Published online August 26, 2005

4828–4837 *Nucleic Acids Research, 2005, Vol. 33, No. 15*
doi:10.1093/nar/gki793

Identifying cooperative transcriptional regulations using protein–protein interactions

Nobuyoshi Nagamine, Yuji Kawada and Yasubumi Sakakibara*

Department of Biosciences and Informatics, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan

Received April 6, 2005; Revised July 2, 2005; Accepted August 9, 2005

ABSTRACT

Cooperative transcriptional activations among multiple transcription factors (TFs) are important to understand the mechanisms of complex transcriptional regulations in eukaryotes. Previous studies have attempted to find cooperative TFs based on gene expression data with gene expression profiles as a measure of similarity of gene regulations. In this paper, we use protein–protein interaction data to infer synergistic binding of cooperative TFs. Our fundamental idea is based on the assumption that genes contributing to a similar biological process are regulated under the same control mechanism. First, the protein–protein interaction networks are used to calculate the similarity of biological processes among genes. Second, we integrate this similarity and the chromatin immuno-precipitation data to identify cooperative TFs. Our computational experiments in yeast show that predictions made by our method have successfully identified eight pairs of cooperative TFs that have literature evidences but could not be identified by the previous method. Further, 12 new possible pairs have been inferred and we have examined the biological relevances for them. However, since a typical problem using protein–protein interaction data is that many false-positive data are contained, we propose a method combining various biological data to increase the prediction accuracy.

INTRODUCTION

Promoter regions of higher eukaryotic genes have complex structures to regulate their transcriptional activations and are controlled by multiple transcription factors (TFs). TFs are DNA-binding proteins at the terminals of signal transduction networks, and computational representations and identifications of binding sites for TFs have been widely studied (1).

Recent molecular technologies have produced many kinds of experimental data including whole genome sequences, gene expression profiles and protein–protein interactions. Therefore, several methods have been developed to infer transcriptional regulation networks by combining these various kinds of data. For example, Banerjee *et al*. (2) used expression data and the chromatin immuno-precipitation (ChIP) data to predict cooperativity of TFs, which is a key factor for the analyses of complex transcriptional regulation networks. This line of approach is very significant because according to the fact that there are at most 200 different TFs among totally 6300 genes in yeast organism, there must exist cooperative transcriptional activations to control the expressions of all 6300 genes.

Recently, in addition to gene expression data, protein–protein interaction data have been rapidly generated. In this paper, we propose a method integrating protein–protein interaction data and ChIP data (our strategy is illustrated in Figure 1), and show the effectiveness of exploiting protein–protein interactions to identify cooperative transcriptional activations.

The existence of interaction between two proteins suggests that they contribute to the same or similar biological processes. Many cellular processes and chemical events in organisms such as enzymatic reactions and dimerization involve protein–protein interactions. In addition, these interactions reveal, in some cases, functional similarity of proteins. Schwikowski *et al*. (3) proposed a protein function prediction method by which a protein of unknown function is predicted to have the three most frequent cellular functions represented among its direct interaction partners.

However, gene expression data also have been applied to protein function predictions. Clustering analysis of gene expression data can be used to predict functions of unannotated proteins based on the idea that genes with similar functions are likely to be co-expressed (4,5).

Based on these observations, we may deduce that the existence of protein–protein interaction is strongly related to the correlation of gene expressions from the viewpoint of

*To whom correspondence should be addressed. Tel/Fax: +81 45 566 1791; Email: yasu@bio.keio.ac.jp

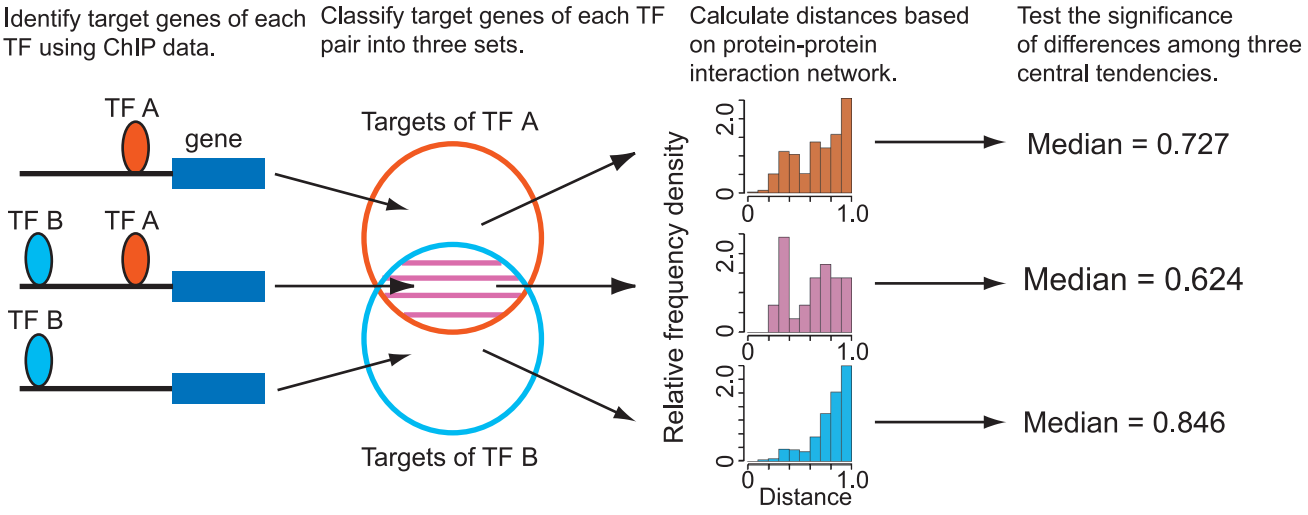

Figure 1. Strategy of identifying cooperative TF regulation using ChIP data and protein–protein interaction data. If the central tendency (we use median) of the overlap set that includes genes which both TF *A* and TF *B* bind to is significantly lower than those of the other two sets, we conclude that TF *A* and TF *B* are cooperative.

functional similarity. This leads to the fundamental assumption of this paper that proteins that are close to each other in the protein–protein interaction network are likely to be co-regulated by a same set of TFs. In fact, this assumption is supported by the observation made by Jansen *et al*. (6) that protein complexes have a strong relationship with gene expressions. Therefore, we can use the similarities of biological processes measured by protein–protein interactions to identify the cooperative TFs.

In our method, first, the protein–protein interaction networks are used to calculate the similarity of biological processes that the genes contribute to. Second, we integrate the similarity of biological processes based on protein–protein interactions and ChIP data to identify synergistic binding of TFs. Our computational experiments in yeast show that predictions made by our method based on protein–protein interactions have successfully identified eight pairs of cooperative TFs that have literature evidences and could not be identified by the previous method based on gene expression data. In addition, 12 new possible pairs of cooperative TFs have been inferred. From our careful analyses of the biological relevances for those pairs, we suggest a biological observation that some metabolism is regulated rather on translation level than on transcription level.

Furthermore, when using protein–protein interaction data, a typical problem is their noisiness, i.e. the data contains many false-positives. Integrating various kinds of data is one solution to this problem (7). In addition, it may enable us to take many biological perspectives into consideration. In this paper, we propose a method integrating cellular localization data and function data with protein–protein interaction data for precise predictions of TF cooperativity.

MATERIALS AND METHODS

Selecting target genes of TF pairs from ChIP data

We use Lee *et al*.'s (8) ChIP data, a genome-wide binding data of 113 yeast regulators, to determine target genes of each TF.

