

Large-Scale Genetic Perturbations Reveal Regulatory Networks and an Abundance of Gene-Specific Repressors

Patrick Kemmeren,^{1,3} Katrin Sameith,^{1,3} Loes A.L. van de Pasch,^{1,3} Joris J. Benschop,^{1,3} Tineke L. Lenstra,^{1,3} Thanasis Margaritis,^{1,3} Eoghan O'Duibhir,¹ Eva Apweiler,¹ Sake van Wageningen,¹ Cheuk W. Ko,¹ Sebastiaan van Heesch,¹ Mehdi M. Kashani,¹ Giannis Ampatzidis-Michailidis,¹ Mariel O. Brok,¹ Nathalie A.C.H. Brabers,¹ Anthony J. Miles,¹ Diane Bouwmeester,¹ Sander R. van Hooff,¹ Harm van Bakel,¹ Erik Sluiters,¹ Linda V. Bakker,¹ Berend Snel,² Philip Lijnzaad,¹ Dik van Leenen,¹ Marian J.A. Groot Koerkamp,¹ and Frank C.P. Holstege^{1,*}

¹Molecular Cancer Research, University Medical Centre Utrecht, Universiteitsweg 100, Utrecht 3584 CG, the Netherlands

²Theoretical Biology and Bioinformatics, Department of Biology, Utrecht University, Utrecht 3584 CG, the Netherlands

³Co-first author

*Correspondence: f.c.p.holstege@umcutrecht.nl

<http://dx.doi.org/10.1016/j.cell.2014.02.054>

SUMMARY

To understand regulatory systems, it would be useful to uniformly determine how different components contribute to the expression of all other genes. We therefore monitored mRNA expression genome-wide, for individual deletions of one-quarter of yeast genes, focusing on (putative) regulators. The resulting genetic perturbation signatures reflect many different properties. These include the architecture of protein complexes and pathways, identification of expression changes compatible with viability, and the varying responsiveness to genetic perturbation. The data are assembled into a genetic perturbation network that shows different connectivities for different classes of regulators. Four feed-forward loop (FFL) types are overrepresented, including incoherent type 2 FFLs that likely represent feedback. Systematic transcription factor classification shows a surprisingly high abundance of gene-specific repressors, suggesting that yeast chromatin is not as generally restrictive to transcription as is often assumed. The data set is useful for studying individual genes and for discovering properties of an entire regulatory system.

INTRODUCTION

Cells depend on many intricate molecular interactions to successfully perform a myriad of functions in an integrative manner. One of the current challenges of molecular biology is to determine and study all interactions important for cellular function (Ideker et al., 2001). This is inspired by increased awareness that complex properties can emerge from combinations of relatively few simple interactions. Systematic interaction analyses

are being realized through high-throughput approaches and are required to understand many aspects of living organisms, including disease (Vidal et al., 2011). Whereas some interactions are physically direct, e.g., protein-protein interactions (Walhout and Vidal, 2001), others can be more abstract, e.g., genetic interactions (Costanzo et al., 2010). Both are informative, either for the function of individual components or for properties of the entire system. Various data sets, generated to different degrees of accuracy and completion, have successfully been applied to study cellular systems. One such system is mRNA expression. To study the regulatory network underlying mRNA expression, it would be useful to determine how different cellular components influence mRNA expression genome-wide.

It is well established that perturbation of individual factors, followed by genome-wide expression analysis, can yield insight into function (DeRisi et al., 1997; Holstege et al., 1998). Regulatory pathways (Roberts et al., 2000) and protein complexes (van de Peppel et al., 2005) can be similarly studied, additionally revealing functional relationships between components. Focusing on functionally uncharacterized genes, a pioneering study of 276 mutants in the yeast *Saccharomyces cerevisiae* first demonstrated the utility of much larger collections of genetic perturbation expression signatures (Hughes et al., 2000). This has been followed by studies of many factors individually, as well as of entire classes of regulators (Hu et al., 2007; van Wageningen et al., 2010; Lenstra et al., 2011) also incorporating other types of perturbation (Chua et al., 2006; Weiner et al., 2012).

Despite many other advances, the number of genetic perturbations analyzed within such studies has not increased significantly since the first compendium (Hughes et al., 2000), likely for logistical reasons. Although many genetic perturbations have been analyzed, analysis of entire systems has been hampered, in particular because of difficulties inherent to properly comparing gene expression data generated across the different conditions, genetic backgrounds, technology platforms, types of controls, and degrees of replication in different studies. Here, we report mRNA expression profiles uniformly generated for deletion of one-quarter of all protein-coding genes

in *S. cerevisiae*. By making particular use of data uniformity and the causal relationships inherent to genetic perturbation, the data are analyzed at different levels of complexity to study fundamental properties of the underlying regulatory system.

RESULTS

mRNA Expression Profiles of 1,484 Deletion Mutants

To systematically investigate the regulatory network of a model organism, expression changes were determined genome-wide for haploid *S. cerevisiae* strains bearing single gene deletions (Giaever et al., 2002). Selection was based on the deleted gene having a (putative) role in regulating gene expression. Selection also included characteristics such as nuclear location or the capacity to modify other proteins. The 1,484 mutants cover many different functional categories, including gene-specific and global transcription factors (TFs), RNA processing and export, ubiquitin(-like) modifications, protein kinases/phosphatases, protein trafficking, cell cycle, meiosis, and DNA replication and repair (Figure S1A and Table S1 available online).

Various strategies were incorporated to ensure a high degree of accuracy and precision (Experimental Procedures). This included four replicates per responsive mutant, robotic procedures optimized with external calibration controls (van Bakel and Holstege, 2004), a common reference design with wild-type (WT) reference RNA applied in dye-swap to each microarray (Figure S1B), as well as dye-bias correction (Margaritis et al., 2009) and spike-in controls to monitor global changes (van de Peppel et al., 2003). Additional WT cultures were processed alongside batches of mutants, with day-specific effects countered by regrowing the entire batch. Statistical modeling results in an average expression profile for each mutant. Each profile consists of p values and average transcript level changes in the mutant relative to 428 WTs. Further controls for consistency, aneuploidy, and correct gene deleted resulted in 101 deletion strains being remade and reprofiled (Table S1). Consistency controls included analysis alongside strains from the same protein complex or pathway, resulting in remaking strains with suspected secondary mutations (Teng et al., 2013). These technical aspects were uniformly applied to the entire data set, some of which has been used previously (Table S1). With coverage of one-quarter of all genes and one-third of all genes not required for viability, this constitutes the largest collection of uniformly generated expression signatures for genetic perturbations.

Response to Genetic Perturbation

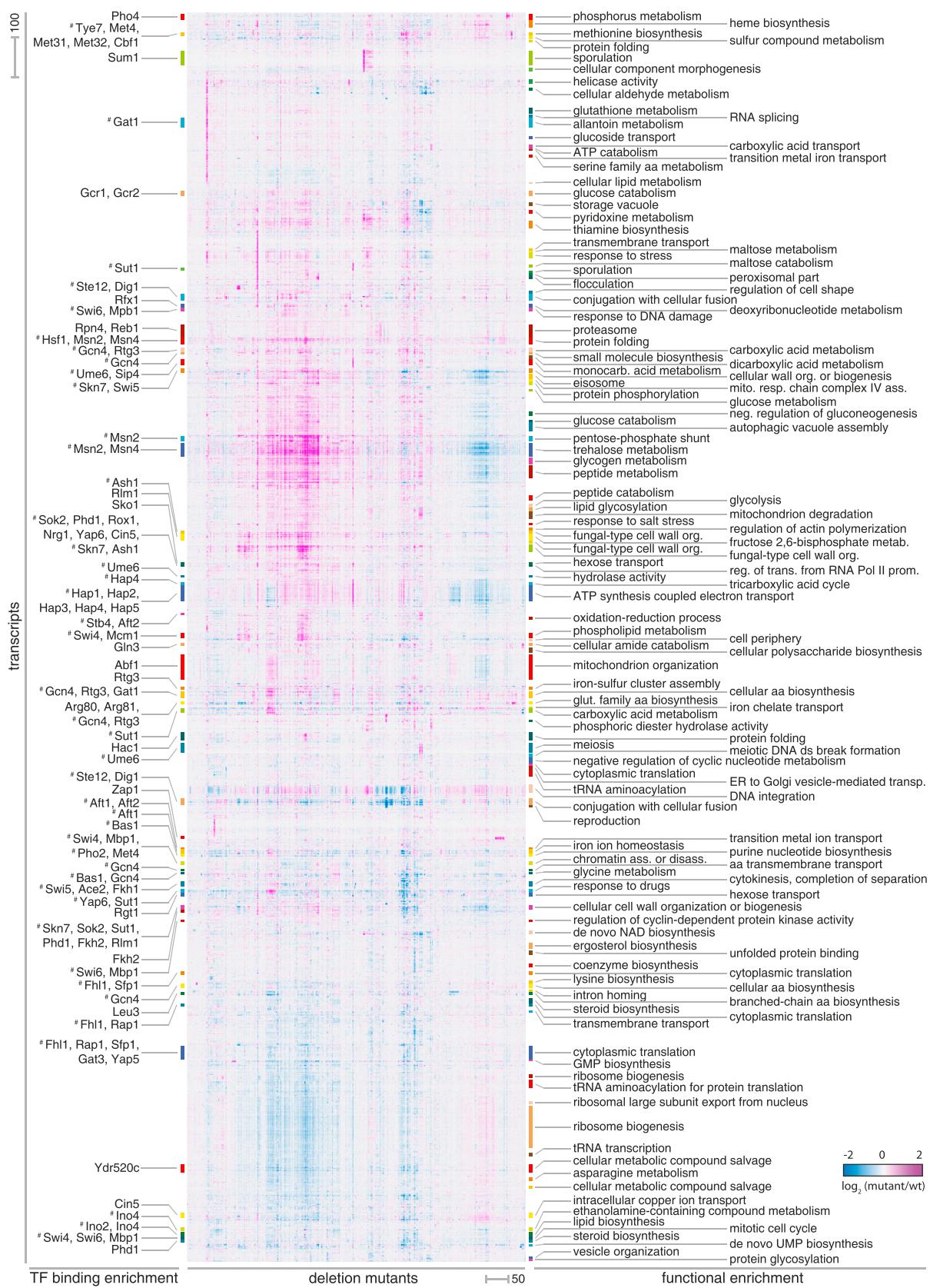
The data set consists of approximately 40 million expression measurements including WTs and replicates. Hierarchical clustering is presented in Figure 1. Although low-magnitude fold-changes [FCs] may have biological relevance, a stringent threshold ($FC > 1.7$, $p < 0.05$) was applied throughout the study to ensure a focus on robust changes more likely to be biologically meaningful. This threshold was based on WT variation. When analyzed collectively, the number of transcripts robustly affected in at least one mutant ($FC > 1.7$, $p < 0.05$) starts leveling off at two-thirds (Figure 2A). Transcripts that do not change are highly enriched for dubious open reading frames (ORFs; $p = 2.6 \times$

10^{-9}) and for genes essential for viability ($p = 7.8 \times 10^{-31}$). Most dubious ORFs are lowly or not expressed in WT (Figure 2B). Combined with their low degree of change, this agrees with their classification as dubious, with most not likely to encode functional proteins (Fisk et al., 2006). Essential genes show much higher WT transcript levels (Figure 2B). The low degree of change observed for essential genes (Figure 2B) indicates that larger changes in their expression are too deleterious for survival. Plateauing of transcripts with altered expression (Figure 2A) suggests that most of the robust gene expression changes compatible with viable genetic perturbation have been covered for this growth condition.

As observed before, strains with reduced growth generally have more transcripts affected, and not all genetic perturbations result in transcriptome changes (Hughes et al., 2000). To focus on mutants with stronger changes, signatures were classified as different from WT (responsive) when at least four transcripts show robust changes. Excluded are a set of 58 transcripts with highly variable behavior in WTs (WT variable genes; Experimental Procedures). These criteria ensure that almost all WTs are classified as having no change and indicate that 53% of mutants are similar to WT (nonresponsive). This is concordant with the previous determination of 43% on a smaller set of deletions using different thresholds (Hughes et al., 2000). Redundancy likely contributes to nonresponding deletions. This is demonstrated by a strong enrichment for genes with a close paralog (Figure S1E). Growth condition-dependency likely also contributes. This is indicated by the larger number of genes with low transcript and undetectable protein levels within the group of nonresponder deletions (Figures S1C and S1D). The information that loss of a gene does not strongly affect expression of other genes is useful for several purposes, including modeling regulatory networks (Macneil and Walhout, 2011). Taking essential genes into account (Giaever et al., 2002), the fraction of genes that can be individually removed under a single growth condition with no strong effects on gene expression is 43%.

Protein Complex and Pathway Organization

Functional relationships are revealed by hierarchical clustering of deletion signatures (Figure 1, columns; dendrogram in Data S1). Previous analyses indicate protein complex and pathway membership as the main factors contributing to profile similarity (Hughes et al., 2000; Lenstra et al., 2011). In contrast to coexpression across different conditions, the degree of deletion-profile similarity for different types of interactions has so far not been systematically addressed. We therefore determined signature similarity for all complexes and pathways, including metabolic pathways as well as signaling factors such as protein kinases, ubiquitin(-like) enzymes and their targets. Signature correlation is highest for protein complexes (Figure 2C), in particular for smaller complexes with four or less subunits (examples in Figure 3A). All transcripts that change significantly in any single mutant are depicted in such figures, rather than a subset selected for similar behavior. Highly similar profiles (Figure 3A) indicate disruption of the entire complex upon deletion of any individual subunit. As shown previously for the transcription coregulator Mediator (van de Peppel et al., 2005) and more comprehensively for 30 chromatin complexes (Lenstra et al.,



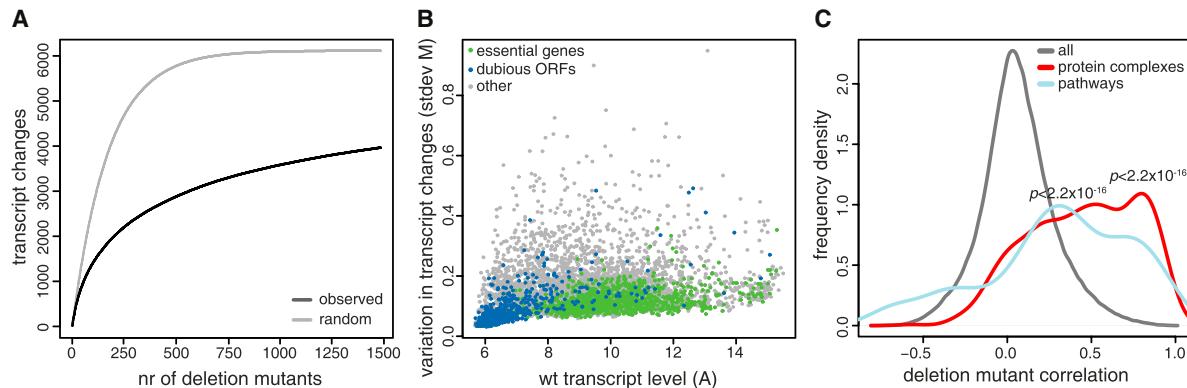


Figure 2. General Properties of Genetic Perturbation

(A) Cumulative plot of transcripts with changed expression as a function of deletion mutants added. The average of 1,000 random orderings of deletion mutants is shown (dark gray) versus the average of 1,000 orderings with the same number of transcripts as in the original profiles but now randomly selected from the entire genome (light gray). The 95% confidence intervals are too close to the average to be visible. The cumulative total of transcripts with changed expression is 3,966 ($FC > 1.7$, $p < 0.05$) and only changes slightly (3,962) when multiple testing correction (Benjamini-Hochberg) is additionally performed across all mutants instead of only for mutants individually.

(B) Variation in transcript level changes in the form of standard deviation of M , the $\log_2(\text{mutant}/\text{WT})$, across all mutants, plotted as a function of A , the \log_2 expression level (fluorescent intensity) from 200 WTs. Essential genes are green, dubious ORFs are blue, and all other transcripts are gray.

(C) Frequency density distribution of correlations between expression profiles of all responsive mutants (dark gray), protein complexes (red), and pathways (blue). Figures S1C–S1F present other general differences between responsive and nonresponsive deletions as well as a direct comparison between all pairwise deletion profile correlations and synthetic genetic interaction (SGI) profile correlations.

2011), larger complexes often reflect a submodular architecture: high similarity within submodules and lower similarity between submodules (examples in Figure 3B, including Mediator with revised data). Besides submodularity, other interesting cases that reduce correlation are subunits shared between different complexes, auxiliary subunits with different function and peripheral subunits with no apparent function under the growth condition analyzed (Lenstra et al., 2011). At least 195 complexes are present in this data set (Experimental Procedures). Their profiles are useful for understanding function, identifying reporter genes, and discriminating between the activities of different submodules.

