

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

Patrick Kemmeren,<sup>1,3</sup> Katrin Sameith,<sup>1,3</sup> Loes A.L. van de Pasch,<sup>1,3</sup> Joris J. Benschop,<sup>1,3</sup> Tineke L. Lenstra,<sup>1,3</sup> Thanasis Margaritis,<sup>1,3</sup> Eoghan O'Duibhir,<sup>1</sup> Eva Apweiler,<sup>1</sup> Sake van Wageningen,<sup>1</sup> Cheuk W. Ko,<sup>1</sup> Sebastiaan van Heesch,<sup>1</sup> Mehdi M. Kashani,<sup>1</sup> Giannis Ampatzidis-Michailidis,<sup>1</sup> Mariel O. Brok,<sup>1</sup> Nathalie A.C.H. Brabers,<sup>1</sup> Anthony J. Miles,<sup>1</sup> Diane Bouwmeester,<sup>1</sup> Sander R. van Hooff,<sup>1</sup> Harm van Bakel,<sup>1</sup> Erik Sluiters,<sup>1</sup> Linda V. Bakker,<sup>1</sup> Berend Snel,<sup>2</sup> Philip Lijnzaad,<sup>1</sup> Dik van Leenen,<sup>1</sup> Marian J.A. Groot Koerkamp,<sup>1</sup> and Frank C.P. Holstege<sup>1,\*</sup>  
<sup>1</sup>Molecular Cancer Research, University Medical Centre Utrecht, Universiteitsweg 100, Utrecht 3584 CG, the Netherlands  
