

# KNOWLEDGE-BASED MODULARITY OF CELLULAR NETWORKS IN CELL CYCLE

<sup>1,2</sup>YANG YUEYING, <sup>2</sup>LIU DI, <sup>1</sup>MENG JUN

<sup>1</sup>Department of Information and Computation Science, Northeast Agricultural University, Harbin, China

<sup>2</sup>Heilongjiang Academy of Agricultural Sciences, Harbin, China

E-mail: [yangyueying2008@yahoo.com.cn](mailto:yangyueying2008@yahoo.com.cn)

## ABSTRACT

A focal point of network studies is the determination of their communities or modules, which help the functional organization and evolution of networks. Modularity can be considered as a function of a dynamic cellular system which performs complex functions in a living cell. How to identify this valuable knowledge resource to build more reliable modules remains one of the most crucial and challenging problems in Bioinformatics. We propose a state-space model combining temporal topology approach to characterize the spatio-temporal modules of the cell cycle process. Our modules discern not only the functionally related sets of genes that are conditionally activated or repressed across cell cycle process in *S.cerevisiae*, but also many different solutions which have evolved for assembling the different molecular components at the right time during the cell cycle. The module map generated by this analysis suggests several hypotheses linking biological processes to specific cell cycle conditions.

**Keywords:** *Cell Cycle, Dynamics, Modules, State-Space Model*

## 1. INTRODUCTION

The systematic application of automated high-throughput molecular biology techniques has provided an unprecedented opportunity to understand countless facets of the functional genome[1-3]. With the increasing amount of high-throughput molecular biology experimental data on gene expression and regulation[4-6], there is a growing need for identifying data relevant to the biology.

Biological systems can be described as dynamic and distributed programs[7] in which a large number of different substances under a coordinated arrangement interact and regulate cellular behaviors and functions. The molecular machinery of cell cycle control has been described as a whole dynamic system, including interconnected signaling pathways[8] and cascades of transcriptional programs[9]. In the classic case of the *S.cerevisiae*, several beautiful measurements have been taken[10,11].

Systematic models[12-15] that link the behaviour of a system to the interactions between its components will have an increasingly important role in post-genomic biology. Since the 1990s, modelling has emerged as a novel tool to handle the rapidly growing information on the underlying genetic architecture and the overwhelmingly

complex interaction circuitry of signalling networks[16-20]. The quantitative data demands quantitative models on the molecular level, from regulation of gene expression and gene function to cellular mechanisms.

As diverse fronts of genomic data continue to grow and mature, it is an opportunity for development bioinformatics, which aims to reconstruct entire biological systems through the modelling of their components. In short, the field of biological systems is still not fully explored, and especially the area of multi-time delay biological systems needs directed research efforts. The state-space model[21-24] can serve as a framework for computing these biological systems. In this paper, we have presented a state-space model combining temporal topology approach for cell cycle processes, in which the periodically expressed genes are viewed as the observation variables and the periodically expressed genes' expression dynamics in cell cycle are governed by a group of the internal variables. We concentrated our efforts on exploring cell cycle related biological features, it offers a description of the network as a continuous time dynamical system that can be used to infer the major mechanisms of the cell cycle network.

## 2. METHODS

Currently most cell cycle signatures are obtained at a specific time, providing only a snapshot of gene expression, it is difficult to understand and differentiate cause and effect from a gene-expression signature. Time-series or condition-specific data are required for further understanding of cellular dynamics. Integrative time-dependent and space-dependent approaches should be developed to infer causality and define directional pathways activated in cell cycle signatures.

The state-space model combining temporal topology approach we adopted will integrate distinct spatio-temporal characters of cell cycle process and elucidate hidden dependencies between variables. The fundamental tenet of our approach for finding modules of cell cycle signatures is that we can replace the difficult task of constructing the genome regulative network as just constructing the modules based on temporal topology of periodically expressed genes. More precisely, the approach is founded on the idea that the instantaneous disposition of the genes of interest can be established from the modules which consist of various molecules exert their relevant functions.

