Suboptimal Global Transcriptional Response Increases the Harmful Effects of Loss-of-Function Mutations

Károly Kovács,[†,1,2] Zoltán Farkas,[†,2] Djordje Bajić,[†,2,3,4,5] Dorottya Kalapis,[1,2] Andreea Daraba,[2] Karola Almási,[2] Bálint Kintses,[6,2,7] Zoltán Bódi,[2] Richard A. Notebaart,[8] Juan F. Poyatos,[3] Patrick Kemmeren,[9] Frank C.P. Holstege,[9] Csaba Pál,*[,2] and Balázs Papp*[,1,2]

[1]HCEMM-BRC Metabolic Systems Biology Lab, Szeged, Hungary
[2]Biological Research Centre, Synthetic and Systems Biology Unit, Institute of Biochemistry, Szeged, Hungary
[3]Logic of Genomic Systems Laboratory, Department of Systems Biology, CNB-CSIC, Madrid, Spain
[4]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT
[5]Microbial Sciences Institute, Yale University West Campus, West Haven, CT
[6]HCEMM-BRC Translational Microbiology Lab, Szeged, Hungary
[7]Department of Biochemistry and Molecular Biology, University of Szeged, Szeged, Hungary
[8]Food Microbiology, Wageningen University & Research, Wageningen, The Netherlands
[9]Princess Máxima Center for Pediatric Oncology, Utrecht, The Netherlands
[†]These authors contributed equally to this work.

***Corresponding authors**: E-mails: cpal@brc.hu; pappb@brc.hu.

Associate editor: Jianzhi Zhang

Abstract

The fitness impact of loss-of-function mutations is generally assumed to reflect the loss of specific molecular functions associated with the perturbed gene. Here, we propose that rewiring of the transcriptome upon deleterious gene inactivation is frequently nonspecific and mimics stereotypic responses to external environmental change. Consequently, transcriptional response to gene deletion could be suboptimal and incur an extra fitness cost. Analysis of the transcriptomes of ~1,500 single-gene deletion *Saccharomyces cerevisiae* strains supported this scenario. First, most transcriptomic changes are not specific to the deleted gene but are rather triggered by perturbations in functionally diverse genes. Second, gene deletions that alter the expression of dosage-sensitive genes are especially harmful. Third, by elevating the expression level of downregulated genes, we could experimentally mitigate the fitness defect of gene deletions. Our work shows that rewiring of genomic expression upon gene inactivation shapes the harmful effects of mutations.

Key words: **genotype–phenotype map, gene deletion, fitness effect of mutations, gene expression regulation.**

Introduction

Genetic perturbations, such as gene deletions, are fundamental for the biological sciences both as natural phenomena and as tools to gain insight into the function of the perturbed genes. The latter approach, reverse genetics, assumes that the cellular response to the perturbation reflects the function of the perturbed gene. More specifically, fitness loss upon deleting a gene is generally interpreted as evidence for the importance of the gene's specific function in the particular environment (Giaever et al. 2002). Several lines of reasoning suggest that the fitness impact of gene disruption is further molded by the cell's response to the absence of the gene. One theory suggests that the cell's regulatory network increases the level of a functionally redundant protein upon gene inactivation and thereby ameliorates its fitness effect (Kafri et al. 2016). Earlier studies found anecdotal evidence for such adaptive regulatory responses in paralog gene pairs (Kafri et al. 2005), but their prevalence remains contentious (He and Zhang 2005; Wong and Roth 2005).

Alternatively, a suboptimal regulatory response to gene deletion might aggravate its fitness cost. Mounting evidence indicates that response to environmental changes often results in suboptimal gene expression patterns in microbes (e.g., Price et al. 2013; Keren et al. 2016). Several molecular mechanisms may underlie suboptimal regulation. First, the majority of expression changes appear to result from global effects and not from specific regulation directly related to the gene's function (Keren et al. 2013; Price et al. 2013). One prominent form of such global indirect effect is growth rate correlated expression, shaping the transcription of about a

Article

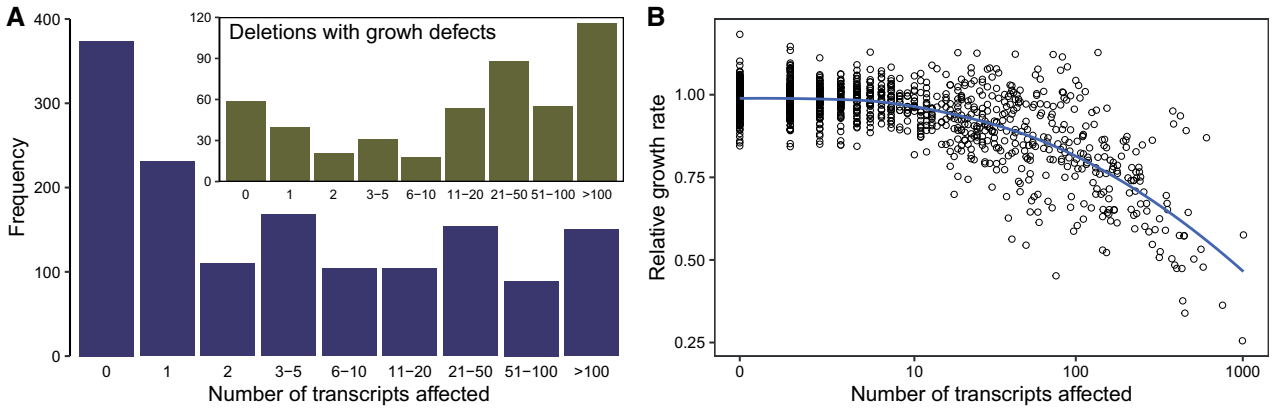

FIG. 1. Extent of transcriptome changes upon gene deletions. (*A*) Distribution of the number of transcript level changes across the ~1,500 single-gene deletion strains. A transcript level change was defined by fold-change (FC > 1.7) and statistical significance thresholds ($P < 0.05$), following Kemmeren et al. (2014). The inset shows number of transcript level changes for the subset of gene deletions causing ≥5% growth defect. (*B*) Growth rate is highly correlated with the number of changing transcripts (Spearman's $\rho = -0.49$, $P < 10^{-15}$). Blue line indicates a loess curve fitted on log(number of transcript level changes + 1), with parameters span = 1, degree = 2. Relative growth rate of deletion strains was obtained from O'Duibhir et al. (2014).

quarter of the protein coding genes in yeast (Brauer et al. 2008), a large part of which can be attributed to shifted residence times in different phases of the cell cycle (O'Duibhir et al. 2014). Second, instead of using accurate regulation when responding to changes in nutrient conditions, bacteria tend to apply general heuristic rules that are beneficial in most, but not all conditions (Towbin et al. 2017). Overall, these observations suggest that regulatory networks have only a limited number of ways to rewire gene expression patterns upon environmental changes. From an evolutionary perspective, responses to environmental changes could be suboptimal because highly accurate and direct regulation might be too costly or too difficult to evolve (Price et al. 2013; Towbin et al. 2017). It is thus feasible that suboptimal gene expression patterns are especially frequent upon genetic perturbations. As most individual mutational events—including deletions of specific genes—are exceedingly rare, natural selection for optimal genomic expression changes upon such mutations is expected to be generally weak under realistic population genetic conditions (Proulx and Phillips 2005). Despite these considerations, it remains unclear how specific the cell's gene expression response to genetic perturbations is and whether the response itself contributes to the fitness defect.

In this study, we investigate the hypothesis that suboptimal gene expression response increases the severity of genetic perturbations. First, by analyzing the transcriptome of hundreds of gene knockouts in the budding yeast *Saccharomyces cerevisiae*, we show that the regulatory circuits of yeast exhibit only a limited number of ways by which the transcriptome can be rewired, thereby gene deletion-specific transcriptomic changes are unlikely to occur. These nonspecific expression responses resemble those displayed upon environmental changes, including, but not limited to, the general environmental stress response (Gasch 2003). Second, we found that gene deletion strains in which the expression level of dosage-sensitive genes (i.e., those sensitive to changed dosage) are altered show an especially low fitness, suggesting suboptimal response. Next, we experimentally probed the fitness impact of gene downregulation by increasing the expression level of specific downregulated genes in slow-growing gene deletion strains. These experiments establish the existence of suboptimally expressed genes in genetically perturbed cells and demonstrate that misregulation of even a single gene can incur a substantial fitness cost. Together, these findings offer a conceptual model to explain why genomic expression changes could be suboptimal upon gene deletion. According to the "stereotypic response" model, the yeast regulatory network has only a limited number of transcriptional response modes (TRMs) that have likely evolved to deal with common environmental perturbations. These response modes are also utilized to cope with genetic perturbations, for lack of a better alternative, often resulting in suboptimal transcriptional states that may incur an extra fitness cost.

Results

Extensive Transcriptional Response upon Deleterious Gene Inactivations

We started by assessing the extent of transcript level changes across deletion strains using data from a previous systematic study (Kemmeren et al. 2014). This unique resource provides genome-wide transcriptome profiles for ~1,500 single-gene deletion yeast strains. Although this set covers only one quarter of protein coding genes in yeast, it is overrepresented in genes having a regulatory role (Kemmeren et al. 2014) and hence expected to capture well null mutations with large transcriptomic effects. Nevertheless, 69% of the investigated genes have no established role in transcriptional regulation and represent various functional categories, including metabolism, protein trafficking, and DNA repair (Kemmeren et al. 2014).

There is a wide variation in the number of genes with significantly altered transcription level across single-gene deletion yeast strains, ranging from 0 to 1,009 (fig. 1A). Approximately, 60% of the gene deletion strains have less

than six transcripts affected, whereas 16% of deletion strains show more than 50 altered transcripts. Importantly, even deletion of genes having no established transcriptional regulatory roles can yield extensive transcriptomic changes affecting numerous genes (supplementary fig. S1, Supplementary Material online). For example, deletion of *ERG2*, a gene encoding a sterol isomerase enzyme involved in ergosterol biosynthesis, triggers transcriptional changes in 91 genes.

