7024–7031 *Nucleic Acids Research, 2003, Vol. 31, No. 23*
DOI: 10.1093/nar/gkg894

Identifying cooperativity among transcription factors controlling the cell cycle in yeast

Nilanjana Banerjee[1,2] and Michael Q. Zhang[1,*]

[1]Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA and [2]George Mason University, School of Computational Sciences, 10900 University Boulevard, Manassas, VA 20110, USA

Received July 1, 2003; Revised September 22, 2003; Accepted October 8, 2003

ABSTRACT

Transcription regulation in eukaryotes is known to occur through the coordinated action of multiple transcription factors (TFs). Recently, a few genome-wide transcription studies have begun to explore the combinatorial nature of TF interactions. We propose a novel approach that reveals how multiple TFs cooperate to regulate transcription in the yeast cell cycle. Our method integrates genome-wide gene expression data and chromatin immunoprecipitation (ChIP-chip) data to discover more biologically relevant synergistic interactions between different TFs and their target genes than previous studies. Given any pair of TFs *A* and *B*, we define a novel measure of cooperativity between the two TFs based on the expression patterns of sets of target genes of only *A*, only *B*, and both *A* and *B*. If the cooperativity measure is significant then there is reason to postulate that the presence of both TFs is needed to influence gene expression. Our results indicate that many cooperative TFs that were previously characterized experimentally indeed have high values of cooperativity measures in our analysis. In addition, we propose several novel, experimentally testable predictions of cooperative TFs that play a role in the cell cycle and other biological processes. Many of them hold interesting clues for cross talk between the cell cycle and other processes including metabolism, stress response and pseudohyphal differentiation. Finally, we have created a web tool where researchers can explore the exhaustive list of cooperative TFs and survey the graphical representation of the target genes' expression profiles. The interface includes a tool to dynamically draw a TF cooperativity network of 113 TFs with user-defined significance levels. This study is an example of how systematic combination of diverse data types along with new functional genomic approaches can provide a rigorous platform to map TF interactions more efficiently.

INTRODUCTION

Precise transcriptional control is one of the major steps in gene expression and regulation. Understanding the mechanisms behind this precise control remains a challenge. The current view is that the cell exerts this control by employing principles of cooperativity and synergy. This process allows small combinations of ubiquitous, signal-specific transcription factors (TFs) binding to promoter DNA, to execute an exponentially larger number of regulatory decisions. The involvement of different kinds of TFs in gene regulation makes possible the integration of several signaling pathways in the nucleus. The interactions between TFs on promoter DNA that underlie this so-called 'combinatorial transcriptional regulation' can be classified into three modes: those between a DNA-binding factor and a non-DNA-binding factor; those between DNA-binding factors adjacently located on the promoter; and those between DNA-binding factors separately or distantly located on the promoter (1).

With advances in experimental approaches and diverse data resources, functional genomics has begun to investigate the more complex, cooperative interactions required by combinations of TFs to properly regulate gene expression (2–4). Pilpel *et al.* (3) screened for cooperatively binding TFs by correlating pairs of computationally derived motif combinations with gene expression data. The presence of computationally derived motif combinations in the promoter, however, does not automatically give direct evidence of TF binding. As a result, such analysis can potentially suffer from a large number of false positives in predicting functional TF-binding sites.

Recently, genome-wide location data (ChIP-chip) (5–7) elucidated the *in vivo* physical interactions of TFs with their chromosomal targets on the genome and as a result can provide a more reliable view of functional TF-binding site interaction. Lee *et al.* (5) and Itamar *et al.* (6) have used genome-wide location analysis to explore the yeast cell cycle gene expression program and showed that TFs that function during one stage of the cell cycle regulate those that function during the next stage. These studies have underscored the need to integrate diverse data resources and construct tools that will assimilate them into biological models (8).

In this paper, we exploit ChIP-chip data (5) (for determining downstream targets of TFs *in vivo*) and genome-wide gene

*To whom correspondence should be addressed. Tel: +1 516 367 8393; Fax: +1 516 367 8461; Email: mzhang@cshl.edu

expression data (9) to rigorously assess cooperativity among TFs in the yeast cell cycle. We generate statistically significant cooperative TFs by exploring the effect of cooperative binding versus independent binding of the TFs on gene expression. The assumption is that if two TFs are cooperative then they should both bind (either directly or through another DNA-binding protein) to the promoters of their target genes and the expression profiles of these target genes would be similar. If they are not cooperative, the TFs will either bind independently or not bind at all to the promoters and the expression profiles will likely not be coherent overall. With our statistical measure we aim to identify the subset of cooperative TFs where the synergistic binding of two TFs influences gene expression. In contrast to analysis on computationally derived putative motifs, our approach exploits direct evidence of where the TFs are binding, to explore the TFs' control over gene expression.

Our results indicate that many cooperative TFs that were previously characterized experimentally indeed have a high cooperativity in our analysis, thus validating the use of this measure as a predictor of potential cooperativity. In addition, we propose several novel predictions of cooperative TFs that might play a role in the cell cycle and other biological processes, hence generating experimentally testable hypotheses. The target genes of the combination of TFs can be assigned a level of confidence in terms of functional relevance based on known binding motifs and expression profiles of known targets.

Integration of diverse genome-wide resources and analysis of regulatory mechanisms, as described above, critically hinges on the accessibility of a computational platform complete with analytical and visualization tools. To that end, we have created a web tool to facilitate exploring the exhaustive list of cooperative TFs and surveying the graphical representation of the target genes' expression profiles. Furthermore, the tool allows dynamic generation of circular cooperativity networks with user-defined significance levels.