In this paper, we view the periodically expressed genes as the observation variables, whose expression values depend on the current internal state variables and any external inputs. Indeed, for a large number of species' cell cycle systems, the change in the internal states depends completely on the current internal states plus any external inputs, if these exist. While each internal variable has a distinct combination of observation variables, observation variables which share the same internal characteristics can be grouped together.

The mathematical description for the state-space model can be compactly represented by the general form:

$$\begin{cases} z(t + \Delta t) = A \cdot z(t) + n_1(t) \\ x(t) = C \cdot z(t) + n_2(t), \end{cases} \quad (1)$$

We constructed an observation variables matrix  $x(t)$  whose rows are genes and columns are time points, in which the  $i$ -th element is the expression level of gene  $i$  at time  $t$ . The matrix  $z(t)$  consists of the internal state variables of the system, and the  $j$ -th element is used to mimic the expression value of internal element  $j$  at time  $t$ .

Variables  $Z(t)$  are usually said to be hidden because they only are accessible indirectly through

$x(t)$ . The matrices  $A$  are the state transition matrices. They provide key information on the influences of the internal variables on each other. The matrix  $C$  is the transformation matrix between the observation variables and the internal state variables. Finally, because experimental data are noisy and incomplete, the values of these parameters in the model are determined by using an optimization scheme that minimizes the system error  $n_1(t)$  and the observation error  $n_2(t)$ . The conceptual basis of this class of models is the idea that the expression level of the genes of interest can be viewed as the observation variables that the expression level associated with these genes are occupied by various internal regulatory variables.

Once the matrix  $z(t)$  characterizing the internal state variables is determined, it is a simple matter to deduce a cluster of observation variables in every internal state variable.

A second key assumption of this approach is the idea that each internal state variable's effect on the expression levels of the periodically expressed genes may not necessarily occur simultaneously. From a biological viewpoint, time delay is ubiquitous in gene regulation. But how cells regulate gene expression programs in response to temporal change is still far from being understood. In this paper, we thought that a cluster of observation variables in each internal state variable exert regulation on each other with time delay.

A certain cluster of observation variables in each internal state variable  $z_i(t)$ ,  $i=1, \dots, p$  ( $p$  is the number of the internal state variables) is represent by  $O_j^i(t) = [o_1^i(t), \dots, o_k^i(t)]^T$ ,  $o_j^i(t) \in \{x_1^i(t) \dots x_n^i(t)\}$ ,  $j=1, \dots, k$  and  $k < n$ . When  $j \neq 1$  then  $o_j^i(t) \neq o_1^i(t)$ ,  $k$  is the number of observation variables in a certain internal state variable.

We proposed the following linear temporal topology equations to model the regulation of a certain cluster of observation variables:

$$X(t+1) = \sum_{\tau=0}^{\tau_{\max}} W_{\tau} \cdot O_j^i(t-\tau) \quad (2)$$

where the vector  $X(t+1)=[x_1(t+1), \dots, x_n(t+1)]^T$  contains the expression levels of overall periodically expressed gene at time point  $t+1$ . The regulatory relationships and degrees among genes are thus captured by the  $W_{\tau} = (\omega_{uv})_{n \times k}$ ,  $(\tau=0, \dots, \tau_{\max})$  in (2) where the matrix element  $\omega_{uv}$  indicates the

power of gene  $v$  at time  $t-\tau$  influences the expression level of gene  $u$  at time  $t+1$ . A positive matrix element leads to the  $u$ -th gene being positively reinforced by the  $v$ -th gene expression level at a pretty time, vice versa.

It is well known that there should be a maximum time delay interval in a cellular network since the time of a cell cycle is limited. In this limit, an activator/inhibitor gene can regulate another gene either instantly, or in the next time slice or up to  $\tau_{\max}$  time slices (The parameter  $\tau_{\max}$  indicates the maximum delay within which a gene can regulate another gene). So the gene expression signal of each gene is correlated with other genes expression signals with a maximum lag of  $\tau_{\max}$  time slices. Then we used the least squares method combined with the control of time-delay to resolve the problem of parameter estimation, i.e. the activation and repression are characterized by these parameters  $W_{\tau}$ .

