OXFORD UNIVERSITY PRESS | Bioinformatics

Improving computational identification of cooperative transcription factors in yeast using TF-gene direct regulation data

Journal:	*Bioinformatics*
Manuscript ID:	BIOINF-2015-0010
Category:	Original Paper
Date Submitted by the Author:	05-Jan-2015
Complete List of Authors:	Wu, Wei-Sheng; National Cheng Kung University, Electrical Engineering Lai, Fu-Jou; National Cheng Kung University, Engineering Science
Keywords:	Algorithms, Bioinformatics, Systems biology, Transcription factors, Transcriptional regulation, Data mining

Systems biology

Improving computational identification of cooperative transcription factors in yeast using TF-gene direct regulation data

Wei-Sheng Wu[1,*] and Fu-Jou Lai[2]

[1]Department of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan

[2]Department of Engineering Science, National Cheng Kung University, Tainan, Taiwan

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXX

ABSTRACT

Motivation: Transcriptional regulation of gene expression in eukaryotes is usually accomplished by cooperative transcription factors (TFs). Computational identification of cooperative TFs is now a hot research topic and many algorithms have been proposed in the literature. A typical cooperative TFs identification algorithm has two steps: (i) define the regulatory targets of each TF under study and (ii) design a measure for calculating the cooperativity of a TF pair based on the regulatory targets of these two TFs. While different algorithms have distinct sophisticated cooperativity measures, the regulatory targets of a TF are usually defined using ChIP-chip data. However, ChIP-chip data can only provide the binding targets but not necessary the regulatory targets of a TF.

Results: We investigate whether the performance of computational identification of cooperative TFs could be improved by using a more biologically relevant way to define the regulatory targets of a TF. We propose an algorithm which uses experimentally validated TF-gene direct regulation data to define the regulatory targets of a TF. Simulation results show that the performance of computational identification of cooperative TFs is improved when using TF-gene direct regulation data instead of ChIP-chip data to define the regulatory targets of a TF. Strikingly, the proposed simple algorithm outperforms 12 existing sophisticated algorithms which all used ChIP-chip data to define the regulatory targets of a TF. Our study suggests that researchers should put more effort on how to define biologically plausible regulatory targets of a TF rather than totally focus on designing sophisticated cooperativity measures.

Contact: wessonwu@mail.ncku.edu.tw

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

In eukaryotes, cooperativity among several transcription factors (TFs) is known to play an important role in transcriptional regulation. A relatively small number of cooperative TFs can set up very complex spatial and temporal patterns of gene expression. Knowing cooperative TFs is helpful for understanding the mechanisms of transcriptional regulation. Therefore, computational identification of cooperative TFs has become a hot research topic in modern biological research.

Many algorithms have been developed to identify cooperative TF pairs in yeast by integrating multiple high-throughput data sources such as gene expression data, ChIP-chip data, protein-protein interaction data, promoter sequence data, etc. (Balaji *et al*., 2006; Banerjee and Zhang, 2003; Chang *et al*., 2006; Chen *et al*., 2012; Chuang *et al*., 2009; Datta and Zhao, 2008; Elati *et al*., 2007; Harbison *et al*., 2004; He *et al*., 2006; Lai *et al*., 2014a; Nagamine *et al*., 2005; Tsai *et al*., 2005; Wang, 2006; Wang *et al*., 2009; Wu *et al*., 2006; Yang *et al*., 2010; Yu *et al*., 2006). The performance of an algorithm varies under different evaluation criteria (Lai *et al*., 2014b). A typical cooperative TFs identification algorithm has two steps. The first step is to define the regulatory targets of each TF under study and the second step is to design a measure for calculating the cooperativity of a TF pair based on the regulatory targets of these two TFs. While different algorithms propose distinct sophisticated cooperativity measures, the regulatory targets of a TF are usually defined using ChIP-chip data with p-value<0.001. Using this stringent p-value threshold effectively reduces false positives at the expense of ~24% false negatives (Harbison *et al*., 2004). Recently, three algorithms did not use the fixed p-value threshold 0.001 but developed different ways to determine the optimal p-value threshold (Chen *et al*., 2012; Datta and Zhao, 2008; Yang *et al*., 2010). Nevertheless, there is an inherent weakness in using ChIP-chip data to define the regulatory targets of a TF. ChIP-chip analysis can only identify the binding targets of a TF but it cannot distinguish the true regulatory from the binding but non-regulatory targets of a TF (Wu *et al*., 2007).

