

Research article

Open Access

Quantitative inference of dynamic regulatory pathways via microarray data

Wen-Chieh Chang¹, Chang-Wei Li¹ and Bor-Sen Chen*²

Address: ¹Lab. of System Biology, National Tsing Hua University, 101, Sec 2, Kuang Fu Road, Hsinchu, 300, Taiwan and ²Department of Electrical Engineering, National Tsing Hua University, 101, Sec 2, Kuang Fu Road, Hsinchu, 300, Taiwan

Email: Wen-Chieh Chang - cwli@moti.ee.nthu.edu.tw; Chang-Wei Li - d925908@oz.nthu.edu.tw; Bor-Sen Chen* - bschen@ee.nthu.edu.tw

* Corresponding author

Published: 07 March 2005

Received: 17 August 2004

BMC Bioinformatics 2005, 6:44 doi:10.1186/1471-2105-6-44

Accepted: 07 March 2005

This article is available from: <http://www.biomedcentral.com/1471-2105/6/44>

© 2005 Chang et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The cellular signaling pathway (network) is one of the main topics of organismic investigations. The intracellular interactions between genes in a signaling pathway are considered as the foundation of functional genomics. Thus, what genes and how much they influence each other through transcriptional binding or physical interactions are essential problems. Under the synchronous measures of gene expression via a microarray chip, an amount of dynamic information is embedded and remains to be discovered. Using a systematically dynamic modeling approach, we explore the causal relationship among genes in cellular signaling pathways from the system biology approach.

Results: In this study, a second-order dynamic model is developed to describe the regulatory mechanism of a target gene from the upstream causality point of view. From the expression profile and dynamic model of a target gene, we can estimate its upstream regulatory function. According to this upstream regulatory function, we would deduce the upstream regulatory genes with their regulatory abilities and activation delays, and then link up a regulatory pathway. Iteratively, these regulatory genes are considered as target genes to trace back their upstream regulatory genes. Then we could construct the regulatory pathway (or network) to the genome wide. In short, we can infer the genetic regulatory pathways from gene-expression profiles quantitatively, which can confirm some doubted paths or seek some unknown paths in a regulatory pathway (network). Finally, the proposed approach is validated by randomly reshuffling the time order of microarray data.

Conclusion: We focus our algorithm on the inference of regulatory abilities of the identified causal genes, and how much delay before they regulate the downstream genes. With this information, a regulatory pathway would be built up using microarray data. In the present study, two signaling pathways, i.e. circadian regulatory pathway in *Arabidopsis thaliana* and metabolic shift pathway from fermentation to respiration in yeast *Saccharomyces cerevisiae*, are reconstructed using microarray data to evaluate the performance of our proposed method. In the circadian regulatory pathway, we identified mainly the interactions between the biological clock and the photoperiodic genes consistent with the known regulatory mechanisms. We also discovered the now less-known regulations between cryptochrome and phytochrome. In the metabolic shift pathway, the casual relationship of enzymatic genes could be detected properly.

Background

Biological phenomena at different organismic levels have revealed some sophisticated systematic architectures of cellular and physiological activities implicitly. These architectures were built upon the biochemical processes before the emergence of proteome and transcriptome [1-3]. 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 binding to regulatory sites of other genes, as enzymes catalyzing metabolic reactions, or as components of signal transduction pathways [4-6]. 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 [7]. In the experience of engineering field, the systematic architecture and dynamic model could investigate the characteristics of signaling regulatory pathways [8]. Therefore, how to construct the dynamic model of a signaling pathway from the system structure point of view might be the first key to the door of system biology. Most biological phenomena directly or indirectly influenced by genes such as metabolism, stress response, and cell cycle are well studied on the molecular basis. Thus, 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 pathway besides the qualitative analysis. More particularly, the embedded time-course feature of microarray data would promote the system analysis of signal regulatory pathways as well, which is very mature in the field of engineering.

In addition to northern blots and reverse transcription-polymerase chain reaction (RT-PCR), which study a small number of genes in a single assay, the transcriptome analysis has, via DNA microarray technology [9], managed to achieve high-throughput monitoring of the almost genome-wide mRNA expression levels in living cells or tissues. Two types of available microarrays, the spotted cDNA and *in situ* synthesized oligonucleotide [10] chips, which permit the spatiotemporal expression levels of genes to be rapidly measured in a massively parallel way, are used in different experimental requirements and stocked in the databases on net, such as Stanford Microarray Database (SMD) [11], Gene Expression Omnibus(GEO) [12] in NCBI, and ArrayExpress [13] 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 [14-17].

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 co-expression is indicative of the co-regulation, thus clustering may identify genes that have similar functions or are involved in the related biological processes. The most widely used method is the unsupervised hierarchical clustering [18]. This approach has an increasing number of nested classes by similarity measurement and resembles a phylogenetic classification. If we know the number of clusters in advance, the k-means clustering [19] could assign gene elements into a fixed number k of clusters in a way to minimize the overall Pearson or Euclidean distances of each member internally in the same cluster. Other algorithms such as the neural-network-based self-organizing maps (SOM) [20], singular value decomposition (SVD) or principal component analysis (PCA) [21], and fuzzy clustering methods [19] also have their own advantages and limitations. Alternative supervised clustering algorithm of support vector machine [22], which uses prior biological information of cluster for training, would enhance the accuracy of clustering. However, the nature of clustering algorithms apparently cannot uncover the causal interactions between genes just by grouping. Regarding the causality of pathways, the clustering analysis needs to cooperate with sequence motif detection [23]. 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.

A statistical model of Bayesian network [24] was proposed to model genetic regulatory networks. Basically, the technique uses a probabilistic score to evaluate the networks with respect to the expression data and searches for the network with the optimal score. The dynamic Bayesian network [25] was proposed to learn the network structure and parameters by maximizing the posterior probability via Bayes rule of prior probability and marginal likelihood. Another algorithm of Boolean networks [26] can also be employed to model the dynamic evolution of gene expression, where the state of a gene can be simplified to being either active (on, 1) or inactive (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. The validity of dynamic Bayesian networks is evaluated by the sensitivity-specificity score ratios [25], which depend on the training size, the degree of accuracy of prior assumption. A genetic regulatory network based on the first order differential equation with given decay rates was discussed in [27].

In this study, the dynamic system approach could be employed to model how a target gene's expression profile is regulated by its upstream regulatory genes from the system causality point of view. Then, with the causal dynamic model, the upstream regulatory function can be extracted from the expression profile of the target gene by the optimal estimation method, i.e. maximum likelihood estimation. Since merely the second-order differential equation is employed to model the dynamic evolution of the target gene, only a few parameters need to be estimated. Furthermore, the derived regulatory function is closely related to the causal upstream information of the pathway and will create a basis for inferring the regulatory pathway from the system biology point of view.

In either eukaryote or prokaryote, signaling regulatory pathways are considered as responses to the physiological activities or the deviation from homeostasis, which would affect the normal states of an organism. Among these signaling regulatory pathways, cell cycle [17] is one of the most conspicuous features of life which plays an important role in growth and cellular differentiation in all organisms. In plants, the stress-induced pathways [28] are very important to survivability under the abiotic environmental treatment such as drought, salinity and cold [29]. If these critical pathways can be identified from quantitative analysis *in silico*, the defect of biological processes would be predicted and corrected before hand. Our aim is to construct signaling regulatory pathways quantitatively by the system inference approach with a dynamic model and microarray data.

In this study, a second-order differential equation, which has been widely used to model many physical dynamic systems with good characteristics, is proposed to model the time-profile evolutionary behavior of a target gene. The regulatory function is taken as the driving input of the dynamic equation of the target gene. Using the dynamic equation and microarray data, we first extract the regulatory function for each target gene. According to the extracted regulatory function, we deduce their upstream regulators to trace back upstream signaling pathways. Then, upstream regulatory genes are taken as target genes to trace back their upstream regulatory genes. Iteratively, we can construct the whole regulatory pathway to the genome wide using the dynamic regulatory model and microarray data from the system biology point of view. Finally, we give some independent validation of our approach by repeating the analysis with randomly reshuffling the time order of microarray data and see if the proposed pathways are destroyed.

We have applied our dynamic system approach to two genetic regulatory pathways with microarray data sets publicly available on net [15,30]. One is the circadian reg-

ulatory pathway in *Arabidopsis thaliana* [31,32], and the other is the metabolic shift pathway from fermentation to respiration in yeast *Saccharomyces cerevisiae* [33]. The circadian system is an essential signaling pathway that allows organisms to adjust cellular and physiological processes in anticipation of periodic changes of light in the environment [34-38]. According to the synchronously dynamic evolution of microarray data, we have successively identified the core signaling transduction from light receptors to the endogenous biological clock [39,40], which is coupled to control the correlatively physiological activity with paces on a daily basis. On the other hand, the diauxic shift [41] from the exhausted fermentable sugar of anaerobic metabolism to aerobic growth is correlated with widespread changes in the expression of genes involved in fundamental cellular processes such as carbon metabolism, protein synthesis, and carbohydrate storage. [28,31,32,42-47] The architecture of the signaling pathway correlative to glycolysis or gluconeogenesis during the diauxic shift is properly built up. With the dynamic system approach, not only the regulatory abilities between causal genes could be derived, but also the delays of regulatory activity are specified. These quantitative characteristics will help determine the intrinsic frameworks of connectivity in the above interesting pathways from the system biology point of view.

Results

The proposed methods in this study would be divided into four steps. In the first step, a dynamic model using the second-order differential equation is developed to describe the expression profile data as output and the regulatory function as input to denote the implicit characteristics of each gene with some parameters. With the help of the second-order dynamic model, we would then extract the upstream regulatory function from the expression profile of the target gene using the optimal estimation method. In the third step, the regulatory function estimated will help seek the correlative regulatory signals from the upstream paths. Iteratively, we can reconstruct the whole signaling regulatory pathway by linking up the upstream regulatory paths. Finally, some biological filters using available biological knowledge are employed to prune the constructed signaling regulatory pathway to improve the accuracy of the proposed method.

1. Dynamic system description of signaling regulatory model

The second-order differential equation is well used in the description of dynamic system evolved from the causality of gene regulatory function. Let $X_i(t)$ denote the expression profile of the i -th gene at time point t . The following second-order differential equation is proposed to model the expression level of the i -th gene,

$$\ddot{X}_i(t) + a_i \dot{X}_i(t) + b_i X_i(t) = G_i(t) + \varepsilon_i(t) \tag{1.1}$$

where $G_i(t)$ is the upstream regulatory function to influence the expression profile $X_i(t)$ of the i -th gene while a_i , and b_i are the parameters that characterize the dynamic inherent property of the gene like degradation and oscillation, and $\varepsilon_i(t)$ is the noise of current microarray data or the residue of the model. In general, the second-order differential equation has been widely used to model dynamic systems to characterize efficiently the dynamic properties of damping and resonance of systems in physics and engineering.

Obviously, the clue of upstream regulatory pathways is in $G_i(t)$. Thus, the first step is to detect the upstream regulatory function $G_i(t)$ from both dynamic equation in (1.1) and microarray data. However, to detect the input regulatory function $G_i(t)$ from both equation (1.1) and microarray data directly is not easy. In this situation, a Fourier decomposition technique is employed to decompose $G_i(t)$ as a synthesis of some harmonic sinusoid functions so that the signal detection problem of $G_i(t)$ is reduced to a simple parameter estimation problem.

Accordingly, we can decompose $G_i(t)$ by the following Fourier series,

$$G_i(t) = \sum_{n=0}^N [\alpha_n \cos(n\omega t) + \beta_n \sin(n\omega t)] \tag{1.2}$$

Then the detection of $G_i(t)$ becomes how to estimate the Fourier coefficients of α_n and β_n , which are the magnitudes of different harmonics of $\cos(n\omega t)$ and $\sin(n\omega t)$, for $n = 0, \dots, N$ in equation (1.2). In science and engineering, the Fourier series has been widely employed to synthesize any continuous functions with finite energy. The estimation of α_n , β_n and the detection of $G_i(t)$ in equation (1.2) are given in Methods in the sequel.

As a result of parameter estimation in Methods, the detection $\widehat{G}_i(t)$ of regulatory function $G_i(t)$ could be derived as follows,

$$\widehat{G}_i(t) = \sum_{n=0}^N [\widehat{\alpha}_n \cos(n\omega t) + \widehat{\beta}_n \sin(n\omega t)] \tag{1.3}$$

Since the input regulatory function $G_i(t)$ of a target gene is usually due to the transcriptional binding or some physical interactions from the upstream regulatory genes, in the following, we would trace back to the corresponding regulatory genes from input regulatory function $\widehat{G}_i(t)$ of the target gene.

