

# Nonlinear Dynamic Trans/Cis Regulatory Circuit for Gene Transcription via Microarray Data

Yu-Hsiang Chang, Yu-Chao Wang and Bor-Sen Chen

Lab of Control and Systems Biology, Department of Electrical Engineering, National Tsing Hua University, Hsinchu, 300, Taiwan.

**Abstract:** The trans-regulatory circuit is considered as the regulatory interactions between upstream regulatory genes and transcription factor binding site motifs or cis elements. And the cis-regulatory circuit is viewed as a dynamic interactive circuit among binding site motifs with their effective action on the expression scheme of target gene. In brief, gene transcription depends on the trans/cis regulatory circuits. In this study, nonlinear trans/cis regulatory circuits for gene transcription in yeast are constructed using microarray data, translation time delay, and information of transcription factors (TFs) binding sites. We provide a useful nonlinear dynamic modeling and develop a parameter estimating method for the construction of trans/cis regulatory circuits, which is powerful for understanding gene transcription. We apply our method to construct trans/cis regulatory circuits of yeast cell cycle-related genes and successfully quantify their regulatory abilities and find possible cis-element interactions. Not only could the data of yeast be applied by our method, but those of other species also could. The proposed method can provide a quantitative basis for system analysis of gene circuits, which is potential for gene regulatory circuit design with a desired gene expression.

**Keywords:** transcription factor, nonlinear dynamic model, trans/cis regulatory circuit, cell cycle.

## Introduction

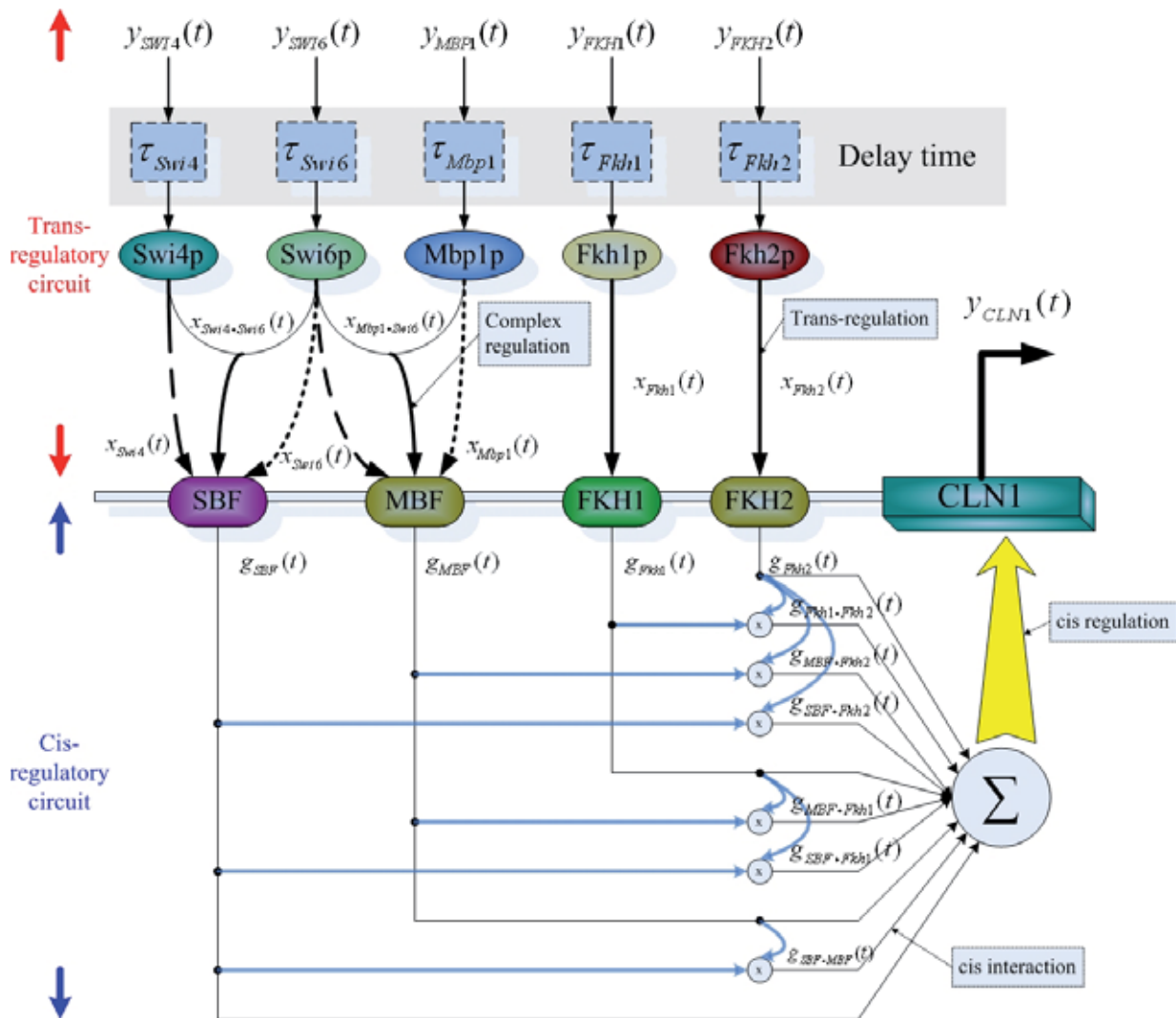
Due to advances in DNA microarray technology and genome sequencing, it has become possible to measure gene expression levels on a genomic scale. Microarray technology provides insight into the transcriptional state of a cell. In recent studies, the expression profiles and motif binding sites of genes in yeast have been revealed, and regulatory networks have been proposed to explain their regulatory functions (Spellman et al. 1998; Iyer et al. 2001; Simon et al. 2001; Hartemink et al. 2002; Lee et al. 2002; Harbison et al. 2004). However, in order to gain more insight into the infrastructure of the regulatory scheme inside the gene transcription, trans/cis regulatory circuit constructed by the view of the systems biology is an important topic.

Genes are always regulated by a number of upstream regulatory genes through their products, for instance, transcription factors bind to specific sites (i.e. cis elements) in the DNA promoter region. The interactions between upstream regulatory genes and cis elements are described by a trans-regulatory circuit (see Fig. 1). The DNA binding motifs, served as anchors to transcription factors, play the role of staying platforms for the assembly of multi-proteins at the output. This mechanism specifies the cis-regulatory circuit, which is viewed as a dynamic interactive circuit among binding site motifs with their effective action on the expression scheme of target gene (Fig. 1). In brief, gene transcription depends on the trans/cis regulatory circuits. Therefore, the comprehension of gene transcriptions needs to recognize the corresponding trans/cis regulatory circuits. Unfortunately, although the cis-regulatory circuit has been widely discussed in *Drosophila* (Berman et al. 2002) and sea urchin (Yuh and Davidson, 1996; Yuh et al. 1996; Yuh et al. 1998), not much is known about trans/cis-regulatory circuits because of their complicated interactive schemes, which are not easily detected directly by conventional experiments. That is why we want to propose a method to construct trans/cis regulatory circuits to understand gene transcription process thoroughly.

Given a processed data set, one expects to be ready to tackle the biological data interpretation problem. It is popular to use clustering, classification, and projection methods to analyze the data set. However, most analysis methods only use microarray data, not both dynamic expression time profiles of microarray data and the information of motif binding sites (Gardner et al. 2003; Chen et al. 2004;

**Correspondence:** Bor-Sen Chen, Department of Electrical Engineering, National Tsing Hua University, Hsinchu, 300, Taiwan. Tel: +886-3-5731155; Email: bschen@ee.nthu.edu.tw

Please note that this article may not be used for commercial purposes. For further information please refer to the copyright statement at <http://www.la-press.com/copyright.htm>



**Figure 1.** The dynamic trans/cis regulatory model of the gene transcription of target gene CLN1.  $x_{ij}(t)$  denotes trans regulatory function of the complex of the transcription factors  $i$  and  $j$  and  $g_{ij}(t)$  denotes the interaction between the cis elements  $i$  and  $j$ .

Chen et al. 2005). Furthermore, the interactions among proteins and delay from transcription to translation are not considered in their models of gene transcription. Recently, systems biology and computational biology methods have been widely considered to describe the biological functions from the dynamic system perspective (Yeung et al. 2002; Davidson et al. 2003; Tegner et al. 2003; Chen et al. 2004; Sontag et al. 2004; Chang et al. 2006). Engineering theory has also been used to know more about biological complexity (Carlson and Doyle, 1999; Yi et al. 2000; Csete and Doyle, 2002; Hasty et al. 2002; Gardner et al. 2003; Tegner et al. 2003). Here, we provide a nonlinear dynamic model to gain more insight into trans/cis regulatory circuits for gene transcription using the knowledge of microarray data, the information

of motif binding sites, translation delays and protein complexes from the systems biology perspective.

In this study, a nonlinear dynamic model is proposed to describe the kinetics of trans/cis regulatory circuits for gene transcriptions of yeast genes. In the trans circuit part (see Fig. 1), the transcription regulatory functions on cis elements are described by the regulatory activities of transcription factors and reaction of complexes among transcription factors. The binding from upstream regulatory genes to the transcription factor binding sites is described by a biological sigmoid function to model the binding (activating the binding function) and no binding (inactivating the binding function) signals beyond or below some threshold of mRNA concentrations of

regulatory genes. The translation time delay from the mRNA of a regulatory gene to transcription factor is also considered in our model. The reactions (cooperations) of complexes of transcription factors are represented by nonlinear interactions (Chang et al. 2006). In this study, nonlinear interactions between two proteins and among three proteins are also considered to mimic the effect of protein complexes on a gene transcription regulation process. In the cis-regulatory circuit part, the interactions of cis elements are also modeled by a nonlinear dynamic equation. The regulatory functions of double and triple nonlinear interactions among the cis elements are considered in this dynamic model. Furthermore, the decay rate is also considered to describe how the output mRNA is degraded in the target gene.