In this study, we would like to investigate whether the performance of computational identification of cooperative TFs in yeast can be improved by using a more biologically relevant way to define the regulatory targets of a TF. Therefore, we propose an algorithm which uses TF-gene direct regulation data (instead of ChIP-chip data) to define the regulatory targets of a TF.

2 METHODS

2.1 The proposed algorithm

*To whom correspondence should be addressed.

The proposed algorithm consists of two steps. The first step is to define the regulatory targets of each yeast TF under study. The biologically plausible regulatory targets of a TF are defined using TF-gene direct regulation data from the YEASTRACT and YTRP databases (Miguel *et al.*, 2014; Yang *et al.*, 2014). We retrieved 8609 TF-gene direct regulation pairs for 151 distinct yeast TFs. Each TF-gene direct regulation pair is supported by both TF binding evidence and TF regulation evidence in the literature. TF binding evidence means that there exist publications with experimental evidence showing that the TF binds to the promoters of the target genes. TF regulation evidence means that there exist publications with experimental evidence showing that the TF regulates the target genes since the TF perturbation (knockout or over-expression) causes a significant change in the expression of the target genes. In summary, the regulatory targets of a TF defined by TF-gene direct regulation data are of biological relevance since they are supported by both TF binding evidence and TF regulation evidence. On the contrary, the regulatory targets of a TF defined by ChIP-chip data (in most existing algorithms) may not be of biological relevance since ChIP-chip analysis can only identify the binding targets of a TF but it cannot distinguish the true regulatory from the binding but non-regulatory targets of a TF.

The second step of the proposed algorithm is to design a measure for calculating the cooperativity of a TF pair based on the regulatory targets of these two TFs. Since the biological role of a cooperative TF pair is to co-regulate the expression of a set of genes, the number of the common regulatory targets of a cooperative TF pair should be significantly higher than that of a random TF pair. In other words, the overlap of the regulatory targets of a cooperative TF pair should be significantly higher than that of a random TF pair. Therefore, following Garten *et al.*'s approach (Garten *et al*., 2005), our algorithm measures the cooperativity of a TF pair based on the statistical significance of the overlap of the regulatory targets of these two TFs. The statistical significance is computed using the hypergeometric test (Wu and Li, 2008) as follows:

$$p\_\text{value} = P(i \geq m) = \sum_{i=m}^{\min(N_1,N_2)} \frac{\binom{N_1}{i}\binom{G-N_1}{N_2-i}}{\binom{G}{N_2}}$$

where $G = 6575$ is the number of genes in the yeast genome, N_1 is the number of regulatory targets of the first TF, N_2 is the number of regulatory targets of the second TF, m is the number of common target genes of these two TFs. Note that the regulatory targets of each TF are defined in the first step of the proposed algorithm. In summary, the smaller the p-value, the higher the chance that a TF pair has cooperativity.

Since we can define the regulatory targets of 151 yeast TFs using the TF-gene direct regulation data, the cooperativity of 11325 (151*150/2) TF pairs can be computed. These 11325 TF pairs were then sorted by their p-values, where the top one TF pair has the smallest p-value and therefore is the most likely cooperative TF pair. In this study, we consider the top 46 TF pairs with p-values less than 10^{-15} as statistically significant cooperative TF pairs (see Supplementary Table S1).

2.2 Four performance comparison indices

We adopted four existing performance comparison indices from the literature to evaluate the performance of an algorithm in identifying cooperative TF pairs. These four indices are introduced in the following subsections.