MATERIALS AND METHODS

Selecting sets of target genes of TF combinations

For all pairs of TFs A and B, we identified three sets of target genes from Lee *et al.*'s (5) genome-wide binding data of 113 yeast regulators. The three sets are selected based on significant binding (binding P-value or $P_B < 0.001$ as described in Lee *et al.*'s paper) of TF A but not B ($A \cap \overline{B}$), TF B but not A ($\overline{A} \cap B$), and both TFs A and B ($A \cap B$), respectively. Each set was required to have at least five genes.

Calculating expression correlation score

To quantify the clustering of expression profiles within a given set of genes we calculate an expression correlation score similar to that used by Pilpel *et al.* (3). We define this expression correlation score, EC_G, as the fraction of gene pairs in a given set G with correlation higher than a threshold λT. The threshold correlation λT was determined to be the 95th percentile correlation value of all the pairwise correlations between 1000 randomly chosen genes from Cho *et al.*'s (9) cell cycle gene expression data.

Assessing significant transcription factor cooperativity

A combination of TFs were considered cooperative if the expression correlation score of genes showing binding of both TFs was significantly greater than any set of genes with binding of either TF alone. To accentuate any existing difference between the two cases mentioned above, we used a model based on multivariate hypergeometric distribution.

We calculate:

$$P(m_a, m_{ab}, m_b \mid n_a, n_{ab}, n_b, N, M) = \frac{C_{m_a}^{n_a} C_{m_{ab}}^{n_{ab}} C_{m_b}^{n_b}}{C_M^N}$$

where,

$$C_M^N = \binom{N}{M} = \frac{N!}{M!(N-M)!},$$

n_a = number of gene pairs in set $A \cap \overline{B}$, n_b = number of gene pairs in set $\overline{A} \cap B$, n_{ab} = number of gene pairs in set $A \cap \overline{B}$, m_a = number of correlated gene pairs in set $\overline{A} \cap B$, m_b = number of correlated gene pairs in set $A \cap B$, m_{ab} = number of correlated gene pairs in set $A \cap B$, $N = n_a + n_b + n_{ab}$, $M = m_a + m_b + m_{ab}$.

Then, we consider all possible combinations of x_a, x_b, x_{ab}, such that

$$\sum_{i=\{a,b,ab\}} x_i = \sum_{i=\{a,b,ab\}} m_i$$

and sum all probabilities calculated as above where $x_{ab} > m_{ab}$.

$$P_C = P(x_{ab} \geq m_{ab}) = \sum_{x_{ab} \geq m_{ab}} P(m_a, m_{ab}, m_b \mid n_a, n_{ab}, n_b, N, M)$$

In Table 1, we list all combinations of TFs with cooperativity P-values (P_c) < 0.05.

We evaluate multiple hypothesis tests and therefore the cooperativity P-values need to be corrected to represent the true alpha level. After the correction we have at most 0.05 probability of having at least five false positives. For the adjustment of the P-values we use Bonferroni correction and Holm's correction with 261 as the number of true null hypotheses.

Ranking ChIP-enriched target genes according to functional relevance

The binding motif matrix of Mcm1 and Fkh2, obtained from Spellman *et al.* (10) was scanned across the promoters (500 bp upstream of ATG) of each of the Mcm1 and Fkh2 ChIP-positive genes ($P_B < 0.0001$). A core expression profile was created based on some of the known targets: *ACE2*, *SWI5*, *CLB2*, *CDC20* and *BUD4*. Then, the correlation of this core profile to each ChIP-positive target gene's expression profile was determined. Finally, for each ChIP-positive gene the maximum motif score was plotted against correlation to core expression profile.

Table 1. Statistically significant cooperative TF pairs ($PB^a < 0.001$ and $P_c^b \leq 0.05$)

	TF1	TF2	PC	Literature evidence	Protein–protein interaction[c]
1	Mbp1	Swi6	9.2E–59	11	Y
2	Mcm1	Ndd1	2.9E–49	17	N
3	Fkh2	Mcm1	1.5E–45	16	N
4	Fkh2	Ndd1	9.2E–21	17	N
5	Hir1	Hir2	4.0E–17	15	Y
6	Pdr1	Smp1	2.0E–09	NA	N
7	Swi4	Swi6	2.9E–08	11	Y
8	Gat3	Pdr1	1.9E–06	NA	N
9	Fhl1	Gat3	1.2E–05	NA	N
10	Nrg1	Yap6	1.2E–04	NA	N
11	Gat3	Msn4	1.5E–04	NA	N
12	Reb1	Skn7	1.5E–04[d]	NA	N
13	Ace2	Reb1	2.5E–04	NA	N
14	Hsf1	Reb1	2.9E–04[e]	NA	N
15	Gal4	Rgm1	5.7E–04	NA	N
16	Gcn4	Sum1	6.7E–04	NA	N
17	Fkh1	Fkh2	6.7E–04	16,18	N
18	Cin5	Nrg1	9.2E–04	NA	N
19	Smp1	Swi5	1.1E–03	NA	N
20	Fkh1	Ndd1	1.3E–03	17	N
21	Arg80	Arg81	3.4E–03	28	Y
22	Msn4	Yap5	4.2E–03	NA	Y
23	Ace2	Swi5	7.3E–03	19	N
24	Cin5	Yap6	7.8E–03	NA	N
25	Stb1	Swi4	1.8E–02	NA	N
26	Arg81	Gcn4	3.2E–02	NA	N
27	Ndd1	Stb1	3.2E–02	NA	N
28	Dal81	Stp1	4.0E–02	NA	N
29	Ace2	Hsf1	4.3E–02	21,23	Y
30	Hsf1	Skn7	4.5E–02	20	N
31	Nrg1	Phd1	5.5E–02	NA	N

[a]*P*B, *P*-value for TF binding to chromatin as described in Lee *et al*. (5).
[b]P_c, cooperativity *P*-value on the hypothesis that the target genes of synergistic pairs of TFs should be significantly correlated compared with non-synergistic pairs of TFs.
[c]Protein–protein interaction data were obtained from MIPS (http://mips.gsf.de/proj/yeast/CYGD/interaction/).
[d]*P*-value cut-off for significant cooperative TF pairs with Bonferroni correction.
[e]*P*-value cut-off for significant cooperative TF pairs with Holm's correction.