<sup>2</sup>Theoretical Biology and Bioinformatics, Department of Biology, Utrecht University, Utrecht 3584 CG, the Netherlands  
<sup>3</sup>Co-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

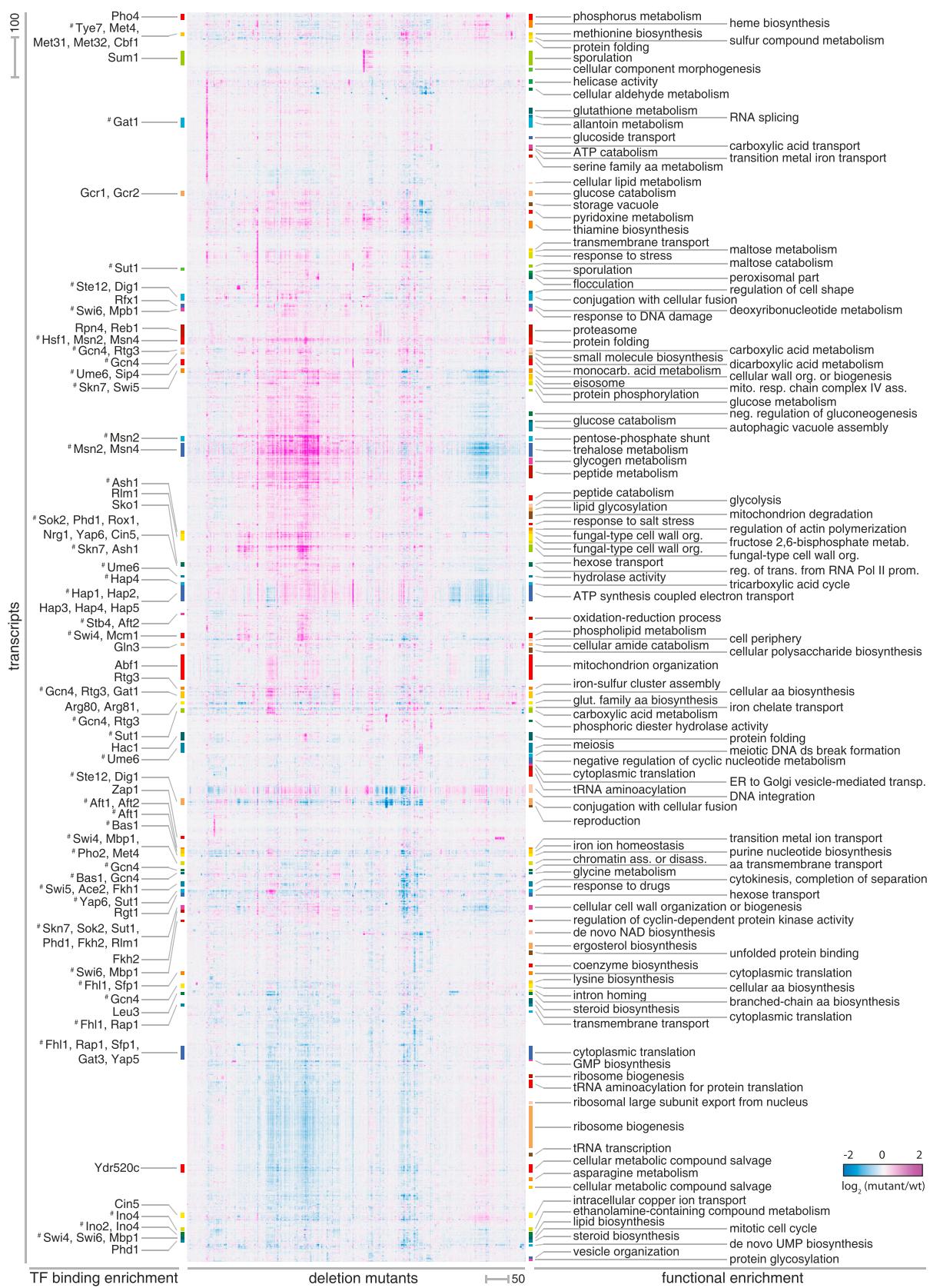
The data set consists of approximately 40 million expression measurements including WTs and replicates. Hierarchical clustering is presented in Figure 1. Although low-magnitude fold-changes [FCs] may have biological relevance, a stringent threshold ( $FC > 1.7$ ,  $p < 0.05$ ) was applied throughout 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;  $p = 2.6 \times$

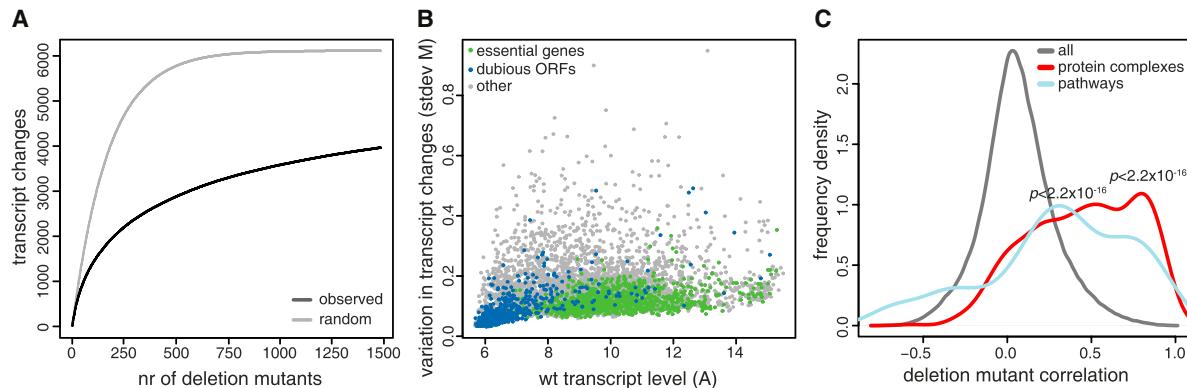
$10^{-9}$ ) and for genes essential for viability ( $p = 7.8 \times 10^{-31}$ ). Most dubious ORFs are lowly or not expressed in WT (Figure 2B). Combined with their low degree of change, this agrees with their classification as dubious, with most not likely to encode functional proteins (Fisk et al., 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.

