GeneNetFinder2: Improved Inference of Dynamic Gene Regulatory Relations with Multiple Regulators

Kyungsook Han and Jeonghoon Lee

Abstract—A gene involved in complex regulatory interactions may have multiple regulators since gene expression in such interactions is often controlled by more than one gene. Another thing that makes gene regulatory interactions complicated is that regulatory interactions are not static, but change over time during the cell cycle. Most research so far has focused on identifying gene regulatory relations between individual genes in a particular stage of the cell cycle. In this study we developed a method for identifying dynamic gene regulations of several types from the time-series gene expression data. The method can find gene regulations with multiple regulators that work in combination or individually as well as those with single regulators. The method has been implemented as the second version of GeneNetFinder (hereafter called GeneNetFinder2) and tested on several gene expression datasets. Experimental results with gene expression data revealed the existence of genes that are not regulated by individual genes but rather by a combination of several genes. Such gene regulatory relations cannot be found by conventional methods. Our method finds such regulatory relations as well as those with multiple, independent regulators or single regulators, and represents gene regulatory relations as a dynamic network in which different gene regulatory relations are shown in different stages of the cell cycle. GeneNetFinder2 is available at http://bclab.inha.ac.kr/GeneNetFinder and will be useful for modeling dynamic gene regulations with multiple regulators.

Index Terms—Gene expression, time-series data, microarray analysis, gene regulatory network, multiple regulators, visualization, data mining, modeling

1 Introduction

Biological processes in a living cell are governed by complex regulatory interactions between genes rather than by a single gene [1]. Regulatory interactions between genes are essential for a cell to develop and adapt to the environment. Recent advances in microarray technology have produced a huge amount of gene expression data. A variety of approaches to gene selection and clustering have been proposed to analyze large-scale gene expression data [23], [24], [25], [26]. In particular, many classification and clustering methods have been applied to tumor microarray data [27], [28], [29], [30]. Another research effort in the analysis of gene expression data is concerned with identifying gene regulatory interactions and the underlying mechanisms [31], [32]. But, most methods that are typically used for finding gene regulatory interactions focus on regulatory relations between individual genes.

In recent studies, some genes are reported to cooperate with others in playing a role of regulator of another gene although each of them does not have a regulatory relation with the target gene independently [2], [3]. This implies that some genes are not regulated by individual genes but rather by a combination of several genes. Most studies for finding gene regulatory relations from gene expression data have focused on identifying gene regulatory interactions between individual genes or co-regulated genes which have the same regulator. Thus, they cannot find gene regulatory relations with multiple activators or inhibitors that jointly regulate the expression level of a target gene.

Another thing that makes gene regulatory interactions complicated is that regulatory interactions are not static, but change over time during the cell cycle [33], [34], [35]. However, many gene regulatory networks available in literatures or databases are represented as static networks in the sense that they are snapshots of gene regulatory relations at a time point or union of successive gene regulations over time. Static gene regulatory networks cannot model the temporal aspects of regulatory interactions such as the order of gene regulations or the pace of gene regulations.

In our previous study [4], we developed a qualitative method that represents dynamic nature of gene regulatory interactions between individual genes. In an effort to extend the previous work, we developed a new method that identifies dynamic regulatory interactions between a group of genes and a single gene. It constructs a dynamic network of gene regulatory interactions of several types, which have not been considered by existing methods. The rest of this paper presents the method for identifying and visualizing dynamic gene regulations and experimental results with actual gene expression data from the Yeast Genome Project [5].

2 Methods

2.1 Scoring Scheme for Gene Regulatory Relationships

Time-series microarray data can be represented as an $M \times N$ matrix $X = [x_{ij}]$, where the ith row $x_i = [x_{i1}, \ldots, x_{iN}]$

• *The authors are with the Department of Computer Science and Engineering, Inha University, Incheon 402-751, Korea. E-mail: khan@inha.ac.kr, sosal@hanmail.net.*

Manuscript received 8 Jan. 2015; revised 24 Mar. 2015; accepted 23 June 2015. Date of publication 29 June 2015; date of current version 3 Feb. 2016. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below. Digital Object Identifier no. 10.1109/TCBB.2015.2450728

corresponds to the expression levels of the ith gene measured at N different time points. Each element x_{ij} of the matrix represents the expression level of the ith gene at a particular time point j.

We analyze the time-series microarray data for similarity between gene expressions. There are several methods for measuring the similarity of gene expression such as Euclidean distance [7], Pearson correlation coefficient [8] and Spearman correlation [9]. To evaluate the regulatory relation between two genes, we modified the Pearson correlation coefficient. $R1(X, Y, i, p)$ in Equation (1) represents the score of a regulation between gene X at time point i and gene Y at time point $i + p$. p is the time delay of the gene regulation

$$R1(X,Y,i,p) = \frac{\sum_{k=1}^{N-p}(X_k - \bar{X})(Y_{k+p} - \bar{Y})}{\sqrt{\left(\sum_{k=1}^{N}(X_k - \bar{X})^2\right)\left(\sum_{k=1}^{N}(Y_k - \bar{Y})^2\right)}}. \quad (1)$$

In Equation (1), N is the total number of time points contained in the time span, X_k and Y_k are the expression levels of genes X and Y at time k, and $\bar{X}$ and $\bar{Y}$ are the average gene expression levels at all time points of the time section. The $R1$ score is in range of $[-1, 1]$. Among the total $i \times p$ candidate regulations, the regulation with the maximum absolute value of $R1(X, Y, t, p)$ above a threshold value is selected as the regulatory relation between X and Y. By default the threshold value (d) is set to include 90 percent of the $R1$ scores of all identified gene pairs, but can be changed to a different value by the user.

Using the $R1$ score, we can determine whether gene X is a candidate activator or inhibitor of gene Y. If the expression level of gene X increases before that of Y in-creases, X is a candidate activator of gene Y; if the expression level of gene X increases before that of Y decreases, X is a candidate inhibitor of Y.

