



Quantitative characterization of the transcriptional regulatory network in the yeast cell cycle

Hong-Chu Chen¹, Hsiao-Ching Lee², Tsai-Yun Lin², Wen-Hsiung Li^{3,4} and Bor-Sen Chen^{1,*}

¹Department of Electrical Engineering and ²Department of Life Science and Institute of Bioinformatics and Structural Biology, National Tsing Hua University, Hsinchu, 300, Taiwan, ³Department of Ecology and Evolution, University of Chicago, USA and ⁴Genomics Research Center, Academia Sinica, Taipei, Taiwan

Received on August 4, 2003; revised on January 10, 2004; accepted on January 14, 2004
Advance Access publication March 25, 2004

ABSTRACT

Motivation: Genome-wide gene expression programs have been monitored and analyzed in the yeast *Saccharomyces cerevisiae*, but how cells regulate global gene expression programs in response to environmental changes is still far from being understood. We present a systematic approach to quantitatively characterize the transcriptional regulatory network of the yeast cell cycle. For the interpretative purpose, 20 target genes were selected because their expression patterns fluctuated in a periodic manner concurrent with the cell cycle and peaked at different phases. In addition to the most significant five possible regulators of each specific target gene, the expression pattern of each target gene affected by synergy of the regulators during the cell cycle was characterized. Our first step includes modeling the dynamics of gene expression and extracting the transcription rate from a time-course microarray data. The second step embraces finding the regulators that possess a high correlation with the transcription rate of the target gene, and quantifying the regulatory abilities of the identified regulators.

Results: Our network discerns not only the role of the activator or repressor for each specific regulator, but also the regulatory ability of the regulator to the transcription rate of the target gene. The highly coordinated regulatory network has identified a group of significant regulators responsible for the gene expression program through the cell cycle progress. This approach may be useful for computing the regulatory ability of the transcriptional regulatory networks in more diverse conditions and in more complex eukaryotes.

Contact: bschen@moti.ee.nthu.edu.tw

Supplementary information: Matlab code and test data are available at http://www.ee.nthu.edu.tw/~bschen/quantitative/regulatory_network.htm

INTRODUCTION

Living cells are produced by a set of specific gene expression programs of transcription of thousands of genes under a coordinated arrangement. The regulation of gene expression programs in eukaryotes depends on the recognition of specific promoter sequences by transcriptional regulatory proteins (Orphanides and Reinberg, 2002; Garvie and Wolberger, 2001) and often occurs through the coordinated action of multiple transcriptional regulators (Fickett and Wasserman, 2000; Halfon *et al.*, 2000). Combinational regulation of transcription has the advantages of including the control of gene expression in response to environmental changes and the use of limited number of transcription factors. Many combinations of regulators are produced and their regulatory activities are modified through diverse sets of conditions. Thus, the expression of each gene relies on the specific processes of a number of regulatory inputs, which are still unknown in most cases.

The molecular machinery of cell cycle control has been described in detail in the budding yeast *Saccharomyces cerevisiae*. The periodic alternation of gene transcript levels in *S.cerevisiae* is a major driving force of the cell cycle. Expression levels of hundreds of genes fluctuate throughout the cell cycle (Spellman *et al.*, 1998; Cho *et al.*, 1998) and the self-degradation rates of mRNA have been precisely measured (Wang *et al.*, 2002). The functional classes of genes that show periodic transcription reflect the events that are occurring throughout the cell cycle, including bud emergence and growth, DNA synthesis, and spindle pole body duplication and migration. Therefore, it is of interest to decipher how these periodic recurring patterns are regulated.

Genome-wide microarray gene expression data relevant to the yeast cell cycle have been collected in a parallel manner (Spellman *et al.*, 1998; Cho *et al.*, 1998; Zhu *et al.*, 2000). Since the gene expression profile data are the network interactions resulting from the regulators and the target genes,

*To whom correspondence should be addressed.

it is reasonable to trace the interaction network from the microarray gene expression data. This treasure trove of data has been analyzed with a variety of clustering methods to group together genes with similar patterns of expression (Eisen *et al.*, 1998; Tamayo *et al.*, 1999; Tavazoie *et al.*, 1999). As mRNA levels are the outputs of gene expression regulatory networks, it is theoretically possible to use expression data to reversely engineer the network architecture. A number of groups have coped with this problem using singular value decomposition (SVD) analysis (Alter *et al.*, 2000; Holter *et al.*, 2001; Yeung *et al.*, 2002). Instead of grouping genes together by their expression pattern, one can group genes by promoter sequence motif and then check whether this group of genes is co-expressed (Bussemaker *et al.*, 2001) or find combinations of transcription factors or binding sites (Pilpel *et al.*, 2001; GuhaThakurta and Stormo, 2001). To understand how genes are controlled by transcription factors, a genome-wide location analysis has been developed to identify which transcription factors bind to which promoters (Iyer *et al.*, 2001; Simon *et al.*, 2001; Lee *et al.*, 2002). The information above provides us a clue as to how to decipher the gene expression regulatory network.

As an alternative approach, we have devised a method for quantitatively characterizing transcriptional regulatory network of the yeast cell cycle from genome-wide time-course expression data. We have not only identified the upstream suspected regulatory genes but also characterized the regulation significance by quantifying the regulatory ability. Our first step is to construct a dynamic model of gene expression using the maximum likelihood method together with the Akaike Information Criterion (AIC). This allows us to measure the transcription rate that controls the synthesis rate of a gene-specific mRNA level. The second step is to select a group of genes with regulatory ability through the cell cycle as regulator candidates and then deduce the relationship between the gene expression profiles of regulators and the transcription rates of specific target genes. Our algorithm iteratively picks out the genes with high correlations to the transcription rate of target genes, and then identifies the associated parameters by minimizing the difference between the combined effects of regulators and the transcription rate of target genes.

We applied our method to a publicly available data set of yeast from microarray experiments on the cell cycle (Spellman *et al.*, 1998), and used the measured mRNA decay rates (Wang *et al.*, 2002) to identify the transcriptional regulators of the cell cycle and characterize their regulatory abilities on specific target genes. Through the quantitative analysis of the yeast transcriptional regulatory network of the cell-cycle-regulated genes, several transcription factors were identified and their regulatory abilities determined. In addition, a few genes known to regulate kinase activity are predicted here to be synergistic in coordinating gene expression. Our transcriptional regulatory network depicts the combined interactions of a small number of transcriptional regulators that are responsible

for the gene expression program under certain conditions. This work provides a new approach for revealing gene regulatory network from gene expression data using a mathematical dynamic model and system identification techniques. Our approach is rather different from most existing statistical clustering methods for analyzing gene expression data, and our results show that this novel method is suitable for deciphering the complex interactions that regulate gene expression.

METHODS

Our method for quantitatively characterizing the transcriptional regulatory network can be divided into two steps. In the first step, based on a dynamic modeling of gene expression, the transcription rate and expression level of each gene are extracted from the raw microarray data. In the second step, based on the transcription rate of the target gene, the upstream regulators will be deduced by correlating the transcription rate of the target gene with the expression level of regulators.

Dynamic modeling of gene expression

The dynamics of gene expression can be modeled by a simple first-order differential equation, which is well established (Singer and Penman, 1973) and analyzed (Šášik *et al.*, 2002). Let $X_i(t)$ denote the expression level of the i -th gene at time t in a particular experiment. The differential equation has the form

$$\dot{X}_i(t) = G_i(t) - \lambda_i X_i(t) + \xi_i(t), \quad (1)$$

where $G_i(t)$ and λ_i are the transcription rate and the self-degradation rate, respectively, of a specific gene i , and the last term $\xi_i(t)$ represents noise to stand for the data uncertainty and the model residuals. There is a time constant delay τ_i , $\tau_i = 1/\lambda_i$, from input $G_i(t)$ to output $X_i(t)$. The dynamic Equation (1) defines the changing rate of the mRNA level as the difference between the transcription rate $G_i(t)$ controlling the synthesis rate and the natural degradation rate of the mRNA molecule, $\lambda_i X_i(t)$. The transcription rate $G_i(t)$ is, in general, a complex function of combined effects of all upstream relevant regulators which determine the transcript level of a specific target gene.

To dissect the machinery controlling the dynamic model, the appropriate extraction of transcription rate $G_i(t)$ plays a critical role. We assume that the transcription rate is a linear combination of time dependent switch functions:

$$G_i(t) = \sum_{j=1}^m a_{ij} S(t, \theta_j), \quad (2)$$

where $S(t, \theta_j)$ is a switch function of time specified by the parameter vector $\theta_j = \{\alpha, \beta_j, \gamma, \delta\}$ which determines the shape and transition rate of the switch function and a_{ij} is an associated parameter that controls the weight of the specific switch function.

$$S(t, \theta_j) = \frac{1}{1 + \exp\{-\alpha[u(t, \beta_j, \delta) - \gamma]\}},$$

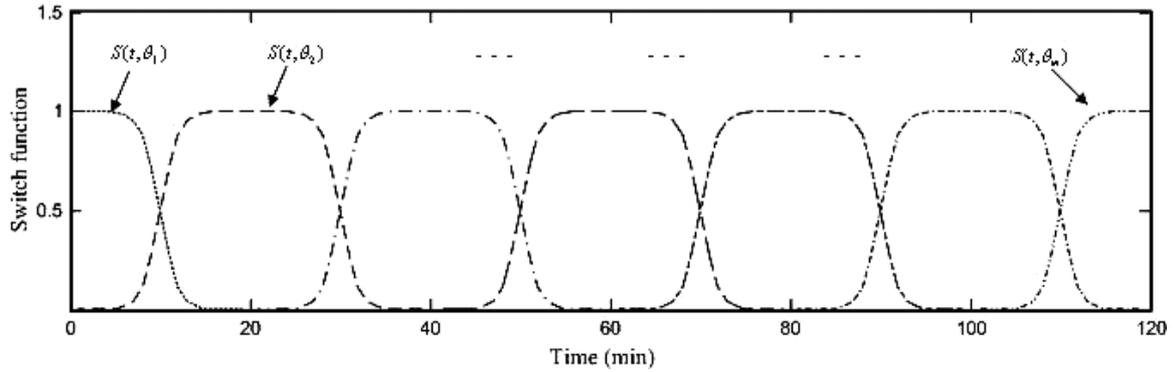


Fig. 1. Illustration of switch functions, different dot lines indicate different switch functions $S(t, \theta_j)$.