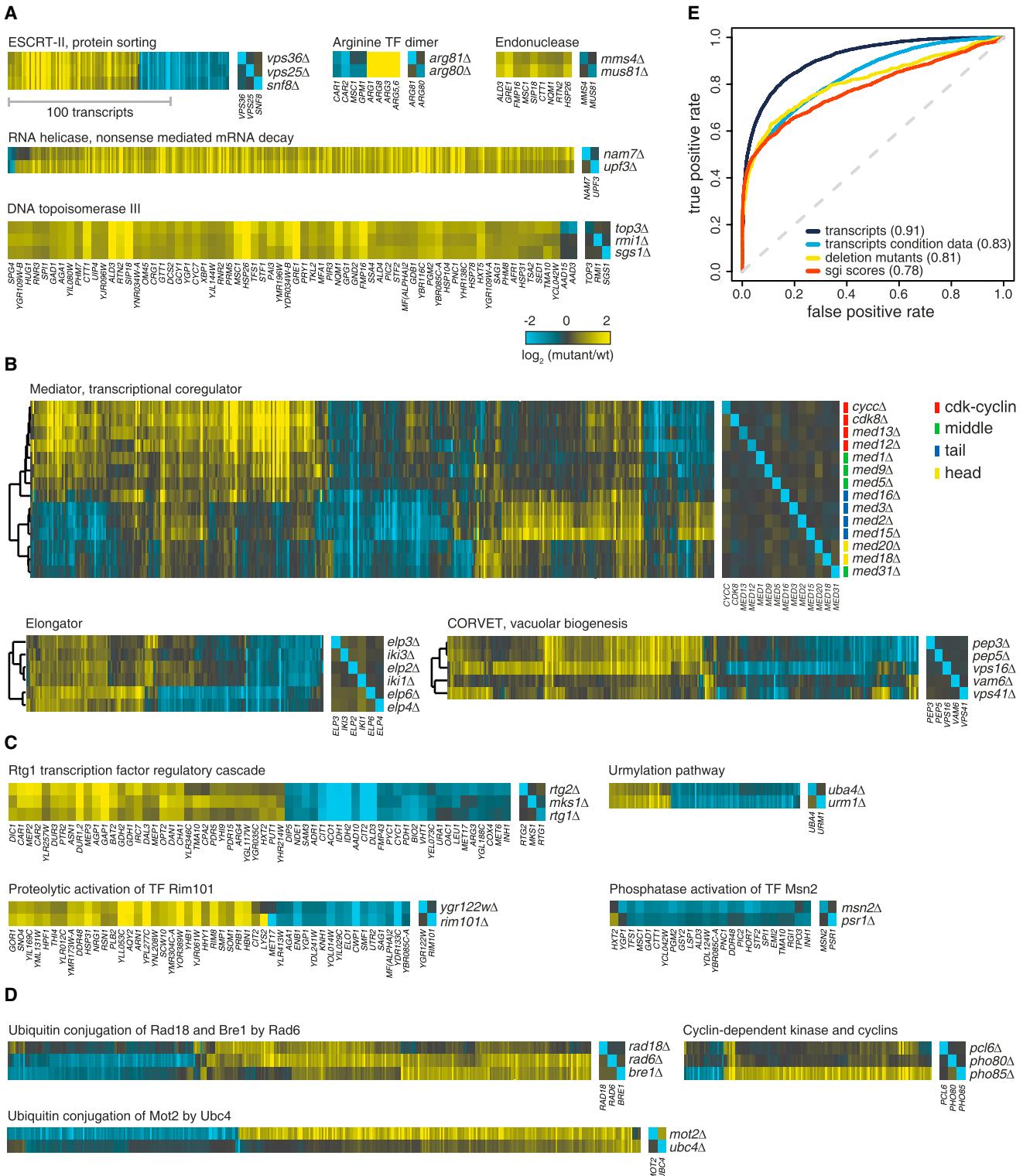
Pathway signature correlation is also significantly high, but lower than for protein complexes (Figure 2C). Pathways that are straight or cyclic, with no branching into distinct active arms, are expected to result in high correlations, similar to small protein complexes. Such apparently unbranched pathways are found (Figure 3C) and include three MAP kinase cascades and one chromatin interaction pathway observed by expression-profiling before (Lenstra et al., 2011; Roberts et al., 2000; van Wageningen et al., 2010). For the majority of pathways, correlation is still high (Figure 2C) but incomplete due to only partially overlapping signatures. Inspection of established pathways indicates that this is caused by branching at nodes to exert different downstream effects (examples in Figure 3D). Most pathways show profiles with partially overlapping signatures, resulting in

reduced correlations (Figure 2C), thereby indicating that most cellular pathways branch. This is a requirement for biological systems to have feedback and to interconnect subsystems. As with protein complexes, partial as well as completely overlapping signatures are, therefore, both useful for exploring pathway relationships.

Clustering the data on transcripts (Figure 1, rows) results in enrichment for very specific processes (Figure 1, right) and for TF binding sites (Figure 1, left), with many cases of combinatorial control indicated. Coexpression across condition-dependent gene expression data can be used to predict similar function (Eisen et al., 1998). We compared the power of different types of correlations to predict similar function. Deletion profile correlation performs slightly better than genetic interaction correlation (Figure 3E). The latter (Costanzo et al., 2010) has higher coverage, and although similar performance is achieved, the data sets are complementary rather than solely overlapping (Figure S1F). Transcript correlation across our deletion data set performs best, also compared to transcript correlation across different conditions (Figure 3E). This is unanticipated given the focus on a single growth condition but likely reflects other properties such as scope of mutants, data quality, and uniformity. The genetic perturbation data are therefore a useful resource for exploring gene function, both by profile correlation and by transcript correlation. A Web-based tool to do both is made available (<http://deleteome.holstegelab.nl/>).

Figure 1. Expression Signatures of Mutants

Hierarchical clustering of all responsive mutants (left-right) and all transcripts (top-bottom) changing in at least two mutants. Enrichment for functional categories in clusters is indicated on the right. GSTF binding site (MacIsaac et al., 2006) enrichment is indicated on the left. #GSTFs enriched in more than one cluster, often in combination with other GSTFs, indicating combinatorial control. The classes of deletion mutants and a schema of the study design are presented in Figures S1A and S1B.

**Figure 3. Complex/Pathway Organization and Function Prediction**

(A) Examples of protein complex subunits with high deletion signature correlation, indicating disruption of the function of the entire complex in each individual deletion. Yellow indicates increase, blue indicates decrease, and black no change in expression versus WT.

(B) Examples of submodularity of large complexes. The coregulator Mediator is a previously well-studied example. The new profiles, some from this study and some from Lenstra et al. (2011), correspond to different (color-coded) submodules previously identified by various interaction assays and by expression profiling

An Abundance of Gene-Specific Repressors

Besides the advantage of uniformity, we also explored the potential advantage of having revised data for previously analyzed mutants. The collection encompasses many different types of regulators that can be analyzed individually or group-wise. As an example, the yeast genome encodes an estimated 182 proteins likely to bind specific DNA sequences (Experimental Procedures), indicative of a role as gene-specific transcription factor (GSTF). GSTFs have been analyzed by expression profiling before (Hu et al., 2007). Other large-scale studies have analyzed the genomic location (Harbison et al., 2004; MacIsaac et al., 2006) or DNA binding specificity of GSTFs (Badis et al., 2008; Zhu et al., 2009). TF binding is not necessarily predictive of function (Spitz and Furlong, 2012; Hughes and de Boer, 2013), and accurate modeling of regulatory networks would benefit from knowing functional TF targets. Previous comparison between genomic location and GSTF deletion profiles yielded a generally low correspondence (Hu et al., 2007). A total of 171 GSTF deletions were successfully profiled here, of which 72 are responsive. Including essential GSTFs, this entails that 54% of GSTFs can be individually removed without strongly affecting gene expression under this growth condition. This is also a reason why a subset has previously been analyzed by overexpression (Chua et al., 2006). Interestingly, the fraction of nonresponders is substantially lower for chromatin factors (25%) and much higher for protein kinases/phosphatases (75%), indicating quite different degrees of condition dependency and redundancy for different classes of regulators.

The GSTF signatures are also more specific compared to chromatin factors. A median of 19 transcripts change in responsive GSTFs compared to 68 for chromatin factors, underscoring their more global role (Figure 4A). Examples of well-studied GSTFs are depicted in Figure 4B. An important issue illustrated here is the distinction between direct and indirect effects. Many profiles show both decreased and increased expression. Comparison with genome-wide binding data (MacIsaac et al., 2006) shows significant enrichment with only one side of the expression response (Figure 4B, black rectangles), in each case confirming the established function (activator or repressor) and agreeing with previously established cellular roles.

Systematic comparison was therefore performed for all GSTFs, also making use of in vitro binding affinities (Badis et al., 2008; Zhu et al., 2009) to increase the number of GSTFs covered. An important outcome is the clear-cut classification of GSTFs into either activators or repressors (Figures S2A–S2C; summarized in Figure 4C). Of the responsive GSTFs for

which systematic DNA binding data are available (88%), significant overlap with the deletion signature is found for 70%. The overlap is one sided in nearly all cases, mapping either to genes with decreased expression (activators, 55%; Figure S2A) or genes with increased expression (repressors, 39%; Figure S2B). Only a few cases of dual activator/repressor function are indicated (Figure S2C). This includes Cbf1, for which such dual function has previously been established (MacIsaac et al., 2012). Importantly, the three different DNA binding data sets support or complement each other with regard to this classification and never conflict (Figure 4C).

The proportion of gene-specific repressors in eukaryotes has not been reported before, and this analysis indicates an unanticipated high abundance. The classification corresponds very well to what was previously known for individual factors (Table S2), taking into account that GSTFs are sometimes called activator or repressor with incomplete evidence for either. For example, the well-studied cell-cycle TF Mbp1 is still frequently referred to as an activator, despite the putative activating and repressive roles originally reported (Koch et al., 1993) and despite clear later evidence for a direct role as repressor (de Bruin et al., 2006; also confirmed here; Figure 4D). Such ambiguities are also relevant for previously poorly characterized GSTFs (Table S2), including Stp3, Stb4, and Rph1, classified here as gene-specific repressors (Figure 4D).

The classification represents a uniformly conducted survey for the activity of GSTFs under a single growth condition. Many GSTFs are condition specific (Hughes and de Boer, 2013). A direct comparison between the growth medium studied here (synthetic complete [SC]) and another commonly used medium (YPD) indicates that repression is not particular to SC. Only 128 genes are differentially expressed between the two conditions, and the differences are balanced (Figure S3A). This indicates that there is as much repression taking place in YPD as in SC. More extremely different growth conditions such as nutrient depletions are accompanied by large general reductions in expression (Radonjic et al., 2005), indicating that the proportion of repressors may be even higher if the survey was carried out under such conditions.

As discussed below, the high abundance of gene-specific repressors (45% including dual-function GSTFs) is especially surprising in light of models of transcription from chromatinized DNA. Many differences in the study design likely contribute to differences with the earlier study (Hu et al., 2007), including measures taken here that result in a lower degree of measurement error or noise (Figures S4A–S4I). The availability of in vitro

before (van de Peppel et al., 2005). The Elongator complex (left) consists of two submodules. Pep3-Vps41 (right) represents the CORVET tethering complex required for protein sorting and vacuolar biogenesis.

(C) Apparently unbranched pathways. The cellular location of the GSTF Rtg1 is regulated by Rtg2, which is in turn regulated by Mks1. Uba4 activates Urm1 prior to urmylation, a ubiquitin-like modification. Ygr122w is proposed to be required for proteolytic activation of the repressor Rim101. Psr1 is a protein phosphatase that activates the gene-specific transcription factor Msn2.

(D) Examples of branched pathways. Rad6 is the ubiquitin-conjugating enzyme (E2) for the ubiquitin-protein ligases Rad18 and Bre1 (E3s). Their profiles are subsets of RAD6 deletion. Pcl6 and Pho80 are cyclins for the Pho85 cyclin-dependent kinase. Mot2 is an E3, activated by Ubc4, which is largely redundant with Ubc5, classified here as nonresponsive.

(E) Receiver operator characteristic (ROC) curves comparing gene function prediction using correlations of genetic interaction profiles (SGI, red), deletion signatures (yellow), coexpression across different conditions (Kemmeren et al., 2002) (blue), and coexpression only across the genetic perturbation data from this study (black). Numbers are area under the curve (AUC). The ROC curve plots the true-positive rate as a function of the false-positive rate. An AUC of 1 would indicate perfect predictions.

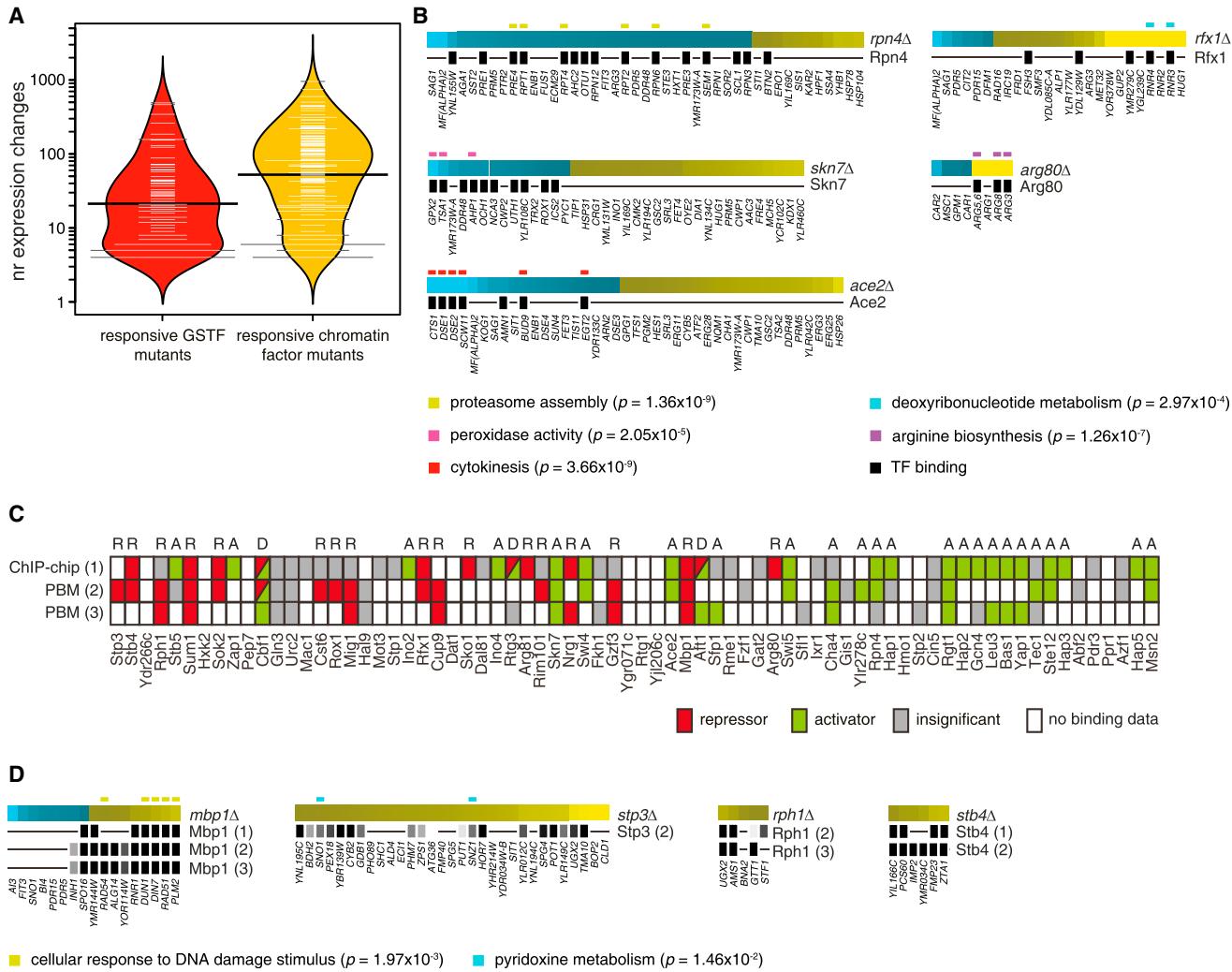


Figure 4. Classification of Gene-Specific Transcription Factors

- (A) Bean plots showing the expression changes in responsive GSTF (red) and chromatin factor mutants (yellow).
- (B) Examples of established GSTFs. Binding [MacIsaac et al., 2006] in the promoters of changed genes is indicated by black rectangles. GO enrichment for genes with expression change and GSTF binding is indicated above each expression profile, with categories listed at the bottom.
- (C) Classification of GSTFs into activator “A,” repressor “R,” or dual function “D” as based on significant enrichment for in vivo binding [(1) (MacIsaac et al., 2006)] or promoter affinity score [(2) (Badis et al., 2008), (3) (Zhu et al., 2009)] in the expression profile.
- (D) Gene-specific repressors, as in (B). All GSTFs are individually depicted in Figures S2 and S3B. Comparison to previously published deletion profiles are in Figures S4A–S4I, and Figure S3A compares the transcriptome in the growth medium used here with another commonly used growth medium.

binding data has also contributed to the classification (Figure 4C). The data set still includes responsive GSTF deletions not covered in any large-scale binding data sets (Figure S3B). This all supports the proposal to revise previously generated large-scale data sets (Hughes and de Boer, 2013). The group-wise analysis of GSTFs employs only a fraction of the entire data set (12%) and also highlights ways in which fundamental aspects of the regulatory system can be discovered from the causal relationships inherent to perturbation data.

The Genetic Perturbation Network

Uniformity and causality can also be employed to study the regulatory system in its entirety. The data can be rendered as a

gene network with directional edges signifying increased or decreased expression in a downstream gene Y due to deletion of an upstream gene X. Including only robust changes (Figure 1) results in a network with 3,476 gene nodes and 50,294 edges (Data S2). The genetic perturbation network (GPN) exhibits a power-law distribution (Barabási, 2009) (Figures S5A and S5B) and the connectivity patterns differ for the various classes of regulators (Figure S5C). Whereas chromatin factors affect many different categories, they themselves do not frequently change in expression, despite the vast majority not being essential for viability. GSTFs have an opposite behavior, with more frequent incoming connectivity. This indicates that in yeast, such regulators frequently serve as downstream drivers of smaller gene

expression programs. Signal transduction cofactors show both facets, with high in- and outdegrees across many different categories. This implies that such genes form important hubs, central for cellular information flow. Central connectivity is stronger for signal transduction cofactors than for protein kinases and phosphatases. This supports the emerging view that cofactors such as scaffold proteins have central roles in cellular regulation (Good et al., 2011).

Besides directionality, another interesting characteristic that can be included is overlap between signatures. When observed in combination with downregulation of one gene (Y) upon deletion of the other (X), overlapping signatures potentially explain (part of) the deletion profile of the upstream gene (X). In the most extreme case the signature of the downstream gene (Y) is completely nested within the signature of the upstream gene (X) (examples in Figure S5D). Nested effects can indicate indirect effects. Although there are many nested effects for individual genes, only a few signatures are nested in their entirety. This fits with the low number of straight, unbranched pathways and is a further reflection of the interconnected nature of biological systems.

Feed-Forward Loop Recurrence and Differential Participation of Regulators

Besides nested effects, more complex motifs (Alon, 2007; Macneil and Walhout, 2011) can also be identified. These include feed-forward loops (FFLs) for which eight types can be envisaged (Mangan and Alon, 2003) (Figure 5A). FFLs can form subcircuits with interesting functionalities such as the persistence and delay circuits associated with coherent type 1 FFLs (C1-FFLs) (Alon, 2007). Previous analyses in yeast have been restricted to GSTF binding data, with relatively low numbers of FFLs available for systematic investigation (49 [Lee et al., 2002] and 56 [Mangan and Alon, 2003]), as well as lack of expression data to study function. FFLs were therefore determined from the GPN. This was based on an operational definition, as observed from the gene expression changes and not requiring edges to be direct. An I2-FFL, for example (Figure 5A), is identified by overlap in genes upregulated upon deletion of two genes (X and Y), whereby Y is also upregulated upon deletion of X (incoherent). To focus on the most consistent FFLs, only X-Y FFL pairs with a significant overlap in downstream Z genes were considered. In the ensuing analyses, unique X-Y FFL pairs were only counted once rather than multiple times for each shared downstream gene Z. This results in 1,120 X-Y FFL pairs (Figure 5B), a vast increase in the number of FFLs available for further analysis.

