Molecular Systems Biology 5; Article number 276; doi:10.1038/msb.2009.33
Citation: *Molecular Systems Biology* 5:276
© 2009 EMBO and Macmillan Publishers Limited All rights reserved 1744-4292/09
www.molecularsystemsbiology.com

REPORT

Backup in gene regulatory networks explains differences between binding and knockout results

Anthony Gitter[1], Zehava Siegfried[2], Michael Klutstein[2], Oriol Fornes[3], Baldo Oliva[4], Itamar Simon[2] and Ziv Bar-Joseph[1,5,*]

[1] Computer Science Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA, [2] Department of Molecular Biology, Hebrew University Medical School, Jerusalem, Israel, [3] Department of Experimental Sciences and Health, Municipal Institute for Medical Research (IMIM-Hospital del Mar), Barcelona, Catalonia, Spain, [4] Department of Experimental Sciences and Health, Pompeu Fabra University, Barcelona, Catalonia, Spain and [5] Machine Learning Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA
* Corresponding author. Computer Science Department, School of Computer Science, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, USA. Tel.: +1 412 268 8595; Fax: +1 412 268 3431; E-mail: zivbj@cs.cmu.edu

Received 28.11.08; accepted 29.4.09

The complementarity of gene expression and protein–DNA interaction data led to several successful models of biological systems. However, recent studies in multiple species raise doubts about the relationship between these two datasets. These studies show that the overwhelming majority of genes bound by a particular transcription factor (TF) are not affected when that factor is knocked out. Here, we show that this surprising result can be partially explained by considering the broader cellular context in which TFs operate. Factors whose functions are not backed up by redundant paralogs show a fourfold increase in the agreement between their bound targets and the expression levels of those targets. In addition, we show that incorporating protein interaction networks provides physical explanations for knockout effects. New double knockout experiments support our conclusions. Our results highlight the robustness provided by redundant TFs and indicate that in the context of diverse cellular systems, binding is still largely functional.

Molecular Systems Biology **5**: 276; published online 16 June 2009; doi:10.1038/msb.2009.33
Subject Categories: simulation and data analysis; chromatin and transcription
Keywords: backup mechanisms; paralogs; protein interactions

Introduction

Many successful studies in systems biology focus on integrating complementary datasets to model systems in the cell. Several computational methods have been developed and applied to combine mRNA expression data and protein–DNA interaction data (using DNA-binding motifs, ChIP-chip experiments, or both) (Lee *et al*, 2002; Liao *et al*, 2003; Beer and Tavazoie, 2004; Yeang *et al*, 2005; Ernst *et al*, 2007). These methods assume that transcript levels are largely driven by binding of transcription factors (TFs) to DNA leading to either expression or repression of the bound genes. Indeed, by some estimates close to 60% of binding sites are actively driving expression of their bound genes (Gao *et al*, 2004).

This assumption was recently challenged by several studies that compared the set of genes bound by a TF with the set of genes affected when that factor is knocked out or knocked down. One of the earliest reports of this phenomenon involved the yeast cell cycle (Horak *et al*, 2002). Using ChIP-chip experiments, researchers looked at the set of genes bound by 11 TFs and concluded that complementary knockout experiments did not affect the same set of genes. In mouse, it was reported that only 11% of those genes that were differentially expressed after glucocorticoid dexamethasone injection were also bound by the glucocorticoid receptor (Phuc Le *et al*, 2005). An estrogen-response study in human reported that 6% of E_2-induced genes were bound by ERα, and 13% of ERα-bound genes were regulated by E_2 (Kwon *et al*, 2007). A human study in which p63 was depleted led to similar conclusions (Yang *et al*, 2006).

Although the above experiments looked at only one, or few, TFs, a recent study in yeast examined the overlap for the entire set of TFs and surprisingly concluded that the overlap was even smaller than the overlaps reported above for individual factors. In a comprehensive analysis of the agreement between binding and knockout experiments, 269 budding yeast TFs were knocked out one at a time (Hu *et al*, 2007), and the differentially expressed gene targets were compared with the protein–DNA binding data generated previously for 188 of those TFs (Harbison *et al*, 2004). It was determined that only 3% of bound genes were affected by the knockout, and similarly only 3% of affected genes were bound by the corresponding TF. Although this low overlap is statistically significant, the percentage is still very low. A possible explanation for this small overlap, which was examined by Hu *et al*, is that indirect regulatory interactions can explain at least one side of this overlap (why genes affected by a knockout are not directly bound by that TF). However, their analysis of these indirect binding effects, in which pathways of protein–DNA binding interactions were allowed as supporting evidence for the knockout effects, resulted in negligible improvements to the overlap and its significance.

