International Journal of Computational Intelligence in Control, 9(1), 2017

System Identification and Robustness Analysis of the Circadian Regulatory Network via ARX Stochastic Interactive Model

Cheng-Wei Li, Wen-Chieh Chang & Bor-Sen Chen

Lab. of Systems Biology, National Tsing Hua University, Hsinchu, 300, Taiwan

Abstract: *The circadian regulatory network is one of the main topics of plant investigations. The intracellular interactions among genes in response to the environmental stimuli of light are related to the foundation of functional genomics in plant. However, to determine what genes and how much they influence each other through transcriptional binding or physical interaction using biological assays is not easy. In this study, we detected the interactive relationships of genes in the circadian regulatory network via a systematically stochastic modeling analysis* in silico. *After the successful construction of the circadian regulatory network, we would analyze the stochastic system under different assumed biological conditions by performing dry experiments and discover the essential robustness (or sensitivity) characteristics of the circadian regulatory network by biological perturbations. In short, we can construct and analyze a genetic regulatory pathway of circadian network from the systems biology viewpoint.*

1. INTRODUCTION

Biological phenomena at different organismic levels have implicitly revealed some sophisticated systematic architectures of cellular and physiological activities. These architectures were built upon the biochemical processes before the emergence of proteome and transcriptome (Kettman *et al.*, 2001; Scheel *et al.*, 2002). Under the molecular machinery, the biochemical processes are mostly interpreted as frameworks of connectivity between biochemical compounds and proteins, which are synthesized from genes to function as transcription factors bound to regulatory sites of other genes, such as enzymes catalyzing metabolic reactions or components of signal transduction pathways (Harkin, 2000). This implies that, in order to understand the molecular mechanism of genes in the control of intracellular or intercellular processes, the scope should be broadened from DNA sequences coding for proteins to the systems of genetic regulatory pathways determining which genes are expressed, when and where in the organism and to which extent (Yanovsky and Kay, 2001). In the experience of engineering field, the systematic architecture and dynamic model could analyze the characteristics of signaling regulatory networks. Therefore, from the system structure point of view how to construct the dynamic model of a signaling regulatory network is an important topic of systems biology. Most biological phenomena such as metabolism, stress response (Motaki *et al.*, 2003), and cell cycle are directly or indirectly influenced by genes and have been well studied on the molecular basis. Thus, the identification of a signal transduction pathway could be traced back to the genetic regulatory level. The rapid advances of genome sequencing and DNA microarray technology make possible the quantitative analysis of signaling regulatory network besides the qualitative analysis (Hughes *et al.*, 1999). Furthermore, the embedded time-course feature of microarray data would improve the system analysis of genetic regulatory networks as well.

In addition to northern blots and reverse transcription-polymerase chain reaction (RT-PCR), which study a small number of genes in a single assay, transcriptome analysis (Velculescu *et al.*, 1997) has, via DNA microarray technology, achieved high-throughput monitoring of the almost genome-wide mRNA expression levels in living cells or tissues. Two types of available microarrays, the spotted cDNA (Schena *et al.*, 1995) and *in situ* synthesized oligonucleotide chips (Lipshutz *et al.*, 1999) are used in different experimental requirements and stocked in the databases on net, such as Stanford Microarray Database (SMD) (Sherlock *et al.*, 2001), Gene Expression Omnibus (GEO) (Edgar *et al.*, 2002) in NCBI, and ArrayExpress (Brazma *et al.*, 2003) in EBI. Microarray experiments are now routinely used to collect large-scale time series data that facilitate quantitative genetic regulatory analysis while qualitative discussion is the traditional thinking (Spellman *et al.*, 1998; Harmer *et al.*, 2000; Causton *et al.*, 2001).

Several analytic methods have been proposed to infer genetic interrelations from gene expression data. In the coarse-scale approach of clustering, the underlying conjecture is that the co-expression is indicative of the

co-regulation, thus the clustering methods may identify genes that have similar functions or are involved in the related biological processes (Soukas *et al*., 2001; Gasch and Eisen, 2002; Tanay *et al*., 2005). The most widely used method is the unsupervised hierarchical clusters (Eisen *et al*., 1998). This approach has an increasing number of the nested classes by the similarity measurement and resembles a phylogenetic classification. Other algorithms such as the neural-network-based self-organizing maps (SOM) (Tamayo *et al*., 1999), singular value decomposition (SVD) or principal component analysis (PCA) (Alter *et al*., 2000), and fuzzy clustering methods (Gasch and Eisen, 2002) also have their own advantages and limitations. Alternative supervised clustering algorithm of support vector machine (SVM) (Brown *et al*., 2000), which uses prior biological information of cluster for training, would enhance the performance of clustering. However, the nature of clustering algorithms is gene grouping and could not be easily used to uncover the causal interactions between genes. Regarding the causality of pathways, the clustering analysis needs to cooperate with the sequence motif detection (Tavazoie *et al*., 1999). It is also important to note that models using clustering analysis are static and thus can not describe the dynamic evolution of gene expression, even in the type of time-course microarray data. Time series analysis with state space models in the context of genetic networks has been well used in human T-Cell (Rangel *et al*., 2004; Beal *et al*., 2005) and yeast cell cycle (Wu *et al*., 2005).

Recently, a statistical model of Bayesian network (Friedman *et al*., 2000) was proposed to model genetic regulatory networks. Basically, the technique used a probabilistic score to evaluate the networks with respect to the expression data and searches for the network with the optimal score. An algorithm of Boolean networks (Huang, 1999) was also employed to model the dynamic evolution of gene expression, where the state of a gene can be simplified as either active state (on, 1) or inactive state (off, 0). The probabilistic nature of Bayesian networks is capable of handling noise inherent in both the biological processes and the microarray experiments. This makes Bayesian networks superior to Boolean networks, which are deterministic in nature. A dynamic model based on the first-order differential equation is proposed for yeast cell cycle (Chen *et al*., 2004). In their model, a transcriptional regulation of a target gene is detected for tracing back upstream regulators. Then these upstream regulators are considered as target genes to trace back their upstream regulators to construct their regulatory network iteratively.

In this study, the stochastic system approach was employed to model how a target gene's expression profile is regulated by its upstream regulatory genes from the system causality viewpoint. The AutoRegressive with eXternal input (ARX) model, which has been widely used to model many physical stochastic systems with several good characteristics, is proposed to model the time-profile evolutional behavior of a target gene under interactive regulations and a random noise environment. Using the interactive ARX model and the microarray data, we can identify the circadian regulatory network from the interactive stochastic system viewpoint.

The interactive ARX stochastic system approach is applied to the circadian regulatory pathway of *Arabidopsis thaliana* (Yanovsky and Kay, 2001; Staiger., 2002) with microarray data sets publicly available on the net (Harmer *et al*., 2000; Schaffer *et al*., 2001). The circadian system is an essential signaling network that allows organisms to adjust cellular and physiological processes in anticipation of periodic changes of light either in the normal environment or in the flowering time. A well known signaling pathway in circadian rhythms of *Arabidopsis* is isolated by mutation method either in the normal environment or in the flowering time (Somers *et al*., 1998; Alabadi *et al*., 2001; Schaffer *et al*., 2001; Toth *et al*., 2001; Yanovsky and Kay, 2001; Staiger, 2002; Hayama and Coupland, 2003; Mass *et al*., 2003). 16 gene pathway (shown in Table 1) in circadian pathway are well roughly constructed in *Arabidopsis* (Yanovsky and Kay, 2001; Hayama and Coupland, 2003). In this study, we use the well isolated probably genetic regulation mechanism in these 16 genes as a biological filter in our dynamic modeling to construct the circadian pathway by using time course microarray measured in constant condition in *Arabidopsis* (Harmer *et al*., 2000). According to the synchronously dynamic evolution of microarray data, we have successively identified the core signaling transduction from light receptors of phytochromes (Casal, 2000) and crytochromes (Fankhauser and Staiger, 2002) to the endogenous biological clock (Alabadi *et al*., 2001), which is coupled to control the correlatively physiological activity with paces on a daily basis in our interactive stochastic system model. With the stochastic system approach, not only the regulatory abilities and random noise effect, but also the oscillatory frequency and the delays of regulatory activity were specified. In addition, the robustness (or sensitivity) analysis is important topic to see more insight into system characteristics of the gene regulatory network (Chen *et al*., 2005). However, the robustness of the circadian system is only at the steady state case. In this study, based on the stochastic model constructed executed by microarray data, a sensitivity analysis is developed for different parameter variations from the dynamic system viewpoint. In this situation, the sensitivities of system genes to different perturbation effects such as *Input* light, *Trans* level, and *Cis* level are also deduced. Moreover, we design several simulation assays with the biological senses to mimic the biological experiments. These quantitative characteristics and assays will help investigate the intrinsic connectivity of the circadian regulatory network in *Arabidopsis*, from the stochastic system viewpoint.

Table 1

The Dynamic Equation Set of the Identified Upstream Regulators and their Regulatory Relationships to the Specific System Genes in the Pathway of Circadian Regulatory Network of Arabidopsis Thaliana. Totally 16 System Genes ($X_1 \sim X_{16}$) are Represented with their ARX Dynamic equations. The estimated 1.0-hr Activation Delay of the ARX(1) is also shown for each Upstream Regulator, i.e. $\widetilde{X}_i(k-\tau_i)$ Denotes $X_i(k)$ with the Delay τ_i and through a Sigmoid Function in equation (1.2). $U(k-\tau_{iu})$ Denotes the Input Light with a Delay τ_{iu} Affecting $X_i(k)$. It is seen that $X_1(k)$ to $X_9(k)$ are Directly Affected by Light and $X_{10}(k)$ to $X_{16}(k)$ are not Affected Directly by Light

(1) $X_1(k) = 0.95523 \cdot X_1(k-0.5) + 0.062396 \cdot U(k)$

(2) $X_2(k) = 1.0115 \cdot X_2(k-0.5) + 0.12859 \cdot \widetilde{X}_1(k-0.5) - 0.15496 \cdot \widetilde{X}_4(k-0.5) + 0.026677 \cdot \widetilde{X}_5(k-0.5) + 0.043613 \cdot \widetilde{X}_6(k-0.5) - 0.02881 \cdot \widetilde{X}_7(k-0.5) + 0.069679 \cdot \widetilde{X}_{12}(k-0.5) + 0.038535 \cdot \widetilde{X}_{13}(k-0.5) - 0.07327 \cdot U(k)$

(3) $X_3(k) = 0.96536 \cdot X_3(k-0.5) + 0.038346 \cdot \widetilde{X}_1(k-0.5) - 0.02186 \cdot \widetilde{X}_4(k-0.5) - 0.02185 \cdot \widetilde{X}_5(k-0.5) - 0.03783 \cdot \widetilde{X}_6(k-0.5) - 0.12463 \cdot \widetilde{X}_7(k-0.5) - 0.05918 \cdot \widetilde{X}_{12}(k-0.5) + 0.257 \cdot \widetilde{X}_{13}(k-0.5) + 0.053959 \cdot U(k)$

(4) $X_4(k) = 0.86142 \cdot X_4(k-0.5) + 0.044109 \cdot \widetilde{X}_1(k-0.5) + 0.010912 \cdot \widetilde{X}_2(k-0.5) + 0.033129 \cdot \widetilde{X}_3(k-0.5) - 0.09248 \cdot \widetilde{X}_8(k-0.5) + 0.002625 \cdot \widetilde{X}_9(k-0.5) + 0.065395 \cdot \widetilde{X}_{12}(k-0.5) - 0.08182 \cdot \widetilde{X}_{13}(k-0.5) + 0.17747 \cdot U(k)$

(5) $X_5(k) = 1.0621 \cdot X_5(k-0.5) + 0.067196 \cdot \widetilde{X}_1(k-0.5) + 0.000164 \cdot \widetilde{X}_2(k-0.5) + 0.015546 \cdot \widetilde{X}_3(k-0.5) - 0.02668 \cdot \widetilde{X}_8(k-0.5) - 0.04628 \cdot \widetilde{X}_9(k-0.5) + 0.052805 \cdot \widetilde{X}_{12}(k-0.5) - 0.05938 \cdot \widetilde{X}_{13}(k-0.5) - 0.05647 \cdot U(k)$

(6) $X_6(k) = 0.98671 \cdot X_6(k-0.5) + 0.009117 \cdot \widetilde{X}_1(k-0.5) + 0.017871 \cdot \widetilde{X}_2(k-0.5) - 0.01325 \cdot \widetilde{X}_3(k-0.5) + 0.021482 \cdot \widetilde{X}_8(k-0.5) + 0.026652 \cdot \widetilde{X}_9(k-0.5) + 0.063417 \cdot \widetilde{X}_{12}(k-0.5) - 0.06054 \cdot \widetilde{X}_{13}(k-0.5) - 0.02682 \cdot U(k)$

(7) $X_7(k) = 1.126 \cdot X_7(k-0.5) + 0.033057 \cdot \widetilde{X}_1(k-0.5) + 0.053598 \cdot \widetilde{X}_2(k-0.5) - 0.08615 \cdot \widetilde{X}_3(k-0.5) - 0.16828 \cdot \widetilde{X}_8(k-0.5) - 0.13794 \cdot \widetilde{X}_9(k-0.5) - 0.22659 \cdot \widetilde{X}_{12}(k-0.5) - 0.19554 \cdot \widetilde{X}_{13}(k-0.5) + 0.29971 \cdot U(k)$

