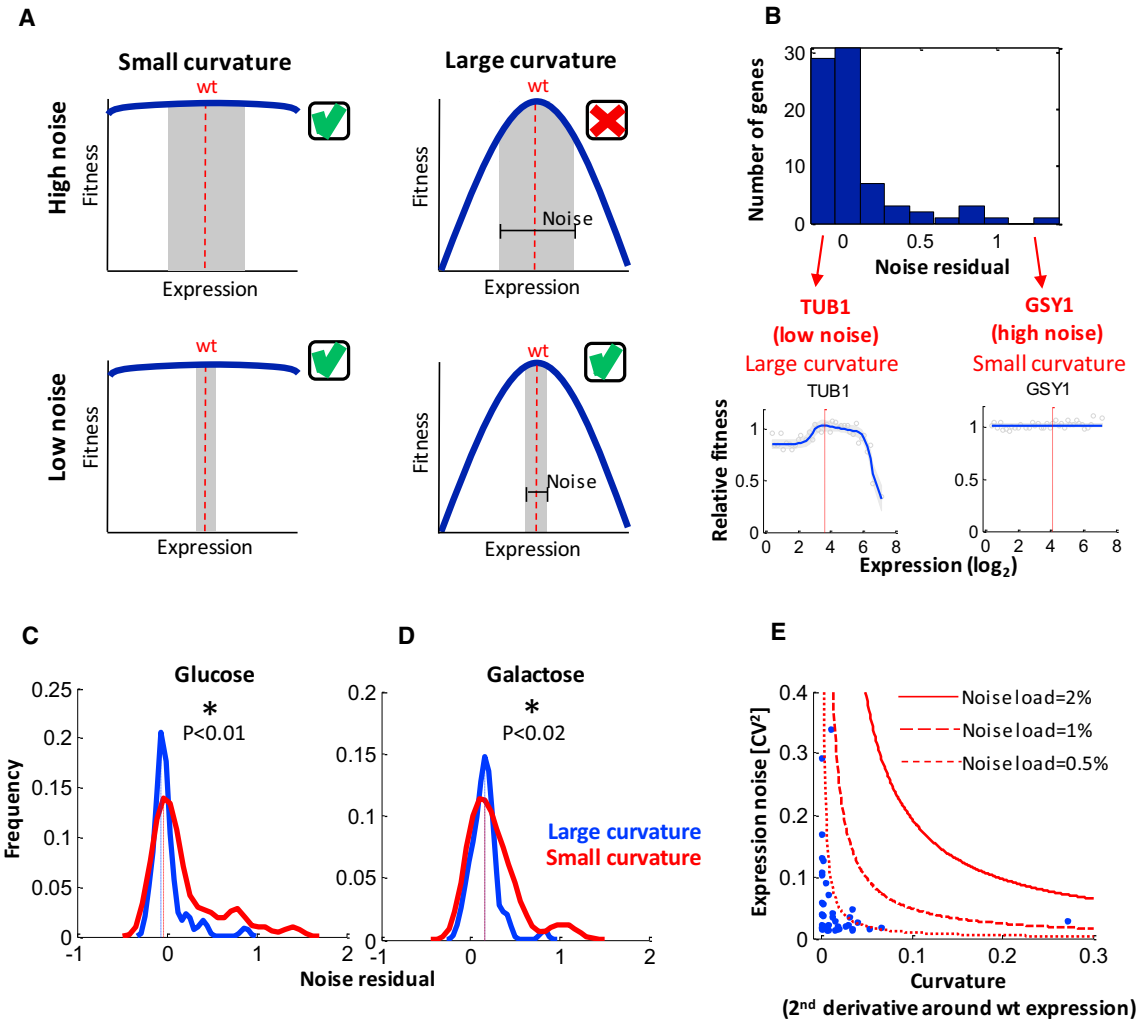


Figure 6. Noise in Expression Is Anti-correlated to the Curvature around Wild-Type Expression

(A) A suggestion for the connection between wild-type noise levels and the shape of the fitness curve. Hypothetical fitness-to-expression curves with small or large curvature around wild-type expression are plotted and marked by a red dashed line. Noise is depicted as the shaded gray area, marking the fluctuations around wild-type expression in the population. A combination of large curvature and high noise results in a reduction in the average fitness of the population.

(B) A histogram of the noise residuals in glucose (Keren et al., 2015) for the 81 genes essayed in this study depicting large differences in noise between different genes. For the genes *GSY1* and *TUB1*, their relative fitness (y axis) as a function of expression (x axis) in glucose are shown as in Figure 1F. The curve for *GSY1* is flat, while the curve for *TUB1* is curved around wild-type expression levels. Arrows denote the locations of the two genes in the noise histogram.

(C and D) All genes were classified, based on their fitness-to-expression curves as having large or small curvature around wild-type expression levels. Densities depicting the noise residuals for both groups in glucose (C) and galactose (D) are shown. High noise levels are less prevalent in genes with large-curvature fitness profiles, as evident by the tail of the distribution, which is larger for the small-curvature genes (red) than for the large-curvature genes (blue).

(E) For all genes in glucose, noise in expression (CV^2, y axis) as a function of their curvature (second derivative around wild-type expression, x axis) is shown. Genes are concentrated in the bottom-left corner and along the axes, and there is no combination of high curvature and high noise. Red lines mark equal noise load, achieved by different combinations of curvature and noise.

See also Figure S7.

cells display heterogeneity in their expression levels (noise), the magnitude of which varies between genes (Newman et al., 2006). We reasoned that such heterogeneity in expression around wild-type expression levels may result in a reduction in fitness, as protein abundance is farther away from the optimum on average. The degree of reduction depends both on the curvature of the fitness curve and on the noise and is maximized for a combination of large curvature and high noise (Figure 6A).

We extracted wild-type noise levels from a dataset in which we previously measured noise for ~900 native yeast promoters in the conditions assayed in this study (Keren et al., 2015). For each gene in each environment, we computed the noise residual, which is the noise of the gene corrected by the expectation from its mean (Figures 6B and S7A; STAR Methods). In addition, for each of our target genes, we calculated the fitness curvature around wild-type expression by computing the fold change

from wild-type expression that results in a 5% reduction in fitness (STAR Methods). We then systematically divided the genes into two groups based on the curvature around wild-type expression levels. In both environmental conditions, genes with high curvature had lower noise (Figures 6C and 6D; Wilcoxon rank-sum test, glucose $p < 0.01$, galactose $p < 0.02$). These results were robust to different threshold choices and to changes in the measure of curvature (Figure S7; STAR Methods).

To determine whether a selective disadvantage could explain the depletion of genes with strong curvature and strong noise, we estimated the noise load—the average fitness penalty caused by fluctuations around optimal expression (Kalisky et al., 2007). By expanding the fitness function around wild-type expression and averaging over fluctuations, one can compute an expression of the noise load $N = (1/2) \times (-f'') \times CV^2$ (STAR Methods), where f'' is the second derivative of the fitness function at wild-type expression and measures the curvature, and CV^2 is the noise. Plotting noise as a function of curvature for all genes (Figure 6E), we find that genes display either high curvature or high noise, but a combination of both is absent. We did not find a single gene for which the noise load is >2% and most genes have a noise load below 0.5%. These results suggest that evolutionary selection may have shaped the observed relationship between curvature and noise (Alon, 2006).

Our results suggest that the sensitivity of fitness to expression levels may serve as a selection force driving lower noise. Genes that are insensitive to changes in expression may be exempt from these selection forces, allowing for their high noise. The degree to which noise levels are indeed under selection and the environmental conditions favoring such selection are yet unknown and constitute a promising focus of future research.

DISCUSSION

In this work, we present a high-throughput method that allows sensitive and systematic interrogation of the effect of expression levels on cellular fitness for many genes simultaneously. Applying our method to a representative selection of yeast genes under two environmental conditions, we found that the effect of gene expression levels on fitness is both gene and environment specific, and the exact quantitative relationship between expression and fitness is a powerful indicator of gene function. We demonstrate that analyzing these relationships within the context of the relevant complexes, pathways, and cellular processes yields information on their functionality, stoichiometry and interplay in different conditions.

Over- and under-expression have a fundamental role in biological research (Gelperin et al., 2005; Giaever et al., 2002; Kemmeren et al., 2014; Sopko et al., 2006). A key advantage of our approach is the high coverage of the entire gene expression spectrum. This allowed us to measure expression levels higher and lower than wild-type in the same experiment and identify functional groups of genes more sensitive to changes in one or the other direction. In addition, it allows us to assay the entire gene expression space with high sensitivity. We found that while different groups of genes may exhibit similar fitness values at extreme expression levels, the shape of the fitness curve at intermediate expression levels may vary drastically. Finally, this experimental system enabled us to investigate fitness around wild-type expression levels and to discover an anti-correlation between sensitivity to changes in expression and noise, which, in agreement with previous works (Metzger et al., 2015), provides insights into selection forces fine-tuning wild-type expression properties.

Our results also pave the way to address fundamental questions of yeast ecology and biotechnology. By evaluating the degree to which wild-type expression levels are optimized for fitness in the different environments we can deduce the relative importance of these environments for overall organismal fitness. For example, we found that wild-type gene expression of laboratory yeast maximizes growth on glucose, whereas it is possible to tune gene expression to increase growth on galactose. Quantifying such data for more environmental conditions may shed light on their relative frequencies and durations in yeast ecology and reveal the constraints shaping wild-type expression levels. From a biotechnological standpoint, such approaches can be used to guide bioengineering of strains toward desirable phenotypes. For example, here we suggest 15 targets for engineering yeast for growth on galactose. Galactose is the major sugar in the abundant marine Macroalgae, and thus optimizing the production of biofuels, such as ethanol, from galactose may have important applications. As genetic engineering experiments are long and laborious, prior knowledge for potential targets to modify may accelerate such endeavors.

Our study lays a foundation for subsequent investigations, for example into the combinatorial effect of changes in the expression levels of different genes (Costanzo et al., 2010). Another challenge will be to mechanistically explain the quantitative magnitude of the effects, for example by profiling genome-wide expression patterns for these strains (Kemmeren et al., 2014). The library of strains presented can now serve as a tool to interrogate the dependence of phenotype on expression for these genes for any quantifiable trait (e.g., cell size, resistance to stress, etc.). Similar approaches can be applied in other organisms and different cell types for any assay yielding a quantifiable output (e.g., proliferation, apoptosis, response to drugs, etc.). Given the increasing speed and ease at which expression data is being generated, it is crucial to advance our ability to quantitatively link expression to function.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Synthetic Promoters and Expression Measurements
  - Construction of Strains in which Synthetic Promoters Drive Expression of Genes
  - Validations of Construction Process
  - Wild-Type
  - Media

- Expression Measurements of Isolated Strains by Quantitative RTPCR
- Isolated Fitness Assay by Pairwise Competition with Wild-Type in a Platereader
- Isolated Growth Assay by Colony Size
- Pooled Competition Assay
- Biological Replicate
- Preparing Samples for Sequencing

- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Data Analysis
  - Calculation of Relative Fitness from Sequencing Data
  - Curve Fitting of Impulse Model
  - Clustering and Analysis of Gene Groups
  - Fitness at Wild-Type Expression Levels
  - Optimality in Wild-Type Expression
  - Analysis of Fitness and Expression Noise
  - Calculating the Noise Load
  - Accounting for Clonal Interference
- DATA AND SOFTWARE AVAILABILITY
  - Data Resources

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2016.07.024.

AUTHOR CONTRIBUTIONS

L.K. conceived the project and experiments, designed the library, constructed the strains, performed experiments, analyzed the data and wrote the manuscript. J.H. analyzed the noise load and contributed to analysis. M.L.-P. constructed the strains. I.V.S. constructed and measured isolated strains. H.A. constructed the strains and performed experiments. S.K. performed experiments. A.W. contributed to the experiments. U.A. provided comments. R.M. provided comments. E.S. conceived the project and experiments, supervised the work, and wrote the manuscript.

ACKNOWLEDGMENTS

This work was supported by grants from the European Research Council (ERC) and the NIH to E.S. and student grants from the Kahn Center and Azrieli Center for Systems Biology to L.K. We thank Michal Levo, Shira Weingarten-Gabbay, Dan Davidi, Niv Antonovski, Martin Mikl, David Zeevi, and Eran Kotler for valuable input.

