Identifying Temporal Trace of Biological Process During Phase Transition

Tao Zeng, Luonan Chen*
Key Laboratory of Systems Biology, SIBS-Novo Nordisk Translational Research Centre for PreDiabetes, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031

***Abstract*—Phase transition widely exists in the biological world, such as the transformation of cell cycle phases, cell differentiation stages, cancer development steps, and so on. These are considered as the conversions of a genetic system from one phenotype/genotype to another. In previous studies, the molecular mechanisms of biological phase transition have attracted much attention, in particular, on the different genotypes related to specific phase but less of focus on the cascade of genes' functions during the phase change. However, it is a fundamental but important mission to track the temporal characteristics of a genetic system during specific phase transition or process, which can offer clues for understanding life and advancing its quality. By overcoming the hurdles of traditional time segmentation and temporal biclustering methods, a causal process model (CPM) in the present work is proposed to study the biological phase transition in a systematic way: boundary gene estimation for gene-specific segmentation and temporal block construction for whole data division. After the computational validation on synthetic data, CPM was used to analyze the well-known *Yeast* cell cycle data to identify the time periods of six phases in two cell cycles, and revealed phase/cycle related biological processes. These primary results demonstrate that CPM is efficient comparing to traditional methods, and has potential to elucidate the genetic mechanism with more complicated phase transitions.**

I. Introduction

In the biological world, a phase transition can be defined as the transformation of a genetic system from one phenotype to another, where different phenotypes can map to distinct genotypes. For example, cell cycle is known to have four distinct phases: G1, S, G2 and M phases; cell differentiation contains different stages like cell proliferation, growth arrest and mature differentiation; and cancer development mainly involves three steps as mutation, promotion and invasion. Obviously, details of biological phase transition will offer valuable clues for understanding life and advancing its quality. Therefore, a fundamental but important mission is to track the temporal characteristics of a genetic system during a particular phase transition or biological process.

In previous studies, the molecular mechanism of biological phase transition has attracted much attention. For instance, by modulating the intracellular redox state and measuring cell cycle progression, the redox cycle within the (mammalian) mouse embryonic fibroblast cell cycle was found to maintain the metabolic processes early in G1 and activate G1-regulatory proteins ahead of entry into S phase [1]. For a well known agricultural pest as migratory locust with a phase transition from the solitary to the gregarious, many down-regulated and some up-regulated genes were found in various organs when arriving to gregarious phase [2], which provides molecular indicators and recovers genetic mechanisms of phase transition in locusts. To determine the dormancy status of raspberry buds whose developmental regulation is helpful to promote the economic values of fruit and horticultural industries, a few significant dormancy-related candidate genes for raspberry buds had been identified by principal component analysis on clones' expressions [3]. Generally speaking, these research works are mainly on the different genotypes under specific phases. Despite of those progresses, however, there is much less of focus on the cascade or sequence of genes' functions during the phase change.

In fact, one gene generally has multiple roles in biological procedures but what role at a specific time is still unclear. Thus, identifying a gene functional group which is composed of cooperative genes in biological processes or pathways is able to reveal the functional specificity of single genes. Nowadays, there is richer information on biological processes than pathways [4], but the information on biological processes lacks dynamic features compared with pathways [5]. That is why we intend to identify the dynamic model of biological processes, in particular on when and what a biological process will be in cooperation during a phase transition. Meanwhile, the newly produced temporal gene expression data indeed gives us the opportunity to unveil mechanisms of dynamic processes behind phenotype changes. Actually, the temporal dynamic model has already shown its ability to detect the presence and absence of stage/phase specific biological processes in *Yeast* cell cycle and metabolic cycle [6]. But, this model limits to the analysis on the time segmentation for all genes, and it simply uses the replicated observations (at most three times due to few data) to infer biological processes' temporal coordination. For these reasons, a bicluster-based temporal segmentation method in this paper is developed to build a causal process model (CPM) for identifying the temporal trace of biological processes during genotype reorganizations.

The construction of the causal process model includes three steps. First, it is to exhaust the specific biclusters with linear patterns and assemble them into temporal blocks representing a group of genes and their time segmentations. Then, a

* to whom correspondence should be addressed.
♭ Emails: T.Zeng: zengtao@sibs.ac.cn; L. Chen: lnchen@sibs.ac.cn.