To further examine these findings and to determine whether the expression and TF-gene binding interaction datasets are indeed complementary, we undertook a systems approach by studying the dependence of their agreement on the TFs' homology relationships and on the protein interaction network context of the TF. As we show, both play a major role in the low overlap. Accounting for these contexts increases both the percentage overlap and its significance, indicating that the difference may be explained by backup mechanisms used when cells lose specific TFs.

Results and discussion

P-value threshold analysis

In both the binding and knockout studies, a *P*-value threshold was used to identify significant genes. We examined the sensitivity of the overlap to these *P*-value thresholds by testing *P*-value cutoffs from $10^{-0.05}$ to 10^{-10} (see Materials and methods). For each *P*-value, we calculated the overlap and its significance using the hypergeometric distribution (Figure 1). The range of *P*-value thresholds from 0.008 to 0.001 yielded the most significant overlaps. Thus, we looked more closely at *P*-values of 0.001 and 0.005, which have been used in the past for binding and expression data (Lee *et al*, 2002; Bar-Joseph *et al*, 2003). The *P*-value threshold of 0.005 generally yields slightly better overlap than the threshold of 0.001, and all overlap values reported hereafter are based on a *P*-value threshold of 0.005 unless otherwise noted.

Each of the two papers used a different method to compute the *P*-values, which may contribute to their disagreement. To test whether this influenced the results, we computed the overlap between the two datasets based on rankings rather than *P*-value cutoffs. We sorted the TF-gene interactions by *P*-value in both datasets and selected the first k (where k ranges from 1 to 1000) from each dataset. The overlap significance peaks when the top 56 interactions per TF are taken to be significant (*P*-value of 10^{-52}, Supplementary Figure 1). This significance value is much worse than that of the threshold-based method (*P*-value 10^{-124}), suggesting that the rank-based method introduces substantial noise because high-ranking interactions may still have insignificant *P*-values for some factors. We also tried several other methods for selecting lists of bound and affected genes but these did not improve the overlap (Supplementary Tables I–III).

The above computations highlight the importance of the results presented by Hu *et al*, indicating that they cannot be explained by issues related to the analysis of the data but are rather likely to represent specific biological phenomena.

Cleaning the data

To lessen the extent to which experimental and biological noise affected the disagreement between the knockout and binding data, we cleaned the datasets in several ways (Supplementary Tables IV–VI; Supplementary Figures 2 and 3).

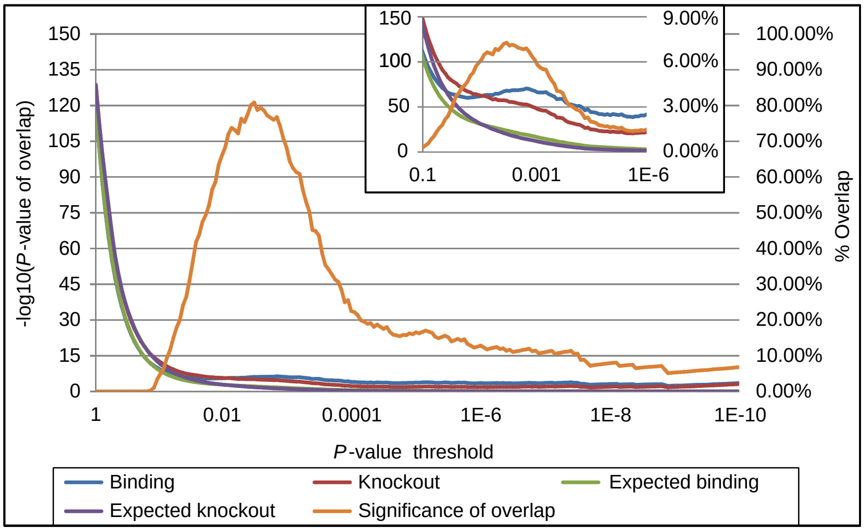

Figure 1 Change in overlap for a range of *P*-value thresholds. The figure plots the overlap as the percentage of the binding interactions and knockout effects (right *y* axis) compared with the expected binding and knockout overlaps. The figure is overlaid with the significance of the overlap calculated using the hypergeometric distribution (left *y* axis). Note the significance peak between *P*-values of 0.001 and 0.005 (inset). Source data is available for this figure at www.nature.com/msb.

We first removed genes that were affected by the knockout of a large number of TFs (see Materials and methods). We termed these 'general KO genes' because they are likely responding to the general stress of the knockout experiments rather than the specific TF deletions, and thus are not expected to be bound by the deleted TFs. In addition, we restricted the set of TF-binding targets to those with sequence motifs conserved in two other species (Harbison *et al*, 2004).

