OMICS A Journal of Integrative Biology
Volume 13, Number 4, 2009

DOI: 10.1089/omi.2009.0040

Research Article

A Network Biology Study on Circadian Rhythm by Integrating Various Omics Data

Yong Wang,[1,*] Xiang-Sun Zhang,[1] and Luonan Chen[2,3,*]

Abstract

Circadian rhythm is fundamentally important in physiological processes of mammals. To reveal its underlying mechanism, we probed functional interactions among genes, motivated by the basic molecular observation on gene expression data in circadian rhythm that a large number of genes oscillate in a coordinated manner. In this study, a reverse-engineering strategy was applied to infer and analyze the structure and function of a circadian rhythm-related gene regulatory network. Specifically, our method integrated four phase-shift time-course gene expression datasets in rat suprachiasmatic nucleus, protein–protein interactions, phosphorylations of a set of key circadian genes, and prior information of *cis*-regulatory elements, to construct the gene regulatory network related to circadian rhythm of the rat. By follow-up analysis, we identified four new regulatory hubs that may play crucial roles in the regulation of circadian rhythm. Furthermore, we found that feedback loop motifs were significantly enriched in the predicted network, which may contribute to the genome-wide oscillations of the circadian clock. Compared to the small-scale gene regulatory network conducted by experimental method, our study provides a system-wide overview on the gene regulations, which not only reveals the global network structure but also gives valuable insights into the essential mechanism of circadian rhythm.

Introduction

Recent biotechnology advances generate multiple types of omics data, such as genomic, transcriptomic, proteomic, and metabolomic datasets (Alon, 2007; Bonneau et al., 2007; Palsson, 2006). Biomolecular interaction networks provide new ways to integrate these heterogeneous data sources, and shed new light into complex physiological processes (Bonneau et al., 2006; Ergun et al., 2007; Lorenz et al., 2009), including the circadian rhythm regulating a wide range of cellular, metabolic, physiological, and behavioral activities. Gene regulatory network is an ideal platform to probe the underlying molecular mechanism of circadian rhythm (Yan et al., 2008). As we know, the 20,000 dissociated neurons consisting of a pair of the mammalian suprachiasmatic nuclei (SCN) display autonomous rhythms in electrophysiological activities. This indicates that the oscillator mechanism resides within individual cells (Kawaguchi et al., 2007). Recent observations revealed that a large number of genes undergo circadian oscillation in their expression levels. Furthermore, extensive studies have identified that a set of key circadian genes utilize the transcriptional-translational autoregulatory loop to generate molecular oscillations of the "central clock."

Identification of the important circadian genes and their roles in relevant processes is the first step in elucidating the complicated mechanism of rhythms at the molecular level. However, the existence of many key circadian genes, functional organization of their interactions, and how they drive the genome-wide circadian oscillation of gene expressions still remain unknown. Thus, reconstruction of the network of regulatory interactions among these circadian genes is in pressing need, and it may further our understanding of the processes that underlie circadian rhythms.

So far, many microarray experiments have been conducted to identify circadian oscillating genes at the system level in

[1]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, People's Republic of China.
[2]Department of Electronics, Information and Communication, Engineering, Osaka Sangyo University, Osaka 574-8530, Japan.
[3]Institute of Systems Biology, Shanghai University, Shanghai 200444, People's Republic of China.
*Corresponding authors: Academy of Mathematics and Systems Science, Chinese Academy of Sciences, No. 55 Zhongguancun East Road, Beijing 100190, People's Republic of China. *Phone:* 86-10-62616659; *E-mail:* ywang@amss.ac.cn (Y. Wang); or Department of Electronics, Information and Communication, Engineering, Osaka Sangyo University, Osaka 574-8530, Japan. *Phone:* 072-875-3001; *Fax:* 072-870-8189; *Email:* chen@eic.osaka-sandai.ac.jp (L. Chen).

various mammalian species, such as the mouse, rat, rhesus macaque, and humans. These experiments are helpful to identify hundreds of circadian oscillating genes and provide valuable information on a genome-wide circadian oscillation of gene expression. However, the gene expression data are intrinsically noisy, scarce, and further heterogeneous, that is, the gene expressions are measured in different experimental conditions and sampling times. On the other hand, the development of reverse engineering techniques to characterize gene regulatory networks from gene expression data has been an active area of research (Chen et al., 2009; De Hoon et al., 2003; Dewey and Galas, 2001; Friedman, 2004; Gardner et al., 2003; Holter et al., 2001; Husmeier, 2003; Nachman et al., 2004; Tegner et al., 2003). A common challenge for all these models is the scarcity of the data, since a typical gene expression dataset consists of a few time points (often less than 30) with respect to a large number of genes (generally over thousands). In other words, the number of genes far exceeds the number of time points, making the problem of determining gene regulatory network structure a difficult and ill-posed one (D'Haeseleer et al., 2000). Recently, we developed a novel method to combine multiple time-course microarray datasets from different conditions for inferring gene regulatory networks (Wang et al., 2006b). The proposed method, called GNR (Gene Network Reconstruction tool), is based on linear programming (LP) and a decomposition procedure. The method ensures the derivation of the network structure that is most consistent with heterogeneous gene expression datasets. As a result, the method not only significantly alleviates the problem of data scarcity, but also markedly improves the prediction reliability. Importantly, we extended the GNR to a new computational strategy to recover gene regulations in a robust and reliable manner by integrating all the available experimental data and prior information in a single framework (Wang et al., 2006a, 2007, 2009). These data sources include multiple gene expression datasets at different conditions and time points, motif occurrence, ChIP–chip data, protein–protein interaction, protein–small molecule interaction, published literature and databases, and knockouts or RNA interference experiments. Furthermore, we incorporated external inputs or perturbations such as small molecules into the formulation so that molecular targets (genes) could be identified in a systematic way.

In this article, we aim to reveal essential gene regulations related to circadian rhythm based on high throughput experimental data of rat suprachiasmatic nucleus by using the improved GNR. Specifically, to improve original GNR, we designed a general framework to incorporate all available information based on linear programming. In the first step, we measure gene expressions to get the profiles of time-course data. Specifically, cultured cells of rat SCN were measured by Affimetrix GeneChip system, where folskolin stimulus was used as a phase-reset agent to induce irregular shift characteristic. In the second step, we integrate these time-course gene expression datasets related to timing of phase-reset stimuli, protein–protein interactions, phosphorylations of a set of key circadian genes, and prior information of *cis*-regulatory elements, to reconstruct a gene regulatory network for the rat circadian rhythm. In the final step, the function and structure of the reconstructed gene regulatory network are analyzed to identify the important regulatory hubs and enriched feedback motifs.

Materials and Methods

Overview of the network inference methodology

The procedure of our improved GNR is summarized as follows. The time-course datasets of microarray experiments from different conditions or perturbations are collected. A gene regulatory network is described by the first-order ordinary differential equations (ODE), and general solution of the ODE is obtained for every single time-course gene expression data. Then the previously known regulations from the heterogeneous sources are collected as prior information, and further incorporated with the multiple expression datasets in a linear programming framework as either soft constraints or hard constraints. In the end, an efficient numerical algorithm is implemented to find the most consistent gene regulatory network.

Describing gene regulation by ordinary differential equations

We need to choose an appropriate model to describe the general regulatory relationships among genes before inferring the gene regulatory network. A number of models have been proposed for reverse engineering gene regulatory network from gene expression data, such as information theory models, Boolean network models, Bayesian network models, and differential equation models. Here, the differential equation model is adopted to describe the gene expression changes as a function of the expression of other genes and environmental factors. It has the advantage to model the dynamic behavior of a gene regulatory network from time-course datasets in a quantitative way.