We suppose that a protein is regulated by a TF if its binding *P*-value $P_B < 0.001$ is satisfied. For all pairs of TFs *A* and *B*, we divide target genes for two TFs into three sets: those of TF *A* but not *B* (i.e. $A \cap \bar{B}$), TF *B* but not *A* ($\bar{A} \cap B$), and both TF *A* and *B* ($A \cap B$). TF pairs whose overlap set ($A \cap B$) has genes less than threshold O_{min} are excluded.

Constructing protein–protein interaction network

We construct a protein–protein interaction network based on the dataset organized by Yu *et al*. (9). It contains data produced by different experiments (compiled from MIPS, BIND and DIP databases), those by large-scale yeast two-hybrid experiments and those by *in vivo* pull-down experiments. It consists of 69592 interactions involving 4957 proteins. The average number of interaction partners per protein is ~27.9.

Calculating distance between two proteins based on protein–protein interaction

We calculate a distance between any two proteins based on a newly defined distance function exploiting the protein–protein interaction network constructed as above.

Two typical distance measures between two proteins based on the protein–protein interaction network are a graph–theoretic distance D_G, and the Czekanowski–Dice distance D_{CD} proposed by Brun *et al*. (10).

For the proteins *i* and *j*, $D_G(i, j)$ is defined as the minimum number of edges needed to traverse from *i* to *j*. However, $D_{CD}(i, j)$ is defined as follows:

$$D_{CD}(i,j) = \frac{|\mathrm{Int}(i)| + |\mathrm{Int}(j)| - 2|\mathrm{Int}(i) \cap \mathrm{Int}(j)|}{|\mathrm{Int}(i)| + |\mathrm{Int}(j)|} \quad \mathbf{1}$$

where Int(*i*) and Int(*j*) are the lists of interactors of the proteins *i* and *j* plus themselves (to decrease the distance between proteins interacting with each other). D_{CD} ranges from 0 to 1. A feature of these distances is that the less the distance between two proteins is, the stronger their biological or functional relatedness is thought to be.

However, a serious problem of these distances is that they cannot express diversity and specificity of distances between proteins adequately.

D_G is a discrete measure and cannot be defined for proteins that are not linked in the network. D_{CD} is <1 only if two proteins are within a distance of 2 in terms of D_G (otherwise, $|\mathrm{Int}(i) \cap \mathrm{Int}(j)|$ in Equation 1 always equals to 0), while $D_G > 2$ for most pair of proteins in the gene sets of Yu *et al*. (9).

To overcome this problem, we extend the Czekanowski–Dice distance as follows:

$$D_I(i,j,l) = \left(\sum_{k=1}^{l} \frac{1}{k} (|\mathrm{Int}_k(i)| + |\mathrm{Int}_k(j)|) - 2 \sum_{n=1}^{l} \sum_{m=1}^{l} \frac{2}{m+n} |\mathrm{Int}_m(i) \cap \mathrm{Int}_n(j)| \right) \Bigg/ \left(\sum_{k=1}^{l} \frac{1}{k} (|\mathrm{Int}_k(i)| + |\mathrm{Int}_k(j)|) \right) \quad \mathbf{2}$$

where $\mathrm{Int}_k(i)$ is a list of proteins whose D_G from the protein i is equal to k [$\mathrm{Int}_1(i)$ includes the protein i itself as in the Czekanowski–Dice distance], and l denotes the range of D_G to be considered. $D_I(i, j, 1)$ is equal to D_{CD}.

Finally, we define the new distance function between the proteins i and j, denoted $D(i, j, l)$, as

$$D(i,j,l) = \min_k D_I(i,j,k), \quad k \leqslant l \quad \mathbf{3}$$

We use $D(i, j, 2)$ as a protein distance and represent it as $D(i, j)$.

Further, we consider the protein pair as biologically significant only when both proteins have direct interactors more than threshold I_{min} (we use 3 as I_{min}).

Evaluating transcription factor cooperativity

We assume that proteins which are close to each other in the protein–protein interaction network, or those which may contribute to the same biological processes, are regulated under the same control mechanism. On this assumption, if TF *A* and TF *B* are cooperative, proteins that are controlled by both TFs must be closer to each other in terms of $D(i, j)$ than those regulated only by TF *A* or TF *B*. To precisely measure this differences of distance among three gene sets, we examine whether the central tendency of the overlap distance set, which is a set of distances $D(i, j)$ for pairs of proteins i, j in the overlap gene set ($A \cap B$), must be significantly lower than those for the two other distance sets, TF *A* distance set for the TF *A* gene set ($A \cap \bar{B}$) and TF *B* distance set for the TF *B* gene set ($\bar{A} \cap B$).

First, we choose TF pairs in which the median of overlap distance set was lower than the other distance sets. We determine the significance of differences by the Mann–Whitney *U*-test. As the distribution of distances is not normal (Figure 2), we use this non-parametric statistical test.

For a pair of TF *A* and TF *B*, if the *P*-value of Mann–Whitney *U*-test for the combination of the overlap distance set and TF *A* distance set and that for the combination of the overlap distance set and the TF *B* distance set satisfy the threshold of 0.05 with Holm's correction, which means that the lower of these two *P*-values must be <0.025 and the other *P*-value must be <0.05, we conclude that TF *A* and TF *B* are cooperative.

The *P*-value for the Mann–Whitney *U*-test is calculated as follows (11). First, the statistic *U* is calculated,

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - \sum_i r_{1i},$$

where n_1 denotes the number of elements in first set, n_2 that of second set and r_{1i} is a rank of i-th element in first set where ranks are assigned according to all the elements. Second, the statistic Z_0 is calculated,

$$Z_0 = \frac{|U - n_1 n_2/2|}{\sqrt{\frac{n_1 n_2}{12(n^2-n)} \left\{ n^3 - n - \sum_{i=1}^{m} (t_i^3 - t_i) \right\}}},$$

where n denotes the number of all the elements, m the number of different kinds of ranks and t_i the number of elements of i-th rank.

Finally, the *P*-value of Z_0 is calculated based on the normal distribution. We execute one-tailed testing so that we only examine that the central tendency of the overlap set is significantly lower than those of the other two sets.

Prediction of cooperative TF triads

When the pair of TF *A* and TF *B* and that of TF *B* and TF *C* are both cooperative, we apply the same method to the triad of TF *A*, TF *B* and TF *C*.

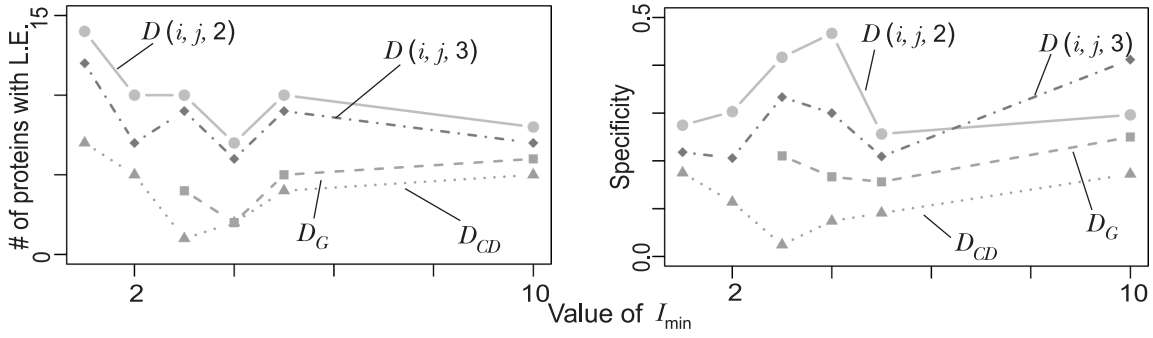

Figure 2. Comparison by other distance functions. $P_B \leqslant 0.001$, $O_{min} = 3$ and $P_{mw} \leqslant 0.05$ are used. The number of predictions with literature evidence is relevant to sensitivity of each condition. Specificity = No. of predictions with literature evidence/No. of predictions. As for D_G and D_{CD}, mean is used instead of median as the central tendency. There are proteins whose distances are not defined in using D_G with $I_{min} = 1$ or 2. Thus, results on D_G with $I_{min} = 1$ and 2 are omitted.