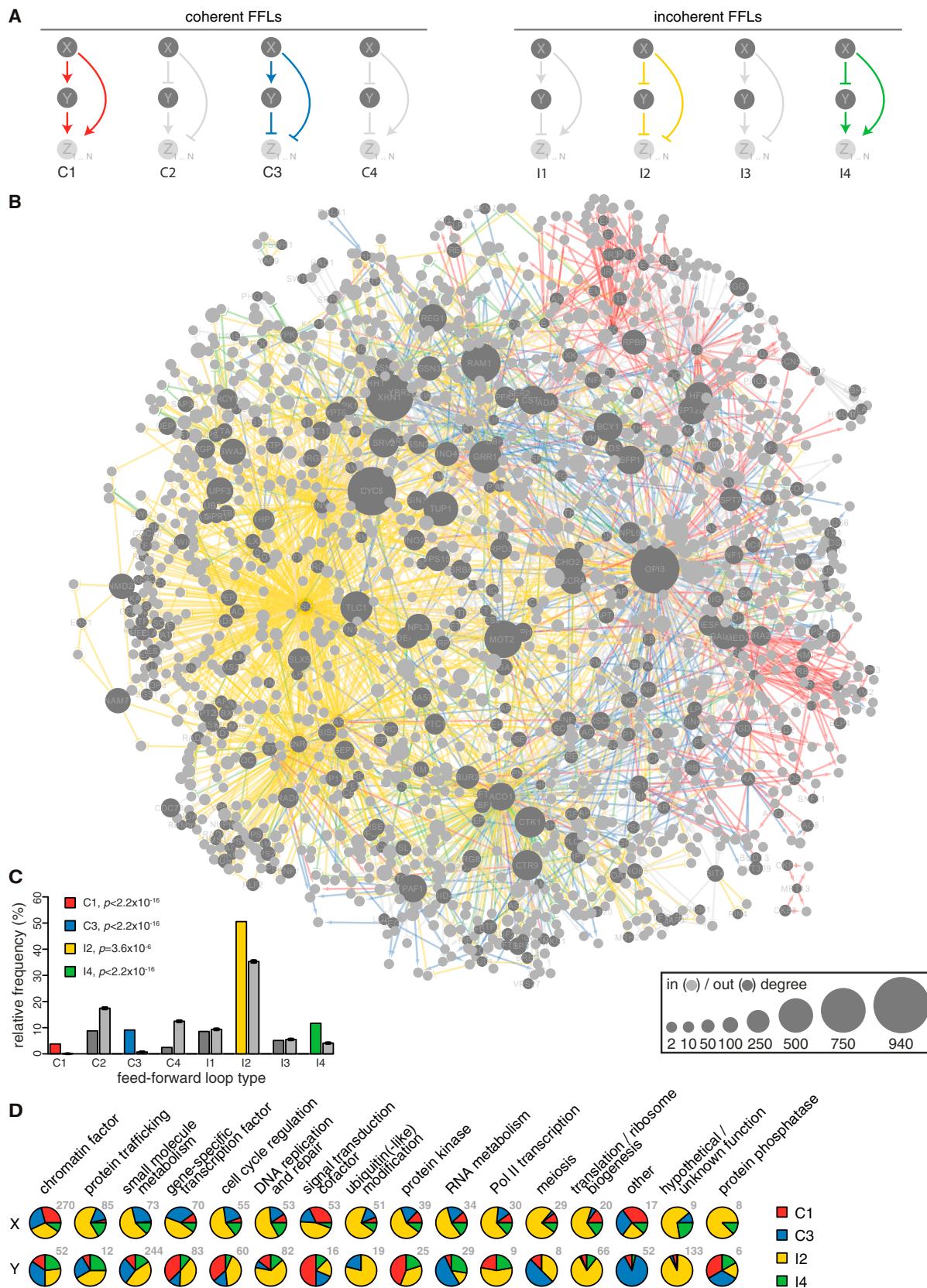
Recurrence of network parts may indicate advantageous regulatory properties and is also interesting from an evolutionary perspective (Sharan and Ideker, 2006). The occurrence of FFL types was therefore compared to 10,000 edge-permuted networks, each conservatively maintaining the network properties of the original (Experimental Procedures). Four of the eight FFL types are overrepresented (C1, C3, I2, and I4; Figure S5E), also after exclusion of nested effects (Figure 5C) or after applying different thresholds for significant change in expression ($FC > 1.7$, $FC > 1.2$, no FC, with $p < 0.001$, $p < 0.01$, $p < 0.05$). Recurrence of C1-FFLs agrees with their enrichment in a sparser DNA binding network (Mangan and Alon, 2003). As with C1-FFLs

(Alon, 2007), regulatory properties of the other overrepresented FFL types will require detailed investigation of individual FFLs, including analysis of kinetics and the combinatorial input function at the downstream gene (Z), not represented in the network.

Incoherent Type 2 FFLs Indicate Feedback in the Perturbation Network

Some interesting general characteristics can be discerned by global analysis of the overrepresented FFL types. For example, participation of different classes of regulators as up- or downstream FFL component is differential (Figure 5D). Chromatin regulators are frequent upstream components and are frequently found within coherent FFLs. This may indicate collaboration with other downstream factors to mutually reinforce gene expression programs. Frequent participation of small-molecule metabolic pathway components as downstream Y nodes, especially in I2-FFLs, is also striking. A number of characteristics indicate that such FFLs may often represent metabolic feedback. First, genes from metabolic pathways feature frequently in I2-FFLs, in particular as downstream Y node (Figure 5D). Second, the same downstream Y node often participates in multiple X-Y I2-FFL pairs. Examples include *RNR4*, required for de novo deoxyribonucleotide biosynthesis (39x); *ACO1*, required for the tricarboxylic acid (TCA) cycle (34x); and *OPI3*, required for phosphatidylcholine biosynthesis (10x). This suggests that such components are involved in downstream events common to many different perturbations. Third, Gene Ontology (GO) enrichment within the I2-FFL downstream Z nodes includes processes such as oxidation-reduction and protein folding/unfolding, also commonly found in different growth condition perturbations (Gasch et al., 2000).

Examination of individual I2-FFLs also indicates metabolic feedback. Two such cases, whereby the GSTF involved was also discerned from the expression data, are shown in Figure 6. The first example consists of the I2-FFL nodes *PFK27* (X) and *ACO1* (Y). *PFK27* and *ACO1* form an incoherent type 2 FFL because the genes upregulated upon their deletion overlap significantly, even though *ACO1* shows increased expression in *pfk27Δ* (Figure 6A). *PFK27* and *ACO1* are functionally connected through the TCA cycle (Broach, 2012), (Figure 6B). In this model, loss of either *ACO1* or *PFK27* reduces output, and due to feedback, other pathway components become upregulated (Figure 6B). This is in part mediated by the heterodimer GSTF Rtg1/3 and its regulator Rtg2 that senses glutamate/glutamine levels (Liu and Butow, 2006). This involvement is confirmed by the effect of their deletion on pathway components (Figure 6B). A similar case that suggests metabolic feedback in phospholipid synthesis is observed for the I2-FFL nodes *CHO2* (X) and *OPI3* (Y), mediated by the GSTF heterodimer Ino2/4 (Figures 6C and 6D). As indicated in these models, such I2-FFLs can better be represented as feedback circuits. Their identification from the GPN indicates a general method for discovering such feedback circuitry. Both the number and type of FFLs made available here is vastly increased compared to what was previously available. The overrepresentation of four FFL types as well as the participation of different regulator classes in distinct FFLs is striking. This further demonstrates



(legend on next page)

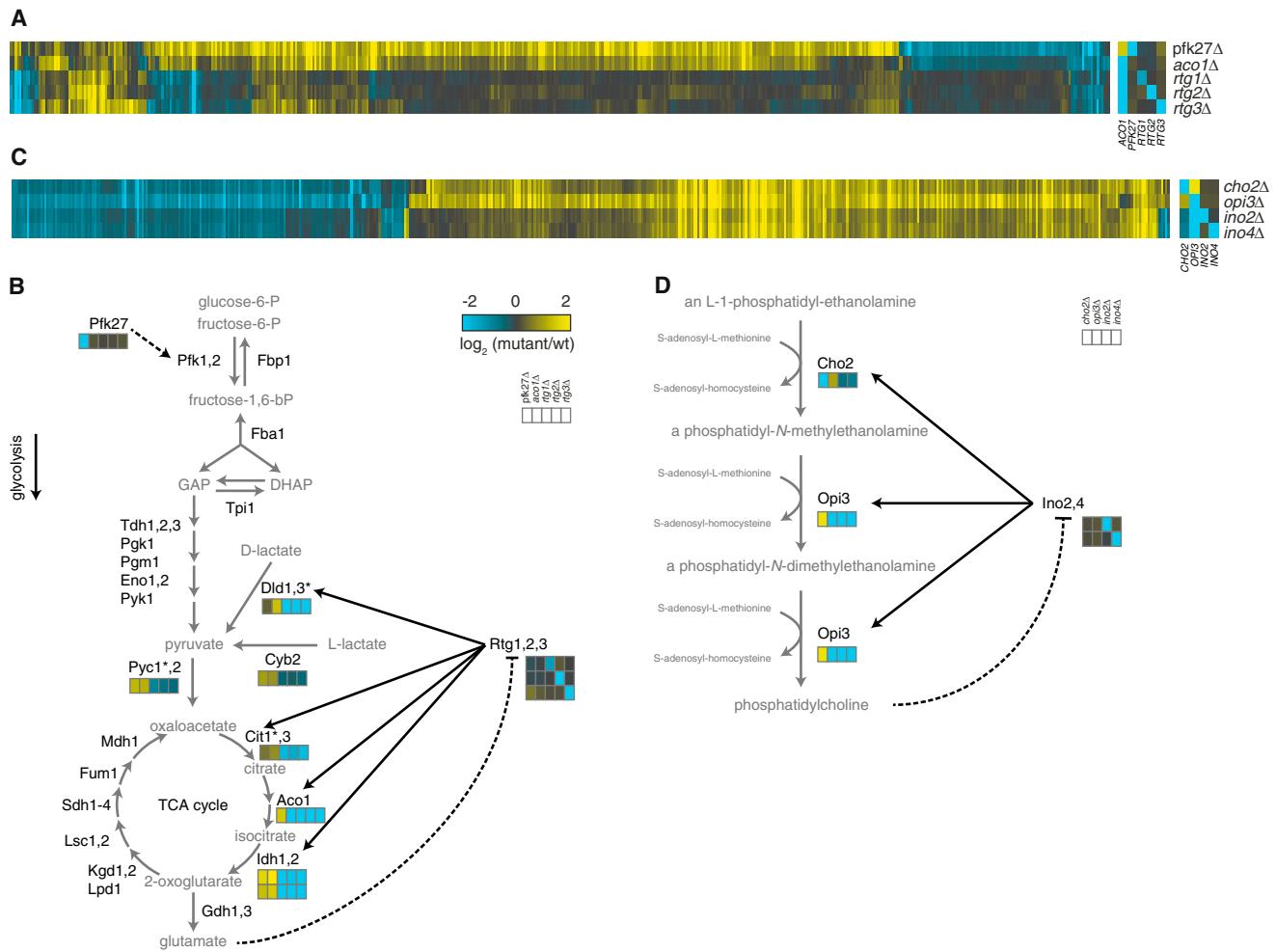


Figure 6. I2-FFLs Indicating Metabolic Feedback

(A) Expression profiles of I2-FFL nodes *PFK27* (X) and *ACO1* (Y) along with GSTF *RTG1/3* (heterodimer) and its regulator *RTG2*.

(B) The glycolysis and TCA cycle with the observed transcript changes for *pfk27Δ*, *aco1Δ*, *rtg1Δ*, *rtg2Δ*, and *rtg3Δ*.

(C) Expression profiles of I2-FFL nodes *CHO2* (X) and *OPI3* (Y) along with the GSTF heterodimer *INO2/4*.

(D) The phosphatidylcholine biosynthesis pathway with the observed transcript changes for *cho2Δ*, *opi3Δ*, *ino2Δ*, and *ino4Δ*.

the utility of the data set for exploring regulatory systems, either globally or parts-wise.

DISCUSSION

The scale and uniformity of the underlying data set allow different general properties of mRNA expression, its regulatory system,

and the response to genetic perturbation to be systematically analyzed. Besides general characterization of the GPN itself, a recurrent theme is the branched and interconnected nature of regulatory networks. This agrees with a large-scale protein interaction study focused on kinases (Breitkreutz et al., 2010) and the propensity of genetic interactions in yeast (Costanzo et al., 2010). It is evident here from the scarcity of straight or cyclic

Figure 5. Feed-Forward Loop Identification in the Genetic Perturbation Network

(A) Different FFL motifs (Mangan and Alon, 2003).

(B) Collapsed GPN with edges colored as in (A). Dark gray nodes represent deletion mutants. Z transcripts are collapsed into a single node (light gray) for each X-Y FFL pair. Node size according to outdegree (dark gray nodes) or number of Z transcripts (light gray nodes). The network layout was generated in Cytoscape (Smoot et al., 2011) using the spring-embedded algorithm.

(C) Relative frequency of FFL types after removing nested effects. Error bars for the 10,000 permuted networks (gray) indicate two times the SEM, the 95% confidence interval.

(D) Participation of different classes of genes as up- (X) or downstream (Y) components in the overrepresented FFL types. Numbers indicate the X or Y nodes per class. Figures S5A and S5B depicts the in- and outdegree frequency distribution for the GPN. Figure S5C compares in- and outdegrees for different classes of genes. Figure S5E shows the frequency of FFL types without filtering for nested effects, and Figure S5D depicts examples of completely nested effects.

pathways, the high number of FFLs, and the low number of entirely nested effects. Other characteristics identified here include the types of genes for which changed transcript levels are compatible with viability, the nature of nonresponsive deletions, and the types of FFLs overrepresented in the GPN. It is also important that the intricacies of pathway and protein complex architectures are reflected in the perturbation signatures, because one aim of this study is to provide a resource for studying gene expression at different levels of complexity.

Focused analysis of one class of regulators is described (GSTFs). More classes are present, and analyzing these individually or in combination (e.g., GSTFs and chromatin factors) (Steinfeld et al., 2007) will be interesting. The abundance of gene-specific repressors is noteworthy. First, such a general finding strengthens the proposal that revision of some early large-scale data sets is worth pursuing (Hughes and de Boer, 2013). Second, systematic classification of GSTF function has not been carried out before. The presence of gene-specific repressors in eukaryotes is known. The extent of their occurrence is not. It is often assumed that GSTFs in eukaryotes are predominantly activators (Fuda et al., 2009; Struhl, 1999). This is in part based on the idea that the chromatinized DNA found in eukaryotes is generally repressive, with activators required for transcription to take place (Hahn and Young, 2011). The relatively high abundance of GSTFs that repress transcription (45% including dual function GSTFs) is therefore surprising. This first systematic classification will likely benefit from improved genomic location data as well as from studies aimed at determining whether differential condition-dependency between activators and repressors alters the survey. The analyses nevertheless indicate that gene-specific repressors are more prevalently active than has previously been assumed. This fits with pervasive transcription throughout the coding and noncoding genome of eukaryotes (David et al., 2006; Jacquier, 2009), also indicating that chromatin is not necessarily generally restrictive to transcription. These observations support the idea that transcription is not always tightly regulated intrinsically (Spitz and Furlong, 2012), frequently requiring additional factors to prevent undesired expression. Similar to some cases of paused RNA polymerase II (Adelman and Lis, 2012), active repression through GSTFs may allow for coordinated and potentially fast upregulation of specific groups of genes upon repressor inactivation.

A high occurrence of gene-specific repressors also raises questions about the dogma that gene-specific activators are required to drive transcription of all genes in eukaryotes. Other aspects of this data set fuel this speculation too. This includes the much lower responsiveness to GSTF deletion compared to chromatin factor deletion, despite their similar abundance, and the general sparsity of GSTF binding site enrichment for the expression changes observed (Figure 1, left). Other likely explanations for these characteristics can also be put forward, and such speculation also requires multiple GSTF redundancy to be considered. The abundance of gene-specific repressors nonetheless raises the possibility that activators may not always be required and that aspects such as the demand rule explanation for prokaryotic activator/repressor promoter configurations (Savageau, 1977) may also hold for eukaryotes.

Compared to purely descriptive data sets, the causality inherent to perturbation adds an additional facet for interpretation, with sometimes surprising outcomes. This is also exemplified by recent demonstrations that a chromatin mark correlating with active transcription actually results in derepression upon its loss (Margaritis et al., 2012; Weiner et al., 2012). The latter study also indicates that temporal and conditional aspects will further improve such approaches, as will large-scale inclusion of other perturbations. Many future analyses can be facilitated by the availability of uniformly collected perturbation data sets, also for other organisms (Bonke et al., 2013). Besides combinatorial analyses with other types of data, this includes the ability to refer to the effects of individual gene mutation in the context of a large-scale reference data set, an aspect made possible, but not highlighted, here.

EXPERIMENTAL PROCEDURES

Full details are provided in [Extended Experimental Procedures](#).

Data Availability

Expression levels (A), ratios (M), and p values are also available as tab-delimited files from <http://deleteome.holstegelab.nl/>. For comparison with other data sets, additional profiles such as YPD versus SC medium, mating type comparison, and diploid versus haploid are included. The data can also be viewed after installing Java TreeView (<http://jtreeview.sourceforge.net/>), by downloading [Data S1](#), extracting the zip file, and opening the .cdt file to yield Figure 1. The GPN can be obtained by downloading [Data S2](#), installing Cytoscape (Smoot et al., 2011) (<http://www.cytoscape.org>), extracting the zip file, and opening the Cytoscape session files. A Web-based tool for exploring mutant and transcript profiles is available at <http://deleteome.holstegelab.nl>.