II. Inference of the regulatory pathway via $\widehat{G}_i(t)$

Apparently the input regulatory function $G_i(t)$ in equation (1.1) contains the driving information for the target gene's expression from the upstream regulatory genes. The identified regulatory function $\widehat{G}_i(t)$ from equation (1.3) could be interpreted as the regulatory connectivity through transcriptional binding or protein-protein interaction imposed on the i -th target gene. Nevertheless, the expression data of protein type which should be considered directly in practice are by now unavailable and unreliable to trace back upstream regulatory genes. Instead, the expression data on mRNA level which is now widely available from microarray assays would make tracing back the upstream regulatory pathway possible under proper assumptions. All along the paper we assume that the expression levels of mRNA transcripts are proportional to the actual number of corresponding proteins in the cell. This assumption is indeed a strong approximation since post-transcription is known to play a very important role in down regulating the number of the transcription factor in the cell.

Before the inference of upstream regulatory genes, it is reasonable to confine the effect of the regulatory genes on the regulated target gene. The saturated activity of expression level reveals that the regulatory ability cannot extend unlimitedly. The sigmoid function is often chosen to express the nonlinear saturation with proper parameters. Here, we apply the sigmoid transformation to represent the 'on' and 'off' activities of the regulatory genes on binding or not to motifs of the target gene. So the regulatory signal $\widetilde{X}_j(t - \tau_j)$ shown below with the parameter set of $\theta_j = \{\gamma, M_j, \tau_j\}$ is the sigmoid transformation of $X_j(t)$, the expression profile of the j -th regulatory gene.

$$\widetilde{X}_j(t - \tau_j) = \frac{1}{1 + e^{-\gamma(X_j(t - \tau_j) - M_j)}} \tag{2.1}$$

where γ is the transition rate, M_j is the mean expression of the j -th regulatory gene's profile, and τ_j is the corresponding signal transduction delay.

The delay activity should be considered in order to describe the signal transduction delay τ_j from the j -th regulatory gene to the target gene. The delay τ_j would be computed by statistical correlation between the regulatory signal $\widetilde{X}_j(t - \tau_j)$ transformed from the j -th regulatory gene and the identified regulatory function $\widehat{G}_i(t)$ of the target gene. The delay τ_j is determined by the following maximum correlation criterion,

$$\tau_j = \max_{\tau} \left\{ r_{\tau} \left[\widetilde{X}_j(t - \tau_j), \widehat{G}_i(t) \right] \right\} \quad (2.2)$$

where r_{τ} is the correlation between $\widetilde{X}_j(t - \tau_j)$ and $\widehat{G}_i(t)$ under variable delay τ . If there are many τ to achieve the maximum correlation in (2.2), then only the smallest one is chosen.

Using the correlation method, we trace back R_i regulatory genes whose regulatory signals $\widetilde{X}_j(t - \tau_j)$ are most correlated with the regulatory function $\widehat{G}_i(t)$ of the i th target gene, i.e. choose R_i genes with maximum correlation but with smaller τ_j in (2.2). The determination of number R_i will be discussed later. Then, we construct the regulatory pathway by tracing back R_i regulatory genes from the identified regulatory function $\widehat{G}_i(t)$ of the target gene as the following kinetic relationship,

$$\widehat{G}_i(t) = c_{i0} + \sum_{j \in R_i} c_{ij} \widetilde{X}_j(t - \tau_j) + e_i(t) \quad (2.3)$$

where c_{ij} is the pathway kinetic parameters from the regulatory gene j to the target gene i , R_i are the searched upstream regulatory genes, the constant c_{i0} represents the basal level to denote the regulatory function other than upstream regulatory genes, and $e_i(t)$ is the residue of the model.

Furthermore, to estimate the pathway kinetic parameters c_{ij} , equation (2.3) for m time points should be written in the following regression form,

$$\widehat{G}_i = B_i \Omega_i + V_i \quad (2.4)$$

where

$$\widehat{G}_i = \begin{bmatrix} \widehat{G}_i(t_1) \\ \widehat{G}_i(t_2) \\ \vdots \\ \widehat{G}_i(t_m) \end{bmatrix}, B_i = \begin{bmatrix} \widetilde{X}_1(t_1 - \tau_1) & \cdots & \cdots & \widetilde{X}_{R_i}(t_1 - \tau_{R_i}) \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \widetilde{X}_1(t_m - \tau_1) & \cdots & \cdots & \widetilde{X}_{R_i}(t_m - \tau_{R_i}) \end{bmatrix}$$

$$, \Omega_i = \begin{bmatrix} c_{i0} \\ c_{i1} \\ \vdots \\ c_{iR_i} \end{bmatrix}, \text{ and } V_i = \begin{bmatrix} e_i(t_1) \\ e_i(t_2) \\ \vdots \\ e_i(t_m) \end{bmatrix}.$$

We assume that each element in the error vector, $e_i(t_k)$, $k = \{1, \dots, m\}$, is an independent random variable with a normal distribution with zero mean and variance σ^2 . By max-

imum likelihood parameters estimation method (see Methods), the estimates of σ^2 and Ω_i are given as follows, which is solved as

$$\sigma^2 = \frac{1}{m} \sum_{k=1}^m \left[\widehat{G}_i - B_i \Omega_i \right]^T \left[\widehat{G}_i - B_i \Omega_i \right] \quad (2.5)$$

and

$$\widehat{\Omega}_i = (B_i^T B_i)^{-1} B_i^T \widehat{G}_i \quad (2.6)$$

It should be noted that with the combination of biological knowledge about the transcriptional factors, protein phosphorylation, post-transcriptional and specific enzyme regulation of target genes, lots of putative and verified genes correlated with the target genes are pruned by this biological filter for the more efficient and accurate searching of R_i upstream regulatory genes in equation (2.3). For example, suppose the expression profile of gene j has a high correlation with regulation function $G_i(t)$ of target gene i . However, if gene j is not a transcription factor, protein phosphorylation, post-transcription or specific enzyme of target gene i , it will be deleted from the candidates of R_i upstream regulatory genes because it may be only a co-expressed gene with the target gene coregulated by the other gene. On the contrary, a verified regulatory gene should be recruited into the candidates even with small correlation with $G_i(t)$.

Finally, we take the well-known Akaike Information Criterion (AIC) into account for determining the number R_i of regulatory signal $\widetilde{X}_j(t - \tau_j)$ [42],

$$AIC = 2 \log \sigma + \frac{2R_i}{m} \quad (2.7)$$

The first term in AIC is the residual variance and the second term R_i is the number of regulatory genes. AIC includes both the estimated residual variance and model complexity in one statistic, which decreases as σ^2 decreases and increases as R_i increases. AIC has been widely employed to determine the complexity of system modeling science and engineering [42]. The optimal number R_i of the upstream regulatory genes will be determined by the minimization of the AIC value in equation(2.7).

Now, for the selected target genes in the interesting pathway, we could search for the optimal R_i upstream regulatory genes by AIC in equation (2.7) after the biological filtering and determine their pathway kinetic parameters c_{ij} of regulatory signal by equation (2.6). After biological filter pruning, if the number of candidates of regulatory genes is still less than R_i determined by AIC, then some

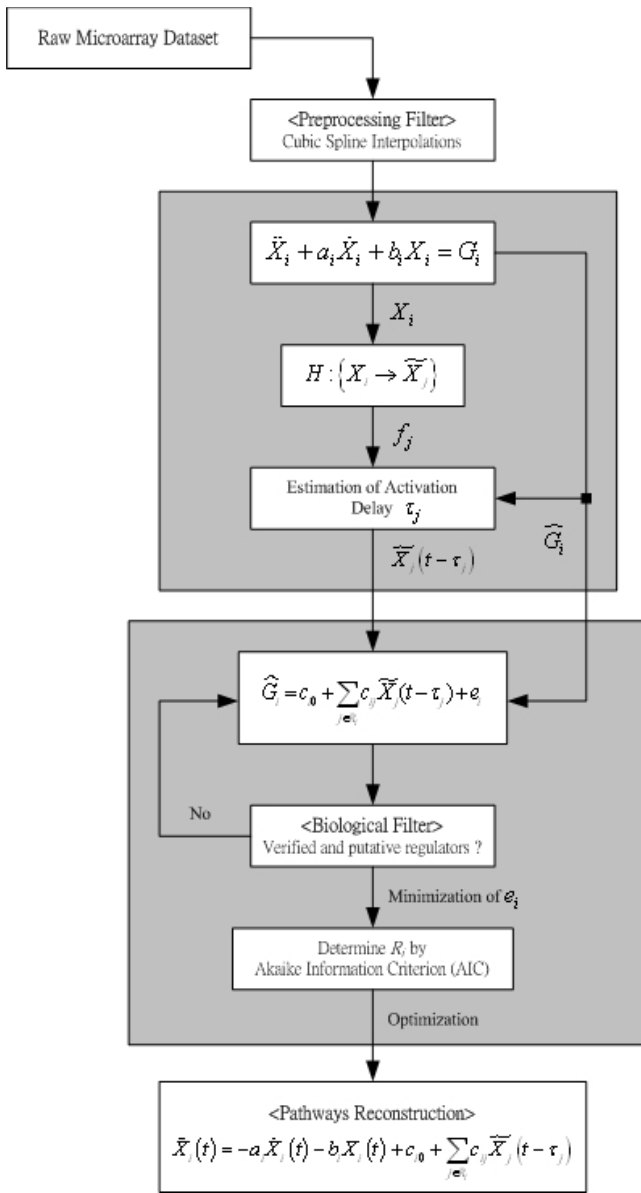


Figure 1
Illustration of the overall flowchart of the pathway inference algorithm.

genes, which are highly correlative to $G_i(t)$ but not of transcription factors or signaling proteins of target gene i , should be recruit into candidates to uncover regulatory relationships that were not suspected to be connected. After the combination of equations (2.3) and (1.1), the whole regulatory pathway is obtained as

$$\ddot{X}_i(t) = -a_i \dot{X}_i(t) - b_i X_i(t) + c_{i0} + \sum_{j \in R_i} c_{ij} \tilde{X}_j(t - \tau_j) \quad (2.8)$$

for $i = \{1, 2, \dots, L\}$, and L is the number of target genes in the pathway. The sub-paths related to the i -th target gene in the interesting pathway could be detected by the inference algorithm. Then, it is natural that the whole regulatory pathway would be constructed by the links of all the sub-paths. We also outline the whole flowchart of our dynamic inferring algorithm as shown in Figure 1 for an overview.

Discussion

Data set of analysis

The two famous modeling organisms, *Arabidopsis thaliana* and yeast *Saccharomyces cerevisiae*, have been well studied biologically and their microarray assays are abundant. Thus, we chose different types of pathways, one is the plant behavior under environmental variation and the other is the cellular metabolism in response to exhaustion of external source, as examples in this study. In other words, two signaling pathways, i.e. circadian regulatory pathway in *Arabidopsis thaliana* and metabolic shift pathway from fermentation to respiration in yeast *Saccharomyces cerevisiae*, are constructed from microarray data to confirm the accuracy of our proposed method.

For cells grown in the light/dark cycle according to circadian rhythm, Harmer and colleagues [15] used highly reproducible oligonucleotide-based arrays representing about 8200 different genes to determine steady-state mRNA levels in *Arabidopsis thaliana* that are measured in replicate hybridization of 12 samples harvested every 4 hours over 2 days. With their investigation on the circadian regulatory system, Harmer et al. have provided an abundance of correlated genes for the regulatory pathway inference.

As for the metabolic pathway, an cDNA microarray assay from DeRisi et al. [30], containing approximately 6400 distinct expression sequence tags (ESTs) in yeast *Saccharomyces cerevisiae*, is harvested at seven successive 2-hour intervals after an initial nine hours of growth under the diauxic shift. Adoption of the diauxic shift data set would make possible the inference of metabolic shift pathways.