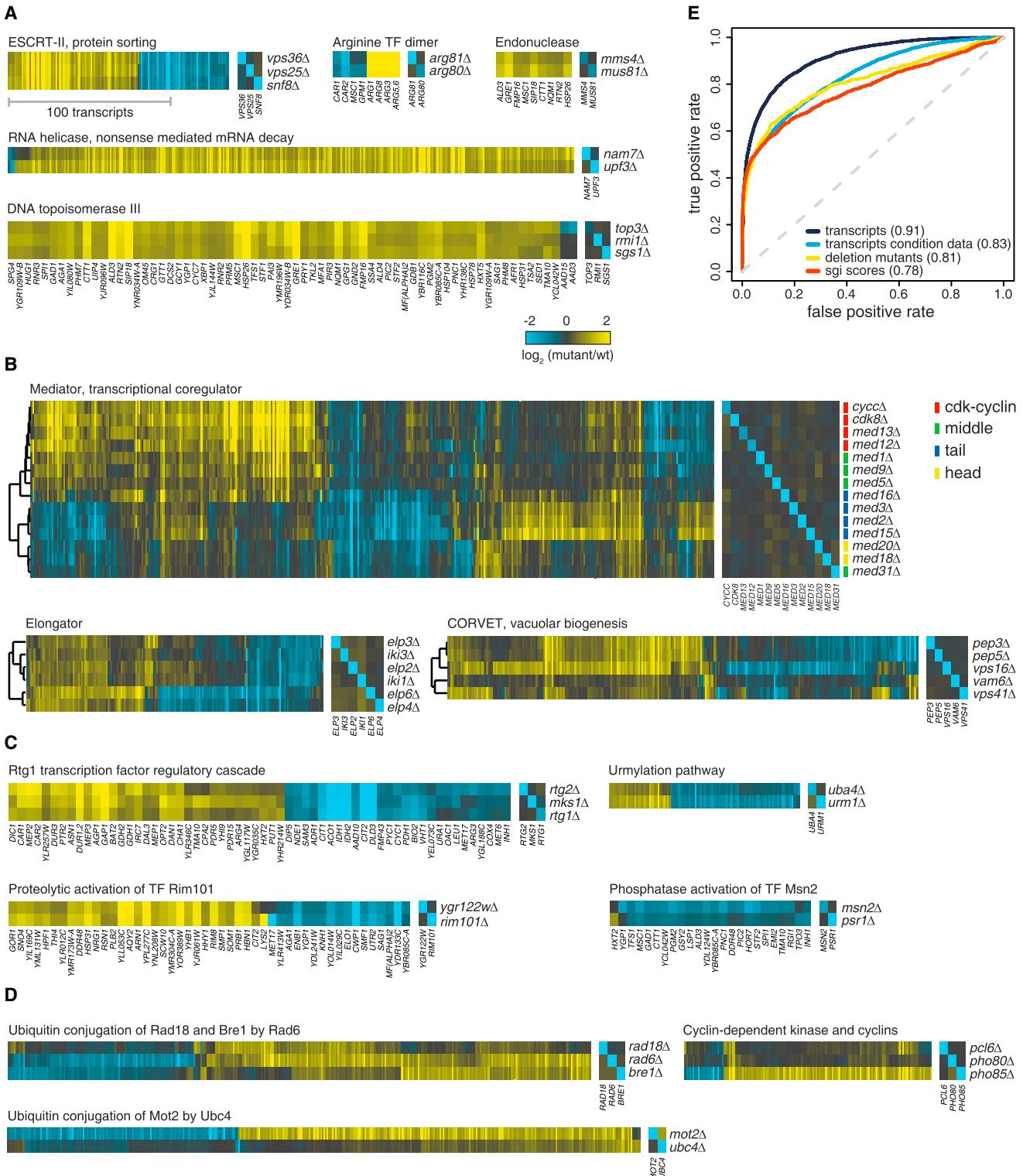
Pathway signature correlation is also significantly high, but lower than for protein complexes (Figure 2C). Pathways that are straight or cyclic, with no branching into distinct active arms, are expected to result