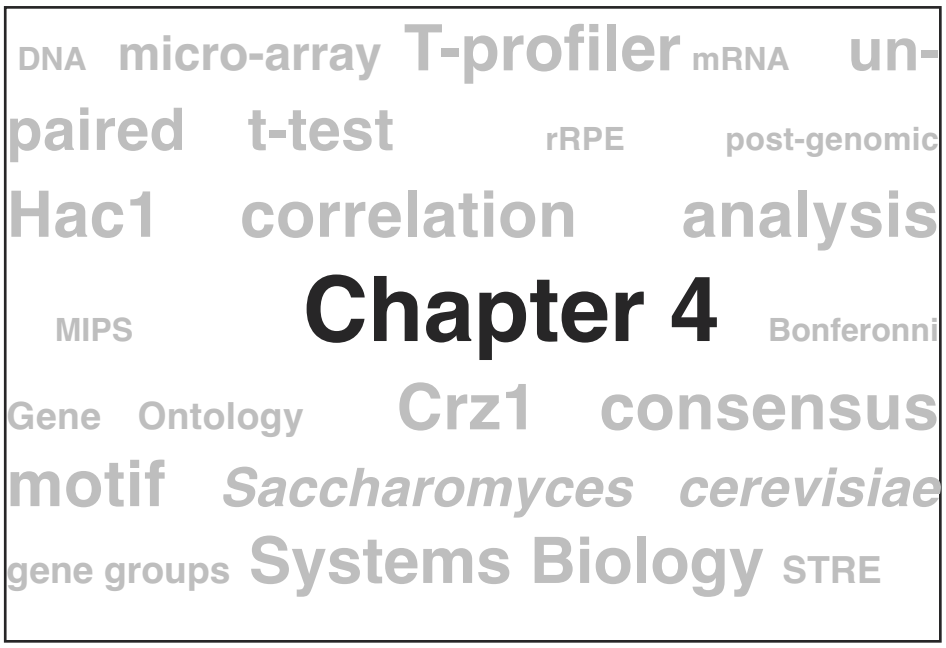
Dissection of transcriptional regulation networks and prediction of gene functions in Saccharomyces cerevisiae

Boorsma, A.


DNA micro-array T-profiler mRNA un-
paired t-test rRPE post-genomic
Hac1 correlation analysis
MIPS
Chapter 4
Bonferonni
Gene Ontology Crz1 consensus
motif Saccharomyces cerevisiae
gene groups Systems Biology STRE

Inferring transcription factor co-modulation networks through regulon-based expression analysis

André Boorsma, Xiang-Jun Lu, Anna Zakrzewska, Ben Snyder,
Frans M. Klis & Harmen J. Bussemaker

Manuscript in preparation

Abstract

A key goal of systems biology is to understand how mRNA expression levels relate to activities of transcription factors (TFs). We recently introduced a statistical method, T-profiler, that scores differential expression at the level of predefined gene sets. Here, we validate experimentally that modulation of TF activity can be inferred by using T-profiler with a collection of "regulons" comprised of the putative targets of various transcription factors (TF) as gene sets. We describe a database (http://www.science.uva.nl/~boorsma/T-base-test/) that quantifies the modulation of TF activity across a large number of conditions in the yeast *S. cerevisiae*, thus facilitating the pathway-centric analysis of regulatory networks. We present evidence that the PAC and rRPE motifs, which antagonize TBP-dependent regulation, function as core promoter elements interacting with the factor NC2. Furthermore, we uncover relationships between regulatory pathways by organizing all TFs into a "co-modulation network" based on correlations between their inferred activity profiles across conditions. Our approach applies to any organism for which expression data and information about putative TF targets is available.

Introduction

A decade ago, simultaneous measurement of the transcript level of all genes in a genome using DNA microarrays became technically feasible [21, 22]. Since then, a large amount of data from such experiments has been accumulated in public repositories [92, 152]. More recently, the marriage of chromatin-immunoprecipitation and microarray technology ("ChIP-chip") [41, 153] has made it possible to measure the genomewide profile of *in vivo* binding by transcription factors (TFs) [62, 154]. Methods for measuring *in vitro* TF-DNA binding affinities have also been developed [155-157].

The rate at which individual genes are transcribed is controlled by DNA-binding transcription factors (TFs) binding to their upstream promoter regions. Knowing the target specificity of each TF, and quantifying how TF activity is modulated by signaling pathways, is therefore crucial for understanding regulatory network function. It is widely recognized that TF activity is often regulated post-translationally, for instance, through translocation between nucleus and cytoplasm. Since it is challenging to measure such changes directly, network inference algorithms often resort to using the mRNA expression level of TFs as a convenient proxy [158, 159]. However, as several studies have shown, it is possible to infer modulation of the "hidden" activity of a TF from the genomewide changes in mRNA expression, by using either motif analysis of upstream promoter sequences [88, 160] or ChIP-chip data [161, 162] to estimate the connectivity between a TF and its target genes. For a review, see [91].

We recently developed T-profiler [83], a microarray data analysis method that scores differential expression of predefined gene sets using the two-sample t-test. T-profiler is closely related to Quontology [81] and PAGE [82], and conceptually simpler than Gene Set Enrichment Analysis [86]. When used with Gene Ontology categories [100], T-profiler has two important advantages over methods that score for enrichment using the hypergeometric distribution [163]: (i) the average expression of a gene set is much more robust than that of individual genes, and (ii) there is no need to impose a threshold on the expression level of individual genes [91]. Another advantage of representing a single expression profile by a relatively small set of statistically robust t-values, which are invariant under additive and multiplicative normalization of the input data, is that it allows for easy comparison between different experiments and laboratories, even across different array platforms.

In this paper, we explore the utility of T-profiler as a simple and practical alternative method for inferring condition-specific modulation of "hidden" TF activity. Rather than treating TF-target connectivity as a quantitative parameter, we treat it as binary, and only require the definition of a set of putative targets or "regulon" for each TF, defined based either on (i) upstream matches to a consensus motif or (ii) the results of a ChIP-chip experiment. We validate experimentally that T-profiler can detect the modulation of TF activity with sufficient sensitivity and specificity. We describe a database, named T-base, which contains t-scores of all predefined gene sets for all experiments in a large compendium of expression data for the yeast *S. cerevisiae*. Querying T-base allows one to determine which TFs are modulated in a given experiment, or conversely, by which environmental conditions a given TF is modulated. We illustrate the util-

ity of our resource by reporting several discoveries regarding transcriptional network function.