2.2 Inferring Dynamic Gene Regulatory Relationships

The gene expression level is represented by the log-ratio (in base 2) of the red and green intensities, so genes with a positive log-ratio are up-regulated and genes with a negative log-ratio are down-regulated. From the gene expression profiles with a series of time points, we identify gene regulations and store them in a regulation list. In the regulation list, $+X(t)$ indicates that gene X is up-regulated at time t, and $-X(t)$ indicates that gene X is down-regulated at time t. The symbol '$\rightarrow$' represents a directional relationship between genes. There are four possible gene regulatory relations between two genes X and Y:

1) $+$ X(t1) $\rightarrow$ $+$ Y(t2): up-regulation of X at time t1 is followed by up-regulation of Y at time t2 (t2 $>$ t1).
2) $-$X(t1) $\rightarrow$ $+$ Y(t2): down-regulation of X at time t1 is followed by up-regulation of Y at time t2 (t2 $>$ t1).
3) $+$ X(t1) $\rightarrow$ $-$Y(t2): up-regulation of X at time t1 is followed by down-regulation of Y at time t2 (t2 $>$ t1).
4) $-$X(t1) $\rightarrow$ $-$Y(t2): down-regulation of X at time t1 is followed by down-regulation of Y at time t2 (t2 $>$ t1).

The regulatory relation between genes X and Y is determined by the sign of the $R1$ score of the genes. A relation with a positive $R1$ score implies that gene X activates gene Y whereas a regulation with a negative $R1$ score implies that X inhibits Y. The $R1$ score of each gene regulation is then iteratively calculated using Equation 1. For genes X and Y, the regulation with the largest absolute $R1$ score that is greater than a threshold value (d) is chosen for the regulation between the genes and represented as $R1(X, Y, t, p)$.

Algorithm 1 provides the top-level description of the algorithm for constructing an initial regulation list using the $R1$ score.

Algorithm 1. Construct an Initial Regulation List

1. For each pair of genes X and Y, $R1(X, Y, t, p)$ at every time point t and time delay $0 < p \leq 8$.
2. Select the regulation with the maximum absolute value of $R1$ (X, Y, t, p) as a candidate regulatory relation.
3. Classify the regulation into one of the four types, $+X(t) \rightarrow +Y(t+p)$, $-X(t) \rightarrow +Y(t+p)$, $+X(t) \rightarrow -Y(t+p)$, $-X(t) \rightarrow -Y(t+p)$, and add it to the regulation list.
4. If the new gene regulation is already in the regulation list, merge it with the previous regulation.
5. Go to step 2 to find the next gene regulation until no more regulation found.
6. Sort the candidate regulations in the regulation list with respect to their $R1$ scores, and remove those with $R1$ score $< d$.

The initial candidate regulations inferred by Algorithm 1 are static relations at a time point t which have the maximum absolute value of the $R1(X, Y, t, p)$ score. In order to infer dynamic gene regulatory interactions, we examine all candidate regulations with different values of time delay p. The experimental results shown in this paper were obtained with eight different values of p ($0 < p \leq 8$). If the absolute value of the R1 score with a different time delay is greater than or equal to a threshold value d, the regulation is considered as another candidate regulation between genes X and Y. In this way, we find all potential dynamic gene regulatory relationships.

2.3 Inferring Gene Regulatory Relationships with Multiple Regulators

We consider two types of multiple regulators in gene regulations. What we call *multiple-separate regulators* have a same target gene, and each of the regulators has an individual regulatory relation with the target gene. The regulators X and Y which satisfy all the conditions below are grouped as multiple-separate regulators, and the regulation with multiple-separate regulators is denoted by {X, Y} $\rightarrow$ Z in the regulation list.

1) The regulators X and Y should have a same target gene Z.
2) The regulations X $\rightarrow$ Z and Y $\rightarrow$ Z must appear at the same time point.
3) The regulations must be of the same type (either activation or inhibition).

In the other type of multiple regulators, the target gene is not regulated by any single gene of the multiple regulators, but rather by a combination of the multiple genes. We call these genes *multiple-combined regulators*, and the regulation with multiple-combined regulators is denoted

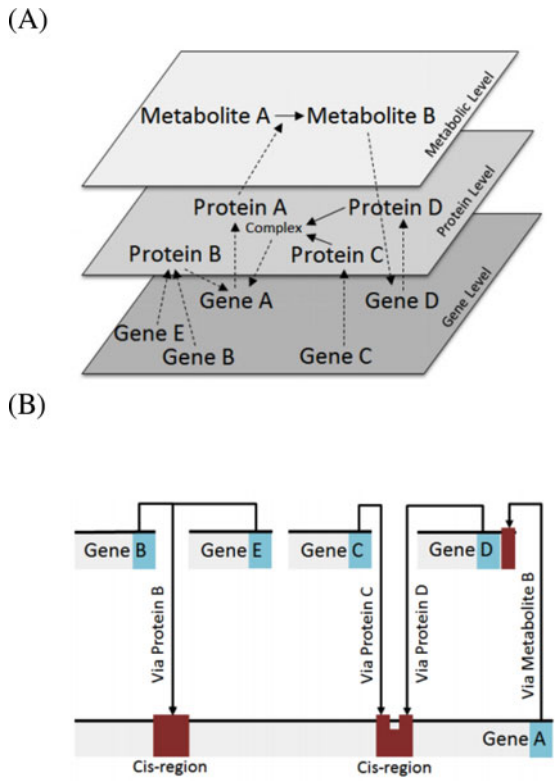

Fig. 1. A hypothetical regulatory network of genes, proteins, and metabolites at different levels (redrawn from [10]). (A) Genes are regulated by other genes, proteins, and metabolites at different levels. (B) Abstraction of the regulatory network of (A). Genes C and D bind to the *cis*-region of gene A, so genes C and D act as a combined activator of gene A.

by {X, Y} → Z. Unlike multiple-separate regulators, regulation by multiple-combined regulators X and Y cannot be found by examining individual gene regulations since individual gene regulations X → Z and Y → Z do not exist. Regulation by multiple-combined regulators is identified using the information on transcription factors of potential regulators and their binding to target genes.