Process of raw microarray data

With the second-order equation and the optimal estimation method, the dynamic model should be developed first for the regulatory scheme of target genes in the signaling regulatory pathway. Because the raw microarray data sample of the biological assays that will be analyzed is small with less than 15 data points for an individual gene, the cubic spline method is used to interpolate the

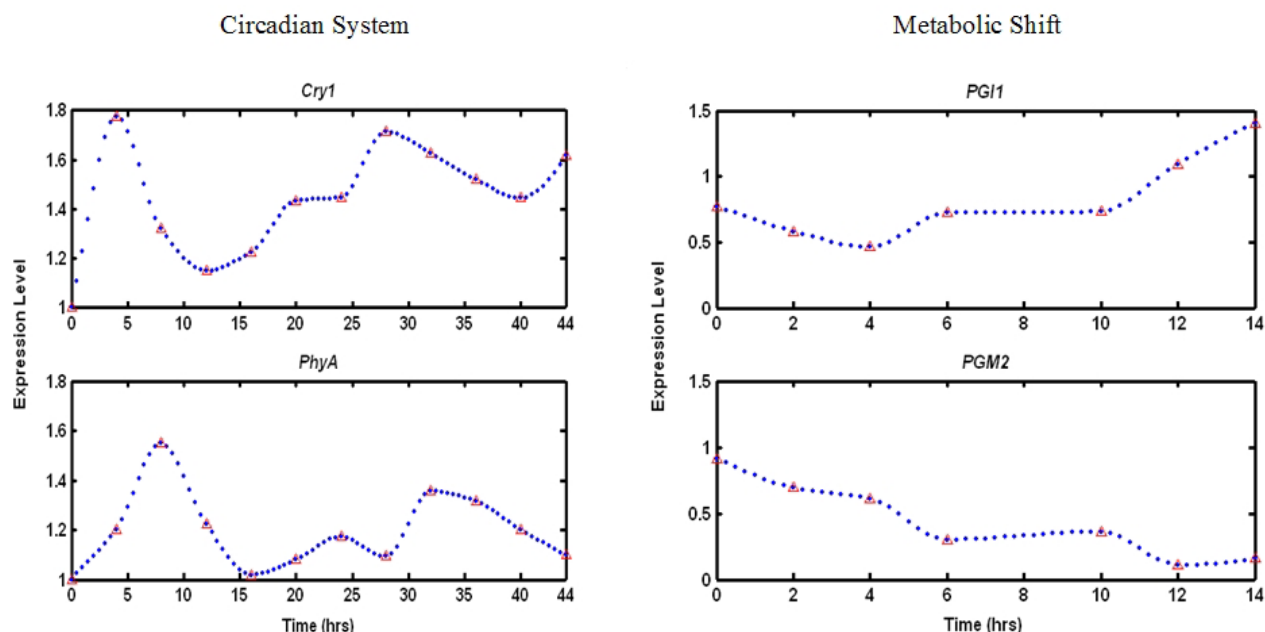


Figure 2
The expression profiles derived from the cubic spline interpolations. The expression profiles after performing the cubic spline interpolations for *Cry1*, *PhyA* (pathway of circadian regulatory system of *Arabidopsis thaliana*) on the left-hand side, and *Pgi1*, *Pgm2* (metabolic shift pathway of yeast *Saccharomyces cerevisiae*) on the right-hand side. The red open triangles are the raw microarray data, and the blue dotted points are the interpolation data.

observed data to increase the data points of each gene's time-course microarray data. As shown in Figure 2, the expression profiles of *Cry1* (CRYTOCHROME 1) and *PhyA* (PHYTOCHROME A) genes in the circadian regulatory pathway of *Arabidopsis thaliana* are interpolated by the cubic spline method among raw data points on the left-hand side. Similarly, *Pgi1* (PHOSPHOGLUCOSE ISOMERASE 1) and *Pgm2* (PHOSPHOGLUCOMUTASE 2) genes in the metabolic shift pathway of yeast *Saccharomyces cerevisiae* are on the right-hand side. After the expression profiles are smoothed by the cubic spline technique, we can obtain the data of the first derivative $\dot{X}_i(t)$ and the second derivative $\ddot{X}_i(t)$ more accurately and abundantly.

Extraction of regulatory information

After data expansion by the cubic spline method, we would have enough data to estimate the parameters of the regulatory dynamic model of the target gene from equation (2.4). Following the dynamic model in equation (3.1), the parameters which characterize the dynamic regulatory mechanism are estimated successfully for each target gene in the pathway. By dynamic model fitting, the expression profiles of the mentioned genes in Figure 2 can be reconstructed in Figure 3 with time progression again.

Hence, we not only could predict the dynamic evolution of the target gene's expression profile accurately, but also deduce the regulatory function $\widehat{G}_i(t)$ simultaneously as the scheme of Figure 4. The regulatory information between target genes and their upstream genes can be extracted properly with this method.

Inference of the regulatory pathway

For illustrations, the inferring strategy is applied to the selected core genes ($X_1 \sim X_{13}$ and $Y_1 \sim Y_{11}$) in two pathways of the circadian regulatory system in *Arabidopsis thaliana* and the metabolic shift pathway in yeast *Saccharomyces cerevisiae* to recognize their upstream regulatory genes, respectively. Their regulatory abilities with signal transduction delays are shown in the form of dynamic equation in Table 1 and Table 2, respectively. These regulatory abilities implying different degrees of influence are converted into a red-colored line as positive regulation (activation) and a blue-colored line as negative regulation (inhibition) for each target gene. Then, according to the dynamic regulatory equations in Table 1 and Table 2, the pathways of the circadian regulatory system and the metabolic shift pathway are described in Figure 5 and Figure 6, respectively.

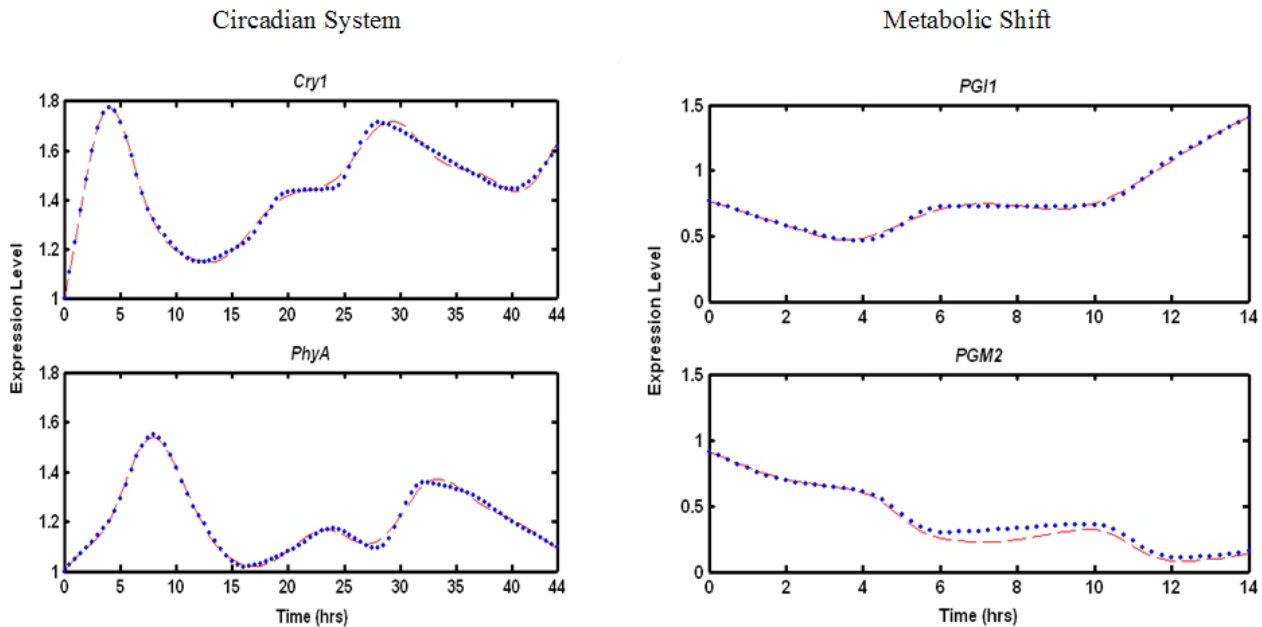


Figure 3
The second-order dynamic model fitting of pathway genes. The expression profile with corresponding second-order dynamic model fitting for *Cry1*, *PhyA* (pathway of circadian regulatory system of *Arabidopsis thaliana*) on the left-hand side, and *PGI1*, *PGM2* (metabolic shift pathway of yeast *Saccharomyces cerevisiae*) on the right-hand side. The blue dotted points are the cubic spline interpolations of microarray data, and the red dashed lines are the estimated dynamic evolution of expression data.

a. Pathway of circadian regulatory system

The circadian rhythm controls processes ranging from cyanobacteria cell division to human wake-sleep cycles. In plant, especially for *Arabidopsis thaliana*, the growth and development have adapted to the diurnal cycling of light and dark [28,31,32,42,44,46-49]. The ability of plants to respond to light is achieved through some photoreceptors. Two classes of photoreceptors are well known to form the photo-transduction pathway under the circadian regulatory system in *Arabidopsis thaliana* [50]. One is the cryptochrome of blue-light photoreceptors, containing *Cry1* and *Cry2*. The other is the phytochrome of mainly red-light photoreceptors, including *PhyA*, *PhyB*, *PhyD* and *PhyE*.

In the photo-transduction related genes (Table 1 and Figure 5), containing both cryptochrome (*Cry1* and *Cry2*) and phytochrome (*PhyA*, *PhyB*, *PhyD* and *PhyE*), *Cry1* [X_6] and *Cry2* [X_{10}] are commonly regulated by *Lhy* [X_3] (LATE ELONGATED HYPOCOTYL) in reciprocal ways with significant values (0.7569 in Eq.(6) and -1.8773, Eq.(10) of Table 1, respectively), implying the essentially regulatory role of *Lhy* on cryptochrome genes. In addition, from

Eq.(10) in Table 1, we further observe that *Ccal* [X_4] (CIRCADIAN CLOCK ASSOCIATED 1) has the greatest positive regulation (2.3465) on *Cry2*, meaning that *Cry2* is jointly regulated by *Lhy* and *Ccal*. Because the binding sites of *Lhy* and *Ccal* found in the promoter regions of *Cry2* [51] are consistent with our inference, the transcriptional binding might be the mechanism of *Cry2* affected by both *Lhy* and *Ccal*. In addition, the mutual activations of phosphorylation between *Cry1* [X_6] and *PhyA* [X_7] in Eq.(6) and Eq.(7) of Table 1 are specifically identified consistent with the previous work [52]. At present, little is known about the nature of interactions between these two classes of photoreceptors. From Eq.(10) in Table 1, *Cry2* [X_{10}] is also positively regulated by *PhyA* [X_7] with 0.5-hr activation delay similar to that in *Cry1* (Eq.(6) in Table 1). Therefore, *PhyA* is considered as a post-transcriptional regulator of phosphorylation to cryptochrome within 1.0-hr after transcription. On the other hand, *PhyB* [X_{11}] down-regulates *Cry2* with a significant effect (-0.7141) while *Cry2* [X_{10}] up-regulates *PhyB* (0.0511) weakly by feedback (see Eqs.(10), (11) in Table 1.). The mutual interactions between *Cry2* and *PhyB* in nuclear speckles that are formed in a light-dependent fashion are also confirmed

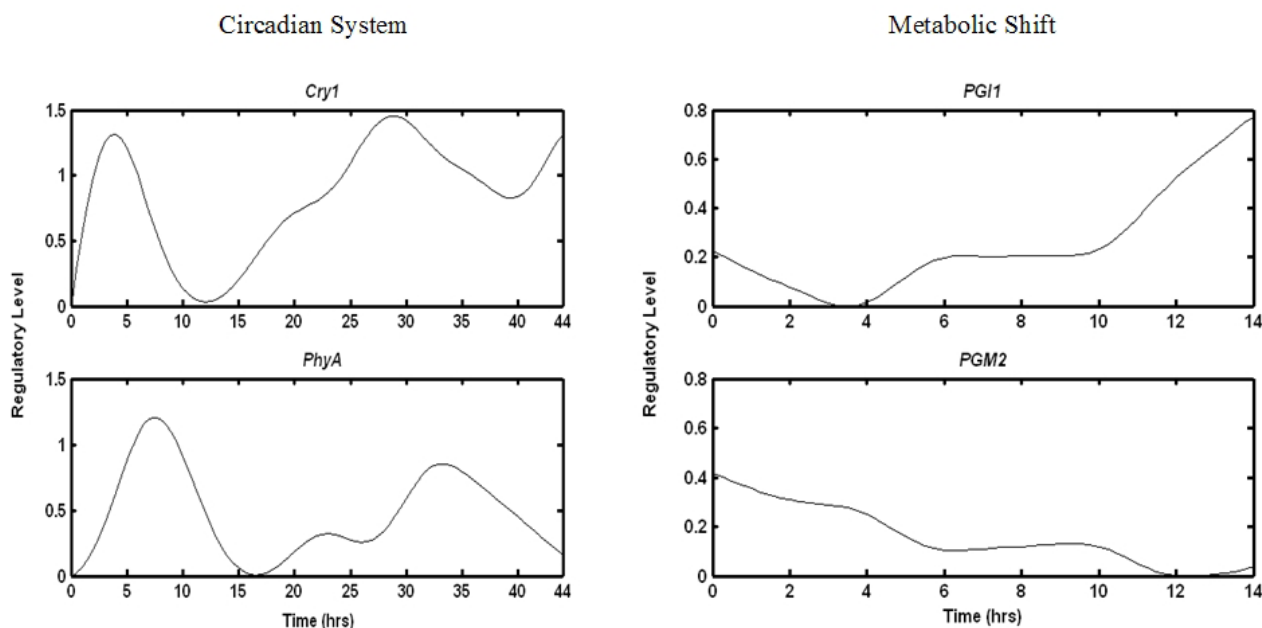


Figure 4

The extracted upstream regulatory functions of pathway genes. The upstream regulatory function $\hat{G}(t)$ extracted from expression profiles of corresponding target genes *Cry1*, *PhyA* (pathway of circadian regulatory system of *Arabidopsis thaliana*) on the left-hand side, and *PGI1*, *PGM2* (metabolic shift pathway of yeast *Saccharomyces cerevisiae*) on the right-hand side.

by Mas *et al.* [48]. Because *Cry1* and *Cry2* are both negatively co-regulated by *PhyD* [X_8] and *PhyE* [X_{12}] significantly (see Eqs.(6), (10) in Table 1), *PhyA* has apparently different behavior from *PhyB*, *PhyD*, and *PhyE* in activating cryptochrome. This might suggest the mechanism that *PhyA* mediates the blue light by up-regulating *Cry1* and *Cry2*, whilst *PhyB*, *PhyD*, and *PhyE* would mediate the red light by inhibiting blue photoreceptors [53,54].