RESULTS

Analysis of cooperative TF combinations

We integrated Lee *et al*.'s genome-wide location data (5) with Cho *et al*.'s mitotic cell cycle gene expression data (9) to identify cooperativity among all possible pairwise interactions among 113 TFs. First, given any of the pairs of TFs *A* and *B* in the ChIP-chip data, we identified three mutually exclusive sets of target genes based on significant binding of *A* but not *B* (i.e. $A \cap \overline{B}$), *B* but not *A* (i.e. $\overline{A} \cap B$) and both *A* and *B* (i.e. $A \cap B$). Now, if the gene expression patterns of the target genes of the TF pair ($A \cap B$) are significantly more tightly correlated compared with the expression patterns of the target genes of $A \cap \overline{B}$ and $\overline{A} \cap B$ separately, then it is likely that the presence of both TFs is needed to influence gene expression (Fig. 1). In this case, TFs *A* and *B* will be assumed to share a cooperative relationship. Using the filtering criterion of at least five target genes in each set we obtained 261 TF pairs. Before *P*-value adjustments, 31 out of the 261 TF pairs showed a significant level of cooperativity. Table 1 shows cooperative TF pairs where the *P*-value threshold for TF binding to chromatin (P_B) was <0.001 and the cooperativity *P*-value threshold (P_c) was 0.05 (see Materials and Methods). Since we evaluate multiple hypothesis tests, we have also adjusted the *P*-values to reflect Bonferroni correction and Holm's correction. Table 1 also shows the subset of the TF pairs that are significant after Bonferroni correction and Holm's correction (11). In this section, we include discussions on TF pairs from Table 1 that are significantly cooperative after Holm's correction and those that were not significant after *P*-value adjustment but had biological evidence to suggest potential cooperativity. As expected, many of the cooperative TF pairs belong to the cell cycle but others span the environmental stress response, pseudohyphal differentiation and metabolism. There are some TF pairs that do not belong to any well defined functional group and would need to be further characterized. In Table 1, we also include evidence of protein–protein interaction for the TF pair as indicated by the MIPS database.

Cell cycle

The G_1 to S phase transition of the eukaryotic cell cycle is a critical point for the coordination of cell cycle progression with cellular growth. This transition is mediated by MluI cell cycle box binding factor (MBF) and Swi4 binding factor (SBF). MBF is composed of Mbp1 and Swi6 while SBF is composed of Swi4 and Swi6. Swi6 does not bind DNA directly and is thought to have regulatory functions (12). As positive controls, our analysis was able to capture this cooperative interaction between the pair Mbp1–Swi6 and the pair Swi4–Swi6.

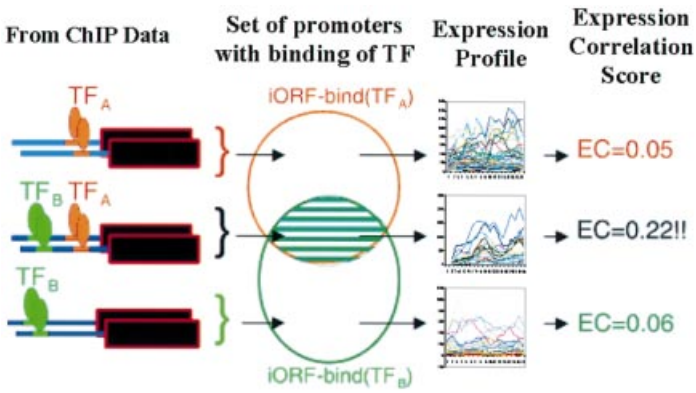

Figure 1. Schematic diagram of the ChIP-based approach of identifying cooperativity. The promoters of the set of genes that show binding of both TFs *A* (TF_A) and *B* (TF_B) show a much higher EC score (0.22) than the promoters of the sets of genes that show binding to only TF *A* (0.05) or TF *B* (0.06). Therefore, TFs *A* and *B* are potentially cooperative.

Sin three-binding protein (Stb1) is also an important regulator of the timing of Start transcription in the absence of CLN3 function and binds to Swi6 *in vitro* (13). Stb1 encodes a protein that was first identified in a two-hybrid screen with the general transcriptional repressor Sin3 even though the role of Stb1 as a transcriptional repressor is still unclear. A recent study suggests that Stb1 may affect MBF-dependent transcription (14). We found Stb1 to be cooperative with Swi4. It is possible that Stb1 regulates SBF in an alternate pathway. Many of the Stb1 ∩ Swi4 ChIP-positive target genes (12/18) also show Mbp1 binding and are mostly G_1 specific (10/18). In addition, we find that Stb1 is cooperative with Ndd1. Ndd1, a ChIP-chip positive target of Swi4, is necessary for the nuclear division process and NDD1 mRNA is most abundant in the S phase (15). The targets of Stb1 ∩ Ndd1 also show binding to Mbp1, Swi4 and Fkh2. The role of Stb1 and Ndd1 in the late S phase will be worth exploring experimentally.

