A Gaussian Graphical Model for Identifying Significantly Responsive Regulatory Networks from Time Series Gene Expression Data

Zhi-Ping Liu*,‖, Wanwei Zhang*,‖, Katsuhisa Horimoto†, and Luonan Chen*
*Key Laboratory of Systems Biology, SIBS-Novo Nordisk Translational Research Centre for Pre-Diabetes, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China
Email: zpliu@sibs.ac.cn; zhangwanwei@sibs.ac.cn; lnchen@sibs.ac.cn
‖These authors contributed equally to this work
†Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo 135-0064, Japan
Email: k.horimoto@aist.go.jp

***Abstract*—With rapid accumulation of functional relationships between biological molecules, knowledge-based networks have been constructed and stocked in many databases. These networks provide the curated and comprehensive information for the functional linkages among genes and proteins, while their activities are highly related with specific phenotypes and conditions. To evaluate a knowledge-based network in a specific condition, measuring the consistency between its structure and the conditionally specific gene expression profiling data is an important criterion. In this work, we propose a Gaussian graphical model to evaluate the documented regulatory networks by the consistency between network architectures and time-series gene expression profiles. By developing a dynamical Bayesian network model, we derive a new method to evaluate gene regulatory networks in both simulated and true time series microarray data. The regulatory networks are evaluated by matching a network structure and gene expressions, which are achieved by randomly rewiring the regulatory structures. To demonstrate the effectiveness of our method, we identify the significant regulatory networks in response to the time series gene expression of circadian rhythm. Moreover, the knowledge-based networks are screened and ranked by their consistencies of structures based on dynamical gene expressions.**

***Index Terms*—Gaussian graphical model, network evaluation, regulatory structure, time series gene expression.**

I. Introduction

Gene regulatory network provides a basic framework for the regulation relationship between transcription factors and their target genes [1], [3]. The network architecture is a lattice of their relationships. It is a promising way to reconstruct the gene regulatory network by reverse engineering [24], [25]. The DREAM (Dialogue for Reverse Engineering Assessment and Methods) challenge provides evidence for the effectiveness of the network reconstruction algorithms [12], but there are still many difficulties in these inferences. It is difficult to assess the sensitivity and specificity of these inference results beyond the curse of dimensionality of the problem. Simultaneously, there is more and more available information about gene regulations. Evaluating a knowledge-based network with gene expression data will provide a valuable alternative to study gene regulatory networks [7], [15]. The documented networks or pathways are often based on the retrieval information about their relationships among genes and proteins [9]. The matching significance between the reference networks and the expression profiles will indicate the enrichment information and functional linkages underlying these genes and their networking activities in the conditional specificity.

The existing approaches of analyzing gene expression generally start from the identification of differentially expressed genes by comparing the expressions in different conditions. They often include statistical tests, such as t-test and SAM [22]. However, genes perform their functions by interacting with each other in the form of network or pathway. Then, there are some methods which have been proposed for the pathway analysis [4], [14]. GSEA [16] and GSA [6] provide the significance test for the predefined gene sets in certain gene expression profiling. Gene set will provide more information for the interrelationship of genes and imply their regulations from the system level [21]. However, network structures or topologies have not been considered in the most of the identification methods [19]. Moreover, the analysis has not conducted to consider the true gene expressions efficiently and there are lots of constraints for network structures in the assessment [5], [13]. The relationship of gene regulations defines the core network architecture underlying these genes during the biological processes [3]. In response to certain conditions, these gene regulatory networks will perform very different biological functions and show obvious structure dynamics [11]. Here, we aim to provide an evaluation method for the documented gene regulatory networks based on the consistency between network structures and time-series gene expression data. The consistency between network structures and measured data is well known in statistical casual hypothesis [15]. The architecture of topological linkages will provide regulatory implications which underlie the gene expression. If we evaluate the functional linkages and their responses matching with the gene expressions, the network activity and importance will be identified. This

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

will provide crucial implications for their biological process. There are some methods have been developed to reconcile network structure and gene expression. [7] provided a linear regression method to measure the consistency, but it cannot handle the amount of large networks in parallel. [5] proposed a scoring scheme for assessing the significance of pathways by their ranks corresponding to gene expressions. However, it ignores the regulatory interactions between genes as well as fails to detect the correspondence between gene expression and network structure. We have provided a method to screen the consistency between network structures and gene expressions, while the method cannot handle the network with cycles and loops [15], [26]. To evaluate biological networks with consideration of general network structures, it is urgent to develop a new method to identify the consistency between regulatory network structures and gene expression profiles.