978-1-4577-1666-9/11/$26.00

temporal block is refined to conduct functional enrichment analysis. Finally, it is to infer the sequential or cascade (causal) relation between temporal blocks by partial correlation among two groups of genes. By numerical experiments, we obtained several interesting results: on synthetic data, the ability of gene-specific temporal segmentation is confirmed; comparing with known CCC-biclustering method, the phase division of CPM is more efficient; and in the analysis of phase-related biological processes, the group of genes actually displays corresponding functional enrichment in different phases. All these findings demonstrate that CPM indeed has potential to unveil the genetic mechanism behind more complicated phase transitions.

II. Method

A. Causal process model: temporal block based on biclusters' assembler

Unlike traditional time segmentation methods requiring the same division on a time period for all genes [6], the gene-specific time segmentation is considered in the present work. That means, for different genes or dissimilar gene groups, they can have different corresponding time segmentations based on their expressions. This is why the biclustering technologies [7], [8], which can group genes and conditions simultaneously, are adopted. However, as discussed in the study of temporal dynamic model [6], the art-of-state CCC-biclustering method [9] has the weakness that it usually cannot cover all/most genes and time points. To overcome this problem, an in-house biclustering method (noted as **EBB**: **E**rror-**B**ounded **B**iclustering, and this algorithm is under another publication) is used to enumerate so-called error-bounded linear patterns (such linear pattern, as traditional shifting pattern and scaling pattern [8], can model a group of genes having similar expression change tendency) and assemble them into proposed temporal blocks by the estimation of following defined boundary genes.

The brief framework of EBB includes three main steps: (1) discretizing the raw data matrix to some 0-1 matrix by a referred element and the given error bound; (2) building a suffix tree based on 0-1 sequences encoded by rows in the above 0-1 matrix; (3) identifying the deepest right-only node in the suffix tree as a potential bicluster with error-bounded linear pattern. In fact, CCC-biclustering is also an exhaustive method [9]. But it adopts a significant trend filtering to handle with the data pre-procession and cannot guarantee to find all potential scaling patterns/linear patterns. These methodology constraints lead to loss of most low-signal patterns and some important kinds of expression patterns (e.g. linear patterns), which prohibit CCC-biclustering to explore the whole data. On the other hand, the suggested EBB method seeks linear patterns covering traditional shifting/scaling patterns [8] so that it can identify all interesting expression patterns in theory currently. Besides, EBB can also keep low-varying signals as many as possible because it uses the error bound but not the tendency bound to discrete the raw data.

As well known to us, biclusters represent similar expression behaviors of a group of genes at the same time points. However, our proposed temporal block gathers the cooperative expression behaviors during a specific time period. Qualitatively speaking, temporal block points to a kind of data sub-matrix in the original data to cover the complete biclusters as many as possible but split the known biclusters as few as possible. According to the following concepts and definitions, the genes on so-called temporal boundary are used to divide the whole data matrix into different matrices named as temporal blocks.

Definition 1 (Boundary gene and set): Given a data matrix $D = \{d_{m,n}\}_{m\in I, n\in J}$ and its a set of gene expression patterns as biclusters $\{P_i = \{(G_i, T_i)|G_i \subseteq I, T_i \subseteq J\}\}_{i=1}^{K}$, a gene g in I is on the temporal boundary at time point t only when its R value is no less than one, where the R is calculated as formula (1). And all boundary genes at every time point consist of a boundary set $\{BG(t) = \{g|R(g,t) \geq 1, g \in I\}\}_{t\in J}$.

$$R(g,t) = \frac{|\{T_i|g \in G_i, t = \min_{\tau\in T_i}\tau\}|max(1, |\{\tau|\tau \in J, \tau < t\}|)}{|\{T_i|g \in G_i, t \in T_i, t \neq \min_{\tau\in T_i}\tau\}|} \quad (1)$$

Definition 2 (Temporal block): Given a matrix data $D = \{d_{m,n}\}_{m\in I, n\in J}$ and its pattern supervised boundary set BG, the temporal block $B_i = \{(G_i, T_i)|G_i \subseteq I, T_i \subseteq J\}$ should satisfy following constraints.