After these cleanup steps, the agreement between the two datasets increases to 6.7% of binding data and 4.5% of knockout data (*P*-value of 10^{-133} versus the original *P*-value of 10^{-114}). As above, we examined different *P*-value cutoffs for the cleaned dataset and found that the results did not improve when using the stricter threshold of 0.001. Similarly, knowledge regarding the function of the TFs did not improve the overlap. Specifically, we found that TFs that are expected to be active only in YPD had the same average overlap as the entire set of TFs (Supplementary Table VII).

Redundancy explains binding interactions absent from the knockout data

We next tested whether redundancy can help explain the small overlap observed. Following Kafri *et al* (2005) we used BLASTP to identify gene pairs with varying levels of homology. We divided the set of TFs into four groups: those with a paralogous TF with an *E*-value of E-20 or less, between E-20 and E-10, between E-10 and E-3 and those with no homolog at E-3 or less (see Materials and methods). TFs with the most similar paralogs had no overlap between their binding and knockout data. In contrast, those with the least similar paralogs had an overlap of more than 12%, nearly twofolds higher than the average overlap. The other groups followed a similar trend in which the overlap increased as the similarity to the closest paralog decreased (Figure 2).

To further test our finding that redundancy impacts the expression outcome we used Pfam, which focuses on the binding domain only, as a measure of similarity and obtained similar division into four groups. As with the BLASTP value, for groups with similar paralogs, the overlap was lower than for those with more distant homologs (4 versus 10%, Supplementary Figures 4 and 5).

Another component that may impact how well one TF can compensate for the loss of another TF is shared protein–protein interactions (PPIs) (Reguly *et al*, 2006). We divided each of the homology groups defined above based on the percentage of protein interaction partners the TF shares with another TF in that homology group (see Materials and methods). Similar to the trend we saw using sequence homology, within each group the overlap decreases as the percentage of shared PPIs increases (Figure 2; Table I). For TFs with the least similar homologs and the fewest shared interactions, we observed an overlap greater than 13%.

For all BLASTP *E*-value thresholds, TFs that shared a larger portion of PPI with their paralog had lower binding overlap. This indicates that putative paralogs with many common PPI are better able to compensate for the deletion effects (Table I).

Protein interaction networks provide physical support for knockout effects

To help explain the other direction (why genes affected by a knockout are not bound by the TF), we used interaction networks. Recent studies have shown that knockout effects can be mediated by PPI networks as well (Yeang *et al*, 2005; Workman *et al*, 2006). Thus, we constructed a network that includes both PPI and protein–DNA edges (Supplementary information). We considered a gene affected by the knockout to be explained by the network if (1) the TF directly binds the gene or (2) there is a path leading from the TF to another TF that directly binds the gene. For the indirect result we vary the maximum path length (number of edges from the initial TF to the last TF). As can be seen in Figure 3, using a path length of 2 leads to an overlap of 22% while significantly increasing the *P*-value of the overlap (from 10^{-133} to 10^{-211}). Path lengths greater than 2 increased the percentage of the overlap but reduced the *P*-value indicating that we are likely overfitting the data. Randomization tests and further analysis using different sets of PPI data confirmed the significance of the increase in overlap due to the PPI network (Supplementary Tables VIII and IX; Supplementary Figures 6 and 7).

As a further test, we repeated the PPI analysis using data from expression and binding experiments in human cells studying the TF p63 (Yang *et al*, 2006). A genome-wide TF-gene binding dataset is not available for human TFs so we used an analysis (Xie *et al*, 2005) to construct an approximate binding dataset for 71 TFs. As with yeast, the overlap greatly increased when using a network with a path length of 2 (*P*-values 10^{-15} and 10^{-5} compared with 10^{-12} and 10^{-2} for the original data, Supplementary Figure 8). Randomization tests of the human PPI network also indicated that the improvement is significant (Supplementary Table X).

Experimental validation

To further validate our results regarding the backup mechanisms used in regulatory networks, we collected expression data from three double knockout experiments involving pairs of factors we predicted could compensate for the loss of each other (Fkh1-Fkh2, Yhp1-Yox1, Ace2-Swi5, all from the E-20 set, Supplementary Table XI). We also carried out new experiments for an additional pair (Pdr1-Pdr3, also in the E-20 set). As predicted, when the paralogous partner is not present to compensate for the effect of a single knockout, the overlap of the knockout and binding data increases significantly. For example, the overlap between genes affected by the single knockout of Yhp1 and Yox1, two cell cycle TFs, and the genes bound by these factors is 0 and 1%, respectively (both are not significant). In contrast, the overlap for the double knockout and the binding targets of Yhp1 and Yox1 is 8 and 9%, respectively (*P*-values of 10^{-4} and $10^{-7.5}$). Similar results were obtained for the other two double knockouts we collected (Supplementary Tables XII). For our Pdr1-Pdr3 double knockout experiment, we again observed a large increase in the percentage of overlap for Pdr1 compared with the single knockout experiment. The overlap increased from 1% (not significant) to 19% (*P*-value of 10^{-5}). For Pdr3 we saw a large increase in binding percentage (from 0 to 4%) though this overlap is still not significant (Supplementary Table XII). Thus,