Prediction of cooperative TF module

When we consider TF *A* and TF *B*, there could be a case that the target genes of each TF are also those of other (third) TF *C* and TF *D* which are possibly cooperative. If the targets of cooperative TFs, TF *C* and TF *D*, are included in the TF *A* set, the TF *A* set can be quite close to each other in terms of $D(i, j)$ due to the influences of TF *C* and *D*. In that case, they may obscure the differences between the TF *A* distance set and the overlap distance set, and so obstruct the detection of their cooperativity.

To treat this problem, in judging cooperativity of TF *A* and TF *B* as above, we redefine the TF *A* gene set as genes that only TF *A* binds to, the TF *B* gene set in the same way, and the overlap gene set as those that both TF *A* and TF *B*, but no other TFs bind to. We apply this method to any combination of TFs whose TF and overlap sets have more than $O_{\min}$ genes. We represent combinations of TFs that are predicted as cooperative by this method as cooperative TF 'modules'.

INTEGRATING OTHER KINDS OF BIOLOGICAL DATA

Cellular localization data

In a high throughput data of protein–protein interactions, many unnatural data such as an interaction between a protein in the nucleus and one in the plasma membrane are included. However, in the living cell, these interactions between proteins existing apart could be never observed. Therefore, we incorporate localization data of proteins to exclude unnatural interactions and improve the reliability of protein–protein interactions.

A straightforward introduction of localization information is that for two proteins which have an interaction reported in experiments, if these two proteins have at least one common localization, then we accept the interaction, otherwise reject the interaction. However, since not all localization data for the proteins are available, this straightforward method is naive. We use mutual information criterion on co-occurrence of localizations in proteins to judge the relatedness of two localizations and verify the possibility of protein–protein interactions. Mutual information I_m is expressed as follows:

$$I_m(i,j) = \sum_{k=\{0,1\}} \sum_{l=\{0,1\}} P\big(X_i = k, X_j = l\big) \times \log_2 \frac{P\big(X_i = k, X_j = l\big)}{P(X_i = k)P\big(X_j = l\big)}, \quad \mathbf{4}$$

where *i* and *j* denote two localizations, $P(X_i = 1)$ denotes the possibility that a protein has localization *i*, $P(X_i = 0)$ the possibility that a protein does not have localization *i*, $P(X_i = 1, X_j = 1)$ the possibility that a protein has both localization *i* and *j*. The mutual information is a method often used to give a quantitative relation between two discrete elements, and produces biologically relevant results applied to biological data (12,13).

For two proteins *i* and *j* that have an interaction reported in experiments, if there exist at least one combination of localization l_i that *i* has and l_j for *j* in which $I_m(l_i, l_j)$ exceeds some threshold I_{loc}, then we adopt the interaction between *i* and *j*.

We use MIPS cellular localization data (14) to calculate mutual information of localizations and to select reliable interactions.

Function data

The existence of protein–protein interaction often reflect functional similarity. Thus, we may incorporate function data annotated for proteins to refine the interaction-based protein distances. Here, we use the logistic regression to achieve this purpose. The logistic regression is a method, often used in medicine (15), to quantitatively evaluate a relation between one discrete element and the others. As we here want to know not a relation between two functions but that between a protein that may have multiple functions and a function, the logistic regression is more appropriate than the mutual information. In addition, its output is more meaningful as probability and more tractable with the range (0, 1) than the linear multiple regression or the mutual information.

By using the logistic regression on co-occurrence of functions in proteins, we can calculate the possibility that a protein has some function *z* as follows:

$$\Pr(z = 1|\boldsymbol{x}) = \frac{1}{1 + \exp\{-(1, \boldsymbol{x})^t \boldsymbol{\beta}\}} \quad \mathbf{5}$$

where a vector $\boldsymbol{x}$ denotes the list of functions that a protein has excluding *z*, in which $x_i = 1$ if a protein has the function *i* and 0 if not, and a vector $\boldsymbol{\beta}$ consists of intercept and coefficients for each element in $\boldsymbol{x}$ and is estimated based on logistic regression model.

For a protein *x* and a function *f*, using the Equation 5, we denote $P(x, f)$ as follows:

$$P(x,f) = \begin{cases} 1 & \text{if } F_f(x) = 1 \\ \Pr\big(F_f(x) = 1|\boldsymbol{F}(x)\big) & \text{else} \end{cases} \quad \mathbf{6}$$

where $\boldsymbol{F}(x)$ denotes a functional vector of *x*.

Exploiting $P(x, f)$, we redefine the Equations 2 and 3 for protein distances as follows:

$$f(\text{Int}_k(i)) = \sum_{x \in \text{Int}_k(i)} P(x,f)$$

$$D_{\text{If}}(i,j,l,f) = \left(\sum_{k=1}^{l} \frac{1}{k} \{f(\text{Int}_k(i)) + f(\text{Int}_k(j))\} - 2 \sum_{n=1}^{l} \sum_{m=1}^{l} \frac{2}{m+n} f(\text{Int}_m(i) \cap \text{Int}_n(j)) \right) \Bigg/ \left(\sum_{k=1}^{l} \frac{1}{k} \{f(\text{Int}_k(i)) + f(\text{Int}_k(j))\} \right)$$

$$D_{\text{If}}(i,j,l,f) = \min_k D_{\text{If}}(i,j,k,f), \quad k \leqslant l$$

We represent $D_f(i, j, 2, f)$ as $D_f(i, j, f)$.

If two proteins *i* and *j* have some common functions that are not related to the functions of the objective TFs, then we calculate $D_f(i, j, f)$ by choosing, among functions which two target proteins have in common, a function *f* that maximizes the distance as the proteins may be strongly influenced by other TFs. We use the mutual information to measure the

'relatedness' of protein functions with some threshold I_{fnc} as we have done in localization data.

We use MIPS function data (14) to construct logistic regression model and calculate functional distance D_f. We exclude proteins whose function is unknown when constructing logistic regression model.

Comparison with a method using expression data

We compare prediction results of our method with those of Banerjee *et al*. (2). The method of Banerjee *et al*. calculates the correlations of expression profiles for the target genes and finds cooperative TFs based on the correlations. Therefore, in order to investigate the relationships between predictions on which two methods agree or disagree, we calculate the relationships between the protein distances by our method based on protein–protein interactions and the correlations of gene expression profiles. We exploit genome-wide cell cycle expression data (16) to obtain the correlation of expression profiles.

RESULTS

Interaction-based protein distance

First, we verify the adequacy of using extended Czekanowski–Dice distance $D(i, j, l)$ as a distance measure.