(a) $\forall g \in G_i, g \in BG(\min_{\tau\in T_i}\tau)$

(b) $\forall g \in G_i, g \in I - BG(\min_{\tau\in T_i}\tau - 1)$
or $\min_{\tau\in T_i}\tau = \min_{\tau\in J}\tau$

(c) $\forall g \in G_i, g \in I - BG(\max_{\tau\in T_i}\tau)$
or $\max_{\tau\in T_i}\tau = \max_{\tau\in J}\tau$

(d) $\forall g \in G_i, g \in BG(\max_{\tau\in T_i}\tau + 1)$
or $\max_{\tau\in T_i}\tau = \max_{\tau\in J}\tau$

(e) $\forall G \subseteq G_i, T \subset T_i$, (G,T) does not satisfy constraints (a)-(d).

(f) $\forall G \subseteq I - G_i, T = T_i$, (G,T) does not satisfy constraints (a)-(d).

Some additional comparisons between the proposed temporal block and traditional bicluster will be discussed in the next section.

B. Causal process model: expansion of temporal block for functional enrichment analysis

Like the temporal segmentation, CPM gives a non-overlapping division on the whole data. That means one gene within one time point can not belong to different temporal blocks, so that, the maximal property of temporal blocks is not always guaranteed in CPM. Using Fig. 1 to be an example, $\{(g_1, g_2, g_3, g_4, g_5, g_6)\}$ have coherent expression on time points $\{(t_3, t_4, t_5, t_6)\}$. In order to reflect the different gene reorganization events happening on time points t_2 and t_3, these genes are divided into two temporal blocks during the co-expression period. This is just the over-division phenomenon in biclustering study which can supply a multi-granularity model for overlapping patterns [10]. When analyzing functional enrichment on temporal blocks, the over-divided genes should be gathered again. This can be easily realized by the expansion of temporal blocks.

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

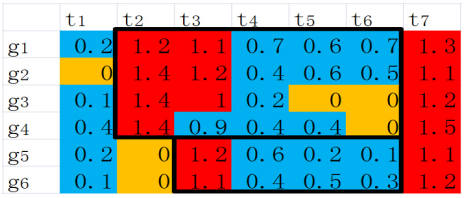

Fig. 1. Illustration of temporal blocks based on estimated boundary genes

Definition 3 (Expanded temporal block): Given a data matrix $D = \{d_{m,n}\}_{m\in I, n\in J}$ and its a temporal block $B_i = \{G_i, T_i | G_i \subseteq I, T_i \subseteq J\}$, the corresponding expanded temporal block $B_i^* = \{G_i^*, T_i^* | G_i^* \subseteq I, G_i^* \supseteq G_i, T_i^* = T_i\}$ satisfies: $\forall x \in G_i^*, \exists y \in G_i, s.t. C_{x,y} \geq p$. Where, $C_{x,y}$ represents the Pearson coefficient correlation between expression profiles of two genes during the time period T_i^*, and p is a threshold with a default value as 0.8.

Thus, the temporal blocks are useful to give a global scheme of the data division, and the expanded temporal blocks are suitable to reflect the local property of large data.

C. Causal process model: temporal block network construction based on partial correlation

In order to extract the cascade of temporal blocks representing the order of biological processes, there is a need to build a network with direction among different blocks whose qualitative connections adopt the partial correlation as the following definitions [11].

Definition 4 (Partial correlation): Given three gene expression profiles or vectors X,Y and Z, the partial correlation between X and Y under condition Z is calculated as:

$$PR(X, Y|Z) = \frac{C_{X,Y} - C_{X,Z}C_{Y,Z}}{\sqrt{1 - C_{X,Z}^2}\sqrt{1 - C_{Y,Z}^2}} \quad (2)$$

Where $C_{.,.}$ represents the Pearson coefficient correlation.

Definition 5 (Link strength between temporal blocks): Given two temporal blocks $B_1 = (G_1, T_1)$ and $B_2 = (G_2, T_2)$, if $\min_{\tau\in T_1} \tau \leq \min_{\tau\in T_2} \tau \leq \max_{\tau\in T_1} \tau + 1$, these two blocks have a link with direction from B_1 to B_2. The link strength between their referred gene expression profiles in the time period $[\min_{\tau\in T_1} \tau, \min(\max_{\tau\in T_1} \tau, \max_{\tau\in T_2} \tau)]$ can be calculated as:

$$LS(B_1, B_2) = \frac{\sum_{X\in G_1} \max_{Y\in G_2}(\min_{Z\in G_2, Z\neq X,Y} |PR(X, Y|Z)|)}{|G_1|} \quad (3)$$

This strength measurement indicates the potential partial relation from genes in a source block B_1 to genes in a target block B_2. It requires that the gene X in a source can directly interact with gene Y in a target, or be indirectly related to Y without the conduction from other target genes.

