

# Construction and Clarification of Dynamic Gene Regulatory Network of Cancer Cell Cycle via Microarray Data

Cheng-Wei Li, Yung-Hsiang Chu and Bor-Sen Chen

Lab. of Systems biology, National Tsing Hua University, Hsinchu, 300, Taiwan.

---

## Abstract

**Background:** Cell cycle is an important clue to unravel the mechanism of cancer cells. Recently, expression profiles of cDNA microarray data of Cancer cell cycle are available for the information of dynamic interactions among Cancer cell cycle related genes. Therefore, it is more appealing to construct a dynamic model for gene regulatory network of Cancer cell cycle to gain more insight into the infrastructure of gene regulatory mechanism of cancer cell via microarray data.

**Results:** Based on the gene regulatory dynamic model and microarray data, we construct the whole dynamic gene regulatory network of Cancer cell cycle. In this study, we trace back upstream regulatory genes of a target gene to infer the regulatory pathways of the gene network by maximum likelihood estimation method. Finally, based on the dynamic regulatory network, we analyze the regulatory abilities and sensitivities of regulatory genes to clarify their roles in the mechanism of Cancer cell cycle.

**Conclusions:** Our study presents a systematically iterative approach to discern and characterize the transcriptional regulatory network in Hela cell cycle from the raw expression profiles. The transcription regulatory network in Hela cell cycle can also be confirmed by some experimental reviews. Based on our study and some literature reviews, we can predict and clarify the E2F target genes in G1/S phase, which are crucial for regulating cell cycle progression and tumorigenesis. From the results of the network construction and literature confirmation, we infer that MCM4, MCM5, CDC6, CDC25A, UNG and E2F2 are E2F target genes in Hela cell cycle.

---

## Introduction

The losses of cellular regulation give rise to most case of cancer. In cells, intricate genetic control systems regulate the balance between cell survival and death in response to growth signals, growth-inhibiting signals, and death signals. When some errors occur in the control systems causing cells to proliferate continuously, tumor just comes into being (Badawi et al. 2005).

The proliferation of cancer cell into two individual cells must go through cell cycle process (Whitfield et al. 2002; Valsesia-Wittmann et al. 2004). Cell cycle entails an ordered series of macromolecular events that lead to cell division and the production of two daughter cells. So that cell cycle is meaningful to the proliferation of cancer cell.

Expression levels of thousands of genes fluctuate throughout the cancer cell cycle (Cho et al. 2001; Ishida et al. 2001; Whitfield et al. 2002). Functional genes show periodic transcription to reflect cell growth, DNA synthesis, spindle pole body duplication and migration through the cell cycle (Cho et al. 1998). These processes and their regulation have been extensively investigated at the molecular level (Stillman, 1996; Nurse, 2000; Shah and Cleveland, 2000; Hinchcliffe and Sluder, 2001; Chen et al. 2004). Systems biology can be described as “integrative biology” with the ultimate goal of being able to predict de novo biological outcomes given the list of the components involved (Liu, 2005). Hence it is the coordinated study by (1) investigating the components of cellular networks and their interactions, (2) applying experimental high-throughput and whole-genome techniques, and (3) integrating computational methods with experimental effort (Klipp et al. 2005). In this situation, characterization of the genome-wide transcriptional program of the cell division cycle in mammalian cells is a critical step toward understanding the basic cell cycle processes and their roles in cancer. Therefore, it is worth investigating how these periodic patterns are regulated in the gene regulatory network of the cancer cell cycle from the systems biology perspective.

Gene expression data of Hela cell cycle have been collected (Whitfield et al. 2002) and analyzed with many clustering methods to organize which gene is associated with the cell cycle (Whitfield et al. 2002;

---

**Correspondence:** Bor-Sen Chen, Lab. of Systems biology, National Tsing Hua University, Hsinchu, 300, Taiwan.

Cho et al. 2001). Theoretically, it is possible to engineer the cell cycle network reversely, if we take cDNA expression levels as the output of gene expression networks, and collect cDNA expression levels of transcription factors as input. In order to realize how genes are regulated by transcription factors, we must also understand the interactions between target genes and their transcription factors (which transcription factor binds to which promoters). With all these information and interactive dynamic model, we get some clues to piece up the gene expression regulatory network in HeLa cell cycle.

In this study, we attempt to devise an interactive dynamic model to characterize transcriptional regulatory network of the HeLa cell cycle from the cDNA expression data of the human cell cycle in tumour (Whitfield et al. 2002). Based on our dynamic regulatory network, we not only predict the upstream regulators but also characterize the significance of the regulators depending on quantifying their regulatory abilities based on the corresponding biochemical kinetic parameters. At first, we construct a discrete-time dynamic model system and calculate the system kinetic parameters as the regulatory ability by using the expression data (Whitfield et al. 2002) and the system identification method (Johansson, 1993). Second, based on the interactive dynamic model, we detect the transcriptional regulatory function of target genes by the maximum likelihood parameter estimation method. Third, we trace back a group of upstream genes that play a role of transcriptional regulators of target genes in the HeLa cell cycle of *Homo sapiens* via deducing the interactive relationship between the expression profiles of regulators and the detected transcription regulation of specific target genes. The pathway kinetic parameters of transcriptional regulatory network of HeLa cell cycle are also estimated by the cDNA expression profiles of target genes and their upstream regulatory genes. Finally, these upstream regulatory genes are considered as the target genes. By a similar method, we construct their upstream regulatory genes one by one. Iteratively, we can construct the whole gene regulatory network of HeLa cell cycle.