In the mainly red-light photoreceptors of phytochrome (*PhyA*, *PhyB*, *PhyD* and *PhyE*) in Figure 5, undoubtedly *Lhy* [X_3] and *Ccal* [X_4], well-known biological clock genes in the circadian system [40,46], are core regulators involved in the transcriptions of both phytochrome (see Eqs.(7), (8), (11), and (12) in Table 1) and cryptochrome (see Eqs.(6), (10) in Table 1) via feedback transcriptional binding. Similarly, *Gi* [X_{15}] (GIGANTEA) in Figure 5 has been identified as a manifested regulator to all the phytochromes (also see Eqs.(7), (8), (11), and (12) in Table 1), although *Gi* sequence lacks any motifs suggesting that it is a transcription factor of phytochromes [55]. Hence, *Gi* might be a post-transcriptional regulatory factor. However, there is another gene *Elf3* [X_{16}] (EARLY FLOWERING 3) opposite to *Gi* on phytochrome, especially for *PhyA*,

PhyB and *PhyE* (Eqs.(7), (11) and (12) in Table 1). Because of lower regulatory ability than transcription factor *Lhy* or *Ccal*, *Elf3* might play the same role as *Lhy* and *Ccal*. Just as expected, *Elf3* contains glutamine-rich motif suggesting that it is a transcription factor [56].

Before entrance of the biological oscillator of the circadian system formed by *Toc1*, *Lhy*, and *Ccal*, a crucial gene of *Pif3* [X_9] (Figure 5) is mediated significantly by *PhyA* [X_7] (-0.7631) and *PhyB* [X_{11}] (0.1223) (see Eq.(9) in Table 1). This is consistent with the post-transcriptional interactions of *Pif3*-*PhyA* and *Pif3*-*PhyB* complexes. As a core gene in the biological oscillator, *Toc1* [X_{13}] is transcriptionally regulated by *Lhy* [X_3] (0.7009) and *Ccal* [X_4] (-1.4704) whilst *Pif3* [X_{16}] (-0.1698) is presumably considered as the bridge between *Toc1* and phytochrome (Eq.(13), Table 1). From *Lhy* and *Ccal* point of view, they are both positively affected simultaneously by *Pif3* implying the regulation on the transcriptional level [57]. In addition, *Toc1* inhibits both *Lhy* and *Ccal* to form the structure of mutual transcriptional regulation (please compare Eqs.(3), (4) with Eq.(13) in Table 1). So we conclude that *Lhy* and *Ccal* function as principal transcription factors.

Table 1: The dynamic equation set of the identified upstream regulators and their regulatory relationships to the specific target genes in the pathway of circadian regulatory system of Arabidopsis thaliana.

$$\begin{aligned}
 (1) \dot{X}_1(t) &= -0.4270\dot{X}_1(t) - 1.2630X_1(t) + 0.3690 - 0.1033 \cdot \bar{X}_4(t-2.5) - 0.2231 \cdot \bar{X}_{13}(t-3.5) \\
 (2) \dot{X}_2(t) &= -0.3951\dot{X}_2(t) - 1.3000X_2(t) + 0.6585 + 10.4196 \cdot \bar{X}_4(t-6.0) - 0.6226 \cdot \bar{X}_{13}(t-3.0) - 9.4639 \cdot \bar{X}_3(t-6.5) \\
 (3) \dot{X}_3(t) &= -0.3696\dot{X}_3(t) - 1.1202X_3(t) + 0.5962 + 0.8880 \cdot \bar{X}_9(t-3.5) - 0.2503 \cdot \bar{X}_4(t-7.0) - 0.3263 \cdot \bar{X}_{13}(t-0.5) \\
 (4) \dot{X}_4(t) &= -0.3757\dot{X}_4(t) - 1.3057X_4(t) + 0.3627 + 0.7002 \cdot \bar{X}_9(t-0.5) - 0.3049 \cdot \bar{X}_{13}(t-1.0) - 0.5689 \cdot \bar{X}_3(t-0.5) \\
 (5) \dot{X}_5(t) &= -0.2034\dot{X}_5(t) - 1.0242X_5(t) + 0.3691 + 2.7078 \cdot \bar{X}_4(t-6.0) - 0.0455 \cdot \bar{X}_1(t-3.5) - 0.1218 \cdot \bar{X}_{13}(t-2.5) - 2.3586 \cdot \bar{X}_3(t-7.0) \\
 (6) \dot{X}_6(t) &= -0.3307\dot{X}_6(t) - 0.9893X_6(t) + 0.1251 + 0.7569 \cdot \bar{X}_3(t-2.0) + 0.2358 \cdot \bar{X}_7(t-1.0) - 0.2020 \cdot \bar{X}_{12}(t-0.5) - 0.2116 \cdot \bar{X}_{14}(t-1.5) - 0.3788 \cdot \bar{X}_6(t-7.0) \\
 (7) \dot{X}_7(t) &= -0.3465\dot{X}_7(t) - 1.0682X_7(t) + 0.0718 + 0.5392 \cdot \bar{X}_3(t-4.5) + 0.3095 \cdot \bar{X}_{12}(t-0.5) - 0.2997 \cdot \bar{X}_6(t-3.5) - 0.2098 \cdot \bar{X}_{16}(t-2.0) - 0.7631 \cdot \bar{X}_4(t-4.5) \\
 (8) \dot{X}_8(t) &= -0.2194\dot{X}_8(t) - 1.1127X_8(t) + 0.1881 + 0.3434 \cdot \bar{X}_3(t-1.0) + 0.1719 \cdot \bar{X}_{13}(t-6.0) - 0.0626 \cdot \bar{X}_{16}(t-7.0) - 0.0124 \cdot \bar{X}_6(t-2.0) - 0.1474 \cdot \bar{X}_{14}(t-1.0) - 0.3037 \cdot \bar{X}_4(t-1.0) \\
 (9) \dot{X}_9(t) &= -0.1457\dot{X}_9(t) - 0.9994X_9(t) + 0.3943 + 0.1223 \cdot \bar{X}_{11}(t-1.5) + 0.0779 \cdot \bar{X}_{12}(t-3.5) - 0.0593 \cdot \bar{X}_{16}(t-2.0) - 0.0514 \cdot \bar{X}_6(t-6.0) - 0.0485 \cdot \bar{X}_6(t-0.5) - 0.1001 \cdot \bar{X}_7(t-1.5) \\
 (10) \dot{X}_{10}(t) &= -0.2899\dot{X}_{10}(t) - 0.8381X_{10}(t) + 0.9862 + 2.3465 \cdot \bar{X}_6(t-5.5) + 0.3928 \cdot \bar{X}_7(t-0.5) - 0.0124 \cdot \bar{X}_{16}(t-6.0) - 0.4084 \cdot \bar{X}_{16}(t-7.0) - 0.6551 \cdot \bar{X}_7(t-7.0) - 0.7141 \cdot \bar{X}_{16}(t-1.5) - 1.8773 \cdot \bar{X}_{16}(t-6.5) \\
 (11) \dot{X}_{11}(t) &= -0.1581\dot{X}_{11}(t) - 0.8327X_{11}(t) + 0.1857 + 0.2978 \cdot \bar{X}_6(t-5.0) + 0.1142 \cdot \bar{X}_{16}(t-0.5) + 0.0511 \cdot \bar{X}_{16}(t-6.5) + 0.0021 \cdot \bar{X}_6(t-1.5) - 0.0905 \cdot \bar{X}_{16}(t-1.0) - 0.1199 \cdot \bar{X}_{16}(t-3.5) - 0.1648 \cdot \bar{X}_{16}(t-5.5) \\
 (12) \dot{X}_{12}(t) &= -0.1477\dot{X}_{12}(t) - 1.0848X_{12}(t) + 1.1463 + 2.9121 \cdot \bar{X}_4(t-7.0) + 0.4623 \cdot \bar{X}_{14}(t-0.5) + 0.4572 \cdot \bar{X}_{16}(t-1.0) + 0.2104 \cdot \bar{X}_{16}(t-0.5) + 0.0326 \cdot \bar{X}_6(t-5.0) - 0.3324 \cdot \bar{X}_{16}(t-5.5) - 3.3699 \cdot \bar{X}_{16}(t-7.0) \\
 (13) \dot{X}_{13}(t) &= -0.1163\dot{X}_{13}(t) - 0.8416X_{13}(t) + 0.4262 + 0.7009 \cdot \bar{X}_6(t-4.5) + 0.3677 \cdot \bar{X}_7(t-7.0) + 0.2575 \cdot \bar{X}_{16}(t-7.0) + 0.1284 \cdot \bar{X}_6(t-1.5) - 0.1698 \cdot \bar{X}_6(t-7.0) - 0.2247 \cdot \bar{X}_{16}(t-7.0) - 1.4704 \cdot \bar{X}_{16}(t-4.0)
 \end{aligned}$$

Target Genes	X_1 Pap1*	X_2 Co	X_3 Lhy*	X_4 Ccal*	X_5 Chs	X_6 Cryl	X_7 PhyA	X_8 PhyD
	X_9 Pif3	X_{10} Cry2	X_{11} PhyB	X_{12} PhyE	X_{13} Tocl*			
Other Genes	X_{14} Fklf	X_{15} Gi	X_{16} Elf3					

We also infer some downstream pathways of *Chs* [X_5] (CHALCONE SYNTHASE), *Pap1* [X_1], and *Co* [X_2] (CONSTANS) in Figure 5. *Chs* is known as correlated with UV-B protection. It seems that *Ccal* and *Lhy* have greater effect (2.7078, -0.7631, respectively) on *Chs* than *Pap1* (-0.0455) as a transcription factor (Eq.(5) in Table 1). This might mean that *Chs* is regulated by *Pap1* in a small scale with amplifying effect on the *cis*-regulatory level. *Co* is recognized as a pivotal gene of photoperiodic regulation of flowering. Indeed, strong regulations from *Ccal* and *Lhy* are identified to show that *Co* is regulated with a large-scale attenuation effect on the *cis*-regulatory level (Eq.(2) in Table 1).

In the overview of the circadian system in Figure 5, most red lines of activating regulation are found in the photo-transduction pathway between phytochrome (light blue ovals) and crytochrome (light yellow ovals) implying the chain interactions after the external light input. By the feedback regulations of *Lhy* and *Ccal* (orange ovals), represented by black lines with more linking to upstream genes, the photo-transduction pathways are stabilized to provide oscillation. On the other hand, more blue lines of

inhibitive interactions are revealed in the biological-clock regulatory pathways relevant to *Co*, *Pap1*, and *Chs* (light green ovals) underlying the anti-phase functional regulation between these output pathways and the oscillator. In addition, the essential signal transduction factors of *Fklf1*, *Gi*, *Elf3*, and *Pif3* (gray ovals) make some critical links between the functional blocks mentioned above in the circadian system[58]. Finally, in order to validate the proposed approach, an independent validation is also given by randomly reshuffling the time order of microarray experiment [see Additional file 2] but with the same choices of target gene and regulatory genes, as shown in Figure 7. It is seen that the proposed circadian regulatory pathway in Figure 5 is destroyed by reshuffling the experimental data.

b. Metabolic shift pathway

Sugars, such as glucose and sucrose, are excellent carbon sources for yeasts and almost all of the energy requirements of the cell can be satisfied by glycolysis [6,45,59-61,63-66]. *Saccharomyces cerevisiae* can switch from fermentation at high levels of glucose to respiration at low levels of glucose with major changes in metabolic

Table 2: The dynamic equation set of the identified upstream regulators and their regulatory relationships to the specific target genes in the metabolic shift pathway of yeast *Saccharomyces cerevisiae*.