Expression Profiling

Each mutant strain (Table S1) was grown twice, from two independently inoculated cultures. Cultures were harvested early during exponential growth in SC medium with 2% glucose. Each culture was expression-profiled in technical replicate to yield four measurements for each profiling mutant. To monitor reproducibility, a common reference design was adopted with a batch of WT RNA applied in dye-swap to one of the channels of each microarray (Figure S1B). Additional WT cultures were grown alongside batches of mutants on each day and profiled in parallel to monitor batch effects and to generate the set of WT transcriptomes for comparison to mutants.

Data Analyses

All correlations are with a standard correlation distance. Hierarchical clustering was by average linkage. During distance calculation for the hierarchical clustering (Figure 1), M values of transcripts with insignificant changes ($p > 0.05$, $FC < 1.7$) were set to zero. Functional enrichment (Figures 1 and 4B) was through hypergeometric testing ($p < 0.01$, Bonferroni corrected). Protein complexes for pairwise deletion mutant correlations (Figure 2C) and for similar function prediction (Figure 3E) were from the “curated consensus + GO” data set describing 518 complexes (Benschop et al., 2010). Pathway definitions (Figure 2C) were derived by merging SGD biochemical pathways with GO pathways (Cherry et al., 2012). Pairwise correlations were calculated for all pairs within a protein complex or pathway if at least two members of the protein complex or pathway are present as a deletion mutant. For the function similarity prediction (Figure 3E), true positives and true negatives were calculated as in Collins et al. (2007).

Classification of GSTFs

GSTFs (Table S1) were compiled from previous studies (Badis et al., 2008; Harbison et al., 2004; Zhu et al., 2009) and augmented with other genes with a domain capable of sequence-specific DNA binding. Classification (Figures S2 and S3; Figure 4C) was by comparison of the expression profiles with

systematic DNA binding sets. Significant overlap between promoter binding and the genes with increased or decreased expression was tested using binarized *in vivo* chromatin immunoprecipitation (ChIP)-chip data (MacIsaac et al., 2006) ($p = 0.005$, no conservation restriction, set 1) and two in vitro-derived promoter affinity scores: sets 2 (Badis et al., 2008) and 3 (Zhu et al., 2009). For ChIP-chip data, significance was determined using a Fisher's exact test. For the promoter affinity scores, a Mann-Whitney test was applied. A GSTF was classified as activator if (1) a significant overlap with its binding targets or (2) significantly higher GSTF promoter affinity scores were observed for genes with decreased expression in the deletion ($p < 0.05$, Benjamini-Hochberg corrected). A GSTF was classified as repressor if one of these two criteria is fulfilled for genes that show increased expression. A GSTF was classified with dual function if either criterion applied to genes that show both decreased and increased expression.

Detection of Feed-Forward Loops

Detection of FFL motifs occurred in four steps. First, individual X-Y-Z subgraphs were extracted by looping through every single node (X) in the GPN and retrieving its successor nodes (Y) and corresponding X-Y shared successor nodes (Z). Second, X-Y-Z subgraphs were classified as a C1, C2, C3, C4, I1, I2, I3, or I4 FFLs depending on the inferred sign of the individual X-Y, X-Z, and Y-Z edges. Third, individual X-Y-Z subgraphs were grouped per unique X-Y pair. Fourth, for each X-Y FFL pair, a hypergeometric test was performed to judge whether the shared Z nodes represent a significant overlap given the number of activating and/or inhibiting edges of the individual X and Y nodes. After Bonferroni correction, FFLs with $p < 0.01$ are considered significant and kept. X-Y FFL pairs were only counted once, regardless of the number of shared Z nodes.

Significant Overrepresentation of FFLs

To test overrepresentation of FFLs, the GPN was permuted 10,000 times, keeping the indegree, outdegree, and mutual degree of every single node in the graph identical to the original GPN and only swapping edges between nodes. The permuted networks therefore contain the exact same background distribution as the original GPN and provide a stringent assessment of the significance of overrepresented FFLs. Over- or underrepresentation was tested against the permuted networks using Z score-transformed counts for each FFL type. p values derived from the Z scores were Bonferroni corrected.

ACCESSION NUMBERS

The ArrayExpress accession number for responsive mutants is E-MTAB-1383. The GEO accession number for responsive mutants is GSE42527. The ArrayExpress accession number for nonresponsive mutants is E-MTAB-1384. The GEO accession number for nonresponsive mutants is GSE42526. The ArrayExpress accession numbers for the WT data sets are E-TABM-773, E-TABM-984, E-TABM-1351, and E-TABM-1352. The GEO accession numbers for the WT data sets are GSE42215, GSE42217, GSE42241, and GSE42240.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, five figures, two tables, and two data files and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.02.054>.

AUTHOR CONTRIBUTIONS

F.C.P.H. and P.K. arranged funding; F.C.P.H., P.K., D.v.L., M.J.A.G.K., A.J.M., H.v.B., E.S., S.R.v.H., P.L., L.V.B., C.W.K., J.J.B., and T.L.L. set up and maintained infrastructure; L.A.L.v.d.P., J.J.B., T.L.L., T.M., E.A., S.v.W., S.v.H., M.M.K., G.A.-M., M.O.B., N.A.C.H.B., D.B., D.v.L., and M.J.A.G.K. carried out and analyzed experiments; F.C.P.H., P.K., K.S., E.O.D., B.S., P.L., L.A.L.v.d.P., J.J.B., and T.L.L. performed large-scale analyses and interpretation; P.K. and K.S. created figures; and F.C.P.H., P.K., K.S., T.L.L., T.M., E.O.D., D.v.L., and L.A.L.v.d.P. wrote the manuscript.

ACKNOWLEDGMENTS

This study was supported by the Netherlands Bioinformatics Centre and the Netherlands Organization of Scientific Research (grants 016108607, 81702015, 05071057, 91106009, and 021002035 to T.L.L.; grants 863.07.007 and 864.11.010 to P.K.; and grant 70057407 to J.J.B.).

Received: October 7, 2013

Revised: December 30, 2013

Accepted: February 25, 2014

Published: April 24, 2014

REFERENCES

- Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* 13, 720–731.
- Alon, U. (2007). Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* 8, 450–461.
- Badis, G., Chan, E.T., van Bakel, H., Pena-Castillo, L., Tillo, D., Tsui, K., Carlson, C.D., Gossett, A.J., Hasinoff, M.J., Warren, C.L., et al. (2008). A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. *Mol. Cell* 32, 878–887.
- Barabási, A.-L. (2009). Scale-free networks: a decade and beyond. *Science* 325, 412–413.
- Benschop, J.J., Brabers, N., van Leenen, D., Bakker, L.V., van Deutekom, H.W.M., van Berkum, N.L., Apweiler, E., Lijnzaad, P., Holstege, F.C.P., and Kemmeren, P. (2010). A consensus of core protein complex compositions for *Saccharomyces cerevisiae*. *Mol. Cell* 38, 916–928.
- Bonke, M., Turunen, M., Sokolova, M., Vähärautio, A., Kivioja, T., Taipale, M., Björklund, M., and Taipale, J. (2013). Transcriptional networks controlling the cell cycle. *G3 (Bethesda)* 3, 75–90.
- Breitkreutz, A., Choi, H., Sharom, J.R., Boucher, L., Neduvia, V., Larsen, B., Lin, Z.-Y., Breitkreutz, B.-J., Stark, C., Liu, G., et al. (2010). A global protein kinase and phosphatase interaction network in yeast. *Science* 328, 1043–1046.
- Broach, J.R. (2012). Nutritional control of growth and development in yeast. *Genetics* 192, 73–105.
- Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., et al. (2012). *Saccharomyces Genome Database*: the genomics resource of budding yeast. *Nucleic Acids Res.* 40 (Database issue), D700–D705.
- Chua, G., Morris, Q.D., Sopko, R., Robinson, M.D., Ryan, O., Chan, E.T., Frey, B.J., Andrews, B.J., Boone, C., and Hughes, T.R. (2006). Identifying transcription factor functions and targets by phenotypic activation. *Proc. Natl. Acad. Sci. USA* 103, 12045–12050.
- Collins, S.R., Kemmeren, P., Zhao, X.-C., Greenblatt, J.F., Spencer, F., Holstege, F.C.P., Weissman, J.S., and Krogan, N.J. (2007). Toward a comprehensive atlas of the physical interactome of *Saccharomyces cerevisiae*. *Mol. Cell. Proteomics* 6, 439–450.
- 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.
- David, L., Huber, W., Granovskia, M., Toedling, J., Palm, C.J., Bofkin, L., Jones, T., Davis, R.W., and Steinmetz, L.M. (2006). A high-resolution map of transcription in the yeast genome. *Proc. Natl. Acad. Sci. USA* 103, 5320–5325.
- de Bruin, R.A.M., Kalashnikova, T.I., Chahwan, C., McDonald, W.H., Wohlschlegel, J., Yates, J., 3rd, Russell, P., and Wittenberg, C. (2006). Constraining G1-specific transcription to late G1 phase: the MBF-associated corepressor Nrm1 acts via negative feedback. *Mol. Cell* 23, 483–496.
- DeRisi, J.L., Iyer, V.R., and Brown, P.O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278, 680–686.
- Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* 95, 14863–14868.

- Fisk, D.G., Ball, C.A., Dolinski, K., Engel, S.R., Hong, E.L., Issel-Tarver, L., Schwartz, K., Sethuraman, A., Botstein, D., and Cherry, J.M.; Saccharomyces Genome Database Project (2006). *Saccharomyces cerevisiae* S288C genome annotation: a working hypothesis. *Yeast* 23, 857–865.
- Fuda, N.J., Ardehali, M.B., and Lis, J.T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* 461, 186–192.
- Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., and Brown, P.O. (2000). Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* 11, 4241–4257.
- 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.
- Good, M.C., Zalatan, J.G., and Lim, W.A. (2011). Scaffold proteins: hubs for controlling the flow of cellular information. *Science* 332, 680–686.
- Hahn, S., and Young, E.T. (2011). Transcriptional regulation in *Saccharomyces cerevisiae*: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics* 189, 705–736.
- Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N.J., MacIsaac, K.D., Danford, T.W., Hannett, N.M., Tagne, J.-B., Reynolds, D.B., Yoo, J., et al. (2004). Transcriptional regulatory code of a eukaryotic genome. *Nature* 431, 99–104.
- Holstege, F.C., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner, C.J., Green, M.R., Golub, T.R., Lander, E.S., and Young, R.A. (1998). Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* 95, 717–728.
- Hu, Z., Killion, P.J., and Iyer, V.R. (2007). Genetic reconstruction of a functional transcriptional regulatory network. *Nat. Genet.* 39, 683–687.
- Hughes, T.R., and de Boer, C.G. (2013). Mapping yeast transcriptional networks. *Genetics* 195, 9–36.
- Hughes, T.R., Marton, M.J., Jones, A.R., Roberts, C.J., Stoughton, R., Armour, C.D., Bennett, H.A., Coffey, E., Dai, H., He, Y.D., et al. (2000). Functional discovery via a compendium of expression profiles. *Cell* 102, 109–126.
- Ideker, T., Galitski, T., and Hood, L. (2001). A new approach to decoding life: systems biology. *Annu. Rev. Genomics Hum. Genet.* 2, 343–372.
- Jacquier, A. (2009). The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. *Nat. Rev. Genet.* 10, 833–844.
- Kemmeren, P., van Berkum, N.L., Vilo, J., Bijma, T., Donders, R., Brazma, A., and Holstege, F.C.P. (2002). Protein interaction verification and functional annotation by integrated analysis of genome-scale data. *Mol. Cell* 9, 1133–1143.
- Koch, C., Moll, T., Neuberg, M., Ahorn, H., and Nasmyth, K. (1993). A role for the transcription factors Mbp1 and Swi4 in progression from G1 to S phase. *Science* 261, 1551–1557.
- Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., et al. (2002). Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298, 799–804.
- Lenstra, T.L., Benschop, J.J., Kim, T., Schulze, J.M., Brabers, N.A.C.H., Margaritis, T., van de Pasch, L.A.L., van Heesch, S.A.A.C., Brok, M.O., Groot Koerkamp, M.J.A., et al. (2011). The specificity and topology of chromatin interaction pathways in yeast. *Mol. Cell* 42, 536–549.
- Liu, Z., and Butow, R.A. (2006). Mitochondrial retrograde signaling. *Annu. Rev. Genet.* 40, 159–185.
- MacIsaac, K.D., Wang, T., Gordon, D.B., Gifford, D.K., Stormo, G.D., and Fraenkel, E. (2006). An improved map of conserved regulatory sites for *Saccharomyces cerevisiae*. *BMC Bioinformatics* 7, 113.
- Macneil, L.T., and Walhout, A.J.M. (2011). Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Res.* 21, 645–657.
- Mangan, S., and Alon, U. (2003). Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. USA* 100, 11980–11985.
- Margaritis, T., Lijnzaad, P., van Leenen, D., Bouwmeester, D., Kemmeren, P., van Hooff, S.R., and Holstege, F.C.P. (2009). Adaptable gene-specific dye bias correction for two-channel DNA microarrays. *Mol. Syst. Biol.* 5, 266.
- Margaritis, T., Oreal, V., Brabers, N., Maestroni, L., Vitaliano-Prunier, A., Benschop, J.J., van Hooff, S., van Leenen, D., Dargemont, C., Gél, V., and Holstege, F.C. (2012). Two distinct repressive mechanisms for histone 3 lysine 4 methylation through promoting 3'-end antisense transcription. *PLoS Genet.* 8, e1002952.
- McIsaac, R.S., Petti, A.A., Bussemaker, H.J., and Botstein, D. (2012). Perturbation-based analysis and modeling of combinatorial regulation in the yeast sulfur assimilation pathway. *Mol. Biol.* 23, 2993–3007.
- Radonjic, M., Andrau, J.-C., Lijnzaad, P., Kemmeren, P., Kockelkorn, T.T.J.P., van Leenen, D., van Berkum, N.L., and Holstege, F.C.P. (2005). Genome-wide analyses reveal RNA polymerase II located upstream of genes poised for rapid response upon *S. cerevisiae* stationary phase exit. *Mol. Cell* 18, 171–183.
- Roberts, C.J., Nelson, B., Marton, M.J., Stoughton, R., Meyer, M.R., Bennett, H.A., He, Y.D., Dai, H., Walker, W.L., Hughes, T.R., et al. (2000). Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. *Science* 287, 873–880.
- Savageau, M.A. (1977). Design of molecular control mechanisms and the demand for gene expression. *Proc. Natl. Acad. Sci. USA* 74, 5647–5651.
- Sharan, R., and Ideker, T. (2006). Modeling cellular machinery through biological network comparison. *Nat. Biotechnol.* 24, 427–433.
- Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.-L., and Ideker, T. (2011). Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27, 431–432.
- Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* 13, 613–626.
- Steinfeld, I., Shamir, R., and Kupiec, M. (2007). A genome-wide analysis in *Saccharomyces cerevisiae* demonstrates the influence of chromatin modifiers on transcription. *Nat. Genet.* 39, 303–309.
- Struhl, K. (1999). Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell* 98, 1–4.
- Teng, X., Dayhoff-Brannigan, M., Cheng, W.-C., Gilbert, C.E., Sing, C.N., Diny, N.L., Wheelan, S.J., Dunham, M.J., Boeke, J.D., Pineda, F.J., and Hardwick, J.M. (2013). Genome-wide consequences of deleting any single gene. *Mol. Cell* 52, 485–494.
- van Bakel, H., and Holstege, F.C.P. (2004). In control: systematic assessment of microarray performance. *EMBO Rep.* 5, 964–969.
- van de Peppel, J., Kemmeren, P., van Bakel, H., Radonjic, M., van Leenen, D., and Holstege, F.C.P. (2003). Monitoring global messenger RNA changes in externally controlled microarray experiments. *EMBO Rep.* 4, 387–393.
- van de Peppel, J., Kettelerij, N., van Bakel, H., Kockelkorn, T.T.J.P., van Leenen, D., and Holstege, F.C.P. (2005). Mediator expression profiling epistasis reveals a signal transduction pathway with antagonistic submodules and highly specific downstream targets. *Mol. Cell* 19, 511–522.
- van Wageningen, S., Kemmeren, P., Lijnzaad, P., Margaritis, T., Benschop, J.J., de Castro, I.J., van Leenen, D., Groot Koerkamp, M.J.A., Ko, C.W., Miles, A.J., et al. (2010). Functional overlap and regulatory links shape genetic interactions between signaling pathways. *Cell* 143, 991–1004.
- Vidal, M., Cusick, M.E., and Barabási, A.-L. (2011). Interactome networks and human disease. *Cell* 144, 986–998.
- Walhout, A.J.M., and Vidal, M. (2001). Protein interaction maps for model organisms. *Nat. Rev. Mol. Cell Biol.* 2, 55–62.
- Weiner, A., Chen, H.V., Liu, C.L., Rahat, A., Klien, A., Soares, L., Gudipati, M., Pfeffner, J., Regev, A., Buratowski, S., et al. (2012). Systematic dissection of roles for chromatin regulators in a yeast stress response. *PLoS Biol.* 10, e1001369.
- Zhu, C., Byers, K.J.R.P., McCord, R.P., Shi, Z., Berger, M.F., Newburger, D.E., Saulrieta, K., Smith, Z., Shah, M.V., Radhakrishnan, M., et al. (2009). High-resolution DNA-binding specificity analysis of yeast transcription factors. *Genome Res.* 19, 556–566.

Supplemental Information

EXTENDED EXPERIMENTAL PROCEDURES

All protocols are described in full below. Protocols related to expression-profiling have also been deposited in ArrayExpress (<http://www.ebi.ac.uk/arrayexpress>) (Parkinson et al., 2011), with accession numbers as indicated in the relevant sections.

Yeast Strains

Haploid MATalpha (BY4742) strains from the gene deletion consortium (Giaever et al., 2002) were profiled. These are isogenic to the sequenced yeast strain S288c and were obtained from two separate copies of the deletion collection: Euroscarf (Frankfurt, Germany) and Open Biosystems (Huntsville, USA). Some strains were unavailable/inviolate as MATalpha. For these strains (0.7%), the MATa strain (BY4741) was profiled and compared to its corresponding WT. Some strains had to be remade as they showed defects upon profiling (described below, quality control on yeast strains). These strains (101, 6.8%) were remade using the kanamycin cassette from pFA6a-kanMX6 (Longtine et al., 1998). The defective strains may be common to all copies of the collection but may also have arisen during generation of copies of the collection. All strains used are described individually in Table S1.

