Detecting Cooperativity between Transcription Factors Based on Functional Coherence and Similarity of Their Target Gene Sets

Wei-Sheng Wu*, Fu-Jou Lai

Department of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan

* wessonwu@gmail.com

Abstract

In eukaryotic cells, transcriptional regulation of gene expression is usually achieved by cooperative transcription factors (TFs). Therefore, knowing cooperative TFs is the first step toward uncovering the molecular mechanisms of gene expression regulation. Many algorithms based on different rationales have been proposed to predict cooperative TF pairs in yeast. Although various types of rationales have been used in the existing algorithms, functional coherence is not yet used. This prompts us to develop a new algorithm based on functional coherence and similarity of the target gene sets to identify cooperative TF pairs in yeast. The proposed algorithm predicted 40 cooperative TF pairs. Among them, three (Pdc2-Thi2, Hot1-Msn1 and Leu3-Met28) are novel predictions, which have not been predicted by any existing algorithms. Strikingly, two (Pdc2-Thi2 and Hot1-Msn1) of the three novel predictions have been experimentally validated, demonstrating the power of the proposed algorithm. Moreover, we show that the predictions of the proposed algorithm are more biologically meaningful than the predictions of 17 existing algorithms under four evaluation indices. In summary, our study suggests that new algorithms based on novel rationales are worthy of developing for detecting previously unidentifiable cooperative TF pairs.

Introduction

Transcription factors (TFs) are a kind of proteins whose biological functions are to transcriptionally regulate the expression of their target genes. In eukaryotic cells, transcriptional regulation of gene expression is usually not achieved by a TF alone but by cooperative TFs which function together to precisely control the location, time and amount of gene expression [1–3]. Therefore, knowing cooperative TFs is crucial for studying the molecular mechanisms of transcriptional regulation of gene expression.

Many algorithms have been proposed to identify cooperative TF pairs in yeast [4–20]. Different algorithms are developed based on different rationales and their performances vary under different evaluation criteria [21–24]. For example, two algorithms [4,6] assume that the

genes bound by both TFs of a cooperative TF pair are more co-expressed or closer in the protein-protein interaction network than are genes bound by either TF alone. Another five algorithms [5,11,14,18,20] assume that for a cooperative TF pair, their binding sites have shorter distance, are more co-depleted of nucleosomes or co-occur more often than expected by chance. Some other algorithms [15,16,18,19] assume that the observed number of the shared target genes of a cooperative TF pair is higher than random expectation (see Table 1 for

Table 1. The rationales of 17 existing algorithms.

Authors	The rationale of the existing algorithm for predicting cooperative TF pairs (CTFPs)	# of predicted CTFPs
Banerjee and Zhang [4]	For a CTFP, the genes bound by both TFs should be more co-expressed than are the genes bound by either TF alone.	31
Harbison et al. [5]	For a CTFP, their binding sites should co-occur more often within the same promoters than would be expected by chance.	94
Nagamine et al. [6]	For a CTFP, the genes bound by both TFs should be closer in the protein-protein interaction network than are the genes bound by either TF alone.	24
Tsai et al. [7]	For a CTFP, their interaction effect (estimated using ANOVA) should significantly influence the expression of genes bound by both TFs.	18
Chang et al. [8]	A stochastic system model is developed to assess TF cooperativity.	55
He et al. [9]	The multivariate statistical method, ANOVA, is used to test whether the expressions of the target genes were significantly influenced by the cooperative effect of their TFs.	30
Wang [10]	Pairwise mixed graphical models or Gaussian graphical models are used for identifying combinatorial regulation of TFs.	14
Yu et al. [11]	An algorithm called Motif-PIE is developed for predicting interacting TF pairs based on the co-occurrence of their binding motifs and the distance between the motifs in promoter sequences.	300
Elati et al. [12]	A data mining technique called LICORN is developed for deriving cooperative regulations.	20
Datta and Zhao [13]	Log-linear models are used to study cooperative bindings among TFs.	25
Chuang et al. [14]	For a CTFP, the distance between their binding sites (in the promoter of their common target genes) should be significantly closer than expected by chance.	13
Wang et al. [15]	A Bayesian network framework is presented to reconstruct a high-confidence whole-genome map of transcriptional cooperativity in Saccharomyces cerevisiae by integrating a comprehensive list of 15 genomic features.	159
Yang et al. [16]	CTFPs are predicted by identifying the most statistically significant overlap of target genes regulated by two TFs in ChIP-chip data and TF knockout data.	186
Chen et al. [17]	A method called simTFBS is developed for inferring TF-TF interactions by incorporating motif discovery as a fundamental step when detecting overlapping targets of TFs based on ChIP-chip data.	221
Lai et al. [18]	For a CTFP, (i) the two TFs should have a significantly higher number of common target genes than random expectation and (ii) their binding sites (in the promoters of their common target genes) should tend to be co-depleted of nucleosomes in order to make these binding sites simultaneously accessible to TF binding.	27
Wu and Lai [19]	For a CTFP, the overlap of the targets (defined by TF binding and TF perturbation data) of these two TFs should be higher than random expectation.	50
Spivak and Stormo [20]	For a CTFP, the distribution of nucleotide spacings between their binding sites should be deviated significantly from random expectation.	1399

doi:10.1371/journal.pone.0162931.t001

details). Apart from the above mentioned algorithms which aim to identify cooperative TF pairs in yeast, several advanced algorithms have been proposed to identify cooperative TF pairs in human [25–27].