and

$$u(t, \beta_j, \delta) = \begin{cases} 0 & t \leq \beta_j - \delta \\ \frac{t - (\beta_j - \delta)}{\delta} & \beta_j - \delta \leq t \leq \beta_j \\ \frac{(\beta_j + \delta) - t}{\delta} & \beta_j \leq t \leq \beta_j + \delta \\ 0 & t \geq \beta_j + \delta. \end{cases}$$

i.e. the different values of the parameter a_{ij} will determine the different forms of the transcription rate $G_i(t)$. The parameter vector θ_j determines the distribution of the switch function $S(t, \theta_j)$, where α and γ represent the transition slope and the crossover threshold of the switch function, respectively, $u(t, \beta_j, \delta)$ represents a triangular function and the parameters $\{\beta_j - \delta, \delta, \beta_j + \delta\}$ determine the three corners of the underlying triangular function. Figure 1 illustrates the distribution of different switch functions $S(t, \theta_j)$. These switch functions are employed to synthesize the transcription rate $G_i(t)$ in Equation (2) by specifying appropriate parameters a_{ij} . If the number of switch functions is increased, the resolution of synthesizing $G_i(t)$ in Equation (2) will be increased simultaneously.

Reconsider Equations (1) and (2). If we have a set of data points $\{\dot{X}_i(t_k), X_i(t_k)\}$, $k \in \{1, \dots, N\}$ and the measured decay rates of λ_i mRNA for the i -th gene expression, the parameters a_{ij} can be estimated for $j = 1, \dots, m$ by the following approach. Using a matrix notation, we can rewrite these two equations in a concise form:

$$\begin{pmatrix} \dot{X}_i(t_1) \\ \dot{X}_i(t_2) \\ \vdots \\ \dot{X}_i(t_N) \end{pmatrix} = \begin{pmatrix} S_1(t_1) & S_2(t_1) & \cdots & S_m(t_1) \\ S_1(t_2) & S_2(t_2) & \cdots & S_m(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ S_1(t_N) & S_2(t_N) & \cdots & S_m(t_N) \end{pmatrix} \times \begin{pmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{im} \end{pmatrix} - \begin{pmatrix} \lambda_i X_i(t_1) \\ \lambda_i X_i(t_2) \\ \vdots \\ \lambda_i X_i(t_N) \end{pmatrix} + \begin{pmatrix} \xi_i(t_1) \\ \xi_i(t_2) \\ \vdots \\ \xi_i(t_N) \end{pmatrix}, \quad (3)$$

where $S_j(t_k) \triangleq S(t_k, \theta_j)$ and λ_i is the measured degradation rate from the study by Wang *et al.* (2002). If λ_i is unavailable, it should be estimated together with the parameters a_{ij} . For simplicity, we can further define the notations, Y_i , A_i and Φ_i , to represent Equation (3) as

$$Y_i = A_i \cdot \Phi_i + E_i. \quad (4)$$

where Y_i denotes an output vector,

$$Y_i = \begin{pmatrix} \dot{X}_i(t_1) + \lambda_i X_i(t_1) \\ \dot{X}_i(t_2) + \lambda_i X_i(t_2) \\ \vdots \\ \dot{X}_i(t_N) + \lambda_i X_i(t_N) \end{pmatrix},$$

A_i denotes a design matrix,

$$A_i = \begin{pmatrix} S_1(t_1) & S_2(t_1) & \cdots & S_m(t_1) \\ S_1(t_2) & S_2(t_2) & \cdots & S_m(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ S_1(t_N) & S_2(t_N) & \cdots & S_m(t_N) \end{pmatrix},$$

Φ_i denotes the parameter vector,

$$\Phi_i = \begin{pmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{im} \end{pmatrix},$$

and E_i denotes the error vector to account for measurement noise and modeling errors.

$$E_i = \begin{pmatrix} \xi_i(t_1) \\ \xi_i(t_2) \\ \vdots \\ \xi_i(t_N) \end{pmatrix}.$$

If λ_i is unavailable, Equation (3) should be modified to include λ_i in the parameter vector Φ_i , e.g. $Y_i = [\dot{X}_i(t_1), \dots, \dot{X}_i(t_N)]^T$, $\Phi_i = [a_{i1}, \dots, a_{im}, \lambda_i]^T$ and A_i

should include the data set $[X_i(t_1), \dots, X_i(t_N)]^T$. We assume that each element in the error vector $\xi_i(t_k), k \in \{1, \dots, N\}$, is an independent random variable with a normal distribution with zero mean and variance σ^2 , and estimate the parameter $\hat{\Phi}_i$ using the maximum likelihood method. The probability distribution of one data point $y_i(t_k)$ in Equation (4), given t_k , is

$$p[y_i(t_k) | t_k; \Phi_i, \sigma^2] = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{[y_i(t_k) - z_i(t_k)]^2}{2\sigma^2}\right\}, \quad (5)$$

where $z_i(t_k) = R_{A_i}(t_k) \cdot \Phi_i$, and $R_{A_i}(t_k)$ is the row vector of A_i at time point t_k . The log-likelihood function for N data points given is then given by

$$L(\Phi_i, \sigma^2) = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{k=1}^N [y_i(t_k) - z_i(t_k)]^2. \quad (6)$$

The necessary condition for the maximum likelihood estimator of variance σ^2 is

$$\frac{\partial L(\Phi_i, \sigma^2)}{\partial \sigma^2} = 0.$$

By solving the preceding equation, we obtain

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{k=1}^N [y_i(t_k) - z_i(t_k)]^2. \quad (7)$$

Substitution of Equation (7) into Equation (6) gives us the following form:

$$L(\Phi_i, \hat{\sigma}^2) = -\frac{N}{2} \ln(2\pi\hat{\sigma}^2) - \frac{N}{2}.$$

This means that we can find the maximum likelihood estimation of $\hat{\Phi}_i$ by minimizing the value of $\hat{\sigma}^2$. It can be achieved by the best choice of the parameter vector $\hat{\Phi}_i$ in Equation (4) using the least squares method, such that

$$\hat{\Phi}_i = (A_i^T A_i)^{-1} A_i^T Y_i. \quad (8)$$

It was shown that the maximum log likelihood has a tendency to overestimate the true value of the expected log likelihood. This tendency becomes more potent for models with a large number of parameters. Consequently, the number of free parameters we choose for Φ_i determines the goodness of fitting the true model. Here, we take the well-known AIC (Akaike, 1974) into account for determining the number of parameter vectors which were used in Equation (3).

$$\begin{aligned} \text{AIC} = & -2 \cdot \left(\begin{array}{l} \text{maximum log likelihood} \\ \text{of the estimated model} \end{array} \right) \\ & + 2 \cdot \left(\begin{array}{l} \text{number of free parameters} \\ \text{of the estimated model} \end{array} \right). \end{aligned} \quad (9)$$

The number of parameters, i.e. in our case the value of m in Equation (3) that minimizes the AIC, is chosen to create

a specific dynamic expression model that is considered to be the most appropriate model for characterizing the expression profile in our analysis.

After the parameters a_{ij} have been estimated, the transcription rate $G_i(t)$ and expression level $X_i(t)$ can be specified from the microarray data by solving Equation (1). The next step in our work is to identify the upstream regulators by correlating the transcription rate $G_i(t)$ of target gene i with the expression level $X_j(t)$ of other genes via their regulatory functions, which will be defined in the next section.

Iterative algorithm for finding regulators

It is practical to consider that specific biochemical reactions are influenced by the concentration of relevant products (Goldbeter and Koshland, 1981; Mestl *et al.*, 1995). For this purpose, we describe the regulatory function of a specific gene as a sigmoid function:

$$f_j[X_j(t), \theta_j] = \frac{1}{1 + \exp\{-r \cdot [X_j(t) - M_j]\}}, \quad (10)$$

where $X_j(t)$ is the expression level of the j -th gene defined in Equation (1), and $\theta_j \triangleq \{r, M_j\}$ indicates the transition rate (r) of the sigmoid function and the mean expression level (M_j) of gene j throughout the experiment, respectively. To further dissect the transcriptional regulatory network from the dynamics of gene expression profiles, we derive upstream regulatory functions from the transcription rate $G_i(t)$ of the target gene in Equation (1), which has been estimated in the previous step. For simplicity, we match the transcription rate of target gene i by a combination of a set of regulatory functions from regulators:

$$G_i(t) = c_{i0} + \sum_{j \in L_i} c_{ij} f_j[X_j(t), \theta_j], \quad (11)$$

where c_{ij} is the regulatory capability from regulator j to target gene i , $f_j[X_j(t), \theta_j]$ is a regulatory function defined in Equation (10) and L_i is a possible set of indices of regulator genes for target gene i . The regulatory parameters c_{ij} are used to mimic modifications such as phosphorylation and dephosphorylation of proteins, and proteolysis through the dynamics of expression profiles, i.e. the activation and repression are characterized by these kinetic parameters c_{ij} . Similar to Equation (3), we can rewrite Equation (11) in the following matrix form:

$$\begin{pmatrix} G_i(t_1) \\ G_i(t_2) \\ \vdots \\ G_i(t_N) \end{pmatrix} = \begin{pmatrix} 1 & f_1(t_1) & f_2(t_1) & \cdots & f_L(t_1) \\ 1 & f_1(t_2) & f_2(t_2) & \cdots & f_L(t_2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & f_1(t_N) & f_2(t_N) & \cdots & f_L(t_N) \end{pmatrix} \cdot \begin{pmatrix} c_{i0} \\ c_{i1} \\ \vdots \\ c_{iL} \end{pmatrix}. \quad (12)$$

Thus, the regulatory parameters c_{ij} can be directly estimated through the least squares method according to Equation (8).

