Available online at www.sciencedirect.com
ScienceDirect
Progress in Natural Science 18 (2008) 945–952
Progress in Natural Science
www.elsevier.com/locate/pnsc

Cell cycle-dependent gene networks relevant to cancer

Yun Xiao [a,1], Xia Li [a,b,*,1], Yueying Yang [a], Yadong Wang [c], Shaoqi Rao [a,b]

[a] *College of Bioinformatics Science and Technology, and Bio-pharmaceutical Key Laboratory of Heilongjiang Province and State, Harbin Medical University, Harbin 150086, China*
[b] *Biomedical Engineering Institute of Capital Medical University, Beijing 10004, China*
[c] *Department of Computer Science, Harbin Institute of Technology, Harbin 150080, China*

Received 27 December 2007; received in revised form 3 February 2008; accepted 1 March 2008

Abstract

The analysis of sophisticated interplays between cell cycle-dependent genes in a disease condition is one of the largely unexplored areas in modern tumor biology research. Many cell cycle-dependent genes are either oncogenes or suppressor genes, or are closely associated with the transition of a cell cycle. However, it is unclear how the complicated relationships between these cell cycle-dependent genes are, especially in cancers. Here, we sought to identify significant expression relationships between cell cycle-dependent genes by analyzing a HeLa microarray dataset using a local alignment algorithm and constructed a gene transcriptional network specific to the cancer by assembling these newly identified gene–gene relationships. We further characterized this global network by partitioning the whole network into several cell cycle phase-specific sub-networks. All generated networks exhibited the power-law node-degree distribution, and the average clustering coefficients of these networks were remarkably higher than those of pure scale-free networks, indicating a property of hierarchical modularity. Based on the known protein–protein interactions and Gene Ontology annotation data, the proteins encoded by cell cycle-dependent interacting genes tended to share the same biological functions or to be involved in the same biological processes, rather than interacting by physical means. Finally, we identified the hub genes related to cancer based on the topological importance that maintain the basic structure of cell cycle-dependent gene networks.

Keywords: Cell cycle; Microarray; HeLa cell; Gene network

1. Introduction

Cell cycling is essential for cell growth, in which gene regulations are highly conserved among the eukaryotes [1]. This process is composed of four phases: G1, S, G2 and M. In the G1 phase, the cell increases in size and prepares to reproduce its DNA. Once all the necessary molecules are collected, the internal clock moves the cell to the S phase, in which the DNA is synthesized and chromosomes are replicated. After the DNA reproduction, a gap called G2 comes, and then the cell divides. The stage in which the cell divides is called M phase for mitosis. The newly generated daughter cells immediately enter G1 to start a new cell cycle. The movement within each phase of a cell cycle and transition from one phase to the next is highly regulated by some genes. For instance, genes involved in cell cycle checkpoints can precisely control cell growth and division in normal cells [2].

In 1981, Hereford and coworkers discovered that yeast histone mRNAs oscillate in abundance during the cell cycle [3]. Such genes that behaved in a periodic manner consistent with the cell cycles are called cell cycle-dependent genes. Many cell cycle-dependent genes are involved in processes that occur only once per cell cycle. Such processes include DNA synthesis, budding and cytokinesis.

* Corresponding author. Tel.: +86 451 8661 5912; fax: +86 451 8666 9617.
E-mail address: lixia@hrbmu.edu.cn (X. Li).
[1] These authors contributed equally to this work.

1002-0071/$ - see front matter © 2008 National Natural Science Foundation of China and Chinese Academy of Sciences. Published by Elsevier Limited and Science in China Press. All rights reserved.
doi:10.1016/j.pnsc.2008.03.003

Additionally, many of these genes are involved in controlling the cell cycle itself. Such regulations might be required for the proper functioning of some mechanisms that precisely maintain the time clock during a cell cycle. Most importantly, understanding these regulations may to a large extent help us understand the pathogenesis of many diseases (e.g. cancer) in which abnormal behaviors of cell cycle are important causative reasons.

DNA microarray technology [4] has enabled researchers to simultaneously monitor the expression levels of thousands of genes. In the experiments of time-series gene expression, a comprehensive transcriptional profile of a gene can be obtained to capture its temporal trends and characteristics by measuring its activities at different time points in a cell cycle. Up to now, the genome-wide transcriptional program during the cell cycle has been investigated in a wide range of the organisms. Some previous studies [5] focused on elucidating the gene–gene relationships based on global correlation pattern over the whole time-course, and identifying clusters of genes whose expression levels simultaneously rise or fall. Obviously, the global clustering algorithm is unable to distinguish the detailed dynamic facets of a gene–gene relationship in a cell cycle. Therefore, in this study we used a local clustering algorithm [6] to identify transcriptional relationships between cell cycle-dependent genes. By applying this algorithm to a human HeLa cell cycle expression dataset [7], we identified four types of temporal relationships between cell cycle-dependent genes: simultaneous relationship, time-shifted relationship, inverted relationship and inverted time-shifted relationship. We further constructed the gene transcriptional network specific to the cancer by assembling these newly identified gene–gene relationships, and then partitioned this global network into several cell cycle phase-specific sub-networks based on the cell cycle phase of each gene. We demonstrated that the dynamic properties and behaviors of these phase-specific sub-networks could be basically depicted by the power-law node-degree distribution. Nevertheless, the average clustering coefficients of these networks were remarkably higher than those of pure scale-free networks, indicating a property of hierarchical modularity. Based on the known protein–protein interactions and Gene Ontology annotation data, the proteins encoded by cell cycle-dependent interacting genes tended to share the same biological functions or to be involved in the same biological processes, rather than interacting by physical means. Finally, by studying the topologies of these sub-networks, we revealed some important genes whose functions are closely related to cancers while some uncharacterized genes may define novel pathways in regulating the cell cycle in cancers.