Materials & Methods

T-profiler. T-profiler is described in [83]. In short:

For a given gene group G, the t-value is given by the following formula:

$$t_G = \frac{\mu_G - \mu_{G'}}{s\sqrt{\frac{1}{N_G} + \frac{1}{N_{G'}}}}$$

where

$$s = \sqrt{\frac{(N_G - 1) * S_G^2 + (N_{G'} - 1) * S_{G'}^2}{N_G + N_{G'} - 2}}$$

Here m_G is the mean expression log-ratio of the N_G genes in gene group G; $m_{G'}$ is the mean expression log-ratio of the remaining $N_{G'}$ genes; and s is the pooled standard deviation, as obtained from the estimated variances for groups G and G'. The associated two-tailed p-value can be calculated from t using the t-distribution with N-2 degrees of freedom. We accounted for multiple testing by computing an E-value equal to the p-value multiplied by the number of gene groups. All groups with a corrected E-value of 0.05 or smaller are considered to be significantly regulated. To reduce the influence of outliers, which may result in false positives or false negatives, we routinely discard the highest and lowest expression value in each gene group.

Motif-based gene groups. Motif-based groups are defined as groups of genes with a match to a particular consensus motif within 600 base pairs upstream of the ORF [104], allowing no overlap between neighboring ORFs. The consensus motifs used in T-profiler [83] are derived from three different sources. First, motifs were extracted from the SCPD database (http://rulai.cshl.edu/SCPD/). Next, motifs were found by comparing the genome sequence of highly related yeast species [66, 107, 164]. Finally, motifs discovered in various microarray experiments by the REDUCE algorithm [88, 106] were added. Most of these motifs are similar or identical to motifs described in the literature. In total, 153 motif groups have been included in T-profiler calculations.

Gene groups based on transcription factor occupancy data. We used the transcription factor occupancy (TFO) data, obtained by Harbison *et al.* ([62]) using ChIP-chip analysis, as input in T-profiler. This data set contains ChIP-chip results of 203 transcription factors from experiments performed in rich medium (YPD). In addition, 84 of these transcription factors were also measured in at least 1 of 12 other environmental conditions. A gene was considered to be part of a TFO group if the p-value reported by the authors was smaller than 0.001. In addition, TFO groups were required to have at least 7 gene members. This resulted in 252 TFO groups that were used for T-profiler analysis.

Expression library of transcription profiles. Our expression library of transcription profiles contains data of 936 hybridization experiments carried out with *S. cerevisiae* from 19 publica-

tions. This expression library contains data from different DNA-array platforms such as Genefilter, Affymetrix, and spotted slides, and includes experiments with gene deletion strains, synchronized cells for cell cycle analysis, sporulating cells, and cells subjected to various physical and chemical stresses. The expression library has been analyzed using T-profiler and the data have been uploaded to a database named T-base, which can be found at (http://www.science.uva.nl/~boorsma/T-base-test/)

Correlation analysis. To quantify the extent to which a gene group follows the average behavior of another gene group, we computed the Pearson correlation *r* between the t-values of the gene groups across all hybridizations expression library. For each value of *r*, the test statistic

$$t = r\sqrt{\frac{G-2}{1-r^2}}$$

was used to obtain a two-tailed p-value using the t-distribution with *G*-2 degrees of freedom, where *G* is the number of genes. We only used gene groups that had significant t-values in at least 5 experiments.

Fluorescence microscopy

Strains: GFP-fused strains, YNL027W (GFP-Crz1p) and YMR037C (GFP-Msn2p) were from Invitrogen. Strain background: EY0986 ATCC 201388: ***MATa*** *his3*Δ1 *leu2*Δ0 *met15*Δ0 *ura3*Δ0 (S288C).

Medium and growth conditions: YPD (1% yeast extract, 2% Bactopeptone, 2% glucose) was used. YPD containing either 0.4 M $CaCl_2$ or 5 mM dithiothreitol (DTT; Boehringer, Manheim) was mixed with an equal volume of YPD to achieve a final concentration of 0.2 M $CaCl_2$ or 2.5 mM DTT. YPD containing 0.4 M $CaCl_2$ was buffered to pH 5.0 with 7.5 mM succinate to prevent precipitation of $CaPO_4$. Cultures were grown at 30°C and shaken at 250-300 rpm. The culture volume did not exceed 25% of the flask capacity. Cultures were grown to an OD of 0.5 before mixing with equal volumes of either $CaCl_2$ or DTT. For $CaCl_2$-treated cells, samples were taken at 0, 5, 15, 30, and 60 minutes, and for DTT-treated cells, samples were taken at 0, 5, 15, 30, 45, 60, 90, 120, and 180 minutes. For both stress conditions, the experiments from the original papers were repeated ($CaCl_2$ [46], DTT [40].

Cell Fixation and Microscopy: 875 μl of culture were combined with 16% EM grade paraformaldehyde to a final concentration of 2% w/v and mixed for 15 minutes at 25°C. The cells were spun down for 2 minutes. The cell pellet was resuspended and washed in 1 ml of a 0.1 M KP_i (pH=7.5)/1 M sorbitol buffer. Finally, the pellet was resuspended in 50 μl of this buffer and stored at 4°C until use.Three μl of cell suspension were mounted on a glass slide under a coverslip. Microscopic imaging was performed using a CoolSnap fx cooled CCD camera, mounted on an Olympus BX60 fluorescence microscope (Olympus, Tokyo, Japan) using a phase-contrast 100× oil-immersion objective with NA = 1.3 (UPlan FI). Fluorescence was excited with a 100 W mercury lamp; for GFP-pictures a U-MNB narrow-band cube (excitation 470–490 nm; emission >515 nm) was used. For DAPI-stained cells, 4′,6-diamidino-2-phenylindole dihydrochloride hydrate (DAPI) was added to a final concentration of 0.5 μg/ml. For DAPI pictures, a U-MWU wide-band cube (excitation 330–385 nm; emission >420 nm) was used.

Results

Creating a database of inferred TF activities

We used T-profiler [83] to create T-base, a database of t-values that quantify the change in mean expression for a large number of predefined gene sets across a large number of experimental conditions (**Figure 1a**). For genes sets, we used both "motif-based" regulons, defined based on matches to specific consensus motifs in their 600-base pair upstream regions, and "ChIP-based" regulons, defined based on measurements of promoter occupancy in different conditions by Harbison *et al.* [62]. We analyzed a wide variety of experiments, including cell cycle [58], various stress response time courses [40], and a collection of gene deletion and gene suppression experiments [59, 165]. The full results are available at http://www.science.uva.nl/~boorsma/T-base-test/. While not discussed in what follows, T-base also contains the results of the expression library analyzed with gene groups based on Gene Ontology (GO) categories [100].