The nonlinear dynamic model is useful to describe how the upstream regulatory genes control a target gene to produce the output expression of mRNA through its trans/cis regulatory circuit. Using the information on transcription factors (Simon et al. 2001) and the experimental profiles on upstream regulatory genes and target gene (Spellman et al. 1998), the model of trans/cis regulatory circuit of target gene can be transformed to an algebraic regression equation for parameter estimation of the dynamic transcription model. With this study, the stochastic noises of microarray data are also considered to describe the uncertainties of the data. This is capable of improving the accuracy of parameter estimations in the dynamic model. The kinetic coefficients of the stochastic dynamic model, the decay rate of target gene's mRNA and the variance of noises are estimated via the maximum likelihood estimation method and the downhill simplex search method. After estimating these parameters via expression profiles of the target gene and its regulatory genes, they will be substituted into the nonlinear dynamic model to confirm the validity of trans/cis regulatory circuits. For illustration, the trans/cis regulatory circuits of the gene CLN1 are constructed in detail by the proposed method from expression profiles in the microarray data (Spellman et al. 1998) and knowledge of binding site motifs of yeast cell cycle (Simon et al. 2001). After constructing the model, we are able to use the mRNA expression profiles of regulatory genes to predict the expression profiles of the target gene and analyze the characteristics of the nonlinear trans/cis regulatory circuits to gain more insight into gene transcription,

which also provides a potential method for gene regulatory circuit design with a desired gene expression.

## Modeling and Identification

First of all, we construct a dynamic trans/cis regulatory circuit with two parts. One simulates the binding effects of regulatory proteins and their complexes, and the other considers interactions of cis elements in the target gene. After that, we apply the downhill simplex search method and the maximum likelihood method to estimate the dynamic parameters of the trans/cis regulatory circuit. Finally, we show the experimental data that have been used and how to preprocess it before system identification of regulatory circuits. The proposed method will provide a way for system analysis of gene regulatory circuits and a method for gene circuit design with a desired gene expression.

## Dynamic model of trans/cis regulatory circuits

First, we consider a trans/cis regulatory network as a system block with several regulatory genes as inputs and a target gene as output. In our model, we use the mRNA expression data of upstream regulatory genes corresponding to transcription factors as the system input and the mRNA expression data of the target gene as the system output. For the illustration of nonlinear dynamic trans/cis regulatory circuit, an example of the target gene CLN1 is shown in Figure 1. From the binding site motif data (Simon et al. 2001), we know that the gene CLN1 has five transcription factors (Swi4, Swi6, Mbp1, Fkh1, and Fkh2). Some of the investigations (Koch et al. 1993; Uetz et al. 2000; Ito et al. 2001) show that some proteins could form a protein complex via protein-protein interaction, and then the protein complex binds to the motif of the target gene. The binding of transcription factors on binding site motifs is described by the sigmoid function to model the binding (ON) and nonbinding (OFF) through some threshold (Goldbeter and Koshland, 1981; Mestl et al. 1995; Gardner et al. 2003; Klipp et al. 2005). In the trans-regulatory circuit, we also consider this protein complex effect. The translated proteins Swi4 and Swi6 will form a protein complex SBF as a transcription factor to bind to the motif in the target gene, and proteins Mbp1 and Swi6 form another complex

MBF (Iyer et al. 2001). Therefore, in Figure 1, we use linear functions and nonlinear functions to express the trans-regulatory functions of the transcription factors and their complexes, respectively. The trans-regulatory functions  $g_{SBF}(t)$ ,  $g_{MBF}(t)$ ,  $g_{Fkh1}(t)$ , and  $g_{Fkh2}(t)$  of the respective cis elements SBF, MBF, Fkh1, and Fkh2 are the regulatory result of the expression profiles of transcription factors and their complexes interactions. In the trans-regulatory circuit, these trans-regulatory functions on cis elements, which are generated from upstream transcription factors and their interactive complexes, are shown as follows

$$\begin{aligned} g_{SBF}(t) &= a_{Swi4}x_{Swi4}(t) + a_{Swi6}x_{Swi6}(t) \\ &\quad + a_{Swi4 \cdot Swi6}x_{Swi4 \cdot Swi6}(t) \\ g_{MBF}(t) &= a_{Mbp1}x_{Mbp1}(t) + a_{Swi6}x_{Swi6}(t) \\ &\quad + a_{Mbp1 \cdot Swi6}x_{Mbp1 \cdot Swi6}(t) \\ g_{Fkh1}(t) &= a_{Fkh1}x_{Fkh1}(t) \\ g_{Fkh2}(t) &= a_{Fkh2}x_{Fkh2}(t) \end{aligned} \quad (1)$$

where  $g_i(t)$  denotes the trans-regulatory function to the cis element  $i$ ,  $x_i(t)$  denotes the reaction of the transcription factor  $i$  on the binding site,  $x_{i,j}(t)$  denotes the complex interaction between transcription factors  $x_i(t)$  and  $x_j(t)$ ,  $a_i$  denotes the regulatory ability of the transcription factor  $x_i(t)$  to the cis elements, and  $a_{i,j}$  denotes the regulatory ability of the complex due to transcription factors  $x_i$  and  $x_j$ .

After considering the trans effects of regulatory proteins and their protein complexes on cis elements, we also take into account the cis effects of regulatory interactions of cis elements on the transcription of coding region of the target gene. We suppose that the interaction function of two cis elements could be represented by the product of trans-regulatory functions from their transcription factors. From Figure 1, the interactions of cis elements are represented as follows:

$$\begin{aligned} g_{SBF \cdot MBF}(t) &= b_{SBF \cdot MBF}x_{Swi4}(t) \cdot x_{Swi6}(t) \cdot x_{Mbp1}(t) \\ g_{SBF \cdot Fkh1}(t) &= b_{SBF \cdot Fkh1}x_{Swi4}(t) \cdot x_{Swi6}(t) \cdot x_{Fkh1}(t) \\ g_{SBF \cdot Fkh2}(t) &= b_{SBF \cdot Fkh2}x_{Swi4}(t) \cdot x_{Swi6}(t) \cdot x_{Fkh2}(t) \\ g_{MBF \cdot Fkh1}(t) &= b_{MBF \cdot Fkh1}x_{Swi6}(t) \cdot x_{Mbp1}(t) \cdot x_{Fkh1}(t) \\ g_{MBF \cdot Fkh2}(t) &= b_{MBF \cdot Fkh2}x_{Swi6}(t) \cdot x_{Mbp1}(t) \cdot x_{Fkh2}(t) \\ g_{Fkh1 \cdot Fkh2}(t) &= b_{Fkh1 \cdot Fkh2}x_{Fkh1}(t) \cdot x_{Fkh2}(t) \end{aligned} \quad (2)$$

where  $g_{i,j}(t)$  denotes the interaction between cis elements  $i$  and  $j$ . The coefficient  $b_{i,j}$  denotes the corresponding interactive ability.

The trans-regulatory functions in Equation (1) and their cis interactions in Equation (2) will lead to the transcription of gene CLN1 with mRNA expression profiles  $y_{CLN1}(t)$  in Figure 1. The transcriptional behavior of cis regulatory circuit will be described by the following stochastic dynamic equation.

$$\dot{y}_{CLN1}(t) = G(t) - \lambda y_{CLN1}(t) + \varepsilon(t) \quad (3)$$

where the nonlinear transcriptional regulatory function  $G(t)$  is denoted as

$$\begin{aligned} G(t) &= g_{SBF}(t) + g_{MBF}(t) + g_{Fkh1}(t) + g_{Fkh2}(t) \\ &\quad + g_{SBF \cdot MBF}(t) + g_{SBF \cdot Fkh1}(t) + g_{SBF \cdot Fkh2}(t) \\ &\quad + g_{MBF \cdot Fkh1}(t) + g_{MBF \cdot Fkh2}(t) + g_{Fkh1 \cdot Fkh2}(t) + c \end{aligned}$$

and  $\lambda$  denotes the decay rate of mRNA expression profiles, which represent the degradation of mRNA. The constant  $c$  denotes the basal level of regulation, which comes from other factors than transcription factors. The gene regulatory function  $G(t)$  in (3) denotes the whole transcriptional regulation from the cis elements due to the binding of transcription factors of the gene CLN1 and the basal level from other factors.  $\varepsilon(t)$  denotes a stochastic noise due to the uncertainty and the fluctuation of mRNA microarray data.