In general, a gene regulatory network can be described by a set of nonlinear differentialequations. Linear ODE is more practical to capture the main features of the network near the steady state and deal with the scarcity of data. Here, a linear differential equation is used to represent the rate of synthesis of a transcript as a function of the concentrations of other transcripts in a cell and the external perturbations:

$$\frac{dx(t)}{dt} = Jx(t) + Pc(t), \quad t = t_1, t_2, \ldots, t_m \tag{1}$$

where $\boldsymbol{x}(t) = [x_1(t), \ldots, x_n(t)]^T \varepsilon R^n$, is an n-dimensional vector and element $x_i(t)$ denotes the expression level (mRNA concentrations) of gene i at time point t. $J = (J_{ij})_{n\times n}$ is an $n\times n$ connectivity matrix with each element J_{ij} representing the effect of gene j on gene i with a positive, zero, or negative sign, indicating activation, no interaction, or repression, respectively. $P = (P_{ij})_{n\times s}$ is an $n\times s$ matrix representing the effect of the s perturbations or s small molecules on n genes, and $c(t) = [c_1(t), \ldots, c_s(t)]^T \varepsilon R^s$ is an s-dimensional vector and represents the external perturbations with s compounds at time t. A nonzero element P_{ij} of P implies that the ith gene is a direct target of the jth perturbation or compound. Identifying P is an important step toward biological function discovery of small molecules and drug design.

We can rewrite Equation (1) or gene regulatory network in a compact form as follows:

$$\frac{d\mathrm{X}}{dt} = J\mathrm{X} + P\mathrm{C} \tag{2}$$

where $X = [\mathbf{x}(t_1), \ldots, \mathbf{x}(t_m)]$ and $dX/dt = [d\mathbf{x}(t_1)/dt, \ldots, d\mathbf{x}(t_m)/dt]$ are $n \times m$ matrices. The first derivative of mRNA concentration is approximated as $dx_i(t_j)/dt = [x_i(t_{j+1}) - x_i(t_j)]/[t_{j+1} - t_j]$ for $I = 1, \ldots, n$; $j = 1, \ldots, m$ in this article. Supposing that there are s external perturbation compounds, then $C = [\mathbf{c}(t_1), \ldots, \mathbf{c}(t_m)]$ is an $s \times m$ matrix representing the s perturbations.

Obtaining the general solution of network structures for each dataset

Equation (2) can be reformulated as:

$$\frac{dX}{dt} = [J, P]\begin{bmatrix} X \\ C \end{bmatrix} \tag{3}$$

We then apply the Singular Value Decomposition (SVD) to $[X^T\ C^T]$:

$$\begin{bmatrix} X \\ C \end{bmatrix}^T_{m\times(n+s)} = U_{m\times(n+s)} S_{(n+s)\times(n+s)} V^T_{(n+s)\times(n+s)} \tag{4}$$

where U is a unitary $m \times (n+s)$ matrix of left eigenvectors, $S = \mathrm{diag}(s_1, \ldots, s_{n+s})$ is a diagonal $(n+s)\times(n+s)$ matrix containing the $(n+s)$ eigenvalues, and V^T is the transpose of a unitary $(n+s)\times(n+s)$ matrix of right eigenvectors. We can then obtain a specific solution of each dataset with the smallest L_2 norm for the Jacobian matrices J and P:

$$[\bar{J}, \bar{P}] = \frac{dX}{dt} U S^{-1} V^T \tag{5}$$

where $S^{-1} = \mathrm{diag}(1s_1, \ldots, 1s_{n+s})$ and $1/s_i$ is set to be zero if $s_i = 0$.

Thus, the general solution of the Jacobian matrix $J = (J_{ij})$ and $P = (P_{ij})$ for each dataset is expressed by

$$[J, P] = [\bar{J}, \bar{P}] + YV^T \tag{6}$$

Equation (6) represents all possible network structures that are consistent with each microarray dataset, depending on variables Y. $Y = (Y_{ij})_{n\times(n+s)}$ is an $n \times (n+s)$ matrix, where Y_{ij} is zero if $s_j \neq 0$, and is otherwise an arbitrary bounded scalar coefficient.

Integrating data to infer a particular solution of network structures by a linear programming framework

We assume that there are N microarray datasets $X^1, X^2, \ldots, X^N$ with $m_1, m_2, \ldots, m_N$ time points respectively for one organism, each of which corresponds to its own general solution in Equation (6). The next step is to find a consistent and also biologically plausible solution by determining variables $Y^k, k = 1, \ldots, N$. In Wang et al. (2006b), we have developed the method of GNR by exploiting the L_1 norm in the formulation of the objective function to infer a sparse and consistent gene network. Note that the gene regulatory network for differential individuals is assumed to be identical provided that they are the same organism.

We further extended GNR to extensively consider other available prior information on gene regulatory relationships. There are essentially three kinds of known regulatory interactions: directed signed, directed unsigned, and nonregulation relationships. Let the gold standard directed and signed relationships be $K = (K_{ij})_{n\times n}$, where K is an $n \times n$ matrix representing the known gene regulation information with signs. Then we store the noisy information into matrix $Z = (Z_{ij})_{n\times n}$, which is an $n \times n$ matrix representing the known gene regulation information without weights or signs. In addition, we treat the nonregulation relationships separately and store them into matrix $E = (E_{ij})_{n\times n}$, which is an $n \times n$ matrix and represents the known gene nonregulation information. These new types of prior information are expected to improve the reliability of the inferred network and reduce the computational complexity.

Specifically, N networks can be separately inferred from N time-course datasets according to Equation (6):

$$[J^k, P^k] = \frac{dX^k}{dt} U^k S^{k^{-1}} V^{k^T} = [\overline{J^k}, \overline{P^k}] + Y^k V^{k^T} \tag{7}$$

where the superscript $k = 1, \ldots, N$ is the index of the kth dataset. We denote $L = [J, P] = (L_{ij})_{n\times(n+s)}$ and $L^k = [J^k, P^k] = (L^k_{ij})_{n\times(n+s)}$ for $k = 1, \ldots, N$. Mathematically the linear programming framework for gene regulatory network inference is formulated as:

$$\begin{aligned} \min_{Y^1, Y^2, \ldots, Y^N, L} \quad & \sum_{k=1}^{N}\sum_{i=1}^{n}\sum_{j=1}^{n+s} \omega_k |L_{ij} - L^k_{ij}| + \lambda \sum_{(i,j)\in\{(i,j)|K_{ij}=0 \text{ or } Z_{ij}=0\}} |L_{ij}| \\ s.t. \quad & L_{ij} > 0 \text{ if } K_{ij} > 0 \quad i,j \in \{1, 2, \ldots, n\} \\ & L_{ij} < 0 \text{ if } K_{ij} < 0 \quad i,j \in \{1, 2, \ldots, n\} \\ & L_{ij} = 0 \text{ if } E_{ij} = 0 \quad i,j \in \{1, 2, \ldots, n\} \end{aligned} \tag{8}$$

Here, L^k_{ij} is a function of Y^k, and $Y = [Y^1, \ldots, Y^N]$ is obtained from Equation (7). The objective function has two terms. The first term is a matching term that forces the matching of L and L^k, whereas the second term is a sparseness term which forces L to be sparse as a result of the minimization of the sum of the L_1 norm. λ is a positive parameter, which balances the matching and sparseness terms in the objective function. The first and second constraints are used to add the directed and signed information, and the third one is used to incorporate the nonregulatory relationship information. All these three sets of constraints are called hard constraints, which are satisfied for any obtained solution. The variables in Equation (8) are L_{ij} and all of nonzero $Y^k_{ij} \cdot \omega^k$ is a positive weight coefficient for the kth dataset and $\sum_{k=1}^{N} \omega^k = 1$. Because different datasets may have different data qualities (e.g., different technologies, the number of repeats in measurements, etc.), the weight coefficient is used to represent the reliability of each dataset. Generally the optimal solution of Equation (8) sets as many $|L_{ij} - L^k_{ij}|$ and $|L_{ij}|$ to zero as possible, thus ensuring a consistent and sparse structure for the inferred gene regulatory network. The soft constraints are represented in the objective function, that is, in $|L_{ij}|$.