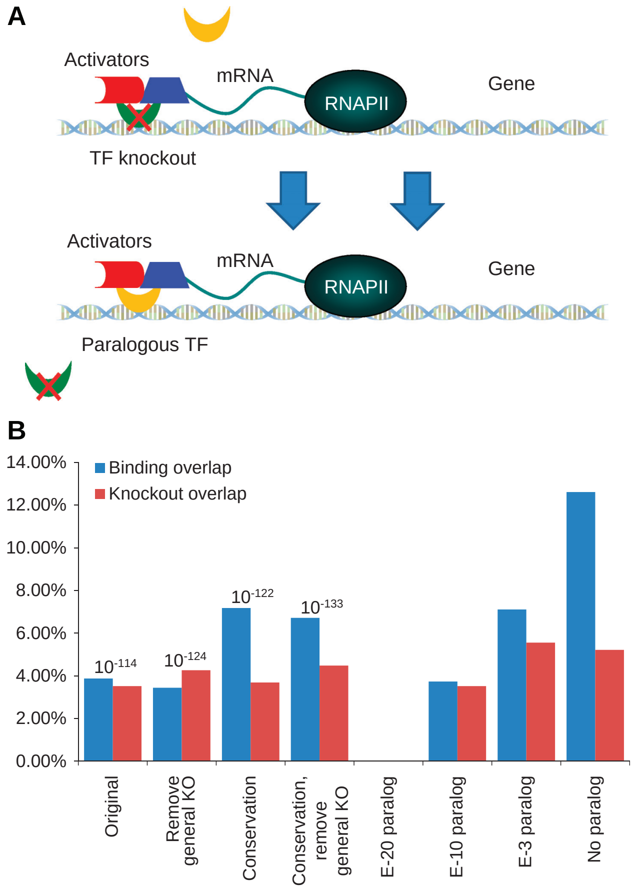

Figure 2 Improved overlap between binding and knockout experiments. (**A**) A schematic view of our analysis. Both sequence homology and shared interactions may lead to one TF compensating for another. Here, the yellow TF can replace the green TF when it is knocked out and is able to recruit the transcription machinery leading to only small overlap between binding and knockout results. (**B**) The binding and knockout overlap for various subsets of the data. The *P*-value of the overlap is given above the columns, which indicate percentage overlap for the whole network analyses. Cleaning the data by removing genes reacting non-specifically to the stress of a knockout ('general KO genes') and interactions not supported by sequence conservation improves the overlap and its significance. In addition, TFs without redundant paralogs have greater overlap. Source data is available for this figure at www.nature.com/msb.

these experiments support our claim of backup provided by these pairs of factors and can also provide clues to the mechanisms used as we discuss below.

Mechanisms leading to TF redundancy

A subset of the homologous TFs we identified bind to an overlapping group of targets, and thus it is not surprising that knocking out one of them has small effect on the expression of its targets. One such example is the two homologous TFs involved in methionine metabolism, Met31 and Met32 (Blaiseau *et al*, 1997). These TFs have a large overlapping set of target genes (>60%), and neither has any target genes that are differentially expressed after deletion. Another example is the two forkhead TFs, Fkh1 and Fkh2. These only bind a partially overlapping set of target genes. However, it has been shown (Hollenhorst *et al*, 2001) that the binding of Fkh1 to Fkh2 targets is enhanced in the absence of Fkh2 and vice versa, suggesting that a compensation can occur beyond the common targets as predicted by our findings.

This type of compensation may happen due to competition between the two TFs that is resolved in the absence of one of them. Another possibility is that the activity of one TF is enhanced in the absence of its homolog due to a feedback mechanism between the two TFs (Kafri *et al*, 2005). To check this idea, we looked at the expression levels of the TFs believed to be compensating for the knockout (most similar based on BLASTP). As expected, we have not found any example in which the expression level of the homologous TF was significantly decreased (Supplementary Table XIII). However, a significant increase was observed in only a few cases. Thus, it appears that these changes are mainly driven by post-transcriptional events, perhaps by the protein interaction networks mentioned above.