Fig. 1A shows a hypothetical gene regulation network [10], [11]. There are no single regulations at the gene level. All genes are regulated by other proteins or metabolites at different levels. For example, genes B and E regulate gene A through protein B, whereas genes C and D regulate gene A through a complex of proteins C and D. Fig. 1B is an abstraction of all interactions shown in Fig. 1A. These interactions represent the binding of transcription factors to DNA during the process of transcription and localized on the cis-regions. For example, genes C and D are binding on the cis-region of gene A. Thus, genes C and D are multiple-combined regulators of gene A.

Genes X and Y that satisfy the following conditions are identified as multiple-combined regulators of their target gene Z.

1) Individual regulatory relations X → Z and Y → Z do not hold (i.e., the relations are not included in the gene regulation list).
2) The transcription factors of X and Y should satisfy one of these:
   - Genes X and Y have a same transcription factor that binds to a target gene (Fig. 2A).

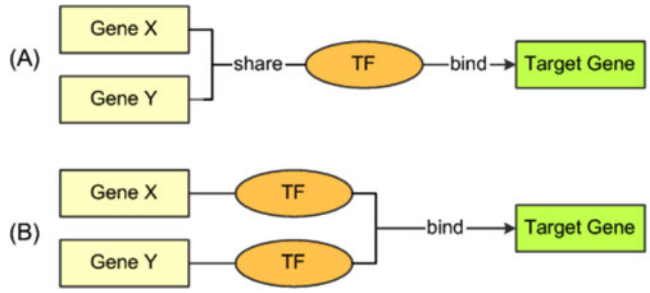

Fig. 2. Two cases of multiple-combined regulators of a target gene. (A) Genes *X* and *Y* have a same transcription factor that binds to a target gene. (B) Genes *X* and *Y* have different transcription factors that constitute a cis-regulatory element and bind to a target gene. TF: Transcription factor.

   - Genes X and Y have different transcription factors that constitute a cis-regulatory element and bind to a target gene in a promoter region (Fig. 2B).

Fig. 3 shows the whole process of finding gene regulations. Given a time-series data of gene expression, it first identifies regulations between individual genes using the R1 score. From the gene regulation list, it then finds the gene regulations with multiple-separate regulators by grouping regulations with a same target gene. For multiple-combined regulators, it first extracts data of transcription factors for all genes left after filtering out housekeeping genes and unexpressed genes. For every pair of genes with transcription factors known, it examines whether there is no regulatory relation between them (i.e., their relation is not included in the gene regulation list) and their transcription factors have a same target gene. If so, the pair of genes is a potential multiple-combined regulator of their target gene.

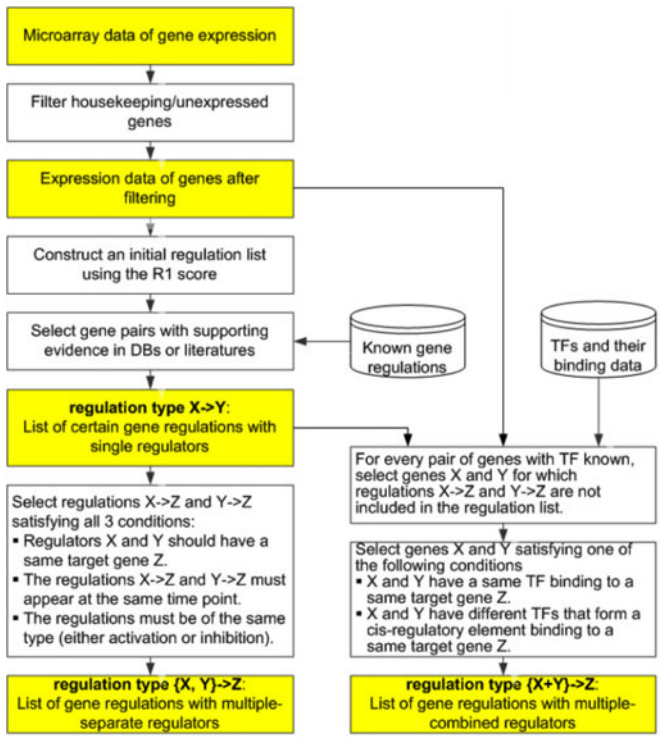

Fig. 3. The whole process of identifying gene regulations. When gene expression values are given, we compute the R1 scores to identify gene regulations between two genes. After verification, we identify gene regulations with multiple activators/inhibitions in two cases.

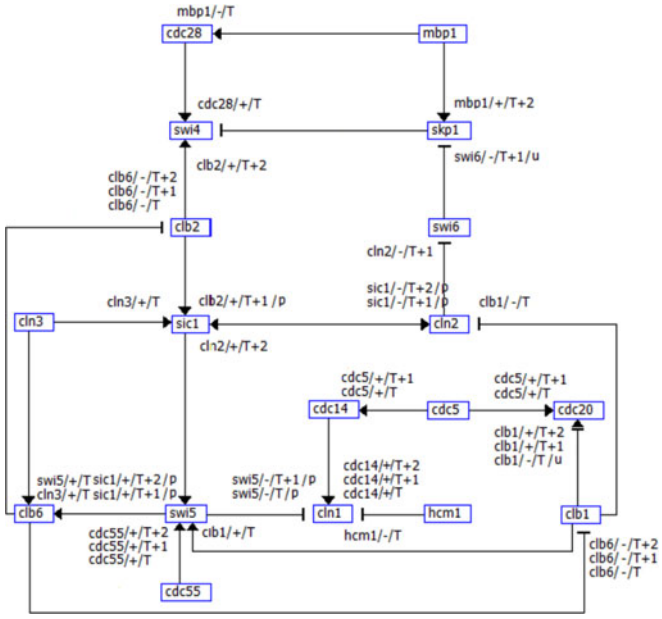

Fig. 4. Gene regulatory network visualized by GeneNetFinder2 for the yeast cell cycle. The gene regulations displayed by the grid layout algorithm. Arrows represent inductive interactions. Each edge is labeled with *R*/*s*/*T*/*p*(*u*) to indicate a regulator gene *R*, sign *s* of the expression level, the time point T of the regulation, sign *p* of the possible regulation or sign u of the uncertain regulations.