The extent of transcriptional rewiring shows a markedly different distribution among gene deletions causing at least a mild (≥5%) growth defect (fig. 1A, inset). Notably, the majority (65%) of such deletions affect the expression of more than ten genes and 35% of them alter more than 50 genes, suggesting that deleterious gene inactivations frequently induce global transcriptome rewiring. Indeed, deletions affecting numerous transcripts have a strong tendency to reduce fitness (Spearman's $\rho = -0.49$, $P < 10^{-15}$, fig. 1B), a finding consistent with previous studies (Hughes et al. 2000; Kemmeren et al. 2014). One may argue that such a strong correlation between transcriptomic and fitness effects might simply reflect regulation through growth rate. Indeed, the expression of numerous yeast genes is strongly influenced by cellular growth rate, independent of the specific environmental factors (Castrillo et al. 2007; Brauer et al. 2008). Shifts in the proportions of cells in different cell cycle phases in slow-growing yeast deletion strains produces a common gene expression signature proportional to the degree of slow growth (O'Duibhir et al. 2014). However, the strong correlation between mutant fitness and the number of altered transcripts remains after removing the shared gene expression signature exhibited by slow-growing mutants (O'Duibhir et al. 2014) (Spearman's $\rho = -0.45$, $P < 10^{-15}$, supplementary fig. S2, Supplementary Material online). Thus, extensive gene expression rewiring is not simply the consequence of the cell cycle phase distribution change of slow-growing strains. Although this finding is consistent with the notion that strongly deleterious gene disruptions are highly pleiotropic, that is, affect many phenotypic traits and molecular processes (Cooper et al. 2007; Szappanos et al. 2011), it also raises the possibility that the transcriptomic response itself may contribute to the observed growth defect. We test this latter possibility below.

Yeast Has a Limited Repertoire of TRMs upon Gene Deletion

Are most transcriptional changes specific to the deleted gene? It has been previously reported that beyond specific expression changes, many gene deletion strains show recurrent expression patterns (O'Duibhir et al. 2014). We hypothesized that yeast has evolved a limited number of fundamentally distinct genome-wide transcriptional responses, likely through adaptation to environmental changes. Such transcriptional responses are recurrently and nonspecifically triggered by gene deletions that lack a more specific response. This "stereotypic response" hypothesis predicts that 1) yeast should exhibit a limited number of TRMs that account for a substantial part of gene expression variation across mutants, and 2) responses to gene deletions should mimic those to environmental changes.

To assess the extent to which gene expression changes can be attributed to stereotypic responses shared across null mutations, we applied principal component analysis (PCA) on the transcriptome profiles of all ~1,500 gene deletion strains. Our analysis expands on the work of O'Duibhir et al. (2014), where the first principal component was analyzed in detail. It has been reported that the first principal component, which we refer to transcriptional response mode #1 (TRM #1), accounts for 24% of all transcriptomic variation in the data set and represents a recurrent expression response common to slow-growing strains. We aimed at systematically identifying additional TRMs, that is, recurrent expression patterns shared by multiple gene deletions. To this end, we applied PCA to both the actual transcriptome data set and to randomized ones and identified those principal components that explain more variation than expected by chance (fig. 2A, Materials and Methods). This analysis revealed 15 uncorrelated TRMs, each of them explaining more gene expression variance than expected by chance, hence capturing recurrent expression responses (fig. 2B). Importantly, together these 15 TRMs explain 58.1% of all transcriptional variation across the entire set of gene knockouts (fig. 2B). These results indicate that a large fraction of genomic expression changes is not specific to individual gene deletions.

We next characterized the functional diversity of gene deletions that trigger each TRM. To this end, we identified sets of genes whose deletion correlates or anticorrelates with a particular TRM (see Materials and Methods and supplementary table S1, Supplementary Material online). Note that a single null mutation can trigger multiple TRMs, however, those cases are not common (26% of all deleted genes). We first focused on TRM #1, which has been studied previously (O'Duibhir et al. 2014). In line with previous observations, the large majority of null mutations triggering TRM #1 act in one particular direction, that is, they induce a common slow growth signature, and tend to cause a large fitness drop (supplementary fig. S3, Supplementary Material online). Importantly, deletion in diverse gene functions, including ribosome biogenesis, cell cycle, signaling and DNA repair initiate this expression signature (supplementary fig. S4 and table S2, Supplementary Material online). Inspecting the functional categories of deletions triggering the other 14 TRMs revealed a similar pattern (supplementary fig. S4 and table S2, Supplementary Material online). Most TRMs are correlated or anticorrelated with deletions in genes spanning multiple functional categories. For example, TRM #2 is triggered by deletions of genes involved in protein alkylation, pseudohyphal/invasive growth, histone modification, or tRNA processing (supplementary fig. S4, Supplementary Material online).

To more systematically investigate the functional diversity of gene deletions triggering the same TRMs, we next employed genetic interaction profile similarity as an unbiased, quantitative measure of functional relationships between genes (Costanzo et al. 2016). For each TRM, we calculated genetic interaction profile correlations between pairs of genes triggering a given TRM when deleted (separately for gene

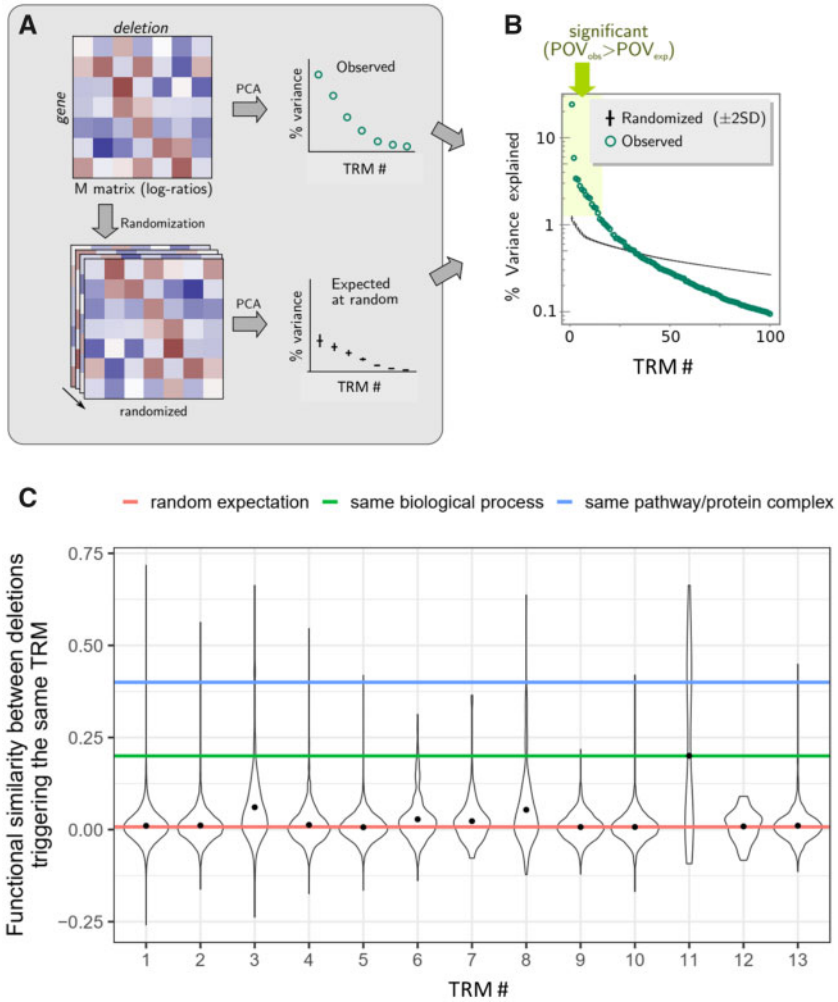

FIG. 2. A handful of TRMs explain most gene expression responses to knockouts. (*A*) Overview of the approach to estimate the number of significant independent TRMs. PCA is performed over the original expression data set, as well as over randomized instances of it (Materials and Methods). The threshold for estimating the number of TRMs is set as the mean + 2 SD in the variance explained by TRM #1 in 600 randomized data sets. (*B*) Results obtained on the transcriptomes of ~1,500 gene deletion strains, in which we estimate that each one of the first 15 TRMs explains more variance than expected by chance. Results are shown for the first 100 TRMs, *y*-axis showing the explained percentage of variance for the randomized (POV_{exp}) and observed (POV_{obs}) data. (*C*) Functionally diverse gene deletions trigger the same TRMs. Functional relatedness between all gene pairs ($N = 98{,}438$) triggering the same TRM is calculated as the Pearson correlation of their genetic interaction profiles (Costanzo et al. 2016). Violin plots show the density distributions of these Pearson correlation coefficients for each TRM, and black dots indicate their mean. Red line shows the randomly expected correlation in our set of genes (0.007), whereas blue and green lines show correlation coefficients indicating functional relatedness (0.2) and shared complex/pathway memberships (0.4), respectively (Costanzo et al. 2016). Only TRMs with more than ten data points are shown.

deletions correlating positively and negatively with a particular TRM). According to Costanzo et al. (2016), Pearson correlation coefficients above 0.2 indicate functional similarity, whereas values above 0.4 indicate that the two genes share the same specific protein complex or pathway. In contrast, the randomly expected correlation coefficient between all possible gene pairs in our gene set is 0.007. Remarkably, the average correlation coefficient within TRMs is 0.01, which is very close to random expectation. Furthermore, only 0.5% of within-TRM gene pairs show clear signs of functional similarity (i.e., values >0.2). This result is consistent with the above finding that, in general, multiple unrelated gene ontology (GO) categories are enriched in gene deletions triggering a given TRM (supplementary fig. S4, Supplementary Material online). Furthermore, it shows that despite such GO enrichments, the vast majority of genes whose null mutations trigger a given TRM are functionally unrelated to each other. This conclusion holds for almost all TRMs (fig. 2C). One notable exception is TRM #11, where a substantial fraction of gene pairs with genetic interaction data (27/62) encode subunits of the histone deacetylase complex, Rpd3L. Notably, excluding these within-complex pairs yields an average correlation coefficient close to random expectation (0.013) for TRM #11 too.