2. Materials and methods

2.1. Gene expression datasets and preprocessing

The human HeLa cell cycling dataset used in this study was part of a large-scale genome-wide program of gene transcriptional profiling during the cell division in a human cancer cell line [7]. Using spotted cDNA microarrays, containing either 22,692 elements representing ~16,322 different human genes or 43,198 elements representing ~29,621 genes, the previous study had identified 1134 clones (representing 874 genes) as cell cycle-dependent genes based on the estimates of "periodicity score", which provided a starting point for further functional discovery. Because the local clustering algorithm (see Section 2) is more successful in analyzing datasets with a longer time series, we thus focused on two datasets which contain 26 and 48 time points (Thy2 and Thy3). Both datasets were synchronized by double thymidine block methods. To validate the observed temporal trends in the HeLa data, we further analyzed an additional dataset for yeast cell cycle analyzed initially by Spellman et al. [8], who obtained the genome-wide transcriptions from the *Saccharomyces cerevisiae* cell cultures that were synchronized by three different methods: cdc15, cdc28 and α-factor. We chose the cdc15 synchronized dataset for the purpose of replication, which measured transcriptional activities of yeast genes at 24 time points. Raw data for each experiment were normalized in a *z*-score transformation [9], so that for each gene the average expression value was zero and the standard deviation was unity.

2.2. Defining four types of expression relationships

Simultaneous relationship: the expression profiles of two genes are synchronous and coincident. Genes with such profiles are expected to be subject to an identical transcriptional regulation, which are sometimes called synexpression [10]. Time-shifted relationship: the profiles of the two genes are similar, but one is time-shifted or out of phase with respect to the other. The expressions of some genes may be delayed compared to others due to a time-lag in their transcription control. Inverted relationship: the profiles of the two genes are inverted. These profiles may exist where the expression of one gene inhibits or suppresses the expression of the other. Inverted time-shifted relationship: this relationship has the characteristics of both time-shifted and inverted correlations. In addition to being inverted, the profile of one gene is staggered with respect to the other.

2.3. Local clustering

Between pairs of genes, the published local clustering algorithm was used to determine the modes of their temporal relationships: simultaneous, time-shifted, inverted and inverted time-shifted, as defined previously. Briefly, this method, which was based on the dynamical programming method for local sequence alignment, was related to conventional gene expression clustering in a fashion analogous to the way that local sequence alignment (the Smith–Waterman algorithm) is derived from global alignment (Needleman–Wunsch) [6].

2.4. Statistical significance determined by permutation tests

In order to identify significant gene relationships, we thought it better to calculate proper p-values from the distribution of scores, as was conventionally done in sequence and structural alignment [11,12]. A permutation statistic method shuffling the normalized expression levels at different time points (i.e. randomly interchanging the expression levels at different time points) was used to determine the p-value for a given score. The p-value is the probability that a score derived from random profiles equal to or larger than the observed one. The smaller the p-value is, the more significant the relationship is. Finally, we selected a p-value 0.0001 as the threshold to determine significant gene relationships.

2.5. Constructing cell cycle-dependent gene networks

To construct cell cycle-dependent gene networks, we first obtained cell cycle phases of cell cycle-dependent genes which had been generated by the previous studies, and then partitioned the whole gene network into two types of cell cycle phase-specific sub-networks based on the cell cycle phase information: between-phase networks in which each pair of genes came from two different phases and within-phase networks in which all genes in the gene pairs were from the same phase. Relationships in the between-phase networks are frequently related to a transition process from one phase to the next phase, whereas relationships in the within-phase networks are associated with internal regulations within a cell cycle phase.

3. Results and discussion

3.1. Construction of cell cycle-dependent gene networks

First, we defined the gene–gene relationships for the HeLa cell cycle expression dataset [7] according to four common modes of temporal relationships (defined in Section 2). In order to obtain more reliable gene expression relationships in HeLa cells, we only selected the gene pairs within the intersection of the results identified at the significance level of p-value 0.0001 using Thy2 and Thy3 (synchronized) datasets, respectively, which consisted of 684 gene pairs (Table 1). Once the list of significantly related gene pairs was generated, we constructed a whole gene expression network including 136 nodes and 684 edges by assembling these pairs. For the network, the maximum node degree was 41, and the average degree was 10.06. As per the information on cell cycle phase(s) that two genes were involved, the network was divided into several phase-specific sub-networks, which can be categorized into two types: within-phase networks where pairs of genes came from the same cell cycle phase and between-phase networks where gene pairs came from two different cell cycle phases.

3.1.1. Between-phase gene networks

Fig. 1 shows the between-phase sub-networks where several phase-specific networks were not included in that there were not enough numbers of significant gene pairs to build the networks (e.g. S–G2). A total of 300 gene pairs were involved in these networks. Obviously, majority of relationships appeared in the G1/S–G2 (45 gene pairs), G1/S–G2/M (55 gene pairs) and G2–G2/M (158 gene pairs) networks. In each between-phase network, some relationship modes were particularly salient. For example, in the G1/S–G2 network, 84.4% gene pairs were of the inverted time-shifted relationships while in the G2–G2/M network, 75.1% gene pairs were of the time-shifted relationships.