(1) $\dot{Y}_1(t) = -1.3288\ddot{Y}_1(t) - 7.8425\dot{Y}_1(t) + 2.5833 - 0.3284 \cdot \ddot{Y}_{22}(t - 1.50) - 1.9639 \cdot \ddot{Y}_{11}(t - 0.25)$

(2) $\dot{Y}_2(t) = -1.6388\ddot{Y}_2(t) - 10.7990\dot{Y}_2(t) + 1.2717 + 0.2147 \cdot \ddot{Y}_{31}(t - 4.00) - 0.7195 \cdot \ddot{Y}_{26}(t - 0.25)$

(3) $\dot{Y}_3(t) = -1.5344\ddot{Y}_3(t) - 8.6412\dot{Y}_3(t) - 0.0085 + 1.0746 \cdot \ddot{Y}_7(t - 0.25) - 1.536 \cdot \ddot{Y}_{30}(t - 4.00)$

(4) $\dot{Y}_4(t) = -1.6373\ddot{Y}_4(t) - 10.5480\dot{Y}_4(t) + 0.7200 + 5.9829 \cdot \ddot{Y}_{15}(t - 0.25) - 5.8763 \cdot \ddot{Y}_{26}(t - 0.25)$

(5) $\dot{Y}_5(t) = -1.7960\ddot{Y}_5(t) - 10.5470\dot{Y}_5(t) + 1.5690 + 0.3532 \cdot \ddot{Y}_{27}(t - 2.25) + 0.2417 \cdot \ddot{Y}_{28}(t - 0.75) - 0.5770 \cdot \ddot{Y}_{29}(t - 3.50) - 0.8506 \cdot \ddot{Y}_9(t - 0.25)$

(6) $\dot{Y}_6(t) = -1.5074\ddot{Y}_6(t) - 10.4820\dot{Y}_6(t) + 0.4572 + 3.1295 \cdot \ddot{Y}_4(t - 0.25) + 0.1164 \cdot \ddot{Y}_{17}(t - 1.00) - 0.0932 \cdot \ddot{Y}_{26}(t - 0.25) - 2.5615 \cdot \ddot{Y}_{13}(t - 0.25)$

(7) $\dot{Y}_7(t) = -1.8942\ddot{Y}_7(t) - 7.1791\dot{Y}_7(t) + 0.3671 + 0.8011 \cdot \ddot{Y}_{25}(t - 4.00) + 0.5045 \cdot \ddot{Y}_3(t - 0.25) + 0.3484 \cdot \ddot{Y}_6(t - 1.25) + 0.2292 \cdot \ddot{Y}_{24}(t - 2.00) - 1.3287 \cdot \ddot{Y}_{18}(t - 1.75)$

(8) $\dot{Y}_8(t) = -1.5560\ddot{Y}_8(t) - 11.8410\dot{Y}_8(t) + 0.3077 + 5.3057 \cdot \ddot{Y}_2(t - 0.25) + 5.2530 \cdot \ddot{Y}_{16}(t - 2.50) + 0.3503 \cdot \ddot{Y}_{17}(t - 1.25) - 5.0414 \cdot \ddot{Y}_{13}(t - 0.25) - 5.1767 \cdot \ddot{Y}_{23}(t - 4.00)$

(9) $\dot{Y}_9(t) = -1.7344\ddot{Y}_9(t) - 10.0450\dot{Y}_9(t) + 1.0756 + 4.7689 \cdot \ddot{Y}_{10}(t - 2.75) + 1.1252 \cdot \ddot{Y}_{11}(t - 4.25) + 0.3681 \cdot \ddot{Y}_{12}(t - 0.25) - 0.3320 \cdot \ddot{Y}_{11}(t - 4.25) - 0.4213 \cdot \ddot{Y}_{11}(t - 0.25) - 5.4206 \cdot \ddot{Y}_3(t - 3.00)$

(10) $\dot{Y}_{10}(t) = -1.9000\ddot{Y}_{10}(t) - 12.0970\dot{Y}_{10}(t) + 0.3121 + 9.9312 \cdot \ddot{Y}_{10}(t - 0.25) + 2.7883 \cdot \ddot{Y}_{11}(t - 0.25) + 0.3995 \cdot \ddot{Y}_{11}(t - 2.00) + 0.1047 \cdot \ddot{Y}_{10}(t - 0.75) - 0.0288 \cdot \ddot{Y}_{10}(t - 4.50) - 5.9365 \cdot \ddot{Y}_{14}(t - 0.25) - 6.5499 \cdot \ddot{Y}_{11}(t - 0.25)$

(11) $\dot{Y}_{11}(t) = -1.6357\ddot{Y}_{11}(t) - 12.1860\dot{Y}_{11}(t) - 0.1556 + 9.9329 \cdot \ddot{Y}_{10}(t - 0.25) + 2.7954 \cdot \ddot{Y}_{10}(t - 3.00) + 2.2388 \cdot \ddot{Y}_{10}(t - 5.50) + 0.9758 \cdot \ddot{Y}_{10}(t - 1.25) - 4.2732 \cdot \ddot{Y}_{10}(t - 5.00) - 9.6430 \cdot \ddot{Y}_{10}(t - 0.25) - 14.4732 \cdot \ddot{Y}_{10}(t - 0.25) - 15.9251 \cdot \ddot{Y}_{10}(t - 0.25)$

Target Genes	Y_1 Fbpl	Y_2 Eno2	Y_3 Adh2	Y_4 Pykl	Y_5 Pgm2	Y_6 Pdc1	Y_7 Adh1	Y_8 Gpm1
	Y_9 Pgil	Y_{10} Tpil	Y_{11} Fbal					
Other Genes	Y_{12} Tdh1	Y_{13} Tdh2	Y_{14} Tdh3	Y_{15} Gcr1*	Y_{16} Gcr2*	Y_{17} Rap1*	Y_{18} Abf1*	Y_{19} Grf1O*
	Y_{20} Ino4*	Y_{21} Pglk	Y_{22} Pfk2	Y_{23} Hsf1*	Y_{24} Pdc5	Y_{25} Pdc6	Y_{26} Pck1	Y_{27} Glik
	Y_{28} Hxkl	Y_{29} Hxk2	Y_{30} Sfp1*	Y_{31} Stp2*				

activity (diauxic shift). In their experiment on the diauxic shift [30], DeRisi et al. inoculated cells from an exponentially growing culture into fresh medium and grew them at 30 for 21 hrs. This offers a resource to infer the possible allosteric regulation of enzymatic activities, protein modification and transcriptional regulation as shown in Table 2. In addition, the scheme of the corresponding inferred pathway is shown in Figure 6. In the overview of the inferring relationships in Table 2, the gluconeogenesis from *Pyk1* to *pgm2* and the partial fermentation from *Pyk1* (PYRUVATE KINASE 1) to *Adh1* and *Adh2* (ETHANOL DEHYDROGENASE ISOZYME 1, 2) are unraveled as a result of the diauxic shift, so two sub-pathways in opposite directions are concluded.

In the fermentation direction, *Pykl* [Y_4] encoding an enzyme, which catalyzes PEP (Phosphoenolpyruvate) to pyruvate, is negatively regulated (-5.8763) by *Pck1* [Y_{26}] (Eq.(4) in Table 2). *Pck1* could be interpreted here as an indirect upstream transcription factor or regulatory gene for *Pyk1* due to its function of decarboxylation and phosphorylation of oxalacetat in the presence of a nucleoside triphosphate and a divalent metal ion to yield PEP. Another *Gcr1* [Y_{15}] gene is also identified as the strongest

positive regulation (5.9829) to *Pyk1* (also see Eq.(4) in Table 2), which is putatively considered as a transcription factor. This candidate transcription factor *Gcr1* of *Pdc1* [Y_6] (PYRUVATE DECARBOXYLASE ISOZYME 1) plays a more essential role (-2.5615, Eq.(6) in Table 2) than *Rap1* [Y_{17}] (0.1164), and *Pyk1* [Y_4] is an upstream regulatory factor coding an enzyme with the most positive effect (3.1295) on *Pdc1* according to the production of acetaldehyde from pyruvate. In the last kernel of the fermentation, *Adh1* [Y_7] and *Adh2* [Y_3] are involved in the ethanol metabolism of carbohydrate storage. *Adh2* is implicated to up-regulate *Adh1* (0.5145, Eq.(7) in Table 2) under the catabolism from ethanol to acetaldehyde and is significantly up-regulated by *Adh1* (1.0746, Eq.(3) in Table 2) to produce ethanol reversely. The mutual regulations of these two isozymes are within a tiny activation delay of 0.5-hr implying their close relationship. In addition, *Gcr2* [Y_{16}] and *Sfp1* [Y_{30}] with consistently dominant negative influences on *Adh2* and *Adh1* respectively would be at the transcriptional level presumably (see Eqs.(3), (7) in Table 2).

In the sub-pathway of glyconeogenesis, *Eno2* [Y_2] (ENO-LASE ISOZYME 2) is regulated by *Pck1* [Y_{26}] (-0.7195) in the same way as *Pyk1* while the main transcription factor

is *Stp2* [Y_{31}] with significantly positive regulation (0.2147, see Eq.(2) in Table 2). As seen in Table 2, a causal cascade of *Eno2*, *Gpm1* [Y_8] (PHOSPHOGLYCERATE MUTASE 1), *Tpi1* [Y_{10}] (TRIOSE-P ISOMERASE 1), *Fba1* [Y_{11}] (ALDOLASE 1), and *Pgi1* [Y_9] (PHOSPHOGLUCOSE ISOMERASE 1) indicates the construction of a trunk of the glyconeogenesis (see Eqs.(8), (9), (10), and (11) in Table 2). Among them, *Rap1* [Y_{17}] and *Gcr1* [Y_{15}] are the common regulators of *Gpm1*, *Pgi1*, and *Tpi1*. This means that *Rap1* and *Gcr1* might be the most important regulators in the glyconeogenesis pathway by the transcriptional binding. Finally, *Pgm2* [Y_5] (PHOSPHOGLUCOMUTASE 2) co-regulated by *Glk1* [Y_{27}] (GLUCOKINASE 1), *Hxk1* [Y_{28}], and *Hxk2* [Y_{29}] (HEXOKINASE 1, 2) significantly confers another pathway leading to the synthesis of UDP-GLU from Glucose-6-P (see Eq.(5) in Table 2).

In the overview of the metabolic shift pathway in Figure 6, extremely significant regulations (vivid red lines or blue lines) from most transcription factors (gray ovals) means that transcriptional regulations are feasibly identified. However, in the fermentation sub-pathway (light blue ovals), the mutual regulations between *Adh1* and *Adh2* are apparent when compared with the obscure relationships in glyconeogenesis (light yellow ovals). Interestingly, three transcription factors *Gcr1*, *Gcr2* and *Rap1* (black-line signals) appear to have very significant effects on the metabolic shift pathway. Finally, in order to validate the proposed method, an independent validation is also given by randomly reshuffling the time order of microarray experiment but with same choices of target gene and regulatory gene, as shown in Figure 8. Obviously, the proposed metabolic shift in Figure 6 is destroyed by reshuffling the experimental data.

Conclusion

Microarray expression analysis by the dynamic system approach offers an opportunity to generate functional regulation interpretation on the genome-wide scale. The crucial ontology behind using dynamic system techniques is that the causality between gene expression profiles could be identified according to the differential equation underlying a dynamic system. Therefore, because the microarray data were harvested with time progression, the simultaneously varied gene expressions implicated in a genetic regulatory system would be detected to infer the regulatory pathways in spite of the versatile interactions such as transcriptional control, protein phosphorylation, or specific enzyme regulation.

The clustering method 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 were used merely for forward probabilistic estimation with 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, which are deterministic with merely two states.