The biological meaning of Equation (3) is that the change of mRNA level in the gene CLN1 is the synthesis due to the transcriptional regulatory function  $G(t)$  and the degradation  $-\lambda y_{CLN1}(t)$ . The mRNA expression of the gene CLN1 will increase if the synthesis regulatory function  $G(t)$  is greater than the degradation  $\lambda y_{CLN1}(t)$ . Otherwise, it will decrease. Substituting Equations (1) and (2) into Equation (3), we get the dynamic equation of trans/cis regulatory circuit in the gene transcription processing of the gene CLN1 as follows

$$\begin{aligned} \dot{y}_{CLN1}(t) &= a_{Swi4}x_{Swi4}(t) + 2a_{Swi6}x_{Swi6}(t) + a_{Mbp1}x_{Mbp1}(t) \\ &\quad + a_{Fkh1}x_{Fkh1}(t) + a_{Fkh2}x_{Fkh2}(t) \\ &\quad + a_{Swi4 \cdot Swi6}x_{Swi4 \cdot Swi6}(t) + a_{Mbp1 \cdot Swi6}x_{Mbp1 \cdot Swi6}(t) \\ &\quad + b_{SBF \cdot MBF}x_{Swi4}(t) \cdot x_{Swi6}(t) \cdot x_{Mbp1}(t) \\ &\quad + b_{SBF \cdot Fkh1}x_{Swi4}(t) \cdot x_{Swi6}(t) \cdot x_{Fkh1}(t) \\ &\quad + b_{SBF \cdot Fkh2}x_{Swi4}(t) \cdot x_{Swi6}(t) \cdot x_{Fkh2}(t) \\ &\quad + b_{MBF \cdot Fkh1}x_{Swi6}(t) \cdot x_{Mbp1}(t) \cdot x_{Fkh1}(t) \\ &\quad + b_{MBF \cdot Fkh2}x_{Swi6}(t) \cdot x_{Mbp1}(t) \cdot x_{Fkh2}(t) \\ &\quad + b_{Fkh1 \cdot Fkh2}x_{Fkh1}(t) \cdot x_{Fkh2}(t) \\ &\quad + c - \lambda y_{CLN1}(t) + \varepsilon(t) \end{aligned} \quad (4)$$



In the above dynamic equation of gene transcription,  $x_i(t)$  denotes the expression profiles of the binding transcription factors. The nonlinear term  $\prod x_i(t)$  denotes the interactions of transcription factors in the trans/cis regulatory circuit. The coefficients in the dynamic gene circuit denote the corresponding regulatory abilities of the corresponding regulatory functions and will be identified later by the corresponding microarray data.

However, at present, it is still not easy to measure directly the expression profiles of transcription factors  $x_i(t)$  in Equation (4). Because the expression profiles of yeast mRNA are available, in this study, the expression levels of these transcription factors will be replaced by expression levels of mRNA microarray data of their upstream regulatory genes but with some translation process delay in the cell. Now, for system identification of trans/cis regulatory circuit, it is more practical to consider the biochemical reactive relation between the transcription factor profiles  $x_i(t)$  at the motif binding sites and their relevant mRNA expression profiles  $y_i(t)$  of the upstream regulatory gene (Goldbeter and Koshland, 1981; Mestl et al. 1995). For this purpose, we describe the binding reaction function  $x_i(t)$  of the transcription factor on its motif binding site as a sigmoid function of mRNA expression profiles of the corresponding regulatory gene (Klipp et al. 2005). Further, it takes time for mRNA to translate into proteins as transcription factors and move to motif binding sites of the following target gene (Arava et al. 2003). We should consider this translation delay time to the sigmoid binding function when using mRNA expression profiles  $y_i(t)$  to replace expression levels  $x_i(t)$  of transcription factors in the trans/cis dynamic model (Table 1).

$$x_i(t) = f_i(y_i(t - \tau_i)) = \frac{1}{1 + \exp\{-r[y_i(t - \tau_i) - M_i]\}} \quad (5)$$

where  $r$  denotes the transition rate of the sigmoid function,  $M_i$  denotes the mean of mRNA expression level of the regulatory gene  $i$ , and  $\tau_i$  denotes the time delay of the translation time from mRNA to transcription factor (protein) for the regulatory gene  $i$ . In Figure 1,  $y_i(t)$  and  $x_i(t)$  represent the mRNA expression profile and the binding regulation function of the corresponding transcription factor on its motif binding site, respectively. The biochemical meaning of Equation (5) is that the regulation of transcription factor  $x_i(t)$  on the binding site motifs is between ON (binding) and OFF (no binding) signal with some transition region dependent on

**Table 1.** Translation delay time of 9 major transcription factors.

Transcription factor	Translation delay time(min)
Fkh1	unavailable
Fkh2	1.056818384
Mcm1	0.050641356
Ace2	0.20647784
Swi5	0.223001589
Ndd1	0.638633059
Swi4	0.495504558
Mbp1	0.408482192
Swi6	0.71182446

beyond or below some threshold of mRNA expression level of the regulatory gene after a time delay  $\tau_i$ , which is available in Arava et al. (Arava et al. 2003). The sigmoid function can also be considered as a method to normalize the expression profiles of regulatory genes between 0 and 1 to model the binding and no binding, which has been successfully employed to describe the binding of regulatory gene (Chen et al. 2004; Klipp et al. 2005; Chang et al. 2006).

Hence, based on the mRNA expression time profiles and the translation delay time, the dynamic equation of trans/cis regulatory circuit for the gene transcription of CLN1 is described by

$$\begin{aligned} \dot{y}_{CLN1}(t) &= a_{Swi4} f_{Swi4}(y_{Swi4}(t - \tau_{Swi4})) \\ &+ 2a_{Swi6} f_{Swi6}(y_{Swi6}(t - \tau_{Swi6})) \\ &+ a_{Mbp1} f_{Mbp1}(y_{Mbp1}(t - \tau_{Mbp1})) \\ &+ a_{Fkh1} f_{Fkh1}(y_{Fkh1}(t - \tau_{Fkh1})) \\ &+ a_{Fkh2} f_{Fkh2}(y_{Fkh2}(t - \tau_{Fkh2})) \\ &+ a_{Swi4-Swi6} f_{Swi4-Swi6}(y_{Swi4}(t - \tau_{Swi4}) \cdot y_{Swi6}(t - \tau_{Swi6})) \\ &+ a_{Mbp1-Swi6} f_{Mbp1-Swi6}(y_{Mbp1}(t - \tau_{Mbp1}) \cdot y_{Swi6}(t - \tau_{Swi6})) \\ &+ b_{SBF-MBF} f_{Swi4}(y_{Swi4}(t - \tau_{Swi4})) \cdot f_{Swi6}(y_{Swi6}(t - \tau_{Swi6})) \\ &\cdot f_{Mbp1}(y_{Mbp1}(t - \tau_{Mbp1})) \\ &+ b_{SBF-Fkh1} f_{Swi4}(y_{Swi4}(t - \tau_{Swi4})) \cdot f_{Swi6}(y_{Swi6}(t - \tau_{Swi6})) \\ &\cdot f_{Fkh1}(y_{Fkh1}(t - \tau_{Fkh1})) \\ &+ b_{SBF-Fkh2} f_{Swi4}(y_{Swi4}(t - \tau_{Swi4})) \cdot f_{Swi6}(y_{Swi6}(t - \tau_{Swi6})) \\ &\cdot f_{Fkh2}(y_{Fkh2}(t - \tau_{Fkh2})) \\ &+ b_{MBF-Fkh1} f_{Swi6}(y_{Swi6}(t - \tau_{Swi6})) \cdot f_{Mbp1}(y_{Mbp1}(t - \tau_{Mbp1})) \\ &\cdot f_{Fkh1}(y_{Fkh1}(t - \tau_{Fkh1})) \\ &+ b_{MBF-Fkh2} f_{Swi6}(y_{Swi6}(t - \tau_{Swi6})) \cdot f_{Mbp1}(y_{Mbp1}(t - \tau_{Mbp1})) \\ &\cdot f_{Fkh2}(y_{Fkh2}(t - \tau_{Fkh2})) \\ &+ b_{Fkh1-Fkh2} f_{Fkh1}(y_{Fkh1}(t - \tau_{Fkh1})) \cdot f_{Fkh2}(y_{Fkh2}(t - \tau_{Fkh2})) \\ &+ c - \lambda y_{CLN1}(t) + \varepsilon(t) \end{aligned} \quad (6)$$

Based on the above analysis, in general, a block diagram to construct trans/cis regulatory circuit of gene transcription is shown in Figure 2, and the nonlinear dynamic model of trans/cis regulatory circuit of gene transcription of any gene of interest in yeast can be described as follows

$$\begin{aligned} \dot{y}(t) = & \sum_{i=1}^N a_i f_i(y_i(t-\tau_i)) + \sum_{i,j} a_{i,j} f_{i,j}(y_i(t-\tau_i) \cdot y_j(t-\tau_j)) \\ & + \sum_{i,j,k} a_{i,j,k} f_{i,j,k}(y_i(t-\tau_i) \cdot y_j(t-\tau_j) \cdot y_k(t-\tau_k)) + \dots \\ & + \sum_p \left[ b_{p,q} \prod_r f_r(y_r(t-\tau_r)) \right] + c - \lambda y(t) + \varepsilon(t) \end{aligned} \quad (7)$$

Unlike the previous dynamic gene regulatory models only with linear regulatory terms (Gardner et al. 2003; Chen et al. 2004; Chen et al. 2005), in our model, nonlinear interactions among two proteins and three proteins are also considered in our gene expression dynamic model. Furthermore, the delay from transcription to translation is also considered in our dynamic model. According to the information on protein complexes and motif binding sites in the promoter region of the gene,  $y(t)$  denotes the mRNA expression profiles of a gene of interest and  $y_i(t)$ ,  $i \in \{1, 2, \dots, N\}$ , denote the mRNA expression profiles of upstream regulatory genes. With sampling expression in the next section, Equation (7) can be rewritten as.