3.1.2. Within-phase gene networks

Fig. 2 shows the within-phase networks in which we observed only few blue lines (i.e. inverted and inverted time-shifted relationships). There were a total of 384 gene pairs in these networks, of which 90.9% gene pairs were

Table 1
Topological statistics of networks

Network name	Number of nodes	Number of edges	Average degree	Degree exponent	Determination coefficient	Average clustering coefficient	Characteristic path length	Diameter
Whole network (HeLa)	136	684	10.06	1.30	0.63	0.41	2.76	8
G2–G2 network	38	167	8.79	0.94	0.59	0.57	1.95	6
G2/M–G2/M network	42	188	8.95	0.91	0.72	0.65	1.96	5
G1/S–G2 network	33	45	2.73	1.45	0.95	0.00	2.29	5
G1/S–G2/M network	42	55	2.62	1.68	0.95	0.00	2.92	7
G2–G2/M network	55	158	5.75	1.63	0.77	0.00	2.64	7
Whole network (yeast)	716	14926	41.69	1.16	0.62	0.43	2.82	9

Several networks with only a few edges were not considered. The characteristic path length of a network is the average value of the minimum distances between two nodes. It describes how closely nodes are connected within the network.

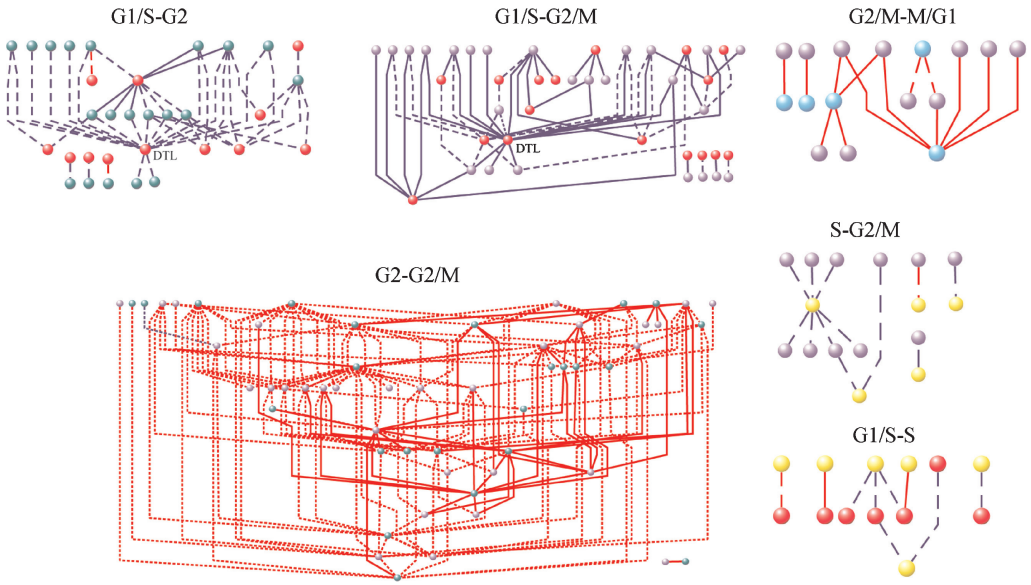

Fig. 1. Between-phase networks. Nodes and lines represent cell cycle-dependent genes and their relationships, respectively. The colors of a node are used to distinguish the cell cycle phases (red = G1/S, yellow = S, green = G2, amethyst = G2/M, cyan = M/G1). The colors for lines describe the modes of relationship: red solid line, simultaneous relationship; red dashed line, time-shifted relationship; blue solid line, inverted relationship; blue dashed line, inverted time-shifted relationship.