2.4 Visualization of Gene Regulatory Networks

The identified gene regulations are visualized as a two dimensional gene regulatory network, in which a node represents a gene and an edge represents gene regulatory relation. There are two types of edges in the network: arrows for activation and blocked arrows for inhibition. The regulator gene, type of regulation (+ for up-regulated of regulator or—for down-regulated of regulator), and time delay of the regulation are annotated as edge labels. Each edge is labeled with R/s/T to indicate a regulator gene(s) R, sign s of the log-ratio of the expression level of R, and the time delay T of the regulation.

Three layout algorithms have been developed to visualize a gene regulatory network: grid layout, layered layout and circular layout. The grid layout algorithm positions all nodes at grid points. The node with the highest degree will be placed at the center grid point. Then, we position all nodes connected to the center node at adjacent grid points. Nodes with a higher degree are positioned earlier than those with a lower degree in the east, north, west, south, northeast, northwest, southwest, and southeast grid point of the current node. Other nodes connected to the positioned nodes are placed in the same manner (see Fig. 4 for an example of grid layout).

The layered layout algorithm puts all nodes to horizontal layers. The node with the maximum degree is assigned to the top layer, and the nodes connected to the node are put in the next layer. If a layer has two nodes connected to each other, it makes a new layer above the layer and moves the node with a smaller degree to the new layer. The layered layout usually takes more time than the grid layout.

The circular layout algorithm places all nodes on a circle. Nodes with a higher degree are placed in an inner circle, and those with a lower degree are placed in an outer circle.

4 Results and Discussion

The algorithms have been implemented as the second version of GeneNetFinder called GeneNet-Finder2 using Microsoft Visual C#. GeneNetFinder2 is executable on any Windows systems, and the program and sample data sets are available at http://bclab.inha.ac.kr/GeneNetFinder. Given a time-series gene expression data, it identifies gene regulatory interactions of several types and visualizes them as a gene regulatory network. Fig. 6 shows the user interface and example run of GeneNetFinder2. As shown in the figure, different gene regulatory relations can be obtained from the same dataset depending on the time points of a cell cycle.

The method has been successfully tested on several datasets, and we discuss the results with the dataset of the yeast

TABLE 1
Example of Dynamic Gene Regulations in the Yeast Cell Cycle

	{X, Y} → Z		
Gene	T	T+1	T+2
CLB1	-CLB1(T) → -SWI5(T+1)		
	-CLB1(T) → -CLN3(T+2)	+CLB1(T+1) → -CLN3(T+3)	
	-CLB1(T) → -CDC20(T+1)	-CLB1(T+1) → -CDC20(T+3)	-CLB1(T+2)→-CDC20(T+4)
CLB2	-CLB2(T) → -SIC1(T+1)	+CLB2(T+1) → SIC1(T+3)	
CLB6	+CLB6(T)→-CLB1(T+1)	+CLB6(T+1)→-CLB1(T+2)	+CLB6(T+2)→-CLB1(T+3)
	+CLB6(T)→-CLB2(T+1)	+CLB6(T+1)→-CLB2(T+2)	+CLB6(T+2)→-CLB2(T+3)
CLN2		-CLN2(T+1) → -CLB1(T+3)	-CLN2(T+2) →+SIC1(T+5)
CLN3	+CLN3(T) →+SIC(T+4)		
	+CLN3(T) →+CLB6(T+2)		
CDC5	+CDC5(T) →+CDC20(T+1)		
		+CDC5(T+1) →-CLN2(T+3)	+CDC5(T+2) →+CLN3(T+5)
HCM1	+HCM1(T) →-CLB5(T+3)		+HCM1(T+2) →-CLB5(T+4)
SIC1		*-SIC1(T+1) →SIW5(T+3)*	*-SIC1(T+2) →SIW5(T+4)*
		-SIC1(T+1) →CLN2(T+3)	*-SIC1(T+2) →CLN2(T+4)*
SWI4	+SWI4(T) →+CLN2(T+1)		
SWI6	+SWI6(T) →-CLB6(T+1)		

Gene regulations identified by GeneNetFinder2 in the time-series expression data of yeast cell cycle. Underlined entries denote the regulations determined by experimental methods, and italicized entries denote the regulations implied by previous studies.