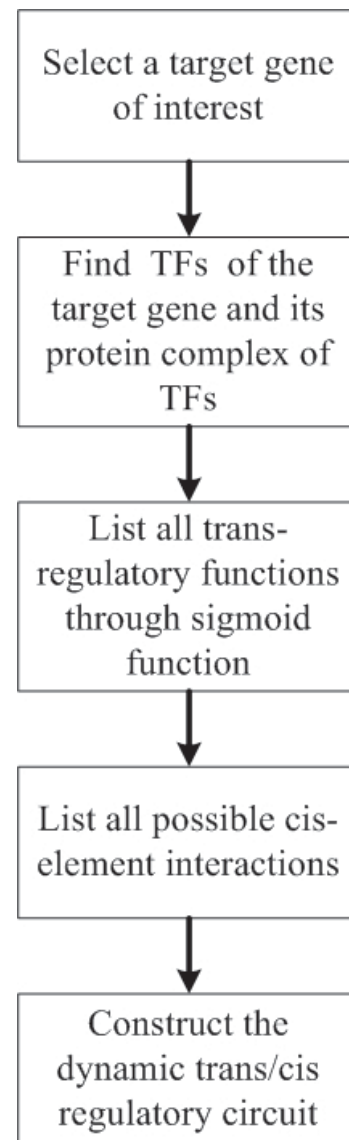
$$\dot{y}(t) + \lambda y(t) = \sum_{i=1}^L \theta_i \xi_i(t) + c + \varepsilon(t) \quad (8)$$

where  $\xi_i(t)$   $i \in \{1, 2, \dots, L\}$  denote all possible regulatory functions,  $\theta_i$   $i \in \{1, 2, \dots, L\}$  denote the corresponding regulatory abilities (coefficients), and  $L$  is the total number of input functions.

The next step is to identify the parameters  $\theta_i$ ,  $\lambda$ ,  $c$ , and the covariance  $\sigma^2$  of noise from mRNA microarray data. The parameter estimations of  $\theta_i$ ,  $\lambda$ ,  $c$ , and  $\sigma^2$  are achieved by the combination of downhill simplex search algorithm and maximum likelihood parameter method using the Methods in the sequel. After the parameter estimation is finished, we could construct a trans/cis regulatory circuit for any gene of interest via mRNA microarray data and the motif binding site information of transcription factors and their complexes.

## Experimental data

We use the yeast microarray hybridization data of Spellman et al. (Spellman et al. 1998) as our mRNA



**Figure 2.** Block diagram to construct a dynamic trans/cis regulatory circuit for gene transcription.

expression profiles. They have many experimental methods to reset the yeast cell cycle to measure mRNA expression profiles for the whole genome comprehensively. Here, we use the experimental cell cycle data of the “ $\alpha$  factor”. The information of motif binding sites in a target gene is from Simon et al. (Simon et al. 2001), Lee et al. (Lee et al. 2002) and Harbison et al. (Harbison et al. 2004), and then the protein complex data are obtained from Simon et al. (Simon et al. 2001) and the MIPS database. In the study of Simon et al. (Simon et al. 2001), there are 9 major transcription factors of cell cycle genes in yeast, so we use these 9 major transcription factors as the regulatory input of our dynamic model, which could also avoid overfitting in

parameter estimation due to only 18 time points in expression profiles. Using the above information, we could construct dynamic models as Equation (7) for cell cycle genes and then identify the parameters from mRNA microarray data by the proposed parameter estimation algorithm methods.

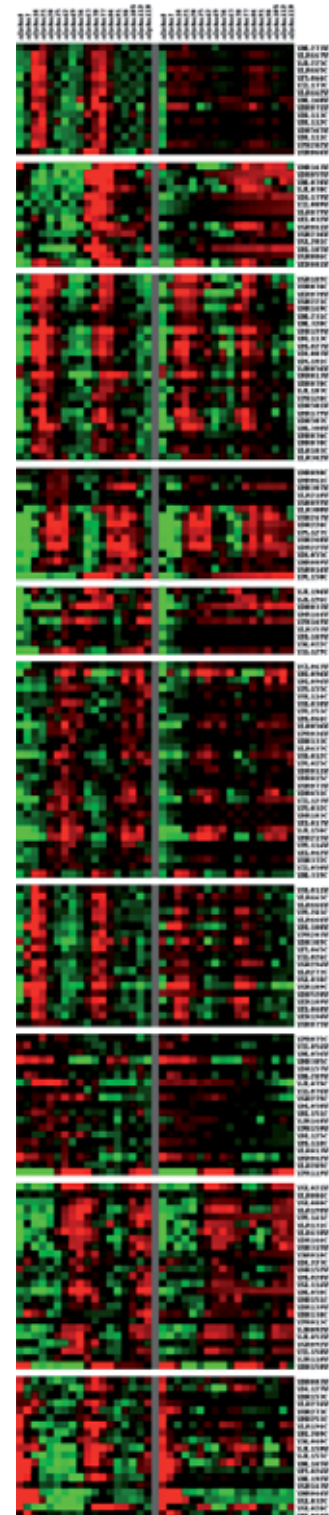
We also include the translation delay time to our dynamic model. The data of translation delay time is available from Arava et al. (Arava et al. 2003).

## Results

### Reconstruction of microarray raw data by the gene regulatory circuits

We use the “ $\alpha$  factor” microarray data of Spellman et al. (Spellman et al. 1998) and information on 9 major transcription factors from Simon et al. (Simon et al. 2001) to construct our nonlinear dynamic trans/cis model. In the study of Simon et al. (Simon et al. 2001), there are 769 cell cycle genes, but only 189 cell cycle genes have at least one motif binding site of 9 major transcription factors (P-value < 0.0015). Therefore, we could construct dynamic trans/cis regulatory circuits for these 189 cell cycle genes.

After estimating the parameters of dynamic trans/cis regulatory circuit for each gene, we compare the actual expression profiles with the constructed expression profiles of 189 genes by the cluster analysis and visualization tool (details can be found in Methods). The comparison can be seen in Figure 3 and the correlation coefficient of these two data is 0.7276. However, we found that 80 cell cycle genes have only one motif binding site of 9 major TFs. These genes may have other motif binding sites which are not of the 9 major transcription factors, or they have post-transcription to dominate their expression profiles. For example, FUN26, SHL7, STE2, AGA2, YBL111C, YBL112C, and YBL113C have many other motif binding sites, which are not in the 9 major transcription factors (Lee et al. 2002; Harbison et al. 2004). Thus these 9 transcription factors may not play dominating roles in these genes, and the performances of predicted expression profiles of these genes by dynamic trans/cis regulatory circuits may not be very satisfactory. After removing 80 genes with one motif binding site from 189 genes, the dynamic trans/cis regulatory circuits of 109 cell cycle genes with at least two motif binding sites of these 9 transcription factors are considered to be more accurate. The actual profiles and the predicted profiles by the dynamic trans/cis



**Figure 3.** Comparison between the experimental mRNA expression profiles and those predicted by the proposed model. The experimental mRNA expression profiles of 189 cell cycle genes are at left side, and the profiles which are generated by the predicted dynamic regulatory circuits are at right side. Genes represented by red tonalities are over expressed and those represented by green ones are down regulated. The correlation coefficient of both profiles is 0.7276.



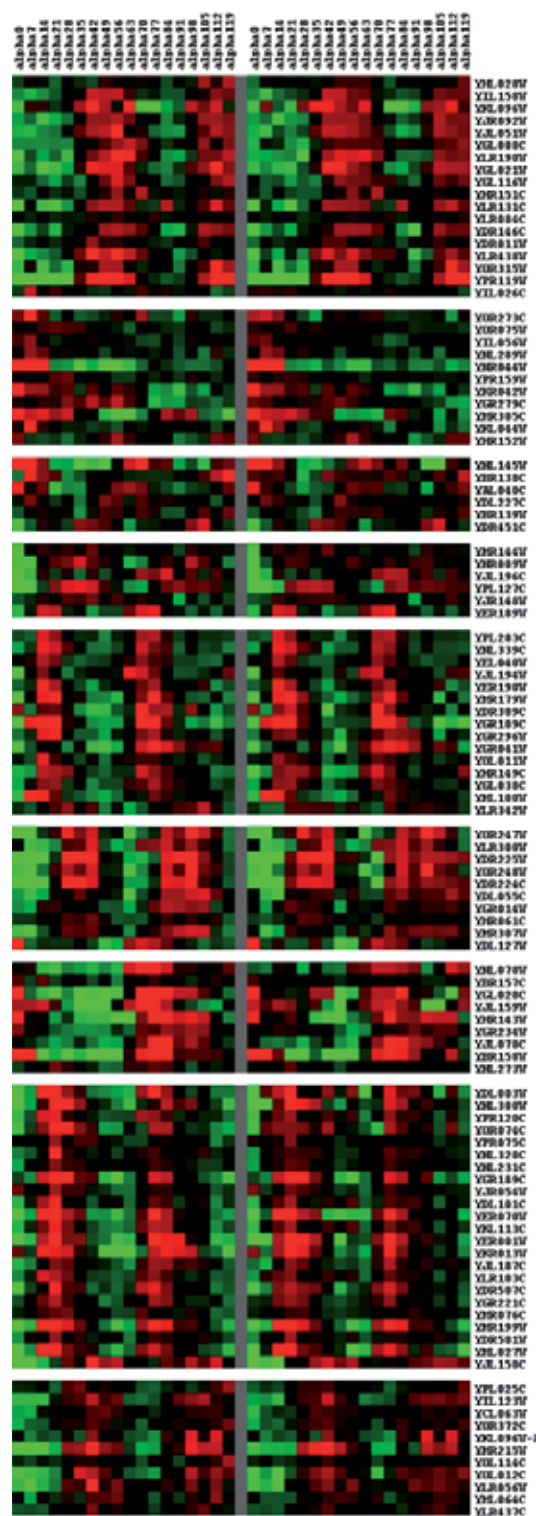
regulatory circuits are shown in Figure 4, and their correlation coefficient is 0.8502. We also use the chi-square method to test the hypothesis of trans/cis circuit by comparing our predicted data and real microarray data. We found that 176 of 189 target genes cannot be rejected by the proposed trans/cis circuit by 95% significance, and other 13 target genes cannot be tested by the chi-square method because the number of their parameters is more than the number of their expression profiles data.