The above analysis indicates that, by quantitatively measuring the regulatory relationships and degrees among a certain cluster of observation variables and the overall periodically expressed genes with time-delay, we can achieve two important goals: 1) by state-space model, one would obtain a quantitative characterization of the internal state variables and the corresponding observation variables in every internal state variable, which is indispensable information for our analysis. As discussed previously, the compact description will facilitate quantitative higher-level study of gene clusters. 2) By describing the temporal topology approach, we can generate hypotheses on the likely mechanisms of temporal topology modules. Thus, the form of the temporal topology modules in gene expression regulation can be an effective tool to character subtle mechanisms of cell cycle systems.

We argue that this class of models is both instructive, predictive and provides an opportunity to learn something. In summary, the objective of the state-space model combining temporal Topology approach (SSTT) will integrate data on the distinct spatio-temporal dynamics of signalling from different cellular compartments and provide new insights into the behaviour of a multicomponent cell cycle system, which taking into consideration the network of interactions between the components of the cell cycle system. The SSTT model will generate new knowledge and provide opportunities to derive underlying

principles of how complex cellular networks are built. Understanding the mechanisms that underlie the functions of cellular networks will provide breakthroughs in the identification of the critical controlling factors in cellular networks.

### 3. RESULTS

Unlike random networks, the components of cellular networks consist of characteristic topological patterns that enable their functionality. These functionally related components often exert on each other, forming modules in cellular networks. It is thus crucial to understand accurately the Dynamic characteristic modules of cellular events. In our work, we used the *S. cerevisiae* dataset of timecourses microarray. For efficiency, the data used in this analysis are characterized by filtering out the total number of 600 genes in *Saccharomyces cerevisiae* which are the best possible list of periodically expressed genes. This criterion was assumed to constitute a good filter for genes which related to cell cycle. In our work, we identified these modules with the SSTT approach. We have determined the parameters of approach (as described in Methods) for the *S. cerevisiae* dataset of periodically expressed genes. The maximum time delay,  $\tau_{\max}$  is set to 3, since it is the time of 1/3 single cell cycle which is a biologically pleasing demand.

#### 3.1. Temporal Module

As in this heading, they should be Times New Roman 11-point boldface, initially capitalized, flush left, with one blank line before, and one after.

In the study that described dynamic gene expression during cell cycles, our approach depicts the simple cell cycle regulated modules which are formed by regulators and target gene at the different internal state variables of the system. The complexity of cell cycle regulated modules not only is of crucial relevance for creating cellular complexity in a developing organism, but also provides evolution with abundant opportunity for creating novelty by tinkering with regulatory information content. To find the basic temporal modules of cell cycles which enable a discussion of various regulatory relationships in a unified way, simple units consisting of a few components were enumerated. These temporal modules were exhibited by figure 1.