Based on the links (edges) with strengths (weights) among temporal blocks (nodes), the temporal block network (TBN) is prepared for deep analysis of dynamic biological processes.

III. Result and Discussion

Ahead of the discussion on experimental results, some additional characteristics of proposed temporal blocks comparing with traditional biclusters are illustrated as follows. Due to the module-in-focus property of biclustering, biclusters always have overlap with each other and have less size than the original data [9]. The redundancy elimination of those overlapping biclusters is still a relevant and open question in the study of biclustering. In the present work, in order to obtain a division of the original data, the temporal blocks instead of biclusters are used to build the dynamic model and constructed by boundary gene estimation, which suffers few effects from possible bicluster redundancy according to the principles of temporal block construction. Noted, temporal block is not a traditional bicluster pattern but a bicluster assembler. In other words, it does not represent the coherent expression solely as a bicluster but the similar expression pattern change events (constraint (a)) as the concept of gene reorganization across the neighbouring time windows [6]. As illustrated in the constraints (b), (c) and (d), temporal block can tolerate the potential disorder period which permits the boundary genes present at continuous time points when block begins. It can also accept the asynchronous ending period which lets those genes do not have to be on temporal boundary when block ends or they even do not belong to any original bicluster pattern yet. These advantages of temporal block are shown in details as above Fig. 1. In the matrix view (with synthetic R values) of this figure, the element in red represents its gene (row) is on the temporal boundary at its time point (column); the element in blue means the gene is not on the temporal boundary but at the starting time point of a few biclusters; the element in orange points the gene is not at the starting time point of any biclusters yet. Clearly, in Fig. 1, the temporal block $\{(g_5, g_6), (t_3, t_4, t_5, t_6)\}$ is a standard one, while the temporal block $\{(g_1, g_2, g_3, g_4), (t_2, t_3, t_4, t_5, t_6)\}$ covers a disorder period as genes (g_1, g_2, g_3) at time points (t_2, t_3) and an asynchronous ending period as genes (g_3, g_4) at time points (t_5, t_6). Constraints (e) and (f) just guarantee the completeness of a temporal block.

Besides, the time cost of CPM is mainly for the bicluster mining of temporal block construction, which is similar to CCC-biclustering with linear time complexity [9].

A. Validation of temporal segmentation by testing on synthetic data

First of all, we analyzed CPM on a synthetic data in a simple but typical strategy as previous studies. It is to produce a random data matrix with 10 rows and 15 columns. Five predefined blocks or patterns with five genes and four continuous time points are embedded into such matrix. As the recovering patterns in above synthetic data are perfect, we used a strict error bound as 0.0001 and minimum bicluster size as 3*3,3*4,4*3,4*4 respectively to run CPM method (hereafter, the annotation x*y points one bicluster contains at least x genes and y time points). Under different parameter settings, the divisions with temporal blocks on the whole synthetic data

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

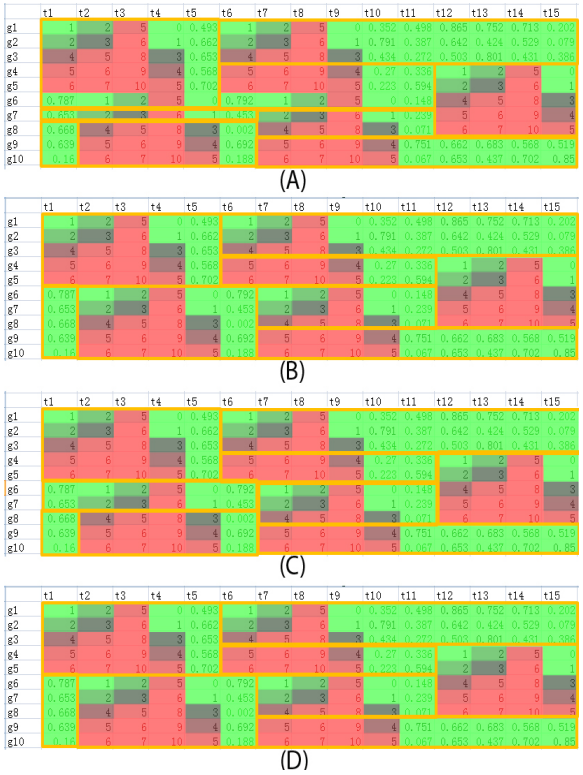

Fig. 2. The temporal blocks on the synthetic data according to CPM with different parameter settings