We applied our method to a publicly available data set of HeLa cell with microarray experiment on cell cycle (Whitfield et al. 2002) to identify the transcriptional regulators of cell cycle and to characterize their regulatory abilities on specific target genes. By means of the quantitative system analysis of the transcriptional regulatory network from HeLa

cell cycle genes, several transcription factors were identified and their regulatory abilities were determined. Further, some genes that may be suspected of regulators of HeLa cell cycle are predicted here to be synergistic in harmonizing gene expression. Our proposed algorithm provides a novel approach to gain insight into the gene regulatory network of HeLa cell by its gene expression data using system identification technique and discrete-time dynamic model. Furthermore, we combine the constructed HeLa cell cycle dynamic network, some experimental reviews and E2F binding site's research (Ren et al. 2002), we not only confirm the reliability of the dynamic network but also find the E2F target genes in HeLa cell cycle progression. Finally, from the results of this study, we infer that E2F directly regulate MCM4, MCM5, CDC6, CDC25A, UNG and E2F2 in HeLa cell cycle progression.

Our approach is so different from the statistical clustering method that it not only provides a suitable interactive dynamic model to decipher the complex signal transduction pathway that regulates gene expressions in HeLa cell cycle but also predicts some potential regulators that have not been found. We can also quantify the regulatory abilities of the transcription factors by the corresponding kinetic activities to their target genes in the HeLa cell cycle regulatory network. We could construct the cell cycle regulatory network in HeLa cells quantitatively and discuss the sensitivity of regulatory genes to the gene regulatory network from the system analysis perspective.

## System Model and Network Identification

The construction of the transcriptional regulatory network of HeLa cell cycle can be divided into two steps. First, the transcriptional regulation should be extracted from the gene expression data by the dynamic discrete-time model. Second, the upstream regulators will be traced back by correlating the transcriptional regulation of target genes with the expression profiles of possible regulators of HeLa cell cycle. In this study, 64 transcription regulators, as shown in Table 1, are used as candidates of upstream regulators to each target gene. Finally, the kinetic parameters of gene transcriptional regulatory network of HeLa cell are estimated by the cDNA expression profiles of target genes and their upstream regulatory genes.

**Table 1. CloneIDs (Whitfield et al. 2002) of 64 transcription factors in human cell cycle are selected as candidates to regulate downstream target genes.**

Gene names are boldly denoted behind or under the CloneIDs in brackets. IEA denotes Inferred from Electronic Annotation in NCBI.

IMAGE:173309 <b>(BCLAF1)</b>	IMAGE:564803( <b>FOXM1</b> ) IEA	IMAGE:590774( <b>MAPK13</b> ) Knebel et al.	IMAGE:202704( <b>SCML1</b> ) IEA
IMAGE:241474( <b>BRCA1</b> ) IEA	IMAGE:149809( <b>GATA2</b> ) IEA	IMAGE:1536451 <b>(MHC2TA)</b>	IMAGE:782622 <b>(SP1)</b>
IMAGE:815287( <b>BRD8</b> ) Monden et al.	IMAGE:135688( <b>GATA2</b> ) IEA	IMAGE:809731( <b>MNT</b> ) Meroni et al.	IMAGE:80318 <b>(SP1)</b>
IMAGE:825210( <b>C14orf106</b> ) IEA	IMAGE:780958 <b>(GTF3C4)</b>	IMAGE:197520( <b>NCOA3</b> ) IEA	IMAGE:840636 <b>(SRF)</b>
IMAGE:726588( <b>C14orf106</b> ) IEA	IMAGE:291827 <b>(GTF3C4)</b>	IMAGE:884438( <b>NFE2L2</b> ) IEA	IMAGE:545503 <b>(STAT1)</b>
IMAGE:1915416( <b>CDK7</b> ) Shiekhattar et al.	IMAGE:344049( <b>HCFC1</b> ) Wysocka et al.	IMAGE:1455463 <b>(NFIC)</b>	IMAGE:134120 <b>(STAT5B)</b>
IMAGE:130242( <b>CDK7</b> ) Shiekhattar et al.	IMAGE:897806 <b>(HIF1A)</b>	IMAGE:265874 <b>(NFIC)</b>	IMAGE:712840 <b>(STAT5B)</b>
IMAGE:268652( <b>CDKN1A</b> ) IEA	IMAGE:878184( <b>HMG20B</b> ) IEA	IMAGE:271198( <b>NR3C1</b> ) Takahashi et al.	IMAGE:132857 <b>(STAT5B)</b>
IMAGE:240367( <b>CTCF</b> ) Filippova et al.	IMAGE:1842250( <b>HMGB2</b> ) Shirakawa et al.	IMAGE:245517( <b>NR5A2</b> ) IEA	IMAGE:272192 <b>(TCERG1)</b>
IMAGE:490728 <b>(DMTF1)</b>	IMAGE:363103( <b>HMGB2</b> ) Shirakawa et al.	IMAGE:43229( <b>PCNA</b> ) IEA	IMAGE:137387 <b>(TFAP2A)</b>
IMAGE:487797 <b>(DR1)</b>	IMAGE:242952( <b>ILF2</b> ) Reichman et al.	IMAGE:789182( <b>PCNA</b> ) IEA	IMAGE:868630( <b>TGFB114</b> ) IEA
IMAGE:566760 <b>(DR1)</b>	IMAGE:838829( <b>JARID1B</b> ) IEA	IMAGE:30114 <b>(PHTF2)</b>	IMAGE:366414( <b>UHRF1</b> ) IEA
IMAGE:884462( <b>DSCR1</b> ) Fuentes et al.	IMAGE:345056( <b>KIAA1404</b> ) IEA	IMAGE:1947972( <b>PKNOX1</b> ) Chen et al.	IMAGE:1550739( <b>UHRF1</b> ) IEA
IMAGE:236142 <b>(E2F1)</b>	IMAGE:510381( <b>KLF6</b> ) Ratziu et al.	IMAGE:2018976( <b>PTTG1</b> ) Dominguez et al.	IMAGE:246869( <b>ZNF207</b> ) Pahl et al.
IMAGE:768260 <b>(E2F1)</b>	IMAGE:302549( <b>KLF9</b> ) Imataka et al.	IMAGE:781089( <b>PTTG1</b> ) Dominguez et al.	IMAGE:296429( <b>ZNF24</b> ) IEA
IMAGE:293331 <b>(E2F2)</b>	IMAGE:35147( <b>LASS6</b> ) Banerjee-Basu et al.	IMAGE:845502( <b>RBPSUH</b> ) Hsieh et al.	IMAGE:280750( <b>ZNF281</b> ) Law et al.

