

# Identifying cooperative transcription factors by combining ChIP-chip data and knockout data

*Cell Research* (2010) 20:1276-1278. doi:10.1038/cr.2010.146; published online 26 October 2010

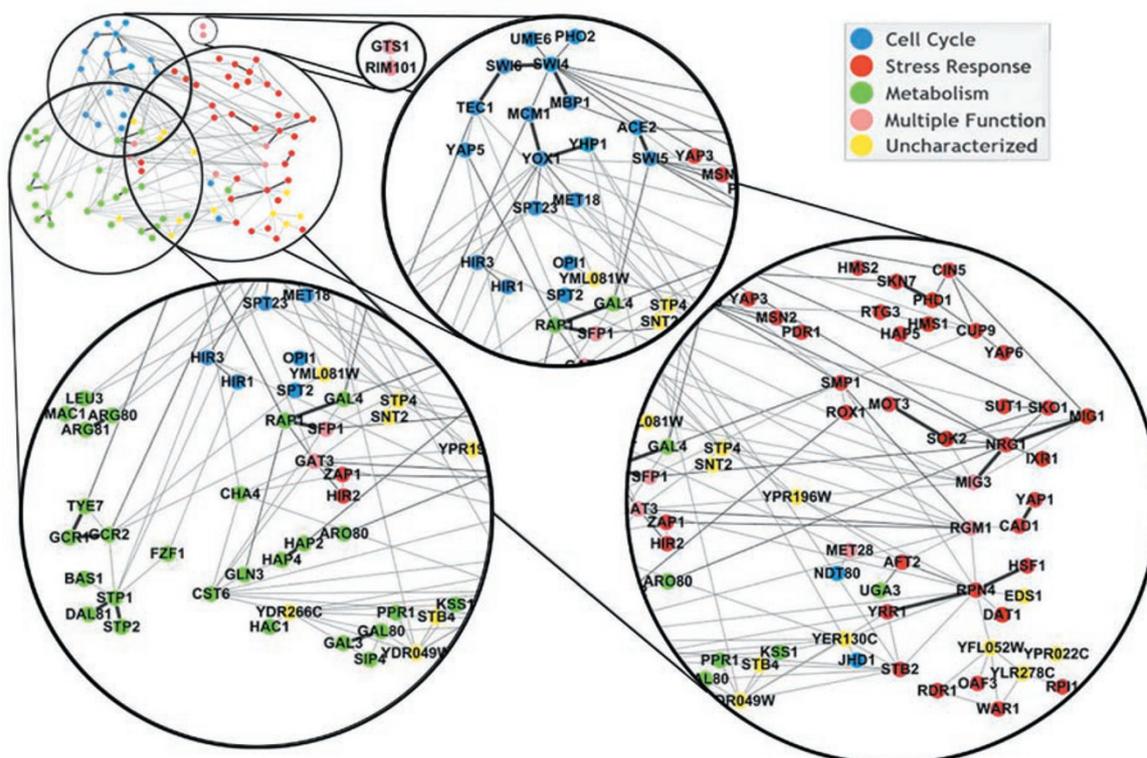
## Dear Editor,

Eukaryotic transcriptional regulation networks are extremely complex. Usually, multiple transcription factors (TFs) bind to the promoter region of a gene and cooperate to control gene expression precisely. Identifying cooperative TFs remains a major challenge in modern biological research. Various types of data, including genomic sequences, expression profiles, ChIP-chip data and protein-protein interactions, have been used to identify mechanisms of cooperative transcriptional regulation. However, because of the noise inherent in these data and the fact that each data source only provides partial information about regulation, combining multiple types of data to improve their ability to infer cooperative TFs is advantageous [1-3].

In our previous work, we successfully integrated ChIP-chip data [4] and expression profiles with individual TF knockout strains [5] to unravel potential relations between TFs and their target genes [6]. This combination of two independent and complementary sources of data improved the accuracy of our prediction. Here, we have extended the work to identify cooperativity between TFs in *Saccharomyces cerevisiae*. We achieved high prediction performance by identifying the most statistically significant overlap of target genes regulated by two TFs in ChIP-chip data and TF knockout data. In addition, we attempted to find the appropriate point to which extent the threshold should be relaxed by looking at the increasing number of cooperative TFs identified within different threshold ranges. Finally, identified TF pairs were ranked using Fisher's combined probability test [7] by combining two independent *P*-values calculated from the ChIP-chip and knockout data (METHODS, Supplementary information, Data S1).