In this work, we proposed a dynamical graphical model to identify the significant regulatory networks from the knowledge-based reference networks in response to conditional gene expressions. Instead of reconstructing gene regulatory networks from high throughput data, the significant regulatory networks are identified from the reference network library by the time-series high throughput data. We validated our method in both simulated data and true time-series circadian rhythm data. The documented network structures of transcription factors and targets are evaluated from the random samplings. The possibility of graph architectures existing in certain conditions was measured by the consistency between network structures and gene expression profiles. In particular, we ranked the referred regulatory networks by the structural consistency in response to specific time-series gene expression profiling data. As shown in the results, the statistical significance as well as the potential regulation architecture provides detail information for the regulatory networks responding to gene expressions in specific conditions.

II. Materials and Methods

A. Framework

Figure 1 shows the framework of our method to identify the significantly responsive regulatory networks by evaluating the consistency between network structures and gene expressions. For the reference regulatory networks documented in databases shown in Figure 1(A), we mapped the time-series gene expression information to these regulatory networks shown in Figure 1(B). By employing a dynamical Bayesian network model, we generated a likelihood value to measure the consistency between the regulatory network structure and the gene expression data (shown in Figure 1(C)). In Figure 1(D), each of the reference regulatory networks is assigned a statistical significance of consistency with the time-series gene expression profiling. The significant networks are the outputs of these identified responsive gene regulatory networks.

B. Data sets

We implemented our method in both simulated data and real time-series gene expression data of circadian rhythm. In the simulation study, we used a gene regulatory network and its expression from the DREAM *in silico* network challenge [12], [17], which is a competition of reverse engineering to infer the simulated regulatory network from its generated gene expression data [17]. For the availabilities of standard regulatory network and its time-series gene expression, we evaluated the consistency between the network structure and its gene expression. One regulatory network and one generated gene expression data were chosen for evaluation. The gene regulatory network contains 10 genes and 12 regulations as shown in Figure 2(A). The gene expression profiling of time-series data contains 21 time points in two conditions, i.e., perturbation and normal [17].

The true time-series gene expression data is about circadian regulation in rat lung which was downloaded from NCBI GEO database (ID:GSE25612) [2]. The gene expression profiles were generated from lungs of Wistar rats by Affymetrix microarray (Rat Genome 230 2.0), which was designed for examining fluctuations in gene expression in lungs within the 24 hour circadian cycle in normal animals [18]. The data contains 18 selected time points in the 12:12 (light:dark) cycle. To build the reference gene regulatory networks for evaluation, we downloaded the KEGG pathways in rat [9]. We built the regulatory networks by the extracted information for every interaction between two genes. The linkages of 'GErel' relationship with activation and repression information are used to construct the regulatory relationships between transcription factors and target genes [9]. In total, there are 207 KEGG pathways which can identify their gene expression information in the rat time-series gene expression data and resulted in 37 gene regulatory networks which contain more than 5 genes. These networks formed the reference regulatory networks which are used to identify the consistency between network structures and gene expressions. Compared to inferring gene regulatory networks in reverse engineering, we identify these significantly responsive regulatory networks in the gene expression profiles of circadian rhythm in a forward manner.