If we integrate Equations (1) and (11) into one equation, the dynamics of the transcriptional regulatory network can be represented as the following nonlinear differential equation

$$\dot{X}_i(t) = c_{i0} + \sum_{j \in L_i} c_{ij} f_j[X_j(t), \theta_j] - \lambda_i X_i(t) + \xi_i(t). \quad (13)$$

In fact, Equation (13) contains the rich information from exploring the regulatory network of each specific target gene. The genes that belong to a specific set L_i represent the potential upstream regulators of target gene i . The regulatory parameter c_{ij} characterizes the type and intensity of the influence of the j -th gene on the i -th target, with a positive sign indicating activation and a negative sign indicating repression, which will be defined as the regulatory ability hereafter. Apart from this, the first term in Equation (13), c_{i0} , refers to the basal expression of a target gene. The goal of reverse engineering the regulatory network is to deduce the possible set of regulators and to identify their associated regulation abilities from the available data set. For this purpose, we devise an iterative algorithm based on the dynamic gene expression model for searching possible upstream regulators and then identifying the relevant regulatory abilities according to Equation (12). This iterative algorithm can be constructed step-wise as follows:

Step 0. Based on the dynamic gene expression model described in Equation (1) estimate the transcription rate $G_i(t)$ and reconstruct the gene expression profile $X_i(t)$ for any interested target gene i .

Step 1. Select around 50 genes that have been confirmed to be transcription factors, cell cycle controlled genes and kinase activity regulators in the literature, and then calculate their regulatory function by Equation (10). Consider these genes as the primary regulator candidates having the putative potential to regulate the periodic expressions of cell-cycle-related genes.

Step 2. For any specific target gene i , examine the correlation between the transcription rate $G_i(t)$ and the regulatory function $f_j[X_j(t), \theta_j]$ stored in the candidate pool using a matched filter (see Appendix for details). Select a few genes from the candidate pool with higher correlation as possible regulators. If there are transcription factors among the selected genes, then consider the transcription factors as putative regulators, and identify the regulatory ability parameters c_{ij} associated with the inferred regulator and the basal expression c_{i0} by the least square method via Equation (12). Finally, an estimated transcription rate $\hat{G}_i(t)$ is generated as described in Equation (11). If there are no transcription factors found in the selected genes, generate the estimated transcription rate $\hat{G}_i(t)$ for each possible regulator and choose the gene with the minimum root mean square error (RMSE) between the estimated $\hat{G}_i(t)$ and $G_i(t)$. The L_i set in Equation (12) contains either the transcription factors or the chosen gene as putative regulators at the moment.

Step 3. In order to find regulators that might play the role of synergistic partners for regulating the target gene, define a new term $\Delta G_i(t) = G_i(t) - \hat{G}_i(t)$, which is used for correlating the remaining genes in the candidate pool instead of the original transcription rate $G_i(t)$ used in Step 2. Use the output scores of the matched filter to help recognize the higher correlation between $\Delta G_i(t)$ and possible regulators. Select the genes with the five highest and five lowest correlations in turn to put into Equation (11) and identify all the parameters again by the least squares method to form a new estimated transcription rate $\hat{G}_i(t)$. Among the 10 estimated transcription rate $\hat{G}_i(t)$ results, choose the gene with the minimum RMSE between the estimated $\hat{G}_i(t)$ and $G_i(t)$, and add the newly found gene into the regulator set L_i defined in Equation (11).

Step 4. Repeat Step 3 until the regulator set L_i contains a predetermined number of regulators (for ease in interpretation, the algorithm identifies five possible regulators). Tabulate all the final regulatory ability parameters, the basal expression level and the possible regulator set L_i . For a new interested target gene, go to Step 1 to start a new program.

After analyzing all the interested target genes according to the above iterative algorithm, we obtained useful information for characterizing the transcriptional regulatory network in the yeast cell cycle. The algorithm aims to investigate the transcriptional regulation in the cell cycle by inferring the relationship between the transcription rate $G_i(t)$ of the target gene and the transformed regulatory function $f_j[X_j(t), \theta_j]$ from the expression pattern of regulators. In addition, for easy interpretation, the algorithm only finds five possible regulators for each target gene and the target gene could be any interested gene in the whole genome, i.e. this algorithm can be applied on a genome-wide scale.

RESULTS

Yeast cell cycle data

To identify the regulators and to characterize the regulatory abilities of yeast cell-cycle-regulated genes, we applied our approach to a publicly available data set of Spellman *et al.* (1998), which contained the expression of 6178 open reading frames in *S.cerevisiae* during the cell cycle (original data sets are available at <http://cellcycle-www.stanford.edu/>). In Spellman *et al.*'s experiments, cell cultures were synchronized with different methods including α factors arrest, temperature arrest of a temperature sensitive mutant and elutriation synchronization. Our analysis was applied on the α factors arrested data set. The raw data were transformed into a linear scale from the original log ratio given by Spellman *et al.* (1998). The raw data certainly contain uncertainty. This study employs the following method to eliminate the effect of uncertainty in the estimation of the transcription rate profiles. The cubic spline was used first for data interpolation and smoothing to obtain a less sensitive first derivative of the expression

pattern and the uncertainty is considered in the noise term $\xi_i(t)$ in Equation (1). Then we used AIC combined with the least squares method to minimize the effect of uncertainty in parameter estimation. The cubic spline method has been used for a long time by engineers to fit curves through given points. A cubic spline interpolation defined and proved that any partition between each given data point is a continuous function that has continuous first and second derivatives everywhere in that corresponding interval and is represented by a polynomial of degree not exceeding three. This well-developed scheme allows us to determine a smooth curve representing the original limited experimental data and having a continuous first derivative function, which is indispensable information for our analysis.

For efficiency, the raw data was preprocessed by filtering out those genes with a flat expression profile (the difference between the maximum and the minimum expression level is <2 -fold). After filtering the original raw data of 6178 genes, 2119 genes were considered as input expression profiles $X_i(t)$ to be examined in our assay. The input to our analysis required at every stage and the entire flowchart are depicted in Figure 2. Figure 3 shows the computed results of the dynamic expression model [Equation (1)] with the experimental expression patterns for some typical cell-cycle-related genes. The transcription rates $G_i(t)$ s, defined in Equation (1), change with time (Fig. 4). According to Equation (1), the transcription rate $G_i(t)$ is the major driving force for regulating the expression profile. The good fit of our computed results and experimental data points (Fig. 3) indicates that the dynamic expression model is suitable for describing the character of gene expression and the transcription rate $G_i(t)$ (Fig. 4). We established the dynamic expression profiles of a group of genes that were previously denoted as cell cycle regulators (Spellman *et al.*, 1998; Simon *et al.*, 2001) and characterized their regulatory function according to Equation (10). Our iterative algorithm can find the most likely genes that participate in regulating the expression of any target gene of interest on a genome-wide scale.

Logical inference from regulatory synergy

An attractive feature of combined synergy between regulatory proteins, such as transcription factors, kinase activity regulators and cyclins is that it accommodates complex mechanisms of gene regulation. Figure 5 shows an example of the combinatorial gene regulation. Consider the following scenario that regulates the expression program of a specific target gene: two activators are capable of causing transcription initiation and a repressor is responsible for reducing the expression level. Although each of the regulators might work at different stages, the overall effect of transcriptional activity caused by these regulators does periodically fluctuate with time and peaks at a specific time period.

Figure 5b illustrates how the expression program is operated under cooperative gene regulation in which two activators

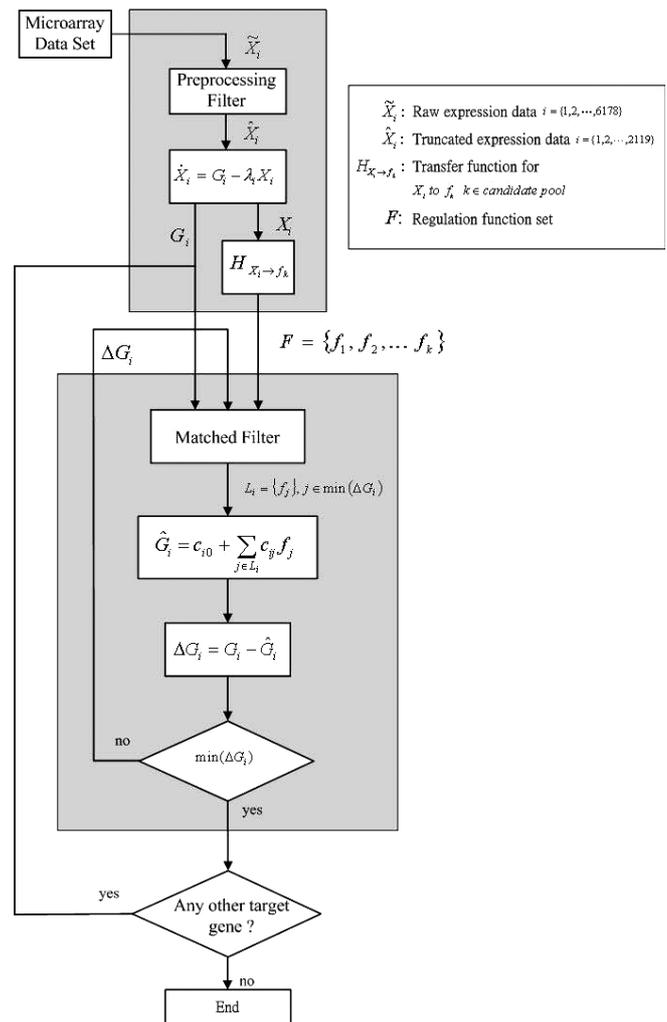


Fig. 2. The overall flowchart of the analysis and data required at every stage.