Although various types of rationales have been used in the existing algorithms, the functional coherence is not yet used. This prompts us to develop a new algorithm based on the functional coherence and similarity of the target gene sets. First, the proposed algorithm assumes that the common target genes of two cooperative TFs have similar functions. This rationale is biologically plausible since co-regulated genes are known to have similar functions [28–30]. Second, the proposed algorithm assumes that two cooperative TFs have similar target gene sets. Since the biological role of two cooperative TFs is to co-regulate the expression of a set of genes, they should have a significant number of shared target genes [5,11,15,16,18,19]. In other words, the target gene sets of two cooperative TFs should be similar to each other.

## Materials and Methods

Data sources

Two data sources were used in this study. First, the experimentally validated target genes of 151 TFs were retrieved from the YEASTRACT database [31]. The association between a TF and its target gene was supported by two types of experimental evidence. One is the TF binding (TFB) evidence from the detailed gene by gene band-shift, foot-printing experiments or the high throughput genome-wide ChIP-chip experiments showing that the TF binds to the promoter of its target gene. The other one is the TF regulation (TFR) evidence from the detailed gene by gene analysis or the genome-wide expression analysis showing that the perturbation (knockout or over-expression) of the TF-encoding gene causes a significant change in the expression of its target gene. Therefore, the target genes of a TF retrieved from the YEASTRACT database are of biological significance since they are validated by two types of experimental evidence.

The second data source used in this study is the functional similarity scores of all gene pairs in yeast retrieved from Yang et al.'s study [32]. Yang et al. proposed an improving Gene Ontology (GO) semantic similarity measure based on downward random walks to calculate the functional similarity score of any gene pair. Their score has been shown to be more biologically meaningful than the other existing functional similarity scores [32].

The proposed algorithm

The proposed algorithm for identifying cooperative TF pairs is based on two rationales (functional coherence and similarity of the target gene sets). First, the proposed algorithm assumes that the common target genes of two cooperative TFs have similar functions. This rationale is biologically plausible since co-regulated genes are known to have similar functions [28–30]. Second, the proposed algorithm assumes that two cooperative TFs have similar target gene sets. Since the biological role of two cooperative TFs is to co-regulate the expression of a set of genes, they should have a significant number of shared target genes [5,11,15,16,18,19]. In other words, the target gene sets of two cooperative TFs should be similar to each other.

Fig 1 depicts the proposed two-step procedure of calculating the cooperativity score of a TF pair (e.g. TF_1-TF_2). The first step is to retrieve the set of TF_1's target genes (denoted as G_1), the set of TF_2's target genes (denoted as G_2) and the set of the common target genes of TF_1 and TF_2 (denoted as G_{12}) from YEASTRACT database [31]. Note that G_1, G_2 and G_{12} are of biological significance since the regulatory associations between a TF and its target genes are validated by two types of experimental evidence (TFB evidence and TFR evidence). The second
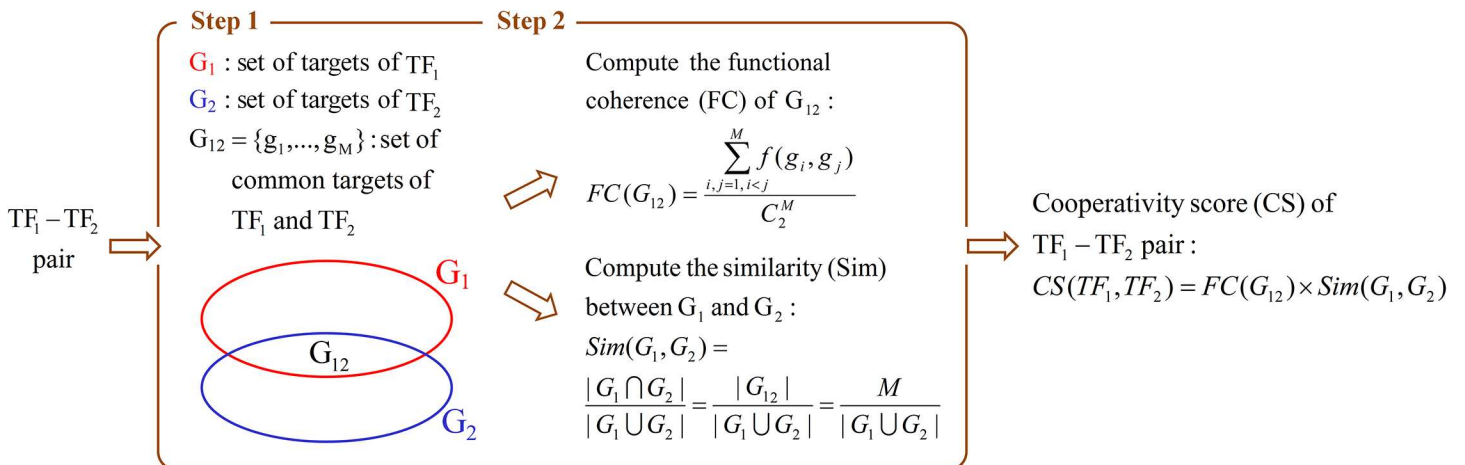
Fig 1. The proposed two-step procedure of calculating the cooperativity score of a TF pair (TF$_1$-TF$_2$).

step is to calculate the cooperative score of TF$_1$-TF$_2$ based on the functional coherence of G$_{12}$ and the similarity between G$_1$ and G$_2$.