## Dynamic signaling regulatory model

The second-order difference equation is used in the description of dynamic system evolved from the causality of gene regulatory function. Let  $X_i(k)$  denote the expression profile of the  $i$ -th gene at time point  $k$ . The following second-order difference equation is proposed to model the cDNA expression level of the  $i$ -th gene,

$$\begin{aligned} X_i(k) + a_i X_i(k-1) + b_i X_i(k-2) \\ = G_i(k) + \varepsilon_i(k) \end{aligned} \quad (1)$$

where  $G_i(k)$  is the upstream transcriptional regulatory function from regulatory genes to influence the expression profile  $X_i(k)$  of the  $i$ -th gene while  $a_i$  and  $b_i$  are the parameters that characterize the dynamic inherent property of the gene like degradation and oscillation, and  $\varepsilon_i(k)$  is the random noise of current microarray data or the residue of the model. In general, the second-order difference equation has been widely used to model dynamic discrete-time systems to efficiently characterize the dynamic properties of damping and resonance of systems in physics and engineering. The reason is that the roots of the characteristic equation of second-order dynamic equation may be a real double root, two real roots or conjugate roots, which could easily describe a system with undamping, overdamping, critical damping, underdamping or oscillation, dependent on the specification of their coefficients (Kreszig, 1993). These characteristics can not be easily described by the first-order dynamic equation. Therefore the second order stochastic equation is employed to characterize the biochemical processing of the gene expression.

Evidently, the transcriptional regulatory function  $G_i(k)$  controls the synthetic rate of cDNA and the clue of upstream regulatory pathway is involved in  $G_i(k)$ . Therefore we emphasize on how to detect the upstream regulatory function  $G_i(k)$  from expression data  $X_i(k)$  and our dynamic model equation in equation (1). In general, it is not easy to extract transcriptional regulatory function  $G_i(k)$ . In order to extract the input regulatory function  $G_i(k)$ , we apply Fourier decomposition method to decompose  $G_i(k)$  to generate some harmonic sinusoid functions. When the extraction problem is reduced to a simple parameter estimation problem,  $G_i(k)$  can be decomposed by the following Fourier series.

$$G_i(k) \approx \sum_{n=0}^N [\alpha_n \cos(nk) + \beta_n \sin(nk)] \quad (2)$$

Then we need to estimate the parameters of Fourier series,  $\alpha_n$  and  $\beta_n$ , that are the magnitudes of different harmonics of sinusoid functions ( $\cos(nt)$  and  $\sin(nt)$ ) for  $n = 0, \dots, N$ . Fourier series is a good tool to synthesize functions with finite energy by harmonic functions in respect of engineering.

## Extraction of the transcriptional regulatory function $G_i(k)$

Since  $G_i(k)$  has been decomposed, we combine equations (1) and (2) to get the following dynamic model equation for the expression profile of the  $i$ -th gene,

$$\begin{aligned} X_i(k) = -a_i X_i(k-1) - b_i X_i(k-2) \\ + \sum_{n=0}^N [\alpha_n \cos(nk) + \beta_n \sin(nk)] + \varepsilon_i(k) \end{aligned} \quad (3)$$

In the above dynamic model equation, the parameters  $a_i$ ,  $b_i$ ,  $\alpha_n$ , and  $\beta_n$  should be estimated by the time profile of expression data of the  $i$ -th gene in linear scale, i.e. these parameters should be specified so that the simulating output  $X_i(k)$  of the dynamic model in equation (3) should match the expression profile of the  $i$ -th gene. The maximum likelihood estimation method is employed to estimate these parameters  $a_i$ ,  $b_i$ ,  $\alpha_n$ , and  $\beta_n$  in equation (3) from the expression profile  $X_i(k)$  in the section Methods.

After the parameters  $\alpha_n$  and  $\beta_n$  of the regulatory function  $G_i(k)$  have been estimated in the section Methods, we can present the regulation detection  $\hat{G}_i(k)$  as follows,

$$\hat{G}_i(k) = \sum_{n=0}^N [\hat{\alpha}_n \cos(nk) + \hat{\beta}_n \sin(nk)] \quad (4)$$

where  $\hat{\alpha}_n$  and  $\hat{\beta}_n$  are the estimates of  $\alpha_n$  and  $\beta_n$ , respectively.

We know that the input transcriptional regulatory function  $G_i(k)$  of the target gene of Hela cell cycle is often relative to the bindings of transcription factors or some interactions from the upstream regulators. In the next step, we will trace back the corresponding regulatory genes from the input regulatory function  $\hat{G}_i(k)$  of the target gene.



## Iterative algorithm for constructing gene regulatory network

In biology, the specific biochemical reactions are usually relative to the concentration of specific products. For this purpose, we describe the regulatory function as the following sigmoid function to describe the binding and nonbinding of transcription factors to motif binding sites (Chen et al. 2004; Klipp et al. 2005)

$$\tilde{X}_j(k) = \frac{1}{1 + e^{-\gamma(X_j(k) - M_j)}} \quad (5)$$

where  $\gamma$  is the transition rate and  $M_j$  is the mean expression of the  $j$ -th regulatory gene's profile  $X_j(k)$ .