Received: May 16, 2016
Revised: July 5, 2016
Accepted: July 18, 2016
Published: August 18, 2016

REFERENCES

Alon, U. (2006). An Introduction to Systems Biology: Design Principles of Biological Circuits (CRC Press).

Bar-Even, A., Paulsson, J., Maheshri, N., Carmi, M., O'Shea, E., Pilpel, Y., and Barkai, N. (2006). Noise in protein expression scales with natural protein abundance. Nat. Genet. *38*, 636–643.

Bauer, C.R., Li, S., and Siegal, M.L. (2015). Essential gene disruptions reveal complex relationships between phenotypic robustness, pleiotropy, and fitness. Mol. Syst. Biol. *11*, 773.

Bernstein, M., Hoffmann, W., Ammerer, G., and Schekman, R. (1985). Characterization of a gene product (Sec53p) required for protein assembly in the yeast endoplasmic reticulum. J. Cell Biol. *101*, 2374–2382.

Blecher-Gonen, R., Barnett-Itzhaki, Z., Jaitin, D., Amann-Zalcenstein, D., Lara-Astiaso, D., and Amit, I. (2013). High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. Nat. Protoc. *8*, 539–554.

Borneman, A.R., Leigh-Bell, J.A., Yu, H., Bertone, P., Gerstein, M., and Snyder, M. (2006). Target hub proteins serve as master regulators of development in yeast. Genes Dev. *20*, 435–448.

Bro, C., Knudsen, S., Regenberg, B., Olsson, L., and Nielsen, J. (2005). Improvement of galactose uptake in Saccharomyces cerevisiae through overexpression of phosphoglucomutase: example of transcript analysis as a tool in inverse metabolic engineering. Appl. Environ. Microbiol. *71*, 6465–6472.

Chambers, A., Packham, E.A., and Graham, I.R. (1995). Control of glycolytic gene expression in the budding yeast (Saccharomyces cerevisiae). Curr. Genet. *29*, 1–9.

Chechik, G., Oh, E., Rando, O., Weissman, J., Regev, A., and Koller, D. (2008). Activity motifs reveal principles of timing in transcriptional control of the yeast metabolic network. Nat. Biotechnol. *26*, 1251–1259.

ENCODE Project Consortium (2004). The ENCODE (ENCyclopedia Of DNA Elements) Project. Science *306*, 636–640.

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L.Y., Toufighi, K., Mostafavi, S., et al. (2010). The genetic landscape of a cell. Science *327*, 425–431.

de Jongh, W.A., Bro, C., Ostergaard, S., Regenberg, B., Olsson, L., and Nielsen, J. (2008). The roles of galactitol, galactose-1-phosphate, and phosphoglucomutase in galactose-induced toxicity in Saccharomyces cerevisiae. Biotechnol. Bioeng. *101*, 317–326.

Dekel, E., and Alon, U. (2005). Optimality and evolutionary tuning of the expression level of a protein. Nature *436*, 588–592.

DeRisi, J.L. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. Science *278*, 680–686.

Deutschbauer, A.M., Jaramillo, D.F., Proctor, M., Kumm, J., Hillenmeyer, M.E., Davis, R.W., Nislow, C., and Giaever, G. (2005). Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. Genetics *169*, 1915–1925.

Dykhuizen, D.E., Dean, A.M., and Hartl, D.L. (1987). Metabolic flux and fitness. Genetics *115*, 25–31.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics *10*, 48.

Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A.S., Zink, F., Zhu, J., Carlson, S., Helgason, A., Walters, G.B., Gunnarsdottir, S., et al. (2008). Genetics of gene expression and its effect on disease. Nature *452*, 423–428.

Fendt, S.-M., and Sauer, U. (2010). Transcriptional regulation of respiration in yeast metabolizing differently repressive carbon substrates. BMC Syst. Biol. *4*, 12.

Gelperin, D.M., White, M.A., Wilkinson, M.L., Kon, Y., Kung, L.A., Wise, K.J., Lopez-Hoyo, N., Jiang, L., Piccirillo, S., Yu, H., et al. (2005). Biochemical and genetic analysis of the yeast proteome with a movable ORF collection. Genes Dev. *19*, 2816–2826.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., et al. (2002). Functional profiling of the Saccharomyces cerevisiae genome. Nature *418*, 387–391.

Gietz, R.D., and Schiestl, R.H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat. Protoc. *2*, 31–34.

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647–661.

Gill, G., and Ptashne, M. (1988). Negative effect of the transcriptional activator GAL4. Nature *334*, 721–724.

Gossen, M., and Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc. Natl. Acad. Sci. USA *89*, 5547–5551.

Hinnebusch, A.G. (1997). Translational regulation of yeast GCN4. A window on factors that control initiator-trna binding to the ribosome. J. Biol. Chem. *272*, 21661–21664.

Holden, H.M., Rayment, I., and Thoden, J.B. (2003). Structure and function of enzymes of the Leloir pathway for galactose metabolism. J. Biol. Chem. *278*, 43885–43888.

Hong, K.-K., Vongsangnak, W., Vemuri, G.N., and Nielsen, J. (2011). Unravelling evolutionary strategies of yeast for improving galactose utilization through integrated systems level analysis. Proc. Natl. Acad. Sci. USA *108*, 12179–12184.

Kafri, M., Metzl-Raz, E., Jona, G., and Barkai, N. (2016). The cost of protein production. Cell Rep. *14*, 22–31.

Kalisky, T., Dekel, E., and Alon, U. (2007). Cost-benefit theory and optimal design of gene regulation functions. Phys. Biol. *4*, 229–245.

Kemmeren, P., Sameith, K., van de Pasch, L.A.L., Benschop, J.J., Lenstra, T.L., Margaritis, T., O'Duibhir, E., Apweiler, E., van Wageningen, S., Ko, C.W., et al. (2014). Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. Cell *157*, 740–752.

Keren, L., Zackay, O., Lotan-Pompan, M., Barenholz, U., Dekel, E., Sasson, V., Aidelberg, G., Bren, A., Zeevi, D., Weinberger, A., et al. (2013). Promoters maintain their relative activity levels under different growth conditions. Mol. Syst. Biol. *9*, 701.

Keren, L., van Dijk, D., Weingarten-Gabbay, S., Davidi, D., Jona, G., Weinberger, A., Milo, R., and Segal, E. (2015). Noise in gene expression is coupled to growth rate. Genome Res. Published online September 9, 2015. http://dx.doi.org/10.1101/gr.191635.115.

Kodadek, T. (1993). How does the GAL4 transcription factor recognize the appropriate DNA binding sites in vivo? Cell. Mol. Biol. Res. *39*, 355–360.

Lee, K.-S., Hong, M.-E., Jung, S.-C., Ha, S.-J., Yu, B.J., Koo, H.M., Park, S.M., Seo, J.-H., Kweon, D.-H., Park, J.C., and Jin, Y.S. (2011). Improved galactose fermentation of Saccharomyces cerevisiae through inverse metabolic engineering. Biotechnol. Bioeng. *108*, 621–631.

Metzger, B.P.H., Yuan, D.C., Gruber, J.D., Duveau, F., and Wittkopp, P.J. (2015). Selection on noise constrains variation in a eukaryotic promoter. Nature *521*, 344–347.

Newman, J.R.S., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L., and Weissman, J.S. (2006). Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature *441*, 840–846.

Ostergaard, S., Olsson, L., Johnston, M., and Nielsen, J. (2000). Increasing galactose consumption by Saccharomyces cerevisiae through metabolic engineering of the GAL gene regulatory network. Nat. Biotechnol. *18*, 1283–1286.

Papp, B., Pál, C., and Hurst, L.D. (2003). Dosage sensitivity and the evolution of gene families in yeast. Nature *424*, 194–197.

Perfeito, L., Ghozzi, S., Berg, J., Schnetz, K., and Lässig, M. (2011). Nonlinear fitness landscape of a molecular pathway. PLoS Genet. *7*, e1002160.

Pierce, S.E., Davis, R.W., Nislow, C., and Giaever, G. (2007). Genome-wide analysis of barcoded Saccharomyces cerevisiae gene-deletion mutants in pooled cultures. Nat. Protoc. *2*, 2958–2974.

Rest, J.S., Morales, C.M., Waldron, J.B., Opulente, D.A., Fisher, J., Moon, S., Bullaughey, K., Carey, L., and Dedousis, D. (2012). Nonlinear fitness consequences of variation in expression level of a eukaryotic gene. Mol. Biol. Evol. *30*, 448–456.

Sellick, C.A., Campbell, R.N., and Reece, R.J. (2008). Galactose metabolism in yeast-structure and regulation of the leloir pathway enzymes and the genes encoding them. Int. Rev. Cell Mol. Biol. *269*, 111–150.

Shalem, O., Carey, L., Zeevi, D., Sharon, E., Keren, L., Weinberger, A., Dahan, O., Pilpel, Y., and Segal, E. (2013). Measurements of the impact of 3′ end sequences on gene expression reveal wide range and sequence dependent effects. PLoS Comput. Biol. *9*, e1002934.

Sharon, E., Kalma, Y., Sharp, A., Raveh-Sadka, T., Levo, M., Zeevi, D., Keren, L., Yakhini, Z., Weinberger, A., and Segal, E. (2012). Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. Nat. Biotechnol. *30*, 521–530.

Sopko, R., Huang, D., Preston, N., Chua, G., Papp, B., Kafadar, K., Snyder, M., Oliver, S.G., Cyert, M., Hughes, T.R., et al. (2006). Mapping pathways and phenotypes by systematic gene overexpression. Mol. Cell *21*, 319–330.

Steinmetz, L.M., Scharfe, C., Deutschbauer, A.M., Mokranjac, D., Herman, Z.S., Jones, T., Chu, A.M., Giaever, G., Prokisch, H., Oefner, P.J., and Davis, R.W. (2002). Systematic screen for human disease genes in yeast. Nat. Genet. *31*, 400–404.

Stoebel, D.M., Dean, A.M., and Dykhuizen, D.E. (2008). The cost of expression of Escherichia coli lac operon proteins is in the process, not in the products. Genetics *178*, 1653–1660.

Tong, A.H.Y., and Boone, C. (2006). Synthetic genetic array analysis in Saccharomyces cerevisiae. Methods Mol. Biol. *313*, 171–192.

Winsor, B., and Schiebel, E. (1997). Review: an overview of the Saccharomyces cerevisiae microtubule and microfilament cytoskeleton. Yeast *13*, 399–434.

Zampar, G.G., Kümmel, A., Ewald, J., Jol, S., Niebel, B., Picotti, P., Aebersold, R., Sauer, U., Zamboni, N., and Heinemann, M. (2013). Temporal system-level organization of the switch from glycolytic to gluconeogenic operation in yeast. Mol. Syst. Biol. *9*, 651.