Finally, we note that 70% of the genes that trigger at least one TRM when deleted have no established direct role in

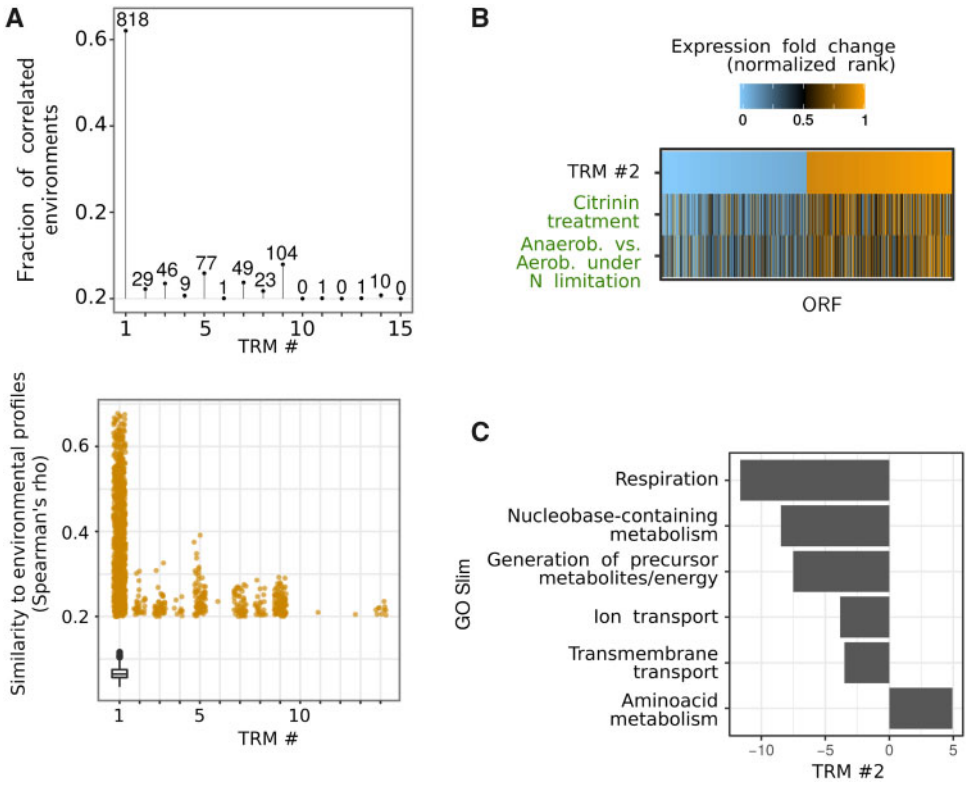

FIG. 3. TRMs mimic genomic expression response to environmental change. (*A*) Upper panel: number and relative proportion of environmental expression responses correlating with TRMs above our threshold (|Spearman ρ| > 0.2). Lower panel: TRMs mimicking environmental expression responses above the same correlation threshold. Boxplot shows the distribution of correlation coefficients between TRM #1 and the most similar environmental profiles for 600 randomized data sets (box indicates the interquartile range [IQR], hinges represent third quartile + 1.5 IQR and first quartile − 1.5 IQR). (*B*) Heatmap showing transcripts (column) with largest absolute fold-change in TRM #2 (top and bottom 8%) and two relevant environments: comparison between anaerobic and anaerobic conditions under nitrogen limitation (Knijnenburg et al. 2007) and citrinin treatment (Iwahashi et al. 2007). For visual clarity, rank-transformed values are shown.(*C*) GO Slim category enrichment in the genes significantly affected by TRM #2 (Materials and Methods). Sign of the enrichment indicates expression direction in the signature profile (i.e., expression profile of the deletion strain with the highest amount of expression change in TRM #2, see Materials and Methods). Only those categories are shown where the enrichment is significant (Fisher's test, FDR-adjusted $P < 0.001$.

transcriptional regulation (Materials and Methods). This observation suggests that more distant regulatory events or indirect control of gene expression (Price et al. 2013; Chagoyen and Poyatos 2019) dominate TRMs.

Overall, we conclude that most of the expression changes across ~1,500 deletion strains can be explained by only a handful of TRMs that are triggered by functionally diverse gene deletions.

Gene Deletions Trigger Transcriptional Responses That Mimic Environmental Responses

Next, we systematically tested whether TRMs upon gene deletion resemble expression responses displayed upon environmental changes. To this end, we calculated correlations between each TRM and published environmental expression profiles from a large compendium of microarrays (Hibbs et al. 2007). We found that the large majority (12 out of 15) of TRMs correlates with at least one environmental response profile at a threshold of Spearman's $\rho > 0.2$ (fig. 3*A* and supplementary table S3, Supplementary Material online, see Materials and Methods). Crucially, correlations observed above this threshold are much stronger than expected based on randomized data (fig. 3*A*), indicating genuine similarities between TRMs upon gene deletion and environmental expression profiles. Reassuringly, TRM #1 correlates with up to two-thirds of the tested environmental response profiles (fig. 3*A*). This is consistent with an earlier report (O'Duibhir et al. 2014) that the slow growth signature captured by this TRM is highly similar to the gene expression response common to diverse environmental perturbations, referred to as the environmental stress response (Gasch 2003). Previously, expression response to heat shock ("Heat shock 20 min after shift from 29°C to 37°C" in Gasch [2003]) was used as a proxy for environmental stress response (O'Duibhir et al. 2014). Importantly, although this environment shows the highest correlation with TRM #1 (Spearman's $\rho = 0.68$), no other TRM shows a correlation higher than our threshold with this environment, (supplementary table S3, Supplementary Material online) suggesting that other TRMs are largely uncorrelated with the environmental stress response program.

This result indicates that other TRMs mimic less general responses to environmental perturbations. For example, the global expression profile of TRM #2 is similar to that observed

in responses to changes in oxygen supply or inhibition of respiration (Iwahashi et al. 2007) (fig. 3B). As a further support, genes that show large expression fold-changes in TRM #2 are enriched in respiration and mitochondria-related functions (fig. 3C and supplementary fig. S5, Supplementary Material online, see Materials and Methods). Together, these results suggest that TRM #2 is linked to transcriptional regulation of aerobiosis. In a similar vein, TRM #4 shows strong correlations with environmental profiles related to leucine/sulfate or histidine limitation (Saldanha et al. 2004; Levy et al. 2007) (supplementary table S3, Supplementary Material online) Indeed, genes affected by TRM #4 are enriched in functions related to cellular amino acid metabolic processes (supplementary fig. S5, Supplementary Material online). Finally, we speculate that those three TRMs showing no clear similarity to environmental response profiles might have evolved as responses to environmental perturbations that remain to be sampled.

To assess whether TRMs may represent regulatory units, we next asked whether genes altered in each TRM are enriched in regulation by any particular transcription factor (TF) or group thereof. As might be expected, TRM #2 shows enrichment in regulation by oxic state-related TFs, such as Skn7p, Sko1p, Hap1p, or the complex formed by Hap2p–Hap5p, which is a global regulator of respiratory gene expression (Pinkham and Guarente 1985). Similarly, TRM #4 shows enrichment in regulation by Met32p, Dal81p, or Gcn4p, consistent with the link of this TRM to amino acid metabolism (supplementary fig. S6, Supplementary Material online). Importantly, TRMs are associated with multiple TFs and vice versa, suggesting a lack of simple correspondence between regulatory units and TRMs in general (supplementary fig. S6, Supplementary Material online).

Taken together, these findings demonstrate that TRMs triggered by gene deletions mimic responses to environmental changes. This further suggests that the regulatory network of yeast might not be able to produce highly specific responses that would be beneficial upon loss of particular gene functions.

Evidence for Nonadaptive Transcriptional Rewiring

Our results so far indicate that a substantial fraction of expression changes upon gene deletion are not gene specific. However, it remains unclear whether these nonspecific responses are beneficial, neutral, or harmful for fitness.