We determine  $R_i$  regulatory genes whose regulatory signals  $X_j(k)$   $j = 1, \dots, R$  are the most correlative to the target gene profile  $X_i(k)$  of the  $i$ -th target gene. Then, we could reconstruct the gene regulatory network by tracing back the upstream regulators from the extracted regulatory function  $\hat{G}_i(k)$ , which are contributed by  $R_i$  regulatory genes, via the following biochemical kinetic relationship,

$$\hat{G}_i(k) = c_{i0} + \sum_{j \in R_i} c_{ij} \tilde{X}_j(k) + e_i(k) \quad (6)$$

where  $c_{ij}$  is the pathway kinetic parameters from the regulatory gene  $j$  to target gene  $i$ ,  $R_i$  represents the number of the searched upstream regulatory genes selected by the absolute value of correlation coefficient between the target gene expression profile and the regulatory gene expression profile which is more than 0.8 based on the 95% confidence of normalized correlation coefficients of expression profiles of total cell cycle-related genes in Whitfield et al. 2002, the constant  $c_{i0}$  is the basal level denoting the regulatory function other than upstream regulatory genes, for example, due to post-transcriptional regulation, and  $e_i(k)$  is the error or the noise of the network model.

Using the maximum likelihood algorithm in Method to estimate the parameters  $c_{i0}$  and  $c_{ij}$  from  $\hat{G}_i(k)$  and  $X_i(k)$ , the regulation from the upstream regulators is identified as

$$\hat{G}_i(k) = \hat{c}_{i0} + \sum_{j \in R_i} \hat{c}_{ij} \tilde{X}_j(k) \quad (7)$$

By combining equation (1) and the above equation, the dynamics of transcriptional

regulatory network of the HeLa cell cycle can be represented by the following identified difference equation,

$$X_i(k) = -\hat{a}_i X_i(k-1) - \hat{b}_i X_i(k-2) + \hat{c}_{i0} + \sum_{j \in R_i} \hat{c}_{ij} \tilde{X}_j(k) \quad (8)$$

where  $i = 1, 2, \dots$  for all profile target genes in HeLa cell cycle.

In fact, equation (8) contains much information for exploring the regulatory network of each specific target gene of the HeLa cell cycle. The regulatory genes, which belong to a specific set,  $R_i$ , represent the potential upstream regulators for target gene  $i$ . The estimated chemical kinetic parameter,  $\hat{c}_{ij}$ , characterizes the type and intensity of the influence of the  $j$ th regulatory gene on the  $i$ th target gene, in which positive sign indicates activation and a negative sign indicates repression, and the magnitude is defined as the regulatory ability. After the regulatory pathways of the  $i$ th target gene is constructed by tracing back their upstream regulatory genes, these  $R_i$  upstream regulators are considered as target genes again to trace back their upstream regulatory genes. Iteratively, we can construct the whole gene regulatory network of the HeLa cell cycle globally. The goal of reverse engineering gene regulatory network is to deduce the possible set of regulators and to identify their associated regulation abilities by the available data set from the dynamic system perspective. For this purpose, we devise a novel algorithm based on the dynamic gene expression model for searching possible upstream regulators and then identifying the relevant regulatory abilities  $\hat{c}_{ij}$  according to equation (8).

## Results

### Data processing and analysis

Data were extracted by superimposing a grid over each array using GenePix 3.0 software (Axon Instruments). Spots of poor quality, determined by visual inspection, were removed from further analysis. Data of HeLa cell collected for each array were stored in the Stanford Microarray Database (SMD) and are available from SMD at <http://genome-www5.stanford.edu/> (Sherlock et al. 2001; Whitfield et al. 2002).

We combine 775 cell cycle-related genes from the Human expression of HeLa cell cycle-regulated genes according to the classification by Whitfield et al. (2002) with Human expression of cell cycle-regulated genes according to the traditional classification as the target genes. After that, we select 64 transcription factors (Table 1) from the 775 cell cycle genes (Whitfield et al. 2002).

The raw data were transformed into a linear scale from the original log ratio and applied to our approach. Following the dynamic model in equation (8), the parameters which characterize the dynamic regulatory mechanism are estimated successfully for each target gene in the pathway. Fig. 1 compares the simulation results of the dynamic expression model in equation (8) with the experimental expression profiles for some important cell cycle-related genes regulated by E2F family (Bracken et al. 2004), such as CDC25A (Stanelle et al. 2002; Muller and Helin, 2000; Ren et al. 2002), MCM6 (Ren et al. 2002; Polager et al. 2002; Ishida et al. 2001), E2F1 (Bracken, 2004), CDC6 (Stanelle et al. 2002; Ren et al. 2002), E2F2 (Ren et al. 2002; Muller et al. 2001), MCM5 (Ren et al. 2002; Ishida et al. 2001), MCM4 (Ren et al. 2002; Ishida et al. 2001), PCNA (Muller and Helin, 2000; Ren et al. 2002; Polager et al. 2002; Ishida et al. 2001), RFC4 (Ren et al. 2002; Polager et al. 2002), and DHFR (Ishida et al. 2001). The extracted regulatory functions,  $\hat{G}_i(k)$  of these genes which are estimated by the maximum likelihood algorithm in Methods from their expression profiles are shown in Fig. 2. The extracted regulatory function  $\hat{G}_i(k)$  in Fig. 2 are employed to estimate the kinetic parameters of gene regulatory network in equation (8) by the parameter estimation scheme in Methods. Our iterative algorithm can find the most likely regulatory genes that may participate in the expression program of HeLa cell cycle genes.

## Inference of the regulatory pathway

For illustrations, the inferring strategy is applied to the E2F target genes (Bracken et al. 2004) in HeLa cell cycle pathways to recognize their upstream regulatory genes. E2F transcription factors are well studied owing to their importance in both cell cycle (Muller and Helin, 2000; Nevins, 2001). Their regulatory abilities are shown in the upstream regulatory functions  $\hat{G}_i(k)$  of dynamic

equation in Table 2. Parameters of regulatory functions  $\hat{G}_i(k)$  in Table 2 represent the regulatory abilities and sensitivities of the relative transcription factors. It is very exciting that E2F1 and E2F2 are found to be active regulators in most E2F target genes listed in Table 2, which agree very well with the previous results (Cam and Dynlacht, 2003; Ivey-Hoyle et al. 1993; Lang et al. 2001). The regulatory abilities of the related regulators implying different degrees of influence are converted to the red-colored lines as positive regulations (activations) and the black-colored lines as negative regulations (inhibitions) for each target gene. Then, based on the dynamic regulatory equations in Table 2 (see detail in Supplementary Table S1), the pathways of E2F target gene in HeLa cell cycle regulatory system are described in Fig. 3. The coefficients of these dynamic regulatory equations represent the kinetic activities of regulatory genes. If a regulatory gene is with a large kinetic parameter in the dynamic regulatory equation, it will play an important role in HeLa cell cycle and is more sensitive to the gene expression of target gene.