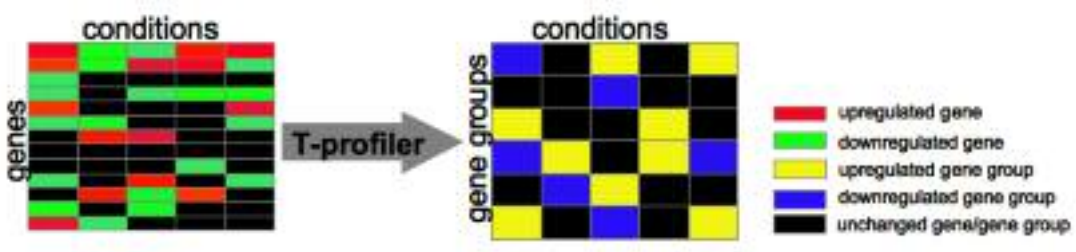


Figure 1 T-base. T-profiler [83] was used to convert each genomewide mRNA expression profile to a set of t-values that quantify the change in regulatory activity for each TF for which a set of putative targets ("regulon") was available. The result was stored in a database available at http://www.science.uva.nl/~boorsma/T-base-test/

Validation of inferred condition-specific TF activity regulation

We first tested the ability of T-profiler to infer changes in TF activity by analyzing experiments in which a transcription factor-encoding gene was either deleted or over-expressed. Yap1p activates genes involved in the response to oxidative stress, while Rox1p represses genes upon oxygen limitation. We monitored the t-values of the ChIP-based Yap1p regulon (72 genes) and the motif-based (YCTATTGTT) Rox1p regulon (95 genes); see **Figure 1b**. In a *YAP1* deletion strain, significant down-regulation (t-value = -4.0; E-value=0.015) of the Yap1 regulon is observed, while over-expression of *YAP1* results in its upregulation (t-value = 5.6; E-value = $6*10^{-6}$). Conversely, deletion of the repressor gene *ROX1* results in upregulation of the Rox1p regulon, while overexpression of *ROX1* causes downregulation. The specificity of our method is demonstrated by the lack of a Yap1 gene set response in H_2O_2-stressed Δ*yap1* cells.

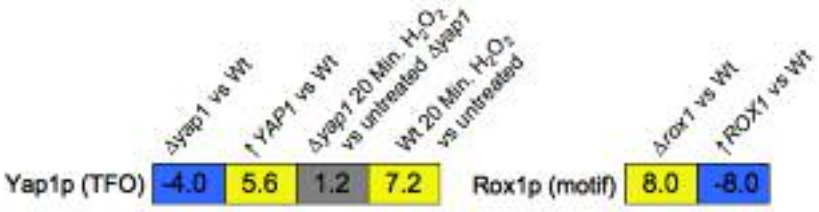


Figure 2 Validation of the inferred TF activities Change in regulatory activity of the activator Yap1p and the repressor Rox1p when the corresponding factors are deleted or overexpressed, as inferred by T-profiler. The t-values for Yap1 are based on the ChIP-based regulon (rich medium), while for Rox1 a motif-based (YCTATTGTT) regulon was used. As expected, a Yap1 regulon response is observed for wild-type cells stressed using H_2O_2, but not for D*yap1* cells in the same condition. More information (size/members) about the regulons can be found at www.science.uva.nl/~boorsma/t-base-test .

The predicted activity of Crz1p and Msn2p corresponds with their translocation to the nucleus

We validated predictions concerning the timing of the transcriptional response of the gene sets controlled by the calcineurin-controlled transcription factor Crz1p and by the general stress activator Msn2p, under two different stress conditions. **Figure 3a** shows the activity of the Crz1 motif (GAGGCT) gene group in response to $CaCl_2$ [46] and dithiothreitol (DTT) [40]. During both $CaCl_2$- and DTT-induced stress, the Crz1 motif-based gene group is activated, but upon $CaCl_2$ stress an immediate response (within 5 minutes) is seen, whereas upon DTT stress the response is considerably delayed. Addition of $CaCl_2$ to the growth medium increases the concentration of cytosolic Ca^{2+} and activates calcineurin, which then dephosphorylates the transcription factor Crz1p. This enables Crz1p to translocate to the nucleus, where it activates its target genes [166]. We made use of a GFP-tagged Crz1 protein (Materials & Methods) to validate our predictions about the activity of Crz1p during $CaCl_2$ and DTT stress. We reproduced the $CaCl_2$ and DTT stress experiments as described in the original papers [40, 46]. The subcellular location of Crz1p was monitored by fluorescence microscopy (Materials and Methods). Upon $CaCl_2$ stress, Crz1p moves to the nucleus within 5 minutes and stays there during the remainder of the time course (**Figure 3b**). In cells treated with DTT, the first cells with nuclear Crz1p appear only after 30 minutes. In both cases the measured responses correspond with the computational predictions of **Figure 3a**. Besides Crz1p, we also measured the sub-cellular localization of an Msn2-GFP fusion protein for both $CaCl_2$ and DTT stress. In $CaCl_2$ stressed cells, Msn2p is translocated to the nucleus within 5 minutes. During DTT stress the first nuclear located Msn2-GFP proteins appear at the 60 minute timepoint. Again in both cases the inferred activation of the Msn2 regulon (STRE motif - AGGGG) corresponded with translocation to the nucleus (data not shown).

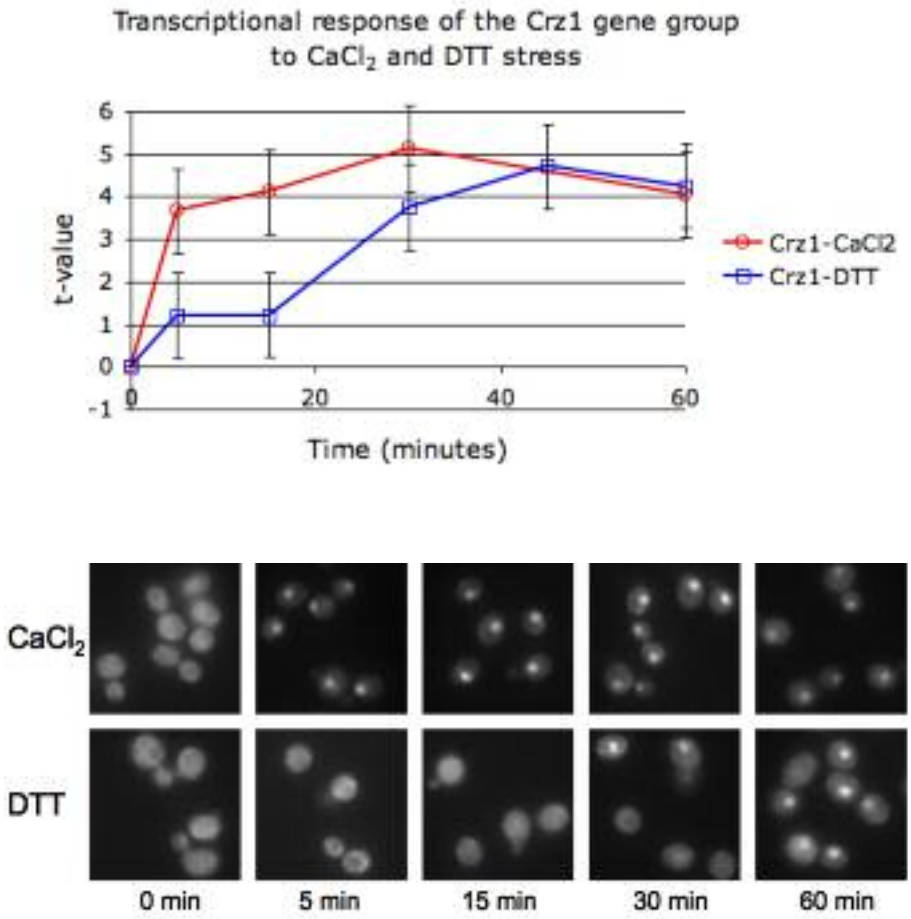
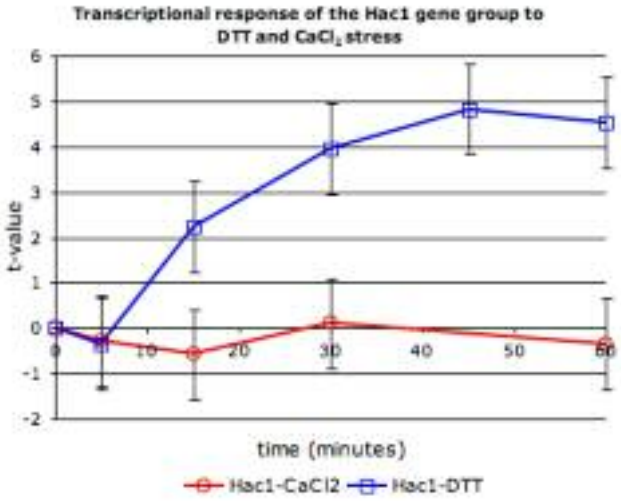