### Test of shuffled microarray data

In order to test the accuracy of our model, we shuffled Spellman et al.'s "α factor" microarray data (Spellman et al. 1998) of expression profiles by random. After shuffling the microarray data, we applied our method to construct another new trans/cis regulatory circuit and then used the new regulatory circuit to generate profiles. For the 189 target genes with at least one motif of 9 major TFs, the correlation coefficient between shuffled profiles and their corresponding profiles predicted by new regulatory circuits via shuffled data is 0.1143 (shown in supplementary Figure S1), but the correlation coefficient between actual profiles and their corresponding predicted profiles is 0.7276. Furthermore, for the 109 target genes with at least two more motifs of 9 major TFs, the correlation coefficient between shuffled profiles and their corresponding profiles predicted by shuffled data is 0.5939 (shown in supplementary Figure S2), but the correlation coefficient between actual profiles and their corresponding predicted profiles is 0.8502. Obviously, only data generated by real biological systems could be identified well by the proposed model. From the view of system identification, a proper system model will lead to a good system identification if overfitting could be avoided. In our case, the proposed model is applicable to the trans/cis regulatory circuit because the correlation coefficient of real microarray data is much larger than that of shuffled microarray data. So we can say that our dynamic model is proper for the trans/cis regulatory circuit of gene expression program of yeast.

### Regulatory activities of TFs in the target gene

From the estimated parameters, we can find which transcription factor activates or represses its target genes and quantify its activity. Therefore, we can see which transcription factor can affect certain



**Figure 4.** Comparison between the experimental mRNA expression profiles and those predicted by the proposed model. The experimental mRNA expression profiles of 109 cell cycle genes are at the left side, and the profiles predicted by the dynamic regulatory circuits are at the right side. And the correlation coefficient of both profiles is 0.8502.



target gene a lot, and why the same transcription factor may cause different effects on different target genes. In Table 2, we can see there are many transcription factors regulating the target gene CLN1. CLN1, a G1/S-specific cyclin, is regulated by SBF, MBF, Fkh1, and Fkh2 (Simon et al. 2001). By the estimated kinetic parameters of MBF (Swi6, Mbp1, and Mbp1 · Swi6), we found that Swi6 dominates the effect of MBF on target gene CLN1, and it may play an important regulatory role in CLN1. The kinetic parameters of Mbp1 and Mbp1 · Swi6 are of smaller scale, which means that Mbp1 may play a certain kind of role but not an important regulatory role in CLN1. Mbp1 may be necessary for CLN1, but has no more affection on CLN1 after the expression of Mbp1 exceeds certain level. This result could reflect that Mbp1 is a DNA-binding component of MBF, and Swi6 may have a regulatory function (Iyer et al. 2001). For parameters of SBF (Swi4, Swi6, and Swi4 · Swi6), Swi4 and Swi6 play major roles of SBF in CLN1, but the role of Swi4 · Swi6 is minor. It shows that the linear combination of Swi4 and Swi6 has significant effect on target gene CLN1. Hence, we consider that Swi6 may have a regulatory function of SBF in CLN1, and Swi4 is also significant on target gene CLN1. From Table 2, Fkh1 and Fkh2 contribute negative regulation to target gene CLN1. Comparing these kinetic parameters, we know that Fkh1, Swi4, and Swi6 affect target gene CLN1 more than Fkh2 and Mbp1 do.

Our model also provides another important estimation of interactions among the cis elements. All possible interactions among the cis elements are considered, and then their kinetic parameters are estimated by expression profiles of microarray data. These estimated kinetic parameters could not only tell us the possibilities of these interactions of cis elements, but also the significance of these interactions. For example, we consider 5 cis-element interactions in target gene CLN1. From the kinetic parameters estimated, there are three parameters larger than the others. We may conclude that these three terms, MBF · Fkh2, SBF · Fkh1, and SBF · Fkh2, could be the three possible interactions among real cis elements.

With the same transcription factor binding to different target genes, different kinetic parameters of this transcription factor determine different regulatory effects on the target genes. A larger parameter of the same transcription factor binding to different target genes means that it plays a more

important role and is more sensitive than those binding to another target gene. From our estimated kinetic parameters, we can compare these effects on different target genes with the same transcription factor. The G<sub>1</sub> cyclin PCL2, which associates with Pho85p cyclin-dependent kinase (Cdk) to contribute to entry into the mitotic cell cycle, is regulated by SBF, Ace2, and Swi5. (Simon et al. 2001). On the kinetic parameters of PCL2 (see Table 2), the effect of transcription factor SBF on PCL2 is dominated by Swi6 and Swi4 · Swi6. For another target gene MNN1, required for addition of alpha1,3-mannose linkages to N-linked and O-linked oligosaccharides, the effect of transcription factor SBF on MNN1 is also dominated by Swi6 and Swi4 · Swi6, which can be seen in SGD database. However, the kinetic parameters of SBF on PCL2 (Simon et al. 2001) are all smaller than those of SBF on MNN1. We can consider that SBF has more regulatory effect on MNN1 than PCL2 and the sensitivity of SBF on MNN1 is more than that on PCL2. After simulation by changing the expression profiles of Swi4 or Swi6, of which SBF is composed, we find that the result is the same as we discussed above.

### Possibilities of cis-element interactions

In order to find the most possible cis-element interactions, we sorted all possible kinetic parameters of cis-element interactions calculated by our method as a distribution and then found those within two-sided 90% confidence interval of the distribution. There are 314 possible cis-element interactions within the 189 cell cycle genes selected before, and then we found 15 cis-element interaction terms in each side of this two-sided interval. Some of these 30 interaction terms appear more frequently, which include complex interactions of Swi4/Mbp1/Swi6, Fkh2/Swi4/Swi6, and Ace2/Swi5. In Table 3, we count the frequencies of these cis-element interactions, which occur more than two times within the two-sided 90% confidence interval.

For the complex interactions of Swi4/Mbp1/Swi6, both SBF, the complex of Swi4 and Swi6, and MBF, the complex of Mbp1 and Swi6, control the transcription of G1/S cyclin genes, and many genes, Cdc6, Swi4, Clb6, Swe1, and Cln1, are both bound by SBF and MBF at the same time. Therefore, it is possible that the activators (Swi4/Mbp1/Swi6) have a cis-element interaction, and the possibility of their relationship has been reported by Kato

**Table 2.** Parameters of dynamic trans/cis regulatory models of CLN1, UTR2, and MNN1 based on Equation (4).

CLN1 $\dot{y}_{CLN1}(t)=$	Terms due to MBF	$-0.009x_{Mbp1}(t-0.4)+0.025x_{Swi6}(t-0.7)-0.0027f_{Mbp1,Swi6}(y_{Mbp1}(t-0.4)\cdot y_{Swi6}(t-0.7))$
	Terms due to SBF	$-0.01x_{Swi4}(t-0.5)+0.02516x_{Swi6}(t-0.7)+0.003435f_{Swi4,Swi6}(y_{Swi4}(t-0.5)\cdot y_{Swi6}(t-0.6))$
	Terms of Fkh1 and Fkh2	$-0.36396x_{Fkh1}(t)-0.024962x_{Fkh2}(t-1)$
	Interaction between MBF and SBF	$-0.0027745x_{Swi4}(t-0.5)x_{Swi6}(t-0.7)x_{Mbp1}(t-0.4)$
	Interaction between SBF and Fkh1	$-0.011154x_{Swi4}(t-0.5)x_{Swi6}(t-0.7)x_{Fkh1}(t)$
	Interaction between SBF and Fkh2	$-0.014026x_{Swi4}(t-0.5)x_{Swi6}(t-0.7)x_{Fkh2}(t-1)$
	Interaction between MBF and Fkh1	$-0.0057164x_{Swi6}(t-0.7)x_{Mbp1}(t-0.4)x_{Fkh1}(t)$
	Interaction between MBF and Fkh2	$-0.046408x_{Swi6}(t-0.7)x_{Mbp1}(t-0.4)x_{Fkh2}(t-1)$
	Interaction between Fkh1 and Fkh2	$-0.0023141x_{Fkh1}(t)x_{Fkh2}(t-1)$
	Decay rate and basal level	$-4.0839-0.20719y_{CLN1}(t)$
PCL2 $\dot{y}_{PCL2}(t)=$	Terms due to SBF	$0.0167x_{Swi4}(t-0.5)-0.0384x_{Swi6}(t-0.7)-0.0464f_{Swi4,Swi6}(y_{Swi4}(t-0.5)\cdot y_{Swi6}(t-0.6))$
	Terms due to Ace2 and Swi5	$+0.070897x_{Ace2}(t-0.21)+0.1044x_{Swi5}(t-0.22)$
	Interaction between Ace2 and Swi5	$+0.045873x_{Ace2}(t-0.21)x_{Swi5}(t-0.22)$
	Interaction between SBF and Ace2	$+0.17035x_{Swi4}(t-0.5)x_{Swi6}(t-0.7)x_{Ace2}(t-0.21)$
	Interaction between SBF and Swi5	$+0.093195x_{Swi4}(t-0.5)x_{Swi6}(t-0.7)x_{Swi5}(t-0.22)$
	Decay rate and basal level	$+1.2225-0.43022y_{PCL2}(t)$
MNN1 $\dot{y}_{MNN1}(t)=$	Terms due to SBF	$-0.0432x_{Swi4}(t-0.5)-0.394x_{Swi6}(t-0.7)-0.142f_{Swi4,Swi6}(y_{Swi4}(t-0.5)\cdot y_{Swi6}(t-0.6))$
	Decay rate and basal level	$+0.81055-0.45975y_{MNN1}(t)$

**Table 3.** The possible cis element interactions which appear within the two-sided 90% confidence interval (there are 30 cis-element interactions within the two-sided 90% confidence interval).