Optimization model (8) provides a general and flexible framework to integrate various data sources in the problem of gene regulatory network inference. Known interactions can be integrated in two ways in Equation (8) based on their data quality. On one hand, reliable interactions and noninteractions with the directed and signed relationship (interactions from literature or high quality prior information) are added as constraints (called hard constraints) in Equaiton (8). Their

signs are guaranteed in the network inference and only weights should be inferred. On the other hand, noisy interactions with the directed and unsigned relationships (protein–protein interaction, ChIP–chip data, or motif–occurrence data) are integrated as a penalty term in the objective function (called soft constraints) of Equation (8). Both their signs and weights need to be inferred in the integration process and their appearances in the final inferred network are preferred but not guaranteed.

Compared with unconstrained LP model in Wang et al. (2006b), the constraints in the above LP provide a flexible and consistent way to integrate all kinds of prior information. Specifically we incorporate reliable signed, noisy unsigned, and nonregulatory data in a systematic way. Because the nature of the gene regulatory network inference problem is underdetermined. In other words, the number of variables far exceeds the equations for which variables are related, the proper incorporation of the prior information from heterogeneous data sources improves the accuracy of the reconstruction.

Collecting multiple time-course expression data for circadian rhythm

In this article, we used the laboratory rat (*Rattus norvegicus*) cultured cells from SCN, and measured gene expression profiles with Affimetrix microarray (Genechip Rat Genome 230 2.0). A common experimental technique for elucidating genetic network architecture is a microarray measurement after different perturbations to the cell. An external perturbation means an experimental treatment that can alter the transcription rate of the genes in the cell. It is well known in

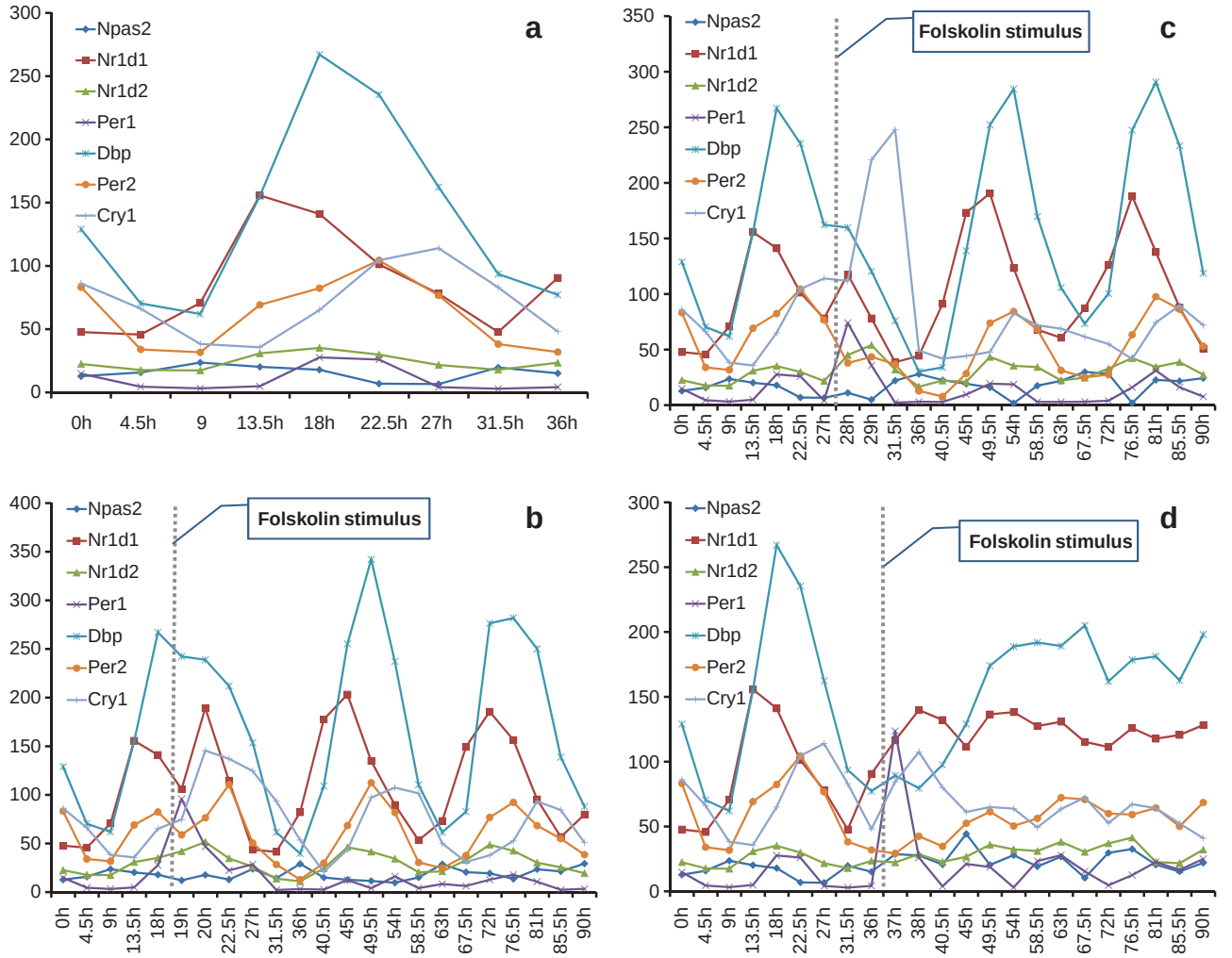

FIG. 1. The oscillation of several key circadian genes in the four experiments. (**a**) The expression of about 30,000 genes was measured in the control experiment in rat. In total, the gene expressions of 14 time points were measured from 0 to 36 h. (**b**) In the second experiment, CT6, the gene expressions of 23 time points were measured from 0 to 90 h. The phase reset stimulus by drug forskolin was applied at the time of 18 h from the time of exchanging culture medium for the cultured cells. Under this perturbation, the master clock gene *Perl* was considered to keep its phase unchanged as in the control condition. (**c**) In the experiment CT14, the gene expressions of 23 time points were measured from 0 h to 90 h. The stimulus was applied at the time of 27 h. In this case, the *Per1* showed phase delay against the control condition. (**d**) In the experiment CT22, the gene expressions of 23 time points were measured from 0 to 90 h and the stimulus was added at 36 h. The *Per1* showed phase advance in comparison with the control condition.

circadian rhythm research that forskolin stimuli can reset the clock of the cultivated cells by phase advance and phase delay. The circadian oscillation of *Per1* expression was immediately phase-shifted by forskolin, resembling the abrupt light-induced phase shift that takes place in SCN. In this study, one control and three kinds of phase-shifted experiments were examined. We normalized each time series of one control and three perturbations by assuming a normal distribution. Figure 1 shows the dynamics of gene expression of 7 key circadian genes in those four conditions.

In the control experiment, the expression of about 30,000 genes was measured in the rat. In total, 14 time points were measured from 0 to 36 h as shown in Figure 1a. The second experiment is called CT6. Twenty-three time points were measured from 0 to 90 h. The phase reset stimulus by drug forskolin was applied at the time of 18 h from the time of exchanging culture medium for the cultured cells. Under this perturbation, the master clock gene *perl* was considered to keep its phase unchanged as in the control condition as shown in Figure 1b. The third experiment is called CT14. Again, 23 time points were measured from 0 to 90 h. The stimulus was applied at the time of 27 h. In this case, the Per1 showed phase delay against the control condition in Figure 1c. The fourth experiment is called CT22. We measured 23 time points from 0 to 90 h and added the stimulus at 36 h. The *per1* showed phase advance in comparison with the control condition in Figure 1d.

Forskolin has been shown to induce *Per1* through the *cis*-regulation element in the *Per1* promoter. Considering the effect of forskolin will disappear 4 h later, we extracted the measured time points after the drug took effect, and discarded these time points before or around the drug stimulus. In this way, we assembled four time-course gene expression datasets. The first one has the 14 time points from the control experiment. The second one contains 16 time points (20 to 90 h) from the CT6 experiment. For the CT14 experiment, we have 14 time points (29 to 90 h). The fourth time-course dataset contains 12 time points (38 to 90 h) from CT22.