C. Significance of networks

In a graphical model, the joint distribution probability of a certain directed network architecture can be represented as a product of the individual density functions with conditions on their parent variables by recursive factorization [8], [15], [20], i.e. $f(G) = f(X_1, X_2, ..., X_n) = \prod_{i=1}^{n} f(X_i|parent\{X_i\})$ in graph G. Let $\mathbf{X}^t = (X_1^t, ..., X_n^t)^{\mathrm{T}}$ be the gene expression vector of n genes at time t. Thus, for the time points $\{1, ..., t, ..., T\}$, under the Markov assumption that $\mathbf{X}^{t+1}$ is independent of $\mathbf{X}^{t'}$ for $t' < t$ given $\mathbf{X}^t$, we have

$$f(\mathbf{X}^1, ..., \mathbf{X}^t, ..., \mathbf{X}^T) = f(\mathbf{X}^1) \prod_{t=2}^{T} \prod_{i=1}^{n} f(X_i^t|parent(X_i^t))$$

in the time series data. Assume $\mathbf{X}^{t+1} = \mathbf{A}\mathbf{X}^t + \mathbf{E}$, where $\mathbf{A} = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$, $a_{i,j} = P(X_i^{t+1}|X_j^t)$;

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

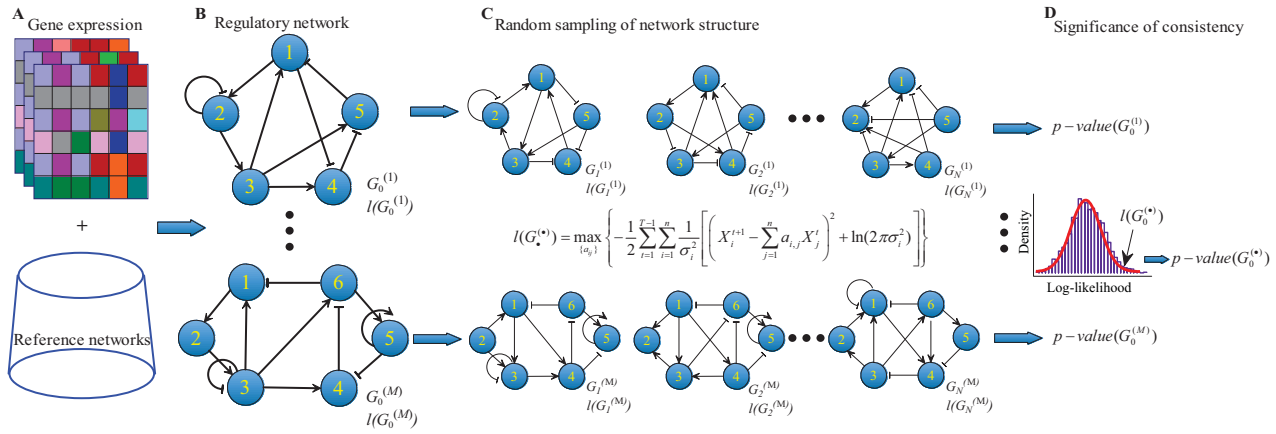


Fig. 1. Framework to identify significant regulatory networks in response to condition-specific gene expressions.

$\mathbf{E} \sim N(0, \Sigma)$, and $\Sigma = \text{diag}(\sigma_1^2, ..., \sigma_n^2)$. According to the linear Gaussian model and

$$f(X_i^{t+1}|\mathbf{X}^t) = \frac{1}{\sqrt{2\pi\sigma_i^2}} exp[-\frac{1}{2\sigma_i^2}(X_i^{t+1} - \alpha_i \mathbf{X}^t)],$$

we have

$$f(\mathbf{X}^1, \mathbf{X}^2, ..., \mathbf{X}^T) = f(\mathbf{X}^1) \prod_{t=1}^{T-1} \prod_{i=1}^{n} f(X_i^{t+1}|\mathbf{X}^t)$$

$$= f(\mathbf{X}^1) \prod_{t=1}^{T-1} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} exp[-\frac{1}{2\sigma_i^2}(X_i^{t+1} - \sum_{j=1}^{n} a_{i,j} X_j^t)^2].$$

Then, the log-likelihood function

$$\ln f(\mathbf{X}^1, ..., \mathbf{X}^t, ..., \mathbf{X}^T) \propto$$
$$-\frac{1}{2} \sum_{t=1}^{T-1} \sum_{i=1}^{n} \frac{1}{\sigma_i^2} [(X_i^{t+1} - \sum_{j=1}^{n} a_{i,j} X_j^t)^2 + \ln(2\pi\sigma_i^2)].$$