in high correlations, similar to small protein complexes. Such apparently unbranched pathways 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

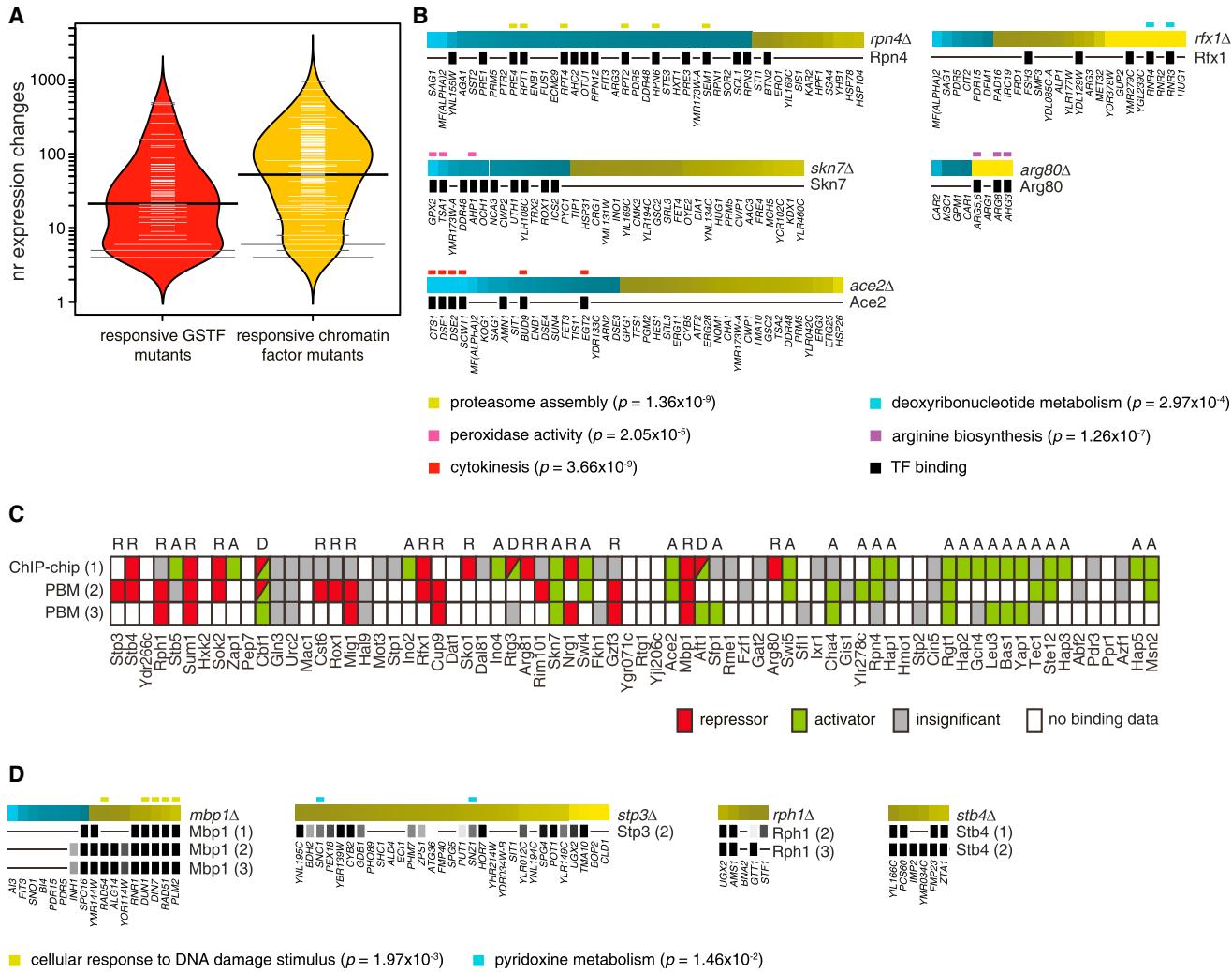
---

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.