Zeevi, D., Sharon, E., Lotan-Pompan, M., Lubling, Y., Shipony, Z., Raveh-Sadka, T., Keren, L., Levo, M., Weinberger, A., and Segal, E. (2011). Compensation for differences in gene copy number among yeast ribosomal proteins is encoded within their promoters. Genome Res. *21*, 2114–2128.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Yeast Nitrogen Base without Amino acids	BD	Cat#291940
Histidine	Sigma	Cat#H6034-25G
Hygromycin B (50mg/mL)	Thermo Fisher	Cat#10687010
PEG	Sigma	Cat#P3640-1KG
LiAc	Sigma	Cat#517992-100G
Salmon sperm	Sigma	Cat#D9156
Critical Commercial Assays		
YeaStar Genomic DNA Kit	ZYMO RESEARCH	Cat#D2002
Herculase II Fusion DNA polymerase	Agilent Technologies	Cat#600675
MasterPure- Yeast RNA Purification Kit	Epicentre	Cat#MPY03100
RBC Real Genomics HiYielsTM Plasmid Mini Kit	RBCBioscience	Cat#YPD100
KOD Hot Start DNA Polymerase	Novagen	Cat#71086
ZR-96 DNA Clean & ConcentratorTM-5	ZYMO RESEARCH	Cat#D4024
SexAI restriction enzyme	Thermo Scientific	Cat#FD2114
PspOMI restriction enzyme (Bsp120I)	Thermo Scientific	Cat#FD0134
QIAquick PCR Purification Kit	QIAGEN	Cat#28104
T4 PolynUcleotide Kinase	Thermo Scientific	Cat#EK0031
Lambda Exonuclease	Epicentre	Cat#LE032K
MinElute PCR Purification Kit	QIAGEN	Cat#28004
Qubit dsDNA HS Assay kit, 500 Assays	Invitrogen	Cat#Q32854
KAPA HiFi Hot Start Ready Mix PCR Kit	Novagen	Cat#KK2601
NextSeq 500 High Output v2 Kit (75 cycles)	Ilumina	Cat#FC-404-2005
SPRI beads Agencourt AMPure XP	Beckman Coulter	Cat#A63881
T4 Polynucleotide Kinase - 2,500 units 10,000 units/ml	NEW ENGLAND BioLabs Inc.	Cat#M0201L
T4 DNA Polymerase - 750 units 3,000 units/ml	NEW ENGLAND BioLabs Inc.	Cat#M0203L
Klenow Fragment (3'-5' exo-) - 1,000 units 5,000 units/ml	NEW ENGLAND BioLabs Inc.	Cat#M0212L
High Sensitivity D1000 ScreenTape	Agilent Technologies	Cat#5067-5584
Deposited Data		
Raw data files for DNA sequencing	NCBI Gene Expression Omnibus	GEO: GSE83936
Experimental Models: Organisms/Strains		
S. cerevisiae: Strain background: Y8205	Tong and Boone, 2006	
S. cerevisiae library: Y8205 can1delta ::STE2pr-Sp_his5 lyp1delta ::STE3pr-LEU2 his3delta1 leu2delta0 ura3delta0 Tef2pr-RFP-HygR-synthetic_promoter	This paper	
S. cerevisiae: BY4741	Euroscarf	Y00000
Recombinant DNA		
pAG32	Addgene	Cat#35122
AgilentLib Hiscore plasmid	Sharon et al., 2012	
pAG60	Addgene	Cat#35128
Sequence-Based Reagents		
Full list of primers is presented in Table S1		
Software and Algorithms		
Data analysis was done using Matlab 2013	Mathworks	

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents may be directed to, and will be fulfilled by the corresponding author Eran Segal (eran.segal@weizmann.ac.il).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All strains used in this study were derivatives of *S. cerevisiae* Y8205 (Tong and Boone, 2006), kindly provided by Charles Boone. Strains with native promoters driving yellow fluorescent protein (YFP) were previously described (Keren et al., 2013; Zeevi et al., 2011). Construction of strains with synthetic promoters driving endogenous genes are described below.

METHOD DETAILS

Synthetic Promoters and Expression Measurements

Synthetic promoters used in this study were previously described (Sharon et al., 2012). All promoters had an identical TATA-containing core promoter (~100bp), differed in their distal promoter elements and contained a unique 10bp molecular barcode. Promoters were integrated upstream of YFP, as previously described. All strains also carried the mCherry gene, driven by the constitutive *TEF2* promoter. From the original pool, 130 strains with promoters spanning an expression range of 500-fold in galactose and 200-fold in glucose with 5% differences in expression were isolated (Sharon et al., 2012). Strains were grown in 96-well plates with 150μl media with either galactose or glucose as carbon sources (see below). At mid-exponential YFP and mCherry were measured by flow cytometry. Expression was calculated as the log2 ratio of YFP to mCherry, and normalized such that the lowest expression level in galactose will equal zero (Figure S2A).

Construction of Strains in which Synthetic Promoters Drive Expression of Genes

Yeast transformations normally result in ~15% off-target integration (Keren et al., 2013; Zeevi et al., 2011). Thus, to ensure integration into the correct genomic location, the construction process of the strains was performed in several steps, as illustrated in Figure S1. In the first step a master strain was created for each target gene, in which half a selection marker (start of *URA3* gene) was inserted upstream of the target gene, followed by a temporary promoter. The resulting master strains were validated for correct integration by PCR and Sanger sequencing. In the second step, the promoter library was ligated to the complimentary half of the selection marker (end of *URA3* gene), amplified with primers containing gene-specific barcodes, mass transformed into the different master strains and selected on plates lacking uracil. Thus, only integrations in the correct site allowed for viable colonies. Integration was confirmed by PCR of 96 colonies, all of which contained the correct site of integration (Figure S2B).

Step 1: Construction of Three Master Plasmids

First, three master plasmids were constructed (Figure S1A) on the backbone of the previously described AgilentLib Hiscore plasmid (Sharon et al., 2012). Briefly, the plasmid contains a mCherry expression cassette followed by a *URA3* selection cassette. A hygromycin resistance cassette was amplified from commercial plasmid pAG32 with primers 5′-ACATACGATTTAGGTGACACTATA GAACGC 3′-AaagcttGCCGCATAGGCCACTAGTGGATC and inserted at HindIII sites. The promoters of *PPA2*, *RPA12* and *RPS13* were amplified from the genome of yeast strain BY4741 with addition of restriction sites for *PspOMI* (forward) and *SexaI* (reverse). Primers: PPA2 forward: ccagagccatgggcccAGGTTAAACATTATCAGAAATTATCTATTCCTA. PPA2 reverse: ccgaattgca cctggtGTTTTTCAGTAACTTCTCTAGAAGGG. RPA12 forward: ccagagccatgggcccTCACTGAATGTCACGATAGAGTTATCGCTG TAAAC. RPA12 reverse: ccgaattgcacctggtTCTTATAAGTACTAACCTGGATTGCTTTAAGCCT. RPS13 forward: ccagagccatgg gcccGTCTATCTTAAACACTAAACTACCTCCT. RPS13 reverse: ccgaattgcacctggtTTTGACTGATTGTTGTTGATTGATTT. Promoters were integrated at the PspOMI and SexaI sites within the URA3 gene, generating three master plasmids, driving either low, medium or high expression, as depicted in figure S1.

Step 2: Construction of 81 Master Strains

Each target gene was assigned to one of the three master plasmids, according to its wild-type expression level (table S3). The appropriate plasmid was chosen such that the expression of the temporary promoter matches as close as possible the expression of the target promoter; to ensure close-to-wt growth rates of all master strains. For each target gene, the mCherry-HygR-URtmp_Promoter fragment was lifted by PCR from the appropriate master plasmid using primers with tails complementary to the promoter (5′) and coding sequence (3′) of the target gene (See full list of primers in Table S1). Fragments were transformed into yeast strain Y8205 using the LiAc method, as previously described (Gietz and Schiestl, 2007), and genomically integrated upstream of the matching target gene, between the endogenous promoter and the coding sequence. Transformants were selected on YPD plates with hygromycin. This process resulted in 81 master strains, one for each target gene. Notably, the wild-type promoter remains intact and the entire cassette is inserted between the native promoter and the target gene. This entails minimal effect to neighboring genes. All master strains were verified by colony PCR (5′ primer- GGAGTTAGTTGAAGCATTAGGTCCC AAAATTTG, gene-specific 3′-primers are listed in Table S1) and Sanger sequencing to ensure integration at the correct genomic location (Figures S1 and S2).

Step 3: Construction of Barcoded Promoter Pools

From 6500 yeast strains carrying synthetic promoters previously described (Sharon et al., 2012), we chose 130 synthetic promoters which span a large range of expression values and have low levels of noise (Figure S2A). The barcode and promoter region from these isolated strains were amplified by PCR (5′- GTTTTTTGGGGACCAGGTGC 3′-CTTTGCCTTCGTTTATCTTGCC), pooled and cleaned (QIAquick PCR Purification Kit). Pooled promoters were amplified in 81 separate reactions using 81 pairs of target-specific primers (Table S1). The forward primer contained a sequence of 7 unique basepairs (bp), which serve as the barcode for the identity of the target gene. The reverse primer contained 40bp of homology to the coding sequence of the target gene. The second half of the *URA3* gene and its terminator (RA) were amplified from plasmid pAG60 using primers: 5′- GGAGTTAGTTGAAGCATTAGGTCCCA AAATTTG. 3′-CACCCAACCCACCAACAACACACCCCGACCCTGATGCGGTATTTTC. The RA fragment was ligated to the barcoded promoter pools using the PIE method as previously described (Zeevi et al., 2011), resulting in the complete promoter constructs (Figure S1).

Brief description of the PIE procedure: forward primers for amplification of promoters and RA fragment were phosphorylated (Thermo Scientific T4 PolynUcleotide Kinase). Barcoded promoter pools and RA fragment were separately amplified, cleaned (QIAGEN MinElute PCR Purification Kit) and digested into single strands (Epicenter Lambda Exonuclease). Barcoded promoter pools and RA fragment were merged and overlapping single strands were elongated (Novagen KOD Hot Start DNA Polymerase) and amplified (Novagen KOD Hot Start DNA Polymerase).

Step 4: Construction of Final Strains for Competition Experiment

Promoter constructs were transformed into their corresponding master strains using the LiAc method as described (Gietz and Schiestl, 2007; Zeevi et al., 2011), resulting in the replacement of the temporary promoters by the synthetic promoter constructs. Colonies were selected on Synthetic Complete Dextrose plates lacking uracil (SCD-Ura). Notably, only integration into the correct genomic location restores the entire *URA3* gene and leads to viable colonies. Selection was performed on plates to allow spatial separation between strains and thus provide an opportunity for less fit strains to be represented in the final pool. Colonies were grown for five days to allow time for colonies of less fit strains to arise. For each target gene 700-3000 colonies were collected, pooled and frozen at −80°c. Frozen pools from successful target genes were mixed in equal amounts and refrozen in aliquots of 2.25 OD (1 OD corresponds to $\sim 2x10^7$ cells).

Validations of Construction Process

Validation of correct promoter mapping was performed by sequencing the final transformants of two target genes (*LCB2* and *ABF1*). The prevalence of synthetic promoters was distributed normally, with ~5-fold difference between the most and least abundant promoters (Figure S2D). Mapping the reads of these single targets against the entire library allowed us to evaluate the error resulting from mistakes in barcode mapping, which we estimated to be $< 10^{-5}$.

Wild-Type

A wild-type strain was constructed similarly to all other strains in the library. In this strain the entire synthetic construct (mCherry-HygR-URA3-barcode_region) was inserted into a dispensable genomic location, within the can resistance cassette. This wild-type was created 1) to account for the fitness cost of integration and expression of the synthetic construct and 2) to allow its detection by next generation sequencing, in the same manner as for the rest of the strains in the library. The growth rate of the wild-type was assayed individually in the same environmental conditions as the pooled competition assay (see below), by measuring OD600. We found that the doubling time of the wild-type was 120 min in SD+His+Hyg and 126 min in SGal+His+Hyg. Including the wild-type in the pooled experiment and measuring its growth rate also in isolation allows to account for clonal interference in the pool and obtain individual fitness values from the pooled format as described in the analysis section below.

Media

Expression measurements and competition experiments were performed in either glucose (SD+His+Hyg: 6.7 g/L yeast nitrogen base (YNB), 2% glucose, 20mg/L His, 300 μg/ml Hygromycin) or galactose (SGal+His+Hyg: 6.7 g/L YNB, 2% galactose, 20mg/L His, 300 μg/ml Hygromycin) media.

Expression Measurements of Isolated Strains by Quantitative RTPCR

For the target gene *LCB2*, 50 isolated clones were constructed separately. Construction was as described above, but instead of pulling all synthetic promoters and doing all steps in a single tube, each synthetic promoter was amplified separately and construction proceeded separately for each promoter in a 96-well-plate. RT-qPCR was performed on 20 strains with primers 5′-ACTATACC CGTGTGCCCCTG 3′-TGGCTTCCCATGACGTGACT.

Strains were inoculated from frozen stocks into 5ml of either glucose or galactose media (described above) and grown at 30°C for 48 h, reaching complete saturation. Cells were then diluted 1:50 in fresh medium and pelleted at mid-exponential phase. RNA was extracted using the EPICENTER Yeast MasterPURE RNA extraction kit, and cDNA was created using random hexamers (sigma). Quantitative PCR was performed by RT–PCR (StepOnePlus, Applied Biosystems) using a ready-mix kit (KAPA, KK4605). For each strain, measurements were performed in two sets of triplicates, measuring both the tested gene and mCherry mRNA (primers

5′- TGTGGGAGGTGATGTCCAACTTGA 3′- AGATCAAGCAGAGGCTGAAGCTGA. Reported values are of mean gene/mCherry from nine replicates derived from three independent experiments.