Assembling key circadian genes, circadian-related genes, and oscillatory gene list

Through many years of molecular and genetic studies (Ko and Takahashi, 2006; Ueda et al., 2005), 18 key circadian genes have been identified and extensively studied in mammals: *Bmal1(Arntl), Clock, Cry1, Cry2, Dbp, Dec1(Bhlhb2), Dec2(Bhlhb3), Nfil3(E4bp4), Npas2, Nr1d1(Rev-erb), Nr1d2, Per1, Per2, Per3, Rora, Rorb, Rorc, Rasd1.* In Table 1, we listed their gene names, probe names, their *cis*-regulatory elements, protein interactions, phosphorylations, degradations, and functions.

In addition, 22 circadian related genes were identified by checking their protein interactions and phosphorylations relationships with the 18 key circadian genes. We presented the list of 22 circadian related genes and their functional relationships with the key circadian genes in Table 2. Furthermore, we extracted 50 oscillatory genes from the expression data over 30,000 genes in rat (shown in Table 3). We first normalized each time series of control and three perturbations by assuming a normal distribution whose mean is zero and variance is 1. Then fast Fouier transform was formulated for each normalized time series (Morioka et al., 2008). We picked those genes with a significantly large spectral variance as oscillating genes because the variance of power spectra of each gene determines whether typical oscillations exist or not. As a result, about 0.17% of the all 30,000 genes were selected as oscillatory genes.

We then assembled the key circadian genes, circadian-related genes, and the oscillatory genes together as the candidate genes in our network inference procedure. In total, 80 genes were selected to construct a medium-scale circadian gene regulatory network. The quality and the coverage of this list ensure the high-confidence network inference results.

Results

Collecting known interactions in circadian gene regulation

Genetic analyses have identified many clock genes and their functional linkages (Ueda et al., 2005). We will incorporate them in our circadian gene regulatory network inference procedure as prior information. To this end we extensively collected the available known interactions among circadian genes and related genes by literature search. We summarized the important prior information as follows and visualized them in Figure 2.

Protein–protein interaction. Proteins are the products of gene transcription and translation, and they play important roles in a cell. Protein–protein interactions occur in many cellular processes, such as signaling cascades and enzyme-complex formation. Previous studies revealed that two basic helix-loop-helix-PAS transcription factors *Clock* and *Bmal1* form heterodimer and subsequently bind *cis*-regulatory element Ebox. Three period genes (*Per1, Per2,* and *Per3*) and two cyptochrome genes (*Cry1* and *Cry2*) form a protein complex. The two basic helix-loop-helix transcription factors *Dec 1* and *Dec2* (also called *Bhlhb2* and *Bhlhb3,* respectively) interact with *Bmal1* detected by the yeast two-hybrid experimental technique. One clock-related gene, *Npas2,* and transcription factor, *Bmal1,* form heterodimer and subsequently bind to the *cis*-regulatory element Ebox (Ueda et al., 2005). These heterodimer and complex interactions are then converted into binary interactions. In total, we get 14 physical protein interactions as prior information (shown in Fig. 2 as blue solid lines).

Phosphorylation. Phosphorylation is the addition of a phosphate (PO4) group to a protein or other organic molecule. Protein phosphorylation in particular plays a significant role in a wide range of cellular processes. The clock gene *Bmal1(Arntl)* is phosphorylated by Casein kinase I epsilon (*CKIe* or *Csnk1e*). The transcription factor *Clock* is phosphorylated by PKG-II (*Prkg2*) and PFK family genes (*Pfkfb2, Pfkfb4, Pfkm, Pfkfb3, Pfkfb1,* and *Pfk1*). The crytochrome gene *Cry1* is phosphorylated by CKIe (*Csnk1e*) and MAPK family genes (*Mapk8ip, Mapk4, Mapk7, Mapk10, Mapk13, Mapl1, Mapk8, Mapk9, Mapk14, Mapk6, Mapk12,* and *Mapkapk2*). The crytochrome gene *Cry2* is phosphorylated by GSK-3b (*Gsk3b*) and MAPK family genes.

The b-Zip family gene *Nfil3 (E4bp4)* is phosphorylated by CKIe (*Csnk1e*). The orphan nuclear hormone receptor *Nr1d1 (Rev-erb)* is phosphorylated by *GSK3-b (Gsk3b*). Two period genes *Per1* and *Per3* are phosphorylated by CKIe (*Csnk1e*). The period gene *Per2* is phosphorylated both by CKIe (*Csnk1e*) and

Table 1. A List of 18 Clock Genes and Their cis-Regulatory Element, Protein Interaction, Phosphorylation, Degradation, and Functions

Clock genes	*Probes*	*Cis-rgulatory element*	*Function of proteins*	*Protein interaction*	*Phosphorylaiton*	*Active degradation*
*Bmal*1(Arntl)	1370510_a_at	RORE	activates Ebox	CLOCK and BMAL1 form heterodimer and subsequently bind Ebox.	Phosphorylated by CKIe	
Clock	1369168_a_at	RORE	activates Ebox	CLOCK and BMAL1 form heterodimer and subsequently bind Ebox.	Phosphorylated by PKG-II, PFKs	
Cry1	1392640_at	Ebox, RORE	inactivates Ebox	Per1, Per2, Per3, Cry1, Cry2 form complex.	Phosphorylated by CKIe, MAPK	
Cry2	1369446_at	Ebox	inactivates Ebox	Per1, Per2, Per3, Cry1, Cry2 form complex.	Phosphorylated by GSK-3b, MAPK	Degraded by b-TrCP after phosphorylation
Dbp	1387874_at	Ebox	activates Dbox			
Dec1(Bhlhb2)	1369415_at, 1379483_at, 1381121_at	Ebox	inactivates Ebox	Dec interact with BMAL1 (Y2H)		
Dec2(Bhlhb3)	1368511_at	Ebox	inactivate Ebox			
Nfil3(E4bp4)	1368488_at	RORE	inactivates Dbox		Phosphorylated by CKIe	
Npas2	1383439_at, 1398024_at	RORE	activates Ebox	Npas2 and BMAL1 form heterodimer and subsequently bind Ebox.		
Nr1d1(Rev-erb)	1370816_at	Ebox, Dbox	inactivates RORE		Phosphorylated by GSK3-b	
Nr1d2	1370540_at, 1370541_at	Ebox, Dbox	inactivates RORE			
Per1	1374855_at	Ebox, Dbox	inactivate Ebox	Per1, Per2, Per3, Cry1, Cry2 form complex.	Phosphorylated by CKIe	Degraded by b-TrCP after phosphorylation
Per2	1368303_at	Ebox, Dbox	inactivate Ebox	Per1, Per2, Per3, Cry1, Cry2 form complex.	Phosphorylated by CKIe, GSK3-b	Degraded by b-TrCP after phosphorylation
Per3	1378745_at	Dbox	inactivate Ebox	Per1, Per2, Per3, Cry1, Cry2 form complex.	Phosphorylated by CKIe	Degraded by b-TrCP after phosphorylation
Rora	1379397_at, 1377029_at, 1388662_at, 1391801_at	Dbox	activates RORE			
Rorb	1371257_at, 1395023_at, 1384814_at	Dbox	activates RORE			
Rorc	1379833_at	Ebox, RORE	activates RORE			
Rasd1	1387908_at		Neuron, 43(5), 715–728, 2004			