In our dynamic system approach, we not only can link qualitatively the upstream genes to the downstream ones iteratively, but also indicate quantitatively their regulatory relationships, including the regulatory abilities and the activation delays. 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. Moreover, the speculation of activation delays benefits the empirical reference by providing us when the upstream regulatory genes might interact with their target genes. Since any gene can be considered as a target gene to trace back its upstream regulators, these regulators are then considered as target genes to trace back their upstream regulators. Iteratively, the genetic regulatory pathway (or network) can be constructed to the genome-wide. According to the qualitative and quantitative features imbedded, two regulatory pathway examples are characterized as in Figure 5 and Figure 6 for the identification of the proposed method. In addition, using the Akaike Information Criterion (AIC), a proper number of regulatory genes would be affirmed. As a result, many links overlap with well-known regulatory and signaling pathways in the previous literature and several putative ones are also found. Furthermore, the activation or repression relationships inferred via the microarray data would distinctly uncover the overall effect of regulatory interactions among casual genes in pathways on the transcriptional level.

In the two pathways under investigation, we have a more detailed understanding about the regulatory interactions among phytochrome, cytochrome and biological clock in the circadian regulatory system. On the other hand, the sophisticated knowledge of the metabolic pathway after the diauxic shift can be unfolded properly in our analysis. Furthermore, the independent validation of our approach is also given by randomly reshuffling the time order of microarray experiment. We found that the proposed pathways in Figures 5 and 6 are all destroyed as shown in Figures 7 and 8, respectively. The successful analysis of these two pathways implies the development of a valid and high-throughput method. All of the programs have been released [see Additional file 1]

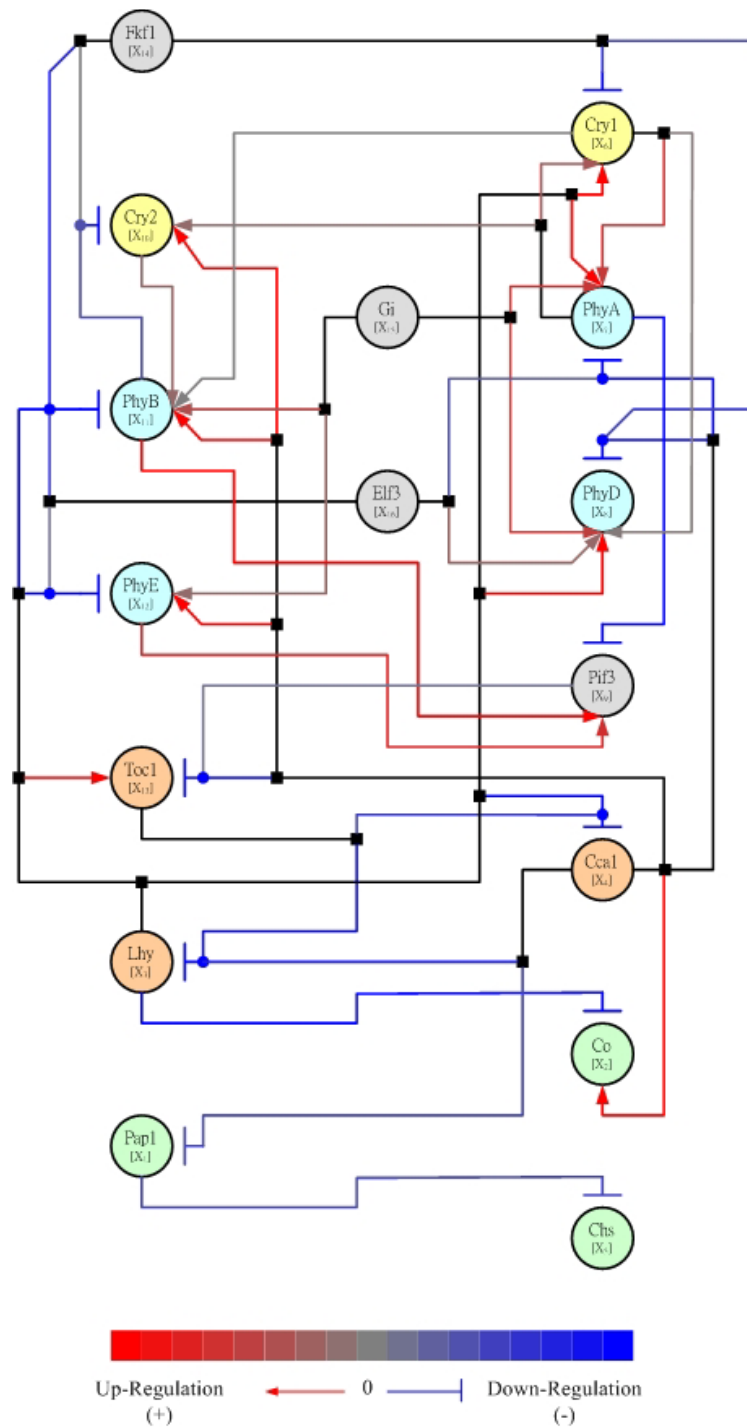


Figure 5
The pathway of circadian regulatory system of *Arabidopsis thaliana* according to the dynamic regulatory modeling in Table I. The related genes are represented as ovals with different colors of light yellow (crytochrome), light blue (phytochrome), orange (biological clock genes), light green (some physiologically light-dependent downstream genes), and gray (other relevant genes not as target genes). There are three types of lines with colors of red (activation), blue (repression), and black. In addition to the black line representing the signal pipes from genes, the red lines are shown depending on the degree of activation whilst the blue lines are shown depending on the degree of repression. The black square symbol attached on the lines is the bifurcate node from the same pipe of signal. The circle symbol attached on the lines is the collecting nodes from different signal sources. The colored degree bar between activation and repression is also shown in the bottom of the figure.

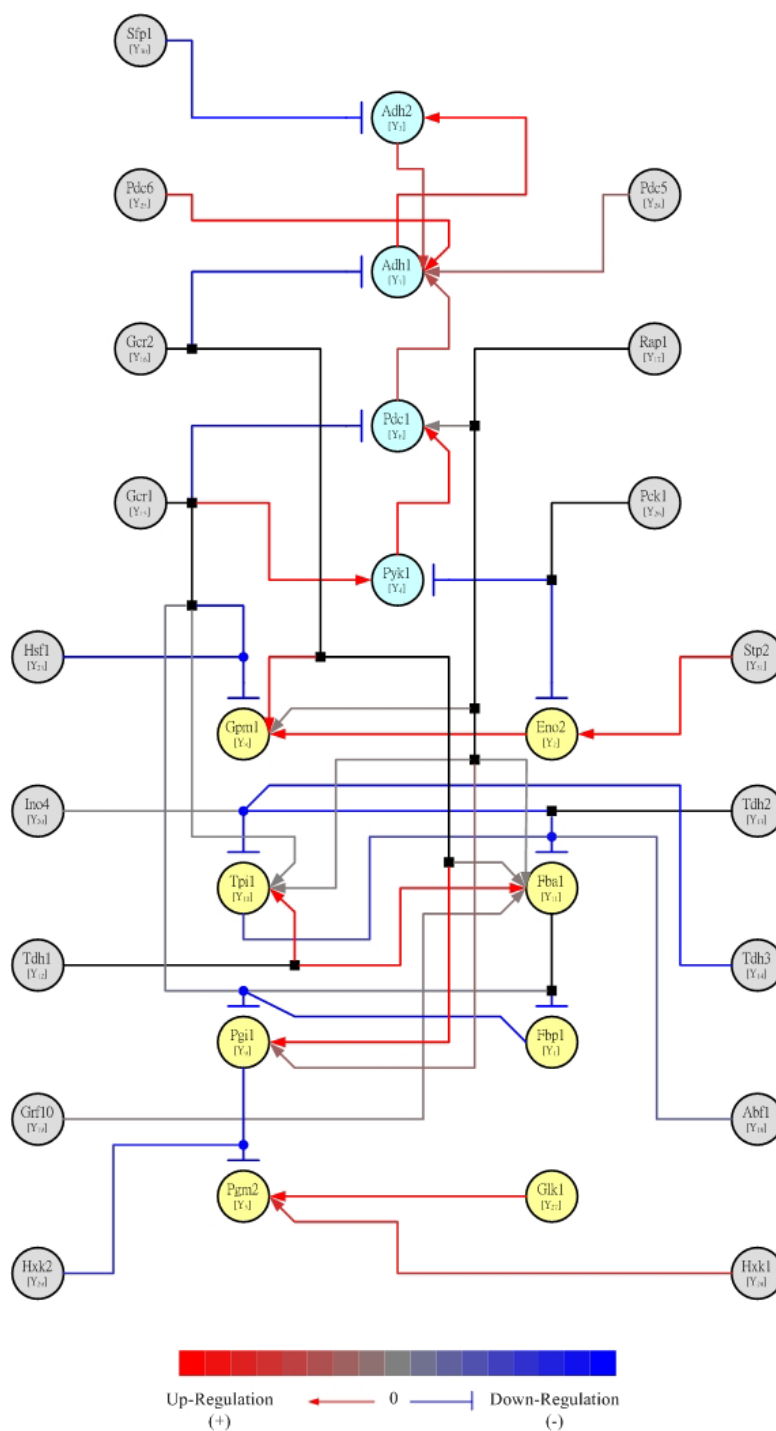


Figure 6
The metabolic shift pathway of yeast *Saccharomyces cerevisiae* according to the dynamic regulatory modeling in Table 2. The related genes are represented as ovals with different colors of light blue (fermentation), light yellow (gluconeogenesis), and gray (other relevant genes not as target genes). There are three types of lines with colors of red (activation), blue (repression), and black. In addition to the black line representing the signal pipes from genes, the red lines are shown depending on the degree of activation whilst the blue lines are shown depending on the degree of repression. The black square symbol attached on the lines is the bifurcate node from the same signal source. And the circle symbol attached on the lines is the collecting nodes from different signal sources. The colored degree bar between activation and repression is also shown in the bottom of the figure.

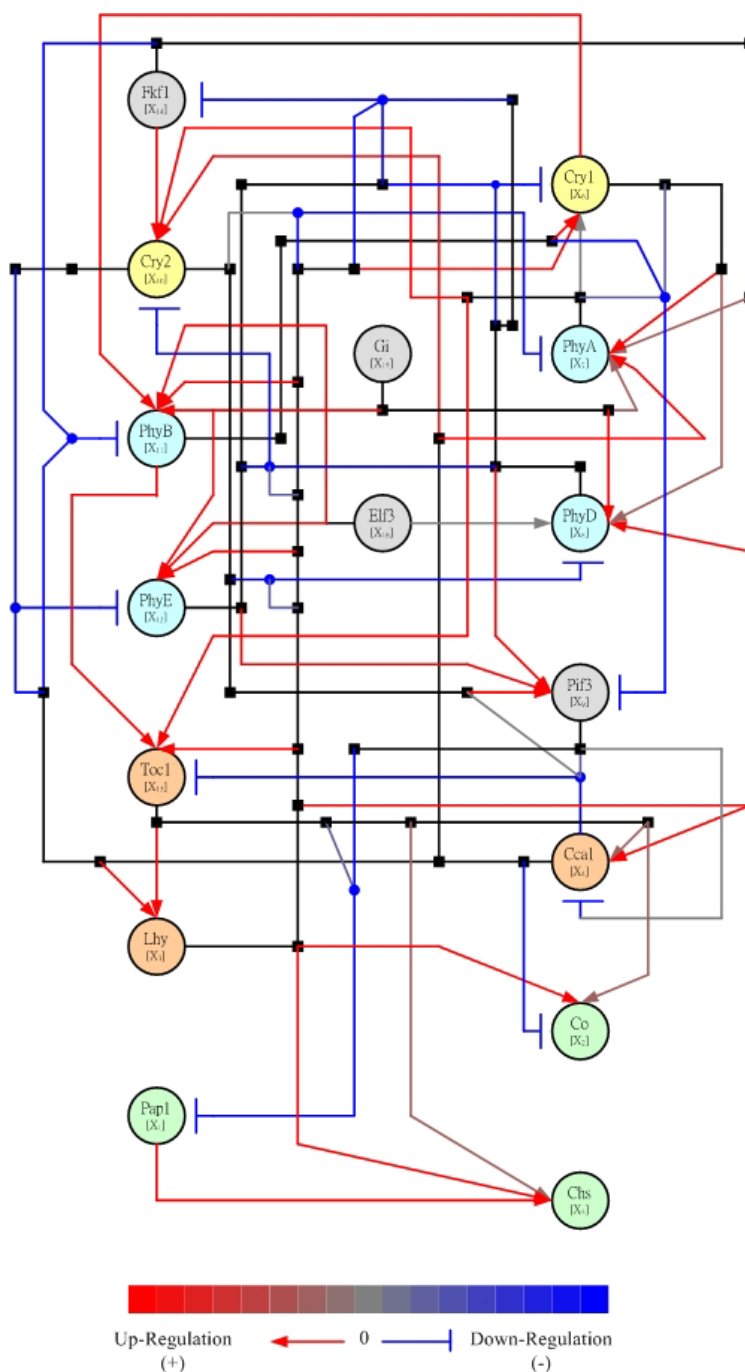


Figure 7
The pathway of circadian regulatory system of Arabidopsis thaliana in Fig. 5 is repeated as independent validation by randomly reshuffling the time order of microarray experiment but with the same choices of target and regulatory genes. Obviously, the proposed pathway in Fig. 5 is destroyed by the reshuffling of experimental data. The related genes are represented as ovals with different colors of light yellow (crytochrome), light blue (phytochrome), orange (biological clock genes), light green (some physiologically light-dependent downstream genes), and gray (other relevant genes not as target genes). There are three types of lines with colors of red (activation), blue (repression), and black. In addition to the black line representing the signal pipes from genes, the red lines are shown depending on the degree of activation whilst the blue lines are shown depending on the degree of repression. The black square symbol attached on the lines is the bifurcate node from the same pipe of signal. The circle symbol attached on the lines is the collecting nodes from different signal sources. The colored degree bar between activation and repression is also shown in the bottom of the figure.