Isolated Fitness Assay by Pairwise Competition with Wild-Type in a Platereader

50 strains in which a different synthetic promoter drives expression of *LCB2* were subjected to a head-to-head competition experiment with wild-type as previously described (Kafri et al., 2016). Briefly, all *LCB2* strains carry a cassette for expression of mCherry under the strong *TEF2* promoter. Strains were competed against a wild-type strain carrying a cassette for expression of YFP under the strong *HXK1* promoter previously described (Keren et al., 2013). The experiment was performed in three replicates in 96-well plates. In each well one *LCB2* strain was mixed with wild-type in a ratio of 10:1, as wild-type is expected to grow faster. Cells were grown in 180ul galactose media (see above) in 30°C for 5 days and every hour measurements of OD_{600}, YFP and mCherry were taken using a robotic system (Tecan Freedom EVO) with a plate reader (Tecan Infinite F500) as previously described (Keren et al., 2013). Every 24 hr wells were diluted 1:40 into fresh media. For each strain, relative fitness was calculated as the slope of a linear fit to the log2 ratio of mCherry to YFP over time.

Isolated Growth Assay by Colony Size

The growth rates of strains with synthetic promoters driving expression of one of four target genes (*TUB1*, *TIM10*, *PFK2* and *SEC27*) on glucose was estimated from colony size as previously described (Costanzo et al., 2010). The pool of synthetic promoters was transformed upstream of these four genes as described above. Transformants of each target were plated on agar plates and grown for five days. Each colony represents a specific strain and colonies are spatially separated and therefore there is no competition or clonal interference. The size of the colony is indicative of the growth rate of the strain and fast-growing strains have larger colonies than slow-growing strains (Figure S3A). Colonies were then scraped from the plates to a pooled population. From these cells genomic DNA was extracted and the barcode region was sequenced, as described in detail in the 'Pooled competition assay' section. Strains that had larger colonies on the plate will have higher representation in this pool, which can be uncovered by sequencing the barcode region of the pool. For each target gene reads were mapped to the synthetic promoters and normalized by the abundance of the synthetic promoters in the pool before the transformation to correct for biases relating to differential representation of different promoters in the pool. Data was binned into 40 logarithmically-equally-spaced expression bins. The resulting normalized relative abundance of each strain reflects the colony size and thus the growth rate (Figure S3B).

Pooled Competition Assay

Pooled competition was performed as described (Pierce et al., 2007). Briefly, 2.25 OD of pooled cells were resuspended in 36ml of either glucose or galactose media. Cultures were grown at 30°C and shaken at 250rpm. After 13 hr of growth in glucose (or 22 in galactose) a time point 0 sample of 5 OD was taken for sequencing (see below) and the culture was diluted 32 fold into 50ml of fresh media to keep the cells in exponential growth. The process was repeated after 23 and 35 hr in glucose or 35 and 49 hr in galactose. To ensure complexity of strains, cultures contained $> 10^7$ cells throughout the experiment.

Biological Replicate

For ten genes (*GAL2,MSN2,RAP1,TUB2,HXT3,FAS1,RPL22A,RPL3,FBA1,PDC1*) a biological replicate of the experiment was performed in galactose media. Experimental times, volume of culture, OD and overall cell numbers were as described above, resulting in an approximately 10-fold increase in the number of cells of each strain over the original experiment which included all genes. The replicates were highly correlated ($R = 0.97$, $p < 10^{-10}$ Figure 1I, S3C), indicating that the data is very reproducible and that the number of cells in the experiment is sufficient to accurately detect fitness changes > 2%.

Preparing Samples for Sequencing

Genomic DNA (gDNA) was purified from all samples using the YeaStar DNA purification kit (Zymo). A fragment of 300bp, containing the strain and promoter barcodes was amplified from the gDNA by PCR. In order to maintain the complexity of the library amplified from gDNA, PCR reactions were carried out on a total of 1200ng of gDNA, an amount calculated to contain $> 10^7$ yeast genomes. To minimize erroneous representation of the strains in the final product due to PCR bias, 12 separate reactions of PCR were performed for each sample (each with 100ng gDNA) and reactions were limited to 24 cycles. PCR was performed with 3 alternatively barcoded 5′ primers, used either to barcode the different experimental time points, or to check for PCR-related biases (Figure S2E). Primers also included 5 random nucleotides at their 5′ to increase sequence complexity and facilitate cluster calling during sequencing. 5_1'- NNNNN*CTACT*TACGTGAACCATCACCCTAATCAA 5_2'- NNNNN*GAGTG*TACGTGAACCATCACCCTAATCAA 5_3'- NNNNN *TCCGA*TACGTGAACCATCACCCTAATCAA 3′- NNNNNCTTTGCCTTCGTTTATCTTGCC. Italic nucleotides denote the time point barcodes. Following PCR, products were united and separated from unspecific fragments by electrophoresis on a 2% agarose gel stained by EtBr, cut from the gel, and cleaned in 2 steps: gel extraction kit (QIAGEN) and SPRI beads (Agencourt AMPure XP). Concentration was measured using a monochromator (Tecan i-control). The samples were assessed for size and purity at the Tapestation, using high sensitivity D1K screenTape (Agilent Technologies). 50 ng DNA were used for library preparation for next generation sequencing; specific Illumina adaptors were added, and DNA was amplified using 8 amplification cycles, protocol

adopted from Blecher-Gonen et al. (Blecher-Gonen et al., 2013). Samples were run on the NextSeq (Illumina) with 75bp single-end kits and > 10^7 reads were obtained per sample.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis was performed using Matlab R2013a (Mathworks). Statistical parameters including precision measures (mean ± SD) and statistical significance are reported in the Main text, Figures and Figure Legends. Data is judged to be statistically significant when $p < 0.05$ by Wilcoxon ranksum test or student's t test, where appropriate. All tests for enrichment were FDR-corrected for multiple hypothesis testing. Pearson correlation coefficients (R) and p-values (P) were calculated using Matlab R2013a (Mathworks).

Data Analysis

NGS reads were mapped to a reference sequence set that contained all barcode combinations of promoter, target gene and experimental time point. Example of a read: TGGGG**TCCGA**TACGTGAACCATCACCCTAATCAA**GGGAGAT**GTTTTTTGGGGACCAGG TGCCGTAAG**CTACACATGT**. Read structure: Random nucleotides (5bp) - experimental time point barcode (5bp) – constant region (24bp) – target gene barcode (7bp) – constant region (26bp) – synthetic promoter barcode (10bp). Barcodes for time points are: 0- CTACT, 1- GAGTG, 2- TCCGA. Barcodes for synthetic promoters were detailed previously (Sharon et al., 2012). List of gene barcodes is in table S1.

Promoters which had overall low numbers of reads at the first experimental time point for all target genes (< 10^4), were removed from further analysis, resulting in 112 promoters in glucose and 120 promoters in galactose. For target genes of transcription factors, synthetic promoters carrying binding sites for these transcription factors were removed from further analysis, to avoid negative/positive feedback loops (Figure S2C). For each target gene the data was binned as follows: the expression space was partitioned into 36 (glucose) or 39 (galactose) logarithmically-spaced expression bins. For each bin the expression level was assigned to be the center of the bin and the reads number to be the median of all strains in the bin. Binning was performed to reduce technical noise and acquire equally-spaced points.

Calculation of Relative Fitness from Sequencing Data

For each strain *(x)*, in each environment *(a)* we calculate the relative fitness for growth $\boldsymbol{F}(\boldsymbol{x},\boldsymbol{a})$ according to:

$$F(x,a)=\frac{\log_2\left(\frac{r_{wt}(t_i)\cdot r_x(t_j)}{r_{wt}(t_j)\cdot r_x(t_i)}\right)}{g_{wt}}+1 \quad (1)$$

where $r_x(t_i)$ is the fraction of strain X in the population at time i and g_{wt} is the number of doublings performed by the wild-type between t_i and t_j. The relative fitness for strain x is the number of doublings performed by x, at the time that wild-type completed a single doubling. Thus, it is comparable between growth conditions. For example, $F(x)=0.9$ indicates that strain x performed 0.9 doublings during the time period that wild-type doubled once. Relative fitness values below 1 indicate slower growth than wt, whereas values above 1 indicate faster growth than wt.

Below we delineate the derivation of Equation 1:

The goal is to calculate the relative growth rates of strains growing in competition, from abundance of reads at two time points, t_1 and t_2.

We note two properties of the data:

1. The representation of the strains at the beginning of the experiment is not identical, and can vary roughly 10-fold.
2. We do not sequence the entire population. At each time a subset of fixed size is taken for the analysis. Thus, we obtain the relative abundances of the strains.

The relative abundance of a strain is dependent not only on the growth rate of this particular strain, but also on the growth rates of other strains in the pool. Therefore, we use knowledge of the known doubling time of the wt, to use as a fixed reference point.

We define:

N – The number of strains in the population.

$R_x(t_1)$ – The reads of strain X at time t_1.

$r_x(t_1)$ – The fraction of strain X in the population at time t_1.

g_x – The number of generations of strain X. This information is known for the *wt*, whose doubling time was measured independently, and should be extracted for all other strains in the competition experiment.

Based on the *wt* reads at t_1, and the known number of generations performed by a *wt* strain between t_1 and t_2, we can calculate the expected number of *wt* reads at t_2, $\widehat{R}_{wt}$, if we were to sample the entire population:

$$R_{wt}(t_1)\cdot 2^{g_{wt}}=\widehat{R}_{wt}(t_2) \quad (2)$$

From the fraction of *wt* reads, we can then calculate the population size at t_2

$$\sum_{x=1}^{N} R_x(t_2) = \widehat{R}_{wt}(t_2)/r_{wt}(t_2) \tag{3}$$

This can now be used to find the number of generations of any strain *X*. Similarly to Equation (2) we write:

$$R_x(t_1)\cdot 2^{g_x} = \widehat{R}_x(t_2) \tag{4}$$

$$R_x(t_1)\cdot 2^{g_x} = \sum_{x=1}^{N} R_x(t_2)\cdot r_x(t_2) \tag{5}$$

We can rewrite (5) as:

$$g_x = \log_2\left(\frac{\sum_{x=1}^{N} R_x(t_2)\cdot r_x(t_2)}{R_x(t_1)}\right) \tag{6}$$

Plugging (2) and (3) into (6) we get:

$$g_x = \log_2\left(\frac{\frac{R_{wt}(t_1)\cdot 2^{g_{wt}}}{r_{wt}(t_2)}\cdot r_x(t_2)}{R_x(t_1)}\right) = \log_2\left(\frac{R_{wt}(t_1)\cdot 2^{g_{wt}}\cdot r_x(t_2)}{r_{wt}(t_2)\cdot R_x(t_1)}\right) = \log_2\left(\frac{r_{wt}(t_1)\cdot 2^{g_{wt}}\cdot r_x(t_2)}{r_{wt}(t_2)\cdot r_x(t_1)}\right) = \log_2\left(\frac{r_{wt}(t_1)\cdot r_x(t_2)}{r_{wt}(t_2)\cdot r_x(t_1)}\right) + \log_2(2^{g_{wt}})$$
$$= \log_2\left(\frac{r_{wt}(t_1)\cdot r_x(t_2)}{r_{wt}(t_2)\cdot r_x(t_1)}\right) + g_{wt} \tag{7}$$

The relative fitness of strain *X* is defined as:

$$F(x) = \frac{g_x}{g_{wt}} \tag{8}$$

It is the number of generations performed by strain *X* during one *wt* generation.
Plugging (7) into (8) we get Equation 1 for the fitness of strain *X*:

$$F(x) = \frac{\log_2\left(\frac{r_{wt}(t_1)\cdot r_x(t_2)}{r_{wt}(t_2)\cdot r_x(t_1)}\right)}{g_{wt}} + 1$$

For all strains in both conditions relative fitness was calculated according to Equation 1 between time points 0 and 2 (22 and 49 hr in galactose, 13 and 35 in glucose). The timeframe of the experiment sets the detection level. Growth experiments involve an inherent tradeoff between detection of very low fitness and detection of very mild fitness effects around wild-type (Pierce et al., 2007). Here we estimate our detection limit at 0.6 for galactose and 0.5 for glucose. In Figure 3B, data after 13 hr of growth is shown to allow detection of lower fitness.