Also in the S phase, Hir1 and Hir2 are known to participate in transcriptional repression through a DNA-binding protein that contacts the negative site at each of the negatively regulated histone gene promoters (16). Cooperativity between Hir1 and Hir2 was in agreement with our results. The other TFs binding to the targets of Hir1 and Hir2 (*HHF1*, *HHF2*, *HHT1*, *HHT2*, *HTB1*, *HTA1*) were Swi4 and Fkh2. Our results indicate that Swi4 and/or Fkh2 could be the DNA-binding protein through which Hir1 and Hir2 achieves specificity.

We also identified the strong cooperative relationship between Mcm1 and Fkh2—known to be important regulators in the G_2/M phase of the cell cycle (17,18). Besides Mcm1 and Fkh2, the other significant cooperative TF pairs in the G_2/M phase include Ndd1–Mcm1. Ndd1 is recruited by Fkh2 (19) in the G_2 stage so its apparent cooperativity with Mcm1 is not surprising. Interestingly, Mcm1 ∩ Fkh2 ∩ Ndd1 target genes do not show cooperativity. This suggests that the three-way interaction is crucial for G_2/M phase-specific cell cycle regulation.

In M/G_1, we found the Ace2 and Swi5 cooperativity to be consistent with the standard model (20). Even though both TFs have been shown to cooperate to induce the expression of a set of target genes, sometimes one partner can have a stronger contribution towards regulation. For example, expression of RME1 (a target of Ace2 and Swi5) is reduced in either Ace2 or Swi5 deletions, but loss of Ace2 has the more dramatic effect (19).

Stress response

Apart from cooperativity among cell cycle-related regulators, our analysis also detects cooperative interactions among regulators representing different biological processes. It is known that Hsf1 and Skn7 cooperate to achieve maximal induction of particular stress-responsive genes (21). We identify an Hsf1–Skn7 synergy but we also find the following cooperative pairs: Ace2–Hsf1, Skn7–Reb1 and Ace2–Reb1. Could there be a functional role of Ace2 and/or Reb1 in assisting Hsf1 and Skn7 during stress response? It has been noted that a protein other than Hsf1 and Skn7 can bind to HSEs *in vitro*, is localized to the nucleus under normal and oxidative stress growth conditions and is required for the full induction of heat shock genes in response to oxidative stress (22). Recent genetic and biochemical studies reveal a cell cycle-specific binding of Hsf1 to nucleosomal DNA (23). It is possible that a coactivator of Hsf1 might be cell cycle regulated, inactive during G_1, and reactivated following entry into S phase (24). Given our analysis, we suggest that a combination of Ace2 and Reb1 could be the additional feature that influences full induction of a subset of the Hsf1 and Skn7 target genes and have a cell cycle-related role in M/G_1. Skn7 is known to have a G_1 transcriptional role, specifically in bud emergence mediated by Mbp1 (25). Such a heat shock-related signal is not surprising since in the Cho *et al*. (9) study, all samples were subjected to heat shock when they were transferred to a 37°C water bath.

Other environmental changes can have a great impact on expression of genes too. Smp1 is thought to play an important role not only in osmostress responses, but also in a Hog1 MAPK pathway required for cell survival in the stationary phase (G_0) (26). It is a MADS-box TF (like human TF MEF2A) and exhibits low DNA bending propensities. Pdr1 on the other hand, affects growth on low-iron medium despite normal high affinity iron uptake (27,28). Our analysis suggests cooperativity between Pdr1 and Smp1. The connection between metal metabolism and ion tolerance is not clear. Further experimental analysis will be needed to elucidate the mechanism and the role in the cell cycle of this cooperativity

(seven out of 21 targets of Pdr1 and Smp1 are cell cycle related).

Metabolism

Arginine metabolism is also regulated by DNA binding and bending along with specific interactions with its cofactors. Arg80 and Mcm1, two members of the MADS box family of DNA-binding proteins, regulate the metabolism of arginine in association with Arg81, the arginine sensor (29). Arg80–Arg81 cooperativity in our results is in agreement with this model. However, we cannot detect cooperativity between Mcm1 and the Arg80–Arg81 complex under the given conditions of ChIP-chip and microarray gene expression experiments.

Uncharacterized functional representation

We also find a group of cooperative TF pairs, where the TFs have not been well studied. Their functional relevance in terms of cooperativity is not intuitive and will need further experimental exploration. Gat3, a member of the GATA family of regulators responsible for the selective use of good nitrogen sources in preference to poor ones (30) has been found to be cooperative with Pdr1, Msn4 and Fhl1 (important in the control of rRNA processing). Many of the Gat3–Pdr1 and Gat3–Msn4 targets are overlapping and belong to the M/G_1 or G_1 phase. Yap5, Rgm1, Smp1, Gal4 and Rap1 are some of the TFs that also bind to many of the targets.

Other cooperative TF pairs include Gal4–Rgm1, Gcn4–Sum1, Smp1–Swi5, Msn4–Yap5, Arg81–Gcn4 and Dal81–Stp1. They are likely to be involved in general regulatory rules (both repressive and inductive) in amino acid biosynthesis and metabolism. Fhl1 plays a key role in the control of rRNA processing, presumably by acting as a transcriptional regulator of genes specifically involved in that process (31). But its association with Gat3 is unclear (30). Similarly, Gat3–Pdr1 cooperativity has not been documented before. Though, if there is cross talk between nitrogen metabolism and metal metabolism, Gat3–Pdr1 cooperativity would suggest functional relevance.