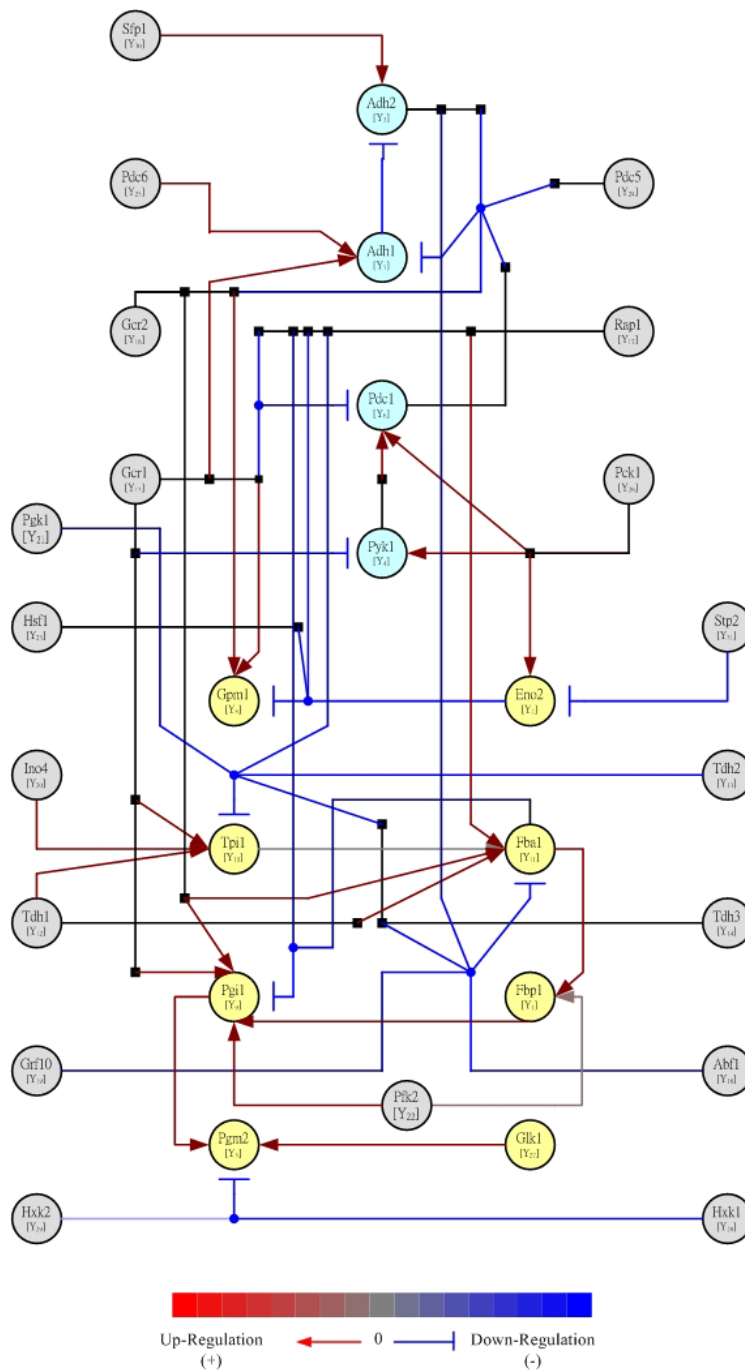


Figure 8
The pathway of metabolic shift regulatory system of yeast *Saccharomyces cerevisiae* in Fig. 6 is repeated as independent validation by randomly reshuffling the time order of microarray experiment but with the same choices of target and regulatory genes. Obviously, the proposed pathway in Fig. 6 is destroyed by the reshuffling of experimental data. The related genes are represented as ovals with different colors of light blue (fermentation), light yellow (gluconeogenesis), and gray (other relevant genes not as target genes). There are three types of lines with colors of red (activation), blue (repression), and black. In addition to the black line representing the signal pipes from genes, the red lines are shown depending on the degree of activation whilst the blue lines are shown depending on the degree of repression. The black square symbol attached on the lines is the bifurcate node from the same signal source. And the circle symbol attached on the lines is the collecting nodes from different signal sources. The colored degree bar between activation and repression is also shown in the bottom of the figure.

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 quick dynamic evolution. A more sampling experiment with respect to the intrinsic turnover rate is expected to have more precise analysis. Secondly, a regulatory gene with larger activation delay would not be recognized because the less activation delay criterion is used, but this might be overcome by properly relaxing the criterion. Thirdly, activation profiles under the proteome should be highly correlated with the transcriptional profiles to elevate the interpretation of our system model. In general, the synchronous time-course microarray assay is more suitable to underlie the transcriptional binding among causal genes, but an inference of physical interactions in the post-transcriptional level also has sufficient feasibility in our study.

In the near future, the most pressing task is to investigate our presumed paths in the laboratory. As the pathway construction algorithms are further developed, we expect this system approach to have immense impact in elucidating the underlying molecular mechanisms of pathways in a variety of organisms, especially after the maturation of the protein chips. Ultimately, we envision that biologists will perform routine pathway inference to seek some novel regulations and to identify the evolutionarily conserved links.

Methods

1. Detection of regulatory function $G_i(t)$ in equation (1.2)

After the decomposition of $G_i(t)$ in equation (1.2), we substitute equation (1.2) into equation (1.1) to obtain the following dynamic equation for the expression profile of the i -th gene,

$$\ddot{X}_i(t) = -a_i \dot{X}_i(t) - b_i X_i(t) + \sum_{n=0}^N [\alpha_n \cos(n\omega t) + \beta_n \sin(n\omega t)] + \varepsilon_i(t) \tag{3.1}$$

In the above dynamic equation, parameters a_i , b_i , α_n , and β_n should be estimated by the time profile of microarray data of the i -th gene, i.e. these parameters should be specified so that the simulating output $X_i(t)$ of the dynamic model in equation (3.1) should meet the empirical expression profile of the i -th gene. The least-squares estimation method is employed to solve this parameter estimation problem.

To make the dynamic model effective, the dynamic equation in (3.1) should meet the expression profile at all time points $t = t_1, \dots, t_m$ and is then arranged in a vector differential form. Consequently, the vector differential form underlined in this equation is applied to m time points in order.

$$\ddot{\underline{X}}_i = -a_i \dot{\underline{X}}_i - b_i \underline{X}_i + \sum_{n=0}^N [\alpha_n \underline{\cos(n)} + \beta_n \underline{\sin(n)}] + \underline{\varepsilon}_i \tag{3.2}$$

where

$$\ddot{\underline{X}}_i = \begin{bmatrix} \ddot{X}_i(t_1) \\ \ddot{X}_i(t_2) \\ \vdots \\ \ddot{X}_i(t_m) \end{bmatrix}, \dot{\underline{X}}_i = \begin{bmatrix} \dot{X}_i(t_1) \\ \dot{X}_i(t_2) \\ \vdots \\ \dot{X}_i(t_m) \end{bmatrix}, \underline{X}_i = \begin{bmatrix} X_i(t_1) \\ X_i(t_2) \\ \vdots \\ X_i(t_m) \end{bmatrix}, \underline{\cos(n)} = \begin{bmatrix} \cos(n\omega t_1) \\ \cos(n\omega t_2) \\ \vdots \\ \cos(n\omega t_m) \end{bmatrix}$$

$$\underline{\sin(n)} = \begin{bmatrix} \sin(n\omega t_1) \\ \sin(n\omega t_2) \\ \vdots \\ \sin(n\omega t_m) \end{bmatrix}, \underline{\varepsilon}_i = \begin{bmatrix} \varepsilon_i(t_1) \\ \varepsilon_i(t_2) \\ \vdots \\ \varepsilon_i(t_m) \end{bmatrix}, \text{ and } m \text{ denotes the}$$

number of time points.

In the next step, formula (3.2) can be translated into a differential matrix equation as follows,

$$Y_i = A_i \Phi_i + E_i \tag{3.3}$$

where $Y_i = \ddot{\underline{X}}_i$, $\Phi_i = [a_i, b_i, \alpha_0, \beta_0, \dots, \alpha_N, \beta_N]^T$, and $E_i = \underline{\varepsilon}_i$ are in vector forms, while

$$A_i = \begin{bmatrix} -\dot{\underline{X}}_i - \underline{X}_i \cos(0) \sin(0) \cdots \cos(N) \sin(N) \end{bmatrix}^T \text{ is a matrix.}$$

To estimate the relevant unknown parameters in Φ_i , the least-squares method below is used to derive the optimal parameters estimation of $\widehat{\Phi}_i$,

$$\widehat{\Phi}_i = \left(A_i^T A_i \right)^{-1} A_i^T Y_i \tag{3.4}$$

Actually, the modeling error could be concluded into E_i as the noise of the gene-expression profile or of the microarray chips. So the consideration of modeling error makes equation (3.3) approach more the reality. By the way, in order to get accurate data of $\ddot{\underline{X}}_i$ and $\dot{\underline{X}}_i$ from the expression profile of the target gene, the cubic spline should be employed to interpolate the time profile of the target gene. Furthermore, the choice of N is based on the trade-off between the accuracy of approximation in (1.2) and the complexity of parameter estimation in (3.4). In this study $N = 6$ is chosen because these harmonics are enough to approximate regulation functions.

II. Maximum likelihood Estimate of kinetic parameters Ω_i in equation (2.4)

Maximum likelihood method for Ω_i in equation (2.4) is given as follows:

$$p(\hat{G}_i | \Omega_i, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{[\hat{G}_i - B_i\Omega_i]^T [\hat{G}_i - B_i\Omega_i]}{2\sigma^2} \right\} \quad (4.1)$$

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

$$L(\Omega_i, \sigma^2) = -\frac{m}{2} \ln[2\pi\sigma^2] - \frac{1}{2\sigma^2} \sum_{k=1}^m [\hat{G}_i - B_i\Omega_i]^T [\hat{G}_i - B_i\Omega_i] \quad (4.2)$$

The necessary condition for the maximum likelihood estimation of variance σ^2 is $\frac{\partial L(\Omega_i, \sigma^2)}{\partial \sigma^2} = 0$, by which equation (2.5) is obtained.

Substituting equation (2.5) into equation (4.2) yields,

$$L(\Omega_i, \sigma^2) = -\frac{m}{2} \ln[2\pi\sigma^2] - \frac{m}{2} \quad (4.3)$$

meaning that we can find the maximum likelihood estimation of Ω_i by minimizing the value of σ^2 in equation(2.5). Then, the maximum likelihood estimate in

equation (2.6) is obtained by $\frac{\partial^2 \sigma}{\partial \Omega_i} = 0$ in equation(2.5).

Authors' contributions

W.C. Chang and Chang-Wei Li carried out the computational studies and analysis. B.S. Chen gave the topic and suggestions. All authors read and approved the final manuscript.

Additional material

Additional File 1

The raw data has been transform to .mat file, which is one of the Matlab files. This file has been divided into three parts: Name, Time, and Profile. [see Additional file 1] Name denotes the names of microarray data. Time denotes the time points of microarray data. Profile denotes the gene profiles of microarray data.

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2105-6-44-S1.rar]

Additional File 2

The 'Shuffled_data' is the Shuffled raw data. [see Additional file 2]

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2105-6-44-S2.xls]

Acknowledgements

We thank the National Science Council, Taiwan for grants NSC 93-3112-B-007-003.