induce expression and one repressor reduces it. The top row shows the regulatory functions of three regulators x , y and z ; and the bottom row shows the overall transcription rate of the target gene expression. If we assume that the relationship of the regulatory function with the expression level is described by Equation (10), then the action period of regulators (i.e. the time period of regulatory function with the magnitude equal to 1) will determine the times of maximum and minimum transcription rates. For example, repressor z acts at three time periods (0–15, 55–75 and 115–120 min) and subsequently causes the valley in the transcription rate that leads to a low level of expression of the target gene. On the other hand, activators x and y work at later time periods; a possible explanation is that activator x decreases the synthesis of repressor z and subsequently promotes the synthesis of another activator y , which leads to an abrupt peak in the transcription rate within the overlapping time period of activators x and y . Thus, the

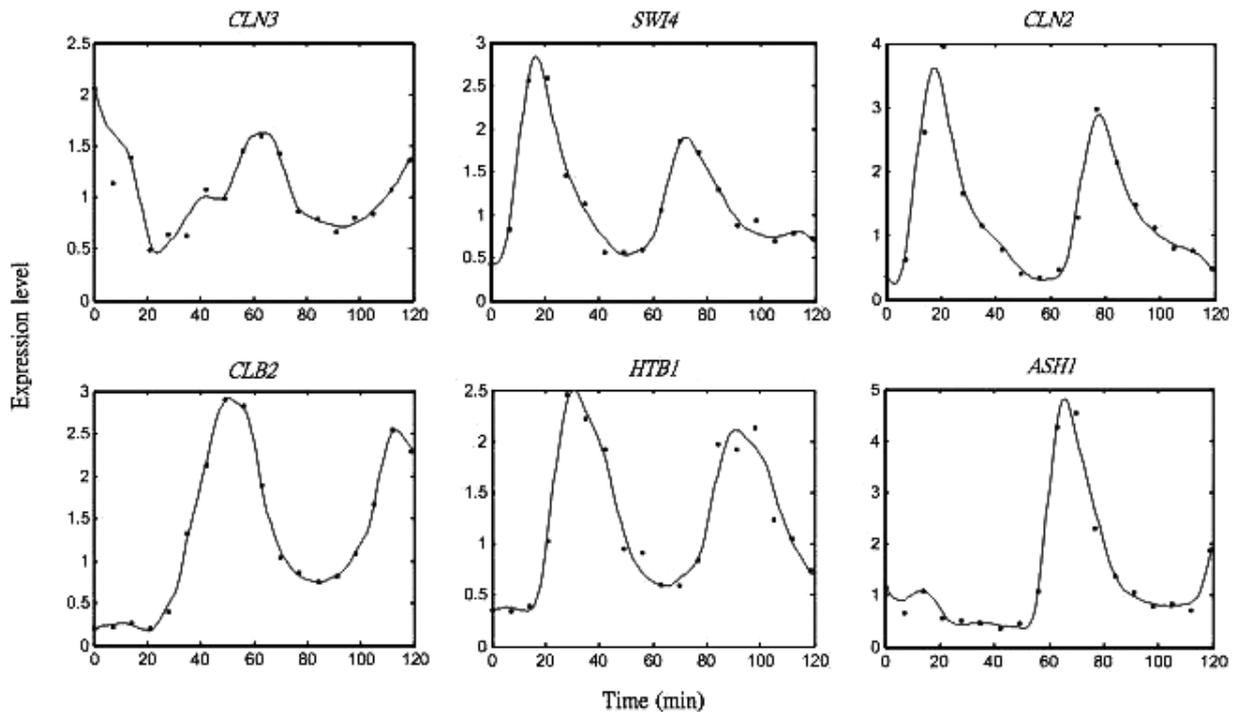


Fig. 3. The expression profiles with corresponding dynamic model fitting. The experimental data are shown as dots and the dynamic modeling results are shown as solid lines. The six cell-cycle-related genes are *CLN3*, *SWI4*, *CLN2*, *CLB2*, *HTB1* and *ASH1*.

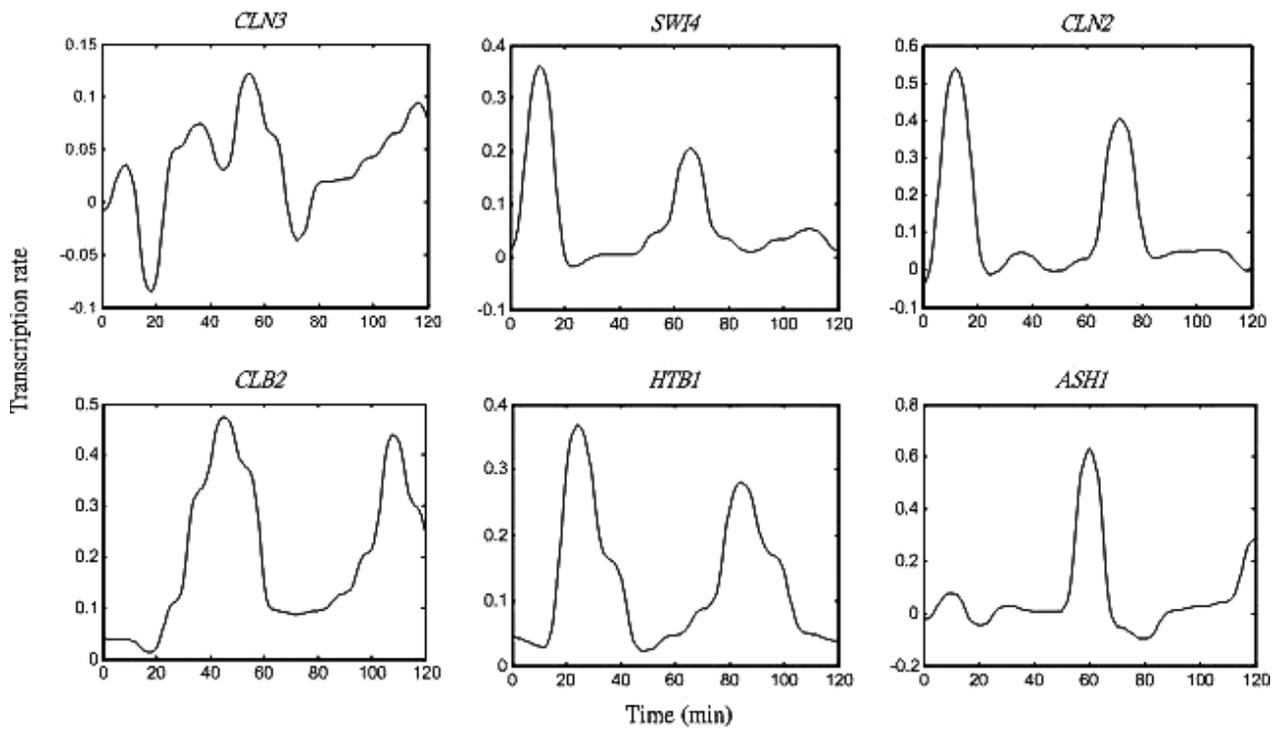


Fig. 4. The identified transcription rates, $G_i(t)$'s, for the six cell cycle-related genes *CLN3*, *SWI4*, *CLN2*, *CLB2*, *HTB1* and *ASH1*.

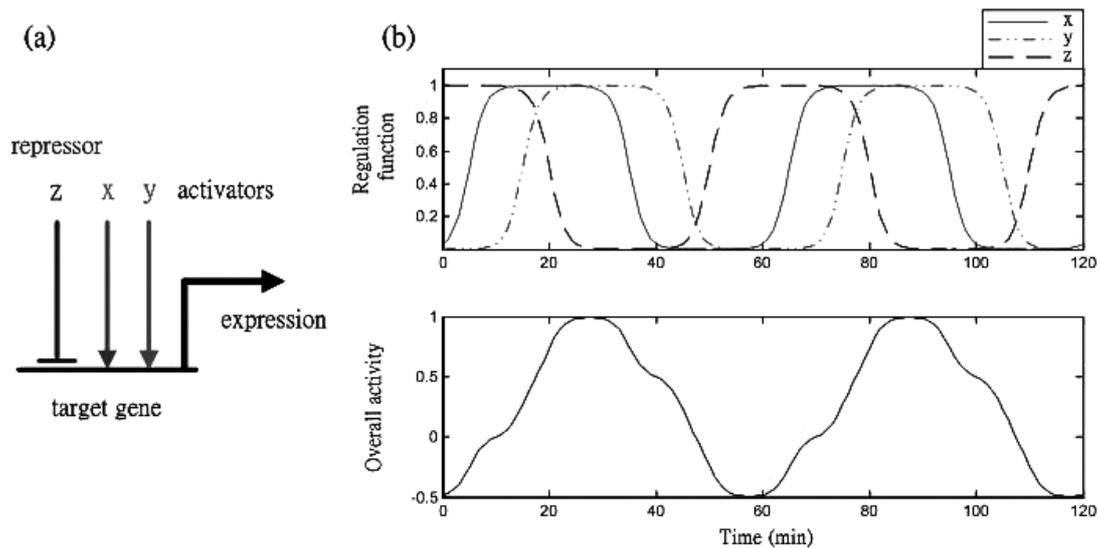


Fig. 5. A simple model of a hypothetical target gene with two activators and one repressor, and three regulators synergistically coordinating the expression program of the target gene. (a) A schematic diagram of the gene expression program: activators *x* and *y* induce the expression of the target gene and a repressor *z* reduces its expression later. (b) Regulatory function of the three transcriptional regulators (top row) and the overall transcription rate caused by the combinatorial regulation of the three different regulators (bottom row). See text for details.

periodically fluctuating mRNA levels or regulatory functions of specific regulators indeed control the expression program of their target genes in a periodic way and usually in a synergistic way. Based on this simple example and the dynamic gene expression model, we were able to examine the periodically expressed genes observed in yeast during the cell division cycle and to dissect the specific expression program temporally mediated by a small number of regulators.