2.2.1 Performance index 1: The existence of physical protein-protein interaction (PPI) between two TFs One important mechanism of the cooperativity between two TFs is through direct physical PPI (Chen *et al*., 2012; Wang *et al*., 2009). Therefore, the existence of physical PPI between two TFs may imply that they participate in the same regulatory mechanism. Five existing algorithms have used the existence of physical PPI to evaluate the biological plausibility of a predicted cooperative TF pair (PCTFP) (Banerjee and Zhang 2003; Chen *et al*., 2012; Elati *et al*., 2007; Tsai *et al*., 2005; Wang *et al*., 2009). The physical PPI data were retrieved from the BioGRID database (Chatr-Aryamontri *et al.*, 2013). Here we use the proportion of PCTFPs which have physical PPIs among all PCTFPs from an algorithm to evaluate the performance of an algorithm. The higher the proportion, the better the performance of an algorithm.

2.2.2 Performance index 2: The shortest path length of two TFs in the physical PPI network The study conducted by Aguilar and Oliva (2008) observed that a cooperative TF pair has a shorter path length in the physical PPI network than random expectation. The physical PPI network is constructed using the physical PPI data retrieved from the BioGRID database (Chatr-Aryamontri *et al.*, 2013). Therefore, we use the average of the reciprocals of the shortest path lengths of all PCTFPs from an algorithm to evaluate the performance of an algorithm. The larger the average, the better the performance of an algorithm.

2.2.3 Performance index 3: The functional similarity of two TFs Since a cooperative TF pair co-regulates the expression of a set of genes, they should have similar functions. Actually functional similarity has been used in several previous studies (Datta and Zhao, 2008; Lai *et al*., 2014a; Lai *et al*., 2014b) to evaluate the biological plausibility of a PCTFP. The functional similarity score of a TF pair, which is calculated based on their GO semantic similarity, was retrieved from Yang et al.'s study (Yang *et al*., 2012). Here we use the average of the functional similarity scores of all PCTFPs from an algorithm to evaluate the performance of an algorithm. The larger the average, the better the performance of an algorithm.

2.2.4 Performance index 4: The statistical significance of the overlap with the benchmark set Yang *et al*. (2010) compiled a benchmark set of 27 known cooperative TF pairs from the MIPS transcription complex catalogues (Mewes *et al*., 2002). Then they computed the statistical significance of the overlap of the PCTFPs from an algorithm with the benchmark set to evaluate the performance of an algorithm. The statistical significance (p-value) is calculated using the Fisher exact test. The larger the –log(p-value), the greater the statistical significance. Therefore, the larger the –log(p-value), the better the performance of an algorithm.

3 RESULTS AND DISCUSSION

Adopting four existing performance comparison indices from the literature, we have the following discoveries.

3.1 Using TF-gene direct regulation data instead of ChIP-chip data can improve the performance of computational identification of cooperative TFs

Here we evaluate the performance of our algorithm under two scenarios. The first is using TF-gene direct regulation data and the second is using ChIP-chip data (with four possible *p*-value thresholds) from Harbison *et al*. (2004) to define the regulatory targets of a TF. Figure 1 shows that using TF-gene direct regulation data instead of ChIP-chip data can improve the performance of computational identification of cooperative TFs. This result is robust no matter which performance index is used and how many top PCTFPs are chosen. Our finding suggests that using a more biologically relevant way to define the regulatory targets of a TF indeed helps to identify cooperative TFs.

3.2 Our algorithm outperforms 12 existing sophisticated algorithms under four performance comparison indices

Here we compare the performances of our algorithm and 12 existing algorithms (Banerjee and Zhang 2003; Chang *et al*., 2006; Chen *et al*., 2012; Chuang *et al*., 2009; Datta and Zhao, 2008; Harbison *et al*., 2004; He *et al*., 2006; Nagamine *et al*., 2005; Tsai *et al*., 2005; Wang, 2006; Yang *et al*., 2010; Yu *et al*., 2006) in the literature. The differences between our algorithm and these 12 existing algorithms are as follows. First, our algorithm used TF-gene direct regulation data but the 12 existing algorithms all used ChIP-chip data to define the regulatory targets of a TF. Second, the cooperativity measures proposed by these 12 existing algorithms are much more sophisticated than that of our algorithm. Figure 2 shows that the proposed simple algorithm outperforms these 12 existing sophisticated algorithms. This result is robust no matter which performance index is used. Our finding suggests that defining the biologically plausible regulatory targets of a TF is more important than designing sophisticated cooperativity measures.