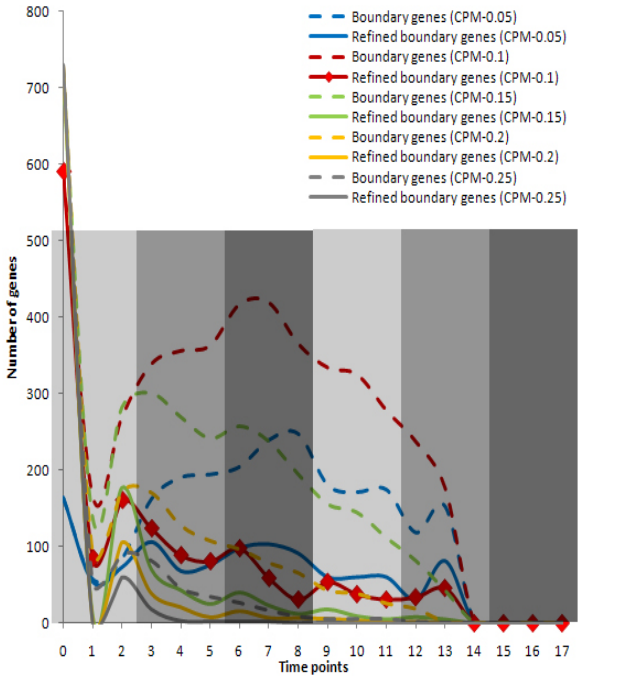

Fig. 3. Statistic view of boundary genes based on CPM with different parameter settings

are shown in Fig. 2, where one temporal block is surrounded by a yellow box. We should emphasize two points on these results. One is, for the effect of minimum bicluster size setting, the biclusters with shorter time period will lead to more sub blocks due to over-division (3*3 in Fig. 2 (A) and 4*3 in Fig. 2 (C)) than those with longer time period (3*4 in Fig. 2 (B) and 4*4 in Fig. 2 (D)), but all blocks are still reasonable. The other is, according to the proposed design principles, each temporal block can cover all time points of a predefined pattern and some asynchronous ending period (e.g. cases shown in Fig. 2), in order to tolerate the noise/error and divide the whole data in a unified way. All in all, CPM can simultaneously group genes and find gene-specific time divisions, which is always not achievable by traditional time segmentation methods. And it can further split the whole data matrix into different blocks which is disregarded in previous biclustering studies.

B. Validation of phase description by comparing with CCC-biclustering based method

Next, we analyzed CPM for the *Yeast* Cell Cycle of α-factor synchronization experiment of Spellman et al. [6], [12]. This dataset comprises two cell cycles, and each cell cycle contains three phases as M/G1, G1&S, and G2&M [6]. Every phase crosses three time points in the experiment with a constant time interval as 7 minutes. After using one-way ANOVA [13] to select genes (Setting the number of sample (time point) groups to be six with prior knowledge in six phases of two cell cycles, and the P-value to be based on the F-distribution with significant threshold as 0.05), remaining data noted as YCC with 730 genes and 18 time points was used for further analysis. Again, we used different error bounds in {0.05, 0.1, 0.15, 0.2, 0.25} and minimum bicluster size as 10*5 (experience values in previous study) to build CPMs on YCC data for extensive evaluation.

As proposed, the boundary genes can be used to track the role-change events of a group of genes, and their number would increase greatly at a time point around the alternation of phases [6]. Due to the need of covering the possible disorder period, a few boundary genes are not effective on the block construction and others are just refined boundary genes locating at the left-end (starting time point) of final temporal blocks. According to the statistic of temporal blocks and their depending boundary genes, Fig. 3 shows two kinds of distributions of boundary gene numbers under different CPM parameter settings, where the dotted line represents the distribution of original boundary genes and the solid line represents the distribution of refined boundary genes. Obviously, the distributions of numbers of refined boundary genes unveil more convincible phase related characteristics than those of original boundary genes so that the temporal block construction is indeed reasonable and boundary genes just point the refined ones in the following discussions. When the error bound is strictly set to 0.05, the peaks of distributions of boundary gene numbers always locate at the middle time of each phase when genes try to maintain their status of steady coordination (Noted, the strictest parameter setting as 0.01 results no bicluster output). When error bound is suitably set to 0.1, the peaks of distributions of boundary gene numbers