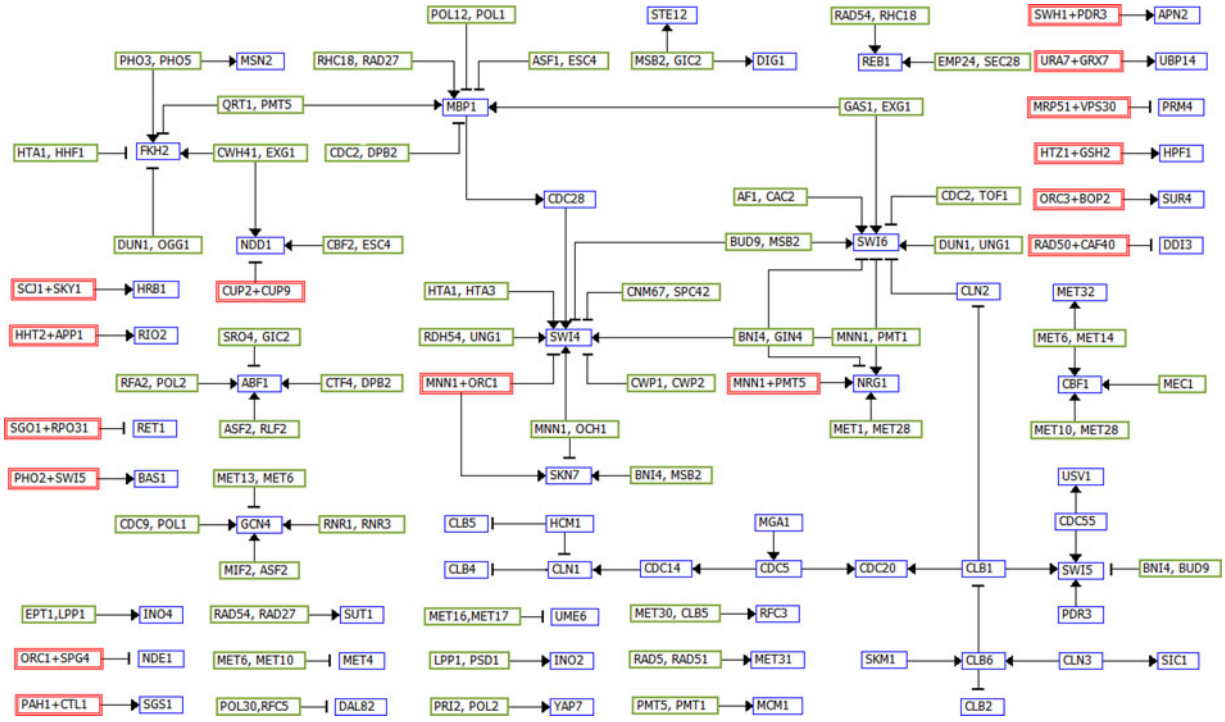

Fig. 5. A regulatory network of yeast genes in the yeast cell cycle. The gene regulations of the network are listed in Tables 1, 2, and 3. Nodes with green borders indicate multiple regulators of type {X, Y}→Z. Nodes with red borders denote multiple regulators of type {X+Y} →Z. Nodes with blue borders denote single genes.

cell cycle in this section. There are 6,178 genes in the yeast cell cycle data obtained from the Yeast Genome Project [5]. After removing housekeeping genes and unexpressed genes, 1,290 yeast genes were left in the yeast cell cycle data and were selected for identification of regulations.

In the microarray gene expression data of the yeast cell cycle, there are 18, 24, 10 time points selected for synchronization in Alpha-factor, Cdc15, and Cdc28, respectively. Cdc mutants (e.g. Cdc14, Cdc15 and Cdc28) and alpha-factor block are induction of synchronization methods for yeast cell populations. Table 1 shows the regulations between 19 yeast genes found by GeneNetFinder2. Fig. 4 shows an example of visualization of a gene regulatory network using a grid layout.

Fig. 5 shows a gene regulatory network of all the regulations listed in Tables 1, 2 and 3. Nodes with green bold border indicate multiple-separate regulators of type {X, Y} → Z. Nodes with double border lines indicate multiple-combined

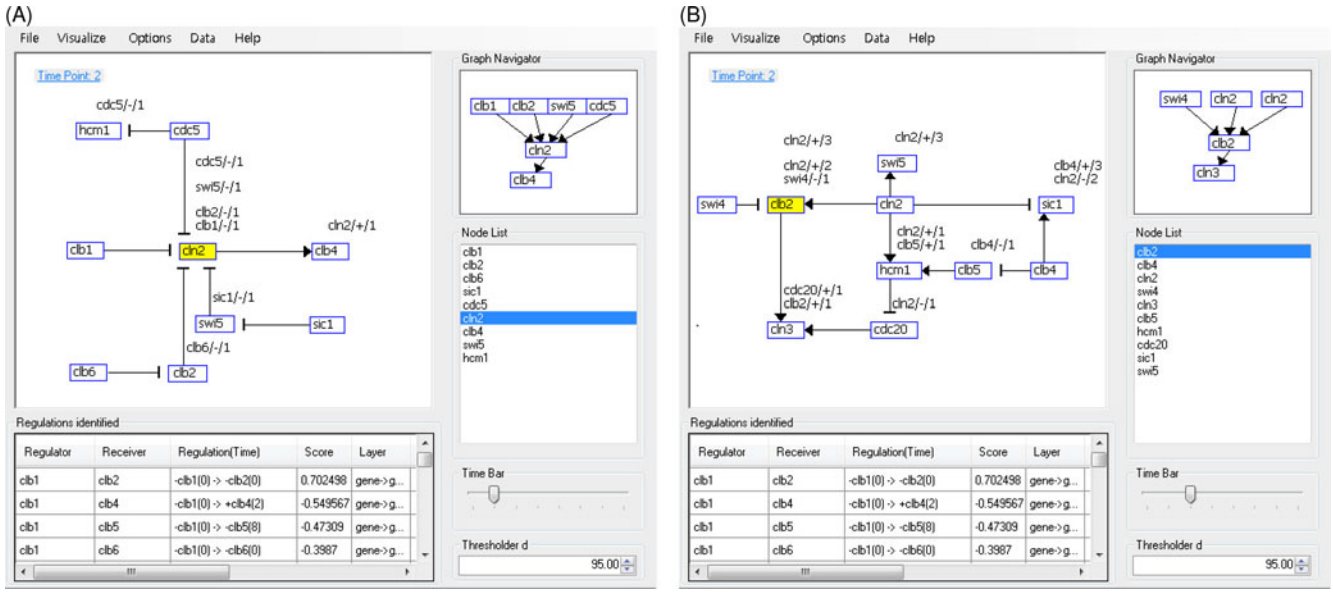

Fig. 6. User interface and example run of GeneNetFinder2. The network shown in (A) shows gene regulatory relations at time point 2 and the network in (B) represents gene regulatory relations at time point 3, which were identified in the yeast cell cycle data. When the user clicks a gene (yellow node in the network), the graph navigator in the top right corner shows all other genes that are regulated by or regulating the gene. A list of gene regulations identified by the program is shown below the network. By moving the time-bar in the lower right corner, the user can see a gene regulatory relation at a specific time point.