Assessing biological relevance of the target genes

Our results also suggest that ChIP-chip data is contributing more strongly towards identifying target genes of biological relevance than previous studies. The related work by Pilpel *et al*. (3) had uncovered a set of 30 target genes with combinations of putative, computationally derived motifs of Mcm1 and Fkh2 in their promoters. We compared this set with the set of 34 target genes of Mcm1 and Fkh2 derived from the ChIP-chip approach utilizing *in vivo* binding evidence and found an overlap of eight genes. To assess the biological relevance of these two sets of target genes of Mcm1 and Fkh2 we examined the extent of overlap of these two sets with the established *CLB2* cluster (10), which is known to be rich in genes with Mcm1 and Fkh2 binding sites. The target genes of our approach had far more overlapping genes with the *CLB2* cluster than the motif- and expression-based approach (Fig. 2).

It is curious that there is such a small number of overlapping target genes between the ChIP-based method and the motif-based method. It is possible that choosing a less stringent *P*-value can potentially identify additional overlapping genes. But this underscores the fact that TFs do not always bind to experimentally established sequence motifs. There might be contextual information near the binding region of the TF, which makes only some of the promoters with known motifs functional. Studying the target genes identified by the ChIP-based method can further help characterize binding specificities.

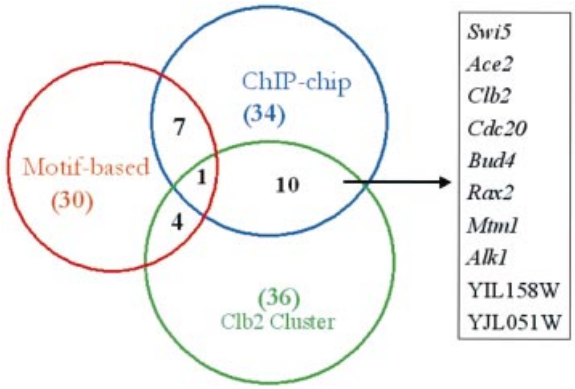

Figure 2. Overlapping ORFs in ChIP, motif and expression analysis. ChIP-based and expression-based approaches appear to identify more experimentally established targets of Mcm1 and Fkh2 than the motif- and expression-based approach described in Pilpel *et al*. (3).

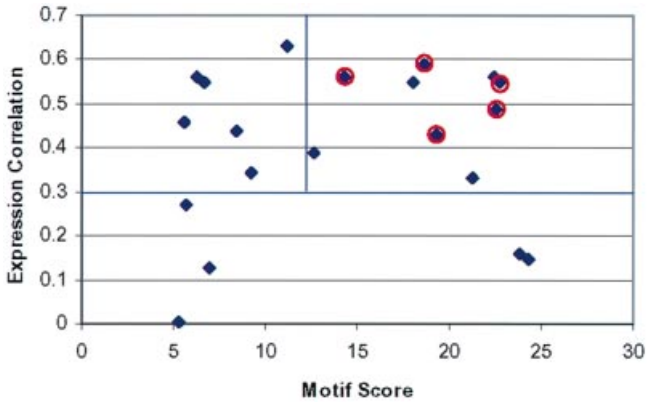

Figure 3. Motif scores and gene expression of ChIP-positive target genes of Mcm1 and Fkh2. The known targets (circled in red) fall in the upper right-hand corner. The genes that are close to the known target genes in expression and motif space are likely to be reliable additional targets of Mcm1 and Fkh2 compared with those genes with low correlation with core expression profile and low motif score.

Among the ChIP-chip positive targets of Mcm1 and Fkh2, many genes are uncharacterized. To get a refined list of target genes that groups the genes according to target specificity of Mcm1 and Fkh2, other functional attributes were incorporated. A core expression profile was created based on expression of some of the known targets (*SWI5*, *CDC20*, *BUD4*, *ACE2* and *CLB2*). The correlation of each ChIP-chip positive target gene's expression to the core expression profile was determined. Also, the binding motif matrix of Mcm1 and Fkh2 (10) was scanned across the promoters of each target gene. For each gene, the correlation value versus the motif score was plotted as shown in Figure 3. The known targets (circled in red) fall in the upper right-hand corner. The genes