Transcriptional regulatory network

Our iterative scheme for finding regulators is conveniently illustrated by the good tracking between the transcription rate $G_i(t)$ and the estimated transcription rate $\hat{G}_i(t)$ for the target gene. Our system provides the information of putative upstream regulators with their regulatory abilities, and depicts the overall transcription rate controlled by the five putative regulators as a function of time for a given target gene (Table 1 and Fig. 6). The extensive studies on the yeast cell cycle permit us to explore the features of the final regulatory network result (Wyrick and Young, 2002; Banerjee and Zhang, 2002; Futcher, 2002). For convenience in illustration, we only studied 20 target genes that are cell-cycle-regulated and are highly expressed at different phases. Five possible regulators of each specific target gene were identified and the regulatory abilities were also characterized (Table 1). For example, the important cell cycle control gene *CLN2*, encoding cyclin protein in conjunction with CDK (Cdc28), forms a complex Cln2/Cdc28 that plays a major role in budding and spindle pole body duplication. Our plots revealed that the peak of the transcription rate of *CLN2* (~15 min; Fig. 4) precedes its expression

peak at the G_1 phase (~20 min; Fig. 3). The five regulators of the target *CLN2* include the positive regulators Cln3, Swi4 and Yhp1, and the negative regulators Clb4 and Clb2. According to the identified regulatory abilities, we propose that Cln3, Swi4 and Yhp1, are the major activators for *CLN2* transcription due to the significant values of their regulatory abilities, and thus Clb4 and Clb2 are the major repressors.

Our results indicate that the expression of *CLN2* was very significantly regulated by Cln3, Swi4 and Yhp1 (0.3211, 0.2745 and 0.2020, respectively; Table 1). The highest regulatory abilities of Cln3 and Swi4 imply that the association of Cln3 and Swi4 mediates *CLN2* transcription. SBF (Swi4-Swi6 cell cycle box binding factor) is a transcription factor binding on the promoters of many genes induced at the G_1/S transition. SBF contains protein Swi6 as a subcomponent and Swi4 as the DNA-binding protein. Therefore, the periodic expression of *SWI4* and the putative activator identified by our method suggest that Swi4 plays an essential role in regulating the expression programs of genes peaked at Start. In addition to *CLN2*, Swi4 was identified as a putative activator associated with several G_1 phase genes, including *SWI4*, *SWE1*, *NDD1*, *SVS1* and *HCM1* (Table 1). All these genes contain an SBF binding site in their promoter regions (Iyer *et al.*, 2001; Simon *et al.*, 2001). In our analysis, the regulatory ability of Swi4 was lower than that of Cln3, suggesting that Swi4 conducts activation succeeding the Cln3 activator. This result coincides with the biological evidence that Swi4 is a DNA-binding subunit of the transcription factor SBF (Wijnen *et al.*, 2002) Cln3 was identified as a positive regulator of target genes *SWI4* and *CLN2* (Table 1). *CLN3* is expressed at the late M phase, and

Table 1. Five possible regulators and their regulatory abilities for 20 selected cell-cycle-related genes

Target genes	Regulators (Regulatory abilities)					Basal level C_0
<i>SWI4</i>	Cln3 (0.2469)	Swi4 ^a (0.0733)	Mbp1 ^a (0.0216)	Swe1 (-0.0839)	Spo12 (-0.1909)	0.0437
<i>CLN2</i>	Cln3 (0.3211)	Swi4 ^a (0.2745)	Yhp1 (0.2020)	Clb4 (-0.1450)	Clb2 (-0.1843)	-0.0943
<i>SWE1</i>	Cln3 (0.0855)	Swi4 ^a (0.0375)	Pcl2 (0.0222)	Spo12 (-0.0463)	Pog1 (-0.0678)	0.0463
<i>HCM1</i>	Swi4 ^a (0.0708)	Cln3 (0.0598)	Rme1 (0.0253)	Mfa1 (-0.0355)	Clb2 (-0.0899)	0.1368
<i>NDD1</i>	Skn7 (0.0525)	Far1 (0.0273)	Swi4 ^a (0.0088)	Rme1 (-0.0126)	Dbf2 (-0.0433)	0.0322
<i>SVS1</i>	Yhp1 (0.1778)	Clb6 (0.1691)	Swi4 ^a (0.0971)	Pcl7 (-0.1209)	Fkh1 (-0.1735)	0.0869
<i>MNN1</i>	Yhp1 (0.1913)	Cln3 (0.1668)	Mfa2 (0.1425)	Swi4 ^a (0.1087)	Clb2 (-0.0543)	-0.2223
<i>CLB6</i>	Cln3 (0.6335)	Yhp1 (0.2449)	Swi4 ^a (0.1851)	Fkh1 (-0.1607)	Spo12 (-0.3443)	-0.0877
<i>HTA1</i>	Clb4 (0.1065)	Swe1 (0.0264)	Swi4 ^a (0.0263)	Yhp1 (-0.1415)	Cln3 (-0.2994)	0.2815
<i>HTB1</i>	Clb4 (0.0803)	Swi4 ^a (0.0524)	Swe1 (0.0135)	Yhp1 (-0.1252)	Cln3 (-0.2218)	0.2176
<i>HHT1</i>	Swe1 (0.1080)	Clb4 (0.0413)	Swi4 (0.0144)	Yhp1 (-0.1425)	Cln3 (-0.2280)	0.2449
<i>CLB4</i>	Ndd1 (0.0368)	Cln2 (0.0168)	Clb1 (-0.0225)	Pcl2 (-0.0303)	Yhp1 (-0.0580)	0.0544
<i>CLB2</i>	Fkh2 ^a (0.1488)	Clb1 (0.1341)	Ndd1 (0.1229)	Cln1 (-0.0269)	Sic1 (-0.0552)	0.0307
<i>CDC20</i>	Spo12 (0.0864)	Ndd1 (0.0487)	Fkh2 ^a (0.0347)	Ash1 (-0.0199)	Mcm1 ^a (-0.0447)	0.0247
<i>SPO12</i>	Fkh2 ^a (0.1079)	Ste12 (0.0990)	Mfa2 (-0.0347)	Mcm1 ^a (-0.0470)	Ndd1 ^a (-0.0941)	0.0461
<i>SWI5</i>	Clb2 (0.0927)	Fkh2 ^a (0.0574)	Ash1 (-0.0232)	Yox1 (-0.0567)	Mob1 (-0.0971)	0.1031
<i>ASH1</i>	Swi5 ^a (0.1207)	Ash1 (0.1135)	Cln2 (-0.1113)	Yhp1 (-0.1337)	Rme1 (-0.1909)	0.1534
<i>SIC1</i>	Pcl9 (0.1193)	Swi5 ^a (0.0261)	Rme1 (-0.0422)	Cln2 (-0.0581)	Yhp1 (-0.0954)	0.1277
<i>CLN3</i>	Sim1 (0.1149)	Swi5 (0.0220)	Mcm1 ^a (0.0177)	Swe1 (-0.0470)	Yhp1 (-0.0624)	0.0075
<i>CDC46</i>	Cln3 (0.1114)	Clb2 (0.0141)	Rme1 (-0.0213)	Clb6 (-0.0269)	Opy2 (-0.0441)	0.0348

The first column indicates the specific target genes, the second to sixth columns outline the identified regulators which were sorted in descending order according to the value of their associated regulation abilities and the last column is the basal level of transcription rate.

^aTranscription factors that were previously reported to bind to the promoters of target genes (Lee et al., 2002), P -value < 0.001.

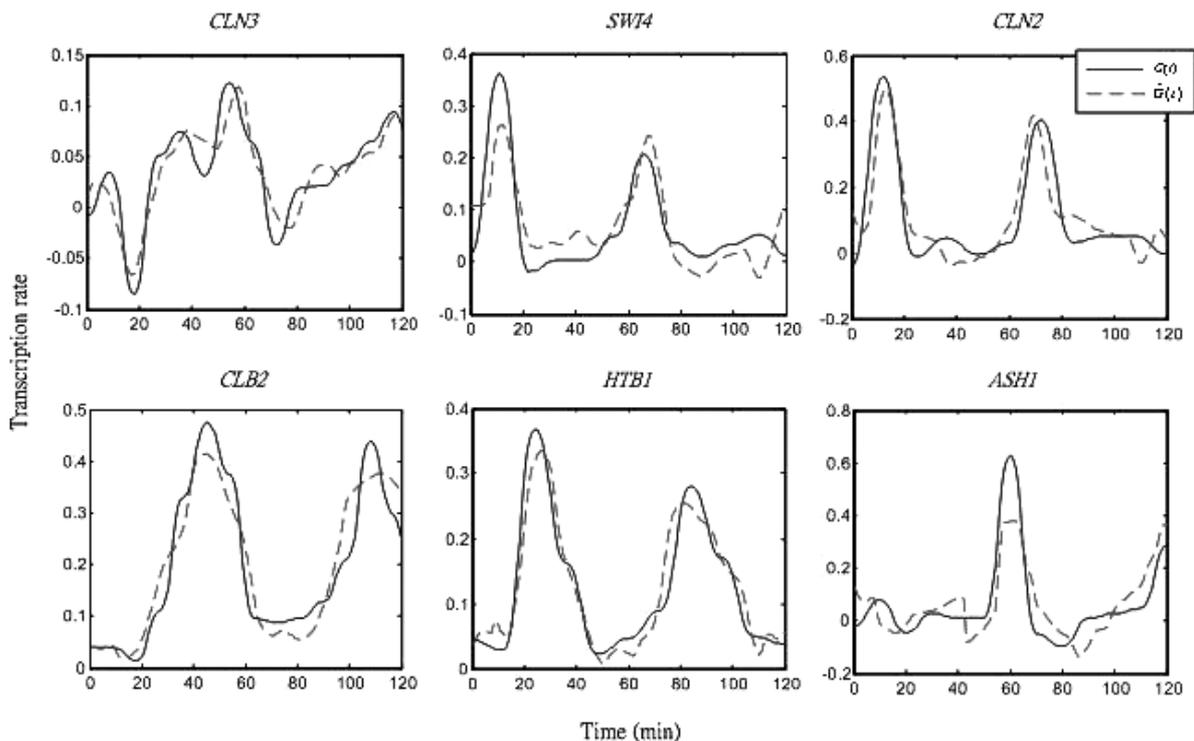


Fig. 6. The comparison between identified transcription rates $G_i(t)$ s and the estimated transcription rates generated by five possible regulators $\hat{G}_i(t)$ s (broken lines) for the six cell-cycle-related genes *CLN3*, *SWI4*, *CLN2*, *CLB2*, *HTB1* and *ASH1*.