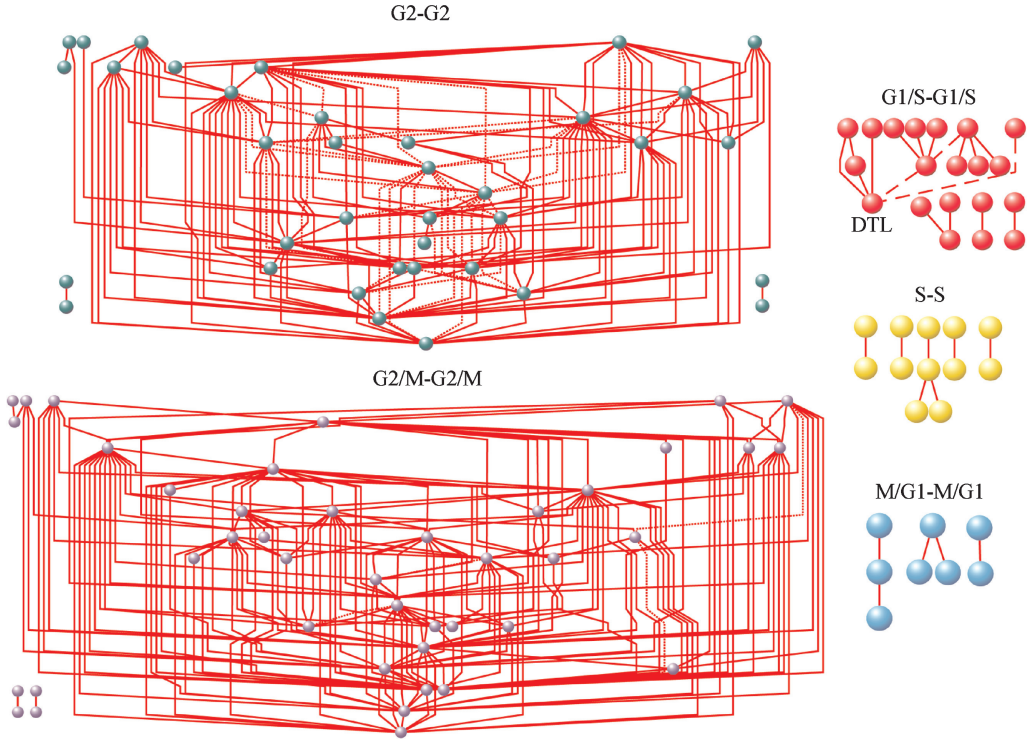

Fig. 2. Within-phase networks. Nodes and lines represent cell cycle-dependent genes and their relationships, respectively. The colors of a node are used to distinguish the cell cycle phases (red = G1/S, yellow = S, green = G2, amethyst = G2/M, cyan = M/G1). The colors for lines describe the modes of relationship: red solid line, simultaneous relationship; red dashed line, time-shifted relationship; blue solid line, inverted relationship; blue dashed line, inverted time-shifted relationship.

of the simultaneous relationships. Moreover, majority of significant gene–gene relationships were observed in the G2–G2 network (167 gene pairs) or the G2/M–G2/M network (188 gene pairs). Other three networks were much less enriched with gene pairs.

3.2. Topological properties of cell cycle-dependent networks

We further investigated many topological properties of several networks including the whole network, some phase-specific networks in HeLa cells and the whole net-

work in yeast without regard to networks with only a few edges using several graph theory-based approaches, such as the degree of nodes, the diameter of networks (Table 1). In order to test the power-law node-degree distribution property in these networks, degree exponent was estimated by measuring the linearity between $\log[P(K)]$ and $\log(K)$, where $P(K)$ is the number of nodes with degree $\geqslant K$. The determination coefficient was then computed for each network, which ranges from 0 to 1. One represents perfect linearity. As shown in Table 1, the results demonstrated that these generated networks exhibited the power-law node-degree distribution, which was consistent with previous studies focusing on various biological networks including protein–protein interaction networks [13], co-expression networks [14], metabolic networks, etc. The networks had the degree exponent between 0.91 and 1.68, and determination coefficient between 0.59 and 0.95 (0.77 on average). Particularly, the degree exponents of within-phase networks that ranged from 0.91 to 0.94 were less than those of between-phase networks ranged from 1.45 to 1.68, and the determination coefficients were also less than those of between-phase networks. In addition, the clustering coefficient defined as the ratio of the number of actual connections between the neighbors of the node to the number of possible connections was calculated. Nodes with clustering coefficient equal to 1 were part of a fully connected subgraph. As for these networks, the average clustering coefficients ranged from 0.41 to 0.65, which were several orders of magnitude greater than that of pure scale-free networks. This result suggested an underlying hierarchical organization of modularity in these networks that was consistent with previously characterized biological networks [15]. For all between-phase networks, the average clustering coefficients were zero (i.e. there were not any connections between node's neighbors) in that we did not take into account within-phase relationships in between-phase networks.

3.3. The distribution of four modes of relationships among the phase-specific networks

To analyze the distribution of four types of relationships, we defined the time interval for a network as the number of the phases separating the peaked transcriptional activities between two genes. For instance, the time interval of within-phase networks was defined to be 0, i.e. the transcriptional dependence between two genes was of no phase delay. Logically, the time interval for the G1/S–S network was shorter than the one for the G1/S–G2 network. As expected, most cell cycle-dependent genes in the within-phase networks were in a simultaneous mode (Fig. 3). With the time interval depicted by the cell cycle-dependent gene networks becoming larger, the number of gene pairs with the simultaneous relationship were reduced gradually and vanished when the time interval was more than one phase. A time-shifted relationship had a similar pattern: declined frequencies of such relationship when the time interval became larger. In contrast, inverted or inverted time-shifted relationships showed different trends from that observed above. Both inverted and inverted time-shifted relationships increased their frequencies when the time interval became larger. Almost all networks consisted of