Figure 3 Timing of the activation of Crz1p corresponds with translocation of Crz1p to the nucleus. (a) Timing of the activation of the Crz1p motif-based gene set during $CaCl_2$ [46] and DTT [40] stress. Note that the t-value for each time point is derived from a distinct genome-wide expression profile. **(b)** Cellular localization of Crz1p during $CaCl_2$ (upper panel) and DTT stress (lower panel) assayed using fluorescence microscopy. The small bright spots in the cells are the nuclei, to which GFP-tagged Crz1p has translocated. More information (size/members) about the gene sets can be found at www.science.uva.nl/~boorsma/t-base-test .

Condition-specific activity of the Hac1 gene group

DTT-stressed cells have to cope with reductive stress resulting in accumulation of misfolded proteins in the endoplasmic reticulum [167]. This leads to the activation of the unfolded protein response, governed by the transcription factor Hac1p [168]. **Figure 4a** shows that the Hac1p-mediated response is independent of the Crz1p response and therefore does not occur during $CaCl_2$ stress. Next, we queried T-base to find experiments that show a significant t-value

for the Hac1 TFO-based gene group. By ranking all experiments according to the t-value of the Hac1p gene group, we found that the Hac1p gene group is specifically activated in expression profiles from DTT-stressed cells or in cells in which certain essential genes have been partially suppressed [165] (**Figure 4b**). According to *Saccharomyces* Genome Database (SGD), *GPI2* and *GWT2* function in GPI-anchor biosynthesis, whereas *GPI16* and *GAB1* are involved in transferring pre-assembled GPI-anchors to a specific class of secretory proteins, called GPI-proteins. When these processes do not function properly, defective GPI-proteins accumulate in the ER. Finally, *PGA1* codes for a protein that localizes to the nuclear periphery, a subregion of the ER. When its activity is repressed, maturation of the GPI-protein Gas1p and of Pho8p, which also follows the secretory pathway, is affected [169], probably resulting in their accumulation in the ER. In other words, activation of the Hac1 regulon seems to occur specifically when defective proteins accumulate in the ER.



Experiment	t-Value	P-value	Mean	ORFs	Reference
GPI2 (TET promoter)	5.29	$< 10^{-12}$	0.78	12	Mnaimneh[3]
2.5mM DTT; 45 min	4.83	3.00E-04	0.69	11	Gasch[4]
2.5mM_DTT; 60 min	4.54	1.40E-03	0.81	9	Gasch
GWT1 (TET promoter)	4.39	2.90E-03	0.91	12	Mnaimneh
2.5mM DTT; 120 min	4.35	3.40E-03	0.82	12	Gasch
GPI16 (TET promoter)	4.16	7.80E-03	1.04	12	Mnaimneh
GAB1 (TET promoter)	4.11	1.00E-02	0.97	12	Mnaimneh
2.5mM_DTT; 90 min	4.04	1.34E-02	0.61	12	Gasch
2.5mM_DTT; 30 min	3.96	1.88E-02	0.55	10	Gasch
PGA1 (TET promoter)	3.95	1.92E-02	1.01	12	Mnaimneh

Figure 4 Condition-specific activity of the Hac1 regulon (a) Transcriptional response of the Hac1p Transcription Factor Occupancy (TFO) gene group during DTT [40] and $CaCl_2$ [46] stress. **(b)** Condition-specific activation of the Hac1p (TFO) gene group. The top ten expression profiles (out of 936) ranked by their t-values from T-base are shown. These expression profiles are either from DTT-stressed [40] cells or from cells with a partially suppressed essential gene under control by the TET-promoter [165].

PAC and rRPE may serve as NC2-dependent core promoter elements

Besides the specific response of the Hac1p gene set to DTT stress, a general transcriptional program known as the Environmental Stress Response (ESR) is triggered [40]. Motifs associated with the ESR include the stress-response element (STRE) motif (AGGGG/CCCCT), which is bound by the transcription factor Msn2p [36], PAC (CGATGAG), and rRPE (AAAATTT), which is associated with genes required for rapid growth [44, 40]. **Figure 5a** shows activity profiles for the corresponding gene sets during DTT stress. Further analysis of the activity profiles of the ESR motifs represented in T-base shows that the antagonism between STRE and PAC/rRPE observed during DTT stress holds over a wide range of cellular states (**Figure 5b,c**). The TATA-box gene set (TATAWAWR) correlates strongly positively with STRE (r = 0.80), consistent with recent observations by Basehoar *et al*. [30] who showed that TATA-box containing genes are activated in response to various stresses.

The strongly coupled, but opposing transcriptional behavior of the STRE/TBP and PAC/rRPE gene sets across many conditions suggests a mechanistic relationship. Currently, it is not known which gene specific transcription factors bind to the PAC and rRPE elements. Similar to the TATA-Box binding Protein (TBP) motif, the PAC and rRPE elements are predominantly found in the first 150 bp upstream from the translational start site [170]. Promoter regions of genes containing PAC and rRPE elements are generally TATA box-less. Beer and Tavazoie [170] found that PAC and rRPE elements correlate with expression only when the PAC element is located downstream of the rRPE element. Similar motif characteristics have been described for regulatory sequences in Drosophila named DPE (Downstream core Promoter Element), which serve as core promoter elements [171]. The DPE is bound by NC2, a bi-functional general transcription factor that differentially regulates gene transcription through DPE or TATA-box motifs [172]. NC2 is a heterodimer of two histone-fold subunits. In *S. cerevisiae*, the α-NC2 subunit consists of Bur6p and Ydr1p and the β-NC2 subunit consist of Ncb2p. **Figure 5d** shows that expression profiles of bur6Δ [173] cells show strong induction of the TBP (TATAWAWR) (t-value=12.3) and STRE (AGGGG) gene sets (t-value=10.8) and strong repression of PAC (CGATGAG) and rRPE (AAAATTT) gene sets (t-values=-7.9 and -11.2, respectively). While the expression profile of a TBP mutant (F182V; [174]) that is unable to bind NC2 shows similar behavior, TBP mutants that are unable to dimerize show an opposite pattern. Together, these observations suggest that the PAC and rRPE sequences might function as core promoter elements with similar properties as DPE, and that *S. cerevisiae* NC2 plays a similar role as in Drosophila, where it activates DPE-driven promoters and represses TATA-box driven promoters [172].