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

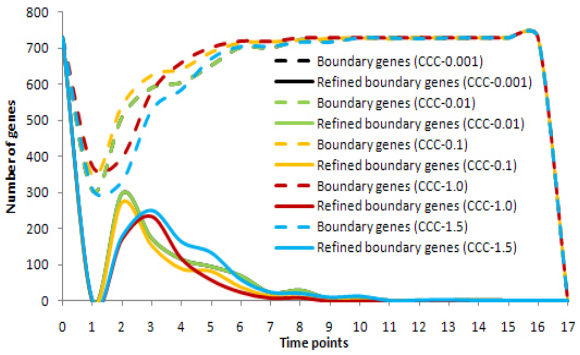

Fig. 4. Statistic view of boundary genes based on CCC-biclustering based method with different parameter settings

always locate at the time point of phase transition because genes usually start cooperation at this time and temporal block can cover the potential beginning disorder period. While, when error bound is set to 0.15 or even larger values, distributions of boundary gene numbers can not keep on their correlations with phases because many noises are introduced to mix up the genes on and not on temporal boundaries. Therefore, CPM can directly use the distributions of boundary gene numbers to track the critical time points of phase transition, whose dependent parameter setting will get benefits both from experience of data analyzers and pattern quality estimation of biclustering researches.

In order to further support the proposed (EBB)bicluster-based segmentation method is actually efficient comparing with other kinds of bicluster-based methods, we used temporal biclusters produced by CCC-biclustering [9] (under five different parameter settings and 1.0 is the default value) to assemble temporal blocks and re-analyze the relation between developmental stages and distribution of boundary gene numbers. Compared with Fig. 3, the results shown in Fig. 4 illustrate CPM is more accurate than traditional temporal biclustering based method. While, the differences between bicluster-based segmentation and traditional temporal segmentation are not discussed more here because they belong to two distinct methodology categories like biclustering and clustering.

C. Validation of phase-related biological process by functional enrichment analysis

Finally, we chosen the temporal blocks obtained under the parameter setting as 0.1 to do functional enrichment analysis. In fact, based on the single phase related temporal block, we might find the similar results as previous dynamic temporal model [6]. In this paper, the multiple phases related temporal blocks are in focus because the differences between two cell cycles after α factor handling were disregarded in previous study [6]. The 1_{st} cell cycle related temporal block TB_1 covers the former three phases with time points 0-8 and has 12 genes expanded to 432 ones. While, the 2_{nd} cell cycle related temporal block TB_2 covers the latter three phases with time points 9-17 and has 42 genes expanded to 400 ones. After functional enrichment analysis [14] on those

TABLE I
BIOLOGICAL PROCESSES ENRICHED IN TWO CELL CYCLES

Biological process	Cell cycle I	Cell cycle II
mannose metabolic process	√	
external encapsulating structure organization	√	√
cell wall organization or biogenesis	√	√
cell wall organization	√	√
cellular cell wall organization or biogenesis	√	√
cellular cell wall organization	√	√
cytokinetic cell separation	√	√
cytokinesis, completion of separation	√	√
cytokinesis	√	√
transition metal ion transport	√	
iron ion transport	√	
chromatin assembly	√	√
nucleosome assembly	√	√
DNA conformation change		√
DNA packaging		√
chromatin assembly or disassembly		√

TABLE II
BIOLOGICAL PATHWAYS ENRICHED IN TWO CELL CYCLES

Pathway	Cell cycle I	Cell cycle II
Amino sugar and nucleotide sugar metabolism	√	√
Steroid biosynthesis	√	
Fructose and mannose metabolism	√	
Regulation of beta-cell development	√	
Regulation of gene expression in beta cells	√	