First, we tested whether transcriptional upregulation upon gene disruption might be beneficial by compensating for the lost function. To this end, we integrated the transcriptional response data with the most recent and comprehensive genome-wide map of genetic interactions in yeast (Costanzo et al. 2016). Specifically, we focused on synthetic sick or lethal genetic interaction, that is, when the single mutants are nearly as fit as the wild-type, but the double mutants display slow growth. As noted earlier, synthetic sick or lethal genetic interaction between null mutations indicates functional compensation (i.e., functional backup) between the gene pair (Hartman et al. 2001). However, only ~2% of upregulations provides backup for the deletion (fig. 4A), a figure only slightly higher than expected by chance

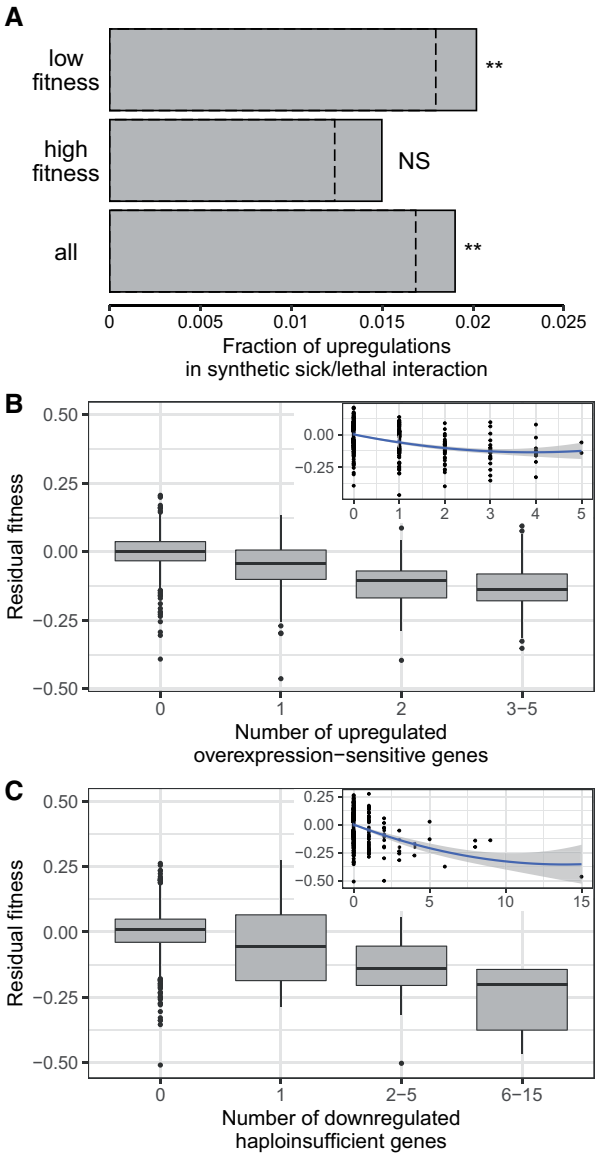

FIG. 4. Nonadaptive gene expression changes in response to gene deletion. (A) Fraction of upregulated genes being in synthetic sick of lethal genetic interaction with the deleted gene. The same fractions are shown for high- and low-fitness deletion strains (based on a 0.95 relative growth rate threshold), and for all strains as well. Dashed lines indicate the randomly expected fractions estimated by randomization (Materials and Methods). Asterisks indicate significant enrichment (*/**/*** indicates P value < 0.05/0.01/0.001, respectively, NS indicates nonsignificant). (B, C) Altered expression of dosage-sensitive genes is associated with lower fitness. Residual fitness of deletion strains negatively correlates with the number of overexpression-sensitive genes upregulated (Spearman correlation $P = 10^{-16}$, panel B) or haploinsufficient genes downregulated (Spearman correlation $P = 7 \times 10^{-11}$, panel C). Residual fitness is the fitness of a deletion strain after controlling for the total number of up- or downregulations. Boxplots show data points being grouped together for multiple ranges of number of up- or downregulations, whereas insets show the same data across the whole range of up- or downregulations with a second order polynomial fit. For details on the statistical analyses, see Materials and Methods.

(1.3-fold enrichment, $P = 0.005$, Materials and Methods). This result holds even when knockouts of TFs are removed from the data set (supplementary fig. S7, Supplementary Material

online) or when considering only high-fitness knockouts, which might be especially well compensated (fig. 4A). Together, our findings are consistent with earlier works suggesting that transcriptional responses rarely compensate for the lost gene functions (Wong and Roth 2005). Thus, most of the expression changes upon gene deletion are unlikely to be beneficial for the cell.

Next, we looked for signatures of harmful responses in the transcriptome data set. We hypothesize that altered expression of dosage-sensitive genes in gene deletion backgrounds is especially likely to reduce fitness. Previous systematic works identified a set of overexpression-sensitive (Makanae et al. 2013) and haploinsufficient (Deutschbauer et al. 2005) genes that are known to have deleterious effects when their dosage is increased and decreased, respectively. The set of overexpression-sensitive genes identified by Makanae et al. is especially well suited for our study as it includes genes for which a mild increase in copy number causes a growth defect. Notably, expression changes in dosage-sensitive genes triggered by gene deletions are also predominantly unspecific (supplementary note, Supplementary Material online), therefore we expect these genes to show signatures of harmful responses.

To test the hypothesis, we examined whether deletion strains with upregulated overexpression-sensitive genes or downregulated haploinsufficient genes have especially low fitness. Gene knockouts with numerous transcriptionally altered genes both have a higher chance that some of those genes are dosage sensitive and also tend to have a low fitness in general (Hughes et al. 2000; Regenberg et al. 2006). Therefore, we controlled for the number of transcriptionally altered genes, by calculating residual fitness values (Materials and Methods). In line with expectation, we found a highly significant negative correlation between the number of upregulated overexpression-sensitive genes in a deletion strain and its residual fitness (fig. 4B and supplementary fig. S8A, Supplementary Material online, Spearman's $\rho = -0.39$, $P = 10^{-16}$). Analogously, the number of downregulated haploinsufficient genes is also negatively correlated with the residual fitness of the gene knockout (fig. 4C and supplementary fig. S8B, Supplementary Material online, Spearman's $\rho = -0.23$, $P = 2 \times 10^{-10}$). Moreover, these correlations remain significant even when considering only those strains with at least one upregulated overexpression-sensitive gene (Spearman's $\rho = -0.37$, $P = 4 \times 10^{-7}$) or one downregulated haploinsufficient gene (Spearman's $\rho = -0.39$, $P = 6 \times 10^{-4}$). Thus, not only the presence but also the number of rewired dosage-sensitive genes is associated with lower fitness. Furthermore, in line with expectations, the number of *downregulated* overexpression-sensitive genes does not show a *negative* correlation with the residual fitness. Rather, a weak positive trend was found between the two variables ($\rho = 0.18$, $P = 2 \times 10^{-7}$). In contrast, the number of *upregulated* haploinsufficient genes shows a weak, but significant negative correlation with residual fitness ($\rho = -0.12$, $P = 0.002$). The latter result is in line with a recent finding indicating an asymmetrical relationship between haploinsufficiency and overexpression phenotype. Although most haploinsufficient genes are sensitive to overexpression, the converse is not true: Most overexpression-sensitive genes are not haploinsufficient (Morrill and Amon 2019).

Overall, these results support the hypothesis that suboptimal transcriptional changes of dosage-sensitive genes contribute to the fitness impact of gene deletions.

Experimental Evidence for Harmful Downregulation upon Gene Disruption

To test the causal link between gene expression response and fitness, we next probed the fitness effect of increasing the expression level of downregulated genes in multiple slow-growing gene deletion backgrounds. Under the assumption that upon gene deletion, downregulation of certain other genes is harmful, restoration of wild-type expression level of these genes should partially restore fitness. Accordingly, mild gene overexpression of specific downregulated genes should be beneficial in the gene deletion, but not in the wild-type genetic background. We tested this scenario by increasing the transcript level of individual genes using a single-copy plasmid library (Molecular Barcoded Yeast [MoBY]-ORF 1.0, Ho et al. 2009). This overexpression system has been used to study the impact of single extra copy increase (Ho et al. 2009; Morrill and Amon 2019) and is well suited to capture mild gene expression changes that are frequently observed in transcriptomic data.

We selected seven slow-growing gene deletion mutants for experimental testing. The deleted genes are functionally diverse and are involved in metabolism, nuclear transport, signaling, and protein degradation (table 1). Importantly, we used the same gene knockout strains that had been subjected to transcriptome profiling and rigorous quality controls in a previous study (Kemmeren et al. 2014). Next, we selected five genes for overexpression. Each of these genes is downregulated in at least two strains carrying deletions in functionally dissimilar genes. Importantly, there is no transcriptional regulator among the deleted genes and there is no known physical interaction between the deleted and downregulated genes (Oughtred et al. 2019). In addition, one of the five downregulated genes (*RPS5*) is haploinsufficient (Deutschbauer et al. 2005). We introduced the corresponding MoBY-ORF plasmids into two to three single-gene knockout strains each, testing 11 gene knockout–downregulated gene combinations in total (table 1). We next estimated fitness using growth curve measurements (see supplementary Methods, Supplementary Material online).

Remarkably, out of these 11 combinations, we found five cases when mild overexpression of the downregulated gene significantly mitigated the fitness cost of the gene knockout (Wilcoxon test, table 1, fig. 5, and supplementary table S5, Supplementary Material online). Quantifying gene expression by quantitative PCR in three randomly selected cases confirmed that the plasmids fully restored the wild-type expression level of the downregulated genes in the corresponding deletion strains (supplementary fig. S10, Supplementary Material online, see supplementary Methods, Supplementary Material online, for details). Crucially, the fitness gains cannot be explained by a general beneficial effect of

Table 1. Experimental Test of Suboptimal Downregulation in Knockouts.