Table 2. Five possible regulators and their influenced target genes

Regulator	Target genes	
	Upregulation	Downregulation
Cln3	<i>SWI4, CLN2, SWE1, HCM1, MNN1, CLB6, CDC46</i>	<i>HTA1, HTB1, HHT1</i>
Swe1	<i>HTA1, HTB1, HHT1</i>	<i>SWI4, CLN3</i>
Spo12	<i>CDC20</i>	<i>SWI4, SWE1, CLB6</i>
Yhp1	<i>CLN2, SVS1, MNN1, CLB6</i>	<i>HTA1, HTB1, HHT1, CLB4, SIC1, ASH1, CLN3</i>
Clb2	<i>SWI5, CDC46</i>	<i>CLN2, HCM, MNN1</i>
Rme1	<i>HCM1</i>	<i>NDD1, ASH1, SIC1, CDC46</i>
Swi4	<i>SWI4, CLN2, SWE1, HCM1, NDD1, SVS1, MNN1, CLB6, HTA1, HTB1, HHT1</i>	
Swi5	<i>ASH1, SIC1, CLN3</i>	
Ndd1	<i>CLB4, CLB2, CDC20</i>	<i>SPO12</i>
Fkh2	<i>CLB2, CDC20, SPO12, SWI5</i>	

The upstream regulators and their relevant target genes. The first column indicates the specific regulators; the second and third columns outline the upregulating target genes and downregulating target genes, respectively

Cln3 might consequently activate the expression of *SWI4* and turn on the activity of SBF. Similar to the immense amount of evidence showing that Cln3 is the most prominent activator of SBF (Tyers *et al.*, 1993; Stuart and Wittenberg, 1995; McInerney *et al.*, 1997), our result displayed that Cln3 significantly regulated the expression of *SWI4* and several G1 phase genes, *CLN2*, *SWE1* and *HCM1* (Table 2).

We selected a set of target genes with peak expression levels at various stages of the cell cycle (Table 1) and diagrammatically depict the regulatory abilities of the regulators on their target genes to illustrate the transcriptional regulatory network of the yeast cell cycle (Fig. 7). It is well known that Clb2 is primarily responsible for inactivating SBF and turning off *CLN2* synthesis as cells enter mitosis. Our analysis also identified that Clb2 is a negative regulator of *CLN2* (-0.1843 ; Table 1), the negative correlation between the transcription rate of *CLN2* and the expression profile of regulator Clb2 explains the relationship of this target–regulator pair. This network reveals that Spo12 had a significantly negative regulatory ability to coordinate the expression of several G1 genes, *SWI4* and *SWE1*. The molecular function of Spo12 is unclear. This protein plays the role of a positive regulator of exit from the M phase in mitosis and meiosis (Stegmeier *et al.*, 2002; Marston *et al.*, 2003; Buonomo *et al.*, 2003). *SPO12* transcripts are expressed in a cell-cycle-regulated manner and peaked at the M phase (Spellman *et al.*, 1998). The strong expression of *SPO12* at this period might cause the downregulation of its target genes through direct or indirect molecular interactions in a temporal manner. Swe1 is a protein having the molecular function of regulating CDK activity and is expressed abruptly during the G₁ phase (Booher *et al.*, 1993). Swe1 represses the transcription of *CLN3* (6.2-fold of the basal

level) and *SWI4* (1.8-fold of the basal level) while the expression of *SWE1* is activated by Cln3 and Swi4 to form a feedback loop chain (Fig. 7 and Table 1). This feedback loop chain might play a sophisticated role in regulating the expression of *CLN3* and *SWI4* and then mediating hundreds of genes whose expression peak at the G₁ phase through binding to SBF subsequently (Iyer *et al.*, 2001; Simon *et al.*, 2001).

Clbs are cyclins crucial for completing the G₂/M transition of mitotic cell cycle and the most typical one is Clb2. Mcm1, Fkh1/2 and Ndd1 are identified as transcription factors binding to the promoter sequence of *CLB2* (Zhu *et al.*, 2000; Iyer *et al.*, 2001; Simon *et al.*, 2001; Koranda *et al.*, 2000). In our analysis, the transcription factors Fkh2 and Ndd1 are recognized as activators with significant regulatory abilities (4.8-fold and 4.1-fold of the basal level, respectively), while the cyclin Clb1 that might activate the transcription by itself is also recognized as another activator (Table 1). The significant regulatory abilities suggest that the regulators might play the primary role in controlling the expression of *CLB2*, concurring with the experimental evidence that Fkh2 and Ndd1 were recruited to the chromatin of the *CLB2* promoter (Koranda *et al.*, 2000). In addition to the activators, a CDK inhibitor Sic1 that binds and inhibits the Clb2–Cdc28 complex was found to downregulate *CLB2* (Schwob *et al.*, 1994). Sic1 did not show a high level of regulation (Table 1) and may act through indirect interaction. Cdc20 is a subcomponent of the anaphase promoting complex (APC), which is required for the onset of anaphase (Zachariae and Nasmyth, 1999). Similar to *CLB2*, the expression of *CDC20* showed a cell-cycle-regulated manner and peaked at the M phase. The two transcription factors Ndd1 and Fkh2 were also identified as positive regulators of *CDC20*. In addition, Spo12 showed a significant regulatory ability, suggesting an activation role of Spo12 in concert with Ndd1 to mediate the expression of *CDC20* (Table 1). As mentioned earlier, Spo12 is thought to be a positive regulator of mitotic exit (Stegmeier *et al.*, 2002; Marston *et al.*, 2003; Buonomo *et al.*, 2003). Spo12 may cooperate with Ndd1 and Fkh2 to turn on the transcription of *CDC20* for subsequently promoting the mitotic metaphase/anaphase transition.

Another crucial regulator for cell cycle control is Sic1. According to Verma *et al.* (1997), when Cln2-associated kinase activity arises at Start, Sic1 is phosphorylated and Sic1p is rapidly presented by Cdc4 to the SCF (SCF is a complex of Skp1, Cdc34, Cdc53 and an F box-containing protein) for ubiquitination and subsequent proteolysis. The expression of *SIC1* reached a peak at the M/G1 phase; two putative activators Pcl9 and Swi5 were found to regulate its transcription (Table 1). Pcl9 is a cyclin-dependent protein kinase whose expression peaked at the M/G1 phase (Wang *et al.*, 2001) and Swi5 is a transcription factor whose expression fluctuated in a cell-cycle-regulated manner and peaked at the M phase (Spellman *et al.*, 1998). Cln2 and Yhp1 were identified as negative regulators for *SIC1* (Table 1). Yhp1 is a negative regulator for transcription and its expression peaked at the

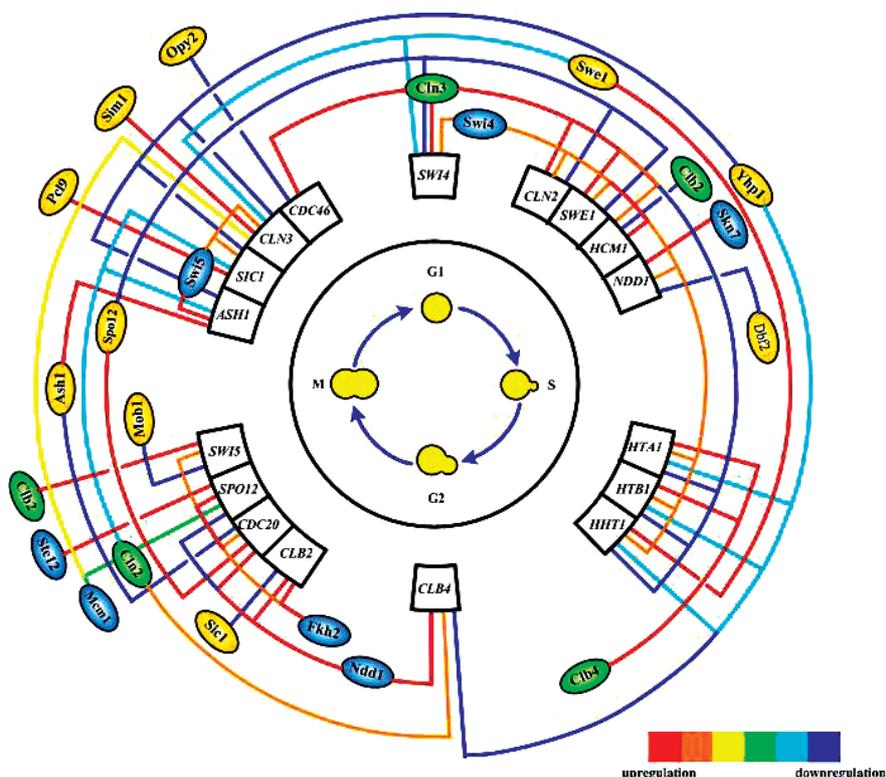


Fig. 7. A transcriptional regulatory network of 17 selected target genes involved in the yeast cell cycle. Each white box represents a target gene and the target gene is allocated in the cell cycle according to the time of its expression peak. Regulators are indicated by ovals of different colors. Blue ovals denote transcription factors, green ovals denote cyclins and orange ovals denote cell cycle control proteins. The target gene and regulators are connected by solid lines. The regulatory abilities of the regulators for a specific target are characterized by colored lines: red for upregulation and blue for downregulation as indicated.