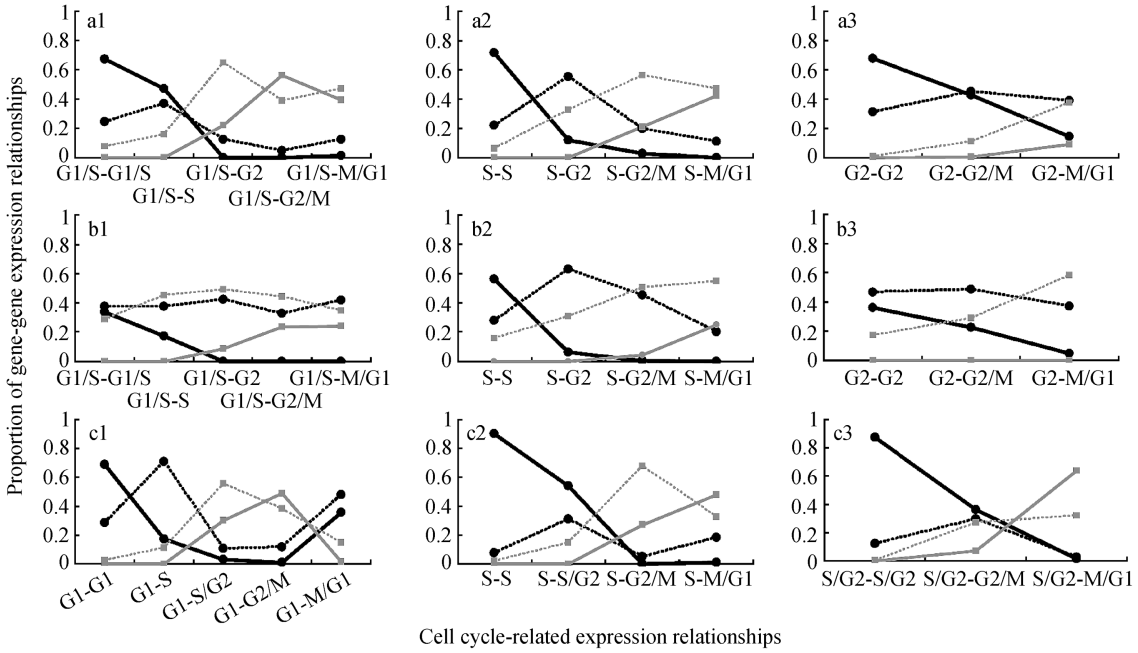

Fig. 3. The distribution of four modes of relationships among the phase-specific networks. The proportions for different types of relationships were calculated from three different time-series datasets: Thy2-HeLa (a), Thy3-HeLa (b) and cdc15-yeast (c). Black solid lines indicate simultaneous relationships; black dashed lines indicate time-shifted relationships; gray solid lines represent inverted relationships; gray dashed lines represent inverted time-shifted relationships.

different modes of relationships. To validate the observed temporal trends, we analyzed an additional dataset for yeast cell cycling (more specific, the cdc15 dataset) [8], which measured 24 time points. There were a total of 13,313 significant gene pairs identified among yeast cell cycle-dependent genes. The distributions in the yeast dataset were similar to the above results for HeLa cells. However, we observed that the proportion of the simultaneous relationship suddenly rise and the proportion of the inverted relationship suddenly fell in the G1–M/G1 network (Fig. 3c1). Because M/G1 is not only the end but also the start point in a cell cycle, and it is often difficult to identify the boundary between M and G1, we suspected that a high proportion of the genes assigned to the M/G1 phase in HeLa datasets were in fact in the M phase, while many genes assigned to the M/G1 phase in yeast dataset were in fact in the G1 phase.

3.4. Biological significance of the constructed networks

To evaluate the biological significance of these transcriptional relationships, we compared them with the known protein–protein interaction (PPI) data. The human PPI data were derived from HRPD [16] (http://www.hprd.org) and BIND databases, in which all the PPIs were manually extracted from the literatures by expert biologists. The yeast PPI data were downloaded from the DIP database [17] (http://dip.doe-mbi.ucla.edu/). Ultimately, 37208 and 17473 PPIs documented in the human and yeast interaction databases were obtained, respectively. We found, however, that only 8 of 668 (1.20%) human gene pairs and 33 of 13,313 (0.25%) yeast gene pairs had direct supports from the PPI data. These data imply that physical contact is not necessary condition for a gene–gene interaction to occur or the current PPI data do not provide such evidence.

We also annotated these gene pairs using Gene Ontology [18]. We found that 227 (33.98%) human gene pairs and 2264 (17.0%) yeast gene pairs shared a biological function(s) or were involved in the same biological processes (excluding process "cell cycle" as it is obvious). In the human HeLa dataset, 148 gene pairs participated in the same biological processes and 123 gene pairs shared an identical function(s). In the yeast dataset, 1051 pairs participated in the same processes and 1871 pairs shared an identical function(s). There were some pairs of genes that both shared a same function(s) and were involved in the same biological processes. Hence, these gene pairs are more likely to have a functional interplay rather than a direct physical binding (as evidenced via PPI). We further analyzed the annotation for each phase-specific network using GO database. The average annotation rate (the number of annotated gene pairs within the same GO terms divided by the number of all significant gene pairs) for all within-phase networks except for S phase network (no any annotation) was 31.48% (G1/S: 41.18%; G2: 21.66%; G2/M: 43.06% and M/G1: 20.0%). In between-phase networks, the average annotation rate was 30.67% (G1/S–S: 11.11%; G1/S–G2: 26.67%; G1/S–G2/M: 36.73%; S–G2/M: 41.67%; G2–G2/M: 39.24% and G2/M–M/G1: 28.57%).