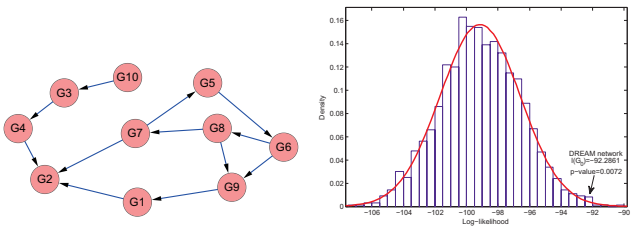
We then employed a quadratic programming method to calculate the likelihood value by optimizing the coefficients $a_{i,j}, (i, j = 1, ..., n)$ in graph G (Shown in Figure 1). Thus, the likelihood value was determined by the time series gene expression data. Based on the log-likelihood value, the significance of a specific network architecture was evaluated by a random sampling process [15]. As shown in Figure 1, for each regulatory network, we randomly generated N networks by rewiring the same number of regulations between the nodes of the evaluating network. After fitting the log-likelihood values of the random network structures by a Gaussian distribution function, we calculated the consistency probability between the evaluating regulatory architecture and the gene expression profiling for each network individually. The statistical significance p-value of one regulatory network $G_0^{(i)} (i = 1, ..., M)$ was calculated by a two-tailed test with the null hypothesis that its log-likelihood is equal to the mean of that of these randomly generated networks. We set $N = 2000$ in this work and the significant threshold of p-value was set as 0.05. All regulatory networks were implemented in the same process to get their impacts of consistency with gene expression profiles individually. The ranking by the significance p-value is clearly able to provide the enrichment measure of these regulatory structures in response to the time-series gene expression profiling of circadian rhythm.

III. Results

A. The simulation study

Firstly, we presented the simulation results to demonstrate our method by evaluating the gene regulatory network in the DREAM challenge [17], [25]. The simulated gene expression data were generated by the designed regulatory network structures shown in Figure 2(A), which contains a cycle of '$gene5 \rightarrow gene6 \rightarrow gene8 \rightarrow gene7$'. We implemented our method to access the consistency between the simulated time-series expression data and the gene regulatory network. After achieving the log-likelihood values of the evaluating gene regulatory network, each randomly rewired regulatory structure was also calculated for its likelihood value of measuring its consistency with the gene expression profiling in the permutation study. For the two conditions of perturbation and normal in the simulated gene expression profiles, the standard network structure achieved its significance p-values of 0.0072 and 0.0034 in the two conditions respectively. In the perturbation data, the distribution of likelihood value in these random samplings of network structures is shown in Figure 2(B). The results provided evidence the high consistency between network structures and its corresponding gene expression data. Compared to the original goal in the *in silico* network challenge of inferring gene regulatory network from simulated expression data, we evaluated the significance of the consistency between the regulatory structures and gene expressions. From the results, we identified the consistency underlying the structure of regulations with the gene expression data. The significant gene regulatory network structure responsive to specific gene expression was identified effectively. The results also indicate the rationale of inferring gene regulatory networks from expression data.

978-1-4673-4398-5/12/$31.00 ©2012 IEEE



(a) (b)

Fig. 2. (a) Network architecture of a gene regulatory network in DREAM challenge. (b) The density distribution of the log-likelihood values between network structures and gene expressions in the perturbation condition.

B. *Significant regulatory networks in real gene expression data*

To test the effectiveness of our method in real time-series gene expression data of circadian rhythm, we implemented the proposed method to identify the significantly responsive regulatory networks enrolled from KEGG [9]. Figure 3 shows the documented regulatory network of circadian rhythm. The gene expression data of circadian rhythm in rat lungs which can be divided into two segments, i.e., light and dark, in the rhythm of 24 hours. Each knowledge-based gene regulatory network was evaluated by calculating its consistency value with the gene expression profiling data in the light and black segments, respectively.