References

- Kettman JR, Frey JR, Lefkovits I: **Proteome, transcriptome and genome: top down or bottom up analysis.** *Biomol Eng* 2001, **18**:207-212.
- Scheel J, Von Brevern MC, Horlein A, Fischer A, Schneider A, Bach A: **Yellow pages to the transcriptome.** *Pharmacogenomics* 2002, **3**:791-807.
- Velculescu VE, Zhang L, Zhou W, Vogelstein J, Basrai MA, Bassett DE Jr, Hieter P, Vogelstein B, Kinzler KW: **Characterization of the yeast transcriptome.** *Cell* 1997, **88**:243-251.
- Eastmond PJ, Graham IA: **Trehalose metabolism: a regulatory role for trehalose-6-phosphate.** *Curr Opin Plant Biol* 2003, **6**:231-235.
- Harkin DP: **Uncovering Functionally Relevant Signaling Pathways Using Microarray-Based Expression Profiles.** *The Oncologist* 2000, **5**:511-517.
- Verdone L, Cesari F, Denis CL, Di Mauro E, Caserta M: **Factors affecting Saccharomyces cerevisiae ADH2 chromatin remodeling and transcription.** *J Biol Chem* 1997, **272**:30828-30834.
- Yanovsky MJ, Kay SA: **Signaling networks in the plant circadian system.** *Curr Opin Plant Biol* 2001, **4**:429-435.
- Sontag E, Kiyatkin A, Kholodenko BN: **Inferring dynamic architecture of cellular networks using time series of gene expression, protein and metabolite data.** *Bioinformatics* 2004, **20**:1877-86.
- Schena M, Shalon D, Davis RW, Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray.** *Science* 1995, **270**:467-470.
- Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ: **High density synthetic oligonucleotide arrays.** *Nat Genet* 1999, **21**:20-24.
- Sherlock G, Hernandez-Boussard T, Kasarskis A, Binkley G, Matese JC, Dwight SS, Kaloper M, Weng S, Jin H, Ball CA, Eisen MB, Spellman PT, Brown PO, Botstein D, Cherry JM: **The Stanford Microarray Database.** *Nucleic Acids Res* 2001, **29**:152-155.
- Edgar R, Domrachev M, Lash AE: **Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.** *Nucleic Acids Res* 2002, **30**:207-210.
- Brazma A, Parkinson H, Sarkans U, Shojatalab M, Vilo J, Abeygunawardena N, Holloway E, Kapushesky M, Kemmeren P, Lara GG, Oezcimen A, Rocca-Serra P, Sansone SA: **ArrayExpress - a public repository for microarray gene expression data at the EBI.** *Nucleic Acids Res* 2003, **31**:68-71.
- Causton HC, Ren B, Koh SS, Harbison CT, Kanin E, Jennings EG, Lee TI, True HL, Lander ES, Young RA: **Remodeling of yeast genome expression in response to environmental changes.** *Mol Biol Cell* 2001, **12**:323-337.
- Harmer SL, Hogenesch JB, Straume M, Chang HS, Han B, Zhu T, Wang X, Kreps JA, Kay SA: **Orchestrated transcription of key pathways in Arabidopsis by the circadian clock.** *Science* 2000, **290**:2110-2113.
- Schaffer R, Landgraf J, Accerbi M, Simon V, Larson M, Wisman E: **Microarray analysis of diurnal and circadian-regulated genes in Arabidopsis.** *Plant Cell* 2001, **13**:113-123.
- Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PQ, Bostein D, Futcher B: **Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization.** *Mol Biol Cell* 1998, **9**:3273-3297.
- Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns.** *Proc Natl Acad Sci USA* 1998, **95**:14863-14868.
- Gasch AP, Eisen MB: **Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering.** *Genome Biol* 2002, **3**:RESEARCH0059.
- Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR: **Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation.** *Proc Natl Acad Sci USA* 1999, **96**:2907-2912.
- Alter O, Brown PO, Botstein D: **Singular value decomposition for genome-wide expression data processing and modeling.** *Proc Natl Acad Sci USA* 2000, **97**:10101-10106.
- Brown MP, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS, Ares M Jr, Haussler D: **Knowledge-based analysis of microarray**

- gene expression data by using support vector machines. *Proc Natl Acad Sci U S A* 2000, **97**:262-267.
23. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM: **Systematic determination of genetic network architecture.** *Nat Genet* 1999, **22**:281-285.
 24. Friedman N, Linial M, Nachman I, Pe'er D: **Using Bayesian networks to analyze expression data.** *J Comput Biol* 2000, **7**:601-620.
 25. Husmeier D: **Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks.** *Bioinformatics* 2003, **19**:2271-82.
 26. Huang S: **Gene expression profiling, genetic networks, and cellular states: an integrating concept for tumorigenesis and drug discovery.** *J Mol Med* 1999, **7**:469-480.
 27. Chen HC, Lee HC, Lin TY, Li WH, Chen BS: **Quantitative characterization of the transcriptional regulatory network in the yeast cell cycle.** *Bioinformatics* 2004, **20**:1914-1927.
 28. Michael TP, McClung CR: **Phase-specific circadian clock regulatory elements in Arabidopsis.** *Plant Physiol* 2002, **130**:627-638.
 29. Motaki S, Ayako K, Kazuko YS, Kazuo S: **Molecular response to drought, salinity and frost: common and different paths for plant protection.** *Curr Opin Biotechnol* 2003, **14**:194-199.
 30. DeRisi JL, Iyer VR, Brown PO: **Exploring the metabolic and genetic control of gene expression on a genomic scale.** *Science* 1997, **278**:680-686.
 31. Staiger D: **Circadian rhythms in Arabidopsis: time for nuclear proteins.** *Planta* 2002, **214**:334-344.
 32. Somers DE, Devlin PF, Kay SA: **Phytochromes and cryptochromes in the entrainment of the Arabidopsis circadian clock.** *Science* 1998, **282**:1488-1490.
 33. Flores CL, Rodriguez C, Petit T, Gancedo C: **Carbohydrate and energy-yielding metabolism in non-conventional yeasts.** *FEMS Microbiol Rev* 2000, **24**:517-529.
 34. Deng XW, Quail PH: **Signalling in light-controlled development.** *Semin Cell Dev Biol* 1999, **10**:121-129.
 35. Eriksson ME, Hanano S, Southern MM, Hall A, Millar AJ: **Response regulator homologues have complementary, light-dependent functions in the Arabidopsis circadian clock.** *Planta* 2003, **218**:159-162.
 36. Green RM, Tingay S, Wang ZY, Tobin EM: **Circadian rhythms confer a higher level of fitness to Arabidopsis plants.** *Plant Physiol* 2002, **129**:576-584.
 37. Hall A, Kozma-Bognar L, Bastow RM, Nagy F, Millar AJ: **Distinct regulation of CAB and PHYB gene expression by similar circadian clocks.** *Plant J* 2002, **32**:529-537.
 38. Hayama R, Coupland G: **Shedding light on the circadian clock and the photoperiodic control of flowering.** *Curr Opin Plant Biol* 2003, **6**:13-19.
 39. Carre IA: **Day-length perception and the photoperiodic regulation of flowering in Arabidopsis.** *J Biol Rhythms* 2001, **16**:415-423.
 40. Alabadi D, Oyama T, Yanovsky MJ, Harmon FG, Mas P, Kay SA: **Reciprocal regulation between TOC1 and LHY/CCA1 within the Arabidopsis circadian clock.** *Science* 2001, **293**:880-883.
 41. Yin Z, Hatton L, Brown AJ: **Differential post-transcriptional regulation of yeast mRNAs in response to high and low glucose concentrations.** *Mol Microbiol* 2000, **35**:553-565.
 42. Johansson R: **System Modeling and Identification.** Prentice-Hall, Englewood Cliffs; 1993.
 43. Mas P, Alabadi D, Yanovsky MJ, Oyama T, Kay SA: **Dual role of TOC1 in the control of circadian and photomorphogenic responses in Arabidopsis.** *Plant Cell* 2003, **15**:223-236.
 44. Mazzella MA, Cerdan PD, Staneloni RJ, Casal JJ: **Hierarchical coupling of phytochromes and cryptochromes reconciles stability and light modulation of Arabidopsis development.** *Development* 2001, **128**:2291-2299.
 45. Ohlmeier S, Kastaniotis AJ, Hiltunen JK, Bergmann U: **The yeast mitochondrial proteome, a study of fermentative and respiratory growth.** *J Biol Chem* 2004, **279**:3956-3979.
 46. Somers DE: **Clock-associated genes in Arabidopsis: a family affair.** *Philos Trans R Soc Lond B Biol Sci* 2001, **356**:1745-1753.
 47. Valverde F, Mouradov A, Soppe W, Ravenscroft D, Samach A, Coupland G: **Photoreceptor regulation of CONSTANS protein in photoperiodic flowering.** *Science* 2004, **303**:1003-1006.
 48. Mas P, Devlin PF, Panda S, Kay SA: **Functional interaction of phytochrome B and cryptochrome 2.** *Nature* 2000, **408**:207-211.
 49. Sharrock RA, Clack T: **Patterns of expression and normalized levels of the five Arabidopsis phytochromes.** *Plant Physiol* 2002, **130**:442-456.
 50. Fankhauser C, Staiger D: **Photoreceptors in Arabidopsis thaliana: light perception, signal transduction and entrainment of the endogenous clock.** *Planta* 2002, **216**:1-16.
 51. Toth R, Kevei E, Hall A, Millar AJ, Nagy F, Kozma-Bognar L: **Orca-dian clock-regulated expression of phytochrome and cryptochrome genes in Arabidopsis.** *Plant Physiol* 2001, **127**:1607-1616.
 52. Ahmad M, Jarillo JA, Smirnova O, Cashmore AR: **The CRY1 blue light photoreceptor of Arabidopsis interacts with phytochrome A in vitro.** *Mol Cell* 1998, **1**:939-948.
 53. Casal JJ: **Phytochromes, cryptochromes, phototropin: photoreceptor interactions in plants.** *Photochem Photobiol* 2000, **71**:1-11.
 54. Hall A, Kozma-Bognar L, Toth R, Nagy F, Millar AJ: **Conditional circadian regulation of PHYTOCHROME A gene expression.** *Plant Physiol* 2001, **127**:1808-1818.
 55. Martinez-Garcia JF, Huq E, Quail PH: **Direct targeting of light signals to a promoter element-bound transcription factor.** *Science* 2000, **288**:859-863.
 56. Carre IA, Kim JY: **MYB transcription factors in the Arabidopsis circadian clock.** *J Exp Bot* 2002, **53**:1551-1557.
 57. Kim JY, Song HR, Taylor BL, Carre IA: **Light-regulated translation mediates gated induction of the Arabidopsis clock protein LHY.** *EMBO J* 2003, **22**:935-944.
 58. Imaizumi T, Tran HG, Swartz TE, Briggs WR, Kay SA: **FKF1 is essential for photoperiodic-specific light signalling in Arabidopsis.** *Nature* 2003, **426**:302-306.
 59. Barnett JA: **A history of research on yeasts 6: the main respiratory pathway.** *Yeast* 2003, **20**:1015-1044.
 60. de Mesquita JF, Zaragoza O, Gancedo JM: **Functional analysis of upstream activating elements in the promoter of the FBPI gene from Saccharomyces cerevisiae.** *Curr Genet* 1998, **33**:406-411.
 61. Dickinson FM, Back S: **The activity of yeast ADH I and ADH II with long-chain alcohols and diols.** *Chem Biol Interact* 2001, **130-132**((1-3)):417-423.
 62. Dickinson JR, Salgado LE, Hewlins MJ: **The catabolism of amino acids to long chain and complex alcohols in Saccharomyces cerevisiae.** *Biol Chem* 2003, **278**:8028-8034.
 63. Fraenkel DG: **The top genes: on the distance from transcript to function in yeast glycolysis.** *Curr Opin Microbiol* 2003, **6**:198-201.
 64. Goncalves PM, Griffioen G, Bebelman JP, Planta RJ: **Signalling pathways leading to transcriptional regulation of genes involved in the activation of glycolysis in yeast.** *Mol Microbiol* 1997, **25**:483-493.
 65. Goncalves P, Planta RJ: **Starting up yeast glycolysis.** *Trends Microbiol* 1998, **6**:314-319.
 66. Scheffler IE, de la Cruz BJ, Prieto S: **Control of mRNA turnover as a mechanism of glucose repression in Saccharomyces cerevisiae.** *Int J Biochem Cell Biol* 1998, **30**:1175-1193.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