This analysis identified 186 cooperative TFs. The identified cooperative TFs, the *P*-value calculated from the ChIP-chip data and knockout data, the combined *P*-value and any previous experimental and computational evidence are listed in Supplementary information, Data S1-Table S1. Figure 1 shows the cooperative network of

TFs, which are colored and clustered according to their functions. This network suggests that different biological processes, such as the cell cycle, stress response pathways and metabolism, are closely connected to each other. We were pleased to find that many previously characterized cooperative TFs showed highly significant cooperativity measures in our results. Of the top 20 predicted pairs with characterized TFs, 16 pairs have been reported in the literature (Supplementary information, Data S1-Table S1) and 9 of these have been experimentally validated (Supplementary information, Data S1-Table S1). For example, SWI4-SWI6, ACE2-SWI5 and MBP1-SWI4 are known cooperative TFs that control the cell cycle. DAL81 facilitates the binding of STP1 to SPS sensor-regulated promoters [8]. The galactose-activated transcription of *GAL* genes occurs when GAL3 binds GAL80 [9]. Another seven pairs, TEC1-TYE7, HIR3-YOX1, SPT23-YOX1, GAT3-RAP1, ACE2-MBP1, GAT3-RGM1 and RAP1-YAP5, have not been experimentally validated; however, they are supported by numerous computational studies (Supplementary information, Data S1-Table S1). For the remaining four pairs, GTS1-RIM101, RPN4-STB2, CHA4-GAT3 and STP4-TEC1, the potential for cooperativity can be inferred from the literature. For example, RPN4-STB2 together with another two pairs, STB2-YRR1 and RPN4-YRR1 (not in the top 20 but listed in Supplementary information, Data S1-Table S1), form a cooperative triad. Researchers have shown the coordinated action of RPN4, PDR3 and YRR1 on the transcriptional activation of *FLR1* when adapting yeast to mancozeb [10]. Both PDR3 and RPD3 control *PDR5* expression [11], indicating their coordinated action, and STB2 has been detected in the protein complex containing RPD3 [12]. Therefore, it is highly probable that RPN4-STB2 and STB2-YRR1 are cooperative. In addition, many predicted cooperative TFs not ranked in the top 20 list have also been experimentally validated; for example, HAP2-HAP4, RPN4-YRR1, STP1-STP2 and YHP1-YOX1 (Supplementary information, Data S1-Table S1). All of these examples suggest that our predicted cooperative TFs are promising



**Figure 1** Cooperative network of TFs. This network is built from the 186 cooperative TFs identified in this study. TFs are colored according to their functions annotated in SGD and CYGD. Blue node: cell cycle; red node: stress response; green node: metabolism; pink node: multiple function; yellow node: uncharacterized. Edges between TFs represent the types of evidence that support the cooperativity. Thick black edge: supported by experimental evidence; thin black edge: supported by computational evidence; thin gray edge: currently unsupported by any evidence.

and interesting subjects for future experiments.

We further compared the power of our method with three existing methods developed by Banerjee and Zhang [1], Nagamine *et al.* [2] and Yu *et al.* [3]. The overlaps between these predictions are low, which may be due to the different sources of data used in each study. We compiled 27 TF pairs from the MIPS transcription complex catalog as our benchmark data set for TF cooperativity (Supplementary information, Data S1-Table S2), which is the only high-quality data set of TF cooperativity currently available. We compared the significance of the overlap of different predictions with this data set using Fisher's exact test. The results showed that our predictions had a more significant overlap with the standard data set than the other three sets of predictions (Supplementary information, Data S1-Table S3), suggesting that the combination of binding and functional data helps improve prediction accuracy.

Using the identified cooperativity between TFs, we predicted functions for 12 uncharacterized TFs: STP4, SNT2, EDS1, STB4, YDR049W, YDR266C, YER130C, YPR196W, YFL052W, YPR022C, YFL278C and YM-

L081W. We assumed that a given TF has a high probability of functioning in the same processes as its cooperative TF partners (Supplementary information, Data S1).

We attribute the reliability of our method to two features. First, ChIP-chip and knockout data are complementary and independent. ChIP-chip data contain information about the binding between a TF and its target(s), whereas TF knockout data provide information about the functional relationship between a TF and the genes it regulates. Thus, by combining the binding and functional data, we can identify TF pairs that both bind to target genes and work as a complex. Second, we used an optimization procedure to calculate the most significant overlap of target genes regulated by two TFs by the stepwise relaxation of *P*-value thresholds. When a stringent *P*-value threshold (0.001) was used, only 20 cooperative TF pairs were identified (Supplementary information, Data S1-Table S4), of which 4 pairs contained uncharacterized TFs. Of the remaining 16 TFs, 13 pairs were supported by literature and only 6 of these pairs were experimentally validated. In comparison, 14 pairs had supporting evidence and 9 of these pairs were experimentally vali-

dated out of the top 16 pairs with characterized TFs in our results. Many well-known cooperative TF pairs were missed when using the stringent threshold, including SWI4-SWI6, GAL3-GAL80 and MCM1-YOX1. When we relaxed the threshold to 0.005 and omitted the optimization, 117 pairs were discovered, of which 44 pairs had evidence, as compared with 68 out of 186 pairs when the optimization was included. Our method also achieved higher Jaccard similarity scores than the method without the optimization (Supplementary information, Data S1-Figure S3). These results suggest that selecting a suitable but not too stringent *P*-value threshold is a feasible way to uncover more interactions and achieve a low false-positive rate. The optimization principle makes sense not only in statistics but also in biology because TFs are independent. Setting the same threshold for each TF does not take this independence into account and thus could exclude some significant cooperative TFs.