Deletion Mutant	Overexpressed Gene	Fitness Change upon Overexpression	Function of Deleted Gene	Function of Overexpressed Gene
Δ*opi3*	*SAM1*	13% (9.1–16%)	Phospholipid biosynthesis	Cofactor biosynthesis
Δ*opi3*	*RPS5*	7% (2.9–8.4%)	Phospholipid biosynthesis	Ribosomal subunit
Δ*seh1*	*SAM1*	7% (4.3–9.6%)	Nuclear transport	Cofactor biosynthesis
Δ*aco1*	*RPS5*	4% (1.1–6.6%)	Tricarboxylic acid cycle	Ribosomal subunit
Δ*sit4*	*EMI2*	4% (0.4–5.2%)	Protein phosphatase/cell cycle	Sporulation
Δ*ubp3*	*ADE12*	2% (0.1–3%)	Protein degradation	Nucleotide biosynthesis
Δ*ubp3*	*AST1*	1% (0–2.4%)	Protein degradation	Protein targeting
Δ*yvh1*	*EMI2*	1% (−0.8% and 1.7%)	Protein phosphatase/ribosome assembly	Sporulation
Δ*ctk1*	*SAM1*	−4% (−6.8% to 1.5%)	Protein kinase	Cofactor biosynthesis
Δ*opi3*	*ADE12*	−12% (−15% to −9.8%)	Phospholipid biosynthesis	Nucleotide biosynthesis
Δ*opi3*	*AST1*	−15% (−2% to −11%)	Phospholipid biosynthesis	Protein targeting

Note.—The table summarizes the results of 11 knockout–downregulated gene pairs that were experimentally tested harmful downregulation. Functional categories of the deleted and overexpressed genes (via the MoBY-ORF system) are indicated in the table. Blue rows indicate harmful downregulation, defined by a significant fitness increase upon overexpression that is also significantly higher than in the wild-type (WT) background (Wilcoxon and bootstrap tests, both at a false discovery rate of 1%). Red rows indicate a significant fitness decrease upon overexpression that is also significantly larger in magnitude than in the WT. White rows indicate no significant fitness change upon overexpression. Numbers indicate the fitness change in the knockout upon overexpression relative to that of the same strain with a control plasmid and 95% confidence intervals based on Wilcoxon tests. For details on the statistical analysis, see Materials and Methods and supplementary table S5, Supplementary Material online.

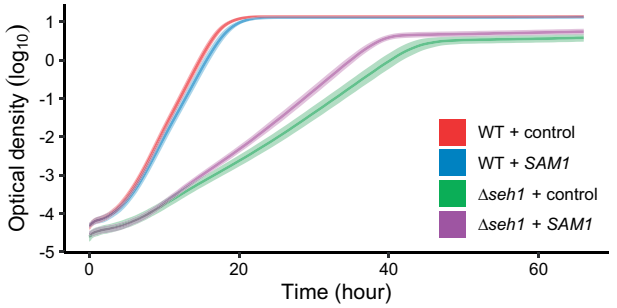

Fig. 5. Experimental test of suboptimal downregulation in knockouts. The figure shows the growth curves of strains with and without SAM1 overexpression from a single-copy plasmid (MoBY) in the wild-type (Δhis3) and Δseh1 backgrounds, respectively. SAM1 overexpression significantly increases the growth rate of Δseh1 background but not that of the WT. Average optical density values with 95% confidence bands based on 15 biological replicates are shown.

increased gene dosage for two reasons. First, overexpressions of the same genes do not increase fitness in the wild-type, indicating genetic interaction between overexpression and gene deletion (bootstrap test, Materials and Methods). Notably, some of the overexpressions were slightly harmful in wild-type background (supplementary fig. S9, Supplementary Material online). Second, for most gene overexpressions, a fitness gain was not observed in all tested deletion strains (table 1). Together, these findings indicate that downregulations of certain genes upon gene deletion are maladaptive and incur a measurable extra fitness cost.

In total, fitness was improved in four out of the seven tested gene knockout strains by at least one of the gene overexpressions (table 1). For example, the metabolic enzyme S-adenosylmethionine synthase (Sam1p) was downregulated upon deleting *SEH1*, a gene involved in regulating the central growth regulator TOR complex. Elevating *SAM1* expression level increased fitness by 7% (fig. 5). Deleting *SEH1* might indirectly trigger oxidative stress by decreasing the activity of the TOR complex (Panchaud et al. 2013), mirroring the oxidative stress-inducing effect of rapamycin (Mülleder et al. 2016). Downregulation of *SAM1* in *SEH1* knockout genetic background exacerbates this problem, as it decreases the production of the central antioxidant glutathione (Ye et al. 2017). This hypothesis needs to be tested in detail by future studies.

In summary, our results provide direct experimental evidence that suboptimal expression changes of even a single gene can incur a substantial fitness cost in gene deletion strains. We do not wish to claim however, that beneficial expression changes upon gene deletion do not occur, and indeed such putative cases appear in our data set (see table 1). Note that the existence of beneficial transcriptional changes in a particular gene deletion strain does not contradict the overall suboptimality of the transcriptional response since we were able to considerably improve fitness by overexpressing other genes in the very same knockout (table 1). Future works should explore how restoration of the wild-type expression level of multiple genes in gene deletion strains affect fitness.

Discussion

In this work, we systematically interrogated the genome-wide transcriptional response to gene deletions in yeast and asked whether it could contribute to the harmful fitness effect of such mutations. We showed that gene deletions frequently induce global transcriptomic rewiring. Most of these gene expression changes are not specific to individual gene deletions but are rather explained by a handful of recurrent expression patterns (i.e., TRMs) shared across null mutations. Specifically, 15 TRMs capture ∼60% of all transcriptional variation across ∼1,500 gene deletion strains (fig. 2). These recurrent TRMs are triggered by deletions in functionally diverse genes (fig. 2) and mimic responses to environmental perturbations (fig. 3). Together, these findings indicate that transcriptional rewiring is largely unspecific to the exact molecular function of the deleted gene and hence unlikely to mitigate the harmful effect of deletion. Indeed, in agreement

with previous works based on less comprehensive data (Wong and Roth 2005), we find that gene upregulations very rarely (2%) provide functional backup in the corresponding gene deletion backgrounds. Moreover, we present two complementary lines of evidences indicating that these transcriptional changes could even be harmful for the mutant cell. First, our systematic analysis revealed that gene deletions that alter the expression of dosage-sensitive genes display especially high-fitness defects (fig. 4). Second, by experimentally elevating the expression level of downregulated genes in specific deletion strains, we provide direct evidence that transcriptional changes contribute to the fitness defect of deletion (table 1 and fig. 5).

Based on the above findings, we propose that gene deletions evoke stereotypic gene expression responses that are repeatedly used across diverse environmental and genetic perturbations. Probably as a consequence of the shortage of knockout-specific regulatory responses, the genomic expression changes are suboptimal and contribute to the fitness impact of gene deletion. We offer two mutually nonexclusive evolutionary reasons why it should be so. First, compared with environmental changes, the rate at which individual loss-of-function mutations arise is generally low. Therefore, the strength of natural selection for the evolution of more specific and adaptive responses upon gene inactivation is generally weak under realistic population genetic settings (Proulx and Phillips 2005). Second, adaptive transcriptional responses to individual gene deletions may demand specific and potentially costly regulatory mechanisms, which could be difficult to evolve. Therefore, the cell may often rely on general and global transcriptional responses that are the result of natural selection to withstand frequent environmental stresses but are suboptimal to genetic perturbations that occur exceedingly rarely. The finding that the regulatory responses of yeast are well captured by a handful of TRMs suggests that optimal transcriptional rewiring upon genetic perturbations might be the exception, rather than the rule.

Our work has limitations that should be addressed in future works. First, although we established the existence of suboptimal expression changes upon gene deletion, its prevalence and overall contribution to fitness defects demand further in-depth studies. For instance, future empirical works should investigate which TRMs are triggered by slow growth and which ones contribute to the fitness defect themselves. Furthermore, it remains to be seen whether beneficial gene deletions (Qian et al. 2012) can be explained by suboptimal use of transcriptional modes in the wild-type. Second, our study focused on changes in gene expression when such changes are *not* needed or inadequate. The other side of the coin is equally interesting: It is feasible that cells do *not* change expression of certain genes when such changes would be beneficial. Third, optimality of the transcriptional response can vary across different types of mutations of the same gene. For example, beneficial expression responses might be more prevalent for mutations that do not prevent transcription of a gene but result in aberrant mRNAs, which could potentially induce the upregulation of compensating genes with similar sequences (Rossi et al. 2015; El-Brolosy et al. 2019; Ma et al. 2019). Fourth, our analysis identified 15 distinct TRMs, but only one of them, the slow growth signature, has been characterized in detail so far (O'Duibhir et al. 2014). The nature and potential functional role of the rest of TRMs remain to be explored. Finally, the specific molecular circuits underlying these transcriptional responses and how they are triggered by gene deletions and environmental changes are generally unknown. Notably, our analysis revealed a complex relationship between regulatory units and TRMs, which might stem from the highly interconnected nature of eukaryotic transcription regulatory networks. This contrasts with prokaryotes where regulatory networks exhibit simpler structures (Yan et al. 2010). Indeed, decomposition of transcriptional responses across different environmental and genetic perturbations in bacterial species resulted in independent modules that frequently represent the direct regulatory effect of a single or few transcriptional regulators (Sastry et al. 2019; Poudel et al. 2020). Thus, deciphering the regulatory interactions underlying eukaryotic TRMs will likely demand more sophisticated approaches.

Our study has relevance for understanding the link between modularity of cellular networks and robustness against mutations (Hartman et al. 2001). Our results indicate that genetic perturbations in various functional modules can elicit similar global transcriptional responses (supplementary fig. S4, Supplementary Material online). These findings are consistent with an earlier report on widespread transcriptomic and proteomic changes upon mutations in a particular metabolic gene (Bershtein et al. 2015). Our work generalizes this case study by showing that mutations in various genes have nonspecific transcriptomic effects. We demonstrated that such extensive regulatory responses can aggravate the fitness cost of the focal mutation. Thus, cellular networks can not only diminish the phenotypic effects of mutations but also contribute to mutational fragility (Diss et al. 2017).