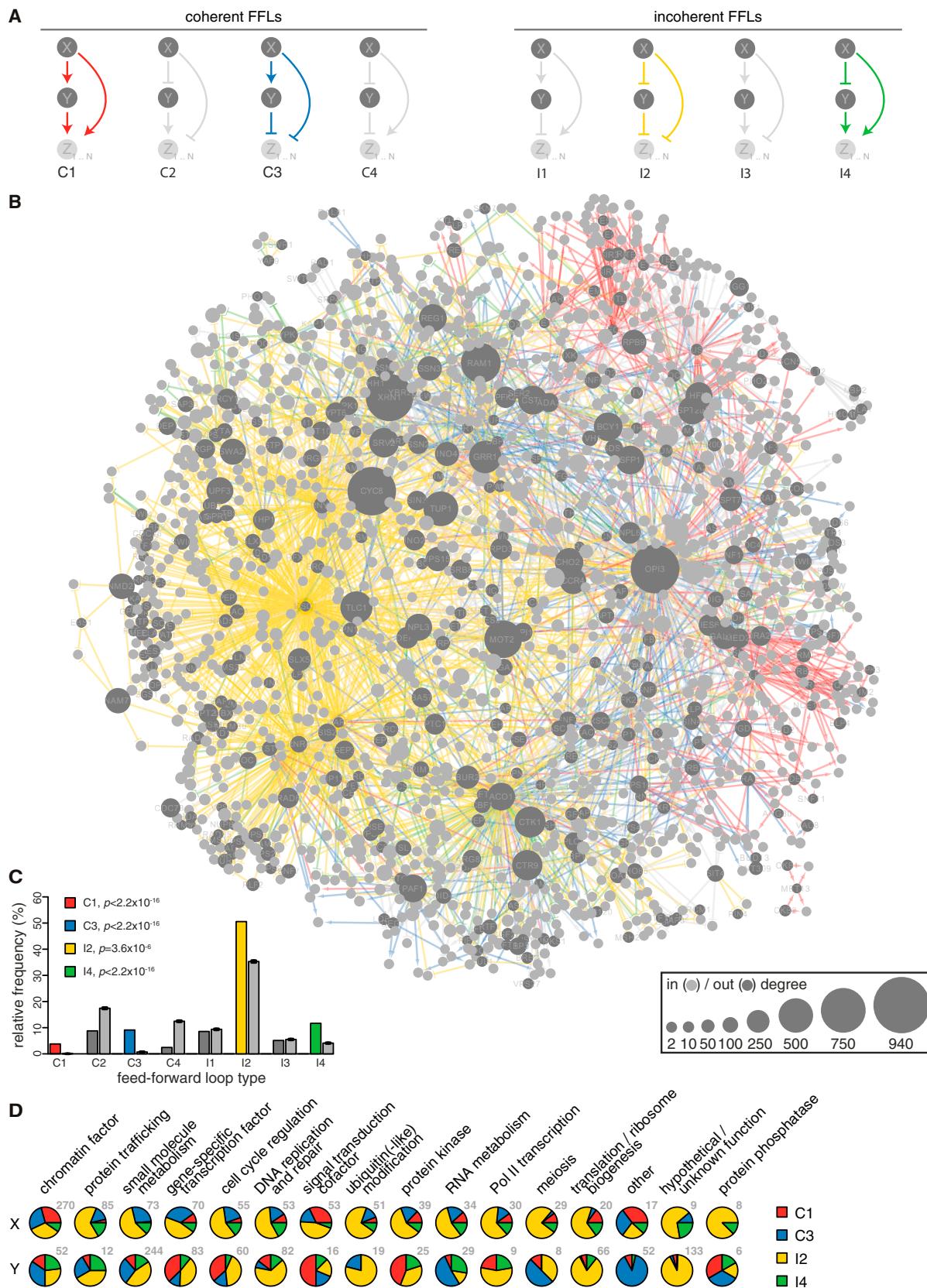
Recurrence of network parts may indicate advantageous regulatory properties and is also interesting from an evolutionary perspective (Sharan and Ideker, 2006). The occurrence of FFL types was therefore compared to 10,000 edge-permuted networks, each conservatively maintaining the network properties of 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.