Table 2. A List of 22 Clock-Related Genes and Their Functions

Gene name	*Probe name*	*Function*
Csnk1e	1384161_at, 1389212_at, 1386693_at, 1369804_a_at	CKIe
Prkg2	1369090_at	PKG-II
Pfkfb2	1369400_a_at, 1388044_at, 1388063_a_at, 1398320_at	related to PFKs
Pfkfb4	1396818_at, 1368719_at	related to PFKs
Pfkm	1384632_at, 1386961_at, 1367864_at	related to PFKs
Pfkfb3	1369794_a_at, 1397082_at	related to PFKs
Pfkfb1	1369467_a_at	related to PFKs
Pfkl	1367743_at	related to PFKs
Mapk8ip	1386973_a_at, 1387971_a_at	related to MAPK
Mapk4	1390092_at, 1378751_at	related to MAPK
Mapk7	1370985_at	related to MAPK
Mapk10	1380878_at, 1377059_at, 1397130_at, 1380648_at, 1369102_at	related to MAPK
Mapk13	1369183_at	related to MAPK
Mapk1	1397822_at, 1398346_at, 1369078_at, 1373426_at, 1375692_at, 1383997_at	related to MAPK
Mapk8	1392418_at	related to MAPK
Mapk9	1378237_at, 1380024_at, 1368646_at	related to MAPK
Mapk14	1377136_at, 1380238_at, 1367697_at	related to MAPK
Mapk6	1368273_at	related to MAPK
Mapk12	1398297_at	related to MAPK
Mapkapk2	1371446_at	related to MAPK
Gsk3b	1370267_at	related to Gsk-3b
Btrc	1374654_at	related to b-TrCP

GSK3-b (*Gsk3b*). All these kinase genes related to the above phosphorylations are added to our circadian-related gene list. In total, we collected 40 phosphorylation interactions. In addition, we know that the clock genes *Cry2, Per1, Per2,* and *Per3* are degraded by *Btrc* after phosphorylation.

Cis-regulatory element. Protein–DNA interactome data concerns the interactions between proteins and DNA, particularly between transcription factors and their target promoters. They fundamentally define the transcriptional regulatory network of the cell. Here we mainly used the known *cis*-regulatory elements or *cis*-elements, which are the regions of DNA that regulates the expression of genes located on that same strand. In total, three *cis*-regulatory elements were well studied in circadian rhythm: Ebox (CACGT[T/G]), DBP/E4BP4 binding elements (Dbox, TTA[T/C]GTAA), and RevErbA/ROR binding elements (RORE, [A/T]A[A/T]NT[A/G]GGTCA). In addition, *Bmal1 (Arntl), Clock, Dbp,* and *Npas2* activate Ebox; *Cry1, Cry2, Dec1(Blhlhb2), Dec2(Blhlhb3), Per1, Per2,* and *Per3* inactivate Ebox; *Dbp* activate Dbox; *Nfil3(E4bp4)* inactivate Dbox; *Rora, Rorb,*and *Rorc* activate PORE; *Nr1d1(Rev-erb)* and *Nrid2* inactivate RORE. Based on these information we established 134 transcriptional regulatory interactions by linking the transcription factor with their target promoter region in the gene level.

Protein–drug interaction. Small molecules can be used to dissect diverse biological processes, such as cellular metabolism, signal transduction, and intracellular protein trafficking. In our microarray experiment, we added folskolin as the stimulus for circadian rhythm process. We picked up the significantly induced and or repressed genes as the potential target of the drug folskolin. This information was incorporated into the network inference as external perturbations.

Reconstructing the circadian gene regulatory network

We integrated all the available prior information described above with the four time-course gene expression datasets by our linear programming framework. In addition to various experimental data sources, we also incorporated prior information about the network structure. From the viewpoint of topology, it is commonly believed that a gene regulatory network is sparse in nature, that is, each gene is only genetically affected by a limited number of genes (Gardner et al., 2003; Yeung et al., 2002). It is straightforward to incorporate this prior information into our inference model. The main idea here is to make the gene regulatory network sparse so that it is biologically plausible. We introduced a parameter λ to tune the sparseness of the inferred network. The larger the λ is, the sparser the inferred network is. When we choose $\lambda = 0$, we inferred 642 regulatory interactions among 80 genes; the average degree is about 8. In this case, all the known interactions were correctly inferred and 505 new regulatory interactions were obtained.

To get a high-confidence set of circadian gene regulatory network in the rat, we chose an even more stringent parameter $\lambda = 0.1$. Using this stringent parameter, we obtained 276 predicted regulatory relationships among 80 circadian related genes. In Figure 3 we visualized this high-confidence circadian gene regulatory network in the rat. We further labeled genes with their gene names, made the node size proportional to their out-degree in the network, and labeled regulatory interactions with their activation or repression role (red solid

Table 3. A List of 50 Oscillatory Genes from the Expression Data with Their Functional Descriptions

Gene name	*Probe name*	*Description*
—	1392490_at	Similar to RIKEN cDNA 1110007C24
LOC291903	1381759_at	hypothetical LOC291903
—	1393405_at	Transcribed locus
RGD1307813	1373813_at	DnaJ (Hsp40) homolog, subfamily C, member 10 (predicted)
Fut4	1373838_at	fucosyltransferase 4
—	1394518_at	—
—	1380594_at	CDNA clone IMAGE:7303954, partial cds
—	1380324_at	—
—	1372544_at	Transcribed locus
—	1376076_at	—
Bbc3	1382993_at	Bcl-2 binding component 3
RGD1307315_predicted	1382320_at	LOC362793 (predicted)
—	1381280_at	Similar to methylenetetrahydrofolate dehydrogenase (NAD) (EC 1.5.1.15)/ methenyltetrahydrofolate cyclohydrolase (EC 3.5.4.9) precursor—mouse
Kif3c	1369637_at	kinesin family member 3C
—	1378117_at	—
RGD1306883	1378160_at	stromal cell derived factor receptor 2 (predicted)
RGD1310899_predicted	1367494_at	similar to CGI-35 protein (predicted)
—	1391808_at	Transcribed locus
—	1395628_at	—
Fosl1	1368489_at	fos-like antigen 1
—	1376602_a_at	Similar to F-box only protein 22
RGD621665	1368144_at	regulator of G-protein signaling 2
—	1373960_at	Similar to transmembrane protein induced by tumor necrosis factor alpha
—	1393547_at	—
Tas1r2	1372277_at	Taste receptor, type 1, member 2
—	1388524_at	Transcribed locus
—	1373786_at	Transcribed locus
RGD1306478	1381836_at	RAB20, member RAS oncogene family (predicted)
—	1384217_at	Transcribed locus
RGD1305020_predicted	1377532_at	similar to Hepatocellular carcinoma-associated antigen 58 homolog (predicted)
—	1385374_at	—
RGD1306620	1388858_at	mitogen activated protein kinase kinase 3 (predicted)
—	1393063_at	Transcribed locus
Dsipi	1367771_at	delta sleep inducing peptide, immunoreactor
RGD727777	1372734_at	small cell adhesion glycoprotein
RGD1307311	1390141_at	formyltetrahydrofolate synthetase domain containing 1 (predicted)
—	1373243_at	—
RGD1304581	1372004_at	heme binding protein 1 (predicted)
—	1387019_at	—
—	1383173_at	Transcribed locus

line and arrow denotes activation; blue line and bar denotes repression). Importantly, we predicted 138 new regulatory relationships that are not in the prior information (73 activation and 65 repressions). Overall, our novel predictions could be grouped into two categories: (1) brand new regulatory relationships, and (2) signs and weights for those functional relationships in the prior information.