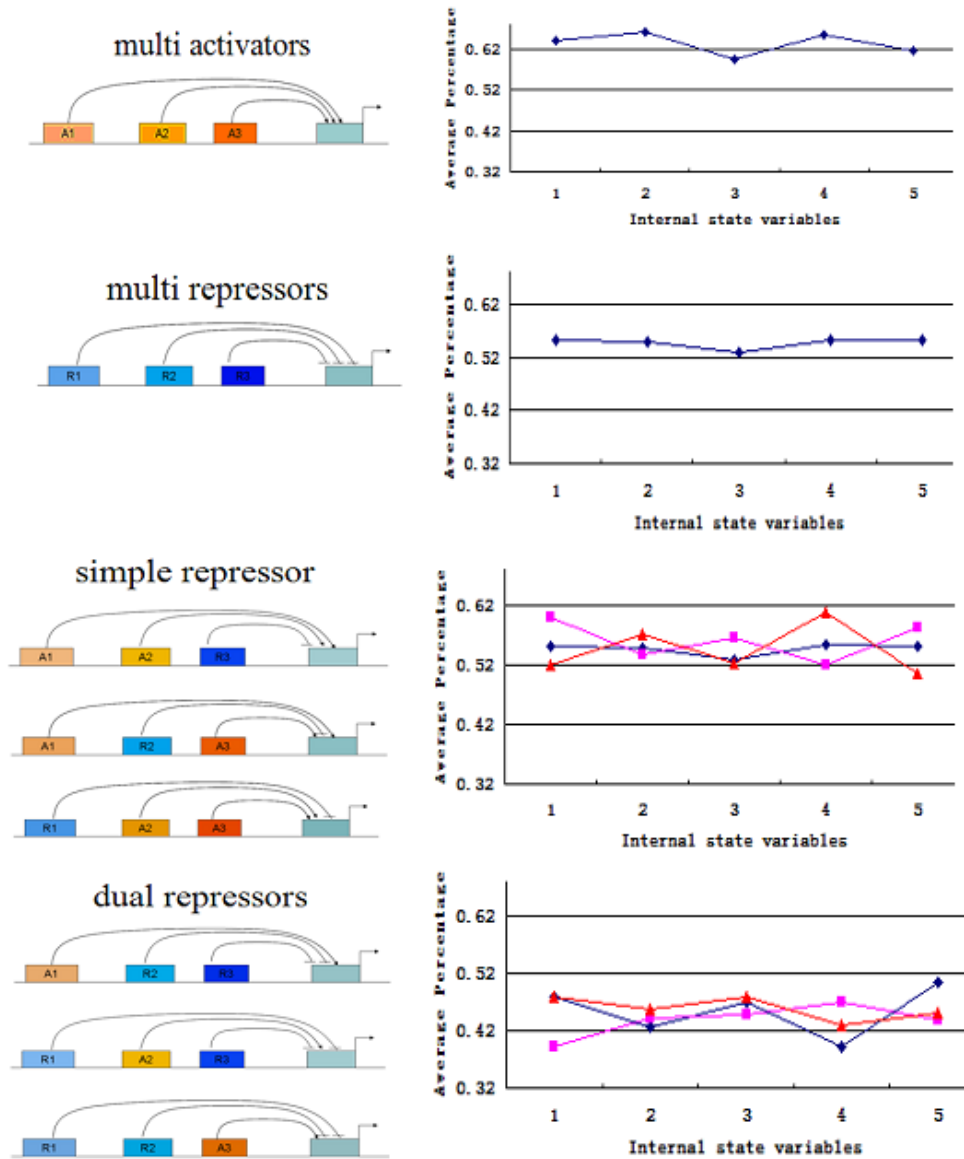


Figure 1. Several different temporal regulatory modules

The figure depicts several different dynamic temporal regulatory modules. As shown in table 1, the inverted 'T' symbol indicates inhibition, arrows represent activation, but the indirect function does not be shown. The first column on the figure 1 includes several possible mechanisms of temporal regulatory modules. The other columns depict the average ratio of the activation and repression regulative effect to target genes in accordingly temporal regulatory modules under the different internal state variables. These modules may maintain certain types of topological properties such as robustness to environmental perturbations and evolutionary conservations, highlighting the different dynamic properties required for each

condition and exhibiting a certain functional autonomy.

In the "multi activators", it seems like that the activation should far greater than the repression, but the result contrary to the expected. The activation effect is not far greater than the repression, it's just stronger a little, and ratio value fluctuates in the vicinity of 0.6. These results seem as unlikely, all activators should show strong activation rather than the average effect. We conclude that genes including in this module have the bistability and switch-like behavior to keep balance of the system. The activation effect should dominant in the "simple repressor" module, but the results are on the contrary. The first and the fifth internal state