Figure 3 shows a distribution of $D(i, j, l)$. As shown in the figure, the distribution mainly depends on l, which determines the extent of interactors to be considered. The bigger l is, the more distant interactors are taken into consideration. $D(i, j, l)$, or $D_{CD}(i, j)$, cannot express variety of protein distances properly in which most distance is equal to 1. However, in using $D(i, j, 3)$, specificity of closeness may be lost as randomly chosen two proteins tend to be rather close to each other. Thus, it is reasonable to use $D(i, j, 2)$ as a protein distance.

Predictions of synergistic binding

From ChIP data of Lee *et al*. (8), with the threshold of binding P-value $P_B < 0.001$, which determined target genes of each TF, we extracted 476 TF pairs satisfying $O_{min} = 3$, the threshold for the number of genes in overlap sets. TF pairs whose overlap set has more than O_{min} genes are considered. From these TF pairs, by calculating $D(i, j, 2)$ for pairs of proteins i and j both of which had more than the threshold $I_{min} = 3$ direct interactors and judging the significance of Mann–Whitney tests on distance sets with $P_{mw} \leqslant 0.05$, we identified 24 pairs as cooperative TFs (Table 1). Figure 2 shows the results of predictions on several parameter values for calculation of protein distances in terms of specificity and sensitivity based on the literature (see Supplementary Materials for details of predictions). It explains the reason to choose these parameters ($P_B < 0.001, O_{min} = 3, I_{min} = 3$), and indicates that parameter values which we chose produce the best result.

Compared with the literature and the predictions using gene expression data (2), about half of our predictions overlap with their results (Table 1 and Figure 4A, C and D).

Overlaps with literature

It is remarkable that cooperative TF pairs Hap2/Hap5, Hap3/Hap5, Arg80/Arg81, Fkh1/Fkh2, Fkh1/Ndd1, Mbp1/Skn7, Gcr1/Gcr2 and Skn7/Yap1 are only detected by our method and have literature evidences. Particularly, the detection of Hap2/Hap3/Hap5 cooperativity (Table 2) meets with the fact that these three TFs form a heterotrimer to be a CCAAT DNA-binding factor (17).

A half of the overlaps, including Hir1/Hir2, Fkh1/Fkh2, Fkh1/Ndd1, Fkh2/Mcm1 and Mbp1/Skn7, are involved in the cell cycle. Fkh1, Fkh2, Mcm1 and Ndd1 are TFs that mainly control the S/G_2, G_2 and G_2/M phases (18). Mbp1 and Skn7 are known to function in the G_1/S phase (19). Hir1 and Hir2, in the S phase, contribute to transcriptional repression (20).

However, Arg80/Arg81, Gcr1/Gcr2 and Skn7/Yap1 are involved in biological processes other than cell cycle. Arg80 and Arg81 regulate the metabolism of arginine (21). Gcr1 and Gcr2 contribute to regulating glycolysis (22,23). Skn7 and Yap1 play a role in the oxidative stress response (24).

In addition, Pho2/Pho4 and Stb1/Swi4/Swi6 are detected as a module (Table 3). Pho2 and Pho4 are known to function in the regulation of phosphate metabolism (25). Stb1, Swi4 and Swi6 regulate START in the G_1 phase of the cell cycle (26).

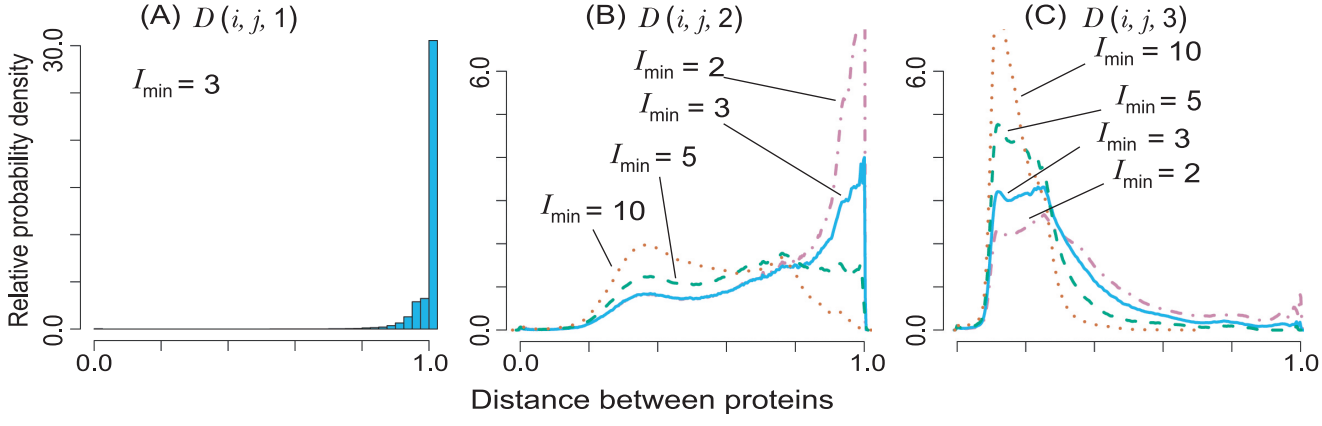

Figure 3. Distributions of distances based on the protein–protein interaction network $D(i, j, l)$. For all possible pairs of proteins that have more than I_{min} interactors, $D(i, j, l)$ is calculated. As for $D(i, j, 2)$ and $D(i, j, 3)$, the probability density on several I_{min} is estimated using Kernel method. $D(i, j, l)$ ranges from 0.0 to 1.0, and its distribution depends on l, the value determining the extent of interactors of protein i (j) to be considered.

Table 1. Predicted cooperative TF pairs (P_B[a] < 0.001, O_{min}[b] = 3, I_{min}[c] = 3, P_{mw}[d] ≤ 0.05)

	TF1	TF2	P_{mw} (versus TF1)	P_{mw} (versus TF2)	Literature evidence	Expression data[e]
1	HIR1	HIR2	5.41E−07	4.14E−07	(30)	Y[f]
2	FHL1	RAP1	3.48E−04	8.10E−34	NA	N
3	MCM1	SWI4	6.53E−04	2.57E−04	NA	N
4	RAP1	YAP5	1.41E−03	1.43E−10	NA	N
5	FKH1	NDD1	2.09E−03	9.63E−04	(46)	y[g]
6	FHL1	PDR1	2.29E−03	2.30E−06	NA	N
7	FKH2	MCM1	3.99E−03	3.02E−04	(20)	Y
8	MBP1	MSN4	4.23E−03	2.88E−03	NA	N
9	ARG80	ARG81	5.61E−03	2.92E−04	(39)	y
10	NRG1	PHD1	6.35E−03	8.02E−05	NA	N
11	RAP1	SFP1	6.42E−03	1.15E−03	NA	N
12	FHL1	GAT3	6.66E−03	3.51E−08	NA	Y
13	FHL1	YAP5	7.52E−03	6.77E−18	NA	N
14	HAP2	HAP5	2.51E−04	1.01E−02	(17)	N
15	HAP3	HAP5	7.96E−04	1.01E−02	(17)	N
16	NRG1	YAP6	4.58E−07	1.17E−02	NA	Y
17	MBP1	MCM1	1.17E−02	2.31E−04	NA	N
18	FKH1	FKH2	3.88E−03	1.43E−02	(20,47)	y
19	HSF1	RAP1	1.44E−02	1.97E−02	NA	N
20	SWI5	YAP5	2.25E−02	1.17E−02	NA	N
21	NDD1	SKN7	3.09E−02	1.42E−02	NA	N
22	MBP1	SKN7	3.59E−02	1.80E−05	(19)	N
23	GCR1	GCR2	4.21E−02	2.41E−02	(22,23)	N
24	SKN7	YAP1	8.15E−03	4.34E−02	(24)	N