One of the first studies looking at the overlap between knockout and binding (Horak *et al*, 2002) hypothesized that redundancy may be part of the disagreement. In that paper, the authors conclude that 'in some cases targets are not significantly affected, presumably because of transcriptional redundancy'. However, it was hard to substantiate this claim without a comprehensive knockout and expression data. The

Table I Analysis of overlap based on paralogs and shared PPIs

BLASTP E-value	Shared PPI	Binding overlap (%)	Knockout overlap (%)
E-3 < E-value ≤ E1	PPI < 20%	13.37	5.37
E-3 < E-value ≤ E1	PPI ≥ 20%	3.30	2.13
E-10 < E-value ≤ E-3	PPI < 20%	7.51	5.85
E-10 < E-value ≤ E-3	PPI ≥ 20%	0.00	0.00
E-20 < E-value ≤ E-10	PPI < 20%	4.20	3.00
E-20 < E-value ≤ E-10	PPI ≥ 20%	2.77	7.48
E-value < E-20	PPI < 20%	0.00	0.00
E-value < E-20	PPI ≥ 20%	0.00	0.00

Transcription factors were divided into four groups based on their most similar TF homolog as determined by the BLASTP *E*-values. These sets were further divided based on the percentage of PPI a TF shared with its paralog. TFs with a putative paralog that share at least 20% PPI are more likely to be redundant and thus exhibit lower overlap.

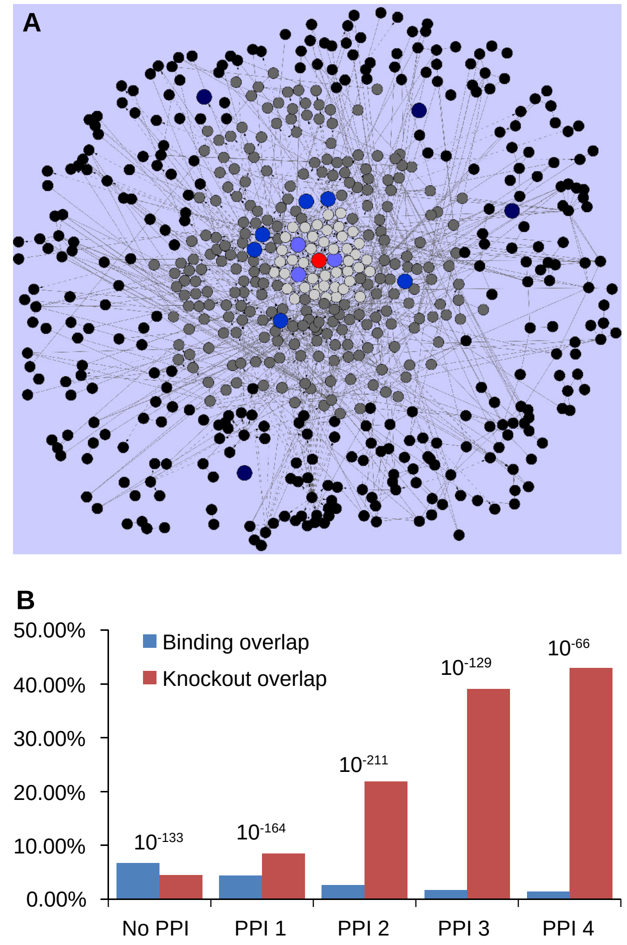

Figure 3 Influence of physical interaction networks. TFs that do not directly bind a gene can exert influence through pathways of PPI and protein–DNA interactions. (**A**) A network consisting of YHR206W (red), its knockout targets (shades of blue), and 20% of all other yeast genes selected at random (shades of gray). Genes are arranged around YHR206W according to the shortest number of interaction edges needed to reach them. The black and dark blue nodes correspond to genes that are three or more interactions away, the medium gray and medium blue genes are two interactions away, and the light gray and light blue genes are a single interaction from YHR206W. In all, 85% of YHR206W's knockout-affected genes are either directly bound by YHR206W or another TF that can be reached through paths of length 1 or 2. (**B**) As longer paths in the network are examined, a much higher percentage of the knockout-affected genes are connected to the deleted TF. The *P*-value of the overlap is given above the columns, which indicate percentage overlap. Source data is available for this figure at www.nature.com/msb.

recent availability of such data allowed us to show that redundancy (both in sequence and in interactions) indeed plays a major role in the disagreement between the two types of data. Our results suggest that paralogs can compensate for the loss of TFs providing backup mechanisms and robustness for eukaryotic cells.

Materials and methods

Naming conventions and gene synonyms

Many of the overlap calculations are sensitive to the manner in which yeast gene names are mapped from their standard name to their systematic name (also known as the ORF name) and vice versa. Although Harbison *et al* (2004) provides a list of ORF names for the standard names used in their experiments, the gene name mapping is constantly evolving, and their list was out-of-date at the time the knockout experiments were run. Therefore, we relied on the SGD (http://www.yeastgenome.org) gene name mapping with the manual addition of 13 retired mappings that appeared in older datasets we used. Any TF targets that could not be mapped to an SGD ORF name were ignored, as were ORFs on the mitochondrial chromosome or the 2-μm plasmid.