Based on the dynamic regulatory modeling, the 152 E2F target genes are found and shown in Table 3. In these 152 E2F target genes, 6 genes match the E2F target genes found by Elkon et al. (2003) from 124 E2F target genes edited by Ren et al. (2002) and 17 genes match the E2F target genes found by Elkon et al. (2003) from 872 periodic genes edited by Whitfield et al. (2002). Ren et al. (2002) has found 124 E2F target genes with E2F binding promoter. The comparisons of these results are shown in Fig. 9.

After finishing the construction of the first layer network of E2F target genes (Bracken et al. 2004), we take these upstream regulators as target genes. By using similar method, we construct the upstream regulatory genes of these target genes. In the second layer network, the regulatory abilities implying different degrees of influence are converted into pink-colored lines as positive regulations (activations) and blue-colored lines as negative regulations (inhibitions). Then, we combine the first layer network and the second layer network together to form a more complete network to E2F target genes in HeLa cell cycle as described in Fig. 4. Iteratively, we can construct the higher layer network to complete the gene regulatory network of HeLa cell cycle.

**Table 2.** Upstream regulatory TFs and their regulatory function  $\hat{G}_i(k)$  on E2F target genes in cancer cell cycle. The positive sign implies activations while the negative sign implies inhibitions for each target gene. The magnitudes indicate their regulatory abilities to the downstream target genes.

$$(1) \hat{G}_{CDC25A}(k) = 12.159 \cdot \overline{X_{DMIT1}}(k) + 11.804 \cdot \overline{X_{E2F1}}(k) - 12.473 \cdot \overline{X_{BRCA1}}(k) - 4.7281 \cdot \overline{X_{PCNA}}(k) - 3.8787 \cdot \overline{X_{E2F2}}(k)$$

$$(2) \hat{G}_{MCM6}(k) = 4.2865 \cdot \overline{X_{PCNA}}(k) + 4.1629 \cdot \overline{X_{E2F1}}(k) + 3.5672 \cdot \overline{X_{MUT}}(k) + 3.5642 \cdot \overline{X_{DMIT1}}(k) + 2.4202 \cdot \overline{X_{E2F2}}(k) - 12.765 \cdot \overline{X_{SP1}}(k)$$

$$(3) \hat{G}_{E2F1}(k) = 19.97 \cdot \overline{X_{E2F2}}(k) + 3.898 \cdot \overline{X_{UNR1}}(k) - 7.0755 \cdot \overline{X_{PCNA}}(k)$$

$$(4) \hat{G}_{CDC6}(k) = 9.2779 \cdot \overline{X_{E2F2}}(k) + 8.1237 \cdot \overline{X_{E2F1}}(k) + 8.0354 \cdot \overline{X_{UNR1}}(k) - 9.5837 \cdot \overline{X_{PCNA}}(k)$$

$$(5) \hat{G}_{E2F2}(k) = 5.557 \cdot \overline{X_{SP1}}(k) + 4.3598 \cdot \overline{X_{BRCA1}}(k) + 3.6731 \cdot \overline{X_{E2F1}}(k) + 3.0949 \cdot \overline{X_{UNR1}}(k) - 12.012 \cdot \overline{X_{BRCA1}}(k) - 9.9406 \cdot \overline{X_{PCNA}}(k) - 2.4365 \cdot \overline{X_{MUT}}(k) - 1.9962 \cdot \overline{X_{DMIT1}}(k)$$

$$(6) \hat{G}_{MCM5}(k) = 6.7316 \cdot \overline{X_{E2F1}}(k) - 9.2368 \cdot \overline{X_{PCNA}}(k)$$

$$(7) \hat{G}_{MCM4}(k) = 9.8915 \cdot \overline{X_{E2F2}}(k) + 8.8902 \cdot \overline{X_{PCNA}}(k) + 6.3504 \cdot \overline{X_{BRCA1}}(k) + 4.7006 \cdot \overline{X_{MUT}}(k) + 2.8582 \cdot \overline{X_{HHEX2}}(k) - 13.12 \cdot \overline{X_{SP1}}(k) - 7.1211 \cdot \overline{X_{DMIT1}}(k) - 4.6503 \cdot \overline{X_{PCNA}}(k) - 4.4148 \cdot \overline{X_{LASS6}}(k)$$

$$(8) \hat{G}_{PCNA}(k) = 8.1178 \cdot \overline{X_{E2F2}}(k)$$

$$(9) \hat{G}_{RBL4}(k) = 18.628 \cdot \overline{X_{DMIT1}}(k) + 15.158 \cdot \overline{X_{PCNA}}(k) + 14.923 \cdot \overline{X_{E2F2}}(k) + 12.721 \cdot \overline{X_{BCLAF1}}(k) + 5.6902 \cdot \overline{X_{RBL2}}(k) - 16.276 \cdot \overline{X_{SP1}}(k) - 11.268 \cdot \overline{X_{LASS6}}(k) - 7.872 \cdot \overline{X_{E2F1}}(k) - 6.8745 \cdot \overline{X_{MUT}}(k) - 6.1689 \cdot \overline{X_{E2F2}}(k) - 2.7893 \cdot \overline{X_{BRCA1}}(k) - 2.4802 \cdot \overline{X_{HHEX2}}(k) - 2.2907 \cdot \overline{X_{PCNA}}(k)$$

$$(10) \hat{G}_{DHFR}(k) = 10.711 \cdot \overline{X_{MUT}}(k) + 6.9038 \cdot \overline{X_{PCNA}}(k) + 5.6045 \cdot \overline{X_{SP1}}(k) + 1.6341 \cdot \overline{X_{E2F1}}(k) - 13.422 \cdot \overline{X_{HHEX2}}(k) - 1.6894 \cdot \overline{X_{BCLAF1}}(k) - 1.5479 \cdot \overline{X_{BRCA1}}(k)$$