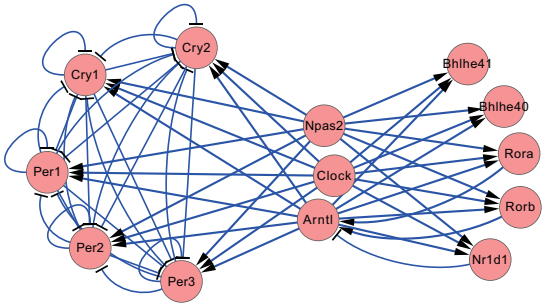
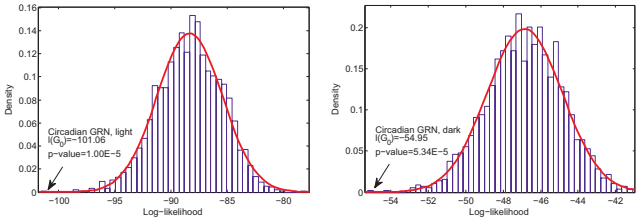


Fig. 3. The documented gene regulatory network of circadian rhythm in KEGG.

Table I lists the significance p-values of these reference gene regulatory networks in response to the circadian rhythm gene expression profiles. They are simply ranked by the significance p-values in the light. In the evaluation, these documented regulatory networks were screened to be significant network structures by their consistency with the gene expression data. The enriched regulatory architectures in response to the gene expression were identified simultaneously. We found that the regulatory network of 'circadian rhythm - mammal' has been identified as one of the most significant networks in both of the segments of light and dark. The significant regulatory network of 'tuberculosis' indicates the active regulations in this pathway under the rhythm transition of day and night in lung cells. It is consistent with the knowledge of circadian oscillation in gene expression in lungs [18]. The significant 'pathways of cancer' illustrates the active regulation associations between genes in cancer pathways in response to the circadian rhythm, which also indicates the importance of circadian rhythm for cancer [23]. These significant networks as well as the 'wnt signal pathway' also imply the interplay among these regulatory networks. The crosstalk between pathways is often crucial to generate complex responses to allow the global regulations for specific mechanisms [10]. The enriched regulatory architectures might be highly related to the circadian rhythm of rat lung cells in the light and dark conditions. Interestingly, we found that 'peroxisome proliferator-activated receptors (PPAR) signaling pathway' is significant in the dark, while is not significant in the light. The pathway is known to be important in the clearance of circulating or cellular lipids [9]. This indicates that the specific regulations are related to lipid metabolism during the night in the rhythm. In contrast, 'Jak-STAT signaling pathway' was identified as significant in the light, while not in the dark. The different significance of them in the two segments hints the different active modulation of regulations in response to the rhythm of different conditions [23].

The architecture of regulations was measured by their consistency with the time-series gene expression data. Each knowledge-based regulatory network achieved its significance evaluated by the p-values of measuring the match between network structures and expression profiles. In our scheme, we randomly rewired the regulatory linkages among these genes by keeping the number of regulations. The significance has been identified in a statistical test framework. Figure 4 shows the log-likelihood density plots for the regulatory network of circadian rhythm in the permutation study. Compared with random samples, we found that the likelihood value of the known regulatory structure is located at the far left of a bell-shape-like normal distribution. The statistical significance of rejecting the null hypothesis was calculated for the known regulation structure. The circadian rhythm regulation relationships are significant in both segments of the light and the black. This provided more evidence for the effectiveness of our method of identifying the consistency between network architecture and gene expression. The results show the importance of these gene regulations during the temporal stages of rhythm. Also, the stability of gene regulation networks has been evaluated in the permutation processes because there are few structures which can achieve higher likelihood values in response to the specific gene expression data.



(a) (b)

Fig. 4. Density distribution of log-likelihood values in the permutation study for regulatory network of circadian rhythm (a) in the light; (b) in the dark.

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

TABLE I
EVALUATION OF GENE REGULATORY NETWORKS IN RESPONSE TO THE GSE25612 GENE EXPRESSION DATA.