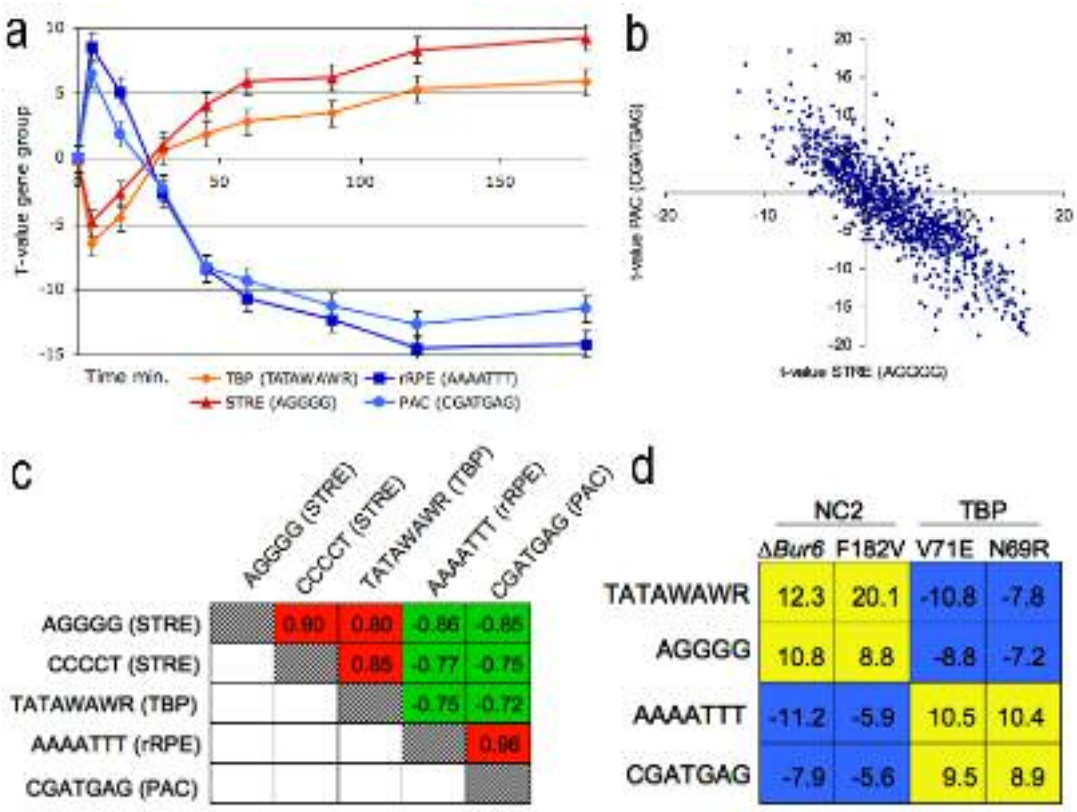


Figure 5 Motif-based dissection of the Environmental Stress Response. (a) Transcriptional response during DTT stress [40] of four regulons based on motifs (see text) associated with the Environmental Stress Response (ESR). **(b)** Scatter plot of the t-values from T-base (936 experiments) for the STRE versus the PAC regulon, showing a strong negative correlation (r = -0.85). **(c)** Strongly coupled antagonism between STRE/TBP and PAC/rRPE motifs. Shown are Pearson correlation coefficients for all pairwise comparisons of the t-value profiles for the four motifs. All r-values correspond to an E-value < 10^{-14}. (**d**) Evidence for interaction between PAC/rRPE and the factor NC2. Shown is the response of the four motif-based regulons to mutations of components of NC2 (*bur6*Δ and F182V (F182V is actually a TBP mutant that is unable to bind NC2) and TBP (V71E and N69R).

Comparison between the transcription factor activity profiles based on gene groups and mRNA levels

The activity of motif- and TFO-based gene groups expressed in t-values may be regarded as a proxy for the activity of transcription factors. In other studies of gene regulatory networks, the activity of transcription factors is often inferred from their mRNA expression levels [74, 175, 176]. However, the activity of many transcription factors is also strongly regulated on a post-transcriptional level, for example, through (de)phosphorylation and translocation from the cytosol to the nucleus and back rather than changes in transcript level. For example, the activation of Crz1p (**figure 3**) is regulated by translocation to the nucleus; indeed no correlation is found between the differential mRNA levels and the gene group level of the TF Crz1p (r = 0.08) over all expression profiles (**figure 6a**). The best correlation that is found between regulation on mRNA level and gene group level is that of Hap4 (0.47) (**figure 6b**).