3.3 Our algorithm is robust against different *p*-value thresholds for determining statistically significant cooperative TF pairs

In this study, our algorithm set 10^{-15} as the *p*-value threshold and reported 46 PCTFPs whose *p*-values are less than the threshold. In the last subsection, we showed that our algorithm performs better than 12 existing algorithms. To check the robustness of our algorithm against different *p*-value thresholds, we evaluate the performance of our algorithm using three other different *p*-value thresholds (10^{-25}, 10^{-20} and 10^{-10}). Figure 3 shows that no matter which *p*-value threshold is used, the performance of our algorithm always outperforms the 12 existing algorithms. This suggests that our algorithm is indeed robust against different *p*-value thresholds.

3.4 Our algorithm predicts nine novel cooperative TF pairs

In this study, our algorithm set 10^{-15} as the *p*-value threshold and reported 46 PCTFPs whose *p*-values are less than the threshold (see Supplementary Table S1). Among them, nine pairs are novel PCTFPs which have not been predicted by any existing algorithms (see Table 1). Strikingly, four of the nine novel pairs are experimentally validated cooperative TF pairs. For the other five novel pairs, the two TFs of each pair both participate in the same biological process, suggesting that they may co-regulate genes involved in that specific biological process.

Two PCTFPs Ifh1-Sfp1 and Ifh1-Rap1 are noteworthy. These two pairs are the top two most statistically significant (ranked first and second) cooperative TF pairs predicted by our algorithm and they have not been predicted by any existing algorithm. Remarkably, these two PCTFPs are known cooperative TF pairs which have been experimentally validated.

It is known that Sfp1 influences the nuclear localization of Ifh1, which binds to ribosomal protein (RP) gene promoters. The absence of Sfp1 causes Ifh1 to localize to nucleolar regions, thus reducing RP gene transcription (Jorgensen *et al.*, 2004). In addition, the RP gene promoter is characterized by upstream binding of the general TF Rap1 followed by the RP gene specific TF Ifh1 via the forkhead-associated domain of Fhl1 (Wang *et al.*, 2004). The fact that only our algorithm but no existing algorithms can predict the four known cooperative TF pairs Ifh1-Sfp1, Ifh1-Rap1, Ifh1-Fhl1, and Rap1-Tup1 convincingly demonstrates the superiority of our algorithm over the existing algorithms.

4 CONCLUSION

In this study, we investigate whether the performance of computational identification of cooperative TFs could be improved by using a more biologically relevant way to define the regulatory targets of a TF. We propose a simple algorithm which uses TF-gene direct regulation data to define biologically plausible regulatory targets of a TF. Each TF-gene direct regulation relationship is supported by both TF binding evidence and TF regulation evidence in the literature. Our algorithm predicts 9 novel PCTFPs which have not been predicted by any existing algorithms. Remarkably, four of the nine novel pairs are experimentally validated cooperative TF pairs, demonstrating the superiority of our algorithm over the existing algorithms. Moreover, by adopting four existing performance comparison indices from the literature, we have two discoveries. First, the performance of computational identification of cooperative TFs is improved when using TF-gene direct regulation data instead of ChIP-chip data to define the regulatory targets of a TF. This suggests that using a more biologically relevant way to define the regulatory targets of a TF indeed helps to identify cooperative TFs. Second, the proposed simple algorithm outperforms 12 existing sophisticated algorithms which all used ChIP-chip data to define the regulatory targets of a TF. This suggests that defining the biologically plausible regulatory targets of a TF is more important than designing sophisticated cooperativity measures. In conclusion, our study shows that how to define the regulatory targets of a TF in a more biologically relevant way than just using ChIP-chip data is critical for successful identification of cooperative TFs.