KEGG ID	Descriptor	Node	Edge	Light		Dark	
				P-value	FDR	P-value	FDR
rno05152	Tuberculosis	42	163	7.59E-08	3.11E-06	1.04E-03	9.40E-03
rno04710	Circadian rhythm – mammal	13	58	1.00E-05	1.37E-04	5.34E-05	1.09E-03
rno05200	Pathways in cancer	82	170	6.99E-05	4.29E-04	9.84E-03	0.0576
rno04630	Jak-STAT signaling pathway	25	90	7.33E-05	4.29E-04	0.9450	0.9587
rno04310	Wnt signaling pathway	13	40	2.52E-03	0.0115	6.61E-03	0.0451
rno05216	Thyroid cancer	7	10	0.0210	0.0860	0.1150	0.2946
rno05213	Endometrial cancer	7	10	0.0252	0.0938	0.1101	0.2946
rno05211	Renal cell carcinoma	13	22	0.0463	0.1582	0.0981	0.2871
rno05212	Pancreatic cancer	9	20	0.0527	0.1661	0.1942	0.3520
rno05143	African trypanosomiasis	6	5	0.0874	0.2561	0.1875	0.3520
rno05215	Prostate cancer	9	7	0.1147	0.2985	0.2408	0.3901
rno04350	TGF-beta signaling pathway	20	26	0.1219	0.2985	0.1865	0.3520
rno04510	Focal adhesion	5	6	0.1238	0.2985	0.0319	0.1191
rno04978	Mineral absorption	6	5	0.1547	0.3375	0.2185	0.3732
rno05217	Basal cell carcinoma	19	38	0.1564	0.3375	0.7837	0.8902
rno04916	Melanogenesis	15	14	0.2001	0.3935	0.3057	0.4643
rno04976	Bile secretion	13	12	0.2102	0.3935	0.4905	0.6487
rno05134	Legionellosis	14	12	0.2244	0.3935	0.4397	0.6217
rno05218	Melanoma	6	5	0.2257	0.3935	0.9124	0.9587
rno04210	Apoptosis	8	12	0.2436	0.3935	0.1367	0.3297
rno04150	mTOR signaling pathway	6	5	0.2560	0.3935	0.7494	0.8778
rno04961	Endocrine and other factor-regulated calcium reabsorption	6	7	0.2614	0.3935	0.8818	0.9514
rno04960	Aldosterone-regulated sodium reabsorption	16	15	0.2645	0.3935	0.1703	0.3520
rno04340	Hedgehog signaling pathway	22	40	0.2773	0.3935	0.7093	0.8778
rno04962	Vasopressin-regulated water reabsorption	7	6	0.2783	0.3935	0.7470	0.8778
rno05031	Amphetamine addiction	13	28	0.2883	0.3941	0.1974	0.3520
rno05222	Small cell lung cancer	25	38	0.2997	0.3964	0.0270	0.1107
rno04910	Insulin signaling pathway	6	5	0.3211	0.4114	0.0200	0.0913
rno04950	Maturity onset diabetes of the young	20	28	0.5106	0.6214	0.9587	0.9587
rno04115	p53 signaling pathway	32	31	0.5153	0.6214	0.4777	0.6487
rno04620	Toll-like receptor signaling pathway	12	11	0.6041	0.7077	0.6175	0.7912
rno05220	Chronic myeloid leukemia	5	3	0.7017	0.7992	0.3982	0.5830
rno04110	Cell cycle	20	26	0.7254	0.8038	0.0495	0.1690
rno04920	Adipocytokine signaling pathway	9	10	0.8524	0.8997	0.8034	0.8902
rno05030	Cocaine addiction	16	22	0.8558	0.8997	0.1461	0.3328
rno03320	PPAR signaling pathway	64	258	0.9756	0.9956	5.26E-13	2.16E-11
rno05221	Acute myeloid leukemia	12	10	0.9956	0.9956	0.2473	0.3901

IV. DISCUSSIONS

In this work, we proposed a novel dynamical Bayesian network model to identify the consistency between the structures of regulatory relationship and the gene expression data. The results show that our method can effectively identify the significantly responsive regulatory networks both in simulated data and in real gene expression data. The simulation study indicates the feasibility and efficiency of our method in the gold standard network and its generated data. The results in real data provide biological regulations which are consistent with the knowledge about circadian rhythm and were also validated by experimental data. Our method can be used to identify large-scale regulatory networks without any constraints of acyclic and loop-less regulations. Note that the regulatory cycles and loops usually exist widely in biological systems.