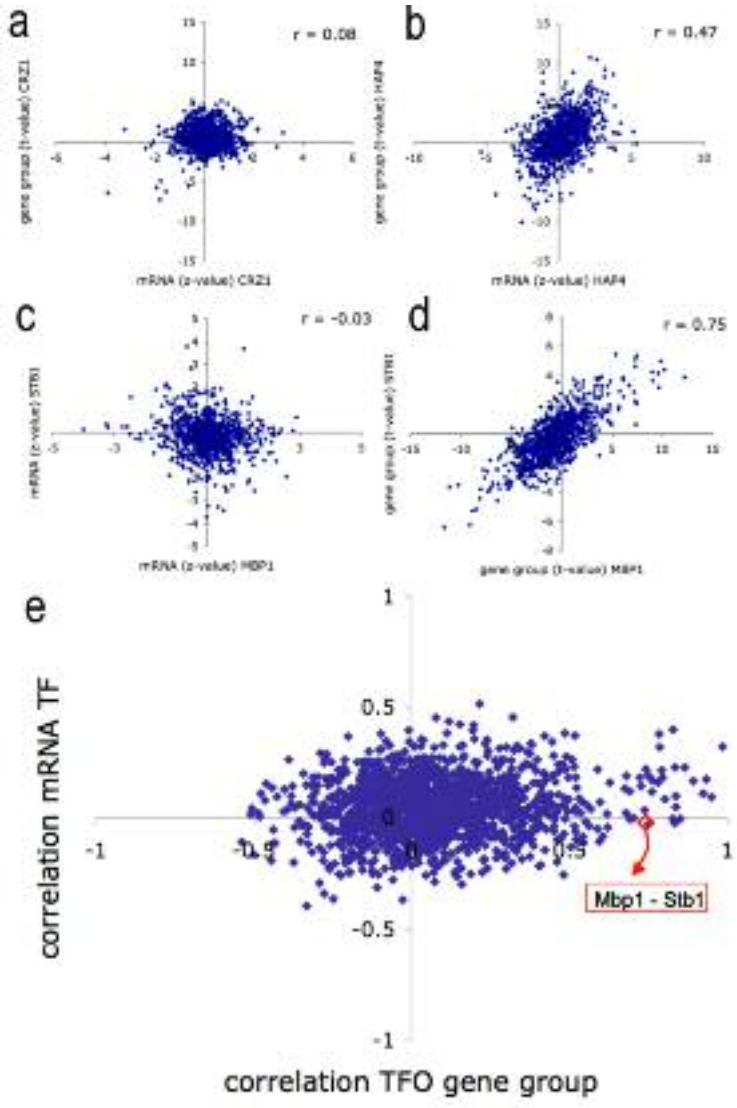


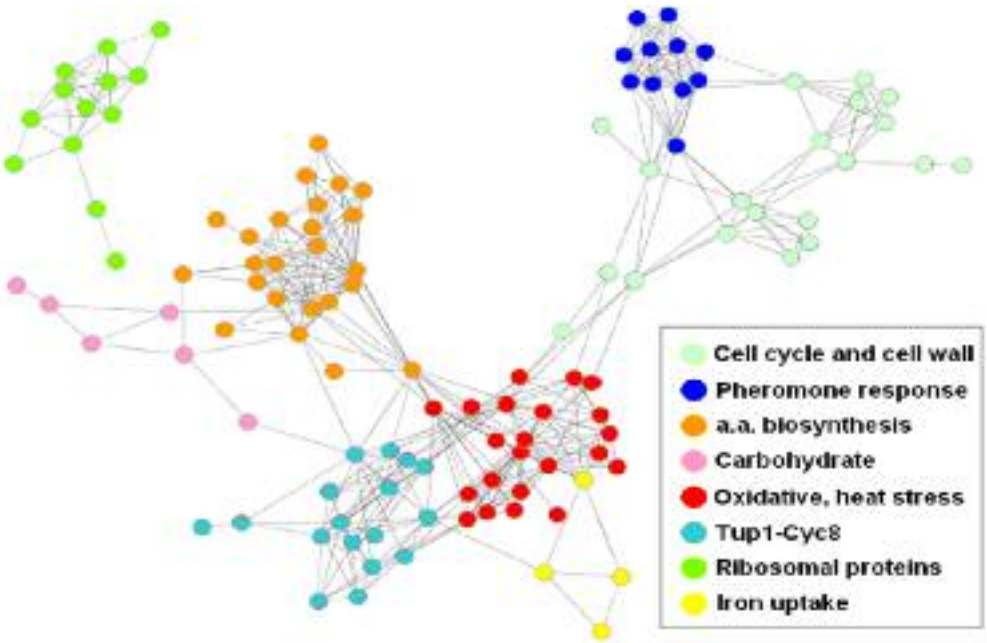
Figure 6 Correlation between transcription factors based on their individual mRNA levels and on the t-values of the corresponding gene groups. Scatter plots Pair-wise Pearson correlation between the differential mRNA levels (z-score) and the gene group levels (t-values) of **(a)** *CRZ1* (no correlation r = 0.03) and **(b)** *HAP4* (correlation r = 0.47). **(c)** Scatter plots Pair-wise Pearson correlation between the differential mRNA levels (z-score) of *STB1* and *MBP1*. **(d)** Scatter plots Pair-wise Pearson correlation between the gene group levels (t-values) of *STB1* and *MBP1*. **(e)** Scatter plot of the pair-wise Pearson correlation between t-values of TFO-based (only YPD condition) gene groups (x-axis) versus pair-wise Pearson correlation between $\log_2$-expression ratios of the genes coding for the corresponding transcription factors (y-axis). The t-values of TFO-based gene groups were used, because they generally show a stronger correlation than motif-based gene groups. The example that is highlighted are the transcription factors Mbp1p and Stb1p that show poor correlation on the mRNA level (*r*=0.03) (see figure 6c) but strong correlation only on the gene group level (*r*=0.75) (see figure 6d).

For each TF, the activity profile over roughly a thousand conditions, as inferred by T-profiler and stored in T-base, represents a highly specific regulatory signature. It is unlikely for two activity profiles to be similar, unless (i) they are derived from strongly overlapping regulons, or (ii) the corresponding TFs are modulated by related signaling networks. The latter suggests a novel approach to organizing the TFs into a network based on co-modulation of their post-translational activity. To motivate this, consider the cell cycle regulators Stb1p and Mbp1p. The correlation between their mRNA expression values (r=0.03) over all conditions in T-base is not statistically significant. However, the t-values scoring the differential expression of the ChIP-based regulons for Mbp1p (188 genes) and Stb1p (63 genes) are highly correlated (r=0.75); even when the 23 genes that occur in both regulons are excluded, the correlation remains high (r=0.54). These results indicate that implicit information about the connectivity between signal transduction pathways and transcription factors can be obtained by comparing the activity profiles of TFs.

Finally, we plotted the pair-wise Pearson correlation between the t-values of TFO gene groups versus the correlation coefficient between z-ratios of the mRNA levels. **Figure 6e** shows that the correlations derived from the t-profiles of two TFO gene groups generally show a stronger correlation and are therefore a more sensitive measure for the co-modulation of the corresponding transcription factors than correlation based on the mRNA expression log-ratios of the same factors.

Organizing TFs into "co-modulation networks" based on their activity profile

Starting from ChIP-based activity profiles for a large number of TFs, and drawing connections between pairs of TFs only when the correlation between their activities exceeded a stringent threshold (r>0.5), we organized all TFs into a "co-modulation network" consisting of eight disjoint sub-networks (**Figure 7**).



In agreement with the findings by Luscombe *et al*. [175], the cell-cycle sub-network and the pheromone response sub-network are found to be separated from the other sub-networks (**Figure 7, blue and green dots**), whereas the oxidative – heat stress sub-network takes a central position (**Figure 7, red dots**). The most highly connected transcription factors are Msn4p (with 21 interactions) and Msn2p, Gcn4p, and Skn7p (each with 20 interactions). These transcription factors might represent central regulators. Within the oxidative-heat stress sub-network (**Figure 8a**) there is a separation between transcription factors involved in oxidative stress (Yap1p, Yap7p and Cad1p) and heat stress (Hsf1p). The sub-network also contains Skn7p that previously has been described as being involved in oxidative, heat and osmotic stress [177].

The sub-network shown in **Figure 8b** reveals the co-modulation of Rap1p, Sfp1p and Fhl1p, known to control the expression of ribosomal protein genes, and Hir1p, Hir2p, and Hir3p, which are co-repressors involved in the cell-cycle-regulated transcription of histone genes. While ribosome biogenesis has been linked to cell division via Sfp1p [178], the parallel activation of the Hir regulon detected by our co-modulation approach provides additional clues about the coupling between these two processes.