Assessing the structure and function of the inferred network

Identifying important regulatory hubs. In the predicted function interactions, a putative regulator affects the expression of its targets through regulatory interactions. Thus, the regulators with many regulatory interactions may have an important role in biological function. We call these regulators

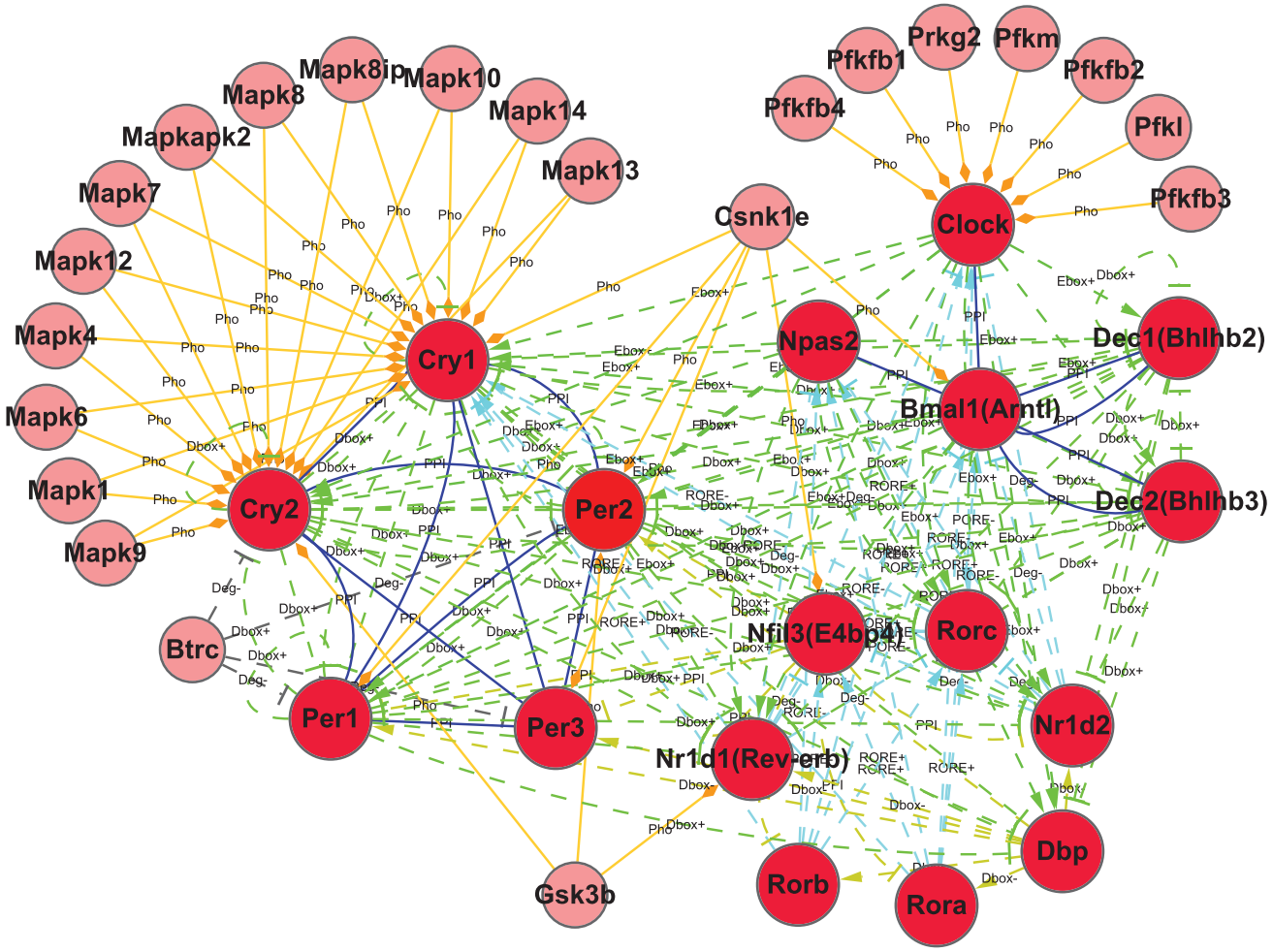

FIG. 2. Known relationships among the key circadian genes and the circadian-related genes. In total, there are nine types of relationships, and the types of the relationships are shown in different edge color and style and labeled in edges. Solid lines denote interaction in the protein level, whereas dashed lines denote regulatory interaction in the gene level. Arrows denote activation and bars denote repression. PPI stands for protein–protein interaction; Pho stands for the phosphorylation; Ebox+ is for the activation through *cis*-regulatory element Ebox; Ebox− means depression through Ebox; Similarly, Dbox+ means activation by *cis*-regulatory element Dbox; Dbox− is for depression through Dbox; RORE+ stands for activation by *cis*-regulatory RORE; RORE− means depression by RORE; Deg− means degradation by *Btrc* after phosphorylation.

with a high out-degree in the predicted network (larger node size in Fig. 3) as regulatory hubs. We performed some analysis on these regulatory hubs in our predicted network.

The well-known transcription factors in circadian rhythms were identified as regulatory hubs in the predicted networks. They include *Clock*, *Npas2*, *Dec2(Bhlhb3)*, *Cry1*, *Nr1d1(Rev-erb)*, *Nfil3(E4bp4)*, *Bmal1(Arntl)*, *Rora*, *Rorb*, *Rorc*, *Dec1(Bhlhb2)*, *Per2*, *Cry2*, *Dbp*, and *Per1*, and regulate many target genes. Surprisingly, we also found four new regulatory hubs: *Dsipi* (regulate 17 target genes), *RGD1310899_predicted* (regulate 29 target genes), *RGD621665* (regulate 20 target genes), and *RGD1307813* (regulate 8 target genes). We then annotated these four regulatory hubs by their GO functions. We find that those important regulatory hubs in the predicted network all have significant functional relationships with circadian rhythm.

The full name of *RGD621665 (Rgs2)* is regulator of G-protein signaling. Its molecular functions include GTPase activator activity and calmodulin binding. The biological process functions include the transmembrane receptor protein tyrosine kinase signaling pathway, the G-protein-coupled receptor protein signaling pathway, the regulation of G-protein-coupled receptor protein signaling pathway, and the negative regulation of the G-protein-coupled receptor protein signaling pathway. *RGD621665* localizes in the nucleus in the cell. The regulatory hub *Dsipi* belongs to the TSC22 domain family protein 3 (glucocorticoid-induced leucine zipper protein). By the GO function, it is a transcription factor. It protects T cells from IL2 deprivation-induced apoptosis through the inhibition of FOXO3A transcriptional activity that leads to the downregulation of the pro-apoptotic factor BCL2L11. In macrophages, it plays a role in the anti-inflammatory and immunosuppressive effects of glucocorticoids and IL10. In T cells, it inhibits anti-CD3-induced NFKB1 nuclear translocation. *In vitro*, it suppresses AP1 and NFKB1 DNA-binding activities. There is relatively little information for the regulatory hubs *RGD1307813* and *RGD1310899_predicted*. *RGD1307813* is related to endoplasmic reticulum,