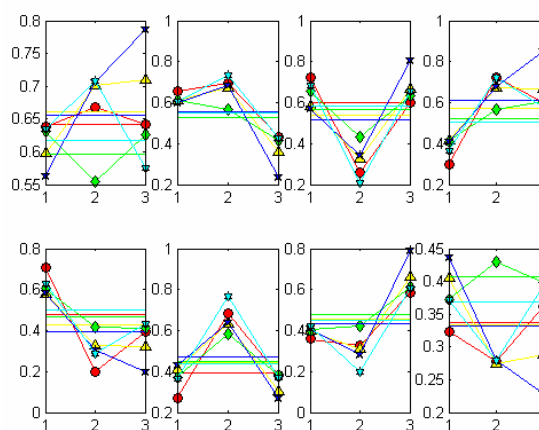
variables of the system, the repression effect is almost stronger than activators compared with the common cases. This is a habit of cell cycle to perform adaptation, desensitization and preservation of homeostasis. It is the same with the "dual repressors" module. This is the reflection of organism self-monitoring control. Specifically, the "multi activators" and "multi repressors" modules are suitable for keeping long-lasting signals to drive multi-staged. The "simple repressor" and "dual repressors" modules are suitable for initiating a quick and coordinated response to external stimuli. It is a biologically pleasing result in the precise spatial and temporal control of gene expression. Temporal regulatory modules reveal dynamic properties that contribute to cellular functions.

### 3.2. Evaluation of Temporal Modules

Using the approach as described in Methods section, we evaluate the performance of the proposed temporal regulatory modules. Each of them bears a distinct regulatory function in cellular networks. In our work, we showed time plays an important role in module regulatory behavior. Thus, analyzing how time affect on modules may shed some insights into understanding temporal regulatory modules' principles in signaling networks. We identified eight types of temporal regulatory modules (figure 1). We mapped target genes onto eight temporal regulatory modules, and discussed each type of module on different time points based on internal state variables above mentioned. For example, the network modules may have three activations to target genes at three time points, or may have two to target genes, or just one or none. For each module, we further calculated the ratio of positive actions to the negative actions (+/-) in each module for five internal state variables at three time point and compared them with the average (+/-) in all the modules, which is shown as a horizontal line in Figure 2.

The figure shows the ratio of the activation to repression regulative effect of target genes in eight temporal regulatory modules under five internal state variables in three time points. For most modules except the first and the eighth, the performance at the three time point is basically consistent. Remarkably, we found that the third internal state variable of the system (which were marked green diamond) in the first and eighth modules is significantly different with the rest internal state variables in the figure. In the first and eighth modules, the third internal state variable show clearly lower at the second time point than at

the other time points. Genes including in this module seems like to have the bistability and switch-like behavior to keep balance of the system. Cln3 was identified as an important gene of the third internal state variable, consequently it belongs to the first module. CLN3 is expressed at the late M phase, Cln3 was identified as a positive regulator of target genes SWI4 and CLN2. Similar to the immense amount of evidence showing that Cln3 is the most prominent activator of SWI4, our result displayed that Cln3 significantly regulated the expression of several G1 phase genes.



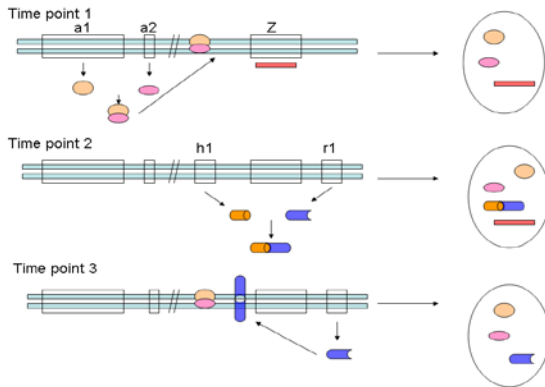
**Figure 2.** The Figure Shows The Ratio Of The Activation To Repression Regulative Power Of Target Genes Which Belong To Eight Temporal Regulatory Modules Under Five Internal State Variables In Three Time Points. (A) 1th Internal State Variable (Red), (B) 2th Internal State Variable (Yellow), (C) 3th Internal State Variable (Green), (D) 4th Internal State Variable (Blue), (E) 5th Internal State Variable (Indigo).

### 3.3. Biological Insights

The organization modules getting from the SSTT approach provides testable hypotheses that lead to biological insights.