TABLE 2
Gene Regulations with Multiple−Separate Regulators in the Yeast Cell Cycle

{X, Y} → Z	
{+ABF1,+CAC2}(T) → +SWI6(T+1)	{+BNI4,−GIN4}(T) → −NRG1(T+1)
{+MIF2,−ASF2}(T) → +GCN4(T+1)	{+MSB2,+GIC2}(T) → −DIG1(T+1)
{+ASF2,+RLF2}(T) → +ABF1(T+1)	{+MSB2,+GIC2}(T) → −STE12(T+1)
{−ASF1,−ESC4}(T) → +MBP1(T+1)	{+SRO4,+GIC2}(T) → +ABF1(T+1)
{+CBF2,+ESC4}(T) → -NDD1(T+1)	{+MNN1,+OCH1}(T) → +SWI4(T+1)
{−HTA1,+HTA3}(T) → +SWI4(T+1)	{+MNN1,+OCH1}(T) → +SKN7(T+2)
{−HTA1,−HHF1}(T) → +FKH2(T+1)	{+MNN1,+PMT1}(T) → −NRG1(T+1)
{+RFA2,+POL2}(T) → +ABF1(T+2)	{+MNN1,+PMT1}(T) → +SWI6(T+1)
{+CTF4,+DPB2}(T) → +ABF1(T+1)	{−QRT1,−PMT5}(T) → +MBP1(T+1)
{−POL30,−RFC5}(T) → −DAL82(T+1)	{−QRT1,−PMT5}(T) → −FKH2(T+2)
{−CDC2,−TOF1}(T) → +SWI6(T+1)	{−PMT5,+PMT1}(T) → −MCM1(T+1)
{+CDC9,+POL1}(T) → −GCN4(T+1)	{+EMP24,+SEC28}(T) → +REB1(T+1)
{−PRI2,−POL2}(T) → −YAP7(T+1)	{+CWH41,+EXG1}(T) → +NDD1(T+1)
{−POL12,−POL1}(T) → +MBP1(T+1)	{+CWH41,+EXG1}(T) → +FKH2(T+2)
{−CDC2,−DPB2}(T) → +MBP1(T+2)	{−GAS1,+EXG1}(T) → +MBP1(T+1)
{+GAS1,+EXG1}(T) → +SWI6(T+1)	{+PHO3,+PHO5}(T) → −FKH2(T+1)
{+CWP1,+CWP2}(T) → +SWI4(T+1)	{+PHO3,+PHO5}(T) → +MSN2(T+2)
{+EPT1,−LPP1}(T) → −INO4(T+1)	{−MET1,−MET28}(T) → −NRG1(T+1)
{+LPP1,+PSD1}(T) → −INO2(T+2)	{+MET16,+MET17}(T) → +UME6(T+1)
{+CNM67,−SPC42}(T) → +SWI4(T+1)	{+MET30,−CLB5}(T) → −RFC3(T+1)
{+RHC18,+RAD27}(T) → +MBP1(T+1)	{−RNR1,−RNR3}(T) → −GCN4(T+1)
{+RAD54,+RAD27}(T) → +SUT1(T+1)	{+BNI4,−BUD9}(T) → −SWI5(T+1)
{+RAD54,+RHC18}(T) → +REB1(T+1)	{+BNI4,+MSB2}(T) → +SKN7(T+1)
{−DUN1,+OGG1}(T) → +FKH2(T+1)	{+BNI4,−GIN4}(T) → +MBP1(T+1)
{−DUN1,−UNG1}(T) → +SWI6(T+1)	{+BNI4,−GIN4}(T) → +SWI6(T+2)
{+RDH54,−UNG1}(T) → +SWI4(T+1)	{−BUD9,+MSB2}(T) → +SWI6(T+1)
{+RAD5,+RAD51}(T) → +MET31(T+1)	{−BUD9,+MSB2}(T) → +SWI4(T+2)
{−BNI4,+GIN4}(T) → +SWI4(T+1)	{+MET6,+MET10}(T) → +MET4(T+1)
{−MET13,+MET6}(T) → −GCN4(T+1)	{+MET6,+MET14}(T) → +CBF1(T+1)
{+MET6,+MET14}(T) → −MET32(T+1)	{+MET10,−MET28}(T) → +CBF1(T+1)

Gene regulations with multiple−separate regulators in the yeast cell cycle. Two genes X and Y have individual regulatory relation with their target gene Z.

regulators of type {X + Y} → Z. Nodes with blue border represent single genes.

For example, the R1 score for CLB1 and SWI5 is 0.737 and both genes show similar expression patterns. The program inferred that up-regulation of CLB1 is followed by up-regulation of SWI5 at the next time point {−CLB1 (T)→−SWI5(T + 1)}, which agrees with the experimental results of Althoefer [12]. SWI5 encodes a transcription factor that activates transcription of genes expressed at the M/G1 boundary and in G1 phase.

TABLE 3
Gene Regulations with Multiple−Combined Regulators in the Yeast Cell Cycle

{X + Y} → Z	
{PHO2+SWI5} → BAS1	{MNN1+PMT5} → NRG1
{ORC1+SPG4} → NDE1	{SWH1+PDR3} →APN2
{MNN1+ORC1} →SKN7	{PAH1+CTL1} → SGS1
{URA7+GRX7} → UBP14	{CUP2+CUP9} → NDD1
{SCJ1+SKY1} → HRB1	{ORC3+BOP2} → SUR4
{HTT2+APP1} → RIO2	{MRP51+VPS30} → PRM4
{RAD50+CAF40} → DDI3	{SGO1+RPO31} → RET1
{HTZ1+GSH2} → HPF1	{ESA1+SLY41} → ECM23
{MNN1+ORC1} → SWI4	

Gene regulations with multiple−combined regulators in the yeast cell cycle. The target gene Z is not regulated by any of X and Y individually, but only by a combination of X and Y.