The presence of nonspecific and suboptimal regulatory responses may also have profound importance on interpreting the results of genetic perturbation studies. First, analyzing the transcriptomes of perturbed signaling and regulatory genes is often employed to dissect pathway memberships (van Wageningen et al. 2010; Lenstra et al. 2011). However, transcriptional responses are not always pathway specific (Apweiler et al. 2012) and can be further confounded by global gene expression signatures (O'Duibhir et al. 2014). Indeed, removing the slow growth signature improves identification of direct target genes of perturbed regulators (O'Duibhir et al. 2014; Chagoyen and Poyatos 2019). Future works should explore whether factoring out additional recurring TRMs would further help reveal the direct functional effects of genetic perturbations. Second, suboptimal transcriptional responses may confound the interpretation of chemical-genetic and genetic interactions screens. A large fitness defect upon deleting a particular gene may indicate its importance associated to the gene's biochemical function or the deleterious side effects of the deletion on genomic expression. Clearly, the relative contribution of these two factors remains to be tested using more systematic measurements of the fitness impacts of expression changes in gene

deletion backgrounds. Third, it has been recently suggested that knowing the degree of evolutionary conservation of the transcriptional responses in gene deletion strains might further help to distinguish between responses that are functionally coupled to the deleted genes' biochemical activities and those that are pleiotropic and evolutionary ad hoc (Liu et al. 2020). Future works should examine whether nonconserved responses contribute to the suboptimality described in the present work.

Our results are consistent with the "induced essentiality" model of synthetic genetic interactions, which posits that deletion of one gene results in a rearrangement of the genetic network into an alternative viable state where a second gene becomes more important for fitness (Tischler et al. 2008). This model predicts that large expression rearrangements should modulate the fitness impact of numerous other genes. Our observation that gene deletions inducing substantial transcriptomic rewiring also frequently show synthetic genetic interactions (Costanzo et al. 2016) supports this idea (supplementary fig. S11, Supplementary Material online). Importantly, this correlation also holds after removing the slow growth gene expression signature (Spearman's $\rho = 0.48$, P value $< 10^{-16}$, Materials and Methods).

More broadly, our work suggests that it is possible to modulate mutational effects through changes in gene expression of functionally unrelated genes. Deciphering the molecular mechanisms by which the phenotype of a mutation is modified by the genetic background is crucial both for our understanding of how genotype determines phenotype in natural populations and for predicting the severity of disease mutations (Nadeau 2003; Vu et al. 2015). Recent works demonstrated that variation in gene expression level between individuals substantially modulates the manifestation of mutations (Vu et al. 2015). In particular, the phenotypic effect of a given mutation is modified by the expression level of other genes within the same functional module. Our results raise the possibility that more global regulatory changes driven by mutations in functionally unrelated genes could also underlie genetic background effects. Future works should elucidate whether such nonspecific regulatory responses contribute to natural variation in the severity of mutant phenotypes, including those of Mendelian disorders in humans.

Materials and Methods

Transcriptome Data Set

To investigate expression changes, we used the genome-wide transcriptome profiles for ~1,500 single-gene deletion strains (Kemmeren et al. 2014). Up- and downregulations were defined based on fold-change (FC $>$ 1.7) and P value cutoff ($P < 0.05$), following Kemmeren et al. (2014). When controlling for the slow growth gene expression signature, we used the transformed gene expression data of O'Duibhir et al. (2014). In this data set, the first principal component, the so called "slow growth signature," was subtracted from the original transcript level changes. Up- and downregulations were defined based on the transformed fold-change values (FC $>$ 1.7). To filter out genes with transcription-related function from the set of gene deletions, we considered functional categories "chromatin factor," "Pol II transcription," "gene-specific transcription factor," and "RNA metabolism" according to Kemmeren et al. (2014).

Growth Rate Data

Growth rates (fitness) of gene deletion strains relative to the wild-type were obtained from (O'Duibhir et al. 2014). In brief, growth rate was determined by growth curve measurement under the same culture condition as for transcriptome analysis (Kemmeren et al. 2014).

Identification of Recurrent TRMs

Following the same approach used in O'Duibhir et al. (2014), we applied a method equivalent to PCA (Singular Value Decomposition) to decompose the transcriptome data set into a weighted sum of simpler expression patterns ("modes"), ordered by how much of the variation in the data they explain. Such TRMs represent "stereotypic" global expression changes, that is, groups of genes that change their expression concomitantly in the same way in a recurrent way across gene deletion strains.

To evaluate the extent to which these TRMs reflect relevant biological responses instead of purely random associations between gene expression changes, we compared the percentage of the variation in the data set explained by each TRM with that obtained from the first principal component of randomized data matrices. The number of significant independent TRMs in the real data was defined as follows. PCA is performed over the original expression data set, as well as over randomized instances of it. The threshold for estimating the number of TRMs is set as the mean + 2 SD in the variance explained by TRM #1 in 600 randomized data sets. Note that the variation explained by the first principal component in the randomized data sets is the maximal amount of variation that any TRM can explain in randomized data sets. Thus, our procedure yields a conservative null model for the identification of significant modes.

The data matrix was randomized assuming that all genes change their expression independently in each gene deletion. Random recurrent associations between specific gene expression changes could be magnified by two characteristics of transcription in yeast. First, some genes have a strong general tendency to change their expression irrespective of the perturbation (e.g., those containing TATA boxes in the promoter). Second, the number of gene expression changes per deletion shows a highly skewed distribution, with a few deletions changing the expression of hundreds of genes, and hundreds of deletions changing the expression of less than three genes. To include these structural constraints in our null model, we devised a method that randomizes the full data matrix of $\log_2$ expression ratios (M matrix of n transcripts $\times$ k deletion strains) while preserving the total amount of gene expression variation per row and column. Our algorithm for this is as follows:

(1) Compute a "gene expression change propensity" matrix W, where the value of each element w_{ij} is a probability

that is proportional to a function of the corresponding sum row values and the sum of column values in the M matrix (gene expression change relative to wild-type) $w_{ij} = f\left(\sum_{j=1}^{K} M_{ij}, \sum_{i=1}^{N} M_{ij}\right)$. We empirically observed that the function f that works best is the product of the row and column sums, $w_{ij} = f\left(\sum_{j=1}^{K} M_{ij}, \sum_{i=1}^{N} M_{ij}\right)^2$. Finally, we normalize the matrix W to sum one, which gives us a matrix of probabilities P.

(2) Construct a vector G with all the values of the M matrix sorted by descending absolute value.

(3) Build an empty matrix R. Sequentially choose entries in R without replacement according to the probabilities in P and assign them the values in G.

This algorithm stochastically assigns the values from the original M matrix to the entries in the randomized R matrix, with the highest values having more probability to be assigned to the entries having the highest gene expression change propensity. By doing this, the values are randomized while preserving the structural constraints of the original matrix, that is, gene expression values in the original matrix, row and column sums (fig. 2A). For computational efficiency, the weighted random sampling in the algorithm was done using the sample_int_crank() function from the wrswoR R package, which implements the efficient sampling method from Efraimidis and Spirakis (2006).

Identification of Deletions Triggering TRMs and Transcripts Affected by TRMs

In order to define gene deletions that trigger a given recurrent TRM, we identified pairs of deletions and TRMs with the following property: the TRM explains more than 10% of the variance in the transcriptomic changes observed in the particular deletion strain. Furthermore, to classify deletions that correlate positively or negatively with a given TRM, we defined a "signature profile" for each TRM following O'Duibhir et al. (2014). The signature profile is the profile of the deletion strain that shows the largest amount of transcriptomic changes (largest Euclidean norm) in a given TRM. Next, for each TRM, we classified deletions into those that correlate positively or negatively with the signature profile. To characterize the functions of these triggering genes, we carried out GO enrichment analysis. We tested separately those subsets of deletions that correlate positively or negatively with the signature profile.

To identify which genes' expression is affected by each TRM, we used the following criterion. We considered that TRM i contributes to the expression change of gene j if it contributes with at least a 1.3 $\log_2$ fold-change in at least one deletion background in which gene i shows a significant expression change. This equals to half of the $\log_2$ 1.7 fold-change threshold. To characterize the functions of the TRM-affected genes, we carried out GO enrichment analysis. We calculated the enrichment separately for genes which were upregulated and downregulated, respectively, in the signature profile. GO Slim processes were downloaded from the SGD database (http://sgd-archive.yeastgenome.org/curation/literature/version, October 25, 2019). Noninformative categories ("biological_process," "other," and "not_yet_annotated") and categories with less than four genes from our data set were removed. One-tailed Fisher's exact test was used to test the enrichments.

Environmental Gene Expression Compendium

We downloaded the expression profiles corresponding to environmental conditions from Saccharomyces Genome Database (SPELL [Hibbs et al. 2007]). Biological replicates were averaged together in each case. In the case of two-channel arrays, original $\log_2$ ratios were kept. In the case of one-channel arrays, log 2 ratios were calculated, taking as reference the array from the same data set that was most similar to a WT control in standard culture conditions. After compiling the data set, profiles where more than 75% of the genes were missing values were removed. In the remaining, KNN-imputation of missing values was performed using the function provided in R package "impute."

In order to compare environmental expression profiles with TRMs, we calculated the absolute Spearman correlation coefficients between each environmental profile and the signature profile of each TRM (supplementary table S3, Supplementary Material online). We applied this rank-based measure of profile similarity due to heterogeneity in the gene expression data sources. Because of the large sample sizes, very small effect sizes appeared statistically significant, therefore we used an absolute correlation coefficient of 0.2 as a conservative threshold to define substantial similarity between TRMs and environmental profiles. Note that this is well above the correlations expected by chance. Randomly expected similarities between TRMs and environmental profiles were obtained by computing the Spearman correlation between TRM #1 from 600 randomized matrices, on the one hand, and the environmental expression profiles, on the other hand.

Analysis of Regulation by TFs

The data set mapping all *S. cerevisiae* TFs to the genes they bind was downloaded from Transfac (Monteiro et al. 2020) (http://www.yeastract.com/) on February 19, 2020. Hypergeometric distribution was used to assess enrichment in the genes responding to each TRM and FDR correction was applied to P values. We calculated the enrichment separately for genes which were upregulated and downregulated, respectively, in the signature profile.