(8) $X_8(k) = 0.9932 \cdot X_8(k-0.5) + 0.01061 \cdot U(k)$

(9) $X_9(k) = 0.96214 \cdot X_9(k-0.5) + 0.05387 \cdot U(k)$

(10) $X_{10}(k) = 0.94208 \cdot X_{10}(k-0.5) - 0.00783 \cdot \widetilde{X}_2(k-0.5) + 0.006922 \cdot \widetilde{X}_3(k-0.5) + 0.054391 \cdot \widetilde{X}_4(k-0.5) + 0.012162 \cdot \widetilde{X}_5(k-0.5) - 0.02955 \cdot \widetilde{X}_6(k-0.5) + 0.00604 \cdot \widetilde{X}_7(k-0.5)$

(11) $X_{11}(k) = 0.98314 \cdot X_{11}(k-0.5) + 0.009773 \cdot \widetilde{X}_2(k-0.5) - 0.04943 \cdot \widetilde{X}_3(k-0.5) + 0.048507 \cdot \widetilde{X}_4(k-0.5) + 0.039943 \cdot \widetilde{X}_5(k-0.5) + 0.049841 \cdot \widetilde{X}_{10}(k-0.5) - 0.05489 \cdot \widetilde{X}_{12}(k-0.5) + 0.015015 \cdot \widetilde{X}_{13}(k-0.5)$

(12) $X_{12}(k) = 1.2126 \cdot X_{12}(k-0.5) + 0.018817 \cdot \widetilde{X}_{10}(k-0.5) + 0.07522 \cdot \widetilde{X}_{11}(k-0.5) - 0.25376 \cdot \widetilde{X}_{13}(k-0.5)$

(13) $X_{13}(k) = 0.84627 \cdot X_{13}(k-0.5) - 0.0523 \cdot \widetilde{X}_{10}(k-0.5) + 0.049786 \cdot \widetilde{X}_{11}(k-0.5) + 0.082987 \cdot \widetilde{X}_{12}(k-0.5)$

(14) $X_{14}(k) = 1.0101 \cdot X_{14}(k-0.5) + 0.012403 \cdot \widetilde{X}_{10}(k-0.5) - 0.01484 \cdot \widetilde{X}_{11}(k-0.5) - 0.05309 \cdot \widetilde{X}_{12}(k-0.5) + 0.037975 \cdot \widetilde{X}_{13}(k-0.5)$

(15) $X_{15}(k) = 1.0008 \cdot X_{15}(k-0.5) - 0.05948 \cdot \widetilde{X}_{10}(k-0.5) + 0.008273 \cdot \widetilde{X}_{11}(k-0.5) - 0.11432 \cdot \widetilde{X}_{12}(k-0.5) + 0.097237 \cdot \widetilde{X}_{13}(k-0.5) + 0.053412 \cdot \widetilde{X}_{14}(k-0.5)$

(16) $X_{16}(k) = 1.1657 \cdot X_{16}(k-0.5) + 0.012415 \cdot \widetilde{X}_{10}(k-0.5) + 0.045795 \cdot \widetilde{X}_{11}(k-0.5) - 0.09915 \cdot \widetilde{X}_{12}(k-0.5) - 0.12059 \cdot \widetilde{X}_{13}(k-0.5)$

	X1	X2	X3	X4	X5	X6	X7	X8
System Genes	Fkf1	Cry1	Cry2	PhyA	PhyB	PhyD	PhyE	E1f3
	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}
	Gi	Pif3	Tocl	•Lhy	Ccal	Papl	Chs	Co

Table 2

The Assays in Silico of the Circadian Regulatory Network in Arabidopsis

Type of assay	*Δ Parameters*	*Sensitivity computation*	*Simulation assays*
Input Light	Light fluences (Amplitude)	150%(+50%) 50% (–50%)	+50% –50%
Trnas level	Trans-sensitivity rate (γ)	+100% –100%	Average of +100% and –100%
	Trans-expression threshold (M)	+100% –100%	Average of +100% and –100%
Cis level	The genetic kinetic parameter $d_{i,N}$, i = 1,2,...16 for the specific gene j	+1.0 and –1.0 of initial gene expression	Average from +100% to –100% with 5% interval $d_{i,N} = 0$, i = 1,2,...,16 for the specific gene j

2. STOCHASTIC SYSTEM MODEL AND IDENTIFICATION METHOD

The proposed circadian regulatory network in this study would be divided into two steps. In the first step, a stochastic system is developed from the interactive ARX model with nonlinear sigmoid activation to describe the expression profile data as output and the upstream genetic signals as input to denote the implicit characteristics of each gene with some parameters. With the help of the optimal estimation method, we can identify the parametric structure of the ARX

interactive stochastic model, which reveals the interactive relationships in a network. After modeling the circadian genetic regulatory system, we then perform the system sensitivity analysis to assess robustness of the circadian network from three aspects of biological perturbations, including the *Input*, *Trans*, and *Cis* levels, *in silico* as the mimic biological assays *in vivo* and *in vitro*. We will unravel the molecular mechanism of the circadian network from the stochastic system viewpoint.

3. STOCHASTIC SYSTEM DESCRIPTION OF CIRCADIAN REGULATORY MODEL

An ARX model is well used in the description of stochastic system evolved from the ontology of causality. However, it is only used to model a stochastic system without interactions with others. We can consider any gene expression profile as a system response or output stimulated by some inputs from other gene expressions and environmental stimuli. Therefore, an interactive ARX model is employed to describe a gene expression through interactive regulations among genes in a circadian regulatory network. According to this description, let $X_i(k)$ denote the expression profile of the i-th gene at time point k. Then the following nonlinear ARX interactive equations are proposed to model the expression level of the i-th gene as the synthesis of n upstream regulatory signals $\widetilde{X}_i(k)$, i = 1, 2, ..., n and an external input light signal u under their τ delays, (see Figure 1)

$$\begin{aligned} X_i(k) = {} & d_{1,i1}\widetilde{X}_1(k-\tau_i) + d_{1,i2}\widetilde{X}_2(k-\tau_i) + \cdots + \\ & d_{1,ii}X_i(k-\tau_i) + \cdots + d_{1,in}\widetilde{X}_n(k-\tau_i) + \\ & d_{2,i1}\widetilde{X}_1(k-2\tau_i) + d_{2,i2}\widetilde{X}_2(k-2\tau_i) + \cdots + \\ & d_{2,ii}X_i(k-2\tau_i) + \cdots + d_{2,in}\widetilde{X}_n(k-2\tau_i) + \\ & d_{q,i1}\widetilde{X}_1(k-Q\tau_i) + d_{q,i2}\widetilde{X}_2(k-Q\tau_i) + \cdots + \\ & d_{q,ii}X_i(k-Q\tau_i) + \cdots + d_{q,in}\widetilde{X}_n(k-Q\tau_i) + \\ & b_i \cdot u(k-\tau_{iu}) + \varepsilon_i(k), \quad i = 1, 2, \cdots, n \end{aligned} \tag{1.1}$$

where $\widetilde{X}_j(k-q\cdot\tau_i)$, $j = 1, 2, \cdots, n$; $q = 1, 2, \cdots, Q$ are the upstream interactive signals transformed by $X_j(k)$ with the q-th order of τ_i delay and through a nonlinear sigmoid activation function to denote the binding of transcription factor on gene i, and the genetic kinetic parameter $d_{q,ij}$ ($i \neq j$) denotes the regulation abilities of transcription factor $\widetilde{X}_j(k)$ on gene i. Meanwhile, $u(k-\tau_{iu})$, which denotes the external input light with a delay τ_{iu} and correlates with the output genetic expression $X_i(k)$ with the input kinetic parameters b_i. $\varepsilon_i(k)$ is the stochastic noise of current microarray data or the residue of the model. Here τ_i and τ_{iu}, which are essential to the activation-time estimation, should be determined previously and will be discussed later. Oscillations exist in a circadian regulatory network through the interactions with other genes if these interactions are limited by nonlinear sigmoid functions to avoid their unstable propagations, which will be discussed by the describing function method of nonlinear limit cycles in control theory (Slotine and Li, 1991) in the sequel.

It should be noted that with the combination of biological knowledge about the transcription factors, protein phosphorylation, and post-transcriptional and specific

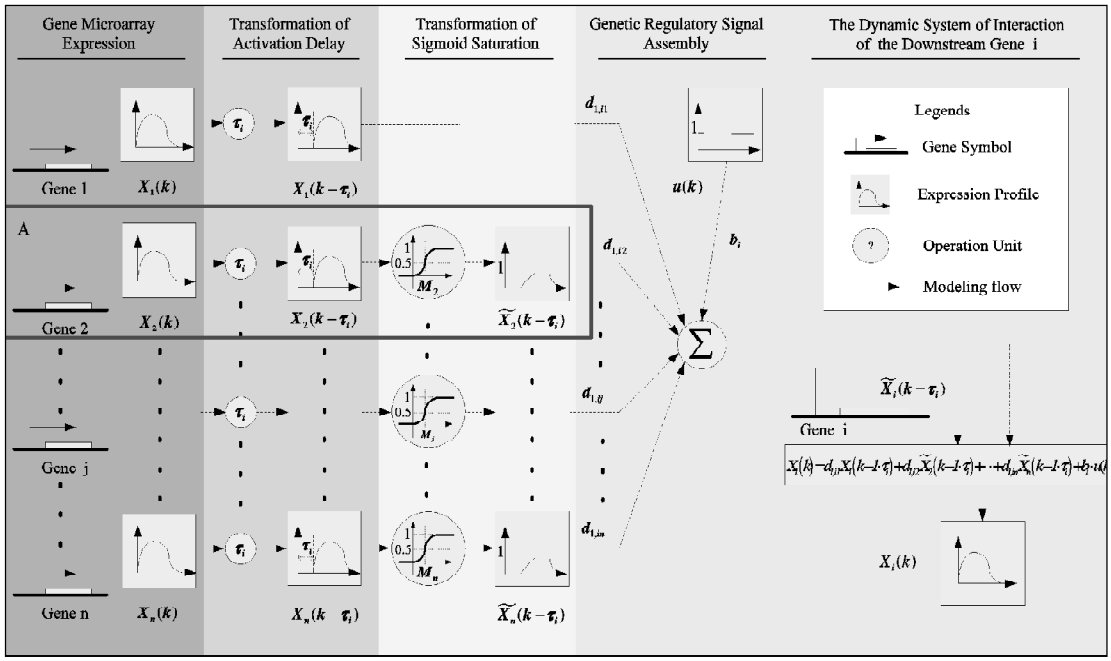

Figure 1: Illustration of the Dynamic System Scheme using the ARX(1) Model. Block A Represents the Transformation of the Genetic Regulatory Signal, $\widetilde{X}_j(k-q\tau_i)$, for $j = 2$ and $q = 1$

enzyme regulations, lots of verified regulatory genes X_j correlated with the target genes X_i via this biological filter are considered to determine whether the kinetic parameters $d_{q,ij}$, $q = 1, 2, ..., Q$; $n = 1, 2, ...$ and b_i should be set to zero previously without estimation. In this way, we determine all the kinetic parameters of the filtered genes that are possible correlated with the output gene biologically.

According to the biological or biochemical principle, the genetic interactions such as transcriptional binding and protein activation start on a threshold of the expression level. Therefore, it is reasonable to confine the effect from the upstream regulation X_j to X_i, which is why we induce the upstream genetic regulatory signal X_j through a sigmoid activation function to be $\widetilde{X_j}$.

For the limited influence expression of $X_j(k - q.\tau_i)$ (see Block A in Figure 1), the sigmoid function is chosen to express the nonlinear 'on' or 'off' activities of physically genetic interactions with parameters $\theta_j = \{\gamma, M_j, \tau_i\}$ as follows,

$$\widetilde{X_j}(k - q\cdot\tau_i) = \frac{1}{1+e^{-\gamma_j(X_j(k-q\cdot\tau_i)-M_j)}} \quad (1.2)$$

Where γ_j is the trans-sensitivity rate, and M_j is the trans-expression threshold derived from the mean of the *j*-th gene's profile. γ_j could determine the transition time of activation between the states of 'off' or 'on' from X_j to X_i, for which a larger γ_j is with a less transition time, to mimic the transient state of the genetic interaction on the *trans* level. M_j can determine the threshold of the half activation level of X_j to X_i, for which a larger M_j is with a less activating ability, to mimic the steady state of the genetic interaction on the *trans* level.

In this study, we use the mRNA expression data of 8200 genes measured in the replicate hybridization of 12 samples harvested every 4 hours over 2 days. The system time delay $q.\tau_i$ we choose must be small than 4 hours. In addition, for the biological reason of small activation delay on mRNA level and less modeling complexity, we can reduce the order of the ARX model to no more than 2, $Q = 1$ (i.e. ARX(1)) or $Q = 2$ (i.e. ARX(2)) in equation (1.1). We will determine an adequate order and delay for our interesting system later. And now we take the second order ARX nonlinear model for illustration as follows,

$$\begin{aligned} X_i(k) = {} & d_{1,i1}\widetilde{X_1}(k-\tau_i) + d_{1,i2}\widetilde{X_2}(k-\tau_i) + \cdots + \\ & d_{1,ii}X_i(k-\tau_i) + \cdots + d_{1,in}\widetilde{X_n}(k-\tau_i) + \\ & d_{2,i1}\widetilde{X_1}(k-2\tau_i) + d_{2,i2}\widetilde{X_2}(k-2\tau_i) + \cdots + \\ & d_{2,ii}X_i(k-2\tau_i) + \cdots + d_{2,in}\widetilde{X_n}(k-2\tau_i) + \\ & b_i\cdot u(k) + \varepsilon_i(k),\ i = 1, 2, \cdots, n \end{aligned} \quad (1.3)$$

Through equation (1.3) is a second order ARX stochastic system, due to several nonlinear feedbacks through other genes, all of the circadian regulatory network could be equivalent to a very high order nonlinear difference equation which maybe exist several oscillations in the circadian regulatory network.