The functional coherence (FC) of G$_{12}$ is calculated using the following formula

$$FC(G_{12}) = \frac{\sum_{i,j=1,i<j}^{M} f(g_i, g_j)}{C_2^M}$$

where M is the number of genes in G$_{12}$ = {$g_1, \cdots, g_M$}, C_2^M is all possible gene pairs formed by genes in G$_{12}$, and $f(g_i, g_j)$ is the functional similarity score of g_i and g_j retrieved from Yang et al.'s study [32]. Note that $FC(G_{12})$ is actually the average of the functional similarity scores of all possible gene pairs formed by genes in G$_{12}$. The higher the $FC(G_{12})$ is, the higher the functional coherence of the genes in G$_{12}$ is.

The similarity (Sim) between G$_1$ and G$_2$ is calculated using Jaccard similarity coefficient

$$Sim(G_1, G_2) = \frac{|G_1 \cap G_2|}{|G_1 \cup G_2|} = \frac{|G_{12}|}{|G_1 \cup G_2|} = \frac{M}{|G_1 \cup G_2|}$$

where $M = |G_{12}|$ is the number of genes in G$_{12}$ and $|G_1 \cup G_2|$ is the number of genes in the union of G$_1$ and G$_2$. The higher the $Sim(G_1, G_2)$ is, the higher the similarity between G$_1$ and G$_2$ is. Then the cooperativity score (CS) of TF$_1$-TF$_2$ is calculated using the following formula

$$CS(TF_1, TF_2) = FC(G_{12}) \times Sim(G_1, G_2) = \frac{\sum_{i,j=1,i<j}^{M} f(g_i, g_j)}{C_2^M} \times \frac{M}{|G_1 \cup G_2|}$$

The higher the $CS(TF_1, TF_2)$ is, the higher the cooperativity between TF$_1$ and TF$_2$ is.

Since we can retrieve the experimentally validated target genes of 151 TFs from YEAS-TRACT database [31], the cooperativity scores of 11325 (151*150/2) TF pairs can be calculated. Finally, these 11325 TF pairs are sorted by their cooperativity scores, where the top one TF pair has the highest cooperativity score and therefore is the most plausible cooperative TF pair. That is, the finally output of the proposed algorithm is a ranked list of 11325 TF pairs, where the top one TF pair is the most plausible cooperative TF pair.

Four existing evaluation indices

To judge the biological significance of the set of predicted cooperative TF pairs (PCTFPs) from an algorithm, here we adopt the following four existing evaluation indices.

Index 1: The statistical significance of the overlap with the benchmark set. Yang et al. [16] proposed to evaluate the performance of an algorithm by calculating the significance of the overlap of its set of PCTFPs with a benchmark set of 27 known cooperative TF pairs collected from MIPS transcription complex catalog [33]. The significance of the overlap is represented as −logP, where P is the p-value computed using Fisher exact test [34]. The higher the −logP is, the better the performance of an algorithm is.

Index 2: The co-regulatory coefficient of a PCTFP. Balaji et al. [35] proposed the co-regulatory coefficient to evaluate the biological plausibility of a PCTFP. The co-regulatory coefficient represents the significance of a PCTFP in regulating common target genes. The greater the co-regulatory coefficient is, the higher the biological plausibility of a PCTFP is. To evaluate the biological significance of the set of PCTFPs from an algorithm, we used the average of the co-regulatory coefficients of all PCTFPs from an algorithm. The higher the average is, the better the performance of an algorithm is.

Index 3: The shortest path length of a PCTFP in the physical protein-protein interaction network. Aguilar and Oliva [36] observed that a cooperative TF pair has a shorter path length in the physical protein-protein interaction (PPI) network (using PPI data from BioGRID database [37]) than expected by random. Therefore, the greater the reciprocal of the shortest path length of a PCTFP in the PPI network is, the higher the biological plausibility of a PCTFP is. To evaluate the biological significance of the set of PCTFPs from an algorithm, we used the average of the reciprocals of the shortest path lengths of all PCTFPs from an algorithm. The higher the average is, the better the performance of an algorithm is.

Index 4: The functional similarity of a PCTFP. Lai et al. [21] proposed to evaluate the biological plausibility of a PCTFP by using the functional similarity between the two TFs of a PCTFP. The functional similarity scores between any two TFs were retrieved from Yang et al.'s study [32]. The higher the functional similarity score between the two TFs of a PCTFP is, the higher the biological plausibility of a PCTFP is. To evaluate the biological significance of the set of PCTFPs from an algorithm, we used the average of the functional similarity scores of all PCTFPs from an algorithm. The higher the average is, the better the performance of an algorithm is.

Results and Discussion

From "The proposed algorithm" subsection, it is known that the final output of the proposed algorithm is a ranked list of 11325 TF pairs, where the top one TF pair is the most plausible cooperative TF pair. Here we consider the top 40 TF pairs as the PCTFPs from the proposed algorithm. Considering the top 40 TF pairs is reasonable because the number of the PCTFPs from most (>10) existing algorithms [4,6–10,12–14,18,19] falls between 13 and 60 (see Table 1).

Validation of the 40 PCTFPs from the proposed algorithm

To judge the biological plausibility of each of the 40 PCTFPs from the proposed algorithm, we provide five types of validation (see Table 2 for details). The five types of validation are (i) whether a PCTFP is predicted by any existing algorithm, (ii) whether a PCTFP has physical or genetic interaction, (iii) whether both TFs of a PCTFP are studied in the same experimental publications, (iv) whether a PCTFP has common GO terms, and (v) whether a PCTFP has common target genes.

Table 2. Five types of validation of the 40 PCTFPs from the proposed algorithm.