[a]P_B, *P*-value for TF binding to chromatin as described by Lee *et al*. (8).
[b]O_{min}, threshold for the number of genes in overlap sets. TF pairs whose overlap set has more than O_{min} genes are considered.
[c]I_{min}, threshold for the number of interactions. $D(i, j, 2)$ is calculated for proteins that have more than I_{min} interactions.
[d]P_{mw}, *P*-value as a result of Mann–Whitney *U*-test.
[e]Predictions by using gene expression data by Banerjee *et al*.(P_B < 0.001) (2).
[f]Capital 'Y' means that *P*-value < 0.05 with Holm's correction by Banerjee *et al*. (2).
[g]Small 'y' means that *P*-value < 0.05, but not significant with Holm's correction by Banerjee *et al*. (2).

Table 2. Predicted cooperative TF triads (P_B < 0.001, O_{min} = 3, I_{min} = 3, P_{mw} ≤ 0.05)

TF1	TF2	TF3	P_{mw} (versus TF1)	P_{mw} (versus TF2)	P_{mw} (versus TF3)
FHL1	PDR1	RAP1	7.55E−03	8.19E−05	2.02E−05
HAP2	HAP3	HAP5	3.52E−04	5.01E−03	1.01E−02
FHL1	RAP1	YAP5	1.67E−02	9.08E−06	1.14E−12
MBP1	MCM1	SWI4	2.47E−02	7.10E−04	1.48E−03
FHL1	RAP1	SFP1	3.77E−02	2.95E−03	2.08E−03
[a]FHL1	GAT3	RAP1	4.73E−02	1.94E−04	1.37E−02
NDD1	SKN7	YAP1	7.61E−03	7.98E−03	7.81E−02
FKH1	FKH2	NDD1	2.34E−02	9.02E−02	6.19E−03
NRG1	PHD1	YAP6	1.38E−03	1.42E−04	1.16E−01
FHL1	GAT3	YAP5	5.06E−01	8.61E−03	7.67E−05

[a]*P*-value boundary upper which all of three Mann–Whitney *U*-tests satisfy the threshold with Holm's correction, and below which two of them satisfy it.

Newly discovered TF pairs

As for newly discovered TF pairs without literature evidence, we may classify them into three groups: cooperative pairs involved in the cell cycle, those concerned with Nrg1 and those including Fhl1, Rap1 and Yap5.

Predictions related to cell cycle

Mbp1/Mcm1, Mbp1/Msn4, Mcm1/Swi4 and Ndd1/Skn7 are thought to be involved in the cell cycle (Figure 4C). Particularly, we infer that Mcm1/Swi4 and Mbp1/Mcm1 cooperatively function in the M/G_1 phase of the cell cycle. In the M/G_1 phase, it is argued that Mcm1 and Swi4 form a feed-forward loop, in which Mcm1 activates Swi4, and both of them activate Clb2 (8). Thus, it is very reasonable to conclude that Mcm1 and Swi4 are cooperative. However, Mbp1 is known to activate Swi4 (18) and may participate in the feed-forward loop. The cooperativity of Mbp1/Mcm1/Swi4 and Mbp1/Mcm1/Swi4/Swi6, in which Mbp1/Swi6 and Swi4/Swi6 are known to be cooperative (27), supports this.

In Mbp1/Msn4 and Ndd1/Skn7, Mbp1 and Ndd1 regulate a progression from the G phase in the cell cycle (G_1 to S and G_2 to M, respectively) (18). However, Msn4 and Skn7 are involved in the response to the oxidative stress (28,29). In addition, in Ndd1/Skn7/Yap1, Skn7/Yap1 contributes to the oxidative stress response (30). From these, Ndd1/Skn7 and Mbp1/Msn4 may contribute to the mechanism of the cell cycle arrestation by stress.