Quality Controls on Yeast Strains

Deleted Gene Not Down

In some deletion strains the deleted gene is not obviously decreased in expression. This is usually due to an already low expression in WT. In all cases whereby the p value for decrease in expression of the deleted gene was insignificant ($p > 0.05$), the deletion strain was further checked by two PCRs, one using two primers outside the presumed deleted gene and another using one primer outside the gene and the other primer in the middle of the selection marker. If the PCR reactions were positive the profile was kept. If PCR validation was negative then the strain was remade and reprofiled.

Aneuploidy

Aneuploidy is revealed in the expression-profiles by analyzing expression changes in the context of chromosome location (Hughes et al., 2000), one of the standard quality controls (QCs) performed on all profiles. Aneuploid mutants were remade and reprofiled. One exception is *asf1Δ*, for which the strain was not remade, because only one of the two duplicates showed a small duplication, affecting less than 15 genes. These genes did not exhibit significant changes in gene expression after averaging the replicates. If the expression-profile of the remade strain also showed aneuploidy, the mutant was excluded from the study.

Spurious Mutations

Some deletion strains passed all QC criteria but had surprising profiles in light of what was previously known about the deleted gene. If established protein complex or pathway members showed the same unanticipated profile, the profiles were kept since verification was through a completely independent strain. If complex or pathway member profiles were different, or if verification through another strain was not possible, then the strain was remade. If the unanticipated gene expression changes were no longer present in the remade strain this indicates the presence of spurious addition mutations in the original. In such cases the old profile was removed from the data set and replaced with the profile from the newly made strain.

Protocols

Unless otherwise specified all water (including water for RNA dilution, measurements, etc.) is fresh milliQ treated water, autoclaved in a blue-top (low silica) Nalgene glass bottle. Some of the profiles in the data set were generated in the course of previous studies (Lenstra et al., 2011; van Wageningen et al., 2010). During the course of those studies all protocols were optimized for higher throughput, switching from manual procedures to the (semi-) automated procedures described here. The manual versions of these protocols are described previously (Lenstra et al., 2011). Extensive comparison of WT and various mutant strain profiles generated with both protocols revealed no systematic differences other than an increase in reproducibility.

Yeast Cultures: ArrayExpress Accession P-UMCU-50

Strains were streaked from -80°C stocks onto 0.2 mg/ml G418 YPD plates and grown at 30°C for 3 days. A few (<3%) severely slow mutants were grown on plate for 4 or 5 days. For these the accompanying WTs were streaked one or two days later so that mutant and WT overnight liquid cultures were inoculated in parallel, with similar sized colonies, fresh from 30°C plates and without temporary storage of the WT colonies at 4°C . Liquid cultures were inoculated with independent colonies from each strain and grown overnight in synthetic complete (SC) medium: 2gr/l amino acid drop-out mix complete, 6.71 gr/l yeast nitrogen base (YNB, without amino acids, without carbohydrate and with ammonium sulfate) all from US Biologicals (Swampscott, USA), with 2% D-glucose. Media was made from a single batch of powder, using an automated, programmable media autoclave for reproducibility of media preparation. The WT doubling-time in this medium is 90 min, virtually identical to growth in YPD. Other recipes of SC +2% glucose can result in longer doubling times. This can be overcome by adding higher concentrations of amino acids.

Overnight cultures were diluted to an OD_{600} of 0.15 in 1.5 ml fresh medium and grown at 30°C in a 24 well plate in a Tecan Infinite F200 plate shaker under continuous shaking. For each mutant strain two cultures from two isolates were grown in batches in parallel with two WT cultures (same day WTs). The plate shaker monitors OD_{600} . Cultures were harvested by centrifugation (3450 g, 3 min) at early mid-log phase at an OD_{600} of 0.6 ± 0.1 and pellets were immediately frozen in liquid nitrogen after removal of supernatant. No more than four cultures were harvested simultaneously to decrease processing time.

OD₆₀₀ measurements of cell numbers are spectrophotometer dependent because it is a reading of light-scattering rather than of absorbance. OD₆₀₀ 0.6 corresponds to early mid-log phase for these cultures. It is essential to harvest at an OD₆₀₀ that corresponds to early mid-log phase. High resolution time-course analysis of the entire growth curve shows that early to mid-log phase represents the window during which expression-profiles are identical along the growth curve. Cultures start showing significant changes in gene expression after mid-log phase and long before growth actually slows down (Radonjic et al., 2005). Adherence to early mid-log OD₆₀₀ is particularly important for (a minority of) mutants that have significantly slower growth compared to WT, since for some of these mutants, OD₆₀₀ of 0.6 is further along their relative growth curve. Problems associated with this are overcome if all mutants are harvested at an OD₆₀₀ that represents early mid-log for WT cultures.

RNA Extraction and Automated Purification: ArrayExpress Accession P-UMCU-51

Total RNA was prepared by phenol extraction and cleaned up on a customized Caliper Sciclone ALH 3000 workstation that included a PCR machine PTC-200 (Bio-Rad Laboratories), SpectraMax 190 spectrophotometer (Molecular Devices) and a magnetic bead-locator (Beckman). For a movie of this automated setup see: <http://microarrays.holstegelab.nl/modx/index.php?id=66>.

RNA Extraction

Frozen cells (-80°C) were resuspended in 500 μl acid phenol-chloroform (5:1, pH 4.7). An equal volume of TES-buffer (TES: 10 mM Tris pH 7.5, 10 mM EDTA, 0.5% SDS) was immediately added. Samples were vortexed very vigorously for 20 s and incubated in a water bath for 10 min at 65°C and vortexed again. Samples were put in a shaking incubator (65°C , 1400 rpm) for 50 min. Samples were centrifuged for 20 min at 20817 g at 4°C . The aqueous phase was extracted and phenol extraction was repeated on this, followed by an otherwise identical chloroform:isoamyl-alcohol (25:1) extraction. RNA was precipitated with 50 μl sodium acetate (NaAc 3M, pH 5.2) and ethanol (96%, -20°C), left at -20°C for at least 30 min and centrifuged for 10 min at 20817 g at 4°C . Pellet was washed with 80% ethanol (-20°C) and dissolved in 90 μl sterile water (milliQ filtered), snapfrozen and stored at -80°C .

RNA Purification

- Concentration of total RNA is measured by mixing 5 μl of the RNA sample with 45 μl water and determining the A₂₆₀ from an absorbance spectrum on the SpectraMax.
- 2.5 μl DNaseI dilution (RNase-free DNase kit, QIAGEN, nr 79254, diluted 1:5 in RDD) is mixed with the remaining 87.5 μl RNA-solution and incubated @ 18°C for 15 min.
- RNA is purified and concentrated with RNAClean (Agencourt, Beckman) according to manufacturers' protocol, to an end volume of 25 μl .
- 5 μl cleaned RNA is mixed with 95 μl water and measured on the SpectraMAX, concentrations are normalized to 0.2 $\mu\text{g}/\mu\text{l}$ in each well.
- 5 μl of cleaned and normalized RNA is used to set up a start plate for RNA amplification.
- 1 μl of cleaned and normalized RNA is used to check integrity by running on a capillary electrophoresis (QiaXcel, QIAGEN) system.
- All plates are snap frozen and stored at -80°C until further use.

External Controls

External control poly-A+ RNAs were added as a mixture to the total RNA to enable monitoring of global transcriptome changes in mutants, as described (van de Peppel et al., 2003). The external controls are EC1-9 and they correspond to *Bacillus subtilis* genes *ycxA*, *yceG*, *ybdO*, *ybbR*, *ybaS*, *ybaF*, *ybaC*, *yacK* and *yabQ* respectively, cloned between the *Xba*I and *Bam*H I sites in pT7T3 (Amersham Pharmacia Biotech) with an additional 30-nucleotide poly(A) sequence between the gene and the *Xho*I site (van de Peppel et al., 2003). For making RNA, plasmids were digested with *Xho*I for use in in vitro transcription reactions using MEGAscript-T7 (Ambion). The external control mixture was made so that each 1 μg of yeast total RNA received the following amounts of the different controls: EC1 0.8 pmol, EC2 0.1 pmol, EC3 0.11 pmol, EC5 0.035 pmol, EC6 0.0075 pmol, EC7 0.0025 pmol, EC9 0.0004 pmol. Probes for the controls (van de Peppel et al., 2003) are present multiple times on each subarray of the microarray, enabling external control normalization (van de Peppel et al., 2003).

Robotic RNA Amplification and Labeling; ArrayExpress Accession P-UMCU-38

All robotic RNA amplification and labeling procedures were performed in 96 well plates (4titude, Bioko) on a customized Sciclone ALH 3000 Workstation (Caliper LifeSciences) that included a PCR PTC-200 (Bio-Rad Laboratories), SpectraMax 190 spectrophotometer (Molecular Devices), and a magnetic bead-locator (Beckman). For a movie of this automated setup see: <http://microarrays.holstegelab.nl/modx/index.php?id=66>.

RNA Amplification

For RNA amplification, total RNA samples were diluted to 0.2 $\mu\text{g}/\mu\text{l}$ and 5 μl is put in a 96-wells plate (Abgene).

All subsequent steps were performed by the robot:

- Mix1 (5 μl) containing 100 ng T7 Mlu VN primer (5' GGA GGC CGG AGA ATT GTA ATA CGA CTC ACT ATA GGG AGA CGC GTG TTT TTT TTT TTT TTT TTT VN 3', whereby V = G, A or C; N = G, A, C or T) and external control RNAs, is added to the 5 μl total RNA and mixed in each well. Plate is incubated at 70°C for 10 min and cooled to 48°C .

2. Mix2 containing 4 µl 5x 1st strand buffer, 2 µl 0.1 M DTT (Invitrogen), 1 µl RNase Inhibitor (Boehringer), 1 µl 20 mM dNTPs (GE Healthcare), 1 µl linear acrylamide, and 1 µl SuperScriptIII (Invitrogen) per sample is prewarmed to 48°C. 10 µl per sample is added and mixed in each well.
3. Plate is incubated at 48°C for 2 hr and cooled to room temperature.
4. 106 µl water and subsequently mix3 containing 15 µl second strand buffer, 3 µl 20 mM dNTPs (GE Healthcare), 1 µl T4 DNA ligase, 4 µl E.coli DNA polymerase I and 1 µl RNaseH (Promega) is added and mixed in each well.
5. Plate is incubated at 16°C for 2 hr, at 65°C for 10 min. Double-stranded cDNA product is purified and concentrated with RNAClean (Agencourt, GC biotech) according to manufacturers' protocol, to an end volume of 25 µl.
6. 8 µl cDNA is put in a 96 well plate and mixed with 12 µl IVT mix containing 2 µl 10x rxn-buffer, 2 µl of each ATP, CTP, GTP, 0.6 µl UTP, 2 µl T7 enzyme mix (MegaScript kit, Ambion, Applied Biosystems), and 2.1 µl 50 mM 5-(3-aminoallyl)-UTP (Ambion, Applied Biosystems).
7. Plate is incubated at 37°C for 4 hr. cRNA product is purified with RNAClean (Agencourt, Beckman) according to manufacturers' protocol.
8. Concentration is measured (SpectraMax 190) and adjusted to 600 ng/µl. An aliquot of the resulting cRNA plates are sampled QC by Qiaxcel, the remainder is snapfrozen and stored at -80°C.

Labeling

1. NHS-ester Cy3 or Cy5 dye (Amersham PA 23001 and 25001): entire tube is resuspended in 100 µl DMSO (Merck 8.02912.10).
2. 8 µl of each cRNA sample (0.6 µg/µl) is put in a 96-well plate (Abgene), the accompanying common reference sample is put in the next column (final step combines Cy3 and Cy5 labeled material).
3. 3 µl 0.5 M Sodium Bicarbonate buffer, pH 9 is added and mixed to all wells.
4. 3 µl Cy-dye solution is added and mixed to the appropriate wells, plate is incubated at 18°C for 1 hr.
5. 4.5 µl 5 M hydroxylamine is added and mixed, incubated at 18°C for 15 min. Labeled cRNA product is purified with RNA Clean (Agencourt, GC biotech) according to manufacturers' protocol.
6. RNA concentration and labeling incorporation are measured (SpectraMax 190).
7. 2.5 µg of each labeled sample and reference cRNA are pooled and subjected to fragmentation according to protocol (Ambion, Applied Biosystems), 15 min at 70°C. Samples are stored at -20°C until hybridization.

QC of RNA and Labeled cRNA

An aliquot of total RNA was QCed by capillary electrophoresis using a Qiaxcel (QIAGEN Benelux, Venlo). QC at this step is based on the size and shape of rRNA bands. Total RNA yield per OD₆₀₀ unit of cells was also monitored. cRNA yield was monitored by A₂₆₀. Additionally an aliquot of cRNA was tested by capillary electrophoresis (Qiaxcel) whereby QC was based on the size distribution of cRNA. Cy3/Cy5 labeling of cRNA was determined with an absorbance scan using A₅₅₀ and A₆₄₉ to monitor incorporation of Cy3 and Cy5 respectively. Minimum incorporation was 1%. Maximum incorporation was 2%.

Experiment Design

The experiment design is depicted in [Figure S1B](#). Two channel microarrays were used. RNA isolated from a large amount of WT yeast was used in this common reference design, in one of the channels for each hybridization and used in the statistical analysis to obtain an average expression-profile for each deletion mutant relative to the WT. Two independent cultures were hybridized on two separate microarrays. For the first hybridization the Cy5 (red) labeled cRNA from the deletion mutant is hybridized together with the Cy3 (green) labeled cRNA from the common reference. For the replicate hybridization from the independent cultures, the labels are swapped. Each gene is represented twice on the microarray, resulting in four measurements per mutant.

DNA Microarrays: ArrayExpress Accession P-UMCU-34, Resulting in Arrays with ArrayExpress Accession A-UMCU-10

DNA microarray slides containing 70-mer oligonucleotides from the Operon Array-Ready Oligo Set (Operon biotechnologies, Huntsville, USA) were printed in a temperature- and humidity-controlled (18°C, 45% humidity), particle-filtered clean-room (ISO7/downflow ISO5), using the following protocol. Oligos are resuspended and adjusted to 10 µM in 150 mM phosphate buffer, pH8.5 in Genetix 384-well plates. Array production is done using a Biorobotics MicroGridII spotter on Codelink activated slides (Surmodics Inc, Eden Prairie, USA). Slide postprocessing is carried out according to the following CodeLink protocol. Printed slides are kept at 20°C, 75% humidity for 24 hr. Residual reactive groups are blocked using prewarmed blocking solution (50 mM ethanolamine, 0.1 M Tris (pH 9) at 50°C for 30 min. Slides are rinsed thoroughly with milliQ water. Slides are washed with 4x SSC, 0.1% SDS (prewarmed to 50°C) for 30 min at 50°C on a shaker. Slides are rinsed with milliQ water, dry-centrifuged and stored desiccated at room temperature. Each gene probe is spotted twice on the array, which contain an additional 2,838 control features for spike-in external control normalization and quality control ([van de Peppel et al., 2003](#)).

Hybridization: ArrayExpress Accession P-UMCU-39

2.5 µg of each labeled sample in a total volume of 60 µl is combined with 60 µl 2x-hybmix, containing 50% formamide, 10xSSC, 0.2% SDS, 200 µg/ml herring sperm DNA. The hybridizations were performed for 16 hr at 42°C in a HS4800Pro hybstation (Tecan,

Männedorf, Switzerland) as detailed below. Hybridization was performed in a temperature- and humidity-controlled laboratory (20°C, 38% humidity), filtered to remove ozone which was monitored to ensure that this was lower than 5 ppb.

1. 60 µl labeled sample is combined with 60 µl 2x-hyb mix, containing 50% formamide, 10xSSC, 0.2% SDS, 200 µg/ml herring sperm DNA.
2. Hybridizations are performed on an HS4800Pro Hybstation (Tecan, Männedorf, Switzerland).
3. Priming (30 s wash and 30 s soak) with 5xSSC, 0.1%SDS at 42°C.
4. Injection of pre-hyb mix: 5xSSC, 25% formamide, 0.1%SDS, 1%BSA, total volume 110 µl.
5. Prehybridization: 45 min at 42°C, agitation frequency medium, other settings off.
6. Wash in milliQ (2 min wash, 1 min soak), at 42°C, 2x.
7. Wash in 5xSSC, 0.1%SDS (45 s), at 42°C.
8. Injection of sample. Volume 110 µl.
9. Hybridization: 16 hr at 42°C, agitation frequency medium, other settings off.
10. Wash (1 min wash, 1 min soak) in 1xSSC, 0.2%SDS at 23°C, 2x.
11. Wash (1 min wash, 1 min soak) in 0.1xSSC, 0.2%SDS at 23°C, 2x.
12. Wash (1 min wash, 1 min soak) in 0.1xSSC at 23°C, 4x.
13. Drying: blow with nitrogen for 3 min at 30°C.

Scanning and Image Analysis: ArrayExpress Accession P-UMCU-40, ArrayExpress Accession P-UMCU-42

Scanning was performed in a temperature- and humidity-controlled laboratory (20°C, 38% humidity), filtered to remove ozone which was monitored to ensure that this was lower than 5 ppb. Slides were scanned using a G2565BA scanner that has a 48 slide carousel (Agilent, California, USA) at 100% laser power and 30% PMT. After scanning, the intensities for the Cy5 (Red) and Cy3 (Green) channels were automatically extracted using the batch-processing module in ImaGene version 8.0.1 (Biodiscovery, California, USA). For spot finding a local flexibility of 2.0 pixels was used. For segmentation the following settings were used: background buffer: 3.0; background width: 3.0; signal percentages: 3% (low), 97% (high); background percentages: 3% (low), 97% (high). Measurements exported: mean, median, total, standard deviation and area of the foreground and background signals. Batch editor configuration files are available upon request.

Microarray QC

Each hybridization performed within this project was subjected to a number of quality controls. Some of these are based on the data from a single hybridization, while others are based on comparing data from multiple hybridizations.