To make the stochastic model effective, the stochastic dynamic equation in equation (1.3) should meet the expression profile at all time points $k = k_1, k_2, ..., k_m$ and is then arranged in a vector difference form. Consequently, the vector underlined in this equation is applied to *m* time points in order.

$$\begin{aligned} \underline{X_i} = {} & d_{1,i1}\underline{\widetilde{X_{1,\tau_i}}} + \cdots + d_{1,ii}\underline{X_{i,\tau_i}} + \cdots + d_{1,in}\underline{\widetilde{X_{n,\tau_i}}} + \\ & d_{2,i1}\underline{\widetilde{X_{1,2\tau_i}}} + \cdots + d_{2,ii}\underline{X_{i,2\tau_i}} + \cdots + d_{2,in}\underline{\widetilde{X_{n,2\tau_i}}} + b_i\cdot\underline{u} + \underline{\varepsilon_i} \end{aligned} \quad (1.4)$$

where

$$\underline{X_i} = \begin{bmatrix} X_i(k_1) \\ X_i(k_2) \\ \vdots \\ X_i(k_m) \end{bmatrix},\ \underline{\widetilde{X_{n,q}}} = \begin{bmatrix} \widetilde{X_n}(k_1-q) \\ \widetilde{X_n}(k_2-q) \\ \vdots \\ \widetilde{X_n}(k_m-q) \end{bmatrix},\ \underline{X_{n,q}} = \begin{bmatrix} X_n(k_1-q) \\ X_n(k_2-q) \\ \vdots \\ X_n(k_m-q) \end{bmatrix},$$

$$\underline{u} = \begin{bmatrix} u(k_1) \\ u(k_2) \\ \vdots \\ u(k_m) \end{bmatrix},\ \underline{\varepsilon_i} = \begin{bmatrix} \varepsilon_i(k_1) \\ \varepsilon_i(k_2) \\ \vdots \\ \varepsilon_i(k_m) \end{bmatrix}, \text{ and } m \text{ denotes the number of}$$

time points. τ_i is the specific activation delay.

In the next step, to estimate the kinetic parameters $d_{q,in}$, $q = 1, 2$; $n = 1, 2, ...$ and b_i, the formula equation (1.4) should be translated into the difference matrix equation as follows,

$$Y_i = A_i\Omega_i + E_i,\ i = 1, 2, ... n \quad (1.5)$$

where $Y_i = \underline{X_i}$, $\Omega_i = \left[d_{1,i1} ... d_{1,in}\quad d_{2,i1} ... d_{2,in}\quad b_i\right]^T$ and $E_i = \underline{\varepsilon_i}$ are in vector forms, while

$A = \left[\underline{\tilde{X}}_{1,\tau_i}\ \ ...\ \ \underline{\tilde{X}}_{n,\tau_i}\ \ \underline{\tilde{X}}_{1,2\tau_i}\ \ ...\ \ \underline{\tilde{X}}_{n,2\tau_i}\ \ \underline{u}_{\tau_{iu}}\right]$ is a matrix.

We assume that each element in the stochastic noise vector, $\varepsilon_i(k_l)$, $l = \{1, ..., m\}$, is an independent random variable with a normal distribution with zero mean and variance σ_i^2, which is unknown and needs to be estimated. Thus, we estimate the parameter $\widehat{\Omega_i}$ using the maximum likelihood method. The likelihood function of Y_i is defined as

$$p(Y_i|\Omega_i, \sigma_i^2) = \frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left\{-\frac{\left[Y_i - A_i\Omega_i\right]^T\left[Y_i - A_i\Omega_i\right]}{2\sigma_i^2}\right\} \quad (1.6)$$

The log-likelihood function for given *m* data points is then described by

$$L(\Omega_i, \sigma_i^2) = -\frac{m}{2}\ln[2\pi\sigma_i^2] - \frac{1}{2\sigma_i^2}\sum_{i=1}^{m}\left[Y_i - A_i\Omega_i\right]^T\left[Y_i - A_i\Omega_i\right] \quad (1.7)$$

The necessary condition for the maximum likelihood estimation of σ_i^2 and Ω_i is to find σ_i^2 and Ω_i to maximize $L(\Omega_i, \sigma_i^2)$. We can obtain the maximum likelihood estimate of σ_i^2 and Ω_i by $\frac{\partial L(\Omega, \sigma^2)}{\partial \sigma^2} = 0$ and, $\frac{\partial L(\Omega, \sigma^2)}{\partial \Omega} = 0$ which are solved as (Johansson, 1993)

$$\hat{\sigma}_i^2 = \frac{1}{m}\sum_{l=1}^{m}\left[Y_i - A_i\hat{\Omega}_i\right]^T \left[Y_i - A_i\hat{\Omega}_i\right] \tag{1.8}$$

$$\hat{\Omega}_i = (A_i^T A_i)^{-1} A_i^T Y_i, \quad i = 1, 2, \ldots n \tag{1.9}$$

We could solve Ω_i first and then substitute $\widehat{\Omega_i}$ into equation (1.8) to find $\hat{\sigma}_i^2$. Here, the modeling error could be concluded into E_i as the noise of the gene-expression profile or the microarray chips. So the consideration of modeling error in equation (1.5) approaches more the reality. We also illustrate the stochastic system approach using a stochastic model ARX interactive model in Figure 1. After the parameter estimation in equation (1.9), substituting $\widehat{\Omega_i}$ in equation (1.9) into stochastic model in equation (1.3) lead to the estimated circadian regulatory network equations in Table 1.

4. BIOLOGICAL ASSAY OF ARX SYSTEM MODEL

The biological assays of the ARX system model are divided into four categories. The first is the confirmation of the oscillation frequency of circadian regulatory network by the oscillatory characteristics of the stochastic circadian regulatory model. The second is the sensitivity analysis with respect to input stimulus changes, the third is the sensitivity analysis under *trans* disturbance, and the last is the sensitivity analysis about the *cis* perturbation. Before these assays, it is necessary to define some measure indexes to evaluate the system characteristics. We considered three most essential features of circadian regulatory networks, i.e. similarity, period, and mean expression. For the measure of similarity, the Pearson correlation coefficient is a widely used similarity metric. For biological assay of the ARX model, we calculated the Pearson correlation coefficient between the genes' mRNA expression profiles of $X_i(k)$ *in vivo* and $\widehat{X_i}(k)$ *in silico* at all time points $k = k_1, k_2, \ldots, k_m$ as follows.

$$r\left(X_i, \widehat{X_i}\right) = \frac{\sum_{l=1}^{m} X_i(k_l)\widehat{X_i}(k_l) - \left(\frac{\sum_{l=1}^{m} X_i(k_l)\sum_{l=1}^{m}\widehat{X_i}(k_l)}{m}\right)}{\sqrt{\left[\sum_{l=1}^{m} X_i^{\,2}(k_l) - \frac{\left[\sum_{l=1}^{m} X_i(k_l)\right]^2}{m}\right]\left[\sum_{l=1}^{m}\widehat{X_i}^{\,2}(k_l) - \frac{\left[\sum_{l=1}^{m}\widehat{X_i}(k_l)\right]^2}{m}\right]}} \tag{2.1}$$

To measure the period of the time-course expression profile, the power spectrum, which has different magnitudes in different frequencies (the reciprocal of periods), is employed to detect which frequency has the largest magnitude. First, we should take the Discrete Fourier Transform of $X_i(k)$ for $k = k_1, k_2, \ldots, k_m$ as follows,

$$X_i(\omega) = \sum_{l=1}^{m} X_i(k_l)\cdot e^{-j\omega k} \tag{2.2}$$

where ω is the radian frequency.

Then we detect the frequency with the maximum magnitude,

$$\omega_i = \arg\max_{\omega}\left|X_i(\omega)\right| = \frac{2\pi}{T_i} \tag{2.3}$$

where T_i is the period of $X_i(k)$ and can be determined from the reciprocal of the detected frequency ω_i. The oscillation frequency ω_i detected by equation (2.3) will be used to assess the oscillation frequency of our stochastic model of the network in the sequel. Furthermore, the measure of mean expression of $X_i(k)$ is important for distinguishing the deviation of expression profile under different assays as follows,

$$M_i = \frac{1}{m}\sum_{l=1}^{m} X_i(k_l) \tag{2.4}$$

5. DETERMINATION OF SYSTEM ORDER AND ACTIVATION DELAY OF CIRCADIAN SYSTEM

In this study, the formulated ARX model should be first assigned with a proper modeling order and an activation delay to analyze the experimental expression data of microarray. According to equation (1.1), we compared the first-order ($Q = 1$) ARX model (i.e. ARX(1)) and the second-order ($Q = 2$) ARX model (i.e. ARX(2)) with different activation delays τ as shown in Figure 2a. We exploited the mean similarity between the raw expression and the simulation of all 16 system genes we concerned in the circadian network of *Arabidopsis thaliana*, which is measured by Pearson correlations, to evaluate the performance of the network model. Owing to the least difference at 0.5-hr delay between ARX(1) and ARX(2), we would prefer the more flexible ARX(1) model with a 0.5-hr activation delay as the system model for the circadian regulatory network. Consequently, the simulation expressions of the derived circadian network model are shown in Figure 2b and exhibit agreeable dynamic data fitting, thus creating a basis for the following system analysis.

On the other hand, the quantitatively regulatory abilities among the system genes in the circadian system can be estimated as the genetic and input kinetic parameters of the system model in equation (1.3). The kinetic parameters of ARX(1) are displayed in Table 1, where the positive value means activation and the negative one means inhibition. Although the upstream genes are selected according to the biological knowledge of relevance such as transcriptional

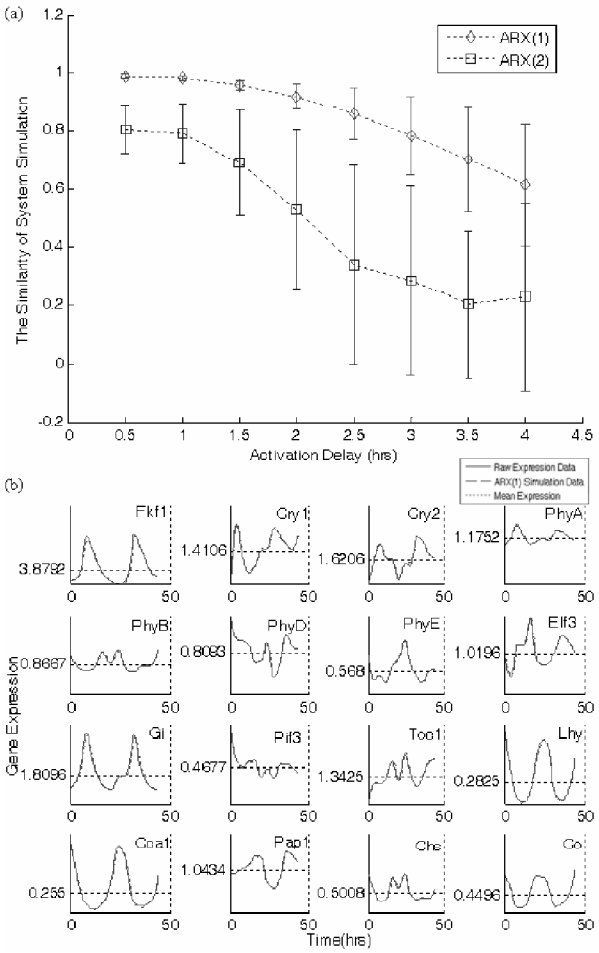

Figure 2: ARX System Modeling with Determination of System Modeling Order and Activation Delay. **(a)** The Average Similarity (Measured by Pearson Correlation) of all System Genes under different Activation Delays. Therefore, ARX(1) Model with Delay τ = 0.5 hr is an Adequate Model with Small Order and Less Delay. **(b)** The Dynamic Data Fitting of 16 Genes in the Circadian Network with ARX(1) Model and 0.5-hr Activation Delay

binding, phosphorylation, and complexes, this quantitative influence could reveal the significance of upstream genes like the ARX dynamic equation set in Table 1. For instance, the mutual interactions of Cry1 [X_2] and PhyA [X_4] have their basis on phosphorylation (Ahmad *et al.*, 1998) (see Eq. (2) and Eq. (4) in Table 1), and we identify that Cry1 has the strongest positive influence on PhyA with activation delay. Meanwhile, Cry1 and PhyA are significantly regulated by Lhy [X_{12}] and Cca1 [X_{13}] known as biological clock genes, which are negative regulated reciprocal, as shown in Eq. (2) and Eq. (3) of Table 1, to highlight their feedback regulatory roles on crytochrome and phytochrome, respectively (also see Eq. (2) and Eq. (4) in Table 1). In this way, we can even recognize the regulatory ability from the input-light signal. In particular, Fkf1 [X_1] that has no regulated by genes other than light stimulus (see Eq. (1) in Table 1). The detection of the static structural characteristics will help construct their hidden significance of cis connectivity as in the signaling transduction network of Figure 3.