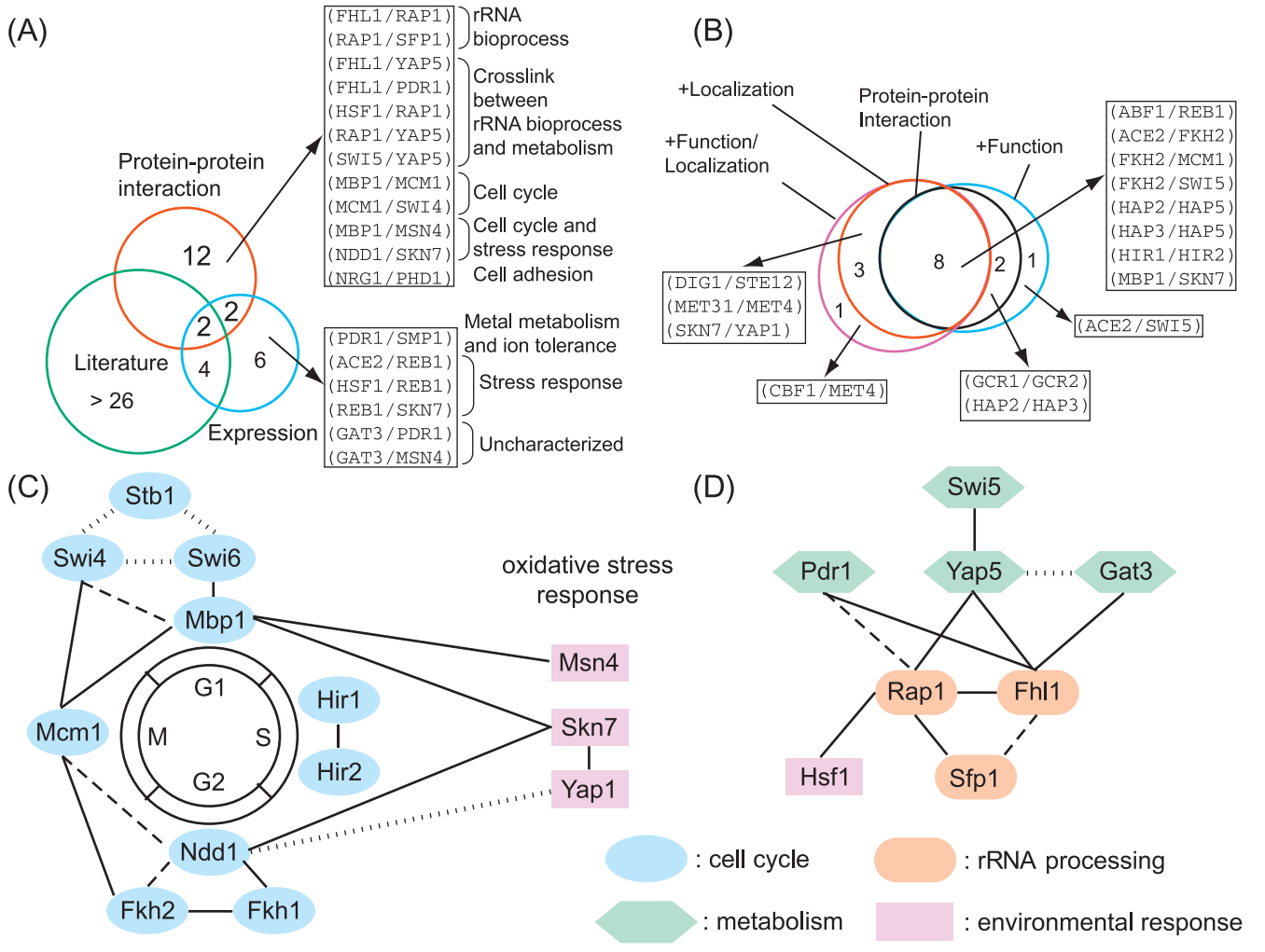

Figure 4. (**A**) Comparison between predictions by protein–protein interaction data (with $P_B < 0.001$, $O_{min} = 3$, $I_{min} = 3$), and those by expression data (2). All TF pairs listed in this diagram are those less than *P*-value cutoff (i.e. *P*-value < 0.05) with Holm's correction. Predictions made only by using protein–protein interaction or only by expression, and their possible functions are shown. The number of predictions by both data depends on parameters and expression or protein–protein interaction data to be used, and it may fluctuate. (**B**) The effect of integrating various biological data. The number of true-positive predictions in each condition, with $P_B < 0.001$, $O_{min} = 3$, $I_{min} = 2$, $P_{mw} \leqslant 0.05$, and the mutual information threshold $I_{loc} = I_{fnc} = 0.01$, are shown. The total number of predictions is 34 in using only protein–protein interaction data, and 37 in the other conditions. (**C**) Predicted cooperative TF clusters. The cluster of TFs involved in cell cycle. The allocations of TFs to four cell cycle phases are determined based on Simon *et al.* (18). (**D**) The cluster of TFs related to rRNA processing. Marks surrounding a protein show a biological process that the protein contributes to. A broken line indicates cooperativity detected as a triad or a module, and a dotted one does weak cooperativity in a triad or a module.

Table 3. Predicted cooperative TF modules ($P_B < 0.001$, $O_{min} = 3$, $I_{min} = 3$, $P_{mw} \leqslant 0.05$)

TF1	TF2	TF3	TF4	P_{mw} (versus TF1)	P_{mw} (versus TF2)	P_{mw} (versus TF3)	P_{mw} (versus TF4)
FHL1	RAP1			<2.2E−308	1.12E−18		
PHO2	PHO4			8.59E−11	2.08E−11		
MBP1	SWI6			8.22E−06	8.89E−04		
FHL1	RAP1	YAP5		3.55E−32	<2.2E−308	<2.2E−308	
FKH1	FKH2	NDD1		1.42E−05	7.67E−104	1.18E−30	
[a]FKH2	MCM1	NDD1		9.16E−08	1.22E−04	2.71E−48	
STB1	SWI4	SWI6		6.07E−08	7.56E−02	3.17E−24	
MBP1	MCM1	SWI4	SWI6	5.79E−02	1.07E−34	5.94E−43	5.21E−34

[a]*P*-value boundary upper which all of the Mann–Whitney *U*-tests satisfy the threshold with Holm's correction, and below which all but one satisfy it.

Predictions involving Nrg1

As for cooperative pairs concerned with Nrg1, a possible function of Nrg1/Phd1 and Nrg1/Yap6 is a control of cell adhesion. Nrg1 and Phd1 are, respectively, thought to be involved in the control of cell adhesion, particularly a regulation of Flo11, which is a key protein of cell adhesion (31–33). Relation of Yap6 to cell adhesion is not yet known. However, given a weak cooperativity of Nrg1/Phd1/Yap6

(two out of three Mann–Whitney *U*-tests satisfy the threshold), it is possible that Nrg1, Phd1 and Yap6 cooperatively control the cell adhesion.

Predictions involving Fhl1, Rap1 and Yap5

The cooperative TF pairs including Fhl1, Rap1 and Yap5 may be classified into two groups: one including Fhl1, Rap1 and Sfp1, and the other consisting of Gat3, Hsf1, Pdr1, Swi5 and Yap5 (Figure 4D). Fhl1, Rap1 and Sfp1 are individually known to play a role in rRNA processing and ribosome biosynthesis (34–36). Fhl1/Rap1/Sfp1 cooperativity indicates the relations among them.

However, Gat3, Pdr1, Swi5 and Yap5 are proteins involved in metabolism. Gat3 is known to be responsible for nitrogen metabolism (37). Hsf1 regulates the heat shock response (38). Pdr1 participates in metal metabolism (39,40). Swi5 and Yap5 may play a role in some drug metabolism (41,42). Hence, according to the relations of TFs controlling rRNA bioprocess to those involved in some types of metabolism (including Fhl1/Rap1/Yap5, Fhl1/Gat3/Rap1 and Fhl1/Pdr1/Rap1 cooperativity), we suggest that 'some metabolism is regulated rather on translation level than on transcription level'.

Effects of integrating other kinds of biological data

Figure 4B shows predictions made by integrating protein localization and function data on conditions of $P_B < 0.001$, $O_{min} = 3$, $I_{min} = 2$, $P_{mw} \leqslant 0.05$ and $I_{fnc} = I_{loc} = 0.01$, the threshold of mutual information that determines which localizations and functions are significantly related to one another. As shown in Figure 4B, integrating other kinds of biological data enables us to detect more true-positive cooperative TF pairs. Among these, Dig1/Ste12 regulates the invasive growth (43), Met31/Met4 is involved in the sulfur amino acid metabolism (44) and Cbf1/Met4 in the glutathione metabolism (45).

While predictions using protein–protein interaction or expression data tend to be related to the cell cycle, integration of various biological data may allow incorporation of many biological aspects and expand the scope of predictions to other biological processes than the cell cycle.

Relationships between protein–protein interaction-based distance and gene expression correlation

In the overlap sets of TF pairs that are predicted as cooperative by our method using protein–protein interaction data, we find that the relationship between the interaction-based protein distance and the expression correlation can be approximated by the following equation:

$$D(i,j) = \frac{-aC(i,j) + c(1/k(i) + 1/k(j)) + d}{1 - bC(i,j)},$$

where $D(i, j)$ denotes the interaction-based distance between proteins i and j, $C(i, j)$ the Pearson correlation coefficient between expression profiles of proteins i and j, $k(i)$ the number of interactors of protein i, and a, b, c, d are appropriate constants ($a, b, c, d > 0$) (Figure 5A).

This equation shows that the closeness of two proteins based on our distance measure is proportional to the expression correlation and the profusion of their interactors. As the fact that an expression similarity often implies an existence of protein complexes is already known in (6), we infer that the detection of cooperativity of two TFs by using protein–protein interactions depends on whether the target proteins regulated cooperatively by those TFs can form a complex or not.