Firstly, the uncharacterized genes or proteins belonging to a given internal variables could be functionally coherent. For instance, a certain internal variable composed of genes and gene-products involved in common functions such as DNA duplication and mitosis, suggesting strong correspondence between network topology and functionality. Moreover, we found that a series of these function concepts such as transcription regulator activity and development belonging to each internal variable will help us to explore the relationship among these genes.

Secondly, module structures provide key regulatory information[25]. In single cell cycle process, an attractive feature of combined synergy, such as transcription factors, kinase activity regulators and cyclins is that it creates extraordinary regulatory complexity of gene expression. The regulatory roles of several previously uncharacterized periodically expressed genes and signaling molecules are inferred to having the potential regulatory synergy functions.

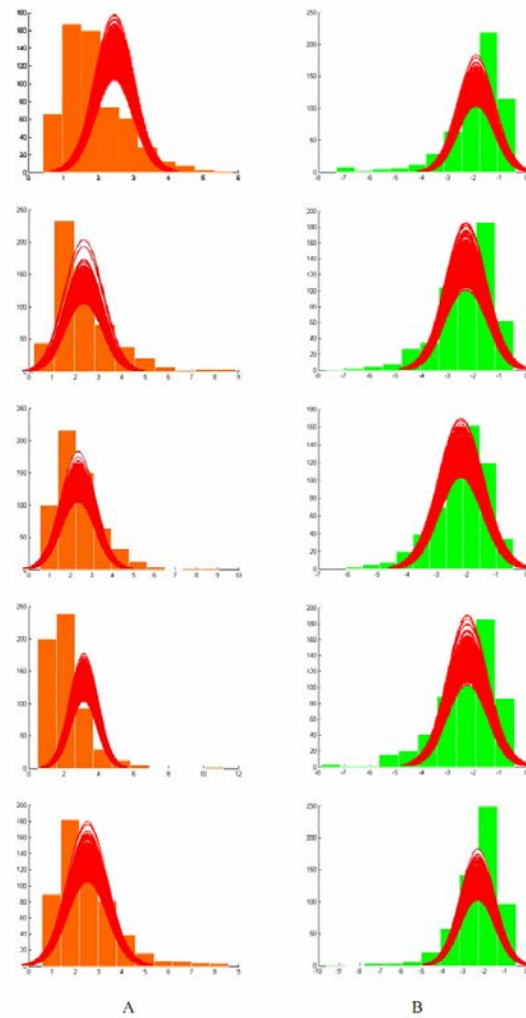


**Figure 3.** An Example Of The Combinatorial Gene Operation Based On Temporal Topology

Figure 3 shows a typical example of the combinatorial gene operation based on temporal topology. Consider the following scenario that regulates the expression program of a specific target gene Z: two activators are capable of increasing the expression level and a repressor is responsible for reducing the expression level at three different time points. Although each of the regulators might work at different time points, the overall effect of target gene activity caused by these regulators does periodically fluctuate with time and peaks at a specific time period. Thus, the periodically fluctuating mRNA levels or regulatory functions of specific regulators indeed control the expression program of their target genes in a periodic way and usually in a synergetic way.

Thirdly, it should be noted that, in cell cycle module identification, the influences in the module indicate different characteristics to five internal variables of system (see Materials and Methods). Figure 3A shows the distribution of genes which were assigned positive action, while Figure 3B shows the distribution of genes which were assigned negative action. Along with the distribution of genes which were assigned positive action (Figure 3A), we also plotted 1000 distributions obtained after randomizing each of the columns in a certain internal variables. In this

randomization we preserved the number of genes which were assigned positive action per element in a certain internal variables, yet assigned actions at random to each gene. The distributions obtained after the randomization differed markedly from the original distribution, both in terms of width and shape.



**Figure 4.** (A) Distribution Of The Number Of Genes Which Were Assigned Positive Action. (B) Distribution Of The Number Of Genes Which Were Assigned Negative Action. The Red Line Represents The Distribution Obtained After Randomizing Each Of The Columns In A Certain Internal Variables. In This Randomization We Preserved The Number Of Genes Which Were Assigned Positive Action Per Element In A Certain Internal Variables, Yet Assigned Actions At Random To Each Gene.