A. From reconstruction to evaluation

Due to the complexity of gene expression, the methods for reconstructing gene regulatory network encounter the difficulties not only from the dimensional curse of high-throughput data, but also from various assumptions underlying these genes [12]. Based on the knowledge-based regulatory networks, we measured the consistency between their architectures and expressions, which provided a powerful alternative to investigate the regulatory relationships from gene expressions. We assessed the significance of these reference networks by their structures. The identified significance of these documented networks provides the implication of responsive regulations in ceratin conditions. In our method, the likelihood of network structures meeting the time-series expression information provides the sequential checks on the architectures of gene regulations. In contrast, network reconstruction is to infer the network structure from the condition-specific gene expression data, which often has inherent barrier of environmental and phenotype flexibility. Based on the knowledge-based regulatory networks, we can identify the significantly regulatory relationships in specific conditions by network screening. Clearly, the forward-like process provide a novel approach to bridge the relationship with phenotypes and molecular data.

B. Effect of network structure

We measured the consistency between the network structure and time-series gene expression in a dynamical Bayesian network formulation. In particular, we tested the significance of the maximum likelihood score between network structure and gene expression data by a random sampling process. The random samples are based on the same gene sets with the random rewiring of linkages between these genes, i.e., the same number of regulations will be assigned in the same gene set. Each generated network was also evaluated for its likelihood of the connection architecture in response to gene expression

978-1-4673-4398-5/12/$31.00 ©2012 IEEE

profiling data. From the likelihood values of each random sample, we got the evaluation value between network structures and expressions, and the documented regulatory networks achieved their consistency measurement with gene expression data. However, in certain gene expression, the significance of gene regulatory network indicates its higher consistency with the gene expressions compared to these random networks with rewiring linkages.

C. Improvement of directed network

The proposed method of Gaussian graphical model in this paper improves our former methods for network screening on acyclic and loop-free networks of gene regulations [15], [26], and it certainly can cover more types of networks for evaluation. Apparently, it is necessary in the future to develop new theoretical model to consider the undirected networks and hybrid networks with both directed and undirected edges. As another research topic, the random samples can also be extended to identify more reasonable network structures and potential regulatory relationship by assessing the generated networks with higher significance given the available gene expressions. The network structure coherent with the expression indicates possible crucial biological meanings, which will provide valuable information for disease mechanism and drug target design.

V. Conclusion

In this work, we developed a novel graphical model to assess the consistency between regulatory network structures and gene expressions. We identified the significant regulatory networks from the documented reference networks in response to circadian rhythm conditions. The directed regulatory networks achieved their significance measurement by the consistency possibility between the network architectures of regulations and gene expression profiles. Clearly, our method provides an alternative way to detect responsive biomolecular networks corresponding to certain conditions and phenotypes. Our model can handle large-scale regulations as well as general directed networks. Moreover, our method can provide potential regulations in the networking genes. The analysis of the dynamics underlying the regulatory networks in circadian rhythm related data provides evidence for the effectiveness of our method as well as biological insights for the rhythm mechanism.

Acknowledgments

This work was partially supported by National Natural Science Foundation of China (NSFC) under Grant Nos. 31100949, 91029301, 61134013 and 61072149; by the Knowledge Innovation Program of CAS with Grant No. KSCX2-EW-R-01; by the Chief Scientist Program of Shanghai Institutes for Biological Sciences (SIBS) of CAS with Grant No. 2009CSP002, by the Shanghai NSF under Grant No. 11ZR1443100; by the Knowledge Innovation Program of SIBS of CAS with Grant No. 2011KIP203. This work was also partially supported by National Center for Mathematics and Interdisciplinary Sciences, CAS.