## Discussion

The losses of cellular regulation give rise to most cases of cancer. The transcription factors are crucial for regulating cell cycle progression and may be related to the development of a cancer. Therefore, to understand these gene regulatory processes, we need to unravel the regulatory mechanisms of these transcription factors in cell cycle. Our study presents a systematically iterative approach to discern and characterize the transcriptional regulatory network of 775 cell cycle-related genes from the raw expression profiles of Hela cell (Whitfield et al. 2002). Because the transcriptional regulatory network of 775 cell cycle-related genes is very complicated, two miniature gene regulatory networks of E2F family during G1/S

phase in Fig. 4 and the other family during G2/M in Fig. 8 are given to illustrate the regulatory mechanism of Hela cell.

Our approach also offers the following advantages. First, based on the dynamic regulatory model, a gene regulatory network of cancer cell could be constructed by the extracted upstream regulatory function through microarray data. Then, the identified regulatory ability for each specific regulator could evaluate the contribution of this regulator; the positive sign stands for activation and the negative sign stands for repression, and the magnitude represents the significance. These advantages of the proposed approach will improve the analysis to cope with rapidly growing microarray data of human. BRCA1 is one of the

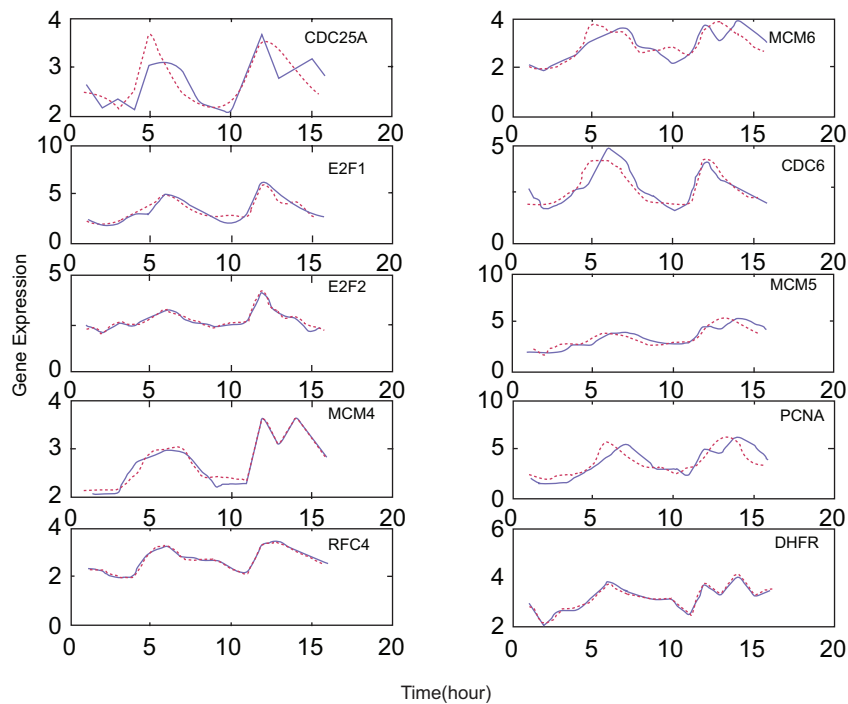
**Supplementary Table S1.** A miniature dynamic model network with the identified upstream regulators and their downstream target genes in the pathway of E2F target genes in cancer cell cycle. The coefficients characterize the corresponding regulatory abilities and sensitivities of the transcription regulations. The positive sign implies activations while the negative sign implies inhibitions for each target gene.