5. Sensitivity Analysis of Circadian System

The sensitivity measure of the circadian system for the analysis of robustness can also be derived from the whole system model, i.e., we should integrate *n* ARX models together. For illustration, we would rearrange equation (1.3) into the following difference matrix equation by which its sensitivity will be investigated later,

$$\underline{Y}(k) = D_1\tilde{\underline{Y}}(k-\tau) + D_2\underline{Y}(k-\tau) + Bu(k) \qquad (3.1)$$

where

$$D_1 = \begin{bmatrix} 0 & d_{1,12} & \cdots & \cdots & d_{1,1n} \\ d_{1,21} & 0 & d_{1,23} & \cdots & \vdots \\ \vdots & d_{1,32} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & d_{1,(n-1)n} \\ d_{1,n1} & \cdots & \cdots & d_{1,n(n-1)} & 0 \end{bmatrix},$$

$$\underline{Y}(k) = \begin{bmatrix} X_1(k) \\ X_2(k) \\ \vdots \\ X_n(k) \end{bmatrix}, \quad \underline{Y}(k-\tau) = \begin{bmatrix} X_1(k-\tau_1) \\ X_2(k-\tau_2) \\ \vdots \\ X_n(k-\tau_n) \end{bmatrix}$$

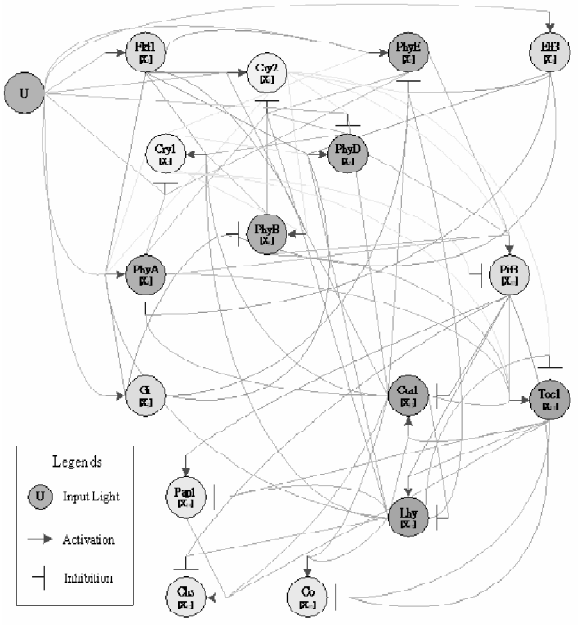

Figure 3: Signaling Transduction Network of System Genes and Input Light in the Circadian Network of *Arabidopsis*. The Colored Circles Indicate the System Genes with their Names and Notations of $X_1 \sim X_{16}$

$$\underline{\tilde{Y}}(k-\tau)=\begin{bmatrix}\widetilde{X_1}(k-\tau_1)\\ \widetilde{X_2}(k-\tau_2)\\ \vdots\\ \widetilde{X_i}(k-\tau_i)\\ \vdots\\ \widetilde{X_n}(k-\tau_n)\end{bmatrix}, B=\begin{bmatrix}b_1\\ b_2\\ \vdots\\ b_n\end{bmatrix}, \underline{U}=\begin{bmatrix}u(k)\\ u(k)\\ \vdots\\ u(k)\end{bmatrix},$$

and n is the number of genes.

A. *Circadian Clock Frequency Assay*

By the existence condition of limit cycle (oscillation) in an interactive stochastic dynamic system with sigmoid (or saturation) interactive feedbacks, we will compare the limit cycles of the interactive stochastic system with the oscillation frequencies calculated by equation (2.2) and (2.3) by Fourier transform of the raw gene expressions to validate the accuracy of the proposed dynamic model in the sequel.

Using the ARX system model as equation (1.3) or equation (3.1) and the definition of measure indexes, our *in silico* biological assays of the circadian regulatory system in *Arabidopsis thaliana* are described in the following. A dynamic system with saturation (or sigmoid function) nonlinear feedback will lead to oscillation (limit cycle) (Slotine and Li, 1991). This oscillation phenomenon can be interpreted by the theory of the describing function, which has been widely used to interpret oscillation in nonlinear saturation feedback systems and will be used to describe the circadian regulatory network of *Arabidopsis thaliana*. According to equation (3.1), we get

$$(I-z^{-\tau}D_2)\underline{Y}(k)=D_1\underline{\tilde{Y}}(k-\tau)+Bu(k) \tag{3.2}$$

where $z^{-\tau}=\begin{bmatrix}z^{-\tau_1} & 0 & \cdots & 0\\ 0 & z^{-\tau_2} & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & z^{-\tau_N}\end{bmatrix}$, and $z^{-\tau_i}$ denotes the delay operator of τ_i

$$\underline{Y}(k)=(I-z^{-\tau}D_2)^{-1}D_1\underline{\tilde{Y}}(k-\tau)+(I-z^{-\tau}D_2)^{-1}Bu(k) \tag{3.3}$$

If the oscillation (limit cycle) occurs in circadian network, then the sigmoid function, $\tilde{Y}(k)$, in equation (1.2) can be approximated by the describing function N(A) as (Slotine and Li, 1991)

$$\underline{\tilde{Y}}(k)=\begin{bmatrix}\tilde{Y}_1(k)\\ \tilde{Y}_2(k)\\ \vdots\\ \tilde{Y}_n(k)\end{bmatrix}\approx N(A)\underline{Y}(k) \tag{3.4}$$

where the describing function matrix

$$N(A)=\begin{bmatrix}N_1(A_1) & 0 & \cdots & 0\\ 0 & N_2(A_2) & 0 & \vdots\\ \vdots & 0 & \ddots & 0\\ 0 & \cdots & 0 & N_n(A_n)\end{bmatrix}, \text{ and}$$

$N_i(A_i)$ denotes the describing function of the i-th gene of oscillation and A_i denotes the amplitude of oscillation of the i-th gene. If a gene j is free of oscillation, then the corresponding $N_j(A_j)$ = 0. From (3.3) and (3.4), we can approximate the circadian network as

$$\underline{Y}(k)=(I-z^{-\tau}D_2)^{-1}D_1N(A)z^{-\tau}\underline{Y}(k)+(I-z^{-\tau}D_2)^{-1}Bu(k) \tag{3.5}$$

There are two rhythms, one is circadian rhythm and another is diurnal rhythm. The first term with gain equal to 1 on the right hand side of equation (3.5) responds for circadian rhythm; and the second term for diurnal rhythm, which is controlled by diurnal cycling of light and dark $u(k)$. Some photoreceptor genes are of second case. Since the oscillation exists in the circadian network, by control theory (Slotine and Li, 1991), the closed loop gain should be lossless during these periods in order to support the oscillation, i.e.

$$(I-z^{-\tau}D_2)^{-1}D_1N(A)z^{-\tau}=I \tag{3.6}$$

or

$$D_1N(A)=z^{\tau}-D_2 \tag{3.7}$$

At frequency domain, we can get

$$e^{jw\tau}-D_2=D_1N(A) \tag{3.8}$$

where $e^{jw\tau}=\begin{bmatrix}e^{-jw_1\tau_1} & 0 & \cdots & 0\\ 0 & e^{-jw_2\tau_2} & 0 & \vdots\\ \vdots & 0 & \ddots & 0\\ 0 & \cdots & 0 & e^{-jw_n\tau_n}\end{bmatrix}$

and w_i is the oscillation frequency of the i-th gene. By the equality of the diagonal terms for each gene, we can get

$$e^{jw_i\tau_i}-d_{2,ii}=\sum_{j\neq i}^{n}d_{1,ij}N_j(A_j) \tag{3.9}$$

By the above describing analysis of nonlinear oscillation (Slotine and Li, 1991), the intersection point of the $e^{jw_i\tau_i}-d_{2,ii}$ and $\sum_{j\neq i}^{n}d_{1,ij}N_j(A_j)$ in equation (3.9) implies the occurrence of oscillation with amplitude A_i and frequency w_i in the i-th gene. If no intersection exists in equation (3.9), there is no oscillation in the i-th gene. For example, for gene

PhyE, we get its oscillation frequency 0.45 from equation (3.9) which approximates the measured oscillation frequency 0.52 by discrete Fourier transform in equation (2.3). For gene Lhy, we get its oscillation frequency 0.21 from equation (3.9), which is near to the oscillation frequency 0.2 by discrete Fourier transform in equation (2.3). These estimation errors in oscillation frequency may be due to the fact that describing function is only a linear approximation of nonlinear operator and the discrete Fourier transform $X_i(w)$ in equation (2.3) has some error because the number of data points is small and the steady state is not achieved yet. However, the describing function method will provide an estimation method to roughly assess the existence of oscillations and their frequencies in nonlinear interactive network.

The physical meaning of equation (3.6) is that if a sigmoid function can limit the increase of feedback interactions such that the loop gain in the circadian regulatory network is equal to 1 to supply the loss of the network due to the degradation of proteins, and then the circadian network will continue to oscillation. The biological meaning is that the concentrations of mRNA of genes in circadian network will continue to oscillate if they are activated to enough supply their degradations by the feedback regulations through the other genes in the network.

B. Input-light Perturbation Assay

In general, the circadian regulatory network is independent of the external light except the photoreceptor, i.e. the expression profiles of genes in the circadian network are less sensitive to the variation of external light except the genes with the role of photoreceptor. Therefore, based on the dynamic model of circadian network, a sensitivity assay to input-light perturbation is discussed in the following. The input signal is the white light containing versatile wave lengths from red light to blue light. The original constant light profile starts at 8:00 A.M. in the morning, and ends at 4:00 A.M. after 44 hours (Harmer *et al.*, 2000). Thus, we assume that the illumination of light as value 1 and the dark state as value 0 to mimic the light profile. The system sensitivity with respect to the input light can be derived from equation (3.1) as follows,

$$\frac{\Delta \underline{Y}}{\Delta \underline{U}} \simeq D_1 \frac{\Delta \underline{\tilde{Y}}}{\Delta \underline{Y}} \cdot \frac{\Delta \underline{Y}}{\Delta \underline{U}} + D_2 \frac{\Delta \underline{Y}}{\Delta \underline{U}} + B \tag{3.10}$$

where Δ means the perturbations. From equation (3.10), we can formulate the sensitivity from the ΔU to ΔY as follows,

$$\frac{\Delta \underline{Y}}{\Delta \underline{U}} \simeq \left(I - D_2 - D_1 \frac{\Delta \underline{\tilde{Y}}}{\Delta \underline{Y}} \right)^{-1} B \tag{3.11}$$

where I is the identity matrix. Then we can measure the sensitivity of the circadian regulatory network with respect to the input-light stimulus. To demonstrate the validity of equation (3.11), we compare the realistic output perturbation ΔY with the estimated output perturbation $B\left(I - D_2 - D_1 \frac{\Delta \underline{\tilde{Y}}}{\Delta \underline{Y}} \right)^{-1} B\Delta u$ to confirm the validation of the network dynamic model as in Figure 4, which represents the system response to the perturbation of the environment. The realistic ΔY for each system gene is close to the computed values so that the sensitivity equation like equation (3.11) could well uncover the sensitivity of the system characteristics. Furthermore, we manipulate a simulation as follows.

Amplitude Simulation of Light: We change the amplitude of light from 100% to 150% (+ 50%) and 50% (– 50%) respectively as input to the circadian system and derive the average measure indexes with the sensitivity for each system gene, which are shown in Figures 5, 6 and Table 3A. Further discussion will be given in the next section.