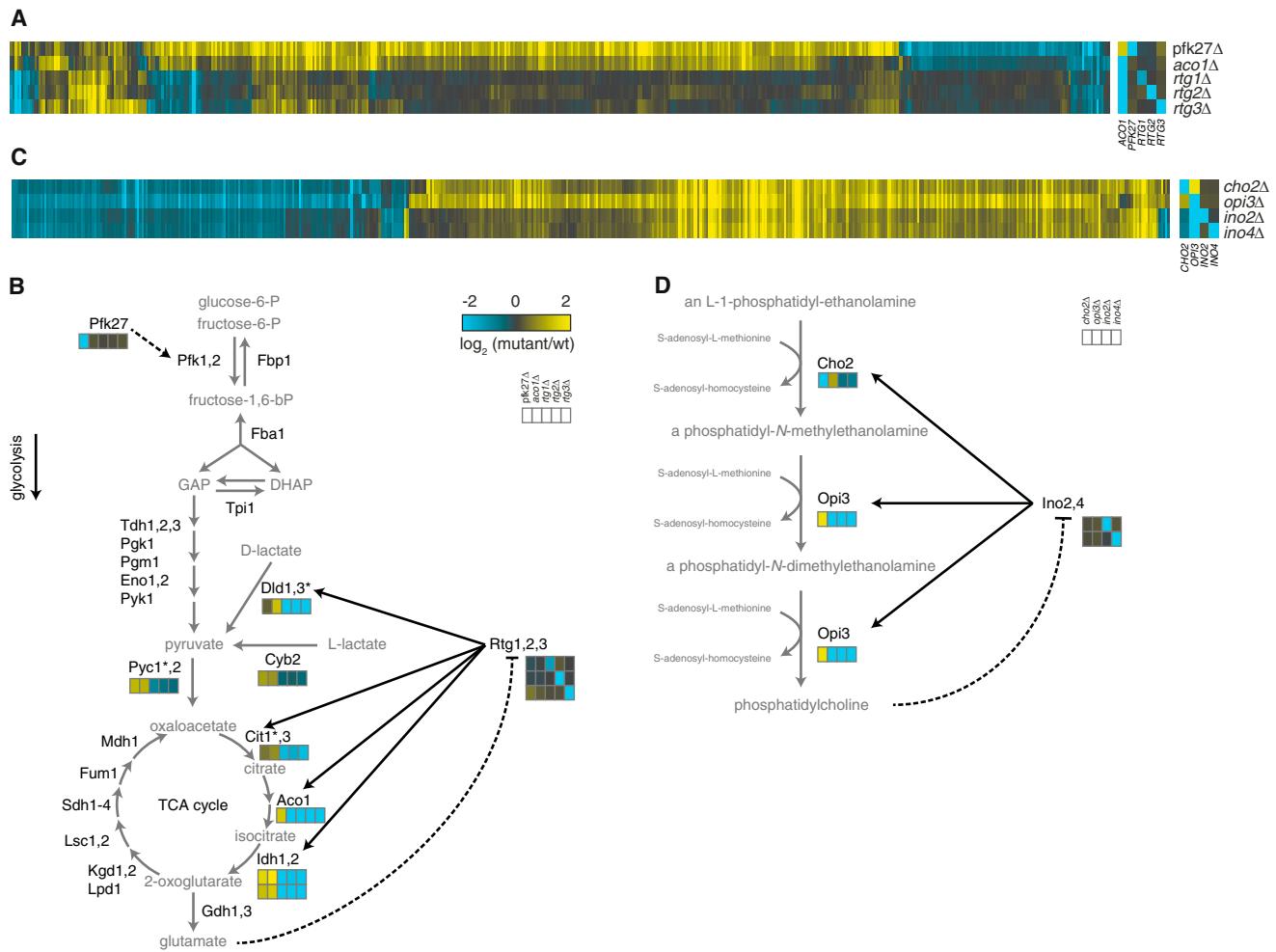
### Incoherent Type 2 FFLs Indicate Feedback in the Perturbation Network

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.