Using the R1 score, 1,378 gene regulations were identified between the 1,290 yeast genes. 1,083 (78.6 percent) out of the 1,378 gene regulations have supporting evidences in literatures or databases [13], [20] and previous studies [21], [22]. The remaining 295 gene regulations are considered uncertain.

From the gene regulations with supporting evidences, we identified 60 multiple-separate regulations by grouping those with the same target gene at the same time point (Table 2). For example, genes ABF1 and CAC2 activate gene SWI6 after a short delay, so their regulatory relation is represented by { + ABF1, + CAC2}(T) → + SWI6(T + 1). For every pair of genes with known transcription factors, we also found multiple-combined regulations in which a target gene is not regulated by any single gene of the multiple regulators but by a combination of them (Table 3).

5 Conclusions

Most methods for finding gene regulatory interactions from microarray data focus on identifying regulatory relations between individual genes without considering the temporal aspects of the regulatory interactions such as the order of gene regulations or the pace of gene regulations. In this study we developed a method for modeling dynamic gene regulatory relations from the time-series data of gene expression. Unlike other methods that find regulations between single genes, our method can identify regulations

by single regulators and those by multiple regulators that work individually or in combination. From the microarray data of gene expression, it infers gene regulatory interactions and the temporal aspects of the regulatory interactions. The identified gene regulatory interactions and their temporal aspects are stored in the regulation list and visualized as a two-dimensional gene regulatory network.

The method described in the paper has been implemented as a working program called GeneNetFinder2 (http://bclab.inha.ac.kr/GeneNetFinder). GeneNetFinder2 is capable of finding complex gene regulatory relations involving multiple regulators, which could not be found by the first version of GeneNetFinder. We tested GeneNetFinder2 on several datasets. In the yeast cell cycle data, GeneNetFinder2 identified a total of 1,378 gene regulations, and 1,083 regulations (78.6 percent) were verified. Seventeen regulations involve multiple-combined regulators, and 60 regulations have multiple-separate regulators. The approach of GeneNetFinder2 would be useful for identifying dynamic gene regulatory interactions, especially those with multiple regulators that work in combination or individually, and for analyzing known regulatory relations.

Acknowledgments

This work was supported by the Basic Science Research Program through the National Research Foundation (NRF) funded by the Ministry of Science, ICT & Future Planning (2015R1A1A3A04001243), and in part by the Ministry of Education (2010-0020163) and INHA UNIVERSITY Research Grant (52057).

References

[1] A. Silvescu and V. Honavar, "Temporal Boolean network models of genetic networks and their inference from gene expression time series," *Complex Syst.*, vol. 13, no. 1, pp. 61–78, 2001.

[2] O. Hobert, "Gene regulation by transcription factors and MicroRNAs," *Science*, vol. 319, no. 5871, pp. 1785–1786, 2008.

[3] B. Snel, V. V. Noort, and M. A. Huynen, "Gene co-regulation is highly conserved in the evolution of eukaryotes and prokaryotes," *Nucleic Acids Res.*, vol. 32, no. 16, pp. 4725–4731, 2004.

[4] Y. Chen, B. Park, and K. Han, "Qualitative reasoning of dynamic gene regulatory interactions from gene expression data," *BMC Genomics*, vol. 11, no. S14, 2010.

[5] P. T. Spellman, G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. O. Brown, D. Bostein, and B. Futcher, "Comprehensive identification of cell cycle-regulated genes of the yeast saccharomyces cerevisiae by microarray hybridization," *Mol. Biol. Cell*, vol. 9, no. 12, pp. 3273–3297, 1998.

[6] M. L. Whitfield, G. Sherlock, A. J. Saldanha, J. I. Murray, C. A. Ball, K. E. Alexander, J. C. Matese, C. M. Perou, M. M. Hurt, P. O. Brown, and D. Botstein, "Identification of genes periodically expressed in the human cell cycle and their expression in tumors," *Mol. Biol. Cell*, vol. 13, no. 6, pp. 1977–2000, 2002.

[7] P. Danielsson, "Euclidean distance mapping," *Comput. Graph. Image Process.*, vol. 14, no. 3, pp. 227–248, 1980.

[8] L. J. Rodgers and W. A. Nicewander, "Thirteen ways to look at the correlation coefficient," *Am. Statistician*, vol. 42, no. 1, pp. 59–66, 1988.

[9] C. Spearman, "The proof and measurement of association between two things," *Am. J. Psychol.*, vol. 15, no. 1, pp. 72–101, 1904.

[10] P. Brazhnik, A. de la Fuente A., and P. Mendes, "Gene networks: How to put the function in genomics," *Trends Biotechnol.*, vol. 20, no. 11, pp. 467–472, 2002.

[11] P. Yitzhak, S. Priya, and M. C. George, "Identifying regulatory networks by combinational analysis of promoter elements," *Nature Genetics*, vol. 29, no. 2, pp. 153–159, 2001.

[12] H. Althoefer, A. Schleiffer, K. Wassmann, A. Nordheim, and G. Ammerer, "Mcm1 is required to coordinate G2-specific transcription in Saccharomyces cerevisiae," *Mol. Cell. Biol.*, vol. 15, no. 11, pp. 5917–5928, 1995.

[13] A. Amon, M. Tyers, B. Futcher, and K. Nasmyth, "Mechanism that help the yeast cell cycle clock tick: G2 cyclins transcriptionally activate G2 cyclins and repress G1 cyclins," *Cell*, vol. 74, no. 6, pp. 993–1007, 1993.

[14] L. Dirick, T. Bohm, and K. Nasmyth, "Roles and regulation of Cln-Cdc28 kinases at the start of the cell cycle of Saccharomyces cerevisae," *EMBO J.*, vol. 14, no. 19, pp. 4803–4813, 1993.

[15] M. Tyers, G. Tokima, and B. Futcher, "Comparison of the Saccaromyces cerevisae G1 cyclins: Cln3 may be an upstream activator of cln1, cln2 and other cyclins," *EMBO J.*, vol. 12, no. 5, pp. 1955–1968, 1993.