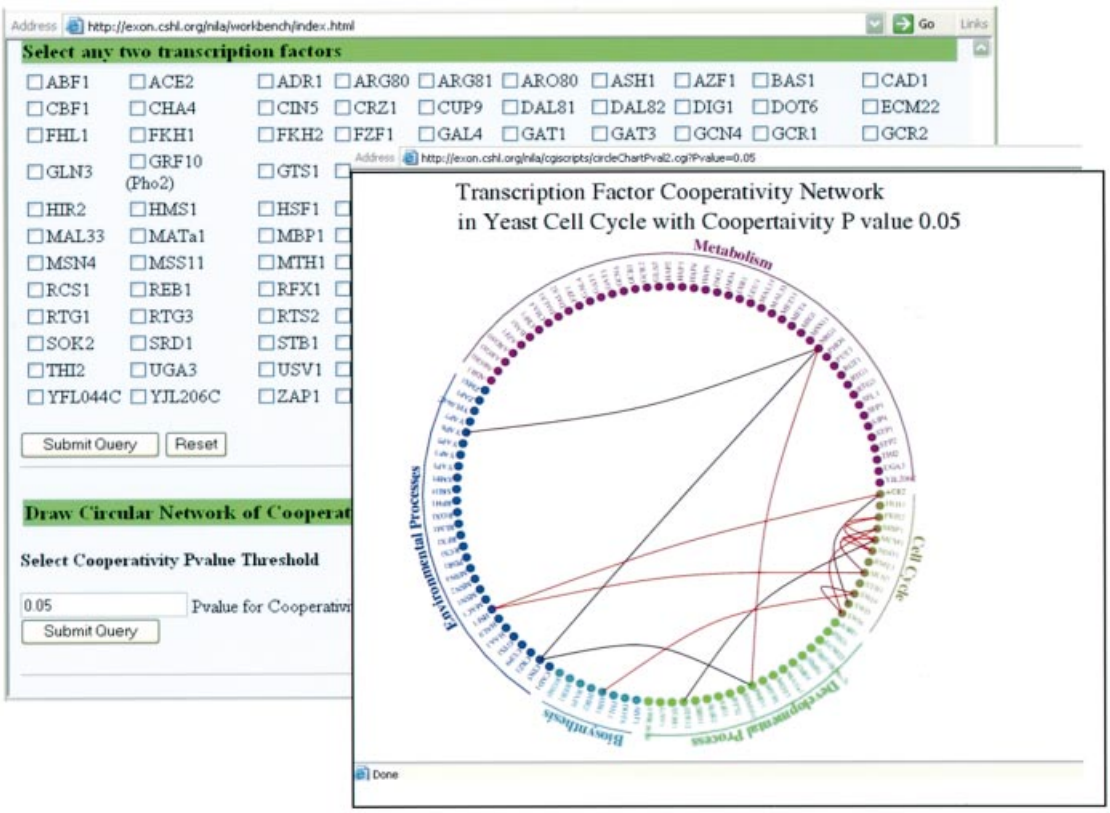

Figure 4. User interface to get a TF combination's target genes and their expression profiles.

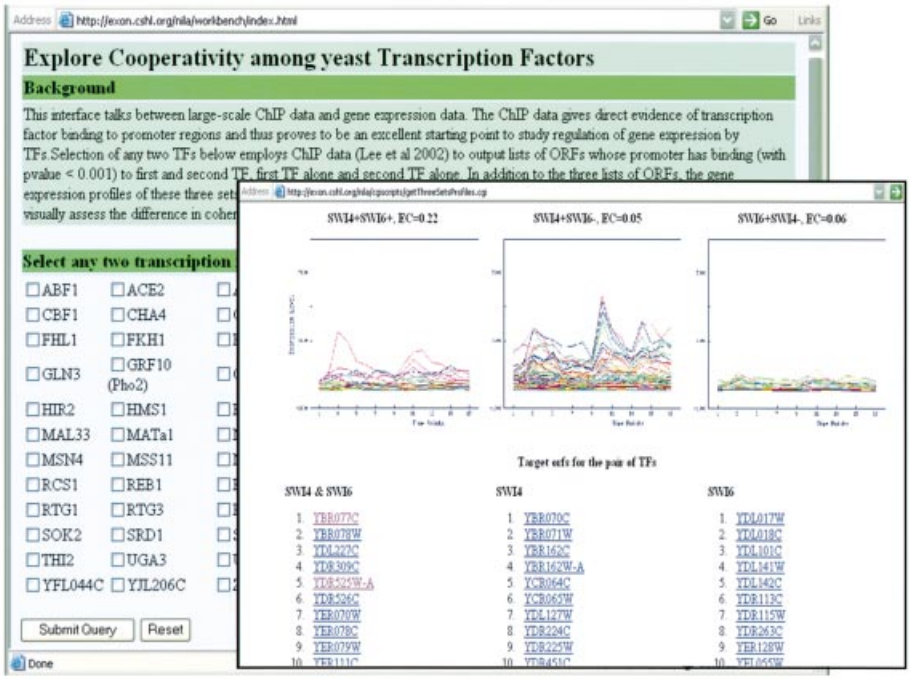

Figure 5. User interface to dynamically generate a TF cooperativity network with user-defined significance values.

that are close to the known target genes in expression and motif space are likely to be reliable additional targets of Mcm1 and Fkh2 (*YJL051W*, *YLR190W*, *YHR151*C, *EKI1*) compared with those genes with low expression correlation with core profile and low motif score (*JSN1*, *PHM5*, *YML053C*, *SPO1* and *YOR315W*).

User interface to explore cooperativity

Large-scale investigation of the interplay between TFs critically hinges on a computational platform where different data sets can be integrated with the aid of analytical and visualization tools. To that end, we have created a web tool at http://exon.cshl.org/nila/workbench/index.html and explore the exhaustive list of cooperative TFs and survey the graphical representation of the target genes' expression profiles. Here researchers can select any pairwise combination of TFs *A* and *B*, and generate three lists of ORFs representing the following combination of TF binding (below a *P*-value of 0.001): TF($A \cap \overline{B}$), TF($A \cap B$) and TF($\overline{A} \cap B$). These sets of ORFs are then used to generate dynamic graphs of their members' gene expression profiles (Fig. 4). This output page gives a stage to visually explore the relationship between pairs of TFs, evaluate how the profiles of the different sets differ from each other and speculate on the biological significance of the results.

In addition, the interface includes a tool to dynamically draw a TF cooperativity network with user-defined significance levels (Fig. 5). The color of a curved edge corresponds to the confidence in cooperativity based on cooperativity *P*-values ranging from 0.05 (black) to 0.0001 (red). The TFs have been clustered according to their functional categories as defined by the MIPS database. The number of edges of the network is determined by the stringency of P_{B}.