S phase (Kunoh *et al.*, 2000; Horak *et al.*, 2002). Our results agreed with previous reports that *SIC1* transcription is upregulated by Swi5 and downregulated by Cln2 (Simon *et al.*, 2001; Verma *et al.*, 1997). The basal level of *SIC1* exceeds the regulatory ability of its regulators (Table 1), suggesting the constitutive expression of *SIC1*. Mutation of *Fkh2* decreased Swi5 transcripts and consequently reduced the *SIC1* expression (Zhu *et al.*, 2000). This scenario can be identified by our network. Swi5 bound to the promoter of a cyclin gene *CLN3*, which plays a critical role for the cell cycle control (Simon *et al.*, 2001). Moreover, Sim1 is identified as the primary activator together with Swi5 and Mcm1 for upregulating the transcription of *CLN3* (Table 1). A possible molecular function of Sim1 is the involvement in control of DNA replication (Mouassite *et al.*, 2000). The expression of *SIM1* peaked at the G₂ phase might cooperate with transcription factors Swi5 and Mcm1 for the onset of *CLN3* transcription and subsequently mediate the cell cycle process.

The three target genes *HTA1*, *HTB1* and *HHT1* (Table 1) are classified as the ‘histone’ cluster whose expressions are highly conserved and peaked at the S phase (Spellman *et al.*, 1998). Histones play an important role in chromatin assembly

and disassembly, and thus are strongly cell cycle-regulated and peaked at the S phase. The three histone genes shared identical putative regulators and had relatively high levels of basal expression (Table 1). The negative regulators Cln3 and Yhp1 had regulatory abilities beyond the basal level and might directly or indirectly inhibit the function of activators at the basal level; thus, histone genes are expressed in a cell-cycle-regulated manner. In addition, the regulatory abilities of the positive regulators Clb4, Swe1 and Swi4 were below the basal level. These positive regulators may help the unknown ubiquitous activators when the expression levels of *CLN3* and *YHP1* are comparatively low. The fact that Swi4 was identified as a positive regulator in our system explains the reason why, through genome-wide binding location analysis, certain histone gene promoters were found to have SBF binding sites (Simon *et al.*, 2001).

Table 1 lists 20 target genes that are expressed strongly at different phases throughout the cell cycle. Other target genes possessing similar expression patterns with these 20 genes might have the same regulators. In order to specify the target genes mediated by specific regulators that were identified in our analysis, we organized Table 2 to list the regulators and

their target genes. The newly found pathways in Table 2 might be useful for dissecting the regulatory mechanism in the yeast cell cycle and could be examined in the future.

DISCUSSION

As a vast accumulation of gene expression microarray data becomes available, there is a growing need for computational approaches to study genome-wide transcriptional regulation. Although the potential of the information contained in the large and diverse databases of genome-wide expression profiles is amply appreciated, the radicalization of biological transcriptional regulation from such data remains a challenging task. This study presents a systematically iterative approach to discern and characterize the transcriptional regulatory network in yeast from the raw expression profiles focused on the cell cycle (Spellman *et al.*, 1998). Our method provides a useful tool for analyzing gene expression data and validates a few regulators with periodic expression that are enough to coordinate the cell cycle program (Fig. 7). Till date, the most widely used method for organizing and visualizing expression data is clustering, which determines a number of gene groups, each holding a similar, particular pattern across the whole data set. There is no doubt that the adequate and accurate clustering of expression profiles is the first step to dissecting the transcription network. However, our study provides another aspect of view for inferring the intertwined relationship between different genes that organize the genome-wide expression program.

Our approach also offers the following advantages. First, we adopt a limited number of sample points and transform the raw data into a mathematical kinetic model that describes the dynamics of the expression profile of any gene of interest and provides valuable information for further analysis. Second, the well-constructed dynamic model allows us to extract the transcription rate that controls the expression of any gene of interest. This calculated term can be used as a touchstone for deducing the transcriptional regulatory network. Third, we characterize the relationship between gene expression pattern and its regulatory function as a sigmoid function and quantifying the transcription rate of each specific target gene that is composed of the combinations of particular regulatory functions contributed by upstream regulators. This is the first time we have used dynamic model to describe transcriptional regulatory networks based on microarray data. Our model explicitly describes a superposition of a smaller number of regulatory effects, including the genes responding to different combinations of common input variables and the regulatory functions that captures the biologically relevant combinatorial regulation. Last, the identified regulatory ability for each specific regulator evaluates the contribution of this regulator; the positive sign stands for activation and the negative sign stands for repression, and the magnitude represents the activity. These advantages of our approach will ameliorate analysis

to cope with rapidly growing data and can be extended to systematic analysis for higher eukaryotes.

A few studies have focused on the gene regulatory network recently using different methods. For example, a fuzzy logic (Woolf and Wang, 2000) approach was proposed to generate a connected network of genes using gene expression data; the fuzzy logic algorithm provides a way to transform expression values into qualitative descriptions that can be evaluated by using heuristic rules, and constructs a model to find triplets of activators, repressors and targets (Woolf and Wang, 2000). A smooth surface response (SRS) algorithm is developed to improve and extend the fuzzy logic algorithm for identifying the regulations of activators, repressors and targets from gene expression data, and constructing a gene regulatory network (Xu *et al.*, 2002). All of the above studies have done a great job of uncovering transcriptional regulatory network from expression data. Compared with these approaches, our model not only qualitatively reveals the regulatory relationship between regulators and target genes but also quantitatively reveals the capabilities of regulators to the specific target genes. Moreover, our regulatory network dissects the synergistic nature of multi-regulation caused by a set of regulators against a simple triplet model of activator, repressor and target. Our analysis provides an independent method for uncovering a transcriptional regulatory network using a mathematical dynamic model and quantitatively characterizing the regulatory ability between regulators and target genes from mRNA transcript data.

The weakness of our model is that the transcriptional control network is over-simplified. In reality, the cell cycle is controlled by a much more complex network that includes not only transcription, but also proteolysis, phosphorylation, compartmentalization and other regulatory machinery. To uncover the real mechanism for regulating the cell cycle, more deliberate experiments and sophisticated computational approaches need to be developed and examined. The limitation of our approach is that all the analyzed information only came from the expression data for cell-cycle-regulated genes of yeast. For example, several crucial proteins promoting the events of cell cycle, such as Cdc28, Swi6, Cdh1 and Cdc14 (Chen *et al.*, 2000) were not analyzed and characterized in our work, because the expression levels of these proteins are relatively consistent throughout the cell cycle (Spellman *et al.*, 1998). It is simplistic to infer the real transcriptional regulatory network from data of mRNA transcripts alone, because the real machinery involves more complicated biological processes, for instance, the post-transcriptional regulation and the protein-protein interactions. Apart from this, the real regulation for transcription might not be a simple combination as we proposed. To our knowledge, our study is the first one done with the application of quantitative dynamic modeling on the assay of the transcription network. The dynamic analysis unraveled gene transcription as well as the contribution of their activators and suppressors. As more detailed

worthwhile experimental results like protein microarray data become available and as a more exhaustive knowledge of transcriptional mechanism is gained, a more accurate model can be developed to augment our approach for deciphering the real transcriptional regulatory network. It is believed that such a systematic approach will be critical for dissecting the complex architecture of transcriptional networks in more complex eukaryotes, in anticipation of an avalanche of microarray data from the human and mouse genome.

ACKNOWLEDGEMENTS

We thank the National Science Council, Taiwan, R.O.C., for grants NSC 91-3112-B-007-002.