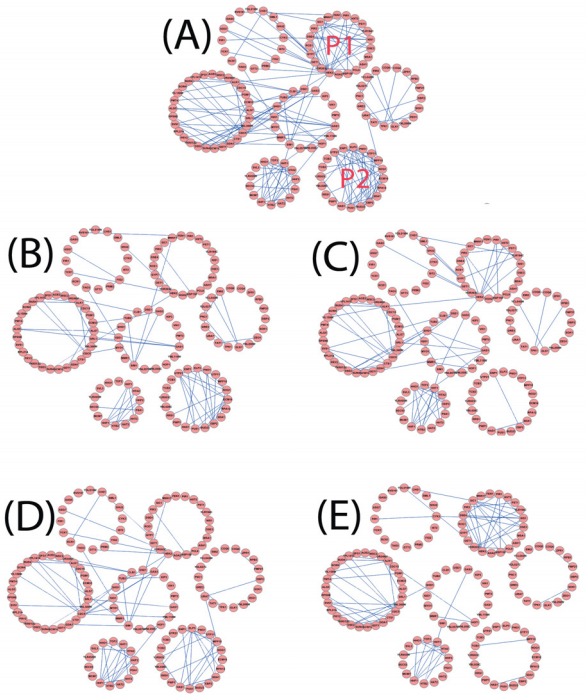

Fig. 5. The co-expression networks related to genes and cell cycles

two expanded gene sets, the significant phase(cycle)-related biological processes and pathways are listed in Table I and II. Obviously, the 1_{st} cell cycle related genes and 2_{nd} cell cycle related genes have shown several different biological processes annotated in GO [15], and the 1_{st} cell cycle related genes are frequently observed in biological pathways annotated in KEGG and Reactome [16], [17]. Therefore the two cell cycles can be just thought as two super-phases with distinct genetic properties, and this viewpoint will be helpful to understand the complicated biological procedure across multiple cell cycles.

Furthermore, the co-expression network [18] was also used to reflect the cell cycle specificity on the structures of protein

978-1-4577-1666-9/11/$26.00 ©2011 IEEE

interaction network (PIN). Given a group of genes and their expression profiles, we obtained these genes/proteins' interactions from STRING database [19]; based on their expression profiles, we calculated the correlation of two proteins with an interaction; again, based on the rank from high correlation value to low value, the top-100 interactions and their proteins consist of a co-expression network.

Thus, we used the genes in TB_1 and TB_2 with their expression profiles during two cell cycles to build four co-expression networks. Fig. 5(A) displays the global network of such four cell cycle specific networks, where several possible complex structures or interaction motifs exist. Fig. 5(B) and (C) show the networks of TB_1 genes in two cell cycles respectively, where motif $P2$ present in the first cell cycle but absent in the next cell cycle. Fig. 5(D) and (E) show the networks of TB_2 genes in two cell cycles correspondingly, where motif $P1$ unobserved in former cell cycle significantly appear in the following cell cycle. Interestingly, these protein interaction motifs can be thought as the dynamic markers (temporal traces) of cell cycles in a transition. And the proposed temporal blocks within causal process model are indeed useful to efficiently uncover such kind of molecular basis of a genetic system transition.

Of course, CPM can give the link strength between above biological processes related to two cell cycles, but the actual biological mechanism behind such candidate causal processes needs further study.

IV. Conclusion

Overcoming the hurdles of traditional time segmentation and temporal biclustering methods, the causal process model was proposed to study the biological phase transitions in a systematic way. This novel method can not only detect potential phase transitions in real biological systems but also identify the candidate temporal traces of biological processes during the transformation of a genetic system.

Acknowledgment

Authors were also supported by NSFC under Grant No. 91029301 and No. 61072149. This work was supported by a fund from National Center for Mathematics and Interdisciplinary Sciences of CAS, the Knowledge Innovation Program of CAS with Grant No. KSCX2-EW-R-01. This research was partially supported by the Chief Scientist Program of SIBS of CAS with Grant No. 2009CSP002, and Aihara Project, the FIRST program from JSPS initiated by CSTP.

References

[1] S. G. Menon, E. H. Sarsour, D. R. Spitz, R. Higashikubo, M. Sturm, H. Zhang, and P. C. Goswami, "Redox regulation of the g1 to s phase transition in the mouse embryo fibroblast cell cycle," *Cancer Research*, vol. 63, no. 9, pp. 2109–2117, 2003.

[2] L. Kang, X. Chen, Y. Zhou, B. Liu, W. Zheng, R. Li, J. Wang, and J. Yu, "The analysis of large-scale gene expression correlated to the phase changes of the migratory locust," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 51, pp. 17 611–17 615, 2004.