$$\begin{aligned}
 (1) \quad & X_{CDC25A}(k) = 0.5096 \cdot X_{CDC25A}(k-1) + 0.3936 \cdot X_{CDC25A}(k-2) - 1.2221 + 12.159 \cdot \widetilde{X_{DMIF1}}(k) + 11.804 \cdot \widetilde{X_{E2F1}}(k) \\
 & - 12.473 \cdot \widetilde{X_{BRCA1}}(k) - 4.7281 \cdot \widetilde{X_{PCNA}}(k) - 3.8787 \cdot \widetilde{X_{E2F2}}(k) \\
 (2) \quad & X_{MCM6}(k) = 0.5223 \cdot X_{MCM6}(k-1) + 0.4037 \cdot X_{MCM6}(k-2) - 4.6131 + 4.2865 \cdot \widetilde{X_{PCNA}}(k) + 4.1629 \cdot \widetilde{X_{E2F1}}(k) + 3.5672 \cdot \widetilde{X_{MNT}}(k) + 3.5642 \cdot \widetilde{X_{DMIF1}}(k) + 2.4202 \cdot \widetilde{X_{E2F2}}(k) \\
 & - 12.765 \cdot \widetilde{X_{SP1}}(k) \\
 (3) \quad & X_{E2F1}(k) = 0.5273 \cdot X_{E2F1}(k-1) + 0.3706 \cdot X_{E2F1}(k-2) - 7.5784 + 19.97 \cdot \widetilde{X_{E2F2}}(k) + 3.898 \cdot \widetilde{X_{UHRF1}}(k) \\
 & - 7.0755 \cdot \widetilde{X_{PCNA}}(k) \\
 (4) \quad & X_{CDC6}(k) = 0.5802 \cdot X_{CDC6}(k-1) + 0.2863 \cdot X_{CDC6}(k-2) - 7.5759 + 9.2779 \cdot \widetilde{X_{E2F2}}(k) + 8.1237 \cdot \widetilde{X_{E2F1}}(k) + 8.0354 \cdot \widetilde{X_{UHRF1}}(k) \\
 & - 9.5837 \cdot \widetilde{X_{PCNA}}(k) \\
 (5) \quad & X_{E2F2}(k) = 0.4709 \cdot X_{E2F2}(k-1) + 0.4298 \cdot X_{E2F2}(k-2) + 3.077 + 5.557 \cdot \widetilde{X_{SP1}}(k) + 4.3598 \cdot \widetilde{X_{NR5A2}}(k) + 3.6731 \cdot \widetilde{X_{E2F1}}(k) + 3.0949 \cdot \widetilde{X_{UHRF1}}(k) \\
 & - 12.012 \cdot \widetilde{X_{BRCA1}}(k) - 9.9406 \cdot \widetilde{X_{PCNA}}(k) - 2.4365 \cdot \widetilde{X_{MNT}}(k) - 1.9962 \cdot \widetilde{X_{DMIF1}}(k) \\
 (6) \quad & X_{MCM5}(k) = 0.5946 \cdot X_{MCM5}(k-1) + 0.3576 \cdot X_{MCM5}(k-2) - 1.1563 + 6.7316 \cdot \widetilde{X_{E2F1}}(k) \\
 & - 9.2368 \cdot \widetilde{X_{PCNA}}(k) \\
 (7) \quad & X_{MCM4}(k) = 0.4833 \cdot X_{MCM4}(k-1) + 0.4359 \cdot X_{MCM4}(k-2) + 0.1858 + 9.8915 \cdot \widetilde{X_{E2F2}}(k) + 8.8902 \cdot \widetilde{X_{PCNA}}(k) + 6.3504 \cdot \widetilde{X_{BRCA1}}(k) + 4.7006 \cdot \widetilde{X_{MNT}}(k) + 2.8582 \cdot \widetilde{X_{NFE2L2}}(k) \\
 & - 13.12 \cdot \widetilde{X_{SP1}}(k) - 7.1211 \cdot \widetilde{X_{DMIF1}}(k) - 4.6503 \cdot \widetilde{X_{PCNA}}(k) - 4.4148 \cdot \widetilde{X_{LASS6}}(k) \\
 (8) \quad & X_{PCNA}(k) = 0.7077 \cdot X_{PCNA}(k-1) - 0.2043 \cdot X_{PCNA}(k-2) - 3.3649 + 8.1178 \cdot \widetilde{X_{E2F2}}(k) \\
 (9) \quad & X_{RFC4}(k) = 0.5079 \cdot X_{RFC4}(k-1) + 0.404 \cdot X_{RFC4}(k-2) - 6.843 + 18.628 \cdot \widetilde{X_{DMIF1}}(k) + 15.158 \cdot \widetilde{X_{PCNA}}(k) + 14.923 \cdot \widetilde{X_{E2F2}}(k) + 12.721 \cdot \widetilde{X_{BCLAF1}}(k) + 5.6902 \cdot \widetilde{X_{HCF1}}(k) \\
 & - 16.276 \cdot \widetilde{X_{SP1}}(k) - 11.268 \cdot \widetilde{X_{LASS6}}(k) - 7.872 \cdot \widetilde{X_{E2F1}}(k) - 6.8745 \cdot \widetilde{X_{KLF9}}(k) - 6.1689 \cdot \widetilde{X_{STAT1}}(k) - 2.7893 \cdot \widetilde{X_{BRCA1}}(k) - 2.4802 \cdot \widetilde{X_{NFE2L2}}(k) - 2.2907 \cdot \widetilde{X_{PCNA}}(k) \\
 (10) \quad & X_{HIRF}(k) = 0.5226 \cdot X_{RFC4}(k-1) + 0.3977 \cdot X_{RFC4}(k-2) - 4.4793 + 10.711 \cdot \widetilde{X_{MNT}}(k) + 6.9038 \cdot \widetilde{X_{PCNA}}(k) + 5.6045 \cdot \widetilde{X_{SRF}}(k) + 1.6341 \cdot \widetilde{X_{E2F1}}(k) \\
 & - 13.422 \cdot \widetilde{X_{NFE2L2}}(k) - 1.6894 \cdot \widetilde{X_{BCLAF1}}(k) - 1.5479 \cdot \widetilde{X_{BRCA1}}(k)
 \end{aligned}$$

important cell cycle-related genes to play as a transcriptional repressor in cell cycle progress (Kennedy et al. 2005). This finding matches the gene regulatory network constructed by dynamic regulatory model (shown in Fig. 4 and Fig. 5). It is clear that E2F regulates the expression of a host of factors that function during G1/S transition and S phase even the whole cell cycle (Bracken et al. 2004). E2F is best known for its role in regulating the transcription of genes that positively affect cell cycle progression (Ortega et al. 2002; Sherr and Roberts, 1999; Di et al. 2003; Stott et al. 1998; Trimarchi and Lees, 2002; Hayashi et al. 2006).