3.5. Identification of key genes for HeLa cell cycle networks

Large-scale mRNA experiments offered global views of biology [19] and thus might provide insights into the role(s) of a newly identified gene involved in the underlying complicated network(s) leading to disease. To assess the importance of a gene in the network topology, we computed its linkage degree (the number of connections), and identified the key genes which were defined here as the nodes with a degree bigger than four in the corresponding networks. There were 65 key genes identified in these networks (Table 2). Majority of these hub genes were in the G2 and G2/M phases. Twenty-nine hub genes were in the G2 phase and 30 hub genes were in the G2/M phase. We found that the hub genes derived from a between-phase network almost always appeared in the corresponding within-phase networks, suggesting their importance in modulating both the within- and between-phase cellular processes. For instance, 13 of 15 G2 phase genes and 14 of 16 G2/M genes in the G2–G2/M network were also observed in the G2–G2 network and the G2/M–G2/M network, respectively. The list of key genes played very important roles in modulating the cellular processes in cancer cells. For example, in the G2 network, gene TOP2A was the most important gene. This gene encodes a DNA topoisomerase, an enzyme that controls and alters the topological states of DNA during transcription. This nuclear enzyme is involved in processes such as chromosome condensation, chromatid separation and the relief of torsional stress that occurs during DNA transcription and replication. Recent studies showed that the expression of this gene was significantly related to various cancers [20–22]. Increased expression of DNA topoisomerase II α is associated with relapses in tumors [23,24]. Gene CDCA8 encoding chromosomal passenger protein is known to play crucial roles in modulating mitosis and cell division. Inappropriate chromosomal segregation and cell division due to dysfunction of this gene may cause an aneuploidy, a known mechanism leading to cancer [25]. Gene CCNA2 binds and activates CDC2 or CDK2 kinases, and thus promotes both cell cycle G1/S and G2/M transitions and its overexpression was significantly associated with a breast cancer [26]. In the G2/M network, gene CENPF encodes a protein that is associated with the centromere–kinetochore complex, and its structure suggests that it may play a role in chromosome segregation during mitosis. Autoantibodies against this protein were found in patients with cancer. More interestingly, we found both genes DTL and PCNA present in almost all of the G1/S-related networks, including G1/S–G1/S, G1/S–G2 and G1/S–G2/M networks. The protein encoded by gene PCNA is found in the nucleus and is a cofactor of DNA polymerase delta. The encoded protein acts as a homotrimer and helps increase the processivity of leading strand synthesis during DNA replication. In response to DNA

Table 2
Hub nodes in cell cycle-dependent sub-networks

Cell cycle networks		Hub genes
Within-phase networks	G1/S–G1/S	$DTL^{5}_{G1/S}$, $PCNA^{5}_{G1/S}$
	G2–G2	$TOP2A^{20}_{G2}$,$CDCA8^{19}_{G2}$,$CCNA2^{17}_{G2}$,$ALKBH^{16}_{G2}$,$UBE2C^{16}_{G2}$,$C9orf100^{16}_{G2}$,$DKFZp762E1312^{15}_{G2}$,$KIF23^{15}_{G2}$,$KPNA2^{14}_{G2}$,$H2AFX^{13}_{G2}$, $CCNF^{12}_{G2}$,$CKAP2^{12}_{G2}$,$KIF11^{12}_{G2}$,$FLJ22624^{12}_{G2}$,$FAM72A^{12}_{G2}$,$NALP2^{11}_{G2}$,$NUSAP1^{9}_{G2}$,$CDC2^{8}_{G2}$,$BRD8^{8}_{G2}$,$BRRN1^{8}_{G2}$,$IQGAP3^{6}_{G2}$, $HLA\text{-}DRA^{5}_{G2}$,$KIFC1^{5}_{G2}$,$FZR1^{5}_{G2}$,$C12orf32^{5}_{G2}$,$C15orf20^{5}_{G2}$
	G2/M–G2/M	$MAPK13^{23}_{G2/M}$,$CENPF^{23}_{G2/M}$, $HMMR^{21}_{G2/M}$,$TPX2^{20}_{G2/M}$,$CENPE^{20}_{G2/M}$,$BUB1^{18}_{G2/M}$,$PLK1^{18}_{G2/M}$,$CDCA1^{17}_{G2/M}$,$TSN^{17}_{G2/M}$,$CKS2^{15}_{G2/M}$,$DLG7^{14}_{G2/M}$,$DEPDC1B^{14}_{G2/M}$, $ZC3HC1^{13}_{G2/M}$,$AURKA^{12}_{G2/M}$,$CNAP1^{12}_{G2/M}$,$CKAP5^{10}_{G2/M}$, $KIF2C^{6}_{G2/M}$,$KIAA1333^{8}_{G2/M}$,$CDC25B^{10}_{G2/M}$,$ARL6IP^{9}_{G2/M}$,$GAS2L3^{8}_{G2/M}$,$BIRC5^{7}_{G2/M}$,$CCNB1^{6}_{G2/M}$,$LRRC17^{6}_{G2/M}$,$TACC3^{6}_{G2/M}$,$BUB1B^{5}_{G2/M}$,$FAM64A^{5}_{G2/M}$,$MKI67^{5}_{G2/M}$
Between-phase networks	G1/S–G2	$DTL^{16}_{G1/S}$, $CCNE1^{9}_{G1/S}$, $PCNA^{5}_{G1/S}$, $KIF23^{5}_{G2}$, $CCNA2^{5}_{G2}$
	G1/S–G2/M	$DTL^{14}_{G1/S}$,$PCNA^{8}_{G1/S}$,$ZNF367^{6}_{G1/S}$, $ARL6IP^{6}_{G2/M}$
	S–G2/M	$FEN1^{7}_{S}$
	G2–G2/M	$CDCA8^{18}_{G2}$,$CCNF^{13}_{G2}$,$KPNA2^{11}_{G2}$,$TOP2A^{10}_{G2}$,$KIF11^{10}_{G2}$,$NUSAP1^{10}_{G2}$,$C9orf100^{10}_{G2}$,$ALKBH^{8}_{G2}$,$UBE2C^{8}_{G2}$, $KIF23^{8}_{G2}$,$CCNA2^{7}_{G2}$,$NALP2^{7}_{G2}$,$DKFZp762E1312^{7}_{G2}$,$MGC57827^{6}_{G2}$,$CDC25^{5}_{G2}$,$GAS2L3^{14}_{G2/M}$,$CDCA1^{12}_{G2/M}$,$TSN^{9}_{G2/M}$;
	G2/M–M/G1	$CDC25B^{8}_{G2/M}$,$CENPF^{8}_{G2/M}$,$TPX2^{8}_{G2/M}$,$ARL6IP^{8}_{G2/M}$,$BUB1B^{8}_{G2/M}$,$HMMR^{7}_{G2/M}$,$CENPE^{7}_{G2/M}$,ch-$TOG^{7}_{G2/M}$,$BUB1^{6}_{G2/M}$,$DLG7^{6}_{G2/M}$,$MKI67^{6}_{G2/M}$,$CIT^{6}_{G2/M}$,$CNAP1^{5}_{G2/M}$
		$UBE2S^{5}_{M/G1}$