PCTFP		Evidence of the cooperativity between TF1 and TF2				
TF1	TF2	Algorithm Evidence	Physical/Genetic Evidence	Co-citations	# of Common GO Terms	# of Common Targets
Arg80	Arg81	7	5	47	5	8
Ifh1	Sfp1	1	0	16	4	82
Met28	Met31	2	1	29	8	11
Hap2	Hap4	5	3	100	8	18
Met32	Met4	5	6	43	6	30
Met31	Met32	6	8	54	14	18
Hap3	Hap5	5	5	65	10	4
Met31	Met4	5	5	42	6	14
Met28	Met4	3	7	35	10	13
Pdc2	**Thi2**	0	0	6	5	2
Met28	Met32	3	0	33	9	14
Mig1	Mig2	3	7	67	14	4
Ifh1	Rap1	1	1	22	4	105
Gcr1	Gcr2	3	10	26	7	8
Hap3	Hap4	1	3	93	8	7
Fhl1	Ifh1	1	7	33	7	26
Rap1	Sfp1	7	0	36	6	113
Hap2	Hap3	3	4	118	9	7
Hap4	Hap5	3	2	59	8	6
Aft1	Aft2	5	9	63	6	15
Stp1	Stp2	2	5	40	9	2
Mbp1	Swi6	13	12	147	7	14
Hot1	**Msn1**	0	0	13	3	2
Gal4	Gal80	3	34	185	6	2
Gcr2	Tye7	1	5	7	3	6
Pdr1	Pdr3	5	12	187	10	30
Dal81	Stp2	1	1	17	3	2
Ino2	Ino4	6	13	117	10	10
Cbf1	Met4	5	5	39	7	24
Ace2	Swi5	12	3	99	9	30
Oaf1	Pip2	4	6	57	13	13
Cbf1	Met32	5	1	38	8	23
Dal80	Dal81	1	0	28	7	4
Dal81	Gln3	2	0	23	7	9
Msn2	Sok2	2	2	36	6	150
Ste12	Tec1	6	12	114	9	171
Msn2	Yap1	3	2	114	7	143
Leu3	**Met28**	0	0	13	6	3
Swi4	Swi6	14	29	256	7	21
Bas1	Pho2	3	7	52	6	7

A PCTFP in boldface means that it is a novel CTFP predicted by the proposed algorithm. "Algorithm Evidence" provides the number of existing algorithms which predict the PCTFP. "Physical/Genetic Evidence" provides the number the experimental papers which suggest that the two TFs of the PCTFP have physical or genetic interaction. "Co-citations" provides the number of experimental papers which study the biological roles of both TFs of the PCTFP. More details could be seen at http://cosbi2.ee.ncku.edu.tw/40TFI/.

doi:10.1371/journal.pone.0162931.t002

Overall speaking, the 40 PCTFPs from the proposed algorithm are likely to be biologically meaningful since (i) 93% (37/40) PCTFPs are also predicted by at least one existing algorithm, (ii) 80% (32/40) PCTFPs have physical or genetic interactions, (iii) the two TFs of each of the 40 PCTFPs are studied in the same experimental publications, (iv) 100% (40/40) PCTFPs have common GO terms, and (v) 100% (40/40) PCTFPs have common target genes.

Among the 40 PCTFPs from the proposed algorithm, three (Pdc2-Thi2, Hot1-Msn1 and Leu3-Met28) are novel predictions, which have not been predicted by any existing algorithms. Strikingly, Thi2 is known to act together with Pdc2 to respond to thiaminediphosphate demand [38]. Moreover, it is known that osmotic stress-induced gene expression requires both Hot1 and Msn1 [39]. The fact that two (Pdc2-Thi2 and Hot1-Msn1) of the three novel predictions have been experimentally validated in the literature [38,39] demonstrates the power of the proposed algorithm.