DISCUSSION

It is clear that to understand the mechanisms behind transcription regulation of the yeast cell cycle, it is imperative to investigate the cooperative nature of TF interactions. But very few computational approaches of studying genome-wide transcription regulation have addressed the role of cooperativity among TFs. In a recent study, Bussemaker *et al*. (32) described a linear model that captured the additive effect of oligomers on the expression levels of individual genes. But they did not assess the effect of addition or subtraction of motifs on gene expression—an important feature in determining cooperativity. This aspect was elegantly studied by Pilpel *et al*. (3). However, their synergistic motif combinations lack direct evidence of TF binding.

Our method integrated genome-wide gene expression data and binding data, and confirmed many of the known cooperative TFs in the cell cycle. In addition, we propose several novel cooperative TFs in the cell cycle (e.g. Ndd1–Stb1, Ace2–Hsf1) and in other biological processes (e.g. Pdr1–Smp1, etc.). It is interesting that cell cycle regulators interact with a strikingly large number of other protein classes. Many different processes in a cell during cell division have to be precisely coordinated with cell cycle regulators. Such cooperativity suggests cross talk that is essential to coordinate different functions.

More sophisticated analysis will likely reveal different nuances and levels of contribution to cooperativity among the combinations of TFs that we are unable to detect so far. Certain cooperative TF pairs might be a combination of a general regulator and a regulator specific for a process while others could be a combination of an activator and a repressor. It is also possible that a TF pair when bound together to a set of target genes produces coherent gene expression and when bound alone also produces coherent expression profiles for its target genes. In our analysis, this case would be a false negative. While there are some elegant motif-finding algorithms that have incorporated ChIP data (33) (Kato,M., Hata,N., Banerjee,N., Futcher,B. and Zhang,M.Q., manuscript submitted for publication) and explored combinatorial regulation, our method has the strength of identifying cooperative interactions of regulators even when the target sequences are not known or their binding motifs are not distinct. This approach can be extended to combinations of more than two TFs fairly easily.

It should be noted that the list of cooperative TFs is dependent on P_B. With a more stringent P_B (0.0001) the cooperative TFs are mostly limited to the cell cycle with a few novel combinations that include Hir1–Swi4. A less stringent *P*-value (0.01) reveals cooperativity among many general regulators or regulators that inherently bind easily to many promoters (e.g. Rap1 which is involved in rRNA processing). For more details, see Tables 1 and 2 of our Supplementary Material. Interestingly, many of the cell cycle-related cooperative TF pairs appear largely insensitive to a wide range of P_B cut-offs.

Our study can be useful for several reasons. First, it demonstrates that when multiple genomic resources are combined the result is an increasingly detailed picture of cooperativity among TFs. Secondly, it ensures a low number of false positives in identifying the target genes of these cooperative TFs. Thirdly, it provides an interface to visually explore cooperativity. In future, this interface will enable researchers to upload their own gene expression and binding data and automatically generate TF network diagrams depicting the relationships between TFs.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to Bruce Futcher, Gengxin Chen, Ivo Grosse, Naoya Hata, Mamoru Kato, John Grefenstette, James Willett, Curtis Jamison and Daniel Carr for helpful comments and discussions. This work has been supported by NIH Grant R01GM060513.

REFERENCES

1. Miller,J.A. and Widom,J. (2003) Collaborative competition mechanism for gene activation *in vivo*. *Mol. Cell. Biol.*, **23**, 1623–1632.
2. Wagner,A. (1999) Genes regulated cooperatively by one or more transcription factors and their identification in whole eukaryotic genomes. *Bioinformatics*, **15**, 776–784.
3. Pilpel,Y., Sudarsanam,P. and Church,G. (2001) Identifying regulatory networks by combinatorial analysis of promoter elements. *Nature Genet.*, **29**, 153–159.
4. Hannenhalli,S. and Levy,S. (2002) Predicting transcription factor synergism. *Nucleic Acids Res.*, **30**, 4278–4284.
5. 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 *S. cerevisiae*. *Science*, **298**, 799–804.