Curve Fitting of Impulse Model

Fitness-to-expression data was fitted with a parametric impulse function (Chechik et al., 2008), which fits two sigmoidal functions to the data. Briefly, we fitted the function as described by Chechick et al. (Chechik et al., 2008), but added an additional parameter, β2, such that the slopes for the two sigmoids may differ; resulting in a total of seven parameters: the initial level (h0), the intermediate level (h1), the final level (h2), the position of the first and second transition (t1 and t2) and the transition slopes (β1 and β2). Together these parameterize an impulse function where:

$$f(x) = \frac{1}{h_1} s_1(x) s_2(x)$$

$$s_1(x) = h_1 + (h_1 - h_0)\mathit{Sigmoid}(\beta_1, t_1)$$

$$s_2(x) = h_2 + (h_2 - h_1)\mathit{Sigmoid}(\beta_2, t_2)$$

$$\mathit{Sigmoid}(\beta, t) = 1/\exp(-\beta(x - t))$$

To avoid overfitting and reduce the effects of outliers, a 10-fold cross-validation scheme was applied, in which for each target gene the function was fitted using 90% of the data, and then evaluated on the remaining 10%. Function was fitted using non-linear regression (matlab nlinfit) and confidence intervals for the fit were obtained using nlpredci. To avoid local solutions, for each target gene, this procedure was repeated with five different starting parameter points and the overall best fit was selected. The impulse function gives good predictions for most strains (Pearson R = 0.86, $p < 10^{-10}$, Figure S2I,S4,S5). However, it does not capture bimodal curves, such

as the one observed for *FBA1* (Figure 2A). Therefore, for *FBA1* a non-parametric approach was applied, by which loess regression was used to smooth the fitness data (matlab smooth).

Clustering and Analysis of Gene Groups

Fitness-to-expression curves in both galactose and glucose were standardized to have a mean of 0 and SD of 1 and hierarchical clustering was performed (matlab clustergram). GO enrichment analysis was performed using Gorilla (Eden et al., 2009). All enrichment tests were FDR corrected for multiple hypothesis testing.

Fitness at Wild-Type Expression Levels

Relative fitness is defined as relative growth rate compared to wild-type. Thus, by definition, it is expected that the fitness curve will equal one at wild-type expression. However, for several genes, we obtain fitness functions in which wild-type expression levels result in different fitness than one, usually below one (Figure S2H). This behavior may be the result of several factors, or a combination of them. First, as wild-type expression is measured by fluorescence in a different experiment, in which the promoter is placed in a different genomic location, driving YFP, there may be changes between our estimated value for wild-type expression and the actual value. Such discrepancies may also occur if the gene is heavily regulated post-transcriptionally. Since our synthetic promoters also replace the 5′UTR of the gene by a constant region (Sharon et al., 2012), if there is significant regulation acting on this sequence, we may be inaccurate in our placement of wild-type expression relative to the expression levels of the synthetic promoters. This may be the case for *GCN4* (Figure S4), which is known to be translationally regulated by uORFs in its 5′UTR (Hinnebusch, 1997).

Genes that are subjected to either autoregulation or exert positive or negative feedback over the synthetic promoters may also deviate between the estimated and actual values of expression of either the synthetic or wild-type promoters. One example for such negative feedback occurred for the transcriptional repressor *GAL80*. When plotting the raw data for *GAL80* we observed that low expression resulted in high fitness, high expression resulted in low fitness, and there was a large, nearly bimodal spread of fitness values for intermediate expression levels (Figure S2C). Coloring the strains by the identity of their synthetic promoters revealed a clear separation in fitness between promoters that have a binding site for *GAL80* and those that do not (compare blue and red strains in Figure S2C). We hypothesize that the strains that carry a synthetic promoter with the *GAL80* binding site underwent negative autoregulation. These promoters may drive intermediate expression levels when upstream of YFP, however when driving expression of *GAL80*, a negative feedback loop is generated, as the gal80 protein represses the promoters. Thus, specifically for *GAL80*, all synthetic promoters containing sites for gal4/gal80 were removed from the analysis.

Noise in expression may also contribute to reducing fitness levels (see analysis in Figure 6 and STAR Methods). For genes with high curvature around wild-type, e.g., high sensitivity to noise, this may result in curves for which the entire fitness curve is below one, as our essay measures for each genetically-identical cell the average fitness in the population. This may be the case for *TUB2* (Figure S4), which exhibits very high curvature, its estimated wild-type expression results in the highest value in the fitness curve, but for which the entire fitness function is below one.

For such genes, while we have high confidence in the shape of the curve, we have lower confidence in the location of wild-type on it. Thus, genes for which the fitness at wild-type expression was very different from one (22 genes in glucose and 25 in galactose) were discarded from all relevant analyses. To facilitate visualization, the fitness axis was shifted such that the fitness at wild-type expression equals one. Original curves are depicted in Figures S4 and S5.

Optimality in Wild-Type Expression

For each gene the fitness at wild-type expression was calculated using the impulse function with appropriate parameters. Genes for which the fitness at wild-type expression was below 0.95 were discarded from further analysis. For the remaining genes, the fitness at the maximal and minimal expression levels was computed and compared to wt. If the fitness at maximal/minimal expression was over 2% higher than wt, the gene was classified as sub-optimal.

We note that the approach presented above for defining genes as sub-optimal is conservative and therefore probably does not have false positives, but may have false negatives. First, genes for which the fitness at wild-type expression was below 0.95 were discarded from further analysis. For such genes we cannot be certain whether indeed wild-type levels are not optimal for fitness in our tested condition or whether there exists an experimental discrepancy between the expression and fitness measurements. Second, we only examined the fitness of wild-type expression compared with extreme expression values. For such cases there exists a monotonic trend in the data (e.g., increased fitness with increased expression), and therefore we can be more certain that our results hold even if the measurement of wild-type expression is inexact. There exist several cases where the optimum of the curve is neither at wild-type nor at any extreme expression value, but rather at another intermediate value in the expression range (e.g., *MSN2* in galactose, Figure S4). Such cases were not taken into consideration by our analysis and thus may add to the pool of genes whose wild-type expression is sub-optimal in the tested conditions. Finally, the threshold of 2% above wild-type is also conservative as it was derived from the sensitivity of the system, without considering the trend along the curve, which leads to increased confidence in small changes. Altogether, the approach was conservative and the results presented may be an underestimation for the degree of non-optimality in both conditions.

Analysis of Fitness and Expression Noise

Noise data for all genes in the current study in both glucose and galactose was taken from Keren et al., 2015 (Keren et al., 2015). For both conditions, the noise versus mean data was fitted using local regression (matlab loess) to obtain the average dependence of mean on noise. For each gene the noise residual was calculated as the log10 difference between its actual noise levels and the average noise at its expression. The noise residual is used to correct for the global dependence of expression noise on mean expression (Bar-Even et al., 2006; Newman et al., 2006).

Two different measures were used to define the curvature of the fitness function around wild-type expression. The first is the second derivative of the fitness function at wild-type expression. This is the exact measure for curvature at wild-type, but due to its local nature may be insensitive to extreme changes in fitness that may occur near, but not exactly at, wild-type expression. Therefore, a second, long-range, measure for fitness was defined as the fold change from wild-type expression that results in a 5% reduction in fitness. Ranging the threshold between 2%–10% did not alter the results significantly (Figures S7C and S7D). To divide the genes into groups of high and low curvature, a threshold of 10-fold change in expression was used. Ranging the threshold between 5-20-fold did not alter the results significantly (Figures S7C and S7E).

Calculating the Noise Load

The noise load N is the reduction in fitness that results from having a population of cells with variable expression of a specific gene, rather than a homogenous population in which all cells express precisely the wild-type level. To compute it, we Taylor expand the fitness function f close to the wild-type expression:

$$f(p) = f_0 + f'(p - p^*) + \frac{1}{2}f''(p - p^*)^2$$

where p is protein abundance, and f_0 is the fitness at wt protein abundance p^*. For genes where wild-type expression provides optimal fitness, the first derivative f' is 0. Averaging fitness over protein fluctuations yields an expression for the average fitness:

$$\langle f(p)\rangle = f_0 + \frac{1}{2}f''\sigma^2$$

Where σ^2 is the variance of protein fluctuations. The noise load N is the average fitness loss:

$$N = f_0 - \langle f(p)\rangle = -\frac{1}{2}\frac{\partial^2 f}{\partial p^2}\sigma^2 = -\frac{1}{2}\frac{\partial^2 f}{\partial(\log p)^2}\left(\frac{\sigma}{p}\right)^2$$

Finally, because all analyses are done in log2 expression and not log, we rewrite the second derivative to obtain

$$N = -\frac{1}{2(\log 2)^2}\frac{\partial^2 f}{\partial(\log_2 p)^2}\left(\frac{\sigma}{p}\right)^2$$

or, put more compactly, $N = (1/2)\cdot(-f'')\cdot CV^2$

where we redefined f'' as the second derivative of the fitness function with respect to log expression, evaluated at wt expression. f'' quantifies the curvature, and CV^2 is the noise. For genes with very flat curves $f'' \to 0$ and the noise load will be small regardless of the noise of the genes. For such genes we do not expect selection to act to reduce noise and no correlation between curvature and noise is expected. However, for genes with high curvature around fitness expression $-f''$ is large and the noise load will depend on the noise of the gene (CV^2).

The analysis was restricted to genes in which fitness at wild-type expression is above 0.95 and thus there is high confidence in the shape of the curve and the position of wild-type expression. Genes for which there is positive curvature around wild-type (e.g., wild-type expression does not result in maximal fitness) were discarded from the analysis.

Accounting for Clonal Interference

In our experiment we conduct a pooled growth assay for 10,000 strains, determine the fractional abundance of each strain at different time points and use these fractional abundances to calculate the fitness of each strain. The relative abundance of a strain is dependent not only on the growth rate of this particular strain, but also on the growth rates of all other strains in the pool, a phenomenon known in evolutionary experiments as 'clonal interference'. In our work we account for clonal interference by: 1) including the wild-type in the experiment and 2) conducting an additional experiment in which we measure the wild-type's growth rate when it is grown alone in the tested environment. The full mathematical account is presented above. Intuitively, we know that the wild-type fitness is 1 by definition. As such, if we observe variations in the fractional abundance of the wild-type over time we know that this is the result of the growth rates of other strains in the pool. For example, if the fractional abundance of the wild-type decreases, this may be the result of 'super growers', which increase their abundance in the population. By measuring the wild-type's growth rate in a non-competition setting and including it in the experiment, we provide our system with a reference point that reveals these 'population effects' and allows to deduce the true fitness of the strain.

To demonstrate how our experimental design and analysis mitigate concerns for confounding of the results by clonal interference, we simulate a population with predefined growth rates and test whether we can reconstruct these growth rates from fractional abundances after time *t*, as we do in the original experiment.

The simulated population consists of three strains: wt (doubling time = 2h), a strain with poor fitness (doubling time = 4h), and a strain with increased fitness (doubling time = 1h). Thus the relative fitness we extract from simulated data should be 1, 0.5 and 2 respectively. For simplicity, we assume that we start with equal abundances of the three strains in the population (although the equation derived above holds for the more general case in which abundances are not equal at the start). At the beginning of the experiment, each strain comprises 0.33 of the population. Assuming exponential growth, after an experiment of 6 hr the fractional abundances of the strains will be 0.11, 0.04 and 0.86 respectively (see more details in the table below). If we were to use these numbers alone to estimate the relative fitness of the three strains, we would indeed be lead to think that the second strain has poor fitness since it decreased in its abundance from 0.33 at the start of the experiment to 0.11. However, knowing that this is the wild-type, we can now correct for the effect of clonal interference using the mathematical framework derived above. Plugging these numbers in Equation 1 for each of the three strains we can accurately retrieve the relative fitness from the fractional abundances of the strains:

$$F(slow) = \frac{\log_2\left(\frac{r_{wt}(t_1)\cdot r_x(t_2)}{r_{wt}(t_2)\cdot r_x(t_1)}\right)}{g_{wt}} + 1 = \frac{\log_2\left(\frac{0.33\ \cdot 0.04}{0.11\cdot 0.33}\right)}{6/2} + 1 = 0.5$$

$$F(wt) = \frac{\log_2\left(\frac{r_{wt}(t_1)\cdot r_x(t_2)}{r_{wt}(t_2)\cdot r_x(t_1)}\right)}{g_{wt}} + 1 = \frac{\log_2\left(\frac{0.33\cdot 0.11}{0.11\cdot 0.33}\right)}{6/2} + 1 = 1$$

$$F(fast) = \frac{\log_2\left(\frac{r_{wt}(t_1)\cdot r_x(t_2)}{r_{wt}(t_2)\cdot r_x(t_1)}\right)}{g_{wt}} + 1 = \frac{\log_2\left(\frac{0.33\cdot 0.86}{0.11\cdot 0.33}\right)}{6/2} + 1 = 2$$

growth	doubling time (h)	number of cells at time zero (arbitrary)	*fractional abundance at time zero*	number of cells after 6 hr (= 100 × $2^{Number\_of\_generations}$)	*fractional abundance after 6 hours*	Calculating relative fitness with Equation 1
slow	4	100	*0.33*	282.84	*0.04*	0.5
wild-type	*2*	100	*0.33*	800.00	*0.11*	1
fast	1	100	*0.33*	6400.00	*0.86*	2

A simulated culture of three strains grown in a pool. Information set in italics indicates data that we measure in the experiment. Information depicted in black underlies the observed distribution of fractional abundances, but is not available to the researchers at the time of the experiment. The last column depicts the relative fitness calculated according to Equation 1, and recapitulates the unknown differences in doubling time of the strains.