Performance comparison of the proposed algorithm with 17 existing algorithms

Using four existing evaluation indices [16,21,35,36], we evaluate the biological significance of the PCTFPs from the proposed algorithms and those from the 17 existing algorithms. The PCTFPs of the 17 existing algorithms were retrieved directly from the corresponding papers [4–20]. Fig 2 shows that the proposed algorithm has the smallest average rank among the 17 compared algorithms, suggesting that the proposed algorithm is the best performing algorithm.
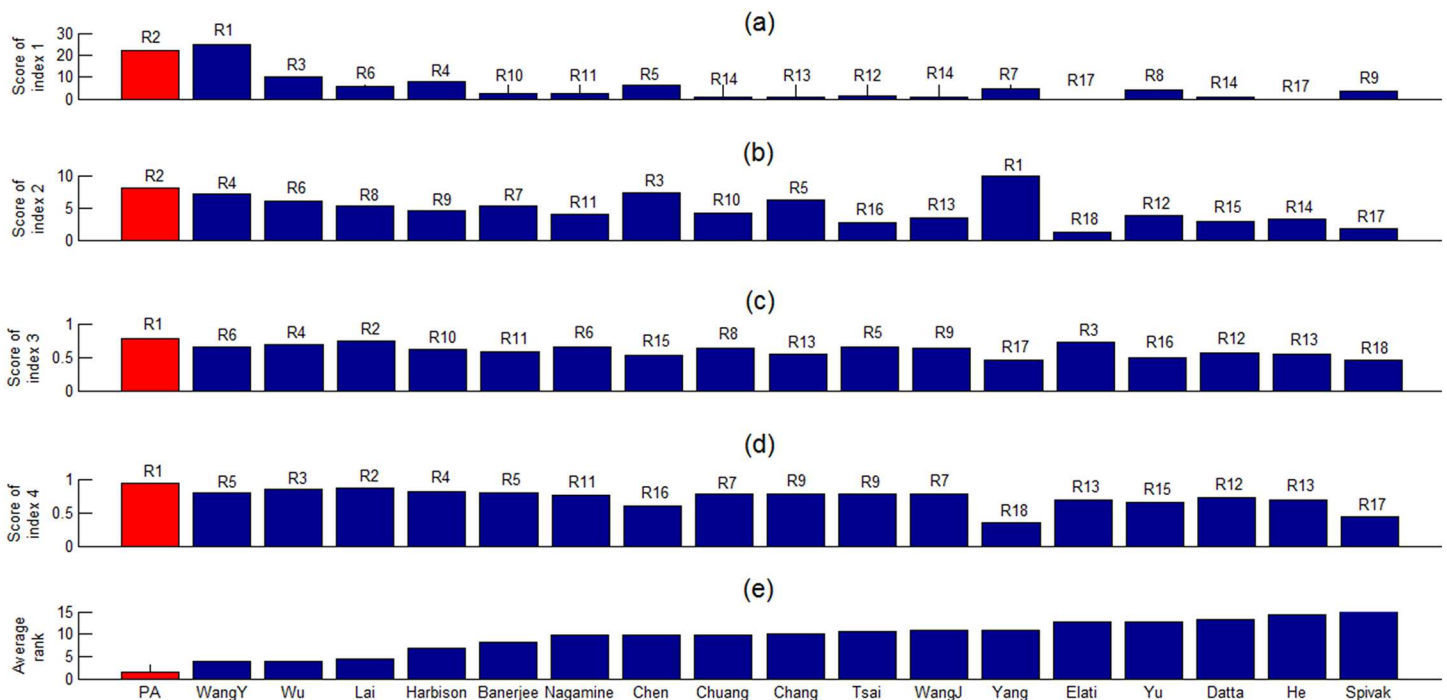
Fig 2. The performance comparison of the proposed algorithm and 17 existing algorithms in the literature. Performance comparison of the proposed algorithm and 17 existing algorithms using four existing evaluation indices. The performance comparison results using (a) index 1, (b) index 2, (c) index 3 and (d) index 4 are shown, where Rj means that the algorithm is ranked j among the 18 compared algorithms. For example, the proposed algorithm ranks first (R1) using the evaluation index 4 since the proposed algorithm has the largest score calculated using index 4. (e) The average rank is used to give the overall performance of an algorithm under four different evaluation indices. The average rank of an algorithm is the average of the ranks of an algorithm under four evaluation indices. For example, the average rank of the proposed algorithm is 1.5 = (2+2+1+1)/4 and the average rank of WangY's algorithm is 4 = (1+4+6+5)/4. The smaller the average rank is, the better the performance of an algorithm is. Since the proposed algorithm has the smallest average rank, the overall performance of the proposed algorithm is the best among all the 18 compared algorithms.
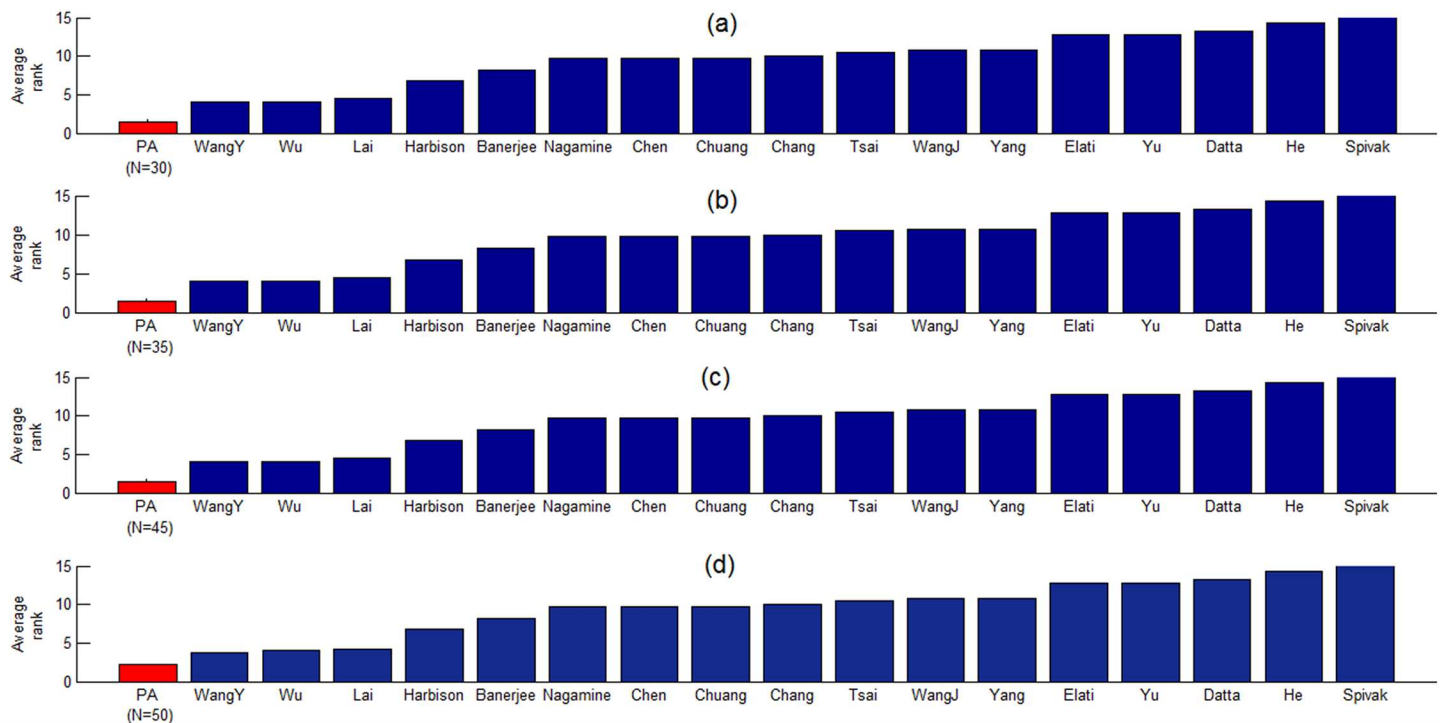
doi:10.1371/journal.pone.0162931.g002

That is, the PCTFPs from the proposed algorithms are more biologically meaningful than are the PCTFPs from the 17 existing algorithms.