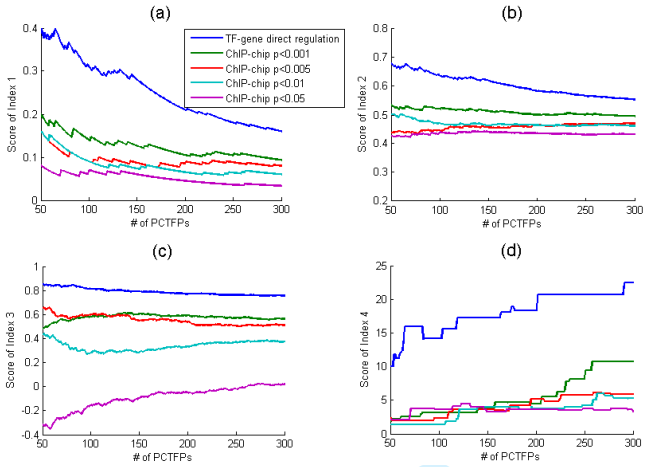

Fig. 1. Performance evaluation of our algorithm using TF-gene direct regulation data or ChIP-chip data (with four possible *p*-value thresholds) from Harbison *et al.* (2004) to define the regulatory targets of a TF. The x-axis is the number of the reported predicted cooperative TF pairs (PCTFPs). For example, 100 means the top 100 PCTFPs in the list of 11325 TF pairs under study. The y-axis is defined as follows. (a) The score of index 1 is the proportion of PCTFPs which have physical PPIs among all PCTFPs from an algorithm. The higher the proportion, the better the performance of an algorithm. (b) The score of index 2 is the average of the reciprocals of the shortest path lengths in the physical PPI network of all PCTFPs from an algorithm. The larger the average, the better the performance of an algorithm. (c) The score of index 3 is the average of the functional similarity scores of all PCTFPs from an algorithm. The larger the average, the better the performance of an algorithm. (d) The score of index 4 is the negative logarithm of the statistical significance (*p*-value) of the overlap of the PCTFPs from an algorithm with the benchmark set. The larger the –log(*p*-value), the better the performance of an algorithm.

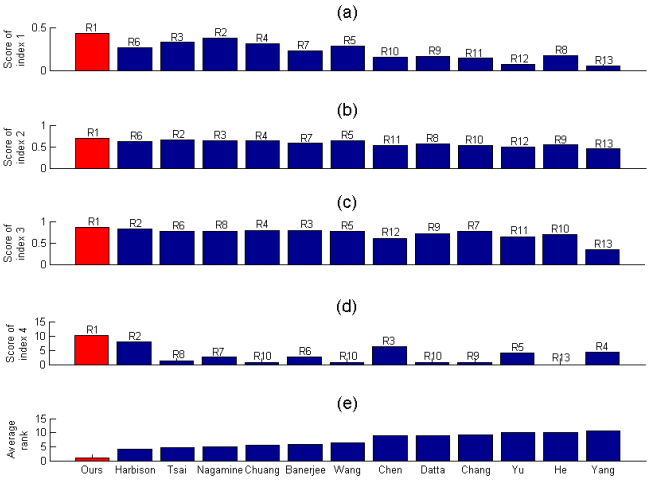

Fig. 2. Performance comparison of our algorithm and 12 existing algorithms using four existing performance indices. The performance comparison results using (a) index 1, (b) index 2, (c) index 3 and (d) index 4 are shown, where Rj means that the algorithm is ranked j among all 13 compared algorithms. For example, our algorithm is ranked first (R1) using the performance index 1 since our algorithm outperforms the other 12 existing algorithms using the performance index 1. (e) The average rank is used to give the overall performance of an algorithm under four different performance indices. The average rank of an algorithm is the average of the ranks of an algorithm under four performance indices. For example, the average rank of our algorithm is $1=(1+1+1+1)/4$ and the average rank of Harbison et al.'s algorithm is $4=(6+6+2+2)/4$. The smaller the average rank, the better the performance of an algorithm. It can be seen that our algorithm has the smallest average rank. Therefore, the overall performance of our algorithm is the best among all the 13 compared algorithms.