Hub genes whose number of degree is bigger than 4 are listed according to cell cycle phase-specific sub-networks that are divided into within-phase networks and between-phase networks. The symbols in the bottom right corner of gene name represents cell cycle phase in which the expression peak appear. The linkage degree of gene in each cell cycle networks is shown in the upper right corner. Several networks with only few nodes were not analyzed (including G1/S–S, G1/S–M/G1, S–G2, S–M/G1, G2–M/G1).

damage, this protein is ubiquitinated and is involved in the RAD6-dependent DNA repair pathway. Data show that PCNA protein expression was significantly increased as the grade of cervical lesion becomes higher from normal epithelium to invasive squamous cell carcinomas. PCNA protein expression shows a significant positive correlation with an increasing histological abnormality [27]. The overexpression of this gene may contribute to the occurrence and progression of pancreatic cancer [28], hepatocellular carcinoma [29] and stomach tumor malignancy [30]. However, we could not find much information about gene DTL from NCBI gene database, which is only shown to code a RA-regulated nuclear matrix-associated protein. Until recently, Higa and his coworkers discovered that DTL and PCNA interact with CUL4/DDB1 complexes and are involved in CDT1 degradation after DNA damage [31]. Banks et al. also found that PCNA, DTL and the DDB1–CUL4A complex play critical and different roles in regulating the protein stability of p53 and MDM2/HDM2 in unstressed and stressed cells [31]. These wet-lab results might partially elucidate the involved mechanism(s) of gene DTL as a hub, consistent with our network analysis.

Because these microarrays analyzed in this study focused on only the cell cycle-dependent genes and do not include all human genes, some important genes which are not periodically expressed remain to be characterized, for example, the Rb and p53 tumor suppressors which are inactivated in HeLa cells as a result of binding the E6/E7 proteins of the human papillomavirus [32]. Therefore, these networks might be incomplete, because of only a limited-size sample analyzed. In addition, while the analysis of the highly scored pairs found by the local clustering can shed light on novel biological relationships, it is limited to procure support from PPI due to the inadequate information available in current databases. There are many ambiguities in the current functional classifications [33] and also there is a problem with the false positives in many of the protein–protein interaction studies, particularly the two-hybrid data [34,35].

In conclusion, we characterized gene–gene relationships between cell cycle-dependent genes into four temporal modes by applying the local clustering algorithm to a human HeLa cell cycle expression dataset. Then, we constructed the cell cycle-dependent gene networks by integrating transcriptional relationships with the cell cycle knowledge. We demonstrated that these cell cycle-dependent gene networks could be characterized well by four modes of relationships that showed distinct temporal patterns. We also found that transcription-related gene pairs were more likely to share some function(s) or to be involved in the same biological processes, but lack sufficient evidence to implicate direct physical interaction (e.g. binding). Finally, we identified the key genes based on their topological roles (connectivity) in the built networks. These function-known hub genes were supported to be closely associated to cancer in the literature. One unknown gene,

DTL, whose functions to be fully characterized, might play an important role in regulating the cell cycle waiting for further wet-lab verification.

Acknowledgements

This work was supported in part by the National High-Tech Development Project of China (Grant No. 2007AA02Z329), National Natural Science Foundation of China (Grant Nos. 30170515, 30370798, 30571034, 30570424 and 60601010), Outstanding Overseas Scientist Grant of the Heilongjiang Province (Grant No. 1055HG009), National Science Foundation of Heilongjiang Province (Grant Nos. ZJG0501, GB03C602-4 and F2004-02) and Health Department of Heilongjiang Province Key Project (2005-39).