Overlap calculation

For a given TF t, we define the set of genes significantly bound by t to be G_B and the set of genes significantly affected by the knockout of t as G_K. The binding overlap B and knockout overlap K are calculated in the following manner:

$$B = \frac{|G_B \cap G_K|}{|G_B|}$$

$$K = \frac{|G_B \cap G_K|}{|G_K|}$$

We use the hypergeometric distribution, also known as the one-tailed version of Fisher's exact test, to calculate a P-value for the overlap of the binding and knockout targets:

$$P\text{-value} = \sum_{o=|G_B \cap G_K|}^{\min(|G_B|,|G_K|)} \left(\frac{\binom{|G_K|}{o} * \binom{|G_A| - |G_K|}{|G_B| - o}}{\binom{|G_A|}{|G_B|}} \right)$$

where $\binom{n}{k}$ is the choose function of n and k, G_A is the set of all possible genes targets in the binding or knockout datasets, and o is the size of the overlap. When calculating the P-value for the entire network of TFs or a subset of TFs sharing some property, we replace G_B in the above equations by I_B, the set of TF-gene binding interactions and similarly replace G_K with I_K, the set of TF-gene knockout effects.

Varying *P*-value thresholds

Data for Figure 1 were generated by considering the set of P-value thresholds defined by:

$$\bigcup_{n=1}^{200} 10^{(-n*0.05)}$$

At each threshold, the interactions in the binding and knockout datasets with significance less than or equal to the threshold were obtained, and the overlap and its significance were calculated. For each threshold, the expected binding and knockout overlaps were calculated using the formulas:

$$E(\text{binding overlap}) = \frac{\sum_{t \in F} \frac{G_B(th,t)*G_K(th,t)}{|G_A|}}{\sum_{t \in F} G_B(th,\, t)}$$

$$E(\text{knockout overlap}) = \frac{\sum_{t \in F} \frac{G_B(th,t)*G_K(th,t)}{|G_A|}}{\sum_{t \in F} G_K(th,\, t)}$$

where th is the threshold, t is a TF in the set of all TFs F, G_A is the set of all genes in the binding and knockout datasets, $G_B(th,t)$ is the number of genes bound by t at threshold th, and $G_K(th, t)$ is the number of genes affected by the knockout of t at threshold th.

Data cleaning

To purge the knockout data of instances in which a gene was differentially expressed due to non-specific effects instead of targeted regulatory mechanisms, we removed genes that were affected by the knockout of a large number of TFs. Specifically, we eliminated 161 'general KO genes' that were differentially expressed in 20 or more TF knockout experiments at a *P*-value of 0.001, which we consider to be genes affected by the stress of the deletions. Of the 14 427 knockout interactions at this *P*-value, 4920 were removed. As any TF-gene binding interaction involving one of the removed general KO genes is guaranteed to no longer have a corresponding knockout effect, we removed these same genes from the binding dataset.

The presence of a sequence motif can provide additional evidence that a protein–DNA binding interaction is functional, especially if that motif is conserved in other species. Therefore, to reduce noise we used Harbison *et al*'s alternate version of the binding dataset in which binding interactions not supported by a motif that is conserved in at least two other yeast species have been removed (Harbison *et al*, 2004). Many of the original 203 TFs do not have a known or conserved motif and were removed from subsequent analysis. Of the 102 remaining TFs, 97 were also knocked out by Hu *et al* (2007). This version of the binding data contains interactions observed in both YPD and non-YPD conditions.

Paralogs

To obtain a list of putative paralogs, the FASTA sequences for the 188 TFs present in both the original binding dataset and the knockout dataset were obtained from SGD (http://www.yeastgenome.org). Next, the NCBI netblast program (http://www.ncbi.nlm.nih.gov/blast/download.shtml) was run on the sequences with the parameters '-p blastp -d refseq_protein -b 200000 -u 'saccharomyces cerevisiae'[Organism]' to use BLASTP (Altschul *et al*, 1997) to search the RefSeq (http://www.ncbi.nlm.nih.gov/RefSeq) database for up to 200 000 sequences from *Saccharomyces cerevisiae* proteins. The default values were used for all other netblast parameters. The netblast results were filtered so that only TF–TF pairs between unique TFs remained, in which a TF was considered to be any of the 284 factors present in the either the binding or the knockout dataset. We post-processed the pairs of putative paralogs to separate them into distinct subsets so that pairs belonging to a set defined by a particular threshold (e.g. E-3) did not also appear in the set of pairs defined by a stricter threshold (e.g. E-10). Any TF that was not a member of the E-20, E-10 or E-3 paralog set was placed in the set of TFs without a paralog. The complete assignment of TFs to paralog sets can be found in Supplementary Table XIV (Excel file).