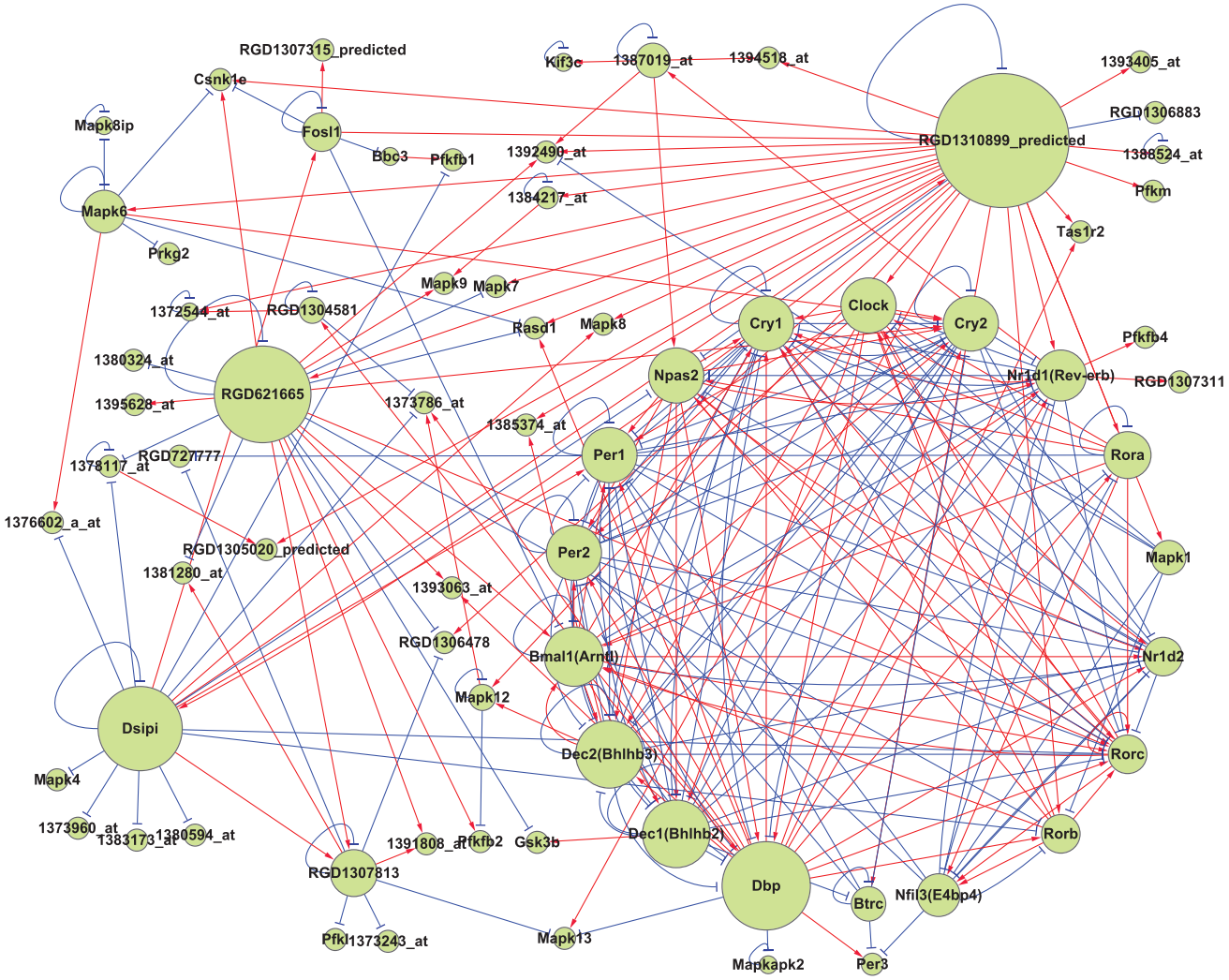

FIG. 3. The predicted gene regulatory network for circadian rhythm in rat. There are total 80 genes with 276 regulatory interactions in the network. Nodes denote the genes and the size of each gene or node is proportional to its out-degree in the network. Red solid line with arrow denotes activation. Blue line with bar denotes repression.

heat-shock protein binding, cell redox homeostasis, and protein folding, whereas *RGD1310899_predicted* is similar to the CGI-35 protein (predicted).

In Figure 4, we plot the mRNA expression profile of hub gene *RGD621665 (Rgs2)* in experiments CT6, CT14, and CT22. It has a relatively high expression level in all the three perturbation experiments, and its dynamic patterns show that the expression level of this regulatory hub is very sensitive to the external drug and has strong oscillation property. From these evidences we deduce that *RGD621665* may play an important role in the regulation of circadian rhythm.

Revealing enriched feedback loop motifs

Motivated by the importance of transcription-translation feedback loops in driving the circadian rhythm, we explored the interesting motifs in the predicted network using MaVisto (Schreiber and Schwobbermeyer, 2005). In this way we broke the large network down into the simplest units of commonly used network architecture which are called network motifs.

Indeed, we found that five types of motifs are significantly enriched in the predicted network (shown in Fig. 5). They all contain the feedback loops. Here, we used the *z*-score and *p*-value to assess the statistical significance of the certain motif in the predicted network. The *z*-score is defined as the difference of the frequency of this motif in the predicted circadian gene regulatory network and its mean frequency in a sufficiently large set of randomized networks (1,000 random networks were generated in our implement), divided by the standard deviation of the frequency values for the randomized networks. The randomized versions of the analyzed networks were generated using a random local rewiring algorithm that preserves the degrees of the vertices. The *p*-value of a motif *m* is defined as the probability *P* that the frequency of *m* in a randomized network is equal to or larger than the frequency of *m* in the predicted network.

Among the total 80 genes in the predicted network, 27 genes are with self-regulatory interaction (Fig. 5a). It means that about 34% of the genes in the circadian gene regulatory network have this one-node feedback loop. They include the important circadian genes such as *Bmal1, Per1, Per2*, etc. This is very significant because in randomly generated networks we seldom can find any self-regulatory loop. The two-node

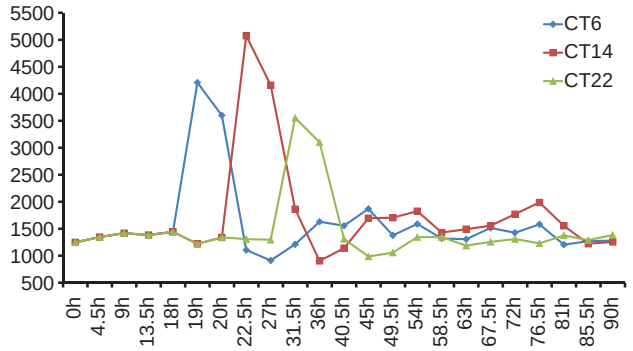

FIG. 4. The mRNA expression profile of regulatory hub *RGD621665 (Rgs2)* in experiments CT6, CT14, and CT22. Its GO function in biological process is regulation of G-protein coupled receptor protein signaling pathway.

feedback loop motif in Figure 5b appears 38 times in the predicted circadian gene regulatory network. The average appearance number in 1,000 randomly generated networks is about 21, with a standard derivation 2.9. The calculated *z*-socre is about 5.6, with a *p*-value <1e-10. As is now a commonly accepted example, *Bmal1* and *Clock* proteins form a complex that positively regulates the transcription of *Per* and *Cry* family genes. *Per* and *Cry* family protein form a complex that inhibits *Bmal1*/*Clock* transcriptional activity. This negative feedback loop is one type of the two-node motif. Similarly, we counted the three-node feedback loop motifs. Again, we found that these motifs are significantly enriched in the predicted network. The motifs in Figures 5c, d, and e appear 345 times, 229 times, and 20 times in the inferred circadian gene regulatory network, respectively. Calculated *z*-scores and *p*-values show that their frequencies of appearance are very significant (*p*-value <1e-10). Because all these motifs contain the feedback loop structures, we claim that the feedback loop structures are greatly enriched in our predicted

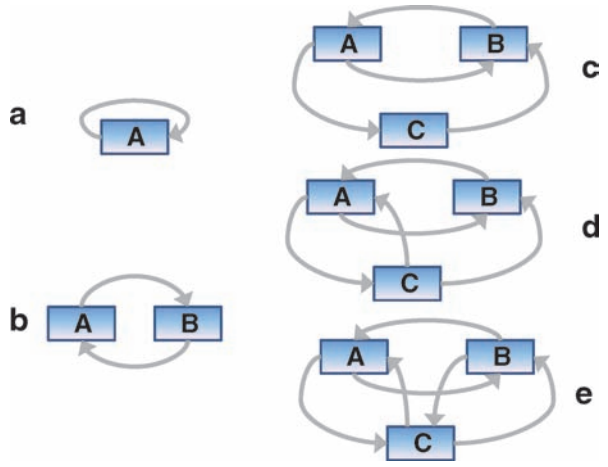

FIG. 5. The enriched feedback loop motifs in the inferred gene regulatory network. (**a**) The self-regulatory interaction motif appears 27 times in the predicted circadian gene regulatory network in the rat. (**b**) The two-node feedback loop motif appears 38 times in the predicted circadian gene regulatory network. The three-node feedback loop motifs appear 345 times (**c**), 229 times (**d**), and 20 times (**e**) in the inferred circadian gene regulatory network, respectively.

circadian gene regulatory network. It also demonstrates that the gene regulation in this important physiological process is more complex than what we understood.