Possible cis element interactions	Counts of appearances within the two-sided 90% confidence interval
Swi4/Mbp1/Swi6	5
Fkh2/Swi4/Swi6	4
Ace2/Swi5	3
Mcm1/Swi5	3
Fkh1/Fkh2	2
Fkh2/Mcm1	2
Fkh2/Mbp1/Swi6	2
Mcm1/Swi4/Swi6	2
Swi4/Swi5/Swi6	2

et al. (Kato et al. 2004). The other two possible cis-element interaction terms (Fkh2/Swi4/Swi6 and Ace2/Swi5) calculated by our method may also exist. A few genomic analyses have indicated the involvement of SBF and Fkh1/Fkh2 in S phase (Lee et al. 2002), and SBF and Fkh2 most probably regulate budding, cell-wall synthesis, and spindle-related genes in S phase (Kato et al. 2004). Ace2 and Swi5 are a pair of TFs of yeast that regulate the expression of many cell cycle-specific genes, including Sic1, an inhibitor of Cdc28 protein kinase, Rme1, a regulator of meiosis, and Ash1, a regulator of meiosis (Doolin et al. 2001). In recent studies, Ace2 and Swi5 cooperate to induce the expressions of a subset of genes, but the antagonistic interaction of Ace2 and Swi5 was found (Doolin et al. 2001). With 82% identical DNA binding domains, Ace2 and Swi5 bind to the same DNA sequence (McBride et al. 1999), and it is possible that proteins compete for access on these promoters, but only one activates transcription (Doolin et al. 2001). Therefore, one partner of Swi5 and Ace2 sometimes can have a stronger contribution towards regulation, and the antagonistic interaction of Ace2 and Swi5 found is not surprising.

## Discussion

Our method uses mRNA expression profiles, protein complexes, translation time delay, and the information of binding site motif to construct a dynamic trans/cis regulatory circuit to gain more insight into the gene expression of yeast. It is more precise to use these kinds of data to construct a nonlinear stochastic regulatory model for the gene

transcription. Furthermore, a stochastic regulatory model can easily describe the properties of change, interaction, and uncertainty in mRNA expression profiles. Recently, Vu and Vohradsky also used nonlinear dynamic model to infer the transcriptional regulators of target genes (Vu and Vohradsky, 2007). They successfully identified possible transcriptional factors of yeast cell cycle, which shows the power of nonlinear dynamic models.

Constructing the trans/cis regulatory circuits of the cell cycle genes of yeast is useful to quantify the influence of transcription factors on their target genes, and then the possibilities of cis element interactions could be found from the statistical perspective. The confirmation of these possible cis element interactions would be a direction of further research. In addition, the proposed method can provide a quantitative basis for system analysis of gene circuit and give a scheme for gene circuit design with a desired gene expression in the future. Not only could the data of *Saccharomyces cerevisiae* be applied by our method, but those of other species also could. Any species with their mRNA expression profiles and the information of binding site motif can be applied using this method to construct their trans/cis regulatory circuits.

However, in this study, we found that the mRNA expression profiles of several genes generated by the predicted dynamic trans/cis regulatory circuit have a little difference from their original mRNA expression profiles. It may be because 9 transcriptional factors are not enough in these cases in which the genes may not be controlled dominantly by 9 TFs. We may use ChIP-chip data (Harbison et al. 2004) to replace the 9 major TFs from Simon et al. (Simon et al. 2001) to construct the transcriptional regulatory circuit. 204 DNA-binding transcriptional factors are found in yeast (Harbison et al. 2004), which may be useful for this elaboration of the transcriptional regulatory circuit. Furthermore, only 18 time points of mRNA expression profiles are used to estimate the parameters of dynamic model in each gene, which are not enough to get precise parameter estimation because of the large number of kinetic parameters to be estimated, which will easily lead to overfitting. If we include all possible 204 TFs into our model, the number of parameters that we want to estimate may be so large that overfitting would happen. Consequently, we only choose the most important 9 TFs in our model to avoid the overfitting in parameter estimation. If there are more time point profiles of mRNA



expression and information of binding site motifs, the accuracy of the dynamic trans/cis regulatory circuit could be improved by the proposed method.

In conclusion, unlike the convention linear dynamic models, this study provides a nonlinear stochastic dynamic trans/cis regulatory circuit for any species via mRNA expression profiles, the information of binding site motif, protein complexes, and translation time delay to gain more insight into transcriptional regulatory infrastructures of gene expression program of yeast. The results are confirmed by a statistical hypothesis test.

## Methods

### Parameter estimation of dynamic trans/cis regulatory circuit

The dynamic trans/cis regulatory circuit of a gene of interest can be described by the dynamic Equation (7) according to the mRNA expression profiles of upstream regulatory genes and their nonlinear interactions. The parameters of the dynamic trans/cis circuit are then specified to meet the practical input/output microarray data of the regulatory genes and the target gene.

In order to avoid the interruption of high frequency noise with the use of the derivative information  $\dot{y}(t)$ , the solution of the dynamic Equation (8) is expressed by the following integration equation.

$$\begin{aligned} y(t) &= y(0) \cdot e^{-\lambda t} + \sum_i \theta_i e^{-\lambda t} \int_0^t e^{\lambda \tau} \xi_i(\tau) d\tau + c e^{-\lambda t} \int_0^t e^{\lambda \tau} d\tau \\ &\quad + e^{-\lambda t} \int_0^t \varepsilon(\tau) e^{\lambda \tau} d\tau \\ &= K e^{-\lambda t} + \sum_i \theta_i e^{-\lambda t} \int_0^t e^{\lambda \tau} \xi_i(\tau) d\tau + k + e(t) \end{aligned} \quad (9)$$

where  $k \equiv \frac{c}{\lambda}$  and  $K \equiv y(0) - \frac{c}{\lambda}$  are constants. Noise of  $y(t)$  is written with  $e(t) \equiv e^{-\lambda t} \int_0^t \varepsilon(\tau) e^{\lambda \tau} d\tau$ , i.e. the mRNA concentration could be generated by dynamic Equation (9) through upstream regulatory mRNA  $y_1(t)$   $y_2(t)$  ...  $y_N(t)$  if their parameters could be estimated.

In Equation (9),  $\xi_i(t)$  is a combinative regulatory function of the regulatory genes. We use a third order polynomial cubic spline method to approximate  $\xi_i(t)$  through microarray data with 18 time points, and then use the partial integration method

to calculate  $e^{-\lambda t} \int_0^t e^{\lambda \tau} \xi_i(\tau) d\tau$  (Faires and Burden, 1998).

For  $\xi_i(t)$  function, we use the cubic spline method to approach it. Cubic spline is a method of using a third order polynomial to approach every four data points (Faires and Burden, 1998). So we can rewrite the function  $\xi_i(t)$  as follows.

$$\xi_i(t) = \alpha_i t^3 + \beta_i t^2 + \gamma_i t + \delta_i \quad (10)$$

From Equations (9) and (10), we can use partial integration to refine Equation (9) as follows.

$$\begin{aligned} y(t) &= \Lambda e^{-\lambda t} + \sum_i \theta_i \cdot \left( \frac{\xi_i(t)}{\lambda} - \frac{\xi_i'(t)}{\lambda^2} + \frac{\xi_i''(t)}{\lambda^3} - \frac{\xi_i'''(t)}{\lambda^4} \right) \\ &\quad + k + e(t) \end{aligned} \quad (11)$$

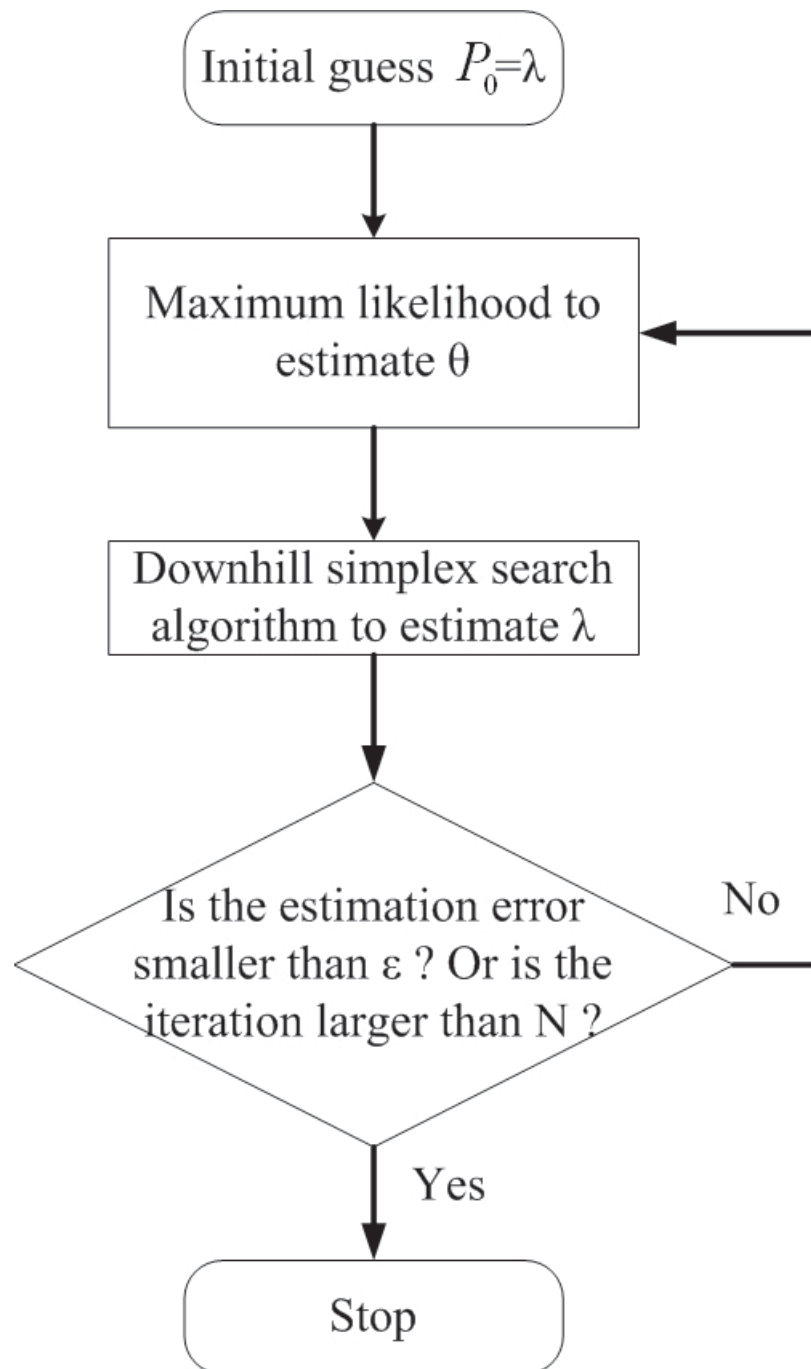
where  $\xi_i'(t) = 3\alpha_i t^2 + 2\beta_i t + \gamma_i$   
 $\xi_i''(t) = 6\alpha_i t + 2\beta_i$ , and  $\xi_i'''(t) = 12\alpha_i$  and