Testing the Enrichment of Upregulations among Synthetic Sick or Lethal Genetic Interactions

To statistically test whether genes that are upregulated upon gene deletions more often show a synthetic sick or lethal genetic interaction with the deleted genes than expected by chance, we carried out a randomization test. We first calculated our test statistics: the fraction of upregulations among gene pairs being in synthetic sick or lethal genetic interaction. We then calculated the same fractions for 1,000 randomized data sets, obtaining a randomized distribution. During randomization, we randomly permuted the links

between the deleted and upregulated genes, thus the number of upregulations for each deletion and transcript change remained constant. We defined *P* value as the number of random data sets where the fraction of gene pairs is higher or equal than the empirical value + 1 divided by the number of randomizations + 1.

A comprehensive set of negative genetic interactions (i.e., synthetic sick or lethal gene pairs) was obtained from Costanzo et al. (2016), using their stringent cutoff. Temperature-sensitive alleles and reciprocal gene pairs showing opposite genetic interactions were removed from the data set.

To examine whether the enrichment of upregulation is significant for both low- and high-fitness knockouts, we calculated the test statistics for high- and low-fitness deletion strains separately after randomizing the gene expression data set. High- and low-fitness deletion strains were defined as those with a relative fitness value higher and lower than 0.95, respectively. Transcription-related genes were defined as described above.

Testing the Association between Knockout Fitness and Altered Expression of Dosage-Sensitive Genes

We used previously defined sets of overexpression-sensitive (Makanae et al. 2013) and haploinsufficient (Deutschbauer et al. 2005) genes to test whether knockout strains with at least one upregulated overexpression sensitive genes or at least one downregulated haploinsufficient genes have an especially low fitness. To control for the positive correlation between the number of upregulated (downregulated) genes with both fitness and the number of upregulated overexpression-sensitive (downregulated haploinsufficient) genes, we fit a loess curve (degree = 2, span = 1) between the $\log_{10}$-scaled number of upregulations (downregulations) and growth rate across those strains without any upregulated overexpression-sensitive (downregulated haploinsufficient) genes and calculated residual fitness values for each strain within this range of total upregulations (downregulations) (supplementary fig. S8, Supplementary Material online). Then, we tested the correlation between the residual fitness values and the number of upregulated overexpression-sensitive (downregulated haploinsufficient) genes by using two-tailed Spearman correlation tests.

Yeast Strains and Media

All strains used in this study were derived from the BY4742 *S. cerevisiae* parental MAT alpha strain. Nonessential single-gene deletion strains from the MAT alpha haploid yeast deletion collection (Open Biosystems; *his3Δ1, leu2Δ0, lys2Δ0, ura3Δ0, xxx::KanMX4* [Giaever et al. 2002]) were selected based on the criteria listed below. Deletion strains and the wild-type strain were transformed with single-copy plasmid clones of the MoBY ORF library (Ho et al. 2009; Open Biosystems) carrying the *URA3* auxotrophic marker. As a control plasmid, the *HO* (*YDL227C*) plasmid clone of the same library was used throughout the study. Yeast transformations were carried out using the standard lithium acetate method (Gietz and Schiestl 2007). Transformants were selected on uracil-dropout synthetic complete medium (SC-uracil: 5 g/l ammonium sulfate, 1.7 g/l Yeast Nitrogen Base, supplemented with 2% glucose and with amino acid mix, without uracil). We used the *HIS3* (*YOR202W*) deletion strain (*YOR202W::KanMX4*) as wild-type strain, for the following reasons: 1) fitness of this strain is indistinguishable from the BY4743 parental wild-type strain (Qian et al. 2012); 2) it carries the same selection marker (*KanMX4*) as all other single-gene deletion strains in the nonfunctional *his3Δ1* allele.

Fitness and Genetic Interaction Measurements

Established protocols were used to measure the fitness effect of increasing the expression level of downregulated genes in populations of slow-growing gene deletion backgrounds (Szamecz et al. 2014). For details, see supplementary Methods, Supplementary Material online.

To define harmful downregulations, we simultaneously applied the following two criteria. First, the deletion strain should have a significantly higher fitness when the tested downregulated gene is overexpressed compared with the case when the deletion strain harbors only a control plasmid. *P* values were calculated and corrected for multiple testing (1% false discovery rate threshold) by two-tailed Wilcoxon rank sum tests. Second, the overexpressed gene and the deleted gene should display a significant positive genetic interaction. Genetic interaction score was calculated as $\varepsilon = (f_{ab} \times f_{WT}) - (f_a \times f_b)$, where f_{WT} is the fitness of the wild-type with control plasmid, f_a is the fitness of the deletion strain with control plasmid, f_b is the fitness of the wild-type strain with overexpressing the tested gene, and f_{ab} is the fitness of the deletion strain with overexpressing the tested gene. A positive ($\varepsilon > 0$) interaction score indicates that the fitness increase caused by the gene overexpression in the deletion strain is higher than expected based on the multiplicative model. We tested the null hypothesis that $\varepsilon = 0$. Two-tailed *P* values were calculated and corrected for multiple testing (1% false discovery rate threshold) using the bootstrap method implemented in the "boot" R package (Canty and Ripley 2020), resampling f_a, f_b, f_{WT}, and f_{ab} separately. In contrast to the above criteria, if overexpressing the downregulated gene resulted in both a significant fitness defect in the deletion strain and a significant negative genetic interaction score, we considered the particular downregulation beneficial in the deletion background (indicated by red in table 1).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

This work was supported by the "Lendület" program of the Hungarian Academy of Sciences (LP2009-013/2012 to B.P., LP 2012-32/2018, and LP-2017-10/2017 to C.P.), the Wellcome Trust WT (098016/Z/11/Z and 084314/Z/07/Z to B.P. and C.P.), The European Research Council (H2020-ERC-2014-CoG 648364- Resistance Evolution to C.P.), National Research, Development and Innovation Office Élvonal

Program (KKP 126506 to C.P., GINOP-2.3.2-15-2016-00014 to C.P., GINOP-2.3.2-15-2016-00020 [MolMedEx TUMORDNS] to C.P., and GINOP-2.3.2-15-2016-00026 [iChamber] to B.P.), The European Union's Horizon 2020 research and innovation program (Grant No. 739593 to B.P. and No 739593 to B.K.), the Hungarian Academy of Sciences Postdoctoral Fellowship Program (Postdoc2014-85 to K.K.), National Research, Development and Innovation Office (FK 128775 to Z.F. and FK 128916 to D.K.), Janos Bolyai Research Fellowship from the Hungarian Academy of Sciences (BO/779/20 to Z.F. and BO/352/20 to B.K.), New National Excellence Program of the Ministry of Human Capacities (Bolyai+, UNKP-20-5-SZTE-646 to Z.F. and UNKP-20-5-SZTE-654 to B.K.), The Spanish Ministerio de Economía y Competitividad (Grant No. FIS2016-78781-R to J.F.P.), and European Molecular Biology Organization short-term fellowship (to D.B.).

Data Availability

Scripts used in the analysis of stereotypic modes are available at https://github.com/pappb/Kovacs-et-al-Suboptimal-transcription-MBE.

References

Apweiler E, Sameith K, Margaritis T, Brabers N, van de Pasch L, Bakker LV, van Leenen D, Holstege FC, Kemmeren P. 2012. Yeast glucose pathways converge on the transcriptional regulation of trehalose biosynthesis. *BMC Genomics* 13(1):239.

Bershtein S, Choi J-M, Bhattacharyya S, Budnik B, Shakhnovich E. 2015. Systems-level response to point mutations in a core metabolic enzyme modulates genotype–phenotype relationship. *Cell Rep.* 11(4):645–656.

Brauer MJ, Huttenhower C, Airoldi EM, Rosenstein R, Matese JC, Gresham D, Boer VM, Troyanskaya OG, Botstein D. 2008. Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast. *Mol Biol Cell* 19(1):352–367.

Castrillo JI, Zeef LA, Hoyle DC, Zhang N, Hayes A, Gardner DC, Cornell MJ, Petty J, Hakes L, Wardleworth L, et al. 2007. Growth control of the eukaryote cell: a systems biology study in yeast. *J Biol.* 6(2):4.

Canty A, Ripley BD. 2020. boot: Bootstrap R (S-Plus) Functions. Available from: https://cran.r-project.org/web/packages/boot/.

Chagoyen M, Poyatos JF. 2019. Complex genetic and epigenetic regulation deviates gene expression from a unifying global transcriptional program. *PLoS Comput Biol.* 15(9):e1007353.

Cooper TF, Ostrowski EA, Travisano M. 2007. A negative relationship between mutation pleiotropy and fitness effect in yeast. *Evolution* 61(6):1495–1499.

Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, Tan G, Wang W, Usaj M, Hanchard J, Lee SD, et al. 2016. A global interaction network maps a wiring diagram of cellular function. *Science* 353(6306):aaf1420.

Deutschbauer AM, Jaramillo DF, Proctor M, Kumm J, Hillenmeyer ME, Davis RW, Nislow C, Giaever G. 2005. Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics* 169(4):1915–1925.

Diss G, Gagnon-Arsenault I, Dion-Coté A-M, Vignaud H, Ascencio DI, Berger CM, Landry CR. 2017. Gene duplication can impart fragility, not robustness, in the yeast protein interaction network. *Science* 355(6325):630–634.

Efraimidis PS, Spirakis PG. 2006. Weighted random sampling with a reservoir. *Inf Process Lett.* 97(5):181–185.

El-Brolosy MA, Kontarakis Z, Rossi A, Kuenne C, Günther S, Fukuda N, Kikhi K, Boezio GLM, Takacs CM, Lai S-L, et al. 2019. Genetic compensation triggered by mutant mRNA degradation. *Nature* 568(7751):193–197.