References

[1] Murray AHT. The cell cycle. New York: Oxford University Press; 1993.
[2] Hartwell LH, Weinert TA. Checkpoints: controls that ensure the order of cell cycle events. Science 1989;246(4930):629–34.
[3] Hereford LM, Osley MA, Ludwig 2nd TR, et al. Cell-cycle regulation of yeast histone mRNA. Cell 1981;24(2):367–75.
[4] Duggan DJ, Bittner M, Chen Y, et al. Expression profiling using cDNA microarrays. Nat Genet 1999;21(Suppl. 1):10–4.
[5] Kim S, Dougherty ER, Bittner ML, et al. General nonlinear framework for the analysis of gene interaction via multivariate expression arrays. J Biomed Opt 2000;5(4):411–24.
[6] Qian J, Dolled-Filhart M, Lin J, et al. Beyond synexpression relationships: local clustering of time-shifted and inverted gene expression profiles identifies new, biologically relevant interactions. J Mol Biol 2001;314(5):1053–66.
[7] Whitfield ML, Sherlock G, Saldanha AJ, et al. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. Mol Biol Cell 2002;13(6):1977–2000.
[8] Spellman PT, Sherlock G, Zhang MQ, et al. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. Mol Biol Cell 1998;9(12):3273–97.
[9] Cheadle C, Vawter MP, Freed WJ, et al. Analysis of microarray data using *Z* score transformation. J Mol Diagn 2003;5(2):73–81.
[10] Niehrs C, Pollet N. Synexpression groups in eukaryotes. Nature 1999;402(6761):483–7.
[11] Levitt M, Gerstein M. A unified statistical framework for sequence comparison and structure comparison. Proc Natl Acad Sci USA 1998;95(11):5913–20.
[12] Pearson WR. Empirical statistical estimates for sequence similarity searches. J Mol Biol 1998;276(1):71–84.
[13] Han JD, Dupuy D, Bertin N, et al. Effect of sampling on topology predictions of protein–protein interaction networks. Nat Biotechnol 2005;23(7):839–44.
[14] Carter SL, Brechbuhler CM, Griffin M, et al. Gene co-expression network topology provides a framework for molecular characterization of cellular state. Bioinformatics 2004;20(14):2242–50.
[15] Ravasz E, Somera AL, Mongru DA, et al. Hierarchical organization of modularity in metabolic networks. Science 2002;297(5586):1551–5.
[16] Peri S, Navarro JD, Amanchy R, et al. Development of human protein reference database as an initial platform for approaching systems biology in humans. Genome Res 2003;13(10):2363–71.
[17] Salwinski L, Miller CS, Smith AJ, et al. The database of interacting proteins: 2004 update. Nucleic Acids Res 2004;32(Database issue): D449–51.
[18] Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25(1):25–9.
[19] Lander ES. The new genomics: global views of biology. Science 1996;274(5287):536–9.
[20] Amaya K, Ohta T, Kitagawa H, et al. Angiotensin II activates MAP kinase and NF-kappaB through angiotensin II type I receptor in human pancreatic cancer cells. Int J Oncol 2004;25(4):849–56.
[21] Chekerov R, Klaman I, Zafrakas M, et al. Altered expression pattern of topoisomerase IIalpha in ovarian tumor epithelial and stromal cells after platinum-based chemotherapy. Neoplasia 2006;8(1):38–45.
[22] Knoop AS, Knudsen H, Balslev E, et al. Retrospective analysis of topoisomerase IIa amplifications and deletions as predictive markers in primary breast cancer patients randomly assigned to cyclophosphamide, methotrexate, and fluorouracil or cyclophosphamide, epirubicin, and fluorouracil: Danish Breast Cancer Cooperative Group. J Clin Oncol 2005;23(30):7483–90.
[23] Lazaris AC, Kavantzas NG, Zorzos HS, et al. Markers of drug resistance in relapsing colon cancer. J Cancer Res Clin Oncol 2002;128(2):114–8.
[24] Tinari N, Lattanzio R, Natoli C, et al. Changes of topoisomerase IIalpha expression in breast tumors after neoadjuvant chemotherapy predicts relapse-free survival. Clin Cancer Res 2006;12(5):1501–6.
[25] Chang JL, Chen TH, Wang CF, et al. Borealin/Dasra B is a cell cycle-regulated chromosomal passenger protein and its nuclear accumulation is linked to poor prognosis for human gastric cancer. Exp Cell Res 2006;312(7):962–73.
[26] Aaltonen K, Ahlin C, Amini RM, et al. Reliability of cyclin A assessment on tissue microarrays in breast cancer compared to conventional histological slides. Br J Cancer 2006;94(11):1697–702.
[27] Astudillo H, Lopez T, Castillo S, et al. p53, Bcl-2, PCNA expression, and apoptotic rates during cervical tumorigenesis. Ann NY Acad Sci 2003;1010:771–4.
[28] Yue H, Jiang HY. Expression of cell cycle regulator p57kip2, cyclinE protein and proliferating cell nuclear antigen in human pancreatic cancer: an immunohistochemical study. World J Gastroenterol 2005;11(32):5057–60.
[29] Nan KJ, Guo H, Ruan ZP, et al. Expression of p57(kip2) and its relationship with clinicopathology, PCNA and p53 in primary hepatocellular carcinoma. World J Gastroenterol 2005;11(8):1237–40.
[30] Czyzewska J, Guzinska-Ustymowicz K, Lebelt A, et al. Evaluation of proliferating markers Ki-67, PCNA in gastric cancers. Rocz Akad Med Bialymst 2004;49(Suppl. 1):64–6.
[31] Banks D, Wu M, Higa LA, et al. L2DTL/CDT2 and PCNA interact with p53 and regulate p53 polyubiquitination and protein stability through MDM2 and CUL4A/DDB1 complexes. Cell Cycle 2006;5(15):1719–29.
[32] Scheffner M, Munger K, Byrne JC, et al. The state of the p53 and retinoblastoma genes in human cervical carcinoma cell lines. Proc Natl Acad Sci USA 1991;88(13):5523–7.
[33] Gerstein M. Integrative database analysis in structural genomics. Nat Struct Biol 2000;7(Suppl.):960–3.
[34] Ito T, Tashiro K, Muta S, et al. Toward a protein–protein interaction map of the budding yeast: a comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. Proc Natl Acad Sci USA 2000;97(3):1143–7.
[35] Uetz P, Giot L, Cagney G, et al. A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. Nature 2000;403(6770):623–7.