$$\Lambda = K - \sum_i \theta_i \cdot \left( \frac{\xi_i(0)}{\lambda} - \frac{\xi_i'(0)}{\lambda^2} + \frac{\xi_i''(0)}{\lambda^3} - \frac{\xi_i'''(0)}{\lambda^4} \right).$$

After applying the cubic spline method to solve and rewrite the dynamic Equation (9), we use a search algorithm, downhill simplex search method, for parameter estimations (Jang et al. 1997). However, this search method is not very efficient. Therefore, we combine the downhill simplex search method and maximum likelihood method to estimate the parameters. We use the downhill simplex search method to estimate the nonlinear parameter  $\lambda$  at first and then use the maximum likelihood method to estimate linear parameters  $K$ ,  $\theta_i$ , and  $k$ .

After estimating  $\lambda$  by downhill simplex research method, we use the method of maximum likelihood to estimate the other linear parameters. Equation (11) can be written as the following regression form.

$$\begin{aligned} y(t) &= \left[ e^{-\lambda t} \left( \frac{\xi_i(t)}{\lambda} - \frac{\xi_i'(t)}{\lambda^2} + \frac{\xi_i''(t)}{\lambda^3} - \frac{\xi_i'''(t)}{\lambda^4} \right) \right. \\ &\quad \left. \dots \left( \frac{\xi_i(t)}{\lambda} - \frac{\xi_i'(t)}{\lambda^2} + \frac{\xi_i''(t)}{\lambda^3} - \frac{\xi_i'''(t)}{\lambda^4} \right) 1 \right] \begin{bmatrix} \Lambda \\ \theta_1 \\ \vdots \\ \theta_i \\ k \end{bmatrix} + e(t) \end{aligned} \quad (12)$$



**Figure 5.** The combined method of the downhill simplex search and the maximum likelihood estimation for  $\lambda$  and  $\theta$  iteratively.

$$\equiv \phi(t) \cdot \theta + e(t)$$

where  $\phi(t)$  denotes the regression vector which can be obtained from the above processing.  $\theta$  is the parameter vector of trans/cis regulatory circuit which is to be estimated to make the transcriptional expression of trans/cis circuit meet the microarray data.

If we have a lot of microarray data points to process by the method above to get values of

$$\left\{ y(t_i) \left( \frac{\xi_l(t)}{\lambda} - \frac{\xi_l'(t)}{\lambda^2} + \frac{\xi_l''(t)}{\lambda^3} - \frac{\xi_l'''(t)}{\lambda^4} \right) \right\}.$$

for  $l \in \{1, 2, \dots, 18\}$  and  $i \in \{1, 2, \dots, L\}$  the parameter vector can be estimated by the following approach. Using a matrix notation, Equation (12) at different time points is of the following form

$$\begin{bmatrix} y(t_1) \\ y(t_2) \\ \vdots \\ y(t_{18}) \end{bmatrix} = \begin{bmatrix} \phi(t_1) \\ \phi(t_2) \\ \vdots \\ \phi(t_{18}) \end{bmatrix} \cdot \theta + \begin{bmatrix} e(t_1) \\ e(t_2) \\ \vdots \\ e(t_{18}) \end{bmatrix} \quad (13)$$

For simplicity, we can further define the notations  $Y$ ,  $\Phi$ , and  $E$  to represent Equation (13) as follows.

$$Y = \Phi \cdot \theta + E \quad (14)$$

In Equation (13), we assume noises  $e(t_i)$  at different time points as independent random variables of normal distribution with zero mean and unknown variance  $\sigma^2$ , i.e. the variance of  $E$  is  $\Sigma = E\{EE^T\} = \sigma^2 I$ , where  $I$  is a unit matrix. In this study, a maximum likelihood parameter estimation method will be used to estimate  $\theta$  and  $\sigma^2$  from microarray data of regulatory genes and the target gene (Johansson, 1993). If  $E$  is assumed to be normally distributed with  $N$  elements, its probability density function is of the following form

$$p(E) = \frac{1}{((2\pi)^N \det \Sigma)^{1/2}} \exp \left\{ -\frac{E^T \Sigma^{-1} E}{2} \right\} \quad (15)$$

Under Equation (14) we have the likelihood function

$$p(\theta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{-N/2}} \exp \left\{ -\frac{(Y - \Phi \cdot \theta)^T (Y - \Phi \cdot \theta)}{2\sigma^2} \right\} \quad (16)$$

Equation (16) can be considered as a function of parameters  $\theta$  and  $\sigma^2$ . We want to specify  $\theta$  and  $\sigma^2$  to maximize the likelihood function in (16). It is practical to take the logarithm of their likelihood function, and then we have the following log-likelihood function as follows.

$$\log L(\theta, \sigma^2) = -\frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{k=1}^N [y(t_k) - \phi(t_k) \cdot \theta]^2 \quad (17)$$

where  $y(t_k)$  and  $\phi(t_k)$  are the  $k$ -th elements of  $Y$  and  $\Phi$ , respectively.

Here we expect the log-likelihood function to have the maximum at  $\theta = \hat{\theta}$  and  $\hat{\sigma}^2$ . The necessary condition for maximum likelihood estimates  $\hat{\theta}$  and  $\hat{\sigma}^2$  as follows (Johansson, 1993)

$$\begin{aligned} \frac{\partial \log L(\theta, \sigma^2)}{\partial \theta} \bigg|_{\theta=\hat{\theta}} &= 0 \\ \frac{\partial \log L(\theta, \sigma^2)}{\partial \sigma^2} \bigg|_{\sigma^2=\hat{\sigma}^2} &= 0 \end{aligned} \quad (18)$$

The estimated parameters  $\hat{\theta}$  and  $\hat{\sigma}^2$  are shown below.

$$\hat{\theta} = (\Phi^T \Phi)^{-1} \Phi^T Y \quad (19)$$

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{N} \sum_{k=1}^N [y(t_k) - \phi(t_k) \cdot \hat{\theta}]^2 \\ &= \frac{1}{N} (Y - \Phi \cdot \hat{\theta})^T (Y - \Phi \cdot \hat{\theta}) \end{aligned} \quad (20)$$

where  $Y$  and  $\Phi$  can be obtained from the microarray data of regulatory genes and the target gene.

After estimating  $\hat{\theta}$  by maximum likelihood method, we take  $\hat{\theta}$  into downhill simplex search method to estimate  $\lambda$  again. We iterate these two methods until the differences of real data and estimated data are as small as possible. All iteration processes are shown in Figure. 5.

## Comparison between actual/constructed expression profiles

We use the cluster analysis and visualization tool (Cluster and Treeview) written by Michael Eisen (<http://rana.lbl.gov/EisenSoftware.htm>) to compare the actual expression profiles with the constructed expression profiles. When clustering the expression profiles, we use hierarchical clustering methods (Eisen et al. 1998).