References

[1] Barabasi,A., Oltvai,Z. (2004) Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.*, **5**, 101–113.
[2] Barrett,T., *et al*. (2007) NCBI GEO: mining tens of millions of expression profiles – database and tools update. *Nucleic Acids Res.*, **35**, D760–D765.
[3] Chen,L., Wang,R.S., Zhang,X.S. (2009) Biomolecular Networks: Methods and Applications in Systems Biology, John Wiley & Sons, New York.
[4] Cary,M.P., Bader,G.D., Sander,C. (2005) Pathway information for systems biology. *FEBS Letters*, **579**, 1815–1820.
[5] Draghici,S., *et al*. (2007) A systems biology approach for pathway level analysis, *Genome Res.*, **17**, 1537–1545.
[6] Efron,B., Tibshirani,R. (2007) On testing the significance of sets of genes. *Ann Appl Statist*, **1**, 107–129.
[7] Herrgard,M.J., Covert,M.W., Palsson,B.O. (2003) Reconciling gene expression data with known genome-scale regulatory network structures. *Genome Res.*, **13**, 2423–2434.
[8] Koller,D., Friedman,N. (2009) Probabilistic Graphical Models. MIT Press, Massachusetts, USA.
[9] Kanehisa,M., *et al.* (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **36**, D480–D484.
[10] Liu,Z.P., *et al.* (2010) Identifying dysfunctional crosstalk of pathways in various regions of Alzheimer's disease brains, *BMC Syst. Biol.*, **4(Suppl 2)**, S11.
[11] Luscombe,N.M., *et al.* (2004) Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* **431**, 308–312.
[12] Marbach,D., *et al.* (2010) Revealing strengths and weaknesses of methods for gene network inference. *Proc. Natl. Acad. Sci. USA*, **107**, 6286–6291.
[13] Friedman,N., *et al.*(2000) Using Bayesian networks to analyze expression data. *J. Comput. Biol.*, **7**, 601–620.
[14] Rahnenfuhrer,J., *et al.* (2004) Calculating the statistical significance of changes in pathway activity from gene expression data. *Stat. Appl. Genet. Mol. Biol.*, **3**, Article16.
[15] Saito,S., Aburatani,S., Horimoto,K. (2008) Network evaluation from the consistency of the graph structure with the measured data. *BMC Syst. Biol.* **2**, 84.
[16] Subramanian,A., *et al.* (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA*, **102**, 15545–15550.
[17] Stolovitzky,G., Monroe,D., Califano,A. (2007) Dialogue on Reverse-Engineering Assessment and Methods: The DREAM of high-throughput pathway inference. *Annals of the New York Academy of Sciences*, **1115**, 11–22.
[18] Sukumaran,S., Jusko,W.J., Dubois,D.C., Almon,R.R. (2011) Light-dark oscillations in the lung transcriptome: implications for lung homeostasis, repair, metabolism, disease, and drug action. *J. Appl. Physiol.* **110**, 1732–1747.
[19] Tarca,A.L., *et al.* (2009) A novel signaling pathway impact analysis. *Bioinformatics*, **25**, 75–82.
[20] Toh,H., Horimoto,K. (2002) Inference of a genetic network by a combined approach of cluster analysis and graphical Gaussian modeling. *Bioinformatics* **18**, 287–297.
[21] Tian,L., *et al.* (2005) Discovering statistically significant pathways in expression profiling studies. *Proc. Natl. Acad. Sci. USA*, **102**, 13544–13549.
[22] Tusher,V.G., Tibshirani,R., Chu,G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA*, **98**, 5116–5621.
[23] Wang,Y., Zhang,X.S., Chen,L. (2009) A network biology study on circadian rhythm by integrating various omics data. *OMICS*, **13**, 313–324.
[24] Wang,Y., *et al.* (2006) Inferring gene regulatory networks from multiple microarray datasets. *Bioinformatics*, **22**, 2413–2420.
[25] Zhang,X.,, *et al.* (2012) Inferring gene regulatory networks from gene expression data by path consistency algorithm based on conditional mutual information. *Bioinformatics*, 28, 98–104.
[26] Zhou,H., *et al.* (2011). Network screening of Goto-Kakizaki rat liver microarray data during diabetic progression. *BMC Syst. Biol.*, **5(Suppl 1)**, S16.