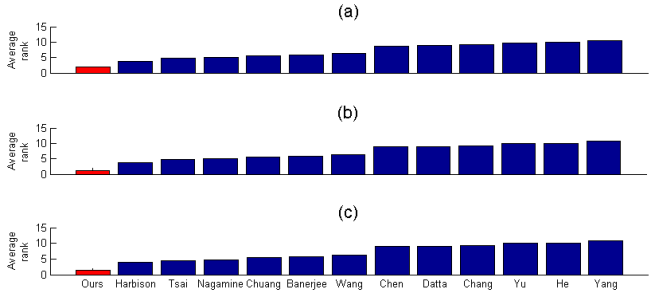

Fig. 3. Robustness analysis of our algorithm. The average rank of our algorithm when using (a) 10^{-25}, (b) 10^{-20} and (c) 10^{-10} as the *p*-value thresholds for determining statistically significant cooperative TF pairs. It can be seen that our algorithm always outperforms the 12 existing algorithms since it always has the smallest average rank no matter which *p*-value threshold is used. This suggests that our algorithm is robust against different *p*-value thresholds.

Table 1. Nine novel PCTFPs which are predicted by our algorithm but not by any existing algorithms

Rank	PCTFP	Experimental evidence	The biological process in which the two TFs both are involved
1	**Ifh1-Sfp1**	Jorgensen *et al.* (2004)	Regulation of ribosomal protein gene transcription
2	**Ifh1-Rap1**	Wang *et al.* (2004)	Regulation of ribosomal protein gene transcription
16	Msn2-Ste12		Stress response
17	Msn2-Tec1		Stress response
20	**Ifh1-Fhl1**	Schawalder *et al.* (2004)	Regulation of ribosomal protein gene transcription
30	Msn2-Pdr1		Stress response
42	Sok2-Ste12		Pseudohyphal growth
44	**Rap1-Tup1**	Roth (1995)	Chromatin-mediated transcription regulation
46	Msn2-Rap1		Stress response

Bold-faced PCTFPs are known cooperative TF pairs which have been experimentally validated in the literature.

ACKNOWLEDGEMENTS

This study was supported by National Cheng Kung University and Ministry of Science and Technology of Taiwan.

Funding: Ministry of Science and Technology of Taiwan [MOST-103-2221-E-006-174-MY2]. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Conflict of Interest: none declared.

REFERENCES

Aguilar,D. and Oliva,B. (2008) Topological comparison of methods for predicting transcriptional cooperativity in yeast. *BMC Genomics*, **9**, 137.

Balaji,S. *et al.* (2006) Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast. *J. Mol. Biol.*, **360**, 213–227.

Banerjee,N. and Zhang,M.Q. (2003) Identifying cooperativity among transcription factors controlling the cell cycle in yeast. *Nucleic Acids Res.*, **31**, 7024–7031.

Chang,Y.H. *et al.* (2006) Identification of transcription factor cooperativity via stochastic system model. *Bioinformatics*, **22**, 2276–2282.

Chatr-Aryamontri,A. *et al.* (2013) The BioGRID Interaction Database: 2013 update. *Nucleic Acids Res.*, **41**, D816–D823.

Chen,M.J.M. *et al.* (2012) *De novo* motif discovery facilitates identification of interactions between transcription factors in *Saccharomyces cerevisiae*. *Bioinformatics*, **28**, 701–708.

Chuang,C.L. *et al.* (2009) Uncovering transcriptional interactions via an adaptive fuzzy logic approach. *BMC Bioinformatics*, **10**, 400.