C. Trans-perturbation Assay

Owing to the saturation of activity of genetic interactions, we employ the sigmoid function to characterize the *trans* expression of upstream genes in equation (1.2). As in the description of equation (1.2), γ is the *trans*-sensitivity rate which is related to the transition time of *trans*-activation and M_j is the *trans*-expression threshold that determines the saturating transformation level of expression. We also induce the corresponding sensitivity in the following,

$$\frac{\Delta \underline{Y}}{\Delta \underline{F}} \simeq D_1 \frac{\Delta \underline{\tilde{Y}}}{\Delta \underline{F}} \tag{3.12}$$

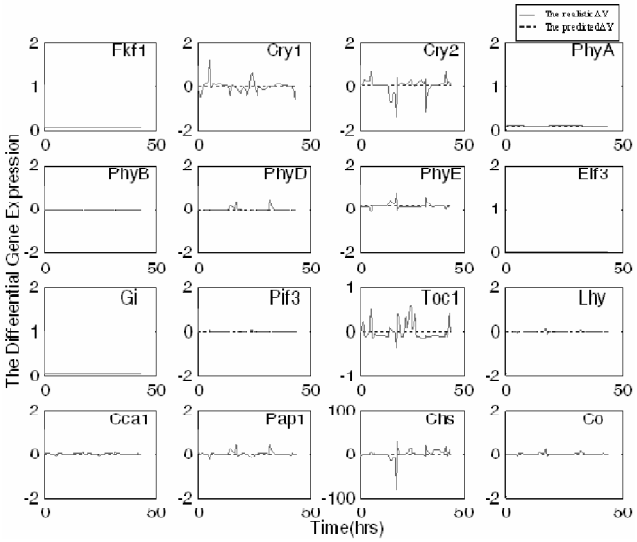

Figure 4: Comparison of the Predicted output Perturbation of $\Delta \underline{Y}$ with the Realistic One to Validate the System in Response to the Environmental Perturbations

Table 3

The Sensitivities of the System Genes in the Circadian Regulatory Network Under Different Perturbations Due to the Input Light, Trans Level, and cis Level. The Bold Values Represent Significant Sensitivities (Less Robustness). In General, Small Sensitivities Implies that the Circadian Regulatory Network is Robust Enough

Perturbation	*Type of Sensitivity*	*Spefification*	*The Sensitivities of the system genes in the circadian regulatory network*															
			Fkfl	*Cryl*	*Cry2*	*PhyA*	*PhyB*	*PhyD*	*PhyE*	*Elf3*	*Gi*	*Pif3*	*Tocl*	*Lhy*	*Ccal*	*Papl*	*Chs*	*Co*
Input Light	A. Fluence rate (Amplitude)	150%	1.3892	3.999	2.442	0.7436	1.2909	0.57575	–0.661	1.5543	1.418	0.4503	4.218	–0.051	–0.118	–0.104	51.74	0.0282
		50%	1.3892	3.999	2.442	0.7436	1.2909	0.57575	–0.661	1.5543	1.418	0.4503	4.218	–0.051	–0.118	–0.104	51.74	0.0282
		+0	0.0112	0.017	0.0119	0.008	0.0104	0.0129	–0.008	0.0112	0.0112	0.0113	0.013	0.0059	0.0106	0.0116	0.01	0.0069
	B. Trans-sensitivity rate (γ)	–100%	0.0112	0.017	0.011	0.0096	0.0112	0.0128	–0.008	0.0112	0.0112	0.0115	0.013	0.0055	0.011	0.0115	0.0096	0.0062
		+100%	0.0112	0.017	0.0121	0.0073	0.0102	0.013	–0.008	0.0112	0.0112	0.0114	0.013	0.006	0.0107	0.0116	0.01	0.0069
Trans level		Mean	0.0112	0.017	0.0117	0.0083	0.0106	0.0129	–0.008	0.0112	0.0112	0.0114	0.013	0.006	0.0107	0.0116	0.01	0.0069
	C. Trans-expression level (M)	+0	0.0112	–0.068	0.0088	0.0474	0.026	0.00403	0.267	0.0112	0.0112	0.0236	–0.031	0.122	0.0116	0.0074	0.0228	0.105
		–100%	0.0112	0.01	–0.007	0.06	0.0301	–0.0003	0.272	0.0112	0.0112	–0.008	–0.031	0.11	0.0011	0.0104	–0.002	0.095
		+100%	0.0112	–0.047	–0.043	0.033	0.0211	0.00812	0.233	0.0112	0.0112	0.0156	0.0002	0.124	–0.002	0.0106	0.0139	0.107
		Mean	0.0112	–0.035	–0.014	0.0468	0.0257	0.00394	0.257	0.0112	0.0112	0.0103	–0.021	0.119	0.0037	0.0095	0.0116	0.102
Cis level	D. The initial value of gene expression (Y_0)	+0	0.0726	0.2039	0.0396	0.0874	0.9874	1.8195	0.4954	0.5974	0.1603	1.364	0.0944	0.9166	0.8634	1.313	0.5527	0.7935
		–10	0.0726	0.2039	0.0394	0.0874	0.9874	1.8195	0.4955	0.5974	0.1603	1.364	0.0944	0.9161	0.8628	1.313	0.5526	0.7936
		+10	0.0726	0.2039	0.0394	0.088	0.9881	1.8196	0.4956	0.5983	0.1604	1.364	0.0945	0.9161	0.8628	1.316	0.5528	0.7936
		Mean	0.0726	0.2039	0.0395	0.0876	0.9876	1.8195	0.4955	0.5977	0.1603	1.364	0.0945	0.9162	0.863	1.314	0.5527	0.7935

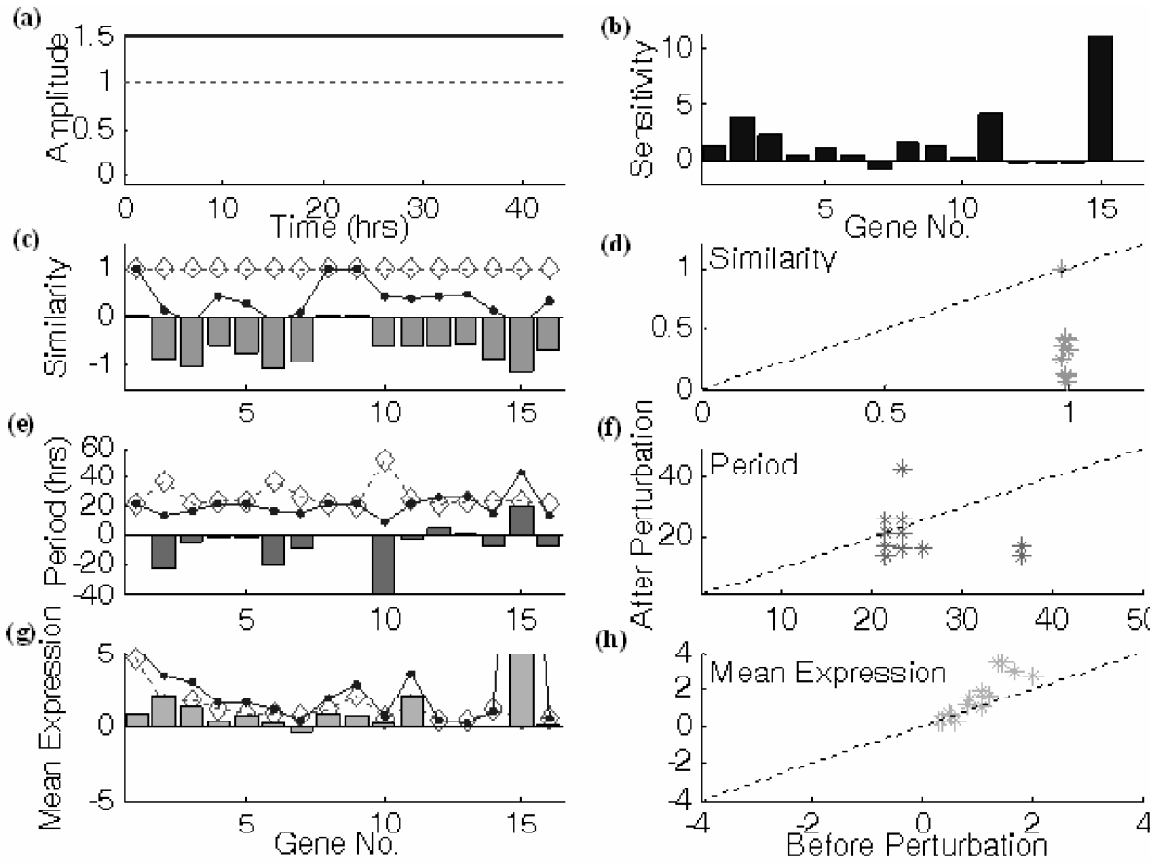

Figure 5: Amplified Perturbation Assay of the Input-light Fluence Rate. **(a)** The Illustration of this Assay with the Change of the White-light Fluence from 100% (Red Dashed Line) to 150% (Blue Solid Line) with Respect to the Original Amount. **(b)** The Measure of Sensitivity for the System Genes Due to this Perturbation. **(c), (e)** and **(g)** Indicate the Three System Measures of Similarity (Pearson Correlation), Period, and Mean Expression, Respectively. The Red Diamonds are Measures of the Original System Whilst the Blue Points are those after Perturbation. The Histogram Represents the Differences for each Measure under Perturbation. **(d), (f)** and **(h)** are the Coordinate Representations of these three Measures before and after these Perturbation Assays Corresponding to **(c)**, **(e)**, **(g)**, Respectively. Gene No. Indicates the Notations of the System Genes $X_1 \sim X_{16}$

where

$$\underline{F} = \left\{\underline{\gamma}, \underline{M}\right\} \frac{\Delta \underline{\tilde{Y}}}{\Delta \underline{\gamma}} \cong \frac{\underline{Y} \cdot e^{-\underline{\gamma}(\underline{Y}-\underline{M})}}{\left(1+e^{-\underline{\gamma}(\underline{Y}-\underline{M})}\right)^2}, \text{ and } \frac{\Delta \underline{\tilde{Y}}}{\Delta \underline{M}} \cong \frac{-\underline{\gamma} \cdot e^{-\underline{\gamma}(\underline{Y}-\underline{M})}}{\left(1+e^{-\underline{\gamma}(\underline{Y}-\underline{M})}\right)^2}$$

Hence we could discuss the sensitivity on the *trans* level like the input sensitivity. By the analysis of these two parameters, we can manipulate two simulations as follows.

Trans-sensitivity Rate γ Simulation of Gene

In a similar way as in input perturbation, we changed γ from 100% to 0% (– 100%) and 200% (+ 100%) of system genes in pathway to compare with their sensitivities to γ, as shown in Table 3B. We also average the three measure indexes of each gene, which are shown in Figures 7.1, 7.2 & 9.

Trans-expression Threshold M_j Simulation of Gene

We varied M_j to 100% lower (– 100%) and higher (+ 100%) than the original mean expression of the *j*-th gene respectively and compared with their sensitivities of M_j, which are shown in Table 3C; and their average measure indexes are shown in Figures 7.3, 7.4 & 9.

D. Cis-perturbation Assay

In equation (1.3), the genetic kinetic parameter $d_{q,ij}$ can be taken as the influence values of upstream gene on the transcriptional binding or physical interaction with the output gene, which is similar to the *cis* activation. Thus, we can discuss the influence of the variations in genetic kinetic parameters, which can be estimated from equation (1.9), using the circadian network. We also induce their equation of sensitivity like the *trans* case as follows,

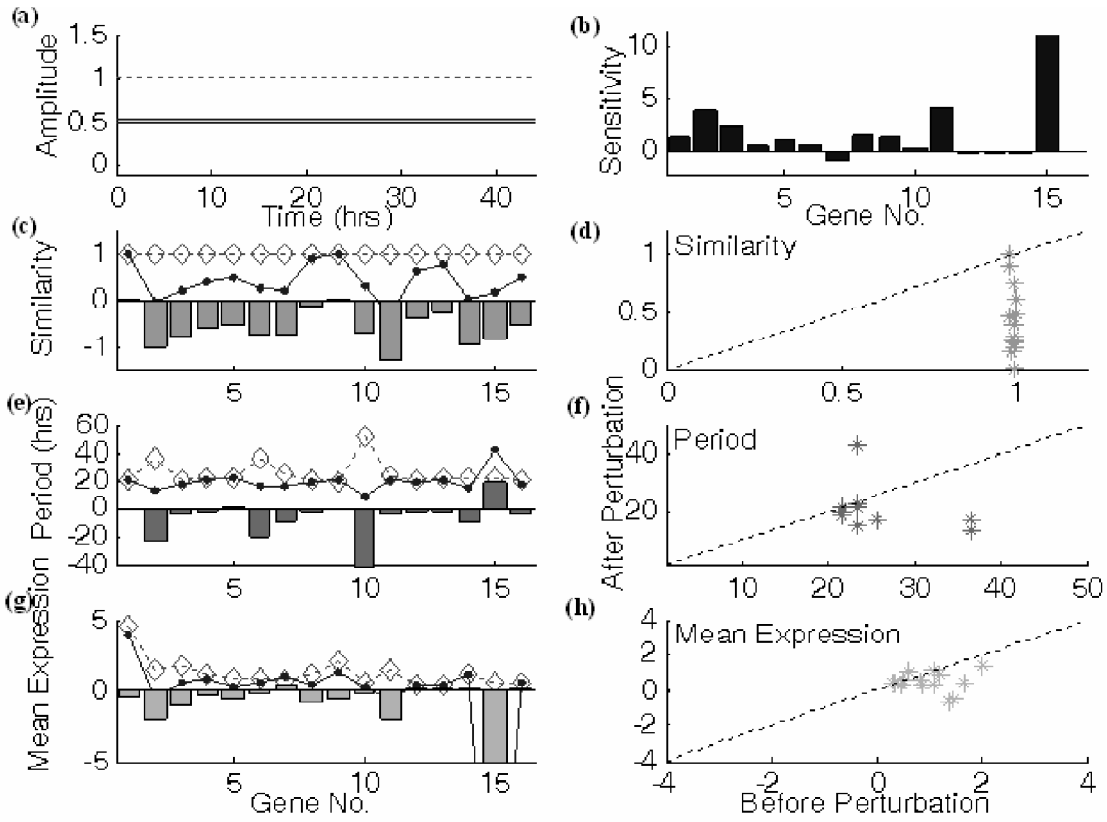

Figure 6: Perturbation Assay under One-half Attenuation of the Input-light Fluence Rate. **(a)** The Illustration of this Assay with the Change of the White-light Fluence from 100% (Red Dashed Line) to 50% (Blue Solid Line) with Respect to the Original Amount. **(b)** The Measure of Sensitivity for the System Genes Due to this Perturbation. **(c), (e)** and **(g)** Indicate the Three System Measures of Similarity (Pearson Correlation), Period, and mean Expression, Respectively. The Red Diamonds are Measures of Original System Whilst the Blue Points are those after Perturbation. The Histogram Represents the Differences for each Measure under Perturbation. **(d), (f)** and **(h)** are the Coordinate Representations of these Three Measures before and after these Perturbation Assays Corresponding to **(c)**, **(e)**, **(g)**, Respectively. Gene No. Indicates the Notations of the System Genes $X_1 \sim X_{16}$

$$\frac{\Delta \underline{Y}}{\Delta \underline{Y}'} \simeq (1 - D_2)^{-1} D_1 \frac{\Delta \underline{\tilde{Y}}}{\Delta \underline{Y}'} \tag{3.13}$$

where $\frac{\Delta \underline{\tilde{Y}}}{\Delta \underline{Y}'} \cong \frac{\underline{\gamma} \cdot e^{-\underline{\gamma}(\underline{Y}' - \underline{M})}}{\left(1 + e^{-\underline{\gamma}(\underline{Y}' - \underline{M})}\right)^2}$, and $\Delta \underline{Y}'$ is the genetic expressions of the perturbation. Hence, we could also discuss the sensitivity on the *cis* level, at which the initial genetic expressions are perturbed to + 1.0 and – 1.0 for the original amount of each system gene and shown in Table 3D.