[16] E. Schwob and K. Nasmyth, "CLB5 and CLB6, a new pair of B cyclins involved in DNA replication in Saccharomyces cerevisae," *Genes Develop.*, vol. 7, no. 7a, pp. 1160–1175, 1993.

[17] B. Andrews and V. Measday, "The cyclin family of budding yeast: Abundant use of a good idea," *Trends Genet.*, vol. 14, no. 2, pp. 66–72, 1998.

[18] P. Nash, X. Tang, S. Orlicky, Q. Chen, F. B. Gertler, M. D. Mendenhall, F. Sicheri, T. Pawson, and M. Tyers, "Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication," *Nature*, vol. 414, no. 6863, pp. 514–521, 2001.

[19] R. Visintin, F. Stegmeier, and A. Amon, "The role of the polo kinase Cdc5 in controlling Cdc14 localization," *Mol. Biol. Cell*, vol. 14, no. 11, pp. 4486–4498, 2003.

[20] J. Ogas, B. J. Andrews, and I. Herskowitz, "Transcriptional activation of CLN1, CLN2, and a putative new G1 cyclin (HCS26) by SWI4, a positive regulator of G1-specific transcription," *Cell*, vol. 66, no. 5, pp. 1015–1026, 1991.

[21] D. Knapp, L. Bhoite, D. J. Stillman, and K. Nasmyth, "The transcription factor Swi5 regulates expression of the cyclin kinase inhibitor p40SIC1," *Mol. Cell. Biol.*, vol. 16, no. 10, pp. 5701–5707, 1996.

[22] A. Lovrics, A. Csikász-Nagy, I. G. Zsély, J. Zador, T. Turanyi, and B. Novak, "Time scale and dimension analysis of a budding yeast cell cycle model," *BMC Bioinformat.*, vol. 7, no. 1, p. 494, 2006.

[23] C. H. Zheng, L. Zhang, V. T. Y. Ng, C. K. Shiu, and D. S. Huang, "Molecular pattern discovery based on penalized matrix decomposition," *IEEE/ACM Trans. Comput. Biol. Bioinformat.*, vol. 8, no. 6, pp. 1592–1603, Nov. 2011.

[24] J. P. Brunet, P. Tamayo, T. R. Golub, and J. P. Mesirov, "Metagenes and molecular pattern discovery using matrix factorization," *Proc. Nat. Acad. Sci. United States of America*, vol. 101, no. 12, pp. 4164–4169, 2004.

[25] C. H. Zheng, D. S. Huang, and L. Shang, "Feature selection in independent component subspace for microarray data classification," *Neurocomputing*, vol. 69, no. 16, pp. 2407–2410, 2006.

[26] Y. Leung and Y. Hung, "A multiple-filter-multiple-wrapper approach to gene selection and microarray data classification," *IEEE/ACM Trans. Comput. Biol. Bioinformat.*, vol. 7, no. 1, pp. 108–117, Jan.–Mar. 2010.

[27] C. H. Zheng, D. S. Huang, D. Zhang, and X. Z. Kong, "Tumor clustering using nonnegative matrix factorization with gene selection," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 4, pp. 599–607, Jul. 2009.

[28] C. H. Zheng, L. Zhang, T. Y. Ng, C. K. Shiu, and D. S. Huang, "Metasample-based sparse representation for tumor classification," *IEEE/ACM Trans. Comput. Biol. Bioinformat.*, vol. 8, no. 5, pp. 1273–1282, Sep./Oct. 2011.

[29] D. S. Huang and C. H. Zheng, "Independent component analysis-based penalized discriminant method for tumor classification using gene expression data," *Bioinformatics*, vol. 22, no. 25, pp. 1855–1862, 2006.

[30] S. L. Wang, Y. H. Zhu, W. Jia, and D. S. Huang, "Robust classification method of tumor subtype by using correlation filters," *IEEE/ACM Trans. Comput. Biol. Bioinformat.*, vol. 9, no. 2, pp. 580–591, Mar. /Apr. 2012.

[31] M. B. Gerstein, A. Kundaje, M. Hariharan, S. G. Landt, K.-K. Yan, C. Cheng, et al., "Architecture of the human regulatory network derived from ENCODE data" *Nature*, vol. 489, no. 7414, pp. 91–100, 2012.

[32] V. Belcastro, et al., "Reverse engineering and analysis of genome-wide gene regulatory networks from gene expression profiles using high-performance computing," *IEEE/ACM Trans. Comput. Biol. and Bioinformatics*, vol. 9, no. 3, pp. 668–674, May/Jun. 2012.

[33] Z. Joseph, A. Gitter, and I. Simon, "Studying and modelling dynamic biological processes using time-series gene expression data," *Nat. Rev. Genetics*, vol. 13, pp. 552–564, 2012.
[34] M. H. Schulz, K. V. Pandit, C. L. Lino Cardenas, N. Amba-lavanan, N. Kaminski, and Z. Bar-Joseph, "Reconstructing dynamic micro-RNA-regulated interaction networks," *Proc. Nat. Acad. Sci. United States of America*, vol. 110, no. 39, pp. 15686–15691, 2013.
[35] O. ElBakry, M. Ahmad, and M. Swamy, "Inference of gene regulatory networks with variable time delay from time-series microarray data," *IEEE/ACM Trans. Comput. Biol. Bioinformat.*, vol. 10, no. 3, pp. 671–687, May/Jun. 2013.

Kyungsook Han received the BS cum laude degree from Seoul National University in 1983, the MS cum laude degree in computer science from Korea Advanced Institute of Science and Technology (KAIST) in 1985, another MS degree in computer science from the University of Minnesota at Minneapolis in 1989, and the PhD degree in computer science from Rutgers University in 1994. She is a professor in the Department of Computer Science and Engineering, Inha University, Incheon, Korea. Her research areas include bioinformatics, visualization, and data mining.

Jeonghoon Lee received the bachelor's degree in information & communication engineering from Inha University in 2014. His research interest is biomedical informatics.