For the Pfam-based putative paralogs, we assigned one or more Pfam domains to a TF by aligning each single sequence with a domain from the Pfam database (a library of HMMs, version 22.0) (Finn *et al*, 2006) using the HMMER software package (2.3.2 release) (Durbin *et al*, 1998). We only considered those alignments involving Pfam domains classified as DNA-binding domains. Moreover, Pfam domains were assigned to one of the three different *E*-value thresholds, E-8, E-3 and E-1, yielding three sets of putative paralogs and a fourth set of TFs without a paralog. As with the BLASTP-based paralogs, we removed paralogs that already belonged to a stricter set from the less strict sets so that the subsets contain distinct TFs.

Protein interactions

For a pair of TFs *T1* and *T2*, we define the percentage of shared PPI to be

$$S(T1,\,T2) = \min\left(\frac{|P_{T1} \cap P_{T2}|}{|P_{T1}|}, \frac{|P_{T1} \cap P_{T2}|}{|P_{T2}|}\right)$$

where S is the percentage of shared PPI and P_T is the set of PPI partners for TF *T*. The min(*) function guarantees the resulting percentage of *T1*'s protein interaction partners are shared by *T2* and vice versa. For each TF that did not have a putative paralog, we calculated shared PPI using approximate paralogs, which are TF–TF pairs with a BLASTP *E*-value greater than E-3 but less than 10.

For the PPI network analysis, we used two different sets of reported PPIs. The first was a literature-curated PPI dataset (Reguly *et al*, 2006) and the second is the BioGRID dataset (version 2.0.48) (Stark *et al*, 2006) that includes data from high throughput interaction studies. For BioGRID, we removed all genetic interactions as well as those inferred from co-localization. In the main text, we report the results for the literature-curated dataset. Results for the BioGRID dataset are reported in the Supplementary information.

RNA extraction and labeling for expression profiling

The double KO strain YYA100 (*pdr1*Δ::*KanMx6*, *pdr3*Δ::*His3Mx6*) was kindly provided by Florian Zwolanek (Schuller *et al*, 2007). Cultures were grown to OD600 1.0–1.5 and total RNA was extracted using MasterPure Yeast RNA Purification Kit (Epicentre Biotechnologies). cDNA synthesis and labeling was performed as in http://www.md.huji.ac.il/units/tzabam/microarray/Labeling.htm. In brief, cDNA was generated with oligo-dt primers (2 μg) (Amersham Biosciences) from total RNA (20 μg) using the reverse transcription enzyme superscript II (Invitrogen). The reverse transcription reaction was carried out at 42°C for 2 h with aminoallyl-dUTP. Removal of unincorporated aa-dUTP and free amines was carried out using Microcon YM-30 (Millipore) filters according to the manufacturer's recommendations. Coupling of aminoallyl labeled cDNA to Cye dye esters was performed in 0.1 M sodium carbonate buffer (pH 8.6) for 1 h at room temperature. Removal of free dyes was accomplished with Qiagen PCR purification columns (Qiagen Ltd). The labeling was then quantified using a ND-1000 spectrophotometer (Nanodrop Ltd). Equal amounts of both samples were mixed and concentrated by speed vacuum for 1 h.

Microarray hybridization

Double spotted microarrays containing 6240 Yeast ORFs printed as cDNA (+ controls, total 6.4 K spots), manufactured by the Genomics Centre, University of Toronto, were pre-hybridized in 5x SSC, 0.1% SDS, 1% BSA for 45 min. The probes were resuspended in hybridization buffer (25% formamide, 5x SSC, 0.1% SDS, 0.4 μg/μl Yeast tRNA) and applied to the slides. Hybridization was carried out overnight at 42°C in a hybridization chamber (Corning). Arrays were scanned using GenePix 4000B scanner (Axon Instruments) with settings adjusted to obtain a similar green and red overall intensity. The resulting images were analyzed using GenePix pro 4.0 (Axon Instruments). The experiment was done in duplicates using dye swapping. Microarray data have been deposited at the ArrayExpress database under accession number E-MEXP-2150.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

We thank Karl Kuchler for the Pdr1/3 double KO strain. We also thank Jason Ernst for helpful discussions about this work. This work was supported in part by NIH grant 1RO1 GM085022 and NSF CAREER award 0448453 to ZBJ. In addition, this material is based on work supported under a National Science Foundation Graduate Research Fellowship to AG.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl Acids Res* **25:** 3389–3402

Bar-Joseph Z, Gerber GK, Lee TI, Rinaldi NJ, Yoo JY, Robert F, Gordon DB, Fraenkel E, Jaakkola TS, Young RA, Gifford DK (2003) Computational discovery of gene modules and regulatory networks. *Nat Biotech* **21:** 1337–1342

Beer MA, Tavazoie S (2004) Predicting gene expression from sequence. *Cell* **117:** 185–198