Variation Simulation of cis Parameters $d_{q,ij}$

For a specific gene *j* in the circadian network, we would alter all the connectivity $d_{q,ij}$, i = 1, 2, ..., n from 0% (– 100%) to 200% (+ 100%) with 5% interval to mimic perturbations of the *cis* circuit and average the three measure indexes for each gene in Figures 8.1, 8.2 & 9.

Mutation Simulation of cis Parameters $d_{q,ij}$

There is a conspicuous perturbation of the *cis* circuit, i.e. the mutation assay. For a specific gene *j* in the circadian network, we also mimic the mutation of all its cis connectivity $d_{q,ij}$, i = 1, 2, ..., n at the same time by setting their values as 0. Using the same way as before, we average their three measure indexes to evaluate the system performance, which are shown in Figures 8.3, 8.4 & 9.

5. RESULTS

Analysis of Data Set

The famous modeling organisms, *Arabidopsis thaliana*, have been well biologically studied and their microarray assays are abundant. The plant behavior in response to the external light, i.e., the circadian regulatory network, which is essential in the physiology and metabolism of plant, has been widely investigated. In this study, we adopted the data set from the works of Harmer (2000) and used the ARX model to construct the circadian pathway and perform the

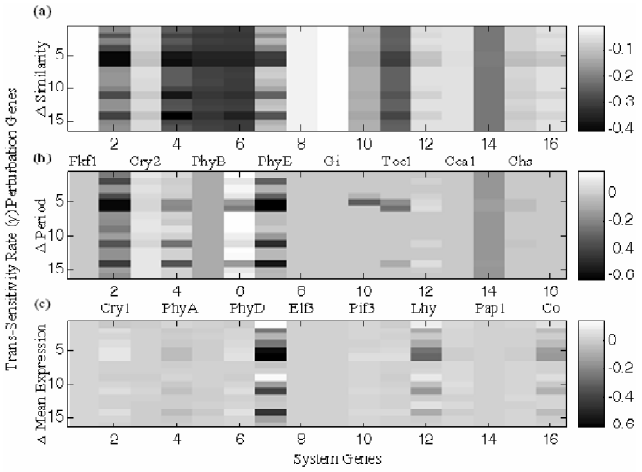

Figure 7.1: Deviation Representations of the System Genes under the Perturbation of Trans-sensitivity Rate γ. The Perturbation is Performed in the Vertical Axis and the Responses of 16 System Genes are shown in the Horizontal Axis, and the Colored Bars of Degree are on the Right-hand Side of each Inset. **(a)** Δ Similarity (Measured by the Pearson Correlation), **(b)** Δ Period, and **(c)** Δ Mean Expression

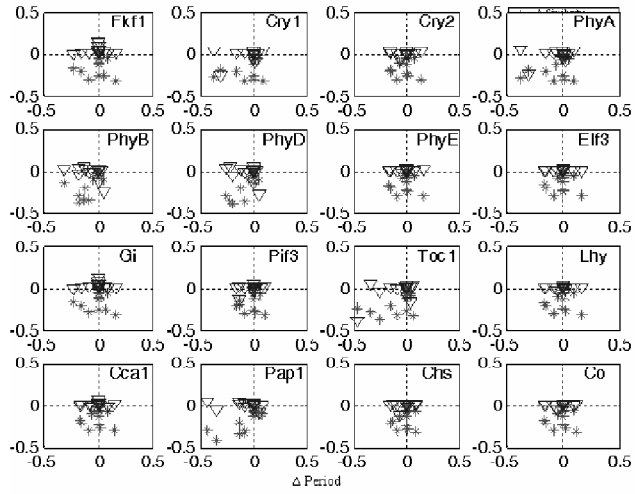

Figure 7.2: Coordinates of the Variations in Similarity and Mean v.s. Period under the Perturbation of *Trans*-sensitivity Rate γ. The Perturbation Genes are shown on the Upper Right Corner of each Inset

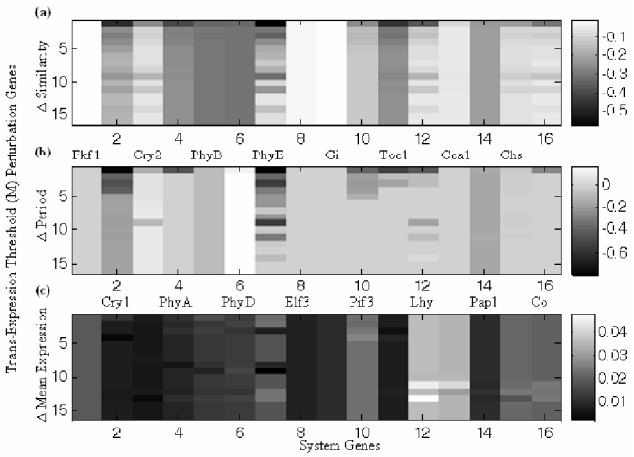

Figure 7.3: Deviation Representations of the System Genes under the Perturbation of Trans-expression Threshold *M*. The Perturbation is Performed in the Vertical Axis and the Responses of 16 System Genes are shown in the Horizontal Axis, and the Colored Bars of Degree are on the Right-hand Side of each Inset. **(a)** Δ Similarity (Measured by the Pearson Correlation), **(b)** Δ Period, and **(c)** Δ Mean Expression

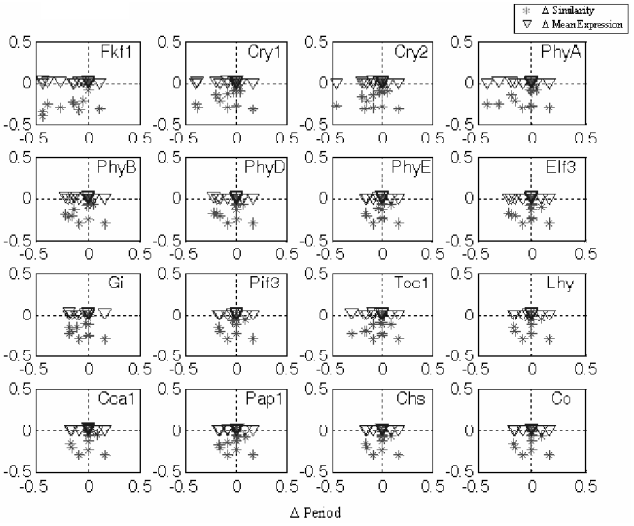

Figure 7.4: Coordinates of Variations in Similarity and Mean v.s. Period under the Perturbation of the Trans-expression Threshold *M*. The Perturbation Genes are shown in each Inset

sensitivity (or robustness) analysis of the circadian regulatory network.

For cells grown in the constant light condition, Harmer and his colleagues used highly reproducible oligonucleotide-based arrays representing about 8200 different genes to determine the steady-state mRNA levels in *Arabidopsis thaliana* that are measured in the replicate hybridization of 12 samples harvested every 4 hours over 2 days. With their investigation of the circadian regulatory system, Harmer *et al.*, have provided an abundance of correlated genes from which we choose 16 core genes of the circadian network with their time-course microarray data for our dynamic system identification and analysis.

Analysis of Circadian Network Model

For the discussion of the circadian network modeled by the stochastic ARX model, we have designed several assays as in the above section, like the biological experiments in the wet laboratory, to discover the genetic responses and

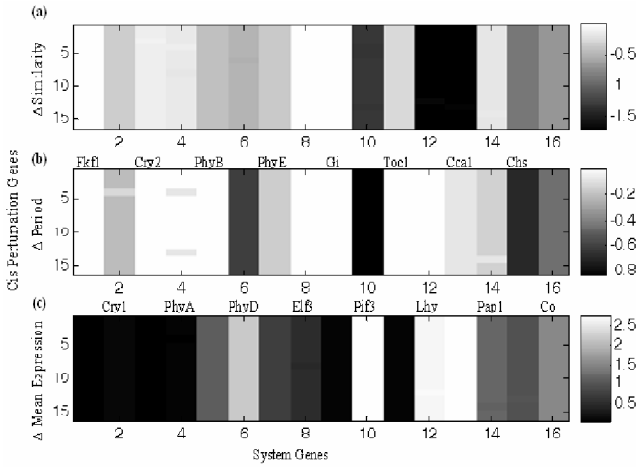

Figure 8.1: Deviation Representations of the System Genes under the *cis* Perturbation. The Perturbation is Performed in the Vertical Axis and the Responses of 16 System Genes are shown in the Horizontal Axis, and the Colored Bars of Degree are on the Right-hand Side of each Inset. **(a)** Δ Similarity (Measured by the Pearson Correlation), **(b)** Δ Period, and **(c)** Δ Mean Expression

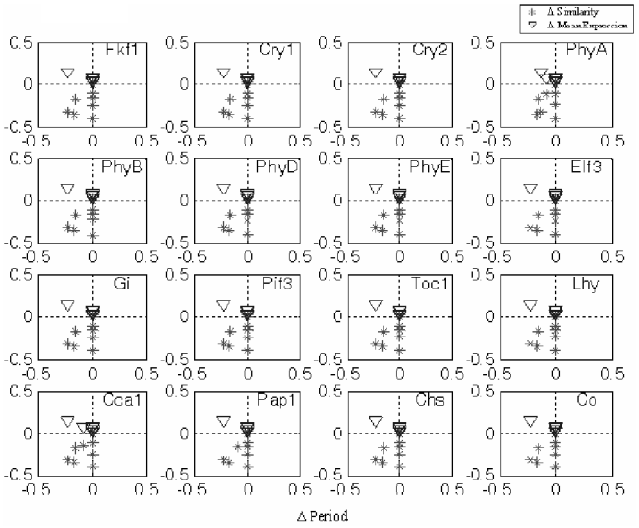

Figure 8.2: Coordinates of Variations in Similarity and mean v.s. Period under the cis Perturbation. The Perturbative Genes are shown in each Inset

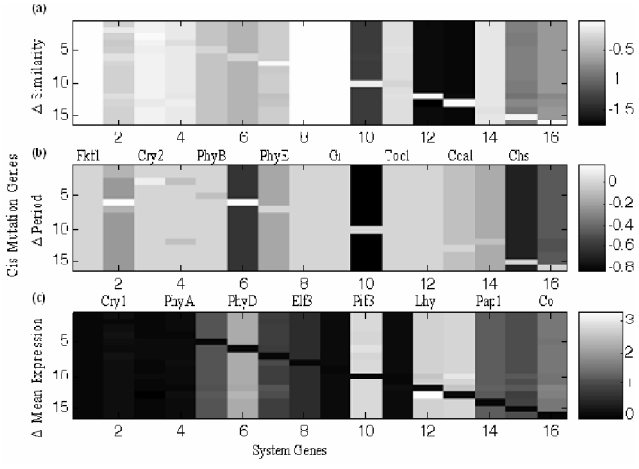

Figure 8.3: Deviation Representations of the System Genes under the *Cis* Mutation. The Mutation is Performed in the Vertical Axis and the Responses of 16 System Genes are shown in the Horizontal Axis, and the Colored Bars of Degree are on the Right-hand Side of each Inset. **(a)** Δ Similarity (Measured by the Pearson Correlation), **(b)** Δ Period, and **(c)** Δ Mean Expression

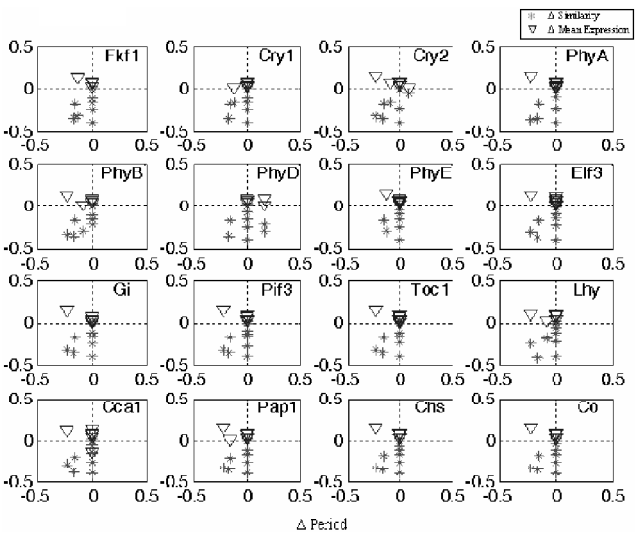

Figure 8.4: Coordinates of Variations in Similarity and Mean v.s. Period under the Cis Mutation. The Mutant Genes are shown in each Inset

elucidate the robustness or sensitivity of the 16 genes in the circadian network.