Altogether, despite the fact that this simulated culture shows clonal interference, including knowledge of the identity and doubling time of the wild-type allows us to retrieve accurate information on the relative fitness of the three strains from the pooled data.

DATA AND SOFTWARE AVAILABILITY

Data Resources

Raw data has been deposited to NCBI GEO: GSE83936. Processed data is available in Tables S2 and S3.

Cell

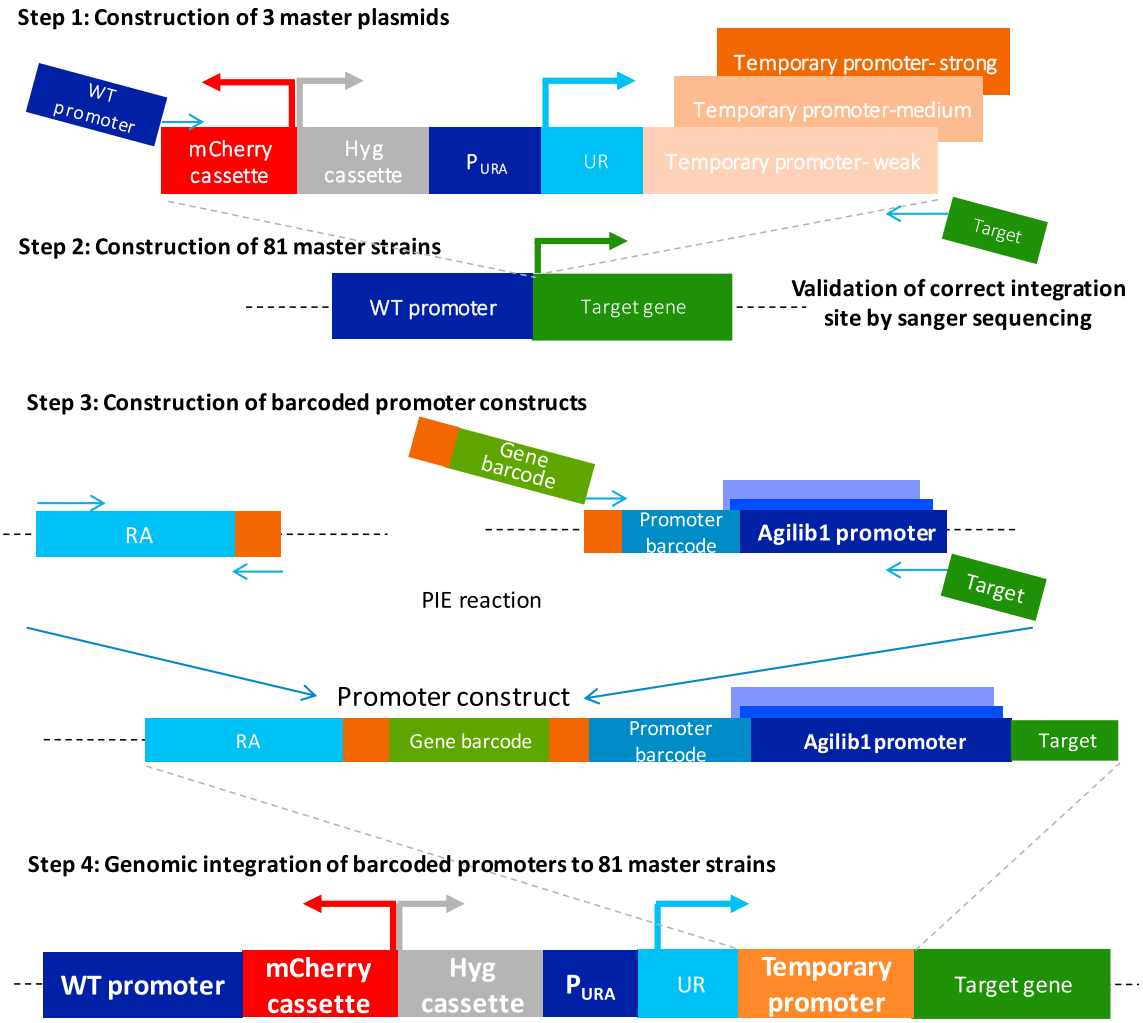
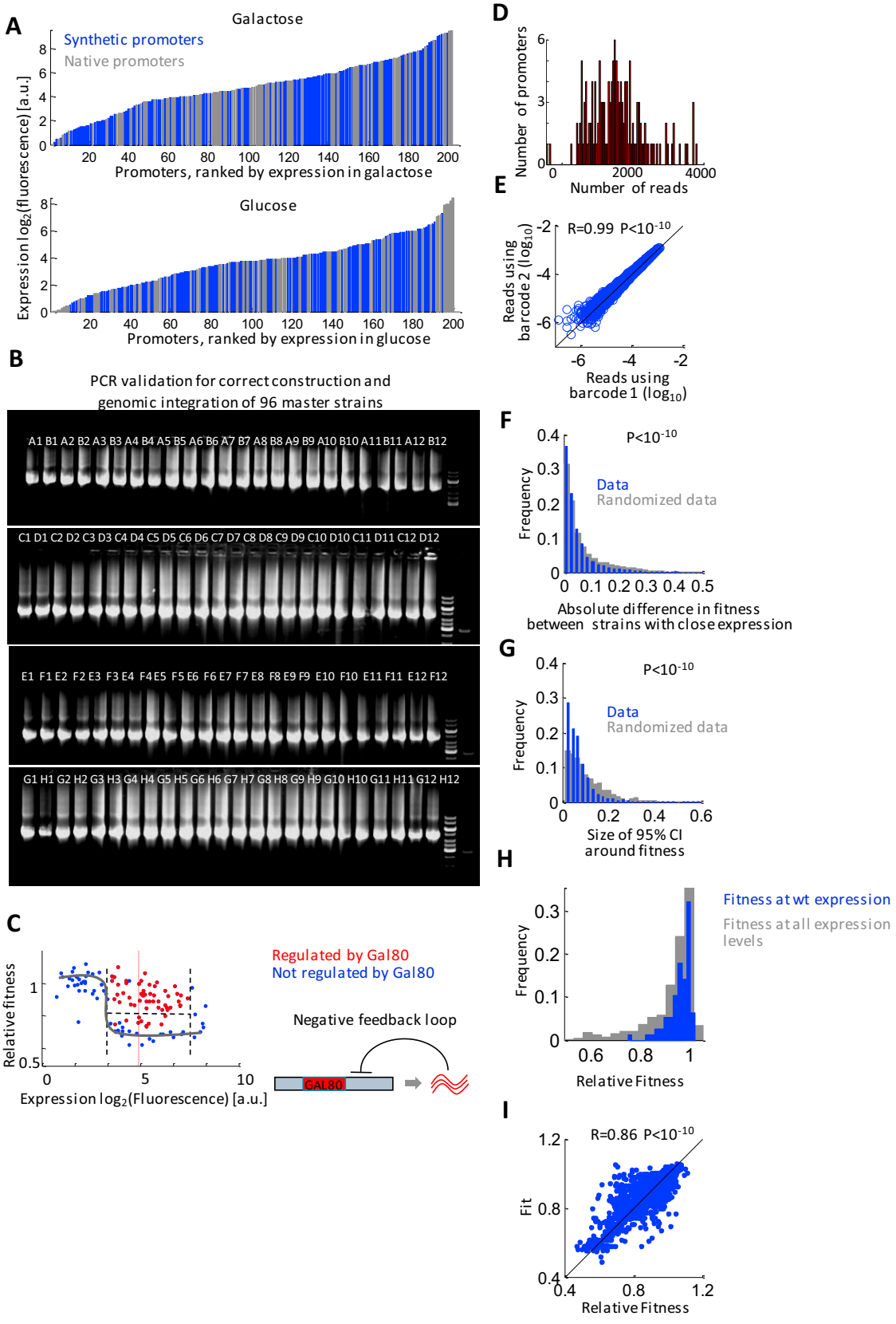


Figure S1. A Two-Step Construction Process Ensures Correct Genomic Integration, Related to Figure 1

Shown is a schematic of the strain construction process. Construction was performed in several steps to ensure correct site for genomic integration. Step 1: Construction of three master plasmids. The master plasmids encode for a mCherry expression cassette, a hygromycin selection cassette, the *URA3* gene promoter followed by half of the *URA3* gene and a temporary promoter (driving either low (*PPA2*pr), medium (*RPA12*pr) or high (*RPS13*pr) expression). Step 2: Construction of 81 master strains. All target genes evaluated in this study were partitioned into three groups according to their native expression levels, either low, medium or high. For each target gene the construct from the appropriate master plasmid was amplified using primers matching the end of the endogenous promoter and the start of the coding region, and was integrated upstream of the target gene, pushing back its endogenous promoter. Transformants were selected on rich (YPD) media supplemented with hygromycin and were validated for correct integration site and sequence by Sanger sequencing. This step resulted in the verified construction of 81 master strains, one per target gene. Step 3: Construction of barcoded promoter constructs. A pool of 130 synthetic barcoded promoters described by Sharon et al. (Sharon et al., 2012), were amplified using gene-specific primers. The forward primer contained 7 nucleotides with a barcode unique to each target gene and the reverse primer complemented the target gene's coding region. Promoters were then joined with the end of the *URA*3 gene by the previously-described PIE method (Zeevi et al., 2011) to obtain the final barcoded, pooled promoter constructs. Step 4: Genomic integration of barcoded promoter pools into the target strains. Each barcoded promoter pool was transformed into its appropriate master strain. Cells were selected on minimal media plates lacking uracil to ensure minimal competition by spatial separation. Cells were grown on plates for 5 days to allow colonies of slow-growing strains to appear. For each target gene at least 700 colonies were collected to ensure promoter complexity in the initial pool and glycerol stocks were frozen at −80°c.



(legend on next page)

Figure S2. Controls for the Construction Process and Measurements, Related to Figure 1

(A) Synthetic promoters span the physiologically-relevant range of expression in small increments. Each promoter in the study, either natively encoded by the yeast genome, or synthetically designed was fused upstream of yellow fluorescent protein (YFP) to drive its expression. Shown are the measured expression values (y axis) for all promoters, ranked by their expression (x axis), in media containing either galactose (top) or glucose (bottom) as carbon source. Native promoters are shown in gray and synthetic promoters in blue. The synthetic promoters span a range of ~500-fold in galactose and ~200-fold in glucose due to their design (Sharon et al., 2012). In galactose the synthetic promoters do not reach the levels of expression of the highest native promoters, for which only downregulation from wild-type expression levels could be evaluated in this study.

(B) Shown are PCR results for the 81 master strains (STAR Methods) validating correct genomic integration site.