Blaiseau PL, Isnard AD, Surdin-Kerjan Y, Thomas D (1997) Met31p and Met32p, two related zinc finger proteins, are involved in transcriptional regulation of yeast sulfur amino acid metabolism. *Mol Cell Biol* **17:** 3640–3648

Durbin R, Eddy S, Krogh A, Mitchison G (1998) *Biological Sequence Analysis, Probabilistic Models of Proteins and Nucleic Acids*. Cambridge, UK: Cambridge University Press

Ernst J, Vainas O, Harbison CT, Simon I, Bar-Joseph Z (2007) Reconstructing dynamic regulatory maps. *Mol Syst Biol* **3:**74

Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer ELL, Bateman A (2006) Pfam: clans, web tools and services. *Nucl Acids Res* **34:** D247–D251

Gao F, Foat B, Bussemaker H (2004) Defining transcriptional networks through integrative modeling of mRNA expression and transcription factor binding data. *BMC Bioinformatics* **5:** 31

Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne J-B, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KT, Lander ES, Gifford DK, Fraenkel E, Young RA (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature* **431:** 99–104

Hollenhorst PC, Pietz G, Fox CA (2001) Mechanisms controlling differential promoter-occupancy by the yeast forkhead proteins Fkh1p and Fkh2p: implications for regulating the cell cycle and differentiation. *Genes Dev* **15:** 2445–2456

Horak CE, Luscombe NM, Qian J, Bertone P, Piccirrillo S, Gerstein M, Snyder M (2002) Complex transcriptional circuitry at the G1/S transition in Saccharomyces cerevisiae. *Genes Dev* **16:** 3017–3033

Hu Z, Killion PJ, Iyer VR (2007) Genetic reconstruction of a functional transcriptional regulatory network. *Nat Genet* **39:** 683–687

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

Kwon Y-S, Garcia-Bassets I, Hutt KR, Cheng CS, Jin M, Liu D, Benner C, Wang D, Ye Z, Bibikova M, Fan J-B, Duan L, Glass CK, Rosenfeld MG, Fu X-D (2007) Sensitive ChIP-DSL technology reveals an extensive estrogen receptor alpha-binding program on human gene promoters. *Proc Natl Acad Sci* **104:** 4852–4857

Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, Zeitlinger J, Jennings EG, Murray HL, Gordon DB, Ren B, Wyrick JJ, Tagne J-B, Volkert TL, Fraenkel E, Gifford DK *et al* (2002) Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science* **298:** 799–804

Liao JC, Boscolo R, Yang Y-L, Tran LM, Sabatti C, Roychowdhury VP (2003) Network component analysis: reconstruction of regulatory signals in biological systems. *Proc Natl Acad Sci USA* **100:** 15522–15527

Phuc Le P, Friedman JR, Schug J, Brestelli JE, Parker JB, Bochkis IM, Kaestner KH (2005) Glucocorticoid receptor-dependent gene regulatory networks. *PLoS Genet* **1:** e16

Reguly T, Breitkreutz A, Boucher L, Breitkreutz B-J, Hon G, Myers C, Parsons A, Friesen H, Oughtred R, Tong A, Stark C, Ho Y, Botstein D, Andrews B, Boone C, Troyanskya O, Ideker T, Dolinski K, Batada N, Tyers M (2006) Comprehensive curation and analysis of global interaction networks in Saccharomyces cerevisiae. *J Biol* **5:** 11

Schuller C, Mamnun YM, Wolfger H, Rockwell N, Thorner J, Kuchler K (2007) Membrane-active compounds activate the transcription factors Pdr1 and Pdr3 connecting pleiotropic drug resistance and membrane lipid homeostasis in Saccharomyces cerevisiae. *Mol Biol Cell* **18:** 4932–4944

Stark C, Breitkreutz B-J, Reguly T, Boucher L, Breitkreutz A, Tyers M (2006) BioGRID: a general repository for interaction datasets. *Nucl Acids Res* **34:** D535–D539

Workman CT, Mak HC, McCuine S, Tagne J-B, Agarwal M, Ozier O, Begley TJ, Samson LD, Ideker T (2006) A systems approach to mapping DNA damage response pathways. *Science* **312:** 1054–1059

Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, Lander ES, Kellis M (2005) Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. *Nature* **434:** 338–345

Yang A, Zhu Z, Kapranov P, McKeon F, Church GM, Gingeras TR, Struhl K (2006) Relationships between p63 binding, DNA sequence, transcription activity, and biological function in human cells. *Mol Cell* **24:** 593–602

Yeang C-H, Mak HC, McCuine S, Workman C, Jaakkola T, Ideker T (2005) Validation and refinement of gene-regulatory pathways on a network of physical interactions. *Genome Biol* **6:** R62