One of the sub-networks (lower left in **Figure 7**) contains Sut1p, Nrg1p, Phd1p, Rim101p and Sok2p (**Figure 8c**). These TFs are involved in a variety of stress responses. However, a shared feature is that most of them are known to repress gene transcription, and interact with the co-repressor Tup1p-Cyc8p (Ssn6p). To further test the hypothesis that this sub-network consists of TFs that depend on Tup1p-Cyc8p to control the expression of their targets genes, we analyzed the expression profiles of both the *tup1* and *cyc8* deletion mutant [59]. We found that almost all of the ChIP-based regulons found in this sub-network are de-repressed in both the *tup1*Δ/wt and the *cyc8*Δ/wt expression profiles (**Table 1**). One of the members of the Tup1p-Cyc8p sub-network is Cin5p, a poorly characterized basic leucine zipper transcription factor of the yAP-1 family, which mediates pleiotropic drug resistance [179]. It is constitutively located in the nucleus. The Cin5p regulon is de-repressed in a *cin5* deletion mutant [59] included in T-base. We therefore predict that Cin5p is a transcriptional repressor that interacts with the Tup1p-Cyc8p co-repressor complex.

Figure 7 (page 82) Co-modulation network of transcription factor (TFO) gene groups. Similar colors represent functionally related transcription factors. Eight sub-networks can be distinguished. The nodes (TFO-based gene groups) are considered to be co-modulated if their t-values are positively correlated over 936 experiments at a level $r > 0.5$.

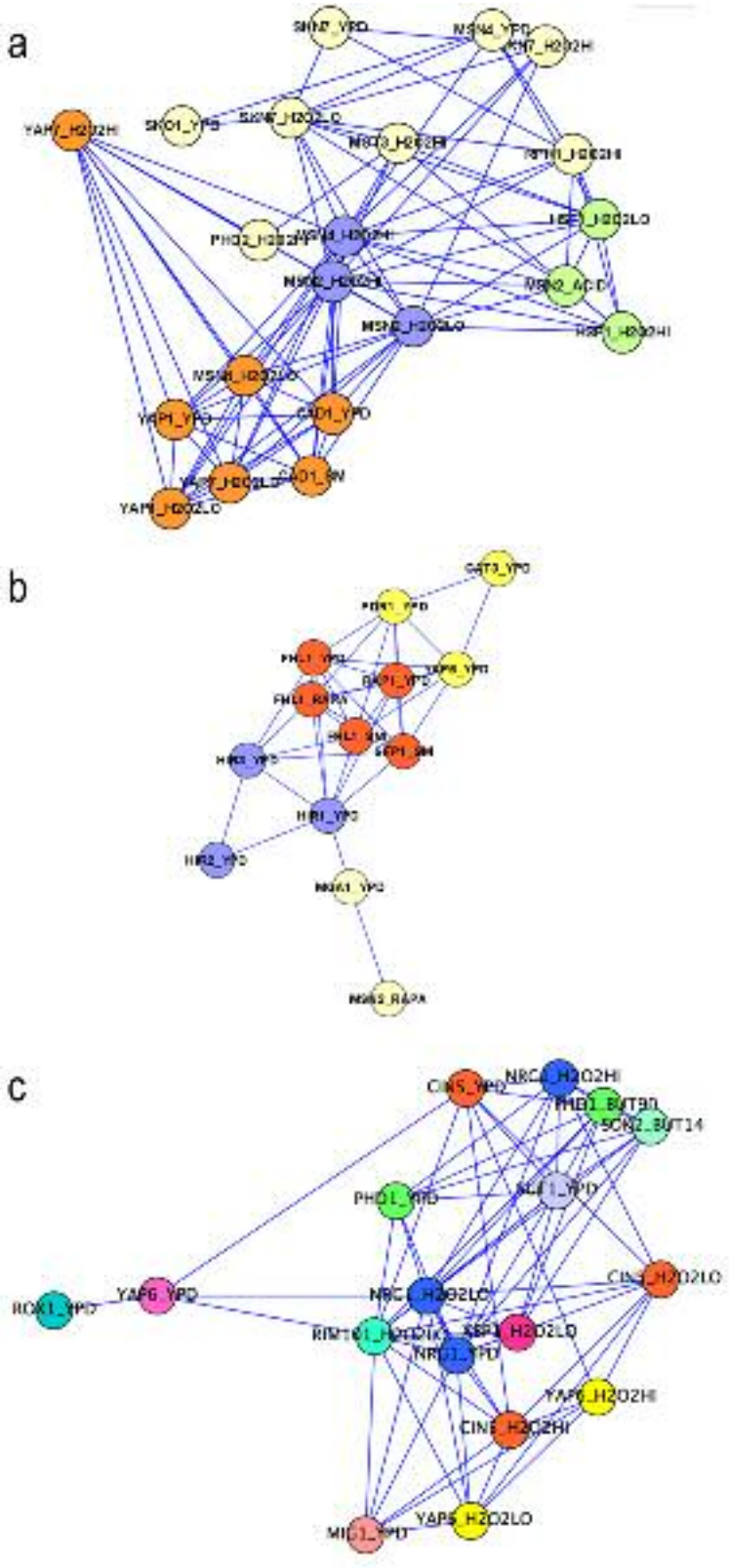

a
b
c

Table 1 T-profiler analysis of the *tup1*Δ and *cyc8* (*ssn6)*Δ transcription profiles. Shown are ChIP-based regulons with a significant t-score (E-value < 0.05) for *tup1*Δ and *cyc8* (*ssn6)*Δ mutant vs. wild-type expression data [59]. Regulons scoring significantly in both mutants are shown in bold. The transcription factors that are part of the Tup1-Cyc8 co-modulation sub-network (Figure 3b) are shown in grey. The condition of the ChIP-chip experiment [62] is shown in parentheses; (YPD; cells grown in YPD, H2O2 low; 0,4 mM hydrogen peroxide, H2O2 high; 4 mM hygrogen peroxide, BUT 14; 1% butanol - 14 hours, BUT 90; 1% butanol 90 minutes)

.