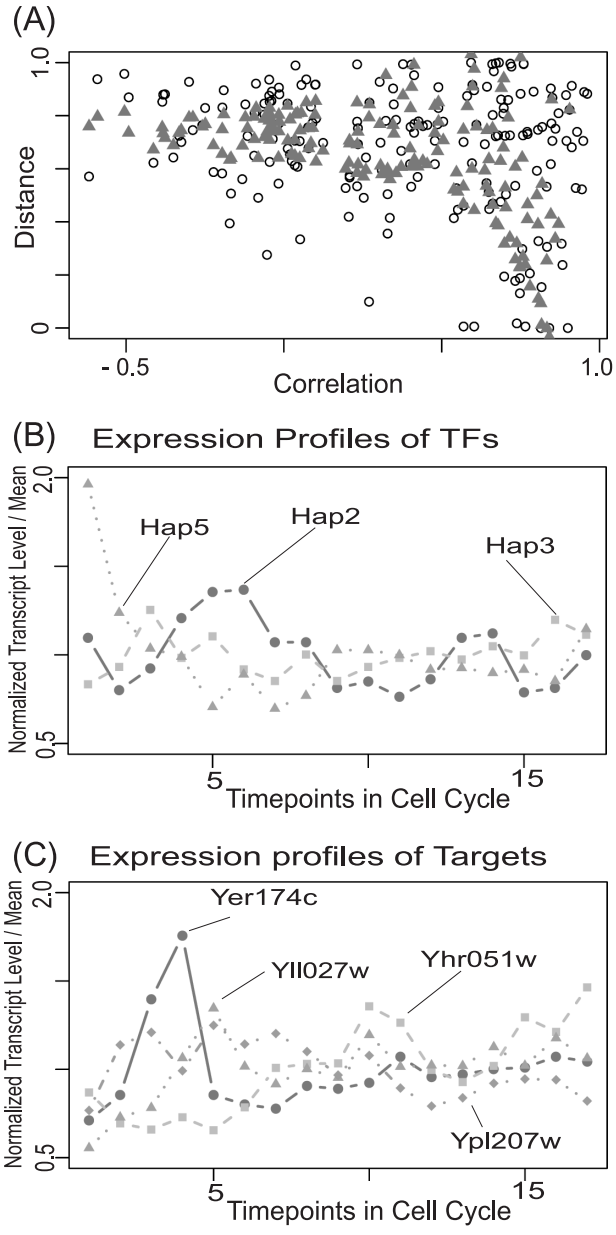

Figure 5. (**A**) Relationships between protein distance $D(i, j)$ and expression correlation coefficient $C(i, j)$ (Pearson correlation coefficient between two expression profiles). For a pair of protein i and j in the overlap sets of TF combinations that are predicted to be cooperative by using protein–protein interaction and expression data, the circle denotes the value of their $C(i, j)$ and $D(i, j)$ on the horizontal and vertical axis, respectively. The triangle denotes approximation of $D(i, j)$ by $C(i, j)$ and number of interactors of proteins i and j. Expression profiles of genes concerned with Hap2, Hap3 and Hap5. Expression profiles of three TFs (**B**) and their target genes which all of these three bind to (**C**) in cell cycle according to Cho *et al.* (16) are shown. Ylr220w, one of targets of these three TFs, is excluded because it has only one interactor and is thought to be less important.

DISCUSSION

We have proposed a novel method to infer cooperativities of TFs based on protein–protein interactions and shown the biological relevance of our predictions (see Predictions of synergistic binding). However, predictions by our method are a bit limited in coverage. As mentioned above (see Relationships between protein–protein interaction-based distance and gene expression correlation), our method may sensitively detect existence of protein complexes. We note that predictions made by our method are not only based on the existence of protein complexes but also capture other biological aspects within protein–protein interactions (see predictions based only on *in vivo* pull-down data, which consist of protein complex data, in Supplementary Materials). Nevertheless, we found that, though predictions made by using protein–protein interactions and those by using expression data overlap on already known cooperative TFs, most novel predictions were specific to each method (Figure 4A). This feature shows the possibility that methods using expression data and protein–protein interaction data, each with a bit limited coverage, can complement each other.

Still, there are some advantages of using protein–protein interaction data.

First, it may enable us to detect locally significant correlations among expression profiles. For example, the expression profiles of target genes of Hap2/Hap3/Hap5, whose cooperativity is detected only by our method, are not similar as a whole, with no higher correlation coefficient between two expression patterns than 0.3 (Figure 5B and C). However, in timepoints 12–15 of the cell cycle, where the expression levels of three TFs are similar and they may form a complex to function, the expression profiles of target genes in this short period are rather related to each other, with three of six possible correlation coefficients exceeding 0.7 and five higher than 0.4. The closeness in the protein–protein interaction network may reflect this locally meaningful relatedness of expression patterns.

Second, a method using protein–protein interaction data has some capability for handling post-transcriptional modifications. Though the post-transcript modification is a key factor for many biological phenomena like diseases, the expression data give no information about it. However, by using interactions between proteins with post-transcript modifications, method using protein–protein interactions can, though indirectly, cooperate post-transcript modifications into its predictions.

Third, the protein–protein interaction network provides a good platform for integrating various biological data. As shown above (see Effects of integrating other kinds of biological data), integration of various data is essential for more comprehensive understanding and prediction of cooperative TFs. Particularly, integration of gene expression data on time course is effective. By selecting time-specific and housekeeping proteins based on expression data and constructing a time-specific protein–protein interaction network from these proteins, our method can be extended to detect time-specific, or dynamic, cooperativity among TFs.

Finally, we must note that both predictions made by protein–protein interaction data and by expression profile data fairly depend on parameters and datasets, and that the same methods may produce different prediction results by using some elaborately selected dataset. As for datasets, data from high-throughput analyses, like *in vivo* pull-down data which we used, contain a lot of false-positives that should be excluded by exploiting other biological data.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank anonymous referees' comments, which help us to improve the quality of our paper. This work is supported in part by Grant-in-Aid for Scientific Research (B) No. 16300095. This work was performed in part through Special Coordination Funds for Promoting Science and Technology from the Ministry of Education, Culture, Sports, Science and Technology, the Japanese Government. Funding to pay the Open Access publication charges for this article was provided by Keio University.

Conflict of interest statement. None declared.

REFERENCES

1. Stormo,G. (2000) DNA binding sites: representation and discovery. *Bioinformatics*, **16**, 16–23.
2. Banerjee,N. and Zhang,M.Q. (2003) Identifying cooperativity among transcription factors controlling the cell cycle in yeast. *Nucleic Acids Res.*, **31**, 7024–7031.
3. Schwikowski,B., Uetz,P. and Fields,S. (2000) A network of protein–protein interactions in yeast. *Nat. Biotechnol.*, **18**, 1257–1261.
4. Eisen,M.B., Spellman,P.T., Brown,P.O. and Bostein,D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA*, **95**, 14863–14868.
5. Brown,M., Grundy,W.N., Lin,D., Cristianini,N., Sugnet,C.W., Furey,T.S., Ares,M.,Jr and Haussler,D. (2000) Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc. Natl Acad. Sci. USA*, **97**, 262–267.
6. Jansen,R., Greenbaum,D. and Gerstein,M. (2002) Relating whole-genome expression data with protein–protein interactions. *Genome Res.*, **12**, 37–46.
7. Jansen,R., Yu,H., Greenbaum,D., Kluger,Y., Krogan,N.J., Chung,S., Emili,A., Snyder,M., Greenblatt,J.F. and Gerstein,M. (2003) A bayesian networks approach for predicting protein–protein interactions from genomic data. *Science*, **302**, 449–453.
8. 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.
9. Yu,H., Zhu,X., Greenbaum,D., Karro,J. and Gerstein,M. (2004) TopNet: a tool for comparing biological subnetworks, correlating protein properties with topological statistics. *Nucleic Acids Res.*, **32**, 328–337.
10. Brun,C., Herrmann,C. and Guenoche,A. (2004) Clustering proteins from interaction networks for the prediction of cellular functions. *BMC Bioinformatics*, **5**, 95.
11. Zar,J.H. (1998) *Biostatistical Analysis*. 4th edn. Prentice Hall International, Upper Saddle River, NJ.
12. Huynen,M., Snel,B., Lathe,W. and Bork,P. (2000) Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. *Genome Res.*, **10**, 1204–1210.
13. Sprinzak,E. and Margalit,H. (2001) Correlated sequence-signatures as markers of protein–protein interaction. *J. Mol. Biol.*, **311**, 681–692.
14. Mewes,H.W., Guldener,U., Mayer,K., Mokrejs,M., Morgenstern,B., Munsterkotter,M., Rudd,S. and Weil,B. (2002) MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.*, **30**, 31–34.