Quality Controls for a Single Hybridization

For each hybridization a quality report was generated that contained a number of quality controls. For all of these quality controls either raw nonbackground corrected mean intensity values were used, or data normalized on all gene probes using the print-tip LOESS (Yang et al., 2002) algorithm (*marray* R package version 1.20.0) with a window span of 0.4 and excluding genes with nearly saturated signals (i.e., mean intensity > 2¹⁵) for the loess curve estimation. The following plots were generated for each hybridization:

1. Spatial distribution plot of all Signal-to-Noise ratios (SNR) according to the position on the microarray for the red and green raw intensity values. SNRs are binned according to the number of standard deviations that the mean intensity signal is above the background and plotted using different colors; SNR < 2: black; 2 ≤ SNR < 3: red; 3 ≤ SNR < 4: orange; SNR > = 4: yellow.
2. Spatial distribution plot of all $M(\log_2(R/G))$ ratios according to the position on the microarray for both raw and normalized data.
3. MA plot ($\log_2(R/G)$ versus $0.5 * \log_2(R^*G)$) for both raw and normalized data. For each data type three different MA plots were generated, one containing all probes present on the microarray, one containing all gene probes present on the microarray and one containing all quality controls and external control probes present on the microarray.
4. Spatial distribution plot of local background intensity values according to the position on the microarray for the red and green channels. Local background intensity values are binned relative to the average background intensity of the entire microarray; background > 2.1*average: black; 1.7*average < background ≤ 2.1*average: red; 1.3*average < background ≤ 1.7*average: orange; background < 1.3*average: yellow.
5. Histograms of the distribution of raw red and green intensities and raw M ratios.
6. Chromosomal location plot of all normalized M ratios according the chromosomal position of the gene represented by the probe on the microarray.
7. Histograms of the distribution of normalized M ratios per chromosome. Individual histograms are colored red if the chromosome-specific M is significantly larger than the average M over all chromosomes by at least 0.15; green if they are similarly smaller and yellow otherwise. Significant in this case means an average Benjamini-Hochberg FDR corrected p value of < 10⁻⁶ as derived from a pairwise t test.

As a first QC, the report generated above was judged. If a hybridization showed an irregular pattern in any of the visual quality control plots due to for instance air bubbles, scratches on the surface, dust, minor printing defects, uneven spread of the hybridization

solution or low signal, the hybridization was excluded from further analysis and the RNA for the deletion mutant was re-hybridized, if based on the available data, the deletion mutant was deemed to have a significant expression-profile.

In addition to the visual quality controls described above, a number of numerical quality measures were calculated for each hybridization as proposed by Chen and coworkers (Chen et al., 2002). Each probe spot present on the microarray was classified as good, marginal or bad depending on the calculated quality measures for the raw intensity values. Good spots have the following characteristics: $\text{SNR} > 4$; background flatness > 6 ; signal consistency > 1.1 ; foreground signal $>$ background signal. Bad spots have the following characteristics: $\text{SNR} \leq 2$; background flatness ≤ 4 ; signal consistency ≤ 0.9 ; foreground signal \leq background signal (black hole). Background flatness is calculated by comparing the background intensity of a probe to the mean background intensity of the subgrid. Signal consistency is calculated from the signal coefficient of variation. An overall percentage of good marginal and bad spots for the gene and control probe groups is reported and judged (see Quality controls for entire project below).

Quality Controls for the Entire Project

To assess the performance of individual hybridizations within this project, hybridizations were ranked according to the percentage of good gene probes. Individual hybridizations for deletion mutants with fewer than 93% good gene probes were considered outliers, were removed and the experiment repeated if based on the available data the mutant had a significant expression-profile.

Each deletion mutant expression-profile was also compared against the expression-profile of the WT grown in parallel to ensure that the effect observed in the deletion mutant is specific for the deletion mutant and is not present in the WT grown in parallel. If a significant overlap was found (hypergeometric test; $p < 0.01$) between the significantly changed genes in the deletion mutant and the WT grown in parallel and the number of genes changing significantly in the deletion mutant expression-profile is more than seven, the hybridizations for the deletion mutant were removed and the experiment repeated.

In addition, replicate deletion mutant expression-profiles were also compared against each other. Limma analyses were performed using the (independent culture) replicates of the mutant, resulting in two expression-profiles per deletion mutant. If a significant overlap was not found (hypergeometric test; $p > 0.01$) between the significantly regulated genes in both deletion mutant expression-profiles and the number of genes changing significantly in the deletion mutant expression-profile is more than seven, the hybridizations for the deletion mutant were removed and the experiment repeated.

Data Normalization: ArrayExpress Accession P-UHCU-43

Microarray data normalization was performed on mean intensity values using print-tip LOESS as described (Yang et al., 2002) and implemented in the *marray* R package version 1.20.0, using no background subtraction and a window span of 0.4. Probes flagged as absent, or with a (nearly) saturated signal (i.e., $> 2^{15}$) in either channel were not considered for the estimation of the LOESS curve. Gene-specific dye bias (GSDB) is not addressed by LOESS, and was corrected after the LOESS normalization using the Gene-And Slide-Specific COrrection (GASSCO) (Margaritis et al., 2009) algorithm implemented in the R package *dyebias* version 1.7.8. The correction consists of two components: the intrinsic dye bias of a specific probe and the degree of dye bias observed in a specific slide. The correction is based on the following formula:

$$M_{ij}^* = M_{ij} + GSDB_{ij} = M_{ij} + iGSDB_i \times F_j$$

Where M_{ij}^* is the measured (biased) \log_2 fold-change of gene i in hybridization j , M_{ij} is its unbiased version, $GSDB_{ij}$ is the gene-specific dye bias component of M_{ij}^* . The latter is the product of the so-called intrinsic gene-specific dye bias of gene i ($iGSDB_i$) and the slide bias of hybridization j (F_j). The corrected \log_2 fold-change M_{ij} is then calculated using:

$$M_{ij} = M_{ij}^* - iGSDB_i \times F_j$$

The $iGSDBs$ are estimated only once in a set of 200 WT versus WT hybridizations (ArrayExpress accession E-TABM-773) and are used for all slides within this project. The slide bias F is estimated for each hybridization separately, using two groups of probes, having the strongest red or green biases, defined as those with an $iGSDB$ in the top and bottom 5th percentiles of $iGSDBs$, respectively. The slide bias is the mean of these red and green probe group's median ($M / iGSDB$)-ratio.

Global Normalization

Default microarray data normalization was performed by LOESS (Yang et al., 2002) followed by a gene-specific dye bias correction (GSDB) (Margaritis et al., 2009) as described above. The mean of all genes is zero after this normalization, since the underlying assumption of LOESS is no global change, even if expression of all genes has been affected. In order to control for possible global changes in expression, the ratio of the spiked-in external controls was checked. These are external control polyA RNAs, spiked in to the total RNA of all samples (as described above) and can be used to monitor and normalize global transcriptome changes (van de Peppel et al., 2003). If upon Loess normalization the ratios of the external controls are zero then this means that there were no global transcriptome changes inadequately normalized by LOESS. Throughout the project the ratio of external controls after Loess normalization was monitored. The range and distribution of external control ratios in the mutants was not significantly different from those in the WTs. In addition, there were no mutant outliers where the average external control ratio was more than 1.7. None of the deletion mutants in this study therefore affected expression on such a global scale that this could not be accurately normalized by LOESS. This is likely due to the exclusive use of viable deletion mutants. For data generated using

temperature-sensitive mutants and for condition perturbations such as excessive stress, external control normalization is often required (van de Peppel et al., 2003).

Statistical Analysis of Expression Profiles

For each mutant the replicate hybridizations from two independent cultures were compared to the WT cultures grown on the same day to control for day-specific effects (described above) and to a pool of WT replicates collected throughout the project to generate the final mutant expression-profile (*p* value and fold-change in mutant versus average WT). For the MATalpha mutants grown on the Tecan plate shaker, the WT pool consisted of 200 MATalpha WTs (BY4742, ArrayExpress accession E-TABM-984, GEO accession GSE42217). For the MATa mutants grown on the Tecan plate shaker, the WT pool consisted of 20 WT-MATa replicates (BY4741, ArrayExpress accession E-MTAB-1351, GEO accession GSE42241). For mutants grown in Erlenmeyers, the MATalpha WT pool consisted of 200 WT-MATalpha replicates (ArrayExpress accession E-TABM-773, GEO accession GSE42215). For MATa mutants grown in Erlenmeyers, the WT pool consisted of 8 WT MATa replicates (ArrayExpress accession E-MTAB-1352, GEO accession GSE42240).

P values were obtained from the *limma* R package version 2.12.0 (Smyth et al., 2005) after Benjamini-Hochberg FDR correction. Genes were considered significantly changed when the fold-change (FC) was > 1.7 and the *p* value < 0.05. A list of genes that changes frequently irrespective of the targeted deletion or perturbation (WT variable genes) was constructed by using the 200 WT versus WT comparisons of the two large WT-MATalpha replicate pools (ArrayExpress accession E-TABM-773 and E-TABM-984). Each WT expression-profile was compared against all other WT expression-profiles through the common reference. After excluding the top 5% of expression-profile outliers, genes that change significantly (*limma*, FC > 1.7, *p* < 0.05) at least two times in any of the WTs is included in the WT variable gene list. This procedure was performed independently for both MATalpha WT pools and the resulting lists were merged to yield the WT variable gene list: *A11*, *A12*, *A14*, *A15_ALPHA*, *A15_BETA*, *ATP8*, *B1O3*, *B1O4*, *B1O5*, *BSC1*, *DDR2*, *FIT2*, *GLK1*, *GSY1*, *HSP12*, *HSP30*, *HSP42*, *HXK1*, *NCE103*, *OL11*, *PHO84*, *PRM7*, *SOL4*, *SPL2*, *SRO9*, *STP4*, *TPS2*, *TSL1*, *VAR1*, *VTC3*, *YDL038C*, *YDR170W-A*, *YDR210C-C*, *YIG1*, *YJR154W*, *YKR075C*, *YNL284C-A*, *YOR343W-B*, *yrO2*, *ZE01*, *AIM33*, *CTR1*, *GPD2*, *GPH1*, *PHO12*, *PKH2*, *RIF2*, *RTC3*, *RTC4*, *VTC1*, *YDL177C*, *YDR210W-B*, *YER053C-A*, *YFL002W-B*, *YMR046C*, *YPR158W-A*, *ZRT1*. *YDL196W* was added to this list as this gene showed significant upregulation in the majority of mutant profiles, likely due to a spurious mutation in the WT culture grown for the common reference RNA. All of these genes were excluded from further downstream analysis.

For determining a threshold to classify mutants with a significant effect on mRNA expression levels (responsive mutants), all deletion mutants and the WT cultures grown in parallel were ranked by the number of significantly changing genes. Less than 6% of the WTs had more than 3 genes changing and based on this the deletion mutants were classified into two groups: responding (> = 4 genes changing) and nonresponding (<4 genes changing). R (template) scripts for running the *limma* statistical analysis are available upon request.

During the course of the project it became clear that a significant fraction of mutants were similar to WT. From that stage on, in order to reduce costs, first only one of the two replicate cultures was processed. Based on this it was decided whether the second culture should be processed. Selection was based on the amount of significant changes (> = 8) in the analysis of the single culture. This entails that for some of the nonresponsive mutants there are four replicate measurements while for others there are only two. The average of all the responsive mutants is based on four measurements.

Correlation Calculation and Hierarchical Clustering

All correlations throughout the study are with a standard correlation distance. Hierarchical clustering (Figure 1) was by average linkage whereby *M* values of transcripts with insignificant changes (*p* > 0.05, FC < 1.7) were set to zero.

Functional Enrichment

Functional enrichment was by a hypergeometric test on Gene Ontology (GO) biological process, GO molecular function, KEGG pathway (Kanehisa et al., 2012) and TF binding (MacIsaac et al., 2006) (Figure 1, *p* < 0.01) or on GO biological process (Figures 4B, and 4D). GO annotations were obtained from SGD (Cherry et al., 2012). The background population was 6,182 and *p* values were Bonferroni corrected.

Protein Complexes and Pathways

Protein complexes (Figures 2C and 3E) were from the curated “consensus + GO” set describing 518 complexes (Benschop et al., 2010). Pathways (Figure 2C) were extracted from GO using the term “pathway” and merged with SGD biochemical pathways (Cherry et al., 2012). Correlations were calculated for all pairs within a complex or pathway profiled as a deletion mutant. For prediction of similar function (Figure 3E), true positive and true negative protein-protein interactions were calculated as described (Collins et al., 2007).

Compilation of GSTFs

The list of GSTFs (Table S1) was first compiled from previous studies (Badis et al., 2008; Harbison et al., 2004; Zhu et al., 2009). These genes were then scanned for common DNA binding domains using PFam (Punta et al., 2012). Other genes with the resulting DNA

binding domains (e.g., fork head, zinc cluster, zinc finger C2H2, myb domain, helix loop helix, GATA, bzip, HMG box, homeobox) were then searched for and included. Two inclusion criteria were then applied: the presence of a DNA binding domain and/or previous evidence for specific DNA binding. Twelve poorly characterized GSTFs did not fulfill the second criteria but were still included. Several well-studied genes with a DNA binding domain that are clearly not GSTFs (e.g., global TFs) were removed. The complete list contains 182 genes. Of these, 11 essential genes were not profiled.

GSTF Promoter Affinity

GSTF promoter affinity scores were calculated from in vitro protein binding microarray (PBM) data (Badis et al., 2008; Zhu et al., 2009). First, signal intensities and enrichment scores for GSTFs measured on multiple PBMs within a single data set were averaged. The affinity by which a GSTF binds to the promoter of a potential target gene was then estimated by adding up the signal intensities of all DNA 8-mer sequences within 600 bp upstream of the translation start site with an enrichment score ≥ 0.45 . The resulting promoter affinity profile of one GSTF to all possible promoters was Z-transformed.

Construction of the Genetic Perturbation Network

The GPN (Data S2) is constructed by taking the gene expression matrix of all responsive mutants (Figure 1) and converting this into a graph. Nodes in this graph represent deletion mutants (columns in the matrix) or transcripts changing in these deletion mutants (rows in the matrix). Edges are drawn for each significant transcript change in the individual deletion mutants and connect deletion mutant nodes with the transcript nodes that change in the corresponding deletion mutant. Activating edges (positive sign, arrowheads) are drawn for downregulated transcripts and indicate that under the standard growth condition, the gene corresponding to the deletion mutant has a positive effect on the expression level of a particular transcript. Inhibiting edges (negative sign, T-heads) are drawn for upregulated transcripts and indicate the opposite behavior.

Determination of the Node In- and Outdegree Distributions

Node in- (Figure S3A) and outdegree (Figure S3B) distributions were determined by constructing the GPN of all deletion mutants and all transcripts changing significantly at least once. The resulting network was loaded in Cytoscape (Smoot et al., 2011) version 2.8.3 and the in- and outdegree distributions were determined by running the “Analyze Network” plugin.

Construction of the Connectivity Matrix between Functional Categories

The connectivity matrix between functional categories (Figure S3C) was constructed by counting the number of edges between the different functional classes (Figure S1A) and dividing this by the theoretically maximum number of edges between the given functional categories. The theoretically maximum number of edges was calculated using $(N \times M) - N$ if N and M are from the same functional category and $N \times M$ otherwise. N and M are the number of deletion mutants in the functional category with outgoing and incoming edges respectively.

Filtering Potential Nested Effects within FFLs

The combinatorial input function on downstream Z target genes can be either additive or Boolean. Both are included in the full set. For additive input function an additional criteria should be fulfilled: that the changes in expression in the downstream Z nodes is greater upon deletion of upstream node X than upon deletion of Y. If this is not the case then a FFL is still possible, but only with a Boolean input function. FFL type frequency was tested on two FFL sets, the complete set and a set filtered for not fulfilling the criteria for an additive FFL (filtered for nested effects). Filtering was through a paired Mann-Whitney test for additional effects on gene expression. C1 and C3 FFLs that have a significant difference ($p < 0.05$, Benjamini-Hochberg corrected) in the mean expression change and show an increase of the absolute mean expression change in all shared Z nodes for X and Y deletion were kept.

SUPPLEMENTAL REFERENCES

- Ambroziak, J., and Henry, S.A. (1994). INO2 and INO4 gene products, positive regulators of phospholipid biosynthesis in *Saccharomyces cerevisiae*, form a complex that binds to the INO1 promoter. *J. Biol. Chem.* 269, 15344–15349.
- Andrews, B.J., and Herskowitz, I. (1989). The yeast SWI4 protein contains a motif present in developmental regulators and is part of a complex involved in cell-cycle-dependent transcription. *Nature* 342, 830–833.
- Apweiler, E., Sameith, K., Margaritis, T., Brabers, N., van de Pasch, L., Bakker, L.V., van Leenen, D., Holstege, F.C., and Kemmeren, P. (2012). Yeast glucose pathways converge on the transcriptional regulation of trehalose biosynthesis. *BMC Genomics* 13, 239.
- Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippe, K.H., Sherman, P.M., et al. (2011). NCBI GEO: archive for functional genomics data sets—10 years on. *Nucleic Acids Res.* 39 (Database issue), D1005–D1010.
- Byrd, C., Turner, G.C., and Varshavsky, A. (1998). The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor. *EMBO J.* 17, 269–277.
- Byrne, K.P., and Wolfe, K.H. (2005). The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res.* 15, 1456–1461.
- Chen, Y., Kamat, V., Dougherty, E.R., Bittner, M.L., Meltzer, P.S., and Trent, J.M. (2002). Ratio statistics of gene expression levels and applications to microarray data analysis. *Bioinformatics* 18, 1207–1215.