Robustness against the number of chosen PCTFPs

In the last subsection, the 40 PCTFPs (i.e. the top 40 TF pairs of the ranked list of 11325 TF pairs) from the proposed algorithm are shown to be more biologically meaningful than those from the 17 existing algorithms in the literature. To check the robustness of the proposed algorithm against the number of chosen PCTFPs, we evaluate the performance of the proposed algorithm when choosing top N (N = 30, 35, 45 or 50) TF pairs as the PCTFPs from the proposed algorithm. Fig 3 shows that no matter which value of N is used, the proposed algorithm always has a smaller average rank than do the 17 existing algorithms in the literature. This suggests that the proposed algorithm is indeed robust against the number of chosen PCTFPs.

Note that our algorithm and most existing algorithms identified less than 100 PCTFPs, but Spivak and Stormo's algorithm [20] identified 1399 PCTFPs (see Table 1). It can be seen in Fig 2, Spivak and Stormo's algorithm performs worst among all the compared algorithms. A possible reason is that their 1399 PCTFPs probably include a large number of false positives. It would be interesting to investigate how the performance of our algorithm evolves with larger N values. As shown in Fig 4, the scores of the four evaluation measures gradually decrease with larger N values, indicating a performance degradation of our algorithm with larger N values. Just like many false positives inside the 1399 PCTFPs from Spivak and Stormo's algorithm, our PCTFPs probably include a large number of false positives with larger N values.

Fig 3. Robustness analysis of the proposed algorithm. The average rank of the proposed algorithm using top N, where (a) N = 30, (b) N = 35, (c) N = 45, and (d) N = 50, TF pairs of the ranked list of 11325 TF pairs as the PCTFPs from the proposed algorithm. It can be seen that no matter which value of N is used, the proposed algorithm always has the smallest average rank. That is, the PCTFPs from the proposed algorithm are always more biologically meaningful than those from the 17 existing algorithms. This suggests that the proposed algorithm is robust against the number of chosen PCTFPs.

doi:10.1371/journal.pone.0162931.g003
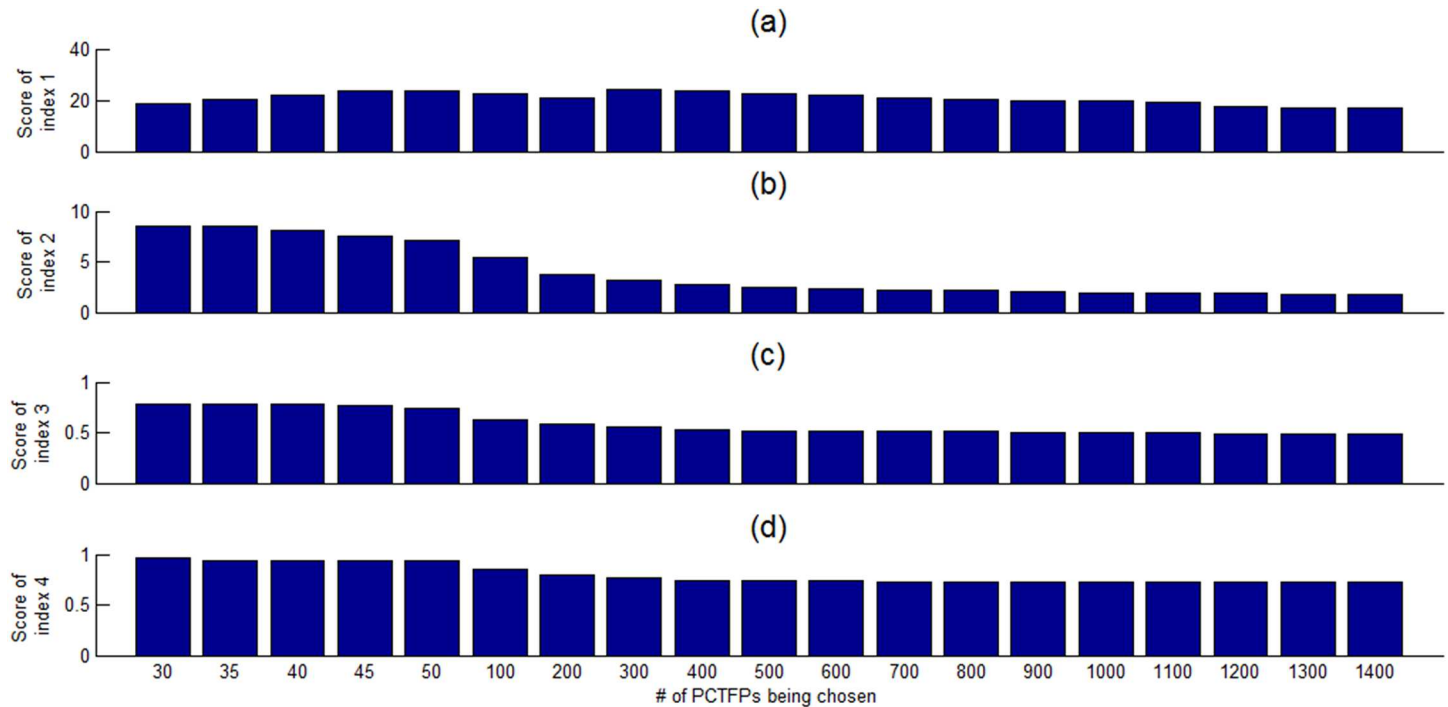
Fig 4. The scores of the four evaluation measures (shown in (a), (b), (c), (d)) with different top N (N = 30, 35, 40, …) chosen for the proposed algorithm.