A. Rhythm Frequency Estimation from Circadian Regulatory Network

The Rhythm exists in the circadian regulatory network via feedback regulations. It can be estimated by equation (3.9) from the identified circadian regulatory network in Table 1. This result can be confirmed by the frequency directly detected by equation (2.3) from microarray data.

B. Simulation of Input-light Perturbation

The paces of the circadian clock are synchronized by environmental cues such as light. The stimulus of the external light could be divided into red light and blue light, which can be absorbed by phytochromes (PhyA [X_4], PhyB [X_5], PhyD [X_6], and PhyE [X_7]) (Martinez-Garcia *et al.*, 2000) and crytochromes (Cry1 [X_2] and Cry2 [X_3]) (Casal, 2000), respectively. Hereby we model the white light with an amplified perturbation on the fluence or amplitude as in Figure 5a. In Figures 5c and 5d, the similarities of the 150%

(+ 50% perturbation) light fluence for system genes show deviation implying a strong influence of light perturbation on the amplified amplitude. However, when considering the periods of expression in Figures 5e and 5f, the significant decrement of period appears in Cry1 [X_2] and Cry2 [X_3] (Somers *et al.*, 1998), indicating its potent dependence on light quantity. There are the elevated mean expressions of Cry1 [X_2] and Cry2 [X_3] in Figure 5g consistent with the extreme sensitivity in Figure 5b and Table 3A (Somers *et al*., 1998). Moreover, the coherent reductions in period can be observed in photoreceptors ([X_1] ~ [X_7]). However, Chs [X_{15}] is found to have the largest sensitivity (see Figure 5b and Table 3A) to both blue and UV light (Deikman and Hammer, 1995). In addition, the attenuated amplitude of light fluence (– 50%, see Figure 6a) will reduce most of the similarities of gene expressions as in the Figures 6c & 6d.

C. Simulation of Trans Perturbation

In the perturbation of *trans*-sensitivity rate (γ), we will discuss whether the transition rate, which determines the transition time of one gene binding to or interacting with another one, affects the system gene's expression in this model system. It seems that the similarity (Figure 7.1a) remains unchanged for most system genes except Cry1 [X_2], PhyA [X_4], PhyD [X_6], and PhyE [X_7] (Aukerman *et al.*, 1997). In addition, the results of PhyB [X_5] and PhyD [X_6] (Figure 7.1a) are much the same as the provided evidence that the role of PhyD [X_6] is similar to PhyB [X_5] (Aukerman *et al.*, 1997). If we consider the periodic variation in Figure 7.1b, Cry2's [X_3] period is lengthened about 10%, whilst that of Cry1 [X_2] and Pap1 [X_{14}] are shortened about 20%, respectively. The diversity and sensitivity of period due to perturbation of the transition time are evident as in Figure 7.1c. The mean expressions of system genes are almost unaltered but PhyE [X_7] (Devlin *et al.*, 1998) is reduced. Further, the results of Figure 7.2 exhibit the robustness of the similarity and mean expressions of the system genes. Therefore, the mimic assay of transition time by the perturbation of *trans*-sensitivity rate has revealed that the period is the only susceptible characteristic in contrast to the similarity and mean expression. From the result of Table 3B, we also found three genes of Cry1 [X_2], PhyD [X_6], and Toc1 [X_{11}] with significant sensitivities with respect to the perturbation of the *trans*-sensitivity rate. Because the largest difference in the mean sensitivity of each gene of Table 3B is about 0.025, we would conclude that the *trans*-sensitivity rate, which determines the transition time indicating the transient state of *trans* activation, has less influence on the circadian system.

In another perturbation of *trans*-expression threshold *M*, we would realize the influence of activation threshold of gene expression on the *trans* level, which determines the activating abilities of upstream genes. There are five genes of Cry1 [X_2], Cry2 [X_3], PhyD [X_6], Pif3 [X_{10}] and Toc1 [X_{11}] with perceptible variations, which have the same behavior in the measures of similarity and period (see Figure 7.3a and 7.3b). However, almost all mean expressions of system genes decay in Figure 7.3c. Thus, we suspect that their deviation in similarity is the result of changes in periods and means of expression profiles. As shown in Figure 7.4, the similarity and mean expression have more deviations in contrast to the period due to the wider distribution on the vertical axis. Nevertheless, the perturbation of Fkf1 [X_1] has different behavior of the stable similarity and mean. This robustness means lower correlation with the regulations of other genes in the circadian network. In the measure of sensitivity in Table 3C, there are three apparent genes of PhyE [X_7], Lhy [X_{12}], and Co [X_{16}] with crucially positive sensitivities, whilst Cry1 [X_2] and Toc1 [X_{11}] have significantly negative ones. Owing the largest difference in the mean sensitivity of each gene of Table 3C being close to 0.29, the circadian network is more sensitive to the perturbation of the *trans*-expression threshold *M* and the activation level of steady state, rather than the *trans*-sensitivity rate γ.

D. Simulation of Cis Perturbation

In the assay of *cis* perturbation, we attempt to alter the values of the regulatory abilities of the genetic kinetic parameters $d_{q,ij}$ in the ARX system model. This could be considered as the variations in the *cis* level. In Figure 8.1a, the similarities of the system genes are mostly the same. Nevertheless, the diagonal has few variations in the similarity to demonstrate the essential role of the genetic kinetic parameter for each gene in the network. We also found that Pif3 [X_{10}] has a significant decrement in period. As in Figure 8.2, the stabilization of the similarity and mean expression is manifested, especially in the perturbation of the biological clock genes of Toc1 [X_{11}], Lhy [X_{12}], and Cca1 [X_{13}] (Mas *et al.*, 2003). Hence, the single perturbation of these three genes will not affect the circadian system due to their well-known close interconnections with potential robustness. This can be demonstrated by the lower sensitivities in Table 3D consistent with Tables 3B and 3C.

In addition to the *cis* perturbation, the mimic genetic mutation on the *cis* level is also a crucial strategy of study, which is very popular in the traditional biological assay. The mutation of each network gene in order has revealed the uncovered deviation of gene expression in the diagonal of the three insets in Figure 8.3. Obviously, in Figure 8.4, the *cis* mutations of Elf3 [X_8], Pap1 [X_{14}], Chs [X_{15}], and Co [X_{16}] could not affect the system genes but their periods. However, the mutants of the remaining genes result in significant deviations of their gene expressions.

In the overview of the perturbation of genes on the *trans* and *cis* levels, we showed these four types of perturbations consistently in Figure 9. The fact that the most diversity appears in the cis mutation indicates that the ability of *cis* activation is essential to the circadian system network. The

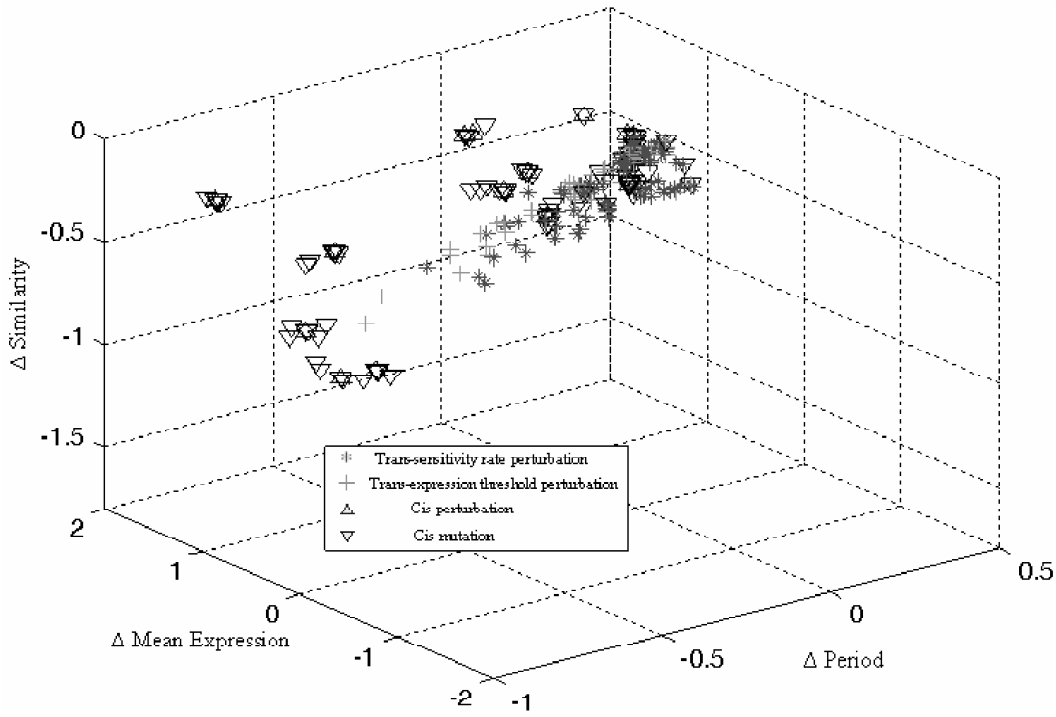

Figure 9: Overview of the Genetic Perturbations on *Trans* and *Cis* Levels in the Circadian System Model with the Measures of Δ Period, Δ Mean Expression and Δ Similarity for all System Genes

other three perturbations are more robust. We presume that the transition time concerning the transient state on the *trans* level is less effective on genetic interactions. The complexity of the circadian network will resist the perturbations of genes on the *cis* level except the *cis* mutation, which is consistent with the results of sensitivities in Tables 3B-3D.

6. DISCUSSION

Microarray analysis using the stochastic system approach offers an opportunity to generate the interpretation of functional influence on a specific genetic regulatory network. The crucial ontology behind using stochastic system techniques is that the framework of the interactivity between gene expression profiles could be recognized quantitatively according to the stochastic process such as ARX underlying a dynamic system under a noisy environment. Therefore, because the microarray data were harvested with time progression (Harmer *et al.*, 2000), the simultaneously varied gene expressions implicated in the circadian regulatory network of *Arabidopsis thaliana* would be detected via interactive stochastic modeling in *silico* in spite of the versatile interactions such as transcriptional control, protein phosphorylation, or specific complex interaction etc.

The clustering method (Soukas *et al.*, 2000 and 2001; Tamayo *et al.*, 1999) answers the problem of what is the functional catalogue of a specific gene by the identification of resembling patterns of gene expressions. Similarly, the co-regulations of upstream genes in our method also imply their concurrent functions. In contrast to the clustering algorithm, the causality of time-course data has been smoothly drawn by our dynamic method. The Bayesian networks (Friedman *et al.*, 2000) were used merely for forward probabilistic estimation with the time transition lacking in the feedback linkages. This unidirectional problem would not happen in our algorithm. Owing to the quantitative regulatory abilities of our model, we have a greater diversity of regulatory influence than the Boolean networks (Huang, 1999), which are deterministic with merely two states.

In our stochastic system approach applied to the circadian network using stochastic ARX, we not only can identify the regulatory abilities via interactive ARX(1) model with activation delays, but also indicate the regulatory strength from the input-light signal. In terms of the regulatory abilities, the comparison between the upstream regulatory genes of a target gene can inspire us to ask which one is significant biologically and whether it is a positive or negative influence on the investigated gene as shown in Figure 3. Further, the speculation of activation delays benefits the experimental reference by providing us when the upstream regulatory genes might interact with their target genes in the circadian regulatory network. The greatest importance of the proposed stochastic model is the convenience of the consequent system analysis, for example, sensitivity analysis, to gain more insight about the circadian regulatory network. From the result of Table 3, the photoreceptor genes and their down stream target genes in circadian regulatory network are more sensitive to input light

perturbations; and the circadian regulatory networks are more sensitive to cis perturbations than to trans perturbations. Furthermore, the activation or repression relationships inferred via microarray data would distinctly uncover the overall effect of regulatory interactions among system genes in the circadian regulatory network on the transcriptional level.

In addition to the stochastic system modeling of the circadian regulatory network *in silico*, the system analysis of the circadian regulatory network by perturbation assays will reveal the practice of systems biology. Especially, we measure systematically the simulation output profiles by three indexes of similarity, period, and mean expression. In the investigation of 16 genes of the circadian regulatory network, we have demonstrated that the period of Chs is very sensitive to the perturbation of input light implying its essentiality among photoreceptors. If we perform the perturbation on the *trans* level, we will discover that the activation threshold of *trans* gene expression is crucial to the circadian system. Cry1, PhyD, PhyE, and Chs are significantly sensitive to the perturbation of the *trans*-expression threshold. This implies that the *trans*-expression threshold, like the activation degree of the protein, should be confined to avoid the disturbance of the genetic system. The robustness of the circadian system is shown in the *cis* perturbation while the *cis* mutation will cause more lethal in the gene expression. The less sensitive genes are Lhy, Cca1 and, Toc1 on the *cis* perturbation or mutation due to their core functions of the initiation of the biological clock. This is why the period of the circadian genes will be altered facilely after the perturbation. From the analysis of the circadian network by system modeling approach *in silico*, we can operate many genetic assays with the external light stimulus that will not be performed easily in reality to discover the responses among the involved genes in the circadian system.