Datta,D. and Zhao,H. (2008) Statistical methods to infer cooperative binding among transcription factors in *Saccharomyces cerevisiae*. *Bioinformatics*, **24**, 545–552.

Elati,M. *et al.* (2007) LICORN: learning cooperative regulation networks from gene expression data. *Bioinformatics*, **23**, 2407–2414.

Garten,Y. *et al.* (2005) Extraction of transcription regulatory signals from genome-wide DNA-protein interaction data. *Nucleic Acids Res.*, **33**, 605–615.

Harbison,C.T. *et al.* (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature*, **431**, 99–104.

He,D. *et al.* (2006) Identifying synergistic transcriptional factors involved in the yeast cell cycle using Microarray and ChIP-chip data. In *Proceedings of the Fifth International Conference on Grid and Cooperative Computing Workshops*, 357–360.

Jorgensen *et al*. (2004) A dynamic transcriptional network communicates growth potential to ribosome synthesis and critical cell size. *Genes Dev.*, **18**, 2491–2505.

Lai,F.J. *et al.* (2014a) Identifying cooperative transcription factors in yeast using multiple data sources. *BMC Systems Biology*, **8**(Suppl. 5), S2.

Lai,F.J. *et al.* (2014b) A comprehensive performance evaluation on the prediction results of existing cooperative transcription factors identification algorithms. *BMC Systems Biology*, **8**(Suppl. 4), S9.

Mewes,H.W. *et al.* (2002) MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.*, **30**, 31–34.

Miguel,C. *et al.* (2014) The YEASTRACT database: an upgraded information system for the analysis of gene and genomic transcription regulation in Saccharomyces cerevisiae. *Nucleic Acids Res.*, **42**, D161–D166.

Nagamine,N. *et al.* (2005) Identifying cooperative transcriptional regulations using protein-protein interactions. *Nucleic Acids Res.*, **33**, 4828–4837.

Roth,S.Y. (1995) Chromatin-mediated transcriptional repression in yeast. C*urr. Opin. Genet. Dev.*, **5**, 168–173.

Schawalder,S.B. *et al*. (2004) Growth-regulated recruitment of the essential yeast ribosomal protein gene activator Ifh1. *Nature*, **432**, 1058–1061.

Tsai,H.K. *et al.* (2005) Statistical methods for identifying yeast cell cycle transcription factors. *Proc. Natl. Acad. Sci. USA*, **102**, 13532–13537.

Wang,J. (2006) A new framework for identifying combinatorial regulation of transcription factors: a case study of the yeast cell cycle. *J. Biomedical Informatics*, **40**, 707–725.

Wang,J.T. *et al.* (2004) The transcription factor Ifh1 is a key regulator of yeast ribosomal protein genes. *Nature*, **432**, 1054–1058.

Wang,Y. *et al.* (2009) Predicting eukaryotic transcriptional cooperativity by Bayesian network integration of genome-wide data. *Nucleic Acids Res.*, **37**, 5943–5958.

Wu,W.S. *et al.* (2006) Computational reconstruction of transcriptional regulatory modules of the yeast cell cycle. *BMC Bioinformatics*, **7**, 421.

Wu,W.S. *et al.* (2007) Identifying regulatory targets of cell cycle transcription factors using gene expression and ChIP-chip data. *BMC Bioinformatics*, **8**, 188.

Wu,W.S. and Li,W.H. (2008) Systematic identification of yeast cell cycle transcription factors using multiple data sources. *BMC Bioinformatics*, **9**, 522.

Yang,H. *et al.* (2012) Improving GO semantic similarity measures using download random walks. *Bioinformatics*, **28**, 1383–1389.

Yang,T.H. *et al.* (2014) YTRP: a repository for yeast transcriptional regulatory pathways. *Database*, **2014**, bau014.

Yang,Y. *et al.* (2010) Identifying cooperative transcription factors by combining ChIP-chip data and knockout data. *Cell Res.*, **20**, 1276–1278.

Yu,X. *et al.* (2006) Genome-wide prediction and characterization of interactions between transcription factors in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **34**, 917–927.