doi:10.1371/journal.pone.0162931.g004

Conclusions

In this study, we develop a new algorithm based on functional coherence and similarity of the target gene sets to identify cooperative TF pairs in yeast. The proposed algorithm provides 40 predicted cooperative TF pairs (PCTFPs) and the biological significance of the PCTFPs is validated by five types of validation. Among the 40 PCTFPs, three (Pdc2-Thi2, Hot1-Msn1 and Leu3-Met28) are novel predictions, which have not been predicted by any existing algorithms. Strikingly, two (Pdc2-Thi2 and Hot1-Msn1) of the three novel predictions have been experimentally validated in the literature, demonstrating the power of the proposed algorithm. Moreover, we show that the predictions of the proposed algorithm are more biologically meaningful than the predictions of 17 existing algorithms under four evaluation indices. In summary, our study suggests that new algorithms based on novel rationales (e.g. functional coherence) are worthy of developing for detecting previously unidentifiable cooperative TF pairs.

Acknowledgments

Author Contributions

Conceptualization: WSW.

Data curation: WSW FJL.

Formal analysis: WSW FJL.

Funding acquisition: WSW.

Investigation: WSW FJL.

Methodology: WSW FJL.

Project administration: WSW.

Resources: WSW.

Software: WSW FJL.

Supervision: WSW.

Validation: WSW FJL.

Visualization: WSW FJL.

Writing – original draft: WSW.

Writing – review & editing: WSW.

References

1. Tanguay RM. Transcriptional activation of heat-shock genes in eukaryotes. Biochem Cell Biol 1998; 66(6):584–593.

2. Nemer G, Nemer M. Regulation of heart development and function through combinatorial interactions of transcription factors. Ann Med 2001; 33(9):604–610. PMID: 11817655

3. Lelli KM, Slattery M, Mann RS. Disentangling the many layers of eukaryotic transcriptional regulation. Annu Rev Genet 2012; 46:43–68. doi: 10.1146/annurev-genet-110711-155437 PMID: 22934649

4. Banerjee N, Zhang MQ. Identifying cooperativity among transcription factors controlling the cell cycle in yeast. Nucleic Acids Res 2003; 31:7024–7031. PMID: 14627835

5. Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, et al. Transcriptional regulatory code of a eukaryotic genome. Nature 2004; 431:99–104. PMID: 15343339

6. Nagamine N, Kawada Y, Sakakibara Y. Identifying cooperative transcriptional regulations using protein-protein interactions. Nucleic Acids Res 2005; 33:4828–4837. PMID: 16126847

7. Tsai HK, Lu HHS, Li WH. Statistical methods for identifying yeast cell cycle transcription factors. Proc Natl Acad Sci USA 2005; 102:13532–13537. PMID: 16157877

8. Chang YH, Wang YC, Chen BS. Identification of transcription factor cooperativity via stochastic system model. Bioinformatics 2006; 22:2276–2282. PMID: 16844711

9. He D, Zhou D, Zhou Y. 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 (GCCW) 2006; 357–360.

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

11. Yu X, Lin J, Masuda T, Esumi N, Zack DJ, Qian J. Genome-wide prediction and characterization of interactions between transcription factors in Saccharomyces cerevisiae. Nucleic Acids Res 2006; 34:917–927. PMID: 16464824

12. Elati M, Neuvial P, Bolotin-Fukuhara M, Barillot E, Radvanyi F, Rouveirol C. LICORN: learning cooperative regulation networks from gene expression data. Bioinformatics 2007; 23:2407–2414. PMID: 17720703

13. Datta D, Zhao H. Statistical methods to infer cooperative binding among transcription factors in Saccharomyces cerevisiae. Bioinformatics 2008; 24:545–552. PMID: 17989095

14. Chuang CL, Hung K, Chen CM, Shieh GS. Uncovering transcriptional interactions via an adaptive fuzzy logic approach. BMC Bioinformatics 2009; 10:400. doi: 10.1186/1471-2105-10-400 PMID: 19961622

15. Wang Y, Zhang XS, Xia Y. Predicting eukaryotic transcriptional cooperativity by Bayesian network integration of genome-wide data. Nucleic Acids Res 2009; 37:5943–5958. doi: 10.1093/nar/gkp625 PMID: 19661283

16. Yang Y, Zhang Z, Li Y, Zhu XG, Liu Q. Identifying cooperative transcription factors by combining ChIP-chip data and knockout data. Cell Res 2010; 20:1276–1278. doi: 10.1038/cr.2010.146 PMID: 20975739

17. Chen MJ, Chou LC, Hsieh TT, Lee DD, Liu KW, Yu CY, et al. De novo motif discovery facilitates identification of interactions between transcription factors in Saccharomyces cerevisiae. Bioinformatics 2012; 28:701–708. doi: 10.1093/bioinformatics/bts002 PMID: 22238267