- Chodosh, L.A., Olesen, J., Hahn, S., Baldwin, A.S., Guarente, L., and Sharp, P.A. (1988). A yeast and a human CCAAT-binding protein have heterologous subunits that are functionally interchangeable. *Cell* 53, 25–35.
- Crespo, J.L., Powers, T., Fowler, B., and Hall, M.N. (2002). The TOR-controlled transcription activators GLN3, RTG1, and RTG3 are regulated in response to intracellular levels of glutamine. *Proc. Natl. Acad. Sci. USA* 99, 6784–6789.
- Daignan-Fornier, B., and Fink, G.R. (1992). Coregulation of purine and histidine biosynthesis by the transcriptional activators BAS1 and BAS2. *Proc. Natl. Acad. Sci. USA* 89, 6746–6750.
- de Boer, M., Nielsen, P.S., Bebelman, J.P., Heerikhuijen, H., Andersen, H.A., and Planta, R.J. (2000). Stp1p, Stp2p and Abf1p are involved in regulation of expression of the amino acid transporter gene BAP3 of *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 28, 974–981.
- Dohrmann, P.R., Butler, G., Tamai, K., Dorland, S., Greene, J.R., Thiele, D.J., and Stillman, D.J. (1992). Parallel pathways of gene regulation: homologous regulators SWI5 and ACE2 differentially control transcription of HO and chitinase. *Genes Dev.* 6, 93–104.
- Dolan, J.W., Kirkman, C., and Fields, S. (1989). The yeast STE12 protein binds to the DNA sequence mediating pheromone induction. *Proc. Natl. Acad. Sci. USA* 86, 5703–5707.
- Dubois, E., and Messenguy, F. (1991). In vitro studies of the binding of the ARGR proteins to the ARG5,6 promoter. *Mol. Cell. Biol.* 11, 2162–2168.
- Friden, P., and Schimmel, P. (1988). LEU3 of *Saccharomyces cerevisiae* activates multiple genes for branched-chain amino acid biosynthesis by binding to a common decanucleotide core sequence. *Mol. Cell. Biol.* 8, 2690–2697.
- Galeote, V.A., Alexandre, H., Bach, B., Delobel, P., Dequin, S., and Blondin, B. (2007). Sfl1p acts as an activator of the HSP30 gene in *Saccharomyces cerevisiae*. *Curr. Genet.* 52, 55–63.
- Gavrias, V., Andrianopoulos, A., Gimeno, C.J., and Timberlake, W.E. (1996). *Saccharomyces cerevisiae* TEC1 is required for pseudohyphal growth. *Mol. Microbiol.* 19, 1255–1263.
- Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. *Nature* 425, 737–741.
- Hanlon, S.E., Rizzo, J.M., Tatomer, D.C., Lieb, J.D., and Buck, M.J. (2011). The stress response factors Yap6, Cin5, Phd1, and Skn7 direct targeting of the conserved co-repressor Tup1-Ssn6 in *S. cerevisiae*. *PLoS ONE* 6, e19060.
- Hinnebusch, A.G., and Fink, G.R. (1983). Positive regulation in the general amino acid control of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 80, 5374–5378.
- Holmberg, S., and Schjerling, P. (1996). Cha4p of *Saccharomyces cerevisiae* activates transcription via serine/threonine response elements. *Genetics* 144, 467–478.
- Huang, M., Zhou, Z., and Elledge, S.J. (1998). The DNA replication and damage checkpoint pathways induce transcription by inhibition of the Crt1 repressor. *Cell* 94, 595–605.
- Jang, Y.K., Wang, L., and Sancar, G.B. (1999). RPH1 and GIS1 are damage-responsive repressors of PHR1. *Mol. Cell. Biol.* 19, 7630–7638.
- Jia, Y., Rothermel, B., Thornton, J., and Butow, R.A. (1997). A basic helix-loop-helix-leucine zipper transcription complex in yeast functions in a signaling pathway from mitochondria to the nucleus. *Mol. Cell. Biol.* 17, 1110–1117.
- Jorgensen, P., Nishikawa, J.L., Breitkreutz, B.-J., and Tyers, M. (2002). Systematic identification of pathways that couple cell growth and division in yeast. *Science* 297, 395–400.
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 40 (Database issue), D109–D114.
- Knight, S.A., Tamai, K.T., Kosman, D.J., and Thiele, D.J. (1994). Identification and analysis of a *Saccharomyces cerevisiae* copper homeostasis gene encoding a homeodomain protein. *Mol. Cell. Biol.* 14, 7792–7804.
- Krems, B., Charizanis, C., and Entian, K.D. (1996). The response regulator-like protein Pos9/Skn7 of *Saccharomyces cerevisiae* is involved in oxidative stress resistance. *Curr. Genet.* 29, 327–334.
- Kuchin, S., Vyas, V.K., and Carlson, M. (2002). Snf1 protein kinase and the repressors Nrg1 and Nrg2 regulate FLO11, haploid invasive growth, and diploid pseudohyphal differentiation. *Mol. Cell. Biol.* 22, 3994–4000.
- Kuge, S., and Jones, N. (1994). YAP1 dependent activation of TRX2 is essential for the response of *Saccharomyces cerevisiae* to oxidative stress by hydroperoxides. *EMBO J.* 13, 655–664.
- Laloux, I., Dubois, E., Dewerchin, M., and Jacobs, E. (1990). TEC1, a gene involved in the activation of Ty1 and Ty1-mediated gene expression in *Saccharomyces cerevisiae*: cloning and molecular analysis. *Mol. Cell. Biol.* 10, 3541–3550.
- Lamb, T.M., and Mitchell, A.P. (2003). The transcription factor Rim101p governs ion tolerance and cell differentiation by direct repression of the regulatory genes NRG1 and SMP1 in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 23, 677–686.
- Larochelle, M., Drouin, S., Robert, F., and Turcotte, B. (2006). Oxidative stress-activated zinc cluster protein Stb5 has dual activator/repressor functions required for pentose phosphate pathway regulation and NADPH production. *Mol. Cell. Biol.* 26, 6690–6701.
- Longtine, M.S., McKenzie, A., 3rd, Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippse, P., and Pringle, J.R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* 14, 953–961.
- Lowry, C.V., and Zitomer, R.S. (1984). Oxygen regulation of anaerobic and aerobic genes mediated by a common factor in yeast. *Proc. Natl. Acad. Sci. USA* 81, 6129–6133.
- Mannhaupt, G., Schnall, R., Karpov, V., Vetter, I., and Feldmann, H. (1999). Rpn4p acts as a transcription factor by binding to PACE, a nonamer box found upstream of 26S proteasomal and other genes in yeast. *FEBS Lett.* 450, 27–34.
- Marion, R.M., Regev, A., Segal, E., Barash, Y., Koller, D., Friedman, N., and O'Shea, E.K. (2004). Sfp1 is a stress- and nutrient-sensitive regulator of ribosomal protein gene expression. *Proc. Natl. Acad. Sci. USA* 101, 14315–14322.
- Martínez-Pastor, M.T., Marchler, G., Schüller, C., Marchler-Bauer, A., Ruis, H., and Estruch, F. (1996). The *Saccharomyces cerevisiae* zinc finger proteins Msn2p and Msn4p are required for transcriptional induction through the stress response element (STRE). *EMBO J.* 15, 2227–2235.

- McNabb, D.S., Xing, Y., and Guarente, L. (1995). Cloning of yeast HAP5: a novel subunit of a heterotrimeric complex required for CCAAT binding. *Genes Dev.* 9, 47–58.
- Mendizabal, I., Rios, G., Mulet, J.M., Serrano, R., and de Larrinoa, I.F. (1998). Yeast putative transcription factors involved in salt tolerance. *FEBS Lett.* 425, 323–328.
- Mitchell, A.P., and Magasanik, B. (1984). Regulation of glutamine-repressible gene products by the GLN3 function in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 4, 2758–2766.
- Nehlin, J.O., and Ronne, H. (1990). Yeast MIG1 repressor is related to the mammalian early growth response and Wilms' tumour finger proteins. *EMBO J.* 9, 2891–2898.
- Ozcan, S., Leong, T., and Johnston, M. (1996). Rgt1p of *Saccharomyces cerevisiae*, a key regulator of glucose-induced genes, is both an activator and a repressor of transcription. *Mol. Cell. Biol.* 16, 6419–6426.
- Park, S.H., Koh, S.S., Chun, J.H., Hwang, H.J., and Kang, H.S. (1999). Nrg1 is a transcriptional repressor for glucose repression of STA1 gene expression in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 19, 2044–2050.
- Parkinson, H., Sarkans, U., Kolesnikov, N., Abeygunawardena, N., Burdett, T., Dylag, M., Emam, I., Farne, A., Hastings, E., Holloway, E., et al. (2011). ArrayExpress update—an archive of microarray and high-throughput sequencing-based functional genomics experiments. *Nucleic Acids Res.* 39 (Database issue), D1002–D1004.
- Proft, M., and Struhl, K. (2002). Hog1 kinase converts the Sko1-Cyc8-Tup1 repressor complex into an activator that recruits SAGA and SWI/SNF in response to osmotic stress. *Mol. Cell* 9, 1307–1317.
- Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., et al. (2012). The Pfam protein families database. *Nucleic Acids Res.* 40 (Database issue), D290–D301.
- Robertson, L.S., and Fink, G.R. (1998). The three yeast A kinases have specific signaling functions in pseudohyphal growth. *Proc. Natl. Acad. Sci. USA* 95, 13783–13787.
- Rowen, D.W., Esiobu, N., and Magasanik, B. (1997). Role of GATA factor Nil2p in nitrogen regulation of gene expression in *Saccharomyces cerevisiae*. *J. Bacteriol.* 179, 3761–3766.
- Slattery, M.G., Liko, D., and Heideman, W. (2006). The function and properties of the Azf1 transcriptional regulator change with growth conditions in *Saccharomyces cerevisiae*. *Eukaryot. Cell* 5, 313–320.
- Smyth, G.K., Michaud, J., and Scott, H.S. (2005). Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* 21, 2067–2075.
- Stillman, D.J., Bankier, A.T., Seddon, A., Groenhout, E.G., and Nasmyth, K.A. (1988). Characterization of a transcription factor involved in mother cell specific transcription of the yeast HO gene. *EMBO J.* 7, 485–494.
- Tu, S., Bulloch, E.M.M., Yang, L., Ren, C., Huang, W.-C., Hsu, P.-H., Chen, C.-H., Liao, C.-L., Yu, H.-M., Lo, W.-S., et al. (2007). Identification of histone demethylases in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 282, 14262–14271.
- Verdière, J., Creusot, F., and Guérineau, M. (1985). Regulation of the expression of iso 2-cytochrome c gene in *S. cerevisiae*: cloning of the positive regulatory gene CYP1 and identification of the region of its target sequence on the structural gene CYP3. *Mol. Gen. Genet.* 199, 524–533.
- Wang, S.S., and Hopper, A.K. (1988). Isolation of a yeast gene involved in species-specific pre-tRNA processing. *Mol. Cell. Biol.* 8, 5140–5149.
- Ward, M.P., Gimeno, C.J., Fink, G.R., and Garrett, S. (1995). SOK2 may regulate cyclic AMP-dependent protein kinase-stimulated growth and pseudohyphal development by repressing transcription. *Mol. Cell. Biol.* 15, 6854–6863.
- Wu, C.-Y., Roje, S., Sandoval, F.J., Bird, A.J., Winge, D.R., and Eide, D.J. (2009). Repression of sulfate assimilation is an adaptive response of yeast to the oxidative stress of zinc deficiency. *J. Biol. Chem.* 284, 27544–27556.
- Xie, J., Pierce, M., Gailus-Durner, V., Wagner, M., Winter, E., and Vershon, A.K. (1999). Sum1 and Hst1 repress middle sporulation-specific gene expression during mitosis in *Saccharomyces cerevisiae*. *EMBO J.* 18, 6448–6454.
- Yamaguchi-Iwai, Y., Dancis, A., and Klausner, R.D. (1995). AFT1: a mediator of iron regulated transcriptional control in *Saccharomyces cerevisiae*. *EMBO J.* 14, 1231–1239.
- Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J., and Speed, T.P. (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res.* 30, e15.
- Zhao, H., and Eide, D.J. (1997). Zap1p, a metalloregulatory protein involved in zinc-responsive transcriptional regulation in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 17, 5044–5052.
- Zhu, G., Spellman, P.T., Volpe, T., Brown, P.O., Botstein, D., Davis, T.N., and Futcher, B. (2000). Two yeast forkhead genes regulate the cell cycle and pseudo-hyphal growth. *Nature* 406, 90–94.

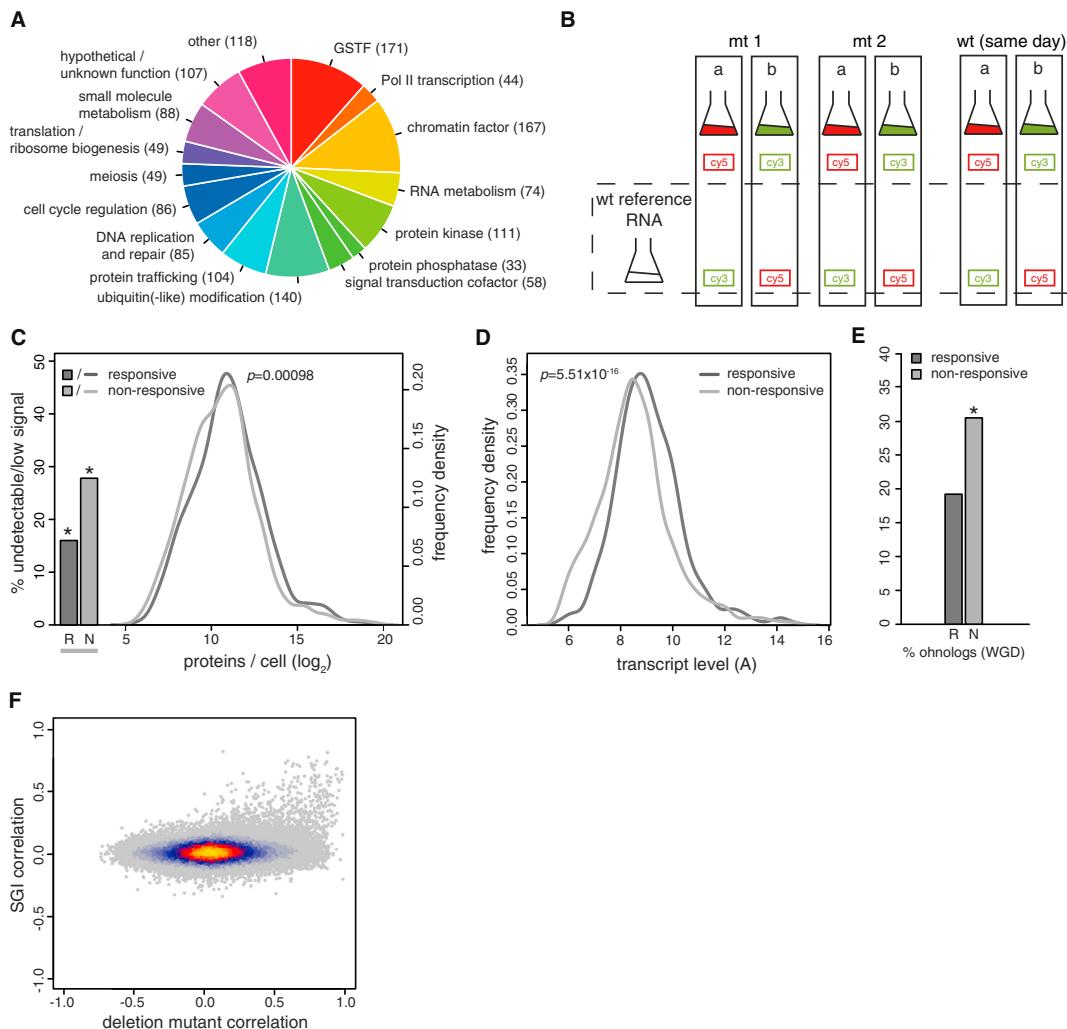


Figure S1. Study Design and General Properties of the Perturbation Data, Related to Figure 1

(A) Distribution of mutants over different functional classes.

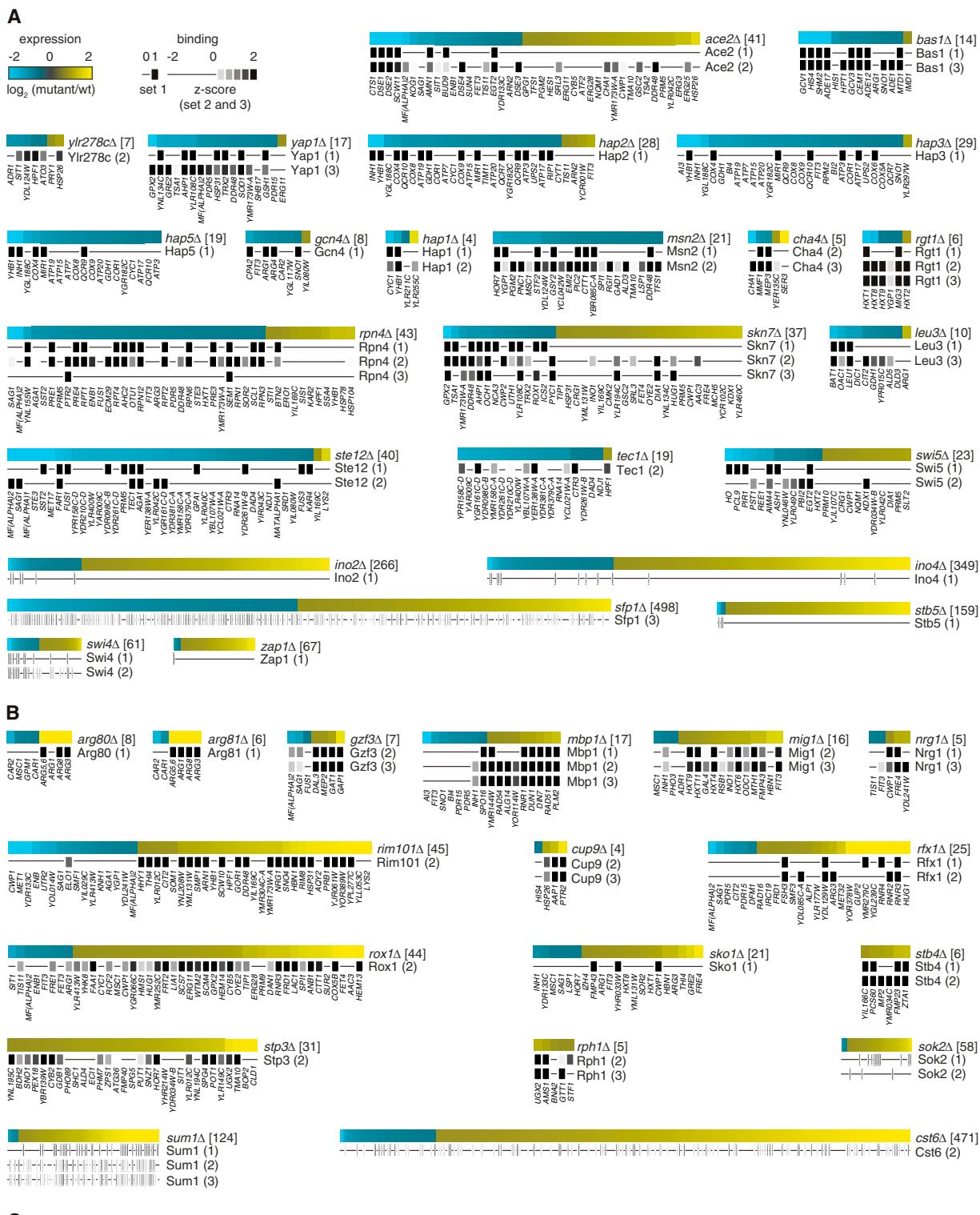
(B) Common reference experiment design. cRNA from replicate mutant cultures (a, b) was cohybridized in dye-swap with cRNA from a batch of common reference WT RNA. Same-day WT cultures were processed alongside mutants using the same common reference WT RNA. The same-day WT profiles formed the large set of WT transcriptomes against which each mutant was compared and also formed a control for day-specific effects.