[3] L. Mazzitelli, R. D. Hancock, S. Haupt, P. G. Walker, S. D. A. Pont, J. McNicol, L. Cardle, J. Morris, R. Viola, R. Brennan, P. E. Hedley, and M. A. Taylor, "Co-ordinated gene expression during phases of dormancy release in raspberry (rubus idaeus l.) buds," *Journal of Experimental Botany*, vol. 58, no. 5, pp. 1035–1045, 2007.

[4] A. Baudot, V. de la Torre, and A. Valencia, "Mutated genes, pathways and processes in tumours," *EMBO reports*, vol. 11, pp. 805–810, 2010.

[5] V. Saxena, D. Orgill, and I. Kohane, "Absolute enrichment: gene set enrichment analysis for homeostatic systems," *Nucleic Acids Res.*, vol. 34, no. 22, p. e151, 2006.

[6] N. Ramakrishnan, S. Tadepalli, L. T. Watson, R. F. Helm, M. Antoniotti, and B. Mishra, "Reverse engineering dynamic temporal models of biological processes and their relationships," *Proceedings of the National Academy of Sciences*, vol. 107, no. 28, pp. 12 511–12 516, 2010.

[7] S. C. Madeira and A. L. Oliveira, "Biclustering algorithms for biological data analysis: A survey," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 1, pp. 24–45, 2004.

[8] J. S. Aguilar-Ruiz, "Shifting and scaling patterns from gene expression data," *Bioinformatics*, vol. 21, pp. 3840–3845, 2005.

[9] S. C. Madeira and A. L. Oliveira, "A polynomial time biclustering algorithm for finding approximate expression patterns in gene expression time series," *Algorithms Mol Biol.*, vol. 4, p. 8, 2009.

[10] J. Gu and J. Liu, "Bayesian biclustering of gene expression data," *BMC Genomics*, vol. 9, no. Suppl 1, p. S4, 2008.

[11] A. Reverter and E. K. F. Chan, "Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks," *Bioinformatics*, vol. 24, no. 21, pp. 2491–2497, 2008.

[12] P. T. Spellman, G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. O. Brown, D. Botstein, and B. Futcher, "Comprehensive identification of cell cycle-regulated genes of the yeast saccharomyces cerevisiae by microarray hybridization," *Mol. Biol. Cell*, vol. 9, no. 12, pp. 3273–3297, 1998.

[13] A. I. Saeed, V. Sharov, J. White, J. Li, W. Liang, N. Bhagabati, J. Braisted, M. Klapa, T. Currier, M. Thiagarajan, A. Sturn, M. Snuffin, A. Rezantsev, D. Popov, A. Ryltsov, E. Kostukovich, I. Borisovsky, Z. Liu, A. Vinsavich, V. Trush, and J. Quackenbush, "TM4: a free, open-source system for microarray data management and analysis." *BioTechniques*, vol. 34, no. 2, pp. 374–378, 2003.

[14] J. Reimand, M. Kull, H. Peterson, J. Hansen, and J. Vilo, "g:profiler-a web-based toolset for functional profiling of gene lists from large-scale experiments," *Nucleic Acids Research*, vol. 35, no. suppl 2, pp. W193–W200, 2007.

[15] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock, "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium." *Nature genetics*, vol. 25, no. 1, pp. 25–29, 2000.

[16] M. Kanehisa and S. Goto, "Kegg: Kyoto encyclopedia of genes and genomes," *Nucleic Acids Research*, vol. 28, no. 1, pp. 27–30, 2000.

[17] I. Vastrik, P. D'Eustachio, E. Schmidt, G. Joshi-Tope, G. Gopinath, D. Croft, B. de Bono, M. Gillespie, B. Jassal, S. Lewis, L. Matthews, G. Wu, E. Birney, and L. Stein, "Reactome: a knowledge base of biologic pathways and processes," *Genome Biology*, vol. 8, no. 3, p. R39, 2007.

[18] S. Horvath and J. Dong, "Geometric interpretation of gene coexpression network analysis," *PLoS Comput Biol*, vol. 4, no. 8, p. e1000117, 2008.

[19] D. Szklarczyk, A. Franceschini, M. Kuhn, M. Simonovic, A. Roth, P. Minguez, T. Doerks, M. Stark, J. Muller, P. Bork, L. J. Jensen, and C. v. Mering, "The string database in 2011: functional interaction networks of proteins, globally integrated and scored," *Nucleic Acids Research*, vol. 39, no. suppl 1, pp. D561–D568, 2011.

978-1-4577-1666-9/11/$26.00 ©2011 IEEE