18. Lai FJ, Jhu MH, Chiu CC, Huang YM, Wu WS. Identifying cooperative transcription factors in yeast using multiple data sources. BMC Systems Biology 2014; 8(Suppl 5):S2. doi: 10.1186/1752-0509-8-S5-S2 PMID: 25559499

19. Wu WS, Lai FJ. Properly defining the targets of a transcription factor significantly improves the computational identification of cooperative transcription factor pairs in yeast. BMC Genomics 2015; 16(Suppl 12):S10. doi: 10.1186/1471-2164-16-S12-S10 PMID: 26679776

20. Spivak AT, Stormo GD. Combinatorial Cis-regulation in Saccharomyces Species. G3 (Bethesda) 2016; 6(3):653–667.

21. Lai FJ, Chang HT, Huang YM, Wu WS. A comprehensive performance evaluation on the prediction results of existing cooperative transcription factors identification algorithms. BMC Systems Biology 2014; 8(Suppl 4):S9. doi: 10.1186/1752-0509-8-S4-S9 PMID: 25521604

22. Lai FJ, Chang HT, Wu WS. PCTFPeval: a web tool for benchmarking newly developed algorithms for predicting cooperative transcription factor pairs in yeast. BMC Bioinformatics 2015; 16(Suppl 18):S2. doi: 10.1186/1471-2105-16-S18-S2 PMID: 26677932

23. Wu WS, Lai FJ, Tu BW, Chang DT. CoopTFD: a repository for predicted yeast cooperative transcription factor pairs. Database 2016; pii:baw092. doi: 10.1093/database/baw092 27242036

24. Wu WS, Hsieh YC, Lai FJ. YCRD: Yeast Combinatorial Regulation Database. PLoS ONE 2016; 11(7): e0159213. doi: 10.1371/journal.pone.0159213 PMID: 27392072

25. Yu X, Lin J, Zack DJ, Qian J. Computational analysis of tissue-specific combinatorial gene regulation: predicting interaction between transcription factors in human tissues. Nucleic Acids Res 2006; 34 (17):4925–4936. PMID: 16982645

26. Hu Z, Gallo SM. Identification of interacting transcription factors regulating tissue gene expression in human. BMC Genomics. 2010; 11:49. doi: 10.1186/1471-2164-11-49 PMID: 20085649

27. Karczewski KJ, Snyder M, Altman RB, Tatonetti NP. Coherent functional modules improve transcription factor target identification, cooperativity prediction, and disease association. PLoS Genet 2014; 10(2): e1004122. doi: 10.1371/journal.pgen.1004122 PMID: 24516403

28. Allocco DJ, Kohane IS, Butte AJ. Quantifying the relationship between coexpression, co-regulation and gene function. BMC Bioinformatics 2004; 5(1):18.

29. Gyenesei A, Wagner U, Barkow-Oesterreicher S, Stolte E, Schlapbach R. Mining co-regulated gene profiles for the detection of functional associations in gene expression data. Bioinformatics 2007; 23 (15):1927–1935. PMID: 17537754

30. Wu WS, Wei ML, Yeh CM, Chang DT. A regulatory similarity measure using the location information of transcription factor binding sites in Saccharomyces cerevisiae. BMC Syst Biol 2014; 8(Suppl 5):S9. doi: 10.1186/1752-0509-8-S5-S9 PMID: 25560196

31. Teixeira MC, Monteiro PT, Guerreiro JF, Gonçalves JP, Mira NP, dos Santos SC, et al. The YEAS-TRACT database: an upgraded information system for the analysis of gene and genomic transcription regulation in Saccharomyces cerevisiae. Nucleic Acids Res 2014; 42:D161–D166. doi: 10.1093/nar/gkt1015 PMID: 24170807

32. Yang H, Nepusz T, Paccanaro A. Improving GO semantic similarity measures by exploring the ontology beneath the terms and modelling uncertainty. Bioinformatics 2012, 28:1383–1389. doi: 10.1093/bioinformatics/bts129 PMID: 22522134

33. Mewes HW, Frishman D, Güldener U, Mannhaupt G, Mayer K, Mokrejs M, et al. MIPS: a database for genomes and protein sequences. Nucleic Acids Res 2002; 30:31–34. PMID: 11752246

34. Yang TH, Wu WS. Identifying biologically interpretable transcription factor knockout targets by jointly analyzing the transcription factor knockout microarray and the ChIP-chip data. BMC Syst Biol 2012; 6:102. doi: 10.1186/1752-0509-6-102 PMID: 22898448

35. Balaji S, Babu MM, Iyer LM, Luscombe NM, Aravind L. Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast. J Mol Biol 2006; 360(1):213–227. PMID: 16762362

36. Aguilar D, Oliva B. Topological comparison of methods for predicting transcriptional cooperativity in yeast. BMC Genomics 2008; 9:137. doi: 10.1186/1471-2164-9-137 PMID: 18366726

37. Chatr-Aryamontri A, Breitkreutz BJ, Oughtred R, Boucher L, Heinicke S, Chen D, et al. The BioGRID interaction database: 2015 update. Nucleic Acids Res 2015; 43:D470–D478. doi: 10.1093/nar/gku1204 PMID: 25428363

38. Mojzita D, Hohmann S. Pdc2 coordinates expression of the THI regulon in the yeast Saccharomyces cerevisiae. Mol Genet Genomics 2006; 276(2):147–161. PMID: 16850348

39. Rep M, Reiser V, Gartner U, Thevelein JM, Hohmann S, Ammerer G, et al. Osmotic stress-induced gene expression in Saccharomyces cerevisiae requires Msn1p and the novel nuclear factor Hot1p. Mol Cell Biol 1999; 19(8):5474–5485. PMID: 10409737