15. Marchand,L.L., Hankin,J.H., Wilkens,L.R., Pierre,L.M., Franke,A., Kolonel,L.N., Seifried,A., Custer,L.J., Chang,W., Lun-Jones,A. *et al*. (2001) Combined effects of welldone red meat and smoking and rapid N-acetyltransferase 2 and CYPIA2 phenotypes in increasing colorectal cancer risk. *Cancer Epidemiol. Biomarkers Prev*., **10**, 1259–1266.
16. 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. *et al*. (1998) A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell. Biol*., **2**, 65–73.
17. McNabb,D.S., Xing,Y. and Guarente,L. (1995) Cloning of yeast HAP5: a novel subunit of a heterotrimeric complex required for CCAAT binding. *Genes Dev*., **9**, 47–58.
18. Simon,I., Barnett,J., Hanett,N., Harbison,C.T., Rinaldi,N.J., Volkert,T.L., Wyrick,J.J., Zeitlinger,J., Gifford,D.K., Jaakkola,T.S. *et al*. (2001) Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell*, **106**, 697–708.
19. 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. Cell. Biol*., **10**, 3389–3400.
20. Spector,M.S., Raff,A., DeSilva,H., 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.
21. 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.
22. Uemura,H., Koshio,M., Inoue,Y., Lopez,M.C. and Baker,H.V. (1997) The role of Gcr1p in the transcriptional activation of glycolytic genes in yeast *Saccharomyces cerevisiae*. *Genetics*, **147**, 521–532.
23. Turkel,S. and Bisson,L.F. (1999) Transcription of the HXT4 gene is regulated by Gcr1p and Gcr2p in the yeast *S. cerevisiae*. *Yeast*, **15**, 1045–1057.
24. Lee,J., Godon,C., Lagniel,G., Spector,D., Garin,J., Labarre,J. and Toledano,M.B. (1999) Yap1 and Skn7 control two specialized oxidative stress response regulons in yeast. *J. Biol. Chem*., **274**, 16040–16046.
25. Berben,G., Legrain,M. and Hilger,F. (1988) Studies on the structure, expression and function of the yeast regulatory gene PHO2. *Gene*, **66**, 307–312.
26. Ho,Y., Constanzo,M., Moore,L., Kobayashi,R. and Andrews,B.J. (1999) Regulation of transcription at the *Saccharomyces cerevisiae* start transition by Stb1 a swi6-binding protein. *Mol. Cell. Biol*., **19**, 5267–5278.
27. Koch,C., Moll,T., Neuberg,M., Ahorn,H. and Nasmyth,K. (1993) A role for the transcriptional factors Mbp1 and Swi4 in progression from G_1 to S phase. *Science*, **261**, 1551–1557.
28. Hasan,R., Leroy,C., Isnard,A.D., Labarre,J., Boy-Marcotte,E. and Toledano,M.B. (2002) The control of the yeast H_2O_2 response by the Msn2/4 transcription factors. *Mol. Microbiol*., **45**, 233–241.
29. Morgan,B.A., Banks,G.R., Toone,W.M., Raitt,D., Kuge,S. and Johnston,L.H. (1997) The Skn7 response regulator controls gene expression in the oxidative stress response of the budding yeast *Saccharomyces cerevisiae*. *EMBO J*., **16**, 1035–1044.
30. Loy,C.J., Ldall,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 gene transcription in *S. cerevisiae*. *Mol. Cell. Biol*., **19**, 3312–3327.
31. Braus,G.H., Grundmann,O., Bruckner,S. and Mosch,H.U. (2003) Aminoacid starvation and Gcn4p regulate adhesive growth and FLO11 gene expression in *Saccharomyces cerevisiae*. *Mol. Cell. Biol*., **14**, 4272–4284.
32. Gancedo,J.M. (2001) Control of pseudohyphae formation in *Saccharomyces cerevisiae*. *FEMS Microbiol. Rev*., **25**, 107–123.
33. Vyas,V.K., Kuchin,S., Berkey,C.D. and Carlson,M. (2003) Snf1 kinases with different beta-subunit isoforms play distinct roles in regulation haploid invasive growth. *Mol. Cell. Biol*., **23**, 1341–1348.
34. Fingerman,I., Nagaraj,V., Norris,D. and Vershon,A.K. (2003) Sfp1 plays a key role in yeast ribosome biogenesis *Eukaryot. Cell*, **2**, 1061–1068.
35. 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.
36. Miyoshi,K., Miyakawa,T. and Mizuta,K. (2001) Repression of rRNA synthesis due to a secretory defect requires the C-terminal silencing domain of Rap1p in *Saccharomyces cerevisiae*. *Nucleic Acids Res*., **29**, 3297–3303.
37. 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.
38. Wiederrecht,G., Seto,D. and Parker,C.S. (1988) Isolation of the gene encoding the *S. cerevisiae* heat shock transcription factor. *Cell*, **54**, 841–853.
39. 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.
40. 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.
41. Butcher,R.A. and Schreiber,S.L. (2004) Identification of Ald6p as the target of a class of small-molecule suppressors of FK506 and their use in network dissection. *Proc. Natl Acad. Sci. USA*, **101**, 7868–7873.
42. Mollapour,M., Fong,D., Balakrishnan,K., Harris,N., Thompson,S., Schuller,C., Kuchler,K. and Piper,P.W. (2004) Screening the yeast deletant mutant collection for hypersensitivity and hyper-resistance to sorbate, a weak organic acid food preservative. *Yeast*, **21**, 927–946.
43. Bardwell,L., Cook,J.G., Zhu-Shimoni,J.X., Voora,D. and Thorner,J. (1998) Differential regulation of transcription: repression by unactivated mitogen-activated protein kinase Kss1 requires the Dig1 and Dig2 proteins. *Proc. Natl Acad. Sci. USA*, **95**, 15400–15405.
44. Blaiseau,P.I., Isnard,A.D., Surdin-Kerjan,Y. and Thomas,D. (1997) Met31p and Met32p, two related zinc finger proteins, are involved in transcriptional regulation of yeast sulfur amino acid metabolism. *Mol. Cell. Biol*., **17**, 3640–3648.
45. Wheeler,G.L., Trotter,E.W., Dawes,L.W. and Grant,C.M. (2003) Coupling of the transcriptional regulation of glutathione biosynthesis to the availability of glutathione and methionine via the Met4 and Yap1 transcription factors. *J. Biol. Chem*., **278**, 49920–49928.
46. 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.
47. Koranda,M., Schleiffer,A., Endler,L. and Ammerer,G. (2000) Forkhead-like transcription factors recruit Ndd1 to the chromatin of G/M-specific promoters. *Nature*, **406**, 94–98.