In conclusion, our work provides an initial step toward identifying cooperative TFs by integrating binding and functional information in a robust manner with few arbitrary thresholds. We successfully identified many cooperative TFs that had previously been experimentally confirmed. In addition, we identified many novel potentially cooperative TFs that could lead directly to new hypotheses for future experiments. The cooperative TF networks we constructed suggest that intensive cross talk occurs between cell cycle, metabolism, protein synthesis and filamentous growth pathways at the level of transcriptional regulation. If the appropriate knock-out expression profiles and genome-wide location data were available, our method could identify cooperative TFs under different conditions in yeast or other species. Although we focused on cooperativity between TFs, our method would work equally well to detect cooperativity between other regulatory factors whose binding sites can be identified (for example, microRNAs) or between TFs and other regulatory factors. Our program is available on request.

## Acknowledgments

This work was supported by the National Basic Research Program of China (Grant Nos. 2009CB918404 and 2006CB910700), International S&T Cooperation Program of China (Grant No. 2007DFA31040) and the National Natural Science Foundation of China (Grant Nos. 30700154 and 31070746).

Yi Yang<sup>1,\*</sup>, Zili Zhang<sup>2</sup>, Yixue Li<sup>1</sup>, Xin-Guang Zhu<sup>3</sup>, Qi Liu<sup>1,\*</sup>

<sup>1</sup>School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China; <sup>2</sup>Faculty of Computer and Information Science, Southwest University, Chongqing 400715, China; <sup>3</sup>CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy Sciences, Shanghai 200031, China

\*These two authors contributed equally to this work.

Correspondence: Yixue Li<sup>a</sup>, Xin-Guang Zhu<sup>b</sup>, Qi Liu<sup>c</sup>

<sup>a</sup>Tel: +86-21-3420-4348; Fax: +86-21-3420-4573

E-mail: yxli@sibs.ac.cn

<sup>b</sup>Tel: +86-21-5492-0486; Fax: +86-21-5492-0451

E-mail: xinguang.zhu@gmail.com

<sup>c</sup>Tel: +86-21-3420-4348; Fax: +86-21-3420-4573

E-mail: liuqi@sjtu.edu.cn

## References

- 1 Banerjee N, Zhang MQ. Identifying cooperativity among transcription factors controlling the cell cycle in yeast. *Nucleic Acids Res* 2003; **31**:7024-7031.
- 2 Nagamine N, Kawada Y, Sakakibara Y. Identifying cooperative transcriptional regulations using protein-protein interactions. *Nucleic Acids Res* 2005; **33**:4828-4837.
- 3 Yu X, Lin J, Masuda T, *et al.* Genome-wide prediction and characterization of interactions between transcription factors in *Saccharomyces cerevisiae*. *Nucleic Acids Res* 2006; **34**:917-927.
- 4 Harbison CT, Gordon DB, Lee TI, *et al.* Transcriptional regulatory code of a eukaryotic genome. *Nature* 2004; **431**:99-104.
- 5 Hu Z, Killion PJ, Iyer VR. Genetic reconstruction of a functional transcriptional regulatory network. *Nat Genet* 2007; **39**:683-687.
- 6 Cheng H, Jiang L, Wu M, Liu Q. Inferring transcriptional interactions by the optimal integration of ChIP-chip and knock-out data. *Bioinform Biol Insights* 2009; **3**:129-140.
- 7 Fisher RA. *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd, 1970.
- 8 Boban M, Ljungdahl PO. Dal81 enhances Stp1- and Stp2-dependent transcription necessitating negative modulation by inner nuclear membrane protein Asil1 in *Saccharomyces cerevisiae*. *Genetics* 2007; **176**:2087-2097.
- 9 Sil AK, Alam S, Xin P, *et al.* The Gal3p-Gal80p-Gal4p transcription switch of yeast: Gal3p destabilizes the Gal80p-Gal4p complex in response to galactose and ATP. *Mol Cell Biol* 1999; **19**:7828-7840.
- 10 Teixeira MC, Dias PJ, Simoes T, Sa-Correia I. Yeast adaptation to mancozeb involves the up-regulation of FLR1 under the coordinate control of Yap1, Rpn4, Pdr3, and Yrr1. *Biochem Biophys Res Commun* 2008; **367**:249-255.
- 11 Borecka-Melkusova S, Kozovska Z, Hikkel I, Dzugasova V, Subik J. RPD3 and ROM2 are required for multidrug resistance in *Saccharomyces cerevisiae*. *FEMS Yeast Res* 2008; **8**:414-424.
- 12 Kasten MM, Dorland S, Stillman DJ. A large protein complex containing the yeast Sin3p and Rpd3p transcriptional regulators. *Mol Cell Biol* 1997; **17**:4852-4858.

(Supplementary information is linked to the online version of the paper on *Cell Research* website.)