6. Itamar,S., Barnett,J., Hannett,N., Harbison,C., Rinaldi,N., Volkert,T.L., Wyrick,J., Zeitlinger,J., Gifford,D., Jaakkola,T. and Young,R. (2001) Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell*, **106**, 697–708.
7. Horak,C.E., Luscombe,N.M., Qian,J., Bertone,P., Piccirrillo,S., Gerstein,M. and Snyder,M. (2002) Complex transcriptional circuitry at the G_1/S transition in *S. cerevisiae*. *Genes Dev.*, **16**, 3017–3033.
8. Banerjee,N. and Zhang,M.Q. (2002) Functional genomics as applied to mapping transcription regulatory networks. *Curr. Opin. Microbiol.*, **5**, 313–317.
9. Cho,R.J., Campbell,M.J., Winzeler,E.A., Steinmetz,L., Conway,A., Wodicka,L., Wolfsberg,T.G., Gabrielian,A.E., Landsman,D., Lockhart,D.J. and Davis,R.W. (1998) A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell*, **2**, 65–73.
10. Spellman,P.T., Sherlock,G., Zhang,M.Q., Iyer,V.R., Anders,K., Eisen,M.B., Brown,P.O., Botstein,D. and Futcher,B. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell*, **9**, 3273–3297.
11. Koch,C., Moll,T., Neuberg,M., Ahorn,H. and Nasmyth,K. (1993) A role for the transcription factors Mbp1 and Swi4 in progression from G_1 to S phase. *Science*, **261**, 1551–1557.
12. Holm,S. (1979) A simple sequentially rejective multiple test procedure. *Scand. J. Stat.*, **6**, 65–67.
13. Yo,H., Costanzo,M., Moore,L., Kobayashi,R. and Andrews,B.J. (1999) Regulation of transcription at the *S. cerevisiae* start transition by Stb1, a Swi6-binding protein. *Mol. Cell. Biol.*, **19**, 5267–5278.
14. Costanzo,M., Schub,O. and Andrews,B.J. (2003) G_1 transcription factors are differentially regulated in *Saccharomyces cerevisiae* by the Swi6-binding protein Stb1. *Mol. Cell. Biol.*, **23**, 5064–5077.
15. Loy,C.J., Lydall,D. and Surana,U. (1999) Ndd1, a high-dosage suppressor of cdc28-1N, is essential for expression of a subset of late-S-phase-specific genes in *S. cerevisiae*. *Mol. Cell. Biol.*, **19**, 3312–3327.
16. Spector,M.S., Raff,A., DeSilva, Heshani., Lee,K. and Osley,M.A. (1997) Hir1p and Hir2p function as transcriptional corepressors to regulate histone gene transcription in the *S. cerevisiae* cell cycle. *Mol. Cell. Biol.*, **17**, 545–552.
17. Kumar,R., Reynolds,D.M., Shevchenko,A., Shevchenko,A., Goldstone,S.D. and Dalton,S. (2000) Forkhead transcription factors, Fkh1p and Fkh2p, collaborate with Mcm1p to control transcription required for M-phase. *Curr. Biol.*, **10**, 896–906.
18. Koranda,M., Schleiffer,A., Endler,L. and Ammerer,G. (2000) Forkhead-like transcription factors recruit Ndd1 to the chromatin of G_2/M-specific promoters. *Nature*, **406**, 94–98.
19. Zhu,G., Spellman,P.T., Volpe,T., Brown,P.O., Botstein,D., Davis,T.N. and Futcher,B. (2000) Two yeast Forkhead genes regulate the cell cycle and pseudohyphal growth. *Nature*, **406**, 90–94.
20. Doolin,M.-T., Johnson,A., Johnston,L. and Butler,G. (2001) Overlapping and distinct roles of the duplicated yeast transcription factors Ace2p and Swi5p. *Mol. Microbiol.*, **40**, 422–432.
21. Raitt,D.C., Johnson,A.L., Erkine,A.M., Makino,K., Morgan,B., Gross,D.S. and Johnston,L.H. (2000) The Skn7 response regulator of *S.cerevisiae* interacts with Hsf1 *in vivo* and is required for the induction of heat shock genes by oxidative stress. *Mol. Biol. Cell*, **11**, 2335–2347.
22. Luft,J.C., Benjamin,I.J., Mestril,R. and Dix,D.J. (2001) Heat shock factor 1-mediated thermotolerance prevents cell death and results in G_2/M cell cycle arrest. *Cell Stress Chaperones*, **6**, 326–336.
23. Nakai,A. and Ishikawa,T. (2001) Cell cycle transition under stress conditions controlled by vertebrate heat shock factors. *EMBO J.*, **20**, 2885–2895.
24. Venturi,C.B., Erkine,A.M. and Gross,D. (2000) Cell cycle-dependent binding of yeast heat shock factor to nucleosomes. *Mol. Cell. Biol.*, **20**, 6435–6448.
25. Bouquin,N., Johnson,A.L., Morgan,B.A. and Johnston,L.H. (1999) Association of the cell cycle transcription factor Mbp1 with the Skn7 response regulator in budding yeast. *Mol. Biol. Cell*, **10**, 3389–3400.
26. de Nadal,E., Casadome,L. and Posas,F. (2003) Targeting the MEF2-like transcription factor Smp1 by the stress-activated Hog1 mitogen-activated protein kinase. *Mol. Cell. Biol.*, **23**, 229–237.
27. Tuttle,M.S., Radisky,D., Li,L. and Kaplan,J. (2003) A dominant allele of PDR1 alters transition metal resistance in yeast. *J. Biol. Chem.*, **278**, 1273–1280.
28. Mamnun,Y.M., Pandjaitan,R., Mahe,Y., Delahodde,A. and Kuchler,K. (2002) The yeast zinc finger regulators Pdr1p and Pdr3p control pleiotropic drug resistance (PDR) as homo- and heterodimers *in vivo*. *Mol. Microbiol.*, **46**, 1429–1440.
29. Jamai,A., Dubois,E., Vershon,A.K. and Messenguy,F. (2002) Swapping functional specificity of a MADS Box protein: residues required for Arg80 regulation of arginine metabolism. *Mol. Cell. Biol.*, **22**, 5741–5752.
30. Cox,K.H., Pinchak,A.B. and Cooper,T.G. (1999) Genome-wide transcriptional analysis in *S. cerevisiae* by mini-array membrane hybridization. *Yeast*, **15**, 703–713.
31. Hermann-LeDenmat,S., Werner,M., Sentenac,A. and Thuriaw,P. (1994) Suppression of yeast RNA polymerase III mutations by FHL1, a gene coding for a fork head protein involved in rRNA processing. *Mol. Cell. Biol.*, **14**, 2905–2913.
32. Bussemaker,H.J., Li,H. and Siggia,E.D. (2001) Regulatory element detection using correlation with expression. *Nature Genet.*, **27**, 167–171.
33. Liu,X.S., Brutlag,D.L. and Liu,J.S. (2002) An algorithm for finding protein–DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. *Nat. Biotechnol.*, **20**, 835–839.