(C) For the gene *GAL80* in galactose shown is the raw data depicting its relative fitness (y axis) as a function of expression (x axis). Strains that carry synthetic promoters with the *GAL1-10* context, which is known to be negatively regulated by gal80, are colored in red. All other promoters are colored blue. For intermediate expression levels there is a clear separation in the fitness values of the red and blue strains. Red dashed line depicts wild-type expression levels. Black dashed lines mark intermediate expression range, with a large span of fitness values. Cartoon depicts negative auto-regulation over promoters regulated by gal80, resulting in lower expression levels for these promoters when they drive *GAL80* than when they drive YFP.

(D) Shown is a histogram depicting the representation of the different synthetic promoters for a single target gene, *LCB2*, following the pooled construction process. Nearly all promoters are represented in the pool, and the difference in frequency between the most abundant promoter and the least abundant promoter is small enough to ensure adequate representation of the promoters at all experimental stages (~5-fold).

(E) Shown are the strains which have the same synthetic promoter driving the same target gene, but with different barcodes. For each strain shown are the number of reads obtained from sequencing of the pool for the first (x axis) and second (y axis) barcodes. Reads are highly correlated and the effect of the barcode is minor.

(F) A randomized dataset was generated by shuffling, for each target gene, the fitness values with respect to the expression values. For each gene, the absolute difference in fitness between neighboring expression values was computed for both datasets. Shown is a histogram for the difference values of the real (blue) and randomized (gray) data. Strains that have similar expression levels of their target genes tend to have similar fitness levels. P-value is for significance in Wilcoxon ranksum test.

(G) Shown is the distribution of confidence intervals (CI) for every point in the real (blue) and randomized (gray) datasets. For over 50% of the points the CI is below 5%.

(H) For each gene, the fitness at wild-type expression was computed by plugging wild-type expression into the impulse function with the gene-specific parameters. Shown are the histograms of fitness values for wild-type expression (blue) and the entire dataset (gray). Fitness levels at wild-type expression fall in a tight histogram, with 70% of the genes exhibiting fitness values between 0.95 and 1.05.

(I) For all strains shown is their measured relative fitness (x axis) and the relative fitness for the same expression levels from the fitted curve (y axis).

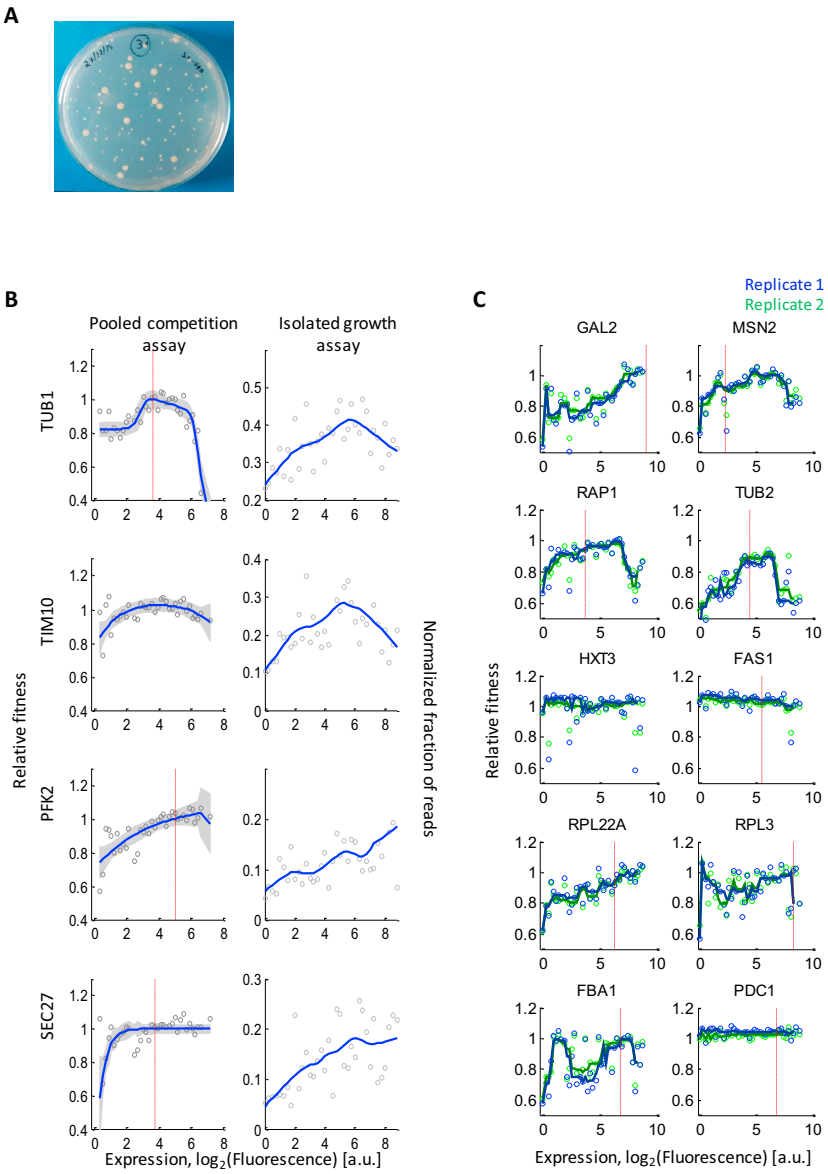


Figure S3. High Correspondence between Pooled and Isolated Fitness Measurements and between Replicates of Pooled Competition Experiment, Related to Figure 1

(A and B) High correspondence between pooled competition experiment and sizes of individually-grown colonies. (A) Synthetic promoters were transformed upstream of the target gene and positive transformants were selected on SCD-Ura plates. After five days colonies of different sizes are visible, reflecting varying degrees of growth rate of their respective strains. (B) Left panels: For four genes from different functional classes, shown is their relative fitness (y axis) as a function of expression (x axis) as obtained from a pooled competition assay. Gray circles are the datapoints, blue lines are the best fit of a parametric impulse model to the data, and shaded gray areas mark the 95% confidence intervals. Dashed red lines mark wild-type expression levels. Right panels: For the same genes shown is the normalized fractional abundance of the strains (y axis) as a function of expression (x axis) when sequencing DNA from a pool of cells that were grown in isolation as depicted in S3A. Promoters were binned into 40 logarithmically-equally- spaced bins and data was plotted as the median of each bin (gray points). Blue line is a running median of the data (matlab 'smooth' with window size 15).

(C) High correspondence between biological replicates of pooled competition experiment. For 10 target genes shown are the relative fitness (y axis) as function of expression (x axis) measured in two biological replicates of the pooled competition experiment in galactose. Replicate 1 is depicted in blue and replicate 2 is depicted in green. Circles are the measured datapoints, blue and green lines are loess fits for the blue and green points, respectively.

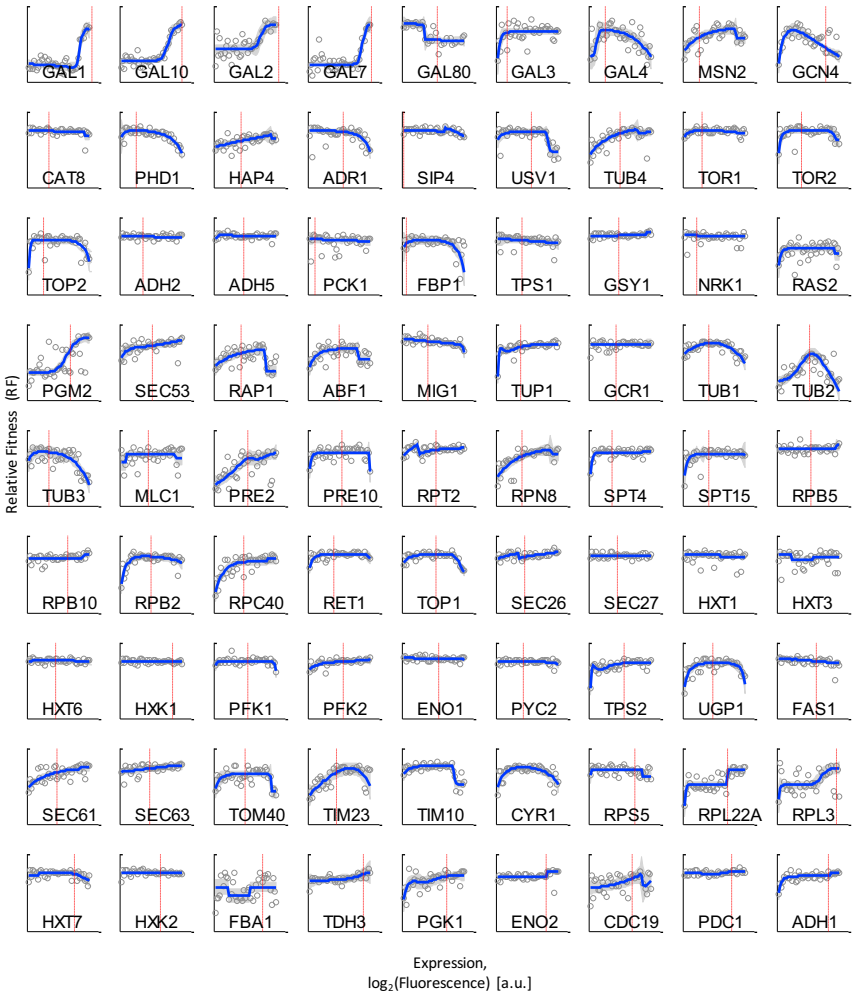


Figure S4. Fitness-to-Expression Curves for 81 Genes in Galactose, Related to Figure 2

For the 81 target genes that succeeded in the construction process and passed all quality controls shown are their relative fitness (y axis) as a function of expression (x axis) in galactose. Gray circles are the datapoints, blue lines are the best fit of a parametric impulse model to the data, and shaded gray areas mark the 95% confidence intervals. Dashed red lines mark wild-type expression levels.