REFERENCES

- Akaike, H. (1974) A new look at the statistical model identification. *IEEE Trans. Automat. Control*, **AC-19**, 716–723.
- Alter, O., Brown, P.O. and Botstein, D. (2000) Singular value decomposition for genome-wide expression data processing and modeling. *Proc. Natl Acad. Sci., USA*, **97**, 10101–10106.
- Banerjee, N. and Zhang, M.Q. (2002) Functional genomics as applied to mapping transcription regulatory networks. *Curr. Opin. Microbiol.*, **5**, 313–317.
- Booher, R.N., Deshaies, R.J. and Kirschner, M.W. (1993) Properties of *Saccharomyces cerevisiae* *wee1* and its differential regulation of p34^{CDC28} in response to G1 and G2 cyclins. *EMBO J.*, **12**, 3417–3426.
- Buonomo, S.B., Rabitsch, K.P., Fuchs, J., Gruber, S., Sullivan, M., Uhlmann, F., Petronczki, M., Toth, A. and Nasmyth, K. (2003) Division of the nucleolus and its release of CDC14 during anaphase of meiosis I depends on separase, SPO12, and SLK19. *Dev. Cell*, **4**, 711–726.
- Bussemaker, H.J., Li, H. and Siggia, E.D. (2001) Regulatory element detection using correlation with expression. *Nat. Genet.*, **27**, 167–171.
- Chen, K.C., Csikasz-Nagy, A., Gyorfy, B., Val, J., Novak, B. and Tyson, J.J. (2000) Kinetic analysis of a molecular model of the budding yeast cell cycle. *Mol. Biol. Cell*, **11**, 369–391.
- Cho, R.J., Campbell, M.J., Winzler, E.A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T.G., Gabrielian, A.E., Landsman, D., Lockhart, D.J. and Davis, R.W. (1998) A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell*, **2**, 65–73.
- Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci., USA*, **95**, 14863–14868.
- Fickett, J.W. and Wasserman, W.W. (2000) Discovery and modeling of transcriptional regulatory regions. *Curr. Opin. Biotechnol.*, **11**, 19–24.
- Futcher, B. (2002) Transcriptional regulatory networks and the yeast cell cycle. *Curr. Opin. Cell Biol.*, **14**, 676–683.
- Garvie, C.W. and Wolberger, C. (2001) Recognition of specific DNA sequences. *Mol. Cell*, **8**, 937–946.
- Goldbeter, A. and Koshland, D.E., Jr (1981) An amplified sensitivity arising from covalent modification in biological systems. *Proc. Natl Acad. Sci., USA*, **78**, 6840–6844.
- GuhaThakurta, D. and Stormo, G.D. (2001) Identifying target sites for cooperatively binding factors. *Bioinformatics*, **17**, 608–621.
- Halfon, M.S., Carmena, A., Gisselbrecht, S., Sackerson, C.M., Jimenez, F., Baylies, M.K. and Michelson, A.M. (2000) Ras pathway specificity is determined by the integration of multiple signal-activated and tissue-restricted transcription factors. *Cell*, **103**, 63–74.
- Holter, N.S., Maritan, A., Cieplak, M., Fedoroff, N.V. and Banavar, J.R. (2001) Dynamic modeling of gene expression data. *Proc. Natl Acad. Sci., USA*, **98**, 1693–1698.
- Horak, C.E., Luscombe, N.M., Qian, J., Bertone, P., Piccirillo, S., Gerstein, M. and Snyder, M. (2002) Complex transcriptional circuitry at the G1/S transition in *Saccharomyces cerevisiae*. *Genes Dev.*, **16**, 3017–3033.
- Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M. and Brown, P.O. (2001) Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature*, **409**, 533–538.
- Koranda, M., Schleiffer, A., Endler, L. and Ammerer, G. (2000) Forkhead-like transcription factors recruit Ndd1 to the chromatin of G2/M-specific promoters. *Nature*, **406**, 94–98.
- Kunoh, T., Kaneko, Y. and Harashima, S. (2000) YHP1 encodes a new homeoprotein that binds to the IME1 promoter in *Saccharomyces cerevisiae*. *Yeast*, **16**, 439–449.
- 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 *Saccharomyces cerevisiae*. *Science*, **298**, 799–804.
- Marston, A.L., Lee, B.H. and Amon, A. (2003) The Cdc14 phosphatase and the FEAR network control meiotic spindle disassembly and chromosome segregation. *Dev. Cell*, **4**, 727–739.
- McInerney, C.J., Partridge, J.F., Mikesell, G.E., Creemer, D.P. and Breeden, L.L. (1997) A novel Mcm1-dependent element in the *SWI4*, *CLN3*, *CDC6*, and *CDC47* promoters activates M/G1-specific transcription. *Genes Dev.*, **11**, 1277–1288.
- Mestl, T., Plahte, E. and Omholt, S.W. (1995) A mathematical framework for describing and analysing gene regulatory networks. *J. Theor. Biol.*, **176**, 291–300.
- Mouassite, M., Guerin, M.G. and Camougrand, N.M. (2000) The SUN family of *Saccharomyces cerevisiae*: the double knock-out of UTH1 and SIM1 promotes defects in nucleus migration and increased drug sensitivity. *FEMS Microbiol. Lett.*, **182**, 137–141.
- Orphanides, G., and Reinberg, D. (2002) A unified theory of gene expression. *Cell*, **108**, 439–451.
- Pilpel, Y., Sudarsanam, P. and Church, G.M. (2001) Identifying regulatory networks by combinatorial analysis of promoter elements. *Nat. Genet.*, **29**, 153–159.
- Proakis, J.G. (2001) *Digital Communications*, 4th edn. McGraw-Hill Publishing, NY.
- Šašik, R., Iranfar, N., Hwa, T. and Loomis, W.F. (2002) Extracting transcriptional events from temporal gene expression patterns during *Dictyostelium* development. *Bioinformatics*, **18**, 61–66.
- Schwob, E., Bohm, T., Mendenhall, M.D. and Nasmyth, K. (1994) The B-type cyclin kinase inhibitor p40^{SIC1} controls the G1 to S transition in *S. cerevisiae*. *Cell*, **79**, 181–184.
- Simon, I., Barnett, J., Hannett, N., Harbison, C.T., Rinaldi, N.J., Volkert, T.L., Wyrick, J.J., Zeitlinger, J., Gifford, D.K., Jaakkola, T.S. and Young, R.A. (2001) Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell*, **106**, 697–708.

Singer, R.H. and Penman, S. (1973) Messenger RNA in HeLa cells: kinetics of formation and decay. *J. Mol. Biol.*, **78**, 321–334.

Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.Q., Bostein, D. and Futcher, B. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell*, **9**, 3273–3297.

Stegmeier, F., Visintin, R. and Amon, A. (2002) Separase, polo kinase, the kinetochore protein Slk19, and Spo12 function in a network that controls Cdc14 localization during early anaphase. *Cell*, **108**, 207–220.

Stuart, D. and Wittenberg, C. (1995) *CLN3*, not positive feedback, determines the timing of *CLN2* transcription in cycling cells. *Genes Dev.*, **9**, 2780–2794.

Tamayo, P., Slonim, D., Mesirov, J., Xhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E.S. and Colub, T.R. (1999) Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc. Natl Acad. Sci., USA*, **96**, 2907–2912.

Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J., Spouge, J.L. and Church, G.M. (1999) Systematic determination of genetic network architecture. *Nat. Genet.*, **22**, 281–285.

Tyers, M., Tokiwa, G. and Futcher, B. (1993) Comparison of the *Saccharomyces cerevisiae* G1 cyclins: Cln3 may be an upstream activator of Cln1, Cln2 and other cyclins. *EMBO J.*, **12**, 1955–1968.

Verma, R., Annan, R.S., Huddleston, M.J., Carr, S.A., Reynard, G. and Deshaies, R.J. (1997) Phosphorylation of Sic1p by G1 Cdk required for its degradation and entry into S phase. *Science*, **278**, 455–460.

Wang, Z., Wilson, W.A., Fujino, M.A. and Roach, P.J. (2001) The yeast cyclins Pc16p and Pc17p are involved in the control of glycogen storage by the cyclin-dependent protein kinase Pho85p. *FEBS Lett.*, **506**, 277–280.

Wang, Y., Liu, C.L., Storey, J.D., Tibshirani, R.J., Herschlag, D. and Brown, P.O. (2002) Precision and functional specificity in mRNA decay. *Proc. Natl Acad. Sci., USA*, **99**, 5860–5865.

Wijnen, H., Landman, A. and Futcher, B. (2002) The G₁ cyclin Cln3 promotes cell cycle entry via the transcription factor Swi6. *Mol. Cell Biol.*, **22**, 4402–4418.

Woolf, P.J. and Wang, Y. (2000) A fuzzy logic approach to analyzing gene expression data. *Physiol. Genomics*, **3**, 9–15.

Wyrick, J.J. and Young, R.A. (2002) Deciphering gene expression regulatory networks. *Curr. Opin. Genet. Dev.*, **12**, 130–136.

Xu, H., Wu, P., Wu, C.F., Tidwell, C. and Wang, Y. (2002) A smooth response surface algorithm for constructing a gene regulatory network. *Physiol. Genomics*, **11**, 11–20.

Yeung, M.K., Tegner, J. and Collins J.J. (2002) Reverse engineering gene networks using singular value decomposition and robust regression. *Proc. Natl Acad. Sci., USA*, **99**, 6163–6168.

Zachariae, W. and Nasmyth, K. (1999) Whose end is destruction: cell division and the anaphase-promoting complex. *Genes Dev.*, **13**, 2039–2058.

Zhu, G., Spellman, P.T., Volpe, T., Brown, P.O., Botstein, D., Davis, T.N. and Futcher B. (2000) Two yeast forkhead genes regulate the cell cycle and pseudohyphal growth. *Nature*, **406**, 90–94.

APPENDIX: MATCHED FILTER

To fully specify the correlation or matching between the transcription rate $G_i(t)$ and regulation function $f_j(t)$, we introduce a powerful technique, matched filter, which was widely used for detecting signals in the communication field and for the correlation classifier in the pattern recognition field. In general, let us suppose that the impulse response of the filter is

$$h_k(t) = s_k(T - t), \quad 0 \leq t \leq T,$$

where $s_k(t)$ is the basis function and $h_k(t) = 0$ outside of the interval $0 \leq t \leq T$. If we let the input signal to the filter be $r(t)$, then the output of this filter is (Proakis, 2001)

$$y_k(t) = \begin{cases} \int_0^t r(\tau)h_k(t - \tau)d\tau \\ \int_0^t r(\tau)s_k(T - t + \tau)d\tau \quad k = 1, 2, \dots, N. \end{cases}$$

Now, if we sample the output of the filter at $t = T$, we obtain

$$y_k(T) = \int_0^T r(\tau)S_k(\tau)d\tau, \quad k = 1, 2, \dots, N$$

Hence, the sampled output of the filter at time $t = T$ is exactly the correlation between $r(t)$ and $s_k(t)$. A filter whose impulse response $h(t) = s(T - t)$, where $s(t)$ is assumed to be confined to the time interval $0 \leq t \leq T$, is called the matched filter to the signal $s(t)$. Based on this concept, we let the input signal as $G_i(t)$ and the impulse response of the filter as $f_k(t)$, where $G_i(t)$ and $f_k(t)$ are transcription rate and regulation function, respectively. Therefore, the sampled output of the filter can be represented as the score that stands for the correlation between $G_i(t)$ and $f_k(t)$.