(C) Percentage of undetectable/low signal proteins (bars, scale left) and the frequency density distribution (lines, scales bottom and right) of the (detectable) number of proteins per WT cell (Ghaemmaghami et al., 2003) for genes corresponding to responsive and nonresponsive mutants. The proportion of undetectable/low signal proteins is much higher for the nonresponsive than for the responsive mutants. Both are underrepresented compared to the average of the protein expression study (36%) that also included many dubious ORFs not profiled here. Asterisks indicate significance of underrepresentation as determined by a hypergeometric test for the responsive ($p < 2.2 \times 10^{-16}$) and nonresponsive mutants ($p = 2.84 \times 10^{-8}$). The p value in the panel is the significance of the difference in distribution of protein levels between responsive and nonresponsive mutants. This is calculated using a two-sided Mann-Whitney test.

(D) Frequency density distribution of WT mRNA levels (average of 200 WTs, A is the \log_2 fluorescent intensity) for genes corresponding to responsive and nonresponsive mutants. P value for the difference in distribution: as in C.

(E) Percentage of genes that have an ohnolog (close paralog) from the whole-genome duplication (WGD) (Byrne and Wolfe, 2005) for responsive and nonresponsive mutants. Asterisk indicates a significant overrepresentation for the nonresponsive mutants ($p < 2.2 \times 10^{-16}$, hypergeometric test). Close paralogs are not significantly differentially represented in the responsive mutants, the genome average being 18%.

(F) Density-colored scatterplot showing the correlation in the deletion signature for pairs of mutants (horizontal) versus the correlation in the SGI profiles (vertical) (Costanzo et al., 2010) for all pairs that exist in both data sets. Whereas there are many pairs that show correlations in both data sets (upper right quadrant), few conflict (lower left quadrant). Complementarity of the two data sets is indicated by pairs with correlating deletion signatures but no large correlation in their SGI profiles. Randomly selected pairs of genes aren't expected to have correlation in either data set on average, resulting in most of the data being centered around zero.



(legend on next page)

Figure S2. Gene-Specific Transcription Factor Classification, Related to Figure 4

Classification of Gene-Specific Transcription Factors (GSTFs) by comparing their deletion expression-profiles with DNA binding sets (1-3) ([Badis et al., 2008](#); [MacIsaac et al., 2006](#); [Zhu et al., 2009](#)).

(A) Activators, classified as such based on a significant overlap between the DNA binding data and the genes with increased expression upon deletion (Experimental Procedures). Square brackets: number of transcripts changing. As with other such figures throughout the study, for each deletion mutant all genes with robust changes ($p < 0.05$, FC > 1.7) are shown. GSTFs with > 50 expression changes are shown on a smaller scale. Rectangles below the profile indicate transcripts with promoter binding. Scales as indicated top left. Set 1 is binary. Sets 2 and 3 are plotted as promoter affinity z-scores.

(B) Repressors.

(C) Dual function.

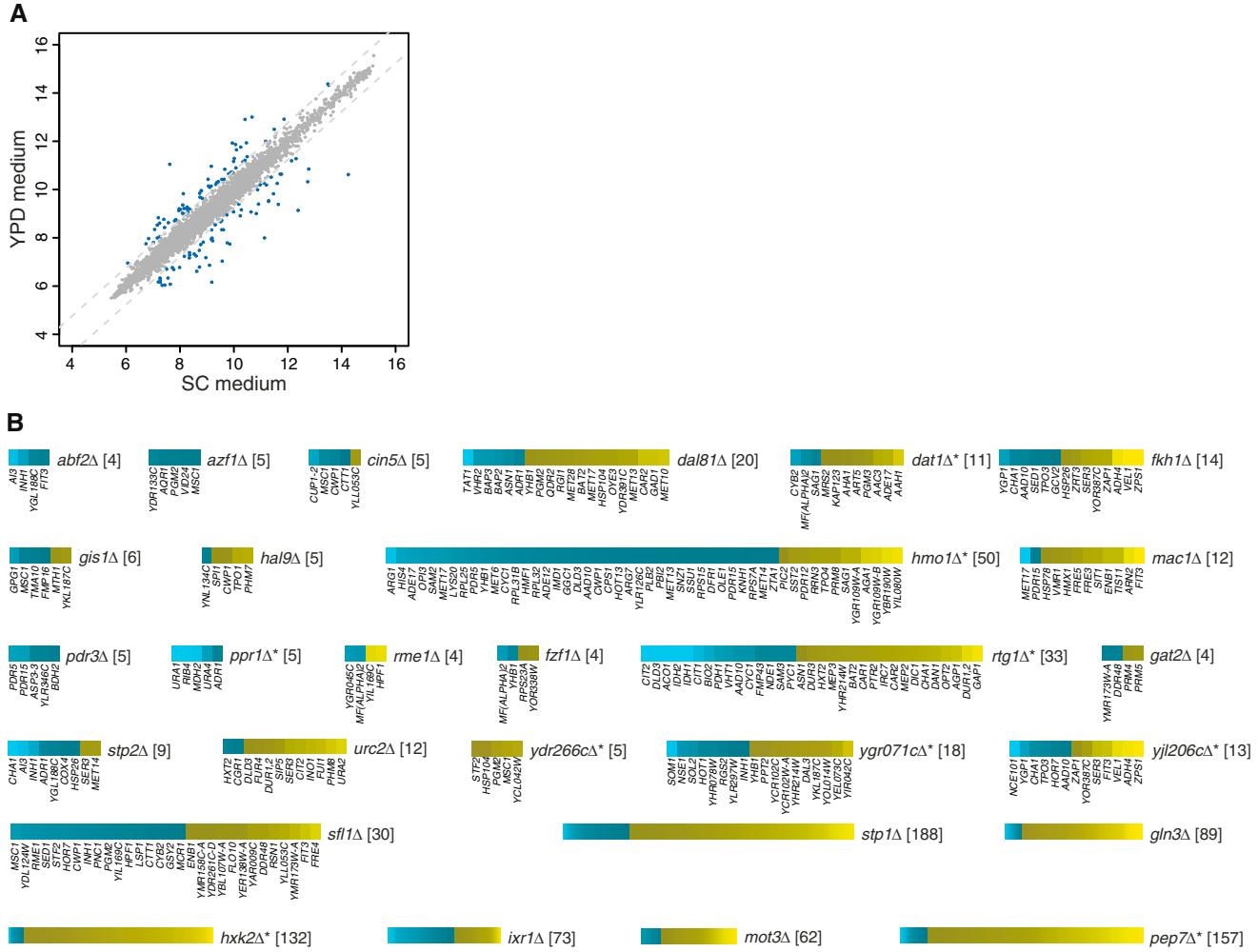


Figure S3. Growth Media Comparison and Unclassified GSTFs, Related to Figure 4

(A) Scatterplot of WT mRNA expression levels (\log_2 fluorescence intensity) in SC medium versus YPD medium. Genes with robust changes (FC 1.7, $p < 0.05$) between the two media are colored blue (128 in total). The dashed lines indicate 1.7 FC.

(B) GSTFs that cannot be classified based on the comparison of their deletion expression-profiles with DNA binding data sets (1-3), either due to no significant overlap or (*) because no binding data were available from the three sets. Color scales as in Figure S2.

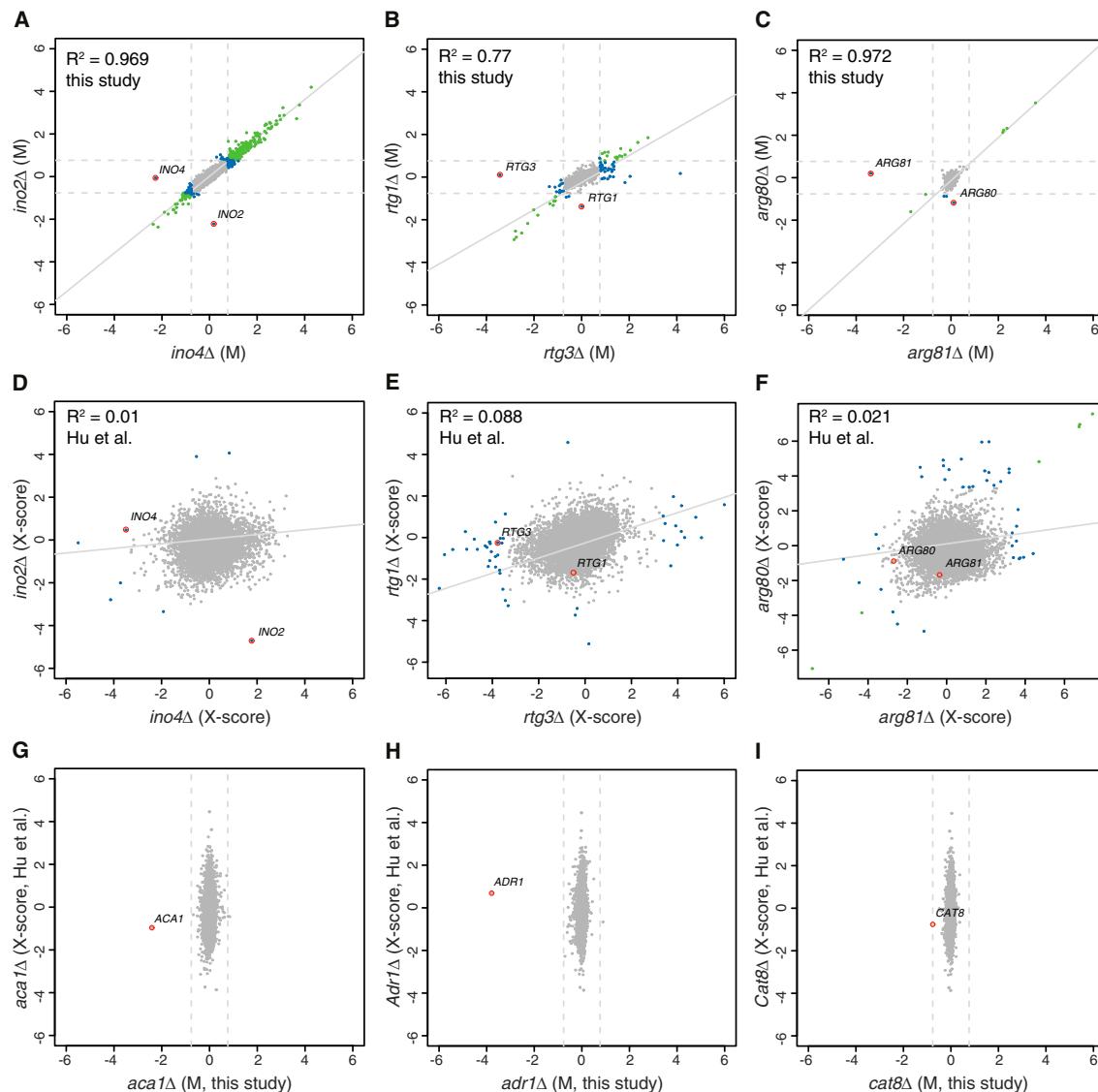


Figure S4. Comparison to Previously Published GSTF Deletion Data, Related to Figure 4

(A–F) Comparison to previously published GSTF deletion data: heterodimer GSTFs. There are many general and individual differences between the expression profiles reported here for GSTF deletion mutants compared to an earlier study (Hu et al., 2007). Heterodimer TFs form an excellent quality control for either data set. Assuming no involvement in other TF complexes, the deletion signatures of the two partners should correspond. This is clearly the case for the three scatterplots of reported fold-changes (M: \log_2 mt/WT) for heterodimer GSTFs from this study: (A) Ino2/Ino4, (B) Rtg1/Rtg3 and (C) Arg80/Arg81, selected based on the range (high to low) in number of reported fold-changes. The dashed lines indicate 1.7 fold-change (FC). Genes with robust changes (FC > 1.7, p < 0.05) in any single deletion mutant are colored blue. Genes with robust changes in both deletion mutants are colored green. The red circles indicate the deleted genes. The R^2 is based on all genes. (D–F) Scatterplots of the same heterodimers from (Hu et al., 2007). X-scores are the error model transformed fold-changes reported in that study, obtained from GEO (Barrett et al., 2011). Genes changing significantly (p < 0.001) in any one deletion mutant are blue. Genes changing significantly in both are green. This is the significance threshold applied in Hu et al. The red circle indicates the deleted gene. Note the lack of green genes, the lack of global correspondence (R^2) and the much larger degree of scatter (noise/measurement error) in all six deletion profiles from Hu et al., in comparison to the current study (E–G).

(G–I) Comparison to previously published GSTF deletion data: nonresponsive deletions. Another way of investigating differences between the two data sets is by comparison of GSTF deletions that are reported as nonresponsive here. Nonresponsive GSTF deletion mutants from this study were directly compared with Hu et al. by plotting the expression changes (M) reported here (horizontal) with the expression changes from Hu et al. (vertical) for (G) Aca1, (H) Adr1 and (I) Cat8. Dotted lines indicate 1.7 FC. The red circle indicates the deleted gene. This is the only gene for which it is 100% certain that it should show a change. In the profiles from this study it is shifted to the left. In Hu et al. there is little or no apparent change (even an apparent upregulation for *ADR1*), despite the large number of other apparent changes (vertical scatter). For this comparison the first three nonresponsive GSTFs were selected by alphabet.

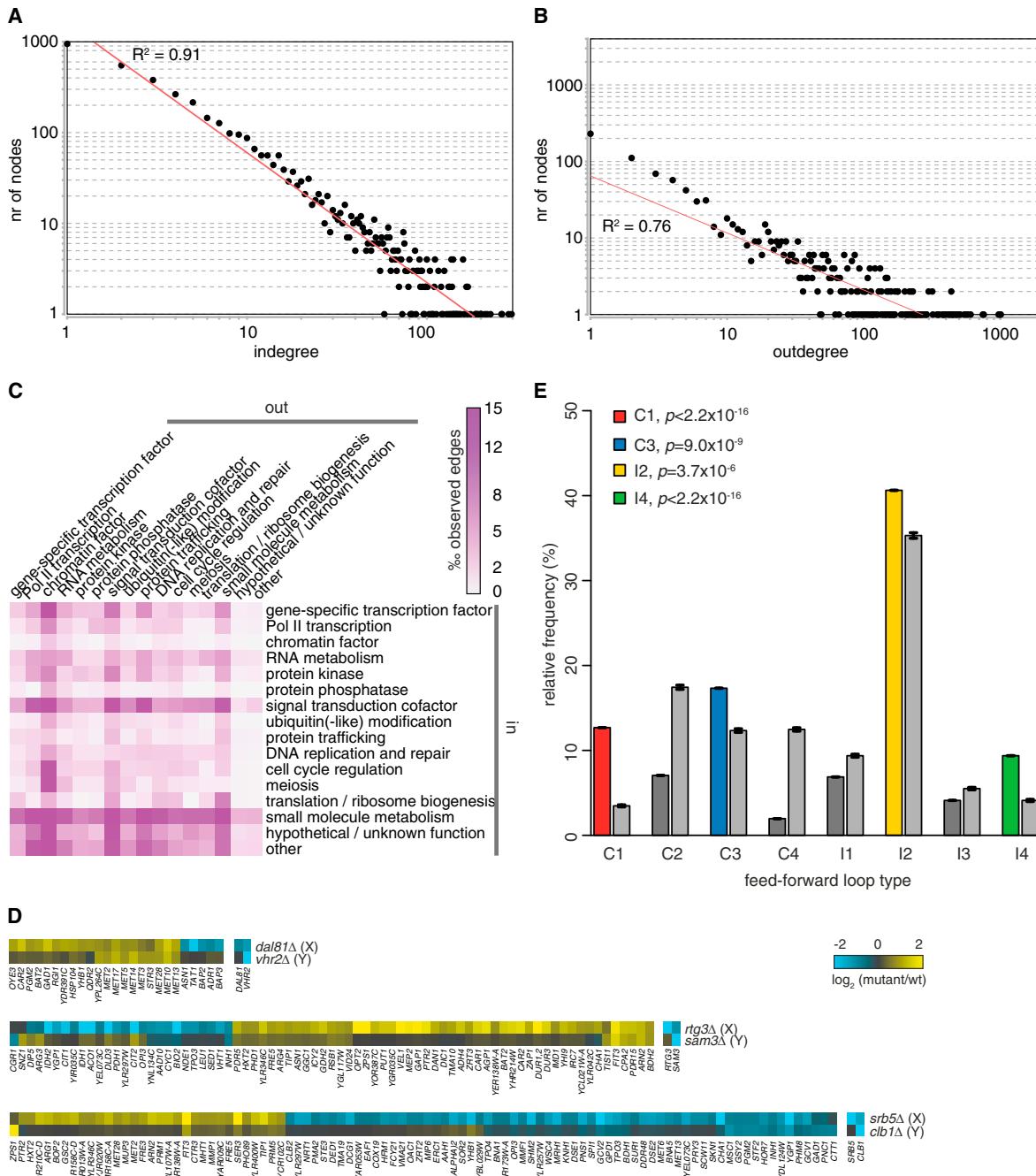


Figure S5. Characteristics of the Genetic Perturbation Network, Related to Figure 5

(A and B) Log-log scale plot of the node indegree (A) and node outdegree (B) distribution of the genetic perturbation network (GPN).

(C) Number of edges observed between genes from different classes, relative to the theoretical maximum of edges between these classes. Color scale as indicated (permillage).

(D) Nested effects. The changes observed upon deletion of the downstream gene (Y) are a subset of the changes observed upon deletion of the upstream gene (X), that also shows decrease in expression of the downstream gene (Y). Such nested effects indicate indirect effects. As discussed in the main text, few profiles are nested in their entirety.

(E) Relative frequency of FFL types in the GPN not filtered for nested effects. Error bars for the 10,000 permuted networks (gray) indicate two times the standard error of the mean (SEM), the 95% confidence interval. P values are derived from Z-score transformed counts and Bonferroni corrected.