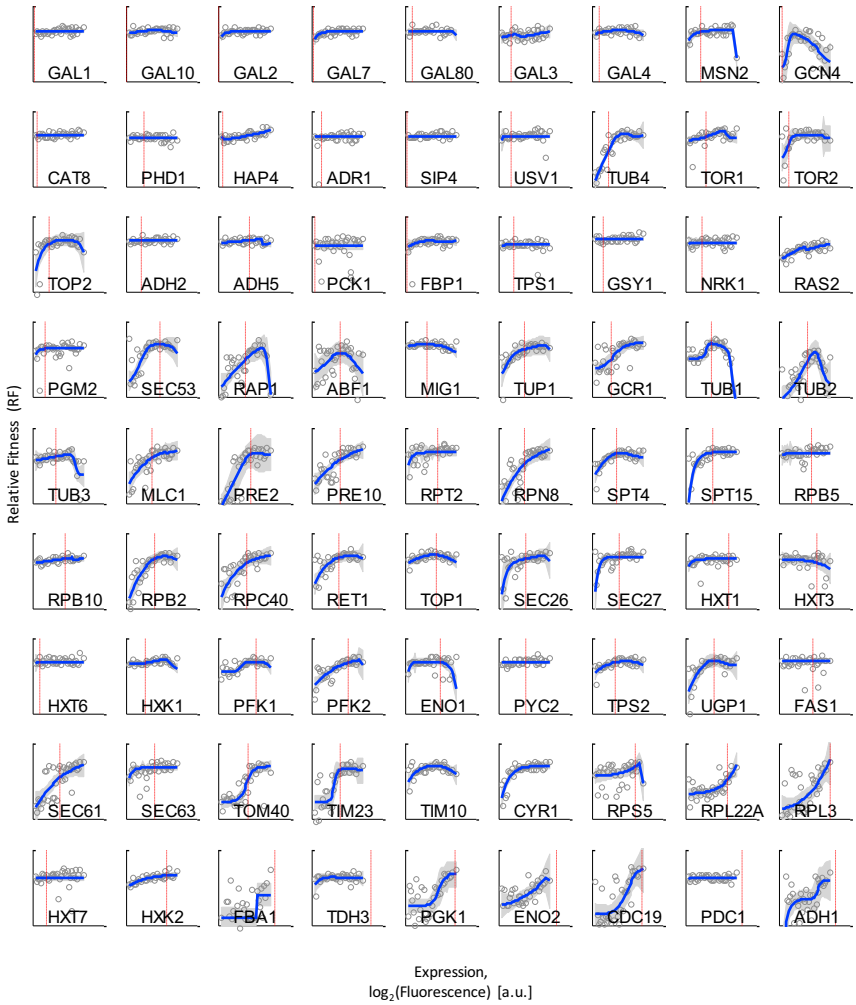
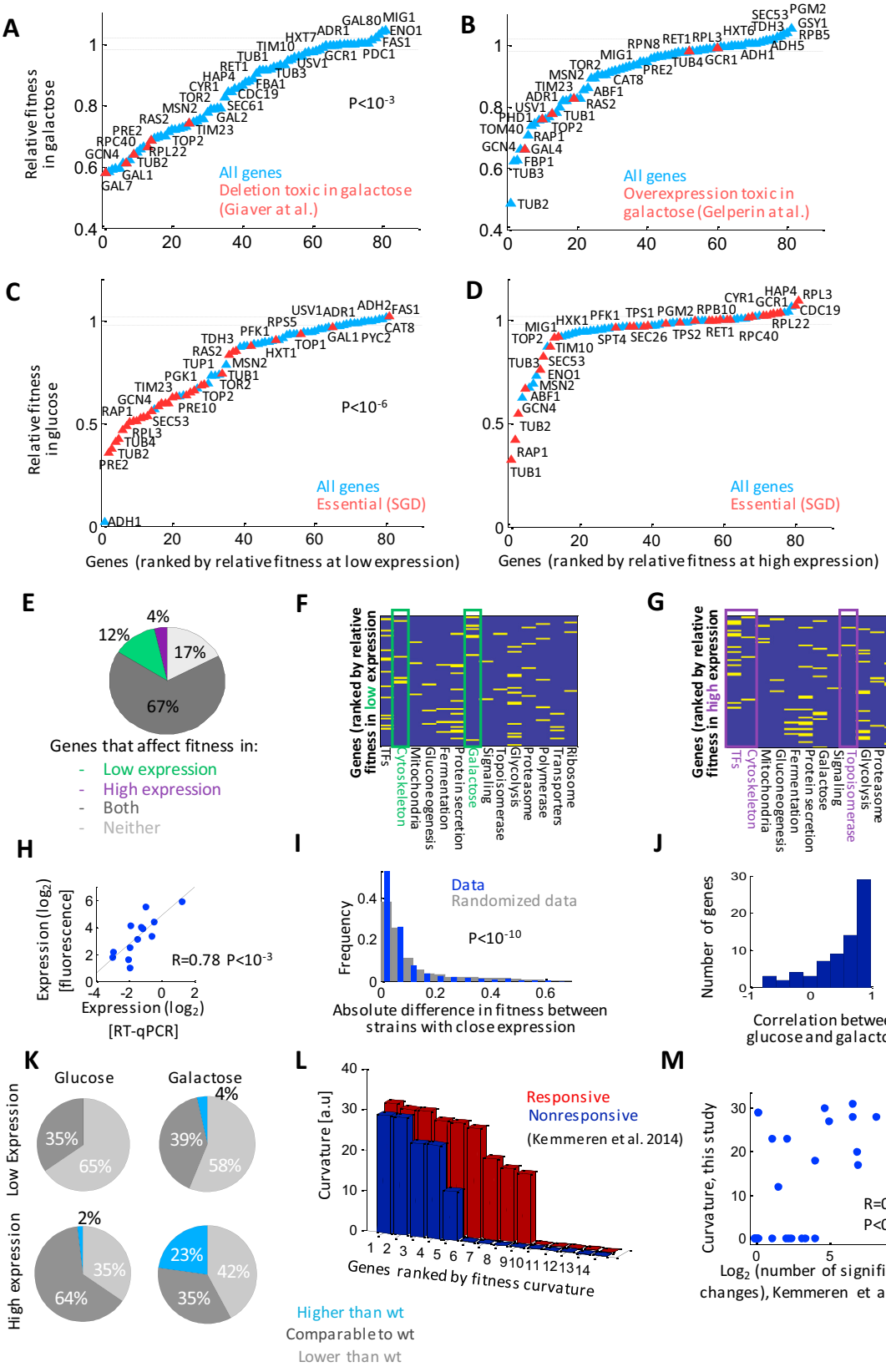


Figure S5. Fitness-to-Expression Curves for 81 Genes in Glucose, Related to Figure 4

For the 81 target genes that succeeded in the construction process and passed all quality controls shown are their relative fitness (y axis) as a function of expression (x axis) in glucose. Gray circles are the datapoints, blue lines are the best fit of a parametric impulse model to the data, and shaded gray areas mark the 95% confidence intervals. Dashed red lines mark wild-type expression levels.



(legend on next page)

Figure S6. Analysis of Fitness-to-Expression Curves in Galactose and Glucose, Related to Figures 2, 4, and 5

(A) Genes were ranked according to their relative fitness in galactose in the lowest expression level measured (y axis) and plotted in ranked order (x axis). Genes whose deletions were previously identified to be toxic in galactose (Giaever et al., 2002) are marked by red triangles. P-value is for Wilcoxon ranksum test. Names of representative genes are indicated on the plot.

(B) Genes were ranked according to their relative fitness in galactose in the highest expression level measured (y axis) and plotted in ranked order (x axis). Genes whose overexpression was previously identified to be toxic in galactose (Gelperin et al., 2005) are marked by red triangles. Names of representative genes are indicated on the plot.

(C) Same as (A), but in glucose. Essential genes (SGD) are marked by red triangles.

(D) Same as (B), but in glucose. Essential genes (SGD) are marked by red triangles.

(E) An upper and lower threshold of 0.98 and 1.02 fitness, respectively, were used to classify genes as having an effect on fitness for a given expression level. Shown are the relative abundances of genes that alter fitness in galactose either at low expression (green), high expression (purple), both (dark gray) or neither (light gray).

(F and G) Shown are all genes in the library (rows) ranked by their fitness in galactose in either low expression (F) or high expression (G). In each column genes belonging to a different functional group are highlighted in yellow. Enriched GO terms in either low or high expression are marked by green or purple boxes, respectively.

(H) For 20 synthetic promoters, driving the expression of the gene *LCB2*, shown is the correlation between mRNA levels as measured by RT-qPCR (x axis) and expression as measured by fluorescent reporters (y axis) for strains grown in glucose.

(I) A randomized dataset was generated by shuffling, for each gene in glucose, the fitness values with respect to the expression values. For each gene, the absolute difference in fitness between neighboring expression values was computed for both datasets. Shown is a histogram for the difference values of the real (blue) and randomized (gray) data. Strains that have similar expression levels of their target genes tend to have similar fitness levels. P-value depicts significance in a ranksum test.

(J) Comparison between fitness-to-expression curves in glucose and galactose. The correlation between the fitness-to-expression curves in glucose and galactose was computed for all genes in the library. Shown is a histogram of the correlation values of all genes.

(K) For all genes in either glucose or galactose, the relative fitness at low expression or high expression was compared to wild-type expression levels. For both low and high expression, in both glucose and galactose, shown are the percentages of genes that have higher fitness than wild-type (blue), lower fitness than wild-type (light gray) or comparable fitness to wild-type (defined as wt ± 0.02, dark gray).

(L) Genes which show larger changes in their expression-profiles relatively to wild-type when deleted have larger fitness curvatures. 27 genes that have been profiled by both Kemmeren et al. (Kemmeren et al., 2014) and this study were divided into 'Responsive' and 'Nonresponsive' according to the definitions of Kemmeren et al. Plotted is the fitness curvature for each of these genes, defined as the distance in expression bins from wild-type expression that results in a 5% reduction in fitness. The genes in the responsive group have overall larger curvatures.

(M) Shown is a positive correlation between the number of genes significantly changed in a deletion mutant (x axis, (Kemmeren et al., 2014)) to the curvature of the gene as a function of expression (y axis, this study). Data was jittered along the x axis (matlab 'scatter', jitter = 0.2) to allow visualization of overlapping points.

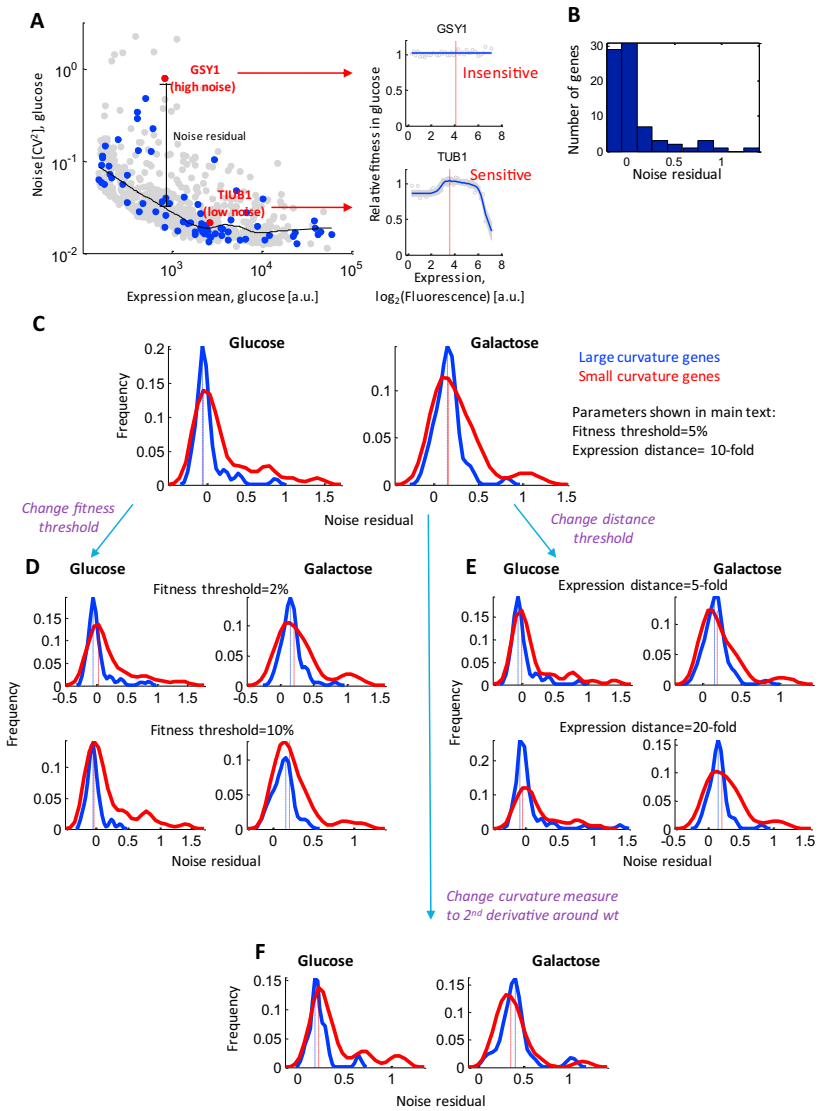


Figure S7. Noise in Gene Expression Is Anti-correlated with Fitness Curvature around Wild-Type Expression, Related to Figure 6

(A) Scatter-plot of the YFP mean (x axis) and noise (CV^2, y axis) for 900 native yeast promoters in glucose, taken from (Keren et al., 2015). Promoters of genes from the current study are highlighted in blue. Black line depicts loess fit to the data. In red are highlighted two genes differing in their noise levels: *GSY1*, which has high noise compared with the expectation from its mean (large noise residual), and *TUB1*, which has low noise compared with the expectation from its mean (small noise residual). For the genes *GSY1* and *TUB1*, shown are their relative fitness (y axis) as a function of expression (x axis) in glucose. Gray circles are the measured datapoints, solid lines are the best fit of a parametric impulse model to the data, and shaded gray areas mark the 95% confidence intervals. Dashed red lines mark wt expression levels.

(B) Histogram of the noise residuals in glucose for the 81 genes essayed in this study, depicting large differences in noise between different genes.

(C) All genes were classified, based on their fitness to expression curves in either galactose or glucose, as having large or small curvature around wild-type expression levels. Genes with large curvature were defined as those for which 10-fold changes from wild-type expression levels resulted in a reduction of at least 5% in fitness. Shown are the density functions of the noise residuals for both groups in either glucose (left) or galactose (right).

(D) Same as in (C), but changing the fitness reduction threshold to either 2% (top) or 10% (bottom).

(E) Same as in (C), but changing the expression threshold to either 5-fold or 20-fold.

(F) Same as in (C), but rather than using a long-range definition for the curvature, the local curvature around wild-type is used, e.g., the second derivative of the fitness function at wild-type expression.