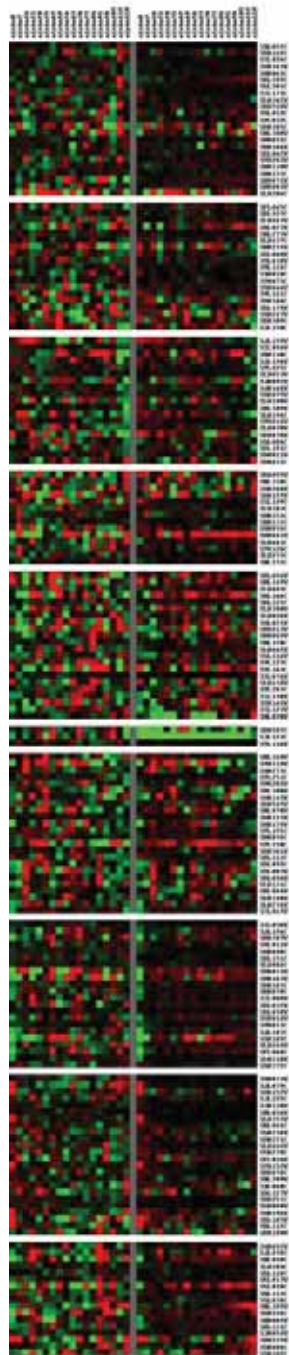


## Reference

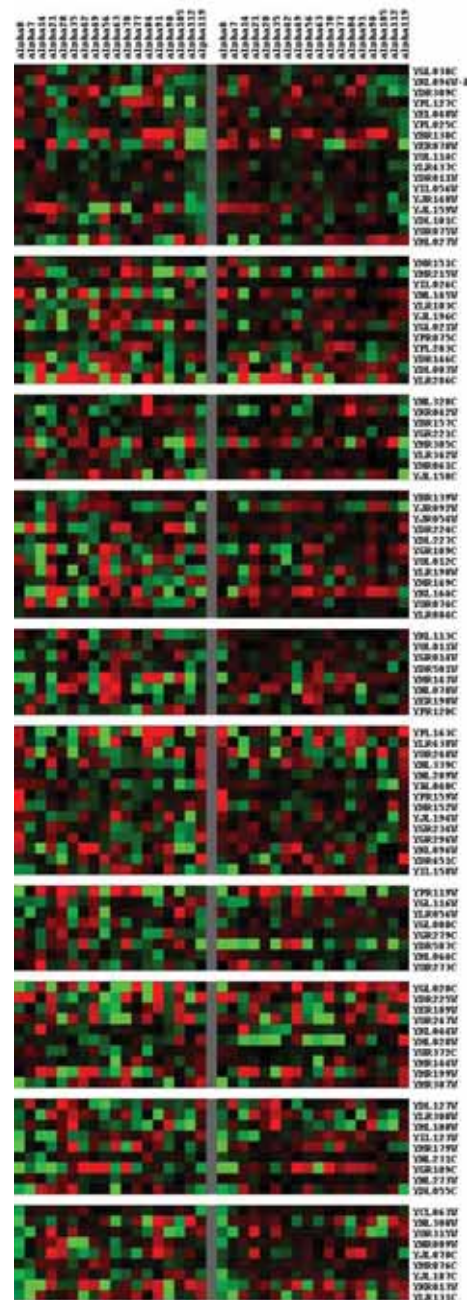
- Arava, Y., Wang, Y., Storey, J.D. et al. 2003. Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*, *Proc. Natl. Acad. Sci. U.S.A.*, 100:3889–3894.
- Berman, B.P., Nibu, Y., Pfeiffer, B.D. et al. 2002. Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the *Drosophila* genome, *Proc. Natl. Acad. Sci. U.S.A.*, 99:757–762.
- Carlson, J.M. and Doyle, J. 1999. Highly optimized tolerance: a mechanism for power laws in designed systems, *Phys. Rev. E Stat. Phys. Plasmas. Fluids Relat Interdiscip Topics*, 60:1412–1427.
- Chang, Y.H., Wang, Y.C. and Chen, B.S. 2006. Identification of transcription factor cooperativity via stochastic system model, *Bioinformatics*, 22:2276–2282.
- Chen, H.C., Lee, H.C., Lin, T.Y., et al. 2004. Quantitative characterization of the transcriptional regulatory network in the yeast cell cycle, *Bioinformatics*, 20:1914–1927.
- Chen, K.C., Wang, T.Y., Tseng, H.H., et al. 2005. A stochastic differential equation model for quantifying transcriptional regulatory network in *Saccharomyces cerevisiae*, *Bioinformatics*, 21:2883–2890.
- Csete, M.E. and Doyle, J.C. 2002. Reverse engineering of biological complexity, *Science*, 295:1664–1669.
- Davidson, E.H., McClay, D.R. and Hood, L. 2003. Regulatory gene networks and the properties of the developmental process, *Proc. Natl. Acad. Sci. U.S.A.*, 100:1475–1480.
- Doolin, M.T., Johnson, A.L., Johnston, L.H., et al. 2001. Overlapping and distinct roles of the duplicated yeast transcription factors Ace2p and Swi5p, *Mol. Microbiol.*, 40:422–432.
- Eisen, M.B., Spellman, P.T., Brown, P.O., et al. 1998. Cluster analysis and display of genome-wide expression patterns, *Proc. Natl. Acad. Sci. U.S.A.*, 95:14863–14868.
- Faires, J.D. and Burden, R. 1998. Numerical methods Pacific Grove, CA: Brooks/Cole Pub. Co.
- Gardner, T.S., di Bernardo, D., Lorenz, D. et al. 2003. Inferring genetic networks and identifying compound mode of action via expression profiling, *Science*, 301:102–105.
- Goldbeter, A. and Koshland, D.E., Jr. 1981. An amplified sensitivity arising from covalent modification in biological systems, *Proc. Natl. Acad. Sci. U.S.A.*, 78:6840–6844.
- Harbison, C.T., Gordon, D.B., Lee, T.I. et al. 2004. Transcriptional regulatory code of a eukaryotic genome, *Nature*, 431: 99–104.
- Hartemink, A.J., Gifford, D.K., Jaakkola, T.S. et al. 2002. Combining location and expression data for principled discovery of genetic regulatory network models, *Pac. Symp. Biocomput.*, 437–449.
- Hasty, J., McMillen, D. and Collins, J.J. 2002. Engineered gene circuits, *Nature*, 420:224–230.
- Ito, T., Chiba, T., Ozawa, R. et al. 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome, *Proc. Natl. Acad. Sci. U.S.A.*, 98:4569–4574.
- Iyer, V.R., Horak, C.E., Scafe, C.S. et al. 2001. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF, *Nature*, 409: 533–538.
- Jang, J.S., Sun, C.T. and Mizutani, E. 1997. Neuro-fuzzy and soft computing Upper Saddle River, NJ Prentice-Hall.
- Johansson, R. 1993. System modeling and identification Englewood Cliffs, NJ Prentice-Hall.
- Kato, M., Hata, N., Banerjee, N. et al. 2004. Identifying combinatorial regulation of transcription factors and binding motifs, *Genome Biol.*, 5: R56.
- Klipp, E., Herwig, R., Kowald, A. et al. 2005. Systems Biology in Practice: Concepts, Implementation and Application. Germany: Wiley-VCH.
- Koch, C., Moll, T., Neuberg, M. et al. 1993. A role for the transcription factors Mbp1 and Swi4 in progression from G1 to S phase, *Science*, 261:1551–1557.
- Lee, T.I., Rinaldi, N.J., Robert, F. et al. 2002. Transcriptional regulatory networks in *Saccharomyces cerevisiae*, *Science*, 298:799–804.
- McBride, H.J., Yu, Y. and Stillman, D.J. 1999. Distinct regions of the Swi5 and Ace2 transcription factors are required for specific gene activation, *J. Biol. Chem.*, 274:21029–21036.
- 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.
- Simon, I., Barnett, J., Hannett, N. et al. 2001. Serial regulation of transcriptional regulators in the yeast cell cycle, *Cell*, 106:697–708.
- Sontag, E., Kiyatkin, A. and Kholodenko, B.N. 2004. Inferring dynamic architecture of cellular networks using time series of gene expression, protein and metabolite data, *Bioinformatics*, 20:1877–1886.
- Spellman, P.T., Sherlock, G., Zhang, M.Q. et al. 1998. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization, *Mol. Biol. Cell*, 9:3273–3297.
- Tegner, J., Yeung, M.K., Hasty, J. et al. 2003. Reverse engineering gene networks: integrating genetic perturbations with dynamical modeling, *Proc. Natl. Acad. Sci., U.S.A.*, 100:5944–5949.
- Uetz, P., Giot, L., Cagney, G. et al. 2000. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*, *Nature*, 403:623–627.
- Vu, T.T. and Vohradsky, J. 2007. Nonlinear differential equation model for quantification of transcriptional regulation applied to microarray data of *Saccharomyces cerevisiae*, *Nucleic Acids Res.*, 35:279–287.
- 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. U.S.A.*, 99:6163–6168.
- Yi, T.M., Huang, Y., Simon, M.I. et al. 2000. Robust perfect adaptation in bacterial chemotaxis through integral feedback control, *Proc. Natl. Acad. Sci. U.S.A.*, 97:4649–4653.
- Yuh, C.H., Bolouri, H. and Davidson, E.H. 1998. Genomic cis-regulatory logic: experimental and computational analysis of a sea urchin gene, *Science*, 279:1896–1902.
- Yuh, C.H. and Davidson, E.H. 1996. Modular cis-regulatory organization of Endo16, a gut-specific gene of the sea urchin embryo, *Development*, 122:1069–1082.
- Yuh, C.H., Moore, J.G. and Davidson, E.H. 1996. Quantitative functional interrelations within the cis-regulatory system of the *S. purpuratus* Endo16 gene, *Development*, 122:4045–4056.

# Nonlinear Dynamic Trans/Cis Regulatory Circuit for Gene Transcription via Microarray Data

## Supplementary materials



**Figure S1.** Comparison between the shuffled experimental mRNA expression profiles and those predicted by the proposed model. The shuffled experimental mRNA expression profiles of 189 cell cycle genes are at the left side, and the profiles predicted by the dynamic regulatory circuits are at the right side. And the correlation coefficient of both profiles is 0.1143.



**Figure S2.** Comparison between the shuffled experimental mRNA expression profiles and those predicted by the proposed model. The shuffled experimental mRNA expression profiles of 109 cell cycle genes are at the left side, and the profiles predicted by the dynamic regulatory circuits are at the right side. And the correlation coefficient of both profiles is 0.5939.