There are some shortcomings in our study. First, although the time-course microarray data are available, its lower samplings will distort the real changes of gene expressions, especially for fast dynamic evolution. A more sampling experiment with respect to the intrinsic turnover rate is expected for a more precise analysis. Second, we formulate our ARX circadian network model using the biological knowledge of the correlations between the circadian genes. In this situation, we should know the exact interactive partners of the system genes. Hence, the more proper knowledge of the correlations can provide, the more accurate the stochastic system model is. In the circadian regulatory network, it is enough for us to construct the stochastic system because of its simulation similarity approaching 0.99 in Figure 2a. Third, a larger network containing more than one hundred genes will make modeling and discussion difficult due to computation complexity. Finally, the activation profiles under the proteome should be highly correlated with the transcriptional profiles to improve the interpretation ability of our system model. In general, the synchronous time-course microarray assay is more suitable to detect the transcriptional binding among genes, so that an inference of physical interactions in the post-transcriptional level will become more feasible in our study.

In the near future, the most pressing task is to investigate our presumed characteristics of the circadian network in the laboratory. As the stochastic system modeling algorithms are further developed, we expect this stochastic system approach to have immense impact in elucidating the underlying molecular mechanisms of network in a variety of organisms besides the circadian network in *Arabidopsis thaliana*, especially after the maturation of the protein chips. Ultimately, we envision that biologists will perform routine system analysis *in silico* to seek some novel genetic characteristics and to identify the robust or sensitive links of system genes.

REFERENCES

[1] M. Ahmad, J. A. Jarillo, O. Smirnova and A. R. Cashmore, The CRY1 Blue Light Photoreceptor of Arabidopsis Interacts with Phytochrome A *in vitro*. *Mol. Cell*, **1**, (1998), 939-948.

[2] D. Alabadi, T. Oyama, M. J. Yanovsky, F. G. Harmon, P. Mas and S. A. Kay, Reciprocal Regulation between TOC1 and LHY/CCA1 within the *Arabidopsis* Circadian Clock. *Science*, **293,** (2001), 880-883.

[3] O. Alter, P. O. Brown and D. Botstein, Singular Value Decomposition for Genome-wide Expression Data Processing and Modeling. *Proc. Natl. Acad. Sci. USA*, **97**, (2000), 10101-10106.

[4] M. J. Aukerman, M. Hirschfeld, L. Wester, M. Weaver, T. Clack, R. M. Amasino and R. A. Sharrock, A Deletion in the *PHYD* Gene of the Arabidopsis Wassilewskija Ecotype Defines a Role for Phytochrome D in Red/ Far-red Light Sensing. *Plant Cell*, **9**, (1997), 1317-1326.

[5] M. J. Beal, F. Falciani, Z. Ghahramani, C. Rangel and D. L. Wild, A Bayesian Approach to Reconstructing Genetic Regulatory Networks with Hidden Factors. *Bioinformatics*., **21**(3), (2005), 349-56.

[6] A. Brazma, H. Parkinson, U. Sarkans, M. Shojatalab, J. Vilo, N. Abeygunawardena, E. Holloway, M. Kapushesky, P. Kemmeren, G. G. Lara, A. Oezcimen, P. Rocca-Serra and S. A. Sansone, ArrayExpress—A Public Repository for Microarray Gene Expression Data at the EBI. *Nucleic Acids Res.*, **31**, (2003), 68-71.

[7] M. P. Brown, W. N. Grundy, D. Lin, N. Cristianini, C. W. Sugnet, T. S. Furey, M. Jr. Ares and D. Haussler, Knowledge-based Analysis of Microarray Gene Expression Data by using Support Vector Machines. *Proc. Natl. Acad. Sci. USA*, **97**, (2000), 262-267.

[8] J. J. Casal, Phytochromes, Cryptochromes, Phototropin: Photoreceptor Interactions in Plants. *Photochem. Photobiol.*, **71**, (2000), 1-11.

[9] H. C. Causton, B. Ren, S. S. Koh, C. T. Harbison, E. Kanin, E. G. Jennings, T. I. Lee, H. L. True, E. S. Lander, and R. A. Young, Remodeling of Yeast Genome Expression in Response to Environmental Changes. *Mol. Biol. Cell*, **12**, (2001), 323-337.

[10] B. S. Chen, Y. C. Wang, W. S. Wu and H. W. Li, A New Measure of the Robustness of Biochemical Networks. Bioinformatics. **21**, (2005), 2698-705.

[11] H. C. Chen, H. C. Lee, T. Y. Lin, W. H. Li, B. S. Chen, Quantitative Characterization of the Transcriptional Regulatory Network in the Yeast Cell Cycle. Bioinformatics. **20**, (2004), 1914-27.

[12] J. Deikman, P. E. Hammer, Induction of Anthocyanin Accumulation by Cytokinins in Arabidopsis Thaliana. *Plant Physiol.* **108**, (1995) 47-57.

[13] P. F. Devlin, S. R. Patel, G. C. Whitelam, Phytochrome E Influences Internode Elongation and Flowering Time in Arabidopsis. *Plant Cell*. **10**, (1998), 1479-87.

[14] R. Edgar, M. Domrachev and A. E. Lash, Gene Expression Omnibus: NCBI Gene Expression and Hybridization Array Data Repository. *Nucleic Acids Res.*, **30**, (2002), 207-210.

[15] M. B. Eisen, P. T. Spellman, P. O. Brown and D. Botstein, Cluster Analysis and Display of Genome-wide Expression Patterns. *Proc. Natl. Acad. Sci. USA*, **95**, (1998), 14863-14868.

[16] C. Fankhauser, D. Staiger, Photoreceptors in Arabidopsis Thaliana: Light Perception, Signal Transduction and Entrainment of the Endogenous Clock. *Planta.*, **216**, (2002), 1-16.

[17] N. Friedman, M. Linial, I. Nachman and D. Pe'er, Using Bayesian Networks to Analyze Expression Data. *J. Comput. Biol.*, **7**, (2000), 601-620.

[18] A. P. Gasch and M. B. Eisen, Exploring the Conditional Coregulation of Yeast Gene Expression Through Fuzzy k-means Clustering. *Genome Biol.*, **3**, (2002), RESEARCH0059.

[19] D. P. Harkin, Uncovering Functionally Relevant Signaling Pathways Using Microarray-Based Expression Profiles. *The Oncologist*, **5**, (2000) 501-507.

[20] S. L. Harmer, J. B. Hogenesch, M. Straume, H. S. Chang, B. Han, T. Zhu, X. Wang, J. A. Kreps and S. A. Kay, Orchestrated Transcription of Key Pathways in *Arabidopsis* by the Circadian Clock. *Science*, **290**, (2000), 2110-2113.

[21] R. Hayama and G. Coupland, Shedding Light on the Circadian Clock and the Photoperiodic Control of Flowering. *Curr Opin Plant Biol.*, **6**(1), (2003), 13-9.

[22] S. Huang, Gene Expression Profiling, Genetic Networks, and Cellular States: An Integrating Concept for Tumorigenesis and Drug Discovery. *J. Mol. Med.*, **7**, (1999), 469-480.

[23] T. S. Hughes, B. Abou-Khalil, P. J. Lavin, T. Fakhoury, B. Blumenkopf, S. P. Donahue, Visual Field Defects after Temporal Lobe Resection: A Prospective Quantitative Analysis. *Neurology.*, **53**, (1999), 167-72.

[24] R. Johansson, *System Modeling and Identification*, PrenticeHall, Englewood Cliffs, NJ, (1993).

[25] J. R. Kettman, J. R. Frey and I. Lefkovits, Proteome, Transcriptome and Genome: Top Down or Bottom up Analysis. *Biomol. Eng.*, **18**, (2001) 207-212.

[26] R. J. Lipshutz, S. P. Fodor, T. R. Gingeras and D. J. Lockhart, High Density Synthetic Olignucleotide Arrays. *Nat. Gent.* **(suppl.) 21**, (1999), 20-24.

[27] J. F. Martinez-Garcia, E. Huq and P. H. Quail, Direct Targeting of Light Signals to a Promoter Element-bound Transcription Factor. *Science*, **288**, (2000), 859-863.

[28] P. Mas, D. Alabadi, M. J. Yanovsky, T. Oyama and S. A. Kay, Dual Role of TOC1 in the Control of Circadian and Photomorphogenic Responses in *Arabidopsis*. *Plant Cell*, **15**, (2003), 223-236.

[29] S. Motaki, K. Ayako, Y. S. Kazuko and S. Kazuo, Molecular Response to Drought, Salinity and Frost: Common and Different Paths for Plant Protection. *Cur. Opn. Biotechnol.* **14**, (2003), 194-199.

[30] C. Rangel, J. Angus, Z. Ghahramani, M. Lioumi, E. Sotheran, A. Gaiba, D. L. Wild and F. Falciani, Modeling T-cell Activation using Gene Expression Profiling and State-space Models. *Bioinformatics*., **20**(9), (2004), 1361-72.

[31] R. Schaffer, J. Landgraf, M. Accerbi, V. Simon, M. Larson, and E. Wisman, Microarray Analysis of Diurnal and Circadian-regulated Genes in *Arabidopsis*. *Plant Cell*, **13**, (2001), 113-123.

[32] J. Scheel, M. C. Von Brevern, A. Horlein, A. Fischer, A. Schneider and A. Bach, Yellow Pages to the Transcriptome. *Pharmacogenomics*, **3**, (2002), 791-807.

[33] M. Schena, D. Shalon, R. W. Davis and P. O. Brown, Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. *Science*, **270**, (1995), 467-470.

[34] G. Sherlock, T. Hernandez-Boussard, A. Kasarskis, G. Binkley, J. C. Matese, S. S. Dwight, M. Kaloper, S. Weng, H. Jin, C. A. Ball, M. B. Eisen, P. T. Spellman, P. O. Brown, D. Botstein and J. M. Cherry, The Stanford Microarray Database. *Nucleic Acids Res.*, **29**, (2001), 152-155.

[35] D. Staiger, Circadian Rhythms in *Arabidopsis*: Time for Nuclear Proteins. *Planta.*, **214**, (2002), 334-344.

[36] J. J. E. Slotine and W. Li, Applied Nonlinear Control Prentice Hall, (1991).

[37] D. E. Somers, P. F. Devlin and S. A. Kay, Phytochromes and Cryptochromes in the Entrainment of the *Arabidopsis* Circadian Clock. *Science*, **282**, (1998), 1488-1490.

[38] A. Soukas, P. Cohen, N. D. Socci, J. M. Friedman, Leptin-specific Patterns of Gene Expression in White Adipose Tissue, *Genes Dev.*, **14**, (2000), 963-80.

[39] A. Soukas, N. D. Socci, B. D. Saatkamp, S. Novelli, J.M. Friedman, Clock-associated Genes in *Arabidopsis*: a Family Affair. *Philos. J Biol Chem.*, **276**, (2001), 34167-743.

[40] P. T. Spellman, G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. Q. Brown, D. Bostein and B. Futcher, Comprehensive Identification of Cell Cycle-Regulated Genes of the Yeast *Saccharomyces Cererisiae* by Microarray Hybridization. *Mol. Biol. Cell.*, **9**, (1998), 3273-3297.

[41] P. Tamayo, D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E. S. Lander and T. R. Golub, Interpreting Patterns of Gene Expression with Self-Organizing Maps: Methods and Application to Hematopoietic Differentiation. *Proc. Natl. Acad. Sci. USA*, **96**, (1999), 2907-2912.

[42] S. Tavazoie, J. D. Hughes, M. J. Campbell, R. J. Cho, and G. M. Church, Systematic Determination of Genetic Network Architecture. *Nat. Genet.*, **22**, (1999), 281-285.

[43] A. Tanay, I. Steinfeld, M. Kupiec and R. Shamir, Integrative Analysis of Genome-wide Experiments in the Context of a Large High-throughput Data Compendium. *Molecular Systems Biology*, (2005), doi: 10.1038/msb4100005.

[44] R. Toth, E. Kevei, A. Hall, A. J. Millar, F. Nagy and L. Kozma-Bognar, Circadian Clock-regulated Expression of Phytochrome and Cryptochrome Genes in Arabidopsis. *Plant Physiol.*, **127**, (2001), 1607-16.

[45] V. E. Velculescu, L. Zhang, W. Zhou, J. Vogelstein, M. A. Basrai, D. E. Jr. Bassett, P. Hieter, B. Vogelstein and K. W. Kinzler, Characterization of the Yeast Transcriptome. *Cell*, **88**, (1997) 243-251.

[46] F. X. Wu, W. J. Zhang and A. J. Kusalik, Modeling Gene Expression from Microarray Expression Data with State-Space Equations. *Pac Symp Biocomput.*, **9**, (2004), 581-92.

[47] M. J. Yanovsky and S. A. Kay, Signaling Networks in the Plant Circadian System. *Curr. Opin. Plant Biol.*, **4**, (2001), 429-435.