The distribution of genes which were affected by different internal variables of system in the original distribution (Figure 4A) has a long tail in contrast to the distributions in the randomized matrices that



looked Gaussian. Along with it is shown a set of distributions obtained by random, namely by randomly assigning effects to each gene, preserving the real number of effects to each target gene. Here, too, the original data significantly higher (or lower) than the number that would be obtained by merely preserving the statistics of number of effects to each target gene.

At last, it should be noted that, in cell cycle module identification, the influences in the module indicate physical attachment but not necessarily activation or repression. The fluctuation of mRNA expression levels may not effectively reflect regulator activities. In addition, post-transcriptional regulation and post-translational modification control protein activities as well. These effects, once they can be quantitatively determined, should be incorporated into the approach. However, actual realization of such modules remains to be explored.

#### 4. CONCLUSION

A major challenge of the post-genomic research is to understand how cellular phenomena arise from the connectivity of genes and proteins. In this paper, we describe a SSTT model that provides a compact description useful for cellular networks studies without the need to invoke the biochemical details of every component. This article, to a large extent, aims to demonstrate how a microscopic picture of the various states of the gene of interest can be mathematized using topology-based state-space mechanics.

Although our preliminary discussion is focused on the spatio-temporal mechanics of periodically expressed gene modules, the framework is the same for generic protein-DNA and protein-protein interactions network. By integrating data from different types of genome-wide experiments we are able to identify novel functional modules involving both proteins and DNA. Therefore, the validity of the derived modules depends on the quality of the experimental data, and is expected to improve as the data improve.

We demonstrate our algorithm on the *S.cerevisiae* cell cycle datasets, showing that our framework does learn to predict topology-based time delay modules. The module suggested a general mechanism where only some genes are regulated during cell cycle process and the module consist of these genes controls the timing of complex assembly.

Specially, it can be used to systematically generate hypotheses on the likely mechanisms over all phases of the cell cycle. The old models are much simpler than ours, and fails to capture important aspects such as taking consider of time delay combinatorial effects on the expression data. Our algorithm could also find highly significant modules in different organisms that it asserts are conservational, providing a strong biological basis for this claim.

#### 5. ACKNOWLEDGMENTS

This work was supported by the Heilongjiang Academy of Agricultural Sciences Project (No. LBH-Z10039), the Natural Science foundation of HeiLongJiang province (No. QC2012C034), the Science and technology projects of Heilongjiang Education Department the Doctoral Research Foundation of Northeast Agricultural University.

#### REFERENCES:

- [1] Daigo, Y. and Nakamura, Y. From cancer genomics to thoracic oncology: discovery of new biomarkers and therapeutic targets for lung and esophageal carcinoma, *Gen Thorac Cardiovasc Surg*, vol. 56, 2008, pp.43-53.
- [2] Ragoussis, J. and Elvidge, G. Affymetrix GeneChip system: moving from research to the clinic, *Expert Rev Mol Diagn*, vol. 6, 2006, pp.145-152.
- [3] Ryu, D.D. and Nam, D.H. Recent progress in biomolecular engineering, *Biotechnol Prog*, vol. 16, 2000, pp. 2-16.
- [4] Bailey, W.J. and Ulrich, R. Molecular profiling approaches for identifying novel biomarkers, *Expert Opin Drug Saf*, vol. 3, 2004, pp.137-151.
- [5] R. Shanmugavadivu and Dr.N.Nagarajan, Unsupervised clustering of gene expression profiles of microarray data using lim and pca in bioinformatics, *Journal of Theoretical and Applied Information Technology*, vol. 43, 2012, pp. 94-102.
- [6] Fischer, H.P. Towards quantitative biology: integration of biological information to elucidate disease pathways and to guide drug discovery, *Biotechnol Annu Rev*, vol. 11, 2005, pp.1-68.
- [7] Xiong, M., Arnett, F.C., Guo, X., Xiong, H. and Zhou, X. Differential dynamic properties of scleroderma fibroblasts in response to perturbation of environmental stimuli, *PLoS ONE*, vol. 3, 2008, pp.e1693.