Discussion

In this article, we have shown that the reverse-engineering approach can be applied to infer a gene regulatory network in complex physiological and metabolic process. We assessed the performance of the methodology by analyzing the structure and function of the predicted network, which demonstrates that many interactions, regulatory hubs, and feedback loop motifs were predicted by our model. These results revealed a greater degree of complexity between regulators and target genes in the network than previously appreciated. Importantly, we provided a medium-scale study (80 nodes) of the circadian gene regulatory network by computational systems biology method. Compared to the relatively small-scale gene regulatory network (16 nodes) by experimental method, the advantage of our study is that this strategy enables a system-wide overview on the gene regulation, which may provide valuable insight onto the essential mechanism of circadian rhythm. However, these predictions need further experimental validation.

Our data integration strategy in the network reconstruction procedure not only improves the reliability of the inferred gene regulatory network, but also can be applied to drug design and many other areas of biomedical research and bioengineering. Specifically, we proposed a framework to combine computational analysis of multiple microarray datasets and other types of biological experiments together for inferring gene regulatory network and further identified small molecule targets of perturbation experiments. In the network inference procedure, available information from heterogeneous data sources is incorporated into the LP as constraints. They can be divided into two classes according to data reliability. For example, the regulatory relationships mined from literature and databases (confirmed by biological experiments) are more reliable. Hence, they can be treated as hard constraints in our linear programming that will be strictly enforced. On the other hand, the regulatory relationships predicted from *cis*-regulatory elements are generally noisy and the related pairs often are unsigned, meaning that the activation or repression role is unknown. In this case, we treat them as soft constraints, which may or may not be satisfied. In this way, our linear programming model provides a flexible prior information learning framework. It finds the most consistent gene regulatory network by balancing among heterogeneous data sources. One major advantage of the proposed method is that it theoretically ensures the derivation of the most consistent network with respect to the available datasets or information, thereby alleviating the problem of data scarcity and improving the reliability. In addition, the convergence of the algorithm is also guaranteed.

Recent studies have revealed that the circadian clock is cell-autonomous and self-sustained not only in the SCN but also in peripheral clock tissues. This finding has led to an increasing effort to better understand the circadian mechanisms in dependent peripheral oscillators. Integration of different circadian microarray datasets in multiple tissues in our work could further elucidate the nature of the mammalian circadian system. However, both data integration

and network inference are challenging problems because the specific contributions of molecular clock components, and interactions among the clock components may vary in a tissue-specific manner.

Acknowledgments

We thank Prof. Hajimi Tei and Dr. Gen Kurosawa on the gene expression data and prior information collection. We also thank them for the helpful discussions. Y.W. is supported by grant No. 10801131 and X.S.Z. is supported by 10631070 from the National Natural Science Foundation of China. Y.W. and X.S.Z. are also supported by grant kjcs-yw-s7 from the Chinese Academy of Sciences. Y.W., X.S.Z., and L.C. are supported by by JSPS and NSFC under JSPS-NSFC collaboration project.

Author Disclosure Statement

The authors declare that no conflicting financial interests exist.

References

Alon, U., 2007. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. (Chapman & Hall/CRC, London).

Bonneau, R., Reiss, D.J., Shannon, P., Facciotti, M., Hood, L., Baliga, N.S., et al. 2006. The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. Genome Biol **7**, R36.

Bonneau, R., Facciotti, M.T., Reiss, D.J., Schmid, A.K., Pan, M., Kaur, A., et al. 2007. A predictive model for transcriptional control of physiology in a free living cell. Cell 131, 1354–1365.

Chen, L., Wang, R-S., and Zhang, X-S. 2009. *Biomolecular Networks: Methods and Applications in Systems Biology*. (John Wiley & Sons, New York).

De Hoon, M.J.L., Imoto, S., Kobayashi, K., Ogasawara, N., and Miyano, S. 2003. Inferring gene regulatory network from time-ordered gene expression data of *Bacillus subtilis* using differential equations. Pac Symp Biocomput 17–28.

Dewey, T.G., and Galas, D.J. 2001. Dynamic models of gene expression and classification. Funct Integrat Genomics 1, 269–278.

D'Haeseleer, P., Liang, S., and Somogyi, R. 2000. Genetic network inference: from co-expression clustering to reverse engineering. Bioinformatics 16, 707–726.

Ergun, A., Lawrence, C.A., Kohanski, M.A., Brennan, T.A., and Collins, J.J. 2007. A network biology approach to prostate cancer. Mol Syst Biol 3, 82.

Friedman, N., 2004. Inferring cellular networks using probabilistic graphical models. Science 303, 799–805.

Gardner, T.S., Di Bernardo, D., Lorenz, D., and Collins, J.J. 2003. Inferring genetic networks and identifying compound mode of action via expression profiling. Science 301, 102–105.

Holter, N.S., Maritan, A., Cieplak, M., Fedoroff, N.V., and Banavar, J.R. 2001. Dynamic modeling of gene expression data. Proc Natl Acad Sci USA 98, 1693.

Husmeier, D. 2003. Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks. Bioinformatics 19, 2271–2282.

Kawaguchi, S., Shinozaki, A., Obinata, M., Saigo, K., Sakaki, Y., and Tei, H. 2007. Establishment of cell lines derived from the rat suprachiasmatic nucleus. Biochem Biophys Res Commun 355, 555–561.

Ko, C.H., and Takahashi, J.S. 2006. Molecular components of the mammalian circadian clock. Hum Mol Genet 15, R271.

Lorenz, D.R., Cantor, C.R., and Collins, J.J. 2009. A network biology approach to aging in yeast. Proc Natl Acad Sci USA 106, 1145.

Morioka, R., Arita, M., Sakamoto, K., Kawaguchi, S., Tei, H., and Horimoto, K. 2008. Phase shifts of circadian transcripts in rat suprachiasmatic nucleus. In *Optimization and Systems Biology*. (World Publishing Corporation, Lijiang, China), pp. 109–114.

Nachman, I., Regev, A., and Friedman, N. 2004. Inferring quantitative models of regulatory networks from expression data. Bioinformatics 20(Suppl. 1), i248–i256.

Palsson, B.O. 2006. *Systems Biology: Properties of Reconstructed Nnetworks*. (Cambridge University Press, New York).

Schreiber, F., and Schwobbermeyer, H. 2005. MAVisto: a tool for the exploration of network motifs. Bioinformatics 21, 3572–3574.

Tegner, J., Yeung, M.K., Hasty, J., and Collins, J.J. 2003. Reverse engineering gene networks: integrating genetic perturbations with dynamical modeling. Proc Natl Acad Sci USA 100, 5944.

Ueda, H.R., Hayashi, S., Chen, W., Sano, M., Machida, M., Shigeyoshi, Y., et al. 2005. System-level identification of transcriptional circuits underlying mammalian circadian clocks. Nat Genet 37, 187–192.

Wang, R.S., Wang, Y., Zhang, X.S., and Chen, L. 2007. Inferring transcriptional regulatory networks from high-throughput data. Bioinformatics 23, 8.

Wang, Y., Joshi, T., Xu, D., Zhang, X., and Chen, L. 2006a. Supervised inference of gene regulatory networks by linear programming. Lecture Notes Comput Sci 4115, 551.

Wang, Y., Joshi, T., Zhang, X.S., Xu, D., and Chen, L. 2006b. Inferring gene regulatory networks from multiple microarray datasets. Bioinformatics 22, 2413.

Wang, Y., Wang, R.-S., Joshi, T., Xu, D., Zhang, X.-S., Chen, L., and Xia, Y. 2009. A linear programming framework for inferring gene regulatory networks by integrating heterogeneous data. In: D.C. Sanjoy, W.H. Hsu, and S.M. Welch (eds.), *Computational Methodologies in Gene Regulatory Networks*. (IGI Global, Hershey, PA).

Yan, J., Wang, H., Liu, Y., and Shao, C. 2008. Analysis of gene regulatory networks in the mammalian circadian rhythm. PLoS Comput Biol 4.

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

Address correspondence to:
Yong Wang
Academy of Mathematics and Systems Science
Chinese Academy of Sciences
No. 55 Zhongguancun East Road
Beijing 100190
People's Republic of China

E-mail: ywang@amss.ac.cn