*tup1*Δ/wt		*cyc8*Δ/wt	
TF (condition)	**t-value**	**TF (condition)**	**t-value**
NRG1 (YPD)	**14.8**	**SOK2 (BUT 14)**	**9.6**
RIM101 (H_2O_2 low)	**14.5**	**NRG1 (YPD)**	**9.6**
CIN5 (H_2O_2 low)	**13.9**	**YAP6 (YPD)**	**8.6**
NRG1 (H_2O_2 low)	**13.6**	**NRG1 (H_2O_2 low)**	**8.6**
YAP6 (H_2O_2 low)	**12.2**	**PHD1 (BUT 90)**	**8.5**
SOK2 (BUT 14)	**11.6**	**CIN5 (H_2O_2 low)**	**8.4**
YAP6 (YPD)	**11.0**	**RIM101 (H_2O_2 Low)**	**8.1**
PHD1 (BUT 90)	**10.6**	**NRG1 (H2O2 high)**	**8.1**
MIG1 (YPD)	**10.6**	**CIN5 (YPD)**	**8.0**
PHD1 (YPD)	**10.6**	**YAP6 (H_2O_2 low)**	**7.9**
NRG1 (H_2O_2 high)	**9.7**	**SUT1 (YPD)**	**7.5**
SUT1 (YPD)	**9.6**	**PHD1 (YPD)**	**7.5**
CIN5 (H_2O_2 high)	**9.3**	**CIN5 (H_2O_2 high)**	**6.8**
YAP6 (H_2O_2 high)	**8.6**	**MIG1 (YPD)**	**6.7**
CIN5 (YPD)	**8.5**	**AFT2 (H_2O_2 low)**	**6.5**
YJL206C (H_2O_2 low)	7.5	**SKN7 (H_2O_2 low)**	**6.4**
SKN7 (H_2O_2 low)	**7.2**	**XBP1 (H_2O_2 low)**	**5.6**
AFT2 (H_2O_2 low)	**7.0**	**SKN7 (H_2O_2 high)**	**5.5**
XBP1 (H_2O_2 low)	**6.5**	**YAP6 (H_2O_2 high)**	**5.4**
CUP9 (YPD)	5.9	**SKN7 (YPD)**	**5.3**
SKN7 (YPD)	**5.7**	RCS1 (H_2O_2 high)	4.6
SKO1 (YPD)	5.7	PUT3 (H_2O_2 low)	4.5
SKN7 (H_2O_2 high)	**5.6**	**ROX1 (YPD)**	**3.9**
YJL206C (YPD)	**5.6**	**YJL206C (YPD)**	**3.8**
ROX1 (YPD)	**4.8**		
YAP1 (H_2O_2 low)	4.1		

Figure 8 (page 84) The oxidative and heat stress and Tup1p-Cyc8p sub-network. (**a**) The oxidative and heat stress sub-network (middle cluster in figure 7). Behind the TF name is the condition presented of the ChIP-chip experiment [62]. Right: TFO based gene groups that mainly contain heat stress genes; in orange: TFO based gene groups that mainly contain oxidative stress genes; the Msn2 /4 TFO based gene groups (middle) interconnect both; (top): the Skn7 TFO gene groups. Finally, the TFO-gene groups in yellow do not have a clear relationship. **(b)** In the “ribosomal protein” sub-network (upper left cluster in Figure 7), the histone-regulating factors Hir1p/Hir2p/Hir3p are connected to the ribosomal protein gene expression factors Rap1p-Fhl1p-Sfp1p. **(c)** The “Tup1p-Cyc8p” sub-network (lower left cluster in Figure 7), in which TFs that rely on this co-repressor to control their transcriptional targets are connected.

Conclustion & Discussion

In this study we used T-profiler to infer changes in the activity of transcription factors with respect to gene groups predicted to be under their control, based either on upstream sequence matches to *cis*-regulatory elements or on occupancy by a specific transcription factor in ChIP-chip experiments. We then created T-base, a database containing t-values quantifying the differential expression of all predefined gene groups under a wide variety of stress conditions and in gene deletion mutants in budding yeast. Finally, we calculated correlations between gene groups over a large set of experiments. Our method was validated in two ways. First, deletion and over-expression of two transcription factors (an activator and a repressor) confirmed the expected up- and down-regulation of their accompanying gene groups. Second, the physical movement of two transcription factors to the nucleus observed during calcium and DTT treatment agreed with the calculated response profiles. DTT stress also activates a specific response of a gene group regulated by the Hac1p transcription factor, a response that does not occur in cells treated with calcium. In fact, querying T-base for experiments, in which the Hac1p-based gene group is activated, only reveals 11 experiments with significant t-values. Four of those originate from the DTT time course, while the others are from transcription profiles of partially suppressed essential genes. Interestingly, these genes are either involved in GPI-anchor biosynthesis, GPI-anchor addition, or in GPI-protein maturation. Although we only present one example of such a condition-specific activity, T-base contains many more examples of condition-specific activity. For example, the Rlm1p-based gene group is mainly activated in experiments related to cell wall perturbation, caused by, for example, Calcofluor white or Zymolysase [51], or in deletion mutants defective in cell wall formation [120].

In contrast to the condition-specific activity of many gene groups, the gene groups based on the STRE motifs (AGGGG/CCCCT) and TBP (TATAWAWR) are activated in 50% of all the experiments and are therefore regulated in a more general manner. Compared to the STRE and TBP-motif gene groups, the PAC and rRPE-based gene groups show opposite transcriptional behavior. We propose that there is a mechanistic relationship between the regulation of these motif gene groups and we provide evidence that NC2, a bi-functional general transcription factor that binds TBP [155] serves as the mechanistic link. The observed bipolar transcriptional regulation in *Saccharomyces cerevisiae* is also found by others [30]. Most probably, it reflects a redistribution of energy fluxes in which the production of ribosomal genes (or intermediates of the ribosomal proteins) and thereby growth is temporarily down regulated and the energy in the form of ATP that is becoming available is used to respond in order to survive the stress that the cell is facing. Interestingly, Basehoar *et al.*, [15] found that approximately 20% of yeast genes contain a TATA box, and similar numbers have also been found for higher eukaryotes. It might be interesting to determine to what extent this form of regulation is conserved in higher eukaryotes. For example, if TBP-NC2 would play a similar role in human cells in the balance between fast and slow growing cells, it might be a potential target for cancer treatment.

Finally, we show that correlation of the activity of transcription factors based on TFO gene groups is generally stronger compared to correlation based on the mRNA levels of the same

factors. The co-modulation network that we built based on the correlation of all the TFO-based gene groups consists of 8 sub-networks. In agreement with the results of Luscombe *et al*. [175] we found the cell-cycle and pheromone sub-network to be separated from the other sub-networks. Interestingly, we found a sub-cluster of transcription factors most of which interact with the co-repressor Tup1p-Cyc8p. Based on the content of this sub-network, we predict the transcription factor Cin5p to be regulated by the Tup1p-Cyc8p complex and therefore to be a repressor of gene expression. Additional evidence in support of this prediction comes from the expression profile of *cin5*D cells; in which the Cin5p-based gene group is de-repressed, consistent with the predicted repressor function of Cin5p.

In summary, we have experimentally validated that a simple regulon-based method can predict condition-specific modulation of TF activity. We created a web-accessible database (T-base) that will serve as a valuable resource for the biological community, and demonstrated its usefulness by generating new hypotheses about the mechanism associated with the elusive PAC and rRPE motifs. Extending our approach to organisms other than yeast should be straightforward.

Acknowledgements
This work was supported by grants from the Netherlands Foundation for Technical Research (STW) to F.K. (APB.5504) and from the National Institutes of Health (HG003008) to HJB. We would like to thank Gabor Halasz, Junbai Wang, Ron Tepper and Daniel Vis for critical reading of the manuscript, and Conrad Woldringh and Wijnand Takkenberg for their help with fluorescence microscopy.