- [8] Li, X., Wang, H., Touma, E., Rousseau, E., Quigg, R.J. and Ryaby, J.T. Genetic network and pathway analysis of differentially expressed proteins during critical cellular events in fracture repair, *J Cell Biochem*, vol. 100, 2007, pp.527-543.
- [9] Luscombe, N.M., Babu, M.M., Yu, H., Snyder, M., Teichmann, S.A. and Gerstein, M. Genomic analysis of regulatory network dynamics reveals large topological changes, *Nature*, vol. 431, 2004, pp. 308-312.
- [10] Rubinstein, A., Gurevich, V., Kasulin-Boneh, Z., Pnueli, L., Kassir, Y. and Pinter, R.Y. Faithful modeling of transient expression and its application to elucidating negative feedback regulation, *Proc Natl Acad Sci U S A*, vol. 104, 2007, pp.6241-6246.
- [11] Vrabioiu, A.M. and Mitchison, T.J. Structural insights into yeast septin organization from polarized fluorescence microscopy, *Nature*, vol. 443, 2006, pp. 466-469.
- [12] Ishii, N., Robert, M., Nakayama, Y., Kanai, A. and Tomita, M. Toward large-scale modeling of the microbial cell for computer simulation, *J Biotechnol*, vol. 113, 2004, pp.281-294.
- [13] Li, H. and Zhan, M. Systematic intervention of transcription for identifying network response to disease and cellular phenotypes, *Bioinformatics*, vol. 22, 2006, pp.96-102.
- [14] Yan Liu, Cheng Xu, Lidong Wang, A generalized multiframe real-time task model upon heterogeneous multiprocessors system, *Journal of Theoretical and Applied Information Technology*, vol. 47, 2013, pp 318-325.
- [15] Saucerman, J.J. and McCulloch, A.D. Mechanistic systems models of cell signaling networks: a case study of myocyte adrenergic regulation, *Prog Biophys Mol Biol*, vol. 85, 2004, pp.261-278.
- [16] Bhalla, and Iyengar, R. Functional modules in biological signalling networks, *Novartis Found Symp*, vol. 239, 2001, pp.45-51.
- [17] Bleakley, K., Biau, G. and Vert, J.P. Supervised reconstruction of biological networks with local models, *Bioinformatics*, vol. 23, 2007, pp.57-65.
- [18] Faeder, J.R., Blinov, M.L., Goldstein, B. and Hlavacek, W.S. Combinatorial complexity and dynamical restriction of network flows in signal transduction, *Syst Biol Stevenage*, vol. 2, 2005, pp.5-15.
- [19] Galperin, M.Y. Bacterial signal transduction network in a genomic perspective, *Environ Microbiol*, vol. 6, 2004, pp.552-567.
- [20] Kollmann, M., Lovdok, L., Bartholome, K., Timmer, J. and Sourjik, V. Design principles of a bacterial signalling network, *Nature*, vol. 438, 2005, pp.504-507.
- [21] Haavisto, O., Hytyniemi, H. and Roos, C. State space modeling of yeast gene expression dynamics, *J Bioinform Comput Biol*, vol. 5, 2007, pp.31-46.
- [22] Wolkenhauer, O. Mathematical modelling in the post-genome era: understanding genome expression and regulation-a system theoretic approach, *Biosystems*, vol. 65, 2002, pp.1-18.
- [23] Wu, F.X. Gene Regulatory Network modelling: a state-space approach, *Int J Data Min Bioinform*, vol. 2, 2008, pp.1-14.
- [24] Wu, F.X., Zhang, W.J. and Kusalik, A.J. Modeling gene expression from microarray expression data with state-space equations, *Pac Symp Biocomput*, 2004, pp.581-592.
- [25] Sim-hui Tee, Computational analysis of the simple sequence repeats in mucin-6 of foveolar cells, *Journal of Theoretical and Applied Information Technology*, vol. 47, 2013, pp 80-85.