Gasch AP. 2003. The environmental stress response: a common yeast response to diverse environmental stresses. In: Hohmann S, Mager WH, editors. Yeast stress responses. topics in current genetics. Berlin/Heidelberg: Springer. p. 11–70.

Giaever G, Chu AM, 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(6896):387–391.

Gietz RD, Schiestl RH. 2007. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc.* 2(1):31–34.

Hartman JL, Garvik B, Hartwell L. 2001. Principles for the buffering of genetic variation. *Science* 291(5506):1001–1004.

He X, Zhang J. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* 169(2):1157–1164.

Hibbs MA, Hess DC, Myers CL, Huttenhower C, Li K, Troyanskaya OG. 2007. Exploring the functional landscape of gene expression: directed search of large microarray compendia. *Bioinformatics* 23(20):2692–2699.

Ho CH, Magtanong L, Barker SL, Gresham D, Nishimura S, Natarajan P, Koh JLY, Porter J, Gray CA, Andersen RJ, et al. 2009. A molecular barcoded yeast ORF library enables mode-of-action analysis of bioactive compounds. *Nat Biotechnol.* 27(4):369–377.

Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai H, He YD, et al. 2000. Functional discovery via a compendium of expression profiles. *Cell* 102(1):109–126.

Iwahashi H, Kitagawa E, Suzuki Y, Ueda Y, Ishizawa Y, Nobumasa H, Kuboki Y, Hosoda H, Iwahashi Y. 2007. Evaluation of toxicity of the mycotoxin citrinin using yeast ORF DNA microarray and Oligo DNA microarray. *BMC Genomics* 8(1):95.

Kafri M, Metzl-Raz E, Jona G, Barkai N. 2016. The cost of protein production. *Cell Rep.* 14(1):22–31.

Kafri R, Bar-Even A, Pilpel Y. 2005. Transcription control reprogramming in genetic backup circuits. *Nat Genet.* 37(3):295–299.

Kemmeren P, Sameith K, van de Pasch LAL, Benschop JJ, Lenstra TL, Margaritis T, O'Duibhir E, Apweiler E, van Wageningen S, Ko CW, et al. 2014. Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. *Cell* 157(3):740–752.

Keren L, Hausser J, Lotan-Pompan M, Vainberg Slutskin I, Alisar H, Kaminski S, Weinberger A, Alon U, Milo R, Segal E. 2016. Massively parallel interrogation of the effects of gene expression levels on fitness. *Cell* 166(5):1282–1294.e18.

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(1):701.

Knijnenburg TA, de Winde JH, Daran J-M, Daran-Lapujade P, Pronk JT, Reinders MJ, Wessels LF. 2007. Exploiting combinatorial cultivation conditions to infer transcriptional regulation. *BMC Genomics* 8(1):25.

Lenstra TL, Benschop JJ, Kim T, Schulze JM, Brabers NACH, Margaritis T, van de Pasch LAL, van Heesch SAAC, Brok MO, Groot Koerkamp MJA, et al. 2011. The specificity and topology of chromatin interaction pathways in yeast. *Mol Cell* 42(4):536–549.

Levy S, Ihmels J, Carmi M, Weinberger A, Friedlander G, Barkai N. 2007. Strategy of transcription regulation in the budding yeast. *PLoS One* 2(2):e250.

Liu L, Liu M, Zhang D, Deng S, Chen P, Yang J, Xie Y, He X. 2020. Decoupling gene functions from knockout effects by evolutionary analyses. *Natl Sci Rev.* 7(7):1169–1180.

Ma Z, Zhu P, Shi H, Guo L, Zhang Q, Chen Y, Chen S, Zhang Z, Peng J, Chen J. 2019. PTC-bearing mRNA elicits a genetic compensation response via Upf3a and COMPASS components. *Nature* 568(7751):259–263.

Makanae K, Kintaka R, Makino T, Kitano H, Moriya H. 2013. Identification of dosage-sensitive genes in *Saccharomyces cerevisiae* using the genetic tug-of-war method. *Genome Res.* 23(2):300–311.

Monteiro PT, Oliveira J, Pais P, Antunes M, Palma M, Cavalheiro M, Galocha M, Godinho CP, Martins LC, Bourbon N, et al. 2020. YEASTRACT+: a portal for cross-species comparative genomics of

transcription regulation in yeasts. *Nucleic Acids Res.* 48(D1):D642–D649.
Morrill SA, Amon A. 2019. Why haploinsufficiency persists. *Proc Natl Acad Sci U S A*. 116(24):11866–11871.
Mülleder M, Calvani E, Alam MT, Wang RK, Eckerstorfer F, Zelezniak A, Ralser M. 2016. Functional metabolomics describes the yeast biosynthetic regulome. *Cell* 167(2):553–565.e12.
Nadeau JH. 2003. Modifier genes and protective alleles in humans and mice. *Curr Opin Genet Dev*. 13(3):290–295.
O'Duibhir E, Lijnzaad P, Benschop JJ, Lenstra TL, Leenen D, Groot Koerkamp MJ, Margaritis T, Brok MO, Kemmeren P, Holstege FC. 2014. Cell cycle population effects in perturbation studies. *Mol Syst Biol*. 10(6):732.
Oughtred R, Stark C, Breitkreutz B-J, Rust J, Boucher L, Chang C, Kolas N, O'Donnell L, Leung G, McAdam R, et al. 2019. The BioGRID interaction database: 2019 update. *Nucleic Acids Res*. 47(D1):D529–D541.
Panchaud N, Péli-Gulli M-P, De Virgilio C. 2013. SEACing the GAP that nEGOCiates TORC1 activation. *Cell Cycle* 12(18):2948–2952.
Pinkham JL, Guarente L. 1985. Cloning and molecular analysis of the HAP2 locus: a global regulator of respiratory genes in *Saccharomyces cerevisiae*. *Mol Cell Biol*. 5(12):3410–3416.
Poudel S, Tsunemoto H, Seif Y, Sastry AV, Szubin R, Xu S, Machado H, Olson CA, Anand A, Pogliano J, et al. 2020. Revealing 29 sets of independently modulated genes in *Staphylococcus aureus*, their regulators, and role in key physiological response. *Proc Natl Acad Sci U S A*. 117(29):17228–17239.
Price MN, Deutschbauer AM, Skerker JM, Wetmore KM, Ruths T, Mar JS, Kuehl JV, Shao W, Arkin AP. 2013. Indirect and suboptimal control of gene expression is widespread in bacteria. *Mol Syst Biol*. 9(1):660.
Proulx SR, Phillips PC. 2005. The opportunity for canalization and the evolution of genetic networks. *Am Nat*. 165(2):147–162.
Qian W, Ma D, Xiao C, Wang Z, Zhang J. 2012. The genomic landscape and evolutionary resolution of antagonistic pleiotropy in yeast. *Cell Rep*. 2(5):1399–1410.
Regenberg B, Grotkjær T, Winther O, Fausbøll A, Åkesson M, Bro C, Hansen LK, Brunak S, Nielsen J. 2006. Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in *Saccharomyces cerevisiae*. *Genome Biol*. 7(11):R107.
Rossi A, Kontarakis Z, Gerri C, Nolte H, Hölper S, Krüger M, Stainier DYR. 2015. Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature* 524(7564):230–233.
Saldanha AJ, Brauer MJ, Botstein D. 2004. Nutritional homeostasis in batch and steady-state culture of yeast. *Mol Biol Cell* 15(9):4089–4104.
Sastry AV, Gao Y, Szubin R, Hefner Y, Xu S, Kim D, Choudhary KS, Yang L, King ZA, Palsson BO. 2019. The *Escherichia coli* transcriptome mostly consists of independently regulated modules. *Nat Commun*. 10:1–14.
Szamecz B, Boross G, Kalapis D, Kovács K, Fekete G, Farkas Z, Lázár V, Hrtyan M, Kemmeren P, Koerkamp MJAG, et al. 2014. The genomic landscape of compensatory evolution. *PLoS Biol*. 12(8):e1001935.
Szappanos B, Kovács K, Szamecz B, Honti F, Costanzo M, Baryshnikova A, Gelius-Dietrich G, Lercher MJ, Jelasity M, Myers CL, et al. 2011. An integrated approach to characterize genetic interaction networks in yeast metabolism. *Nat Genet*. 43(7):656–662.
Tischler J, Lehner B, Fraser AG. 2008. Evolutionary plasticity of genetic interaction networks. *Nat Genet*. 40(4):390–391.
Towbin BD, Korem Y, Bren A, Doron S, Sorek R, Alon U. 2017. Optimality and sub-optimality in a bacterial growth law. *Nat. Commun*. 8:14123.
van Wageningen S, Kemmeren P, Lijnzaad P, Margaritis T, Benschop JJ, de Castro IJ, van Leenen D, Groot Koerkamp MJA, Ko CW, Miles AJ, et al. 2010. Functional overlap and regulatory links shape genetic interactions between signaling pathways. *Cell* 143(6):991–1004.
Vu V, Verster AJ, Schertzberg M, Chuluunbaatar T, Spensley M, Pajkic D, Hart GT, Moffat J, Fraser AG. 2015. Natural variation in gene expression modulates the severity of mutant phenotypes. *Cell* 162(2):391–402.
Wong SL, Roth FP. 2005. Transcriptional compensation for gene loss plays a minor role in maintaining genetic robustness in *Saccharomyces cerevisiae*. *Genetics* 171(2):829–833.
Yan K-K, Fang G, Bhardwaj N, Alexander RP, Gerstein M. 2010. Comparing genomes to computer operating systems in terms of the topology and evolution of their regulatory control networks. *Proc Natl Acad Sci U S A*. 107(20):9186–9191.
Ye C, Sutter BM, Wang Y, Kuang Z, Tu BP. 2017. A metabolic function for phospholipid and histone methylation. *Mol Cell* 66(2):180–193.e8.