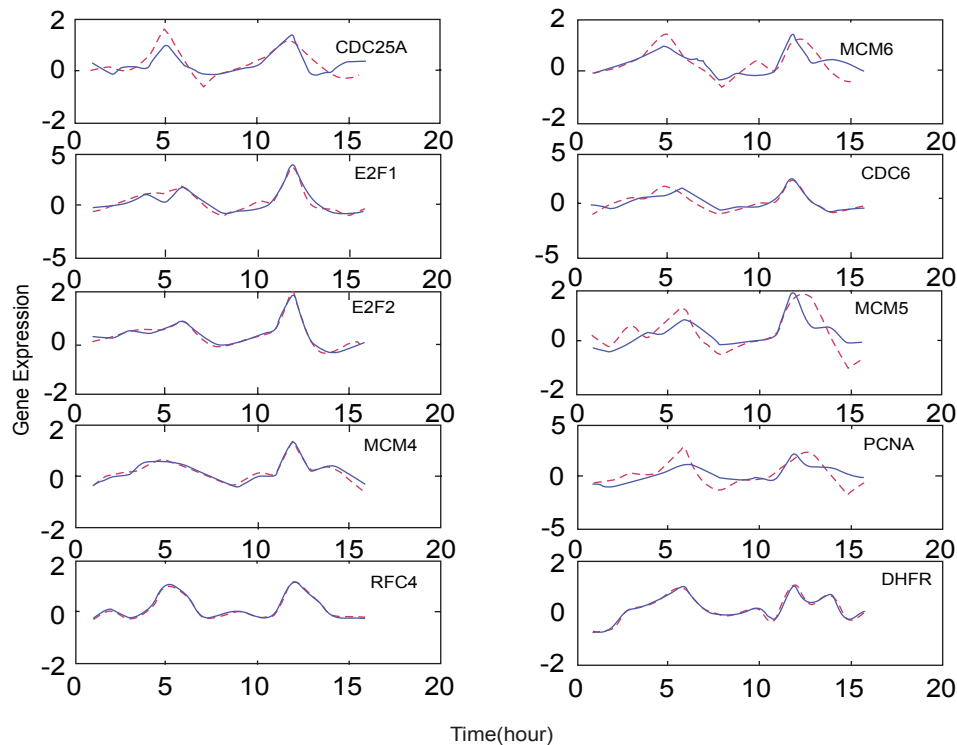
Our results almost match this finding (shown in Fig. 4 and Fig. 5). Hayashi with his collaborator found that E2F1 activates human MCM8. In our study, we found that E2F1 activates human MCM5 and MCM6 (shown in Fig. 4 and Fig. 5). Elkon et al. (2003) and Ren et al. (2002) have found that human MCM5 and MCM6 are two E2F target genes (shown in Table 3). As a result of these studies, we infer that human MCM5 and MCM6 may be positively regulated directly by E2F1 in Hela cell cycle. Accumulating evidence indicates that cdc25A possesses oncogenic properties. Recently, overexpression of cdc25A was





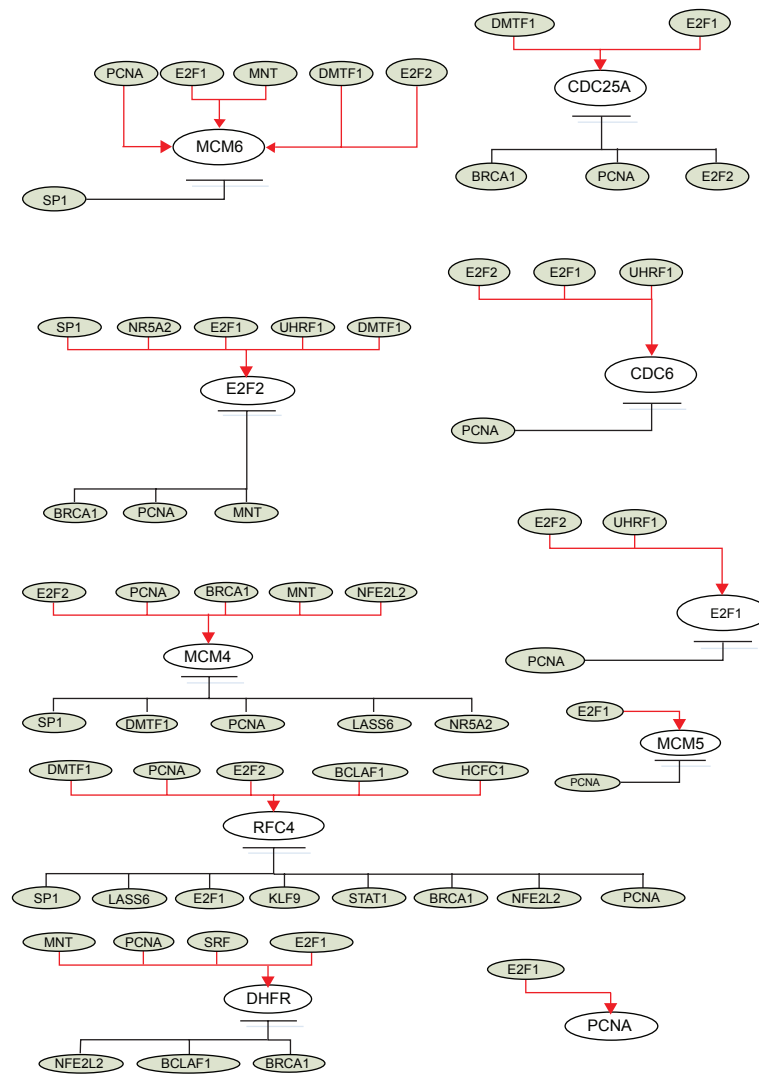
**Figure 1. The second-order dynamic model fitting of E2F target genes.**

The expression profile with the corresponding second-order dynamic model fitting for E2F target genes, CDC25A, MCM6, E2F1, CDC6, E2F2, MCM5, MCM4, PCNA, RFC4, and DHFR. The red dashed lines are the microarray data from Whitfield et al. 2002, and the blue solid lines are the estimated dynamic evolution of expression data by the proposed method. Obviously, the proposed second-order dynamic model could efficiently characterize the dynamic properties like damping and resonance of the gene regulatory network. The data have been transformed from the log scale in Whitfield et al. to linear scale.



**Figure 2. The extracted upstream regulatory functions from expression profiles (Whitfield et al. 2002) and their fitting by upstream regulatory genes.**

The upstream transcriptional regulatory functions  $\hat{G}(t)$  extracted from expression profiles of corresponding E2F target genes are denoted by the red dashed lines, and the blue solid lines are fitted by their upstream transcriptional regulatory functions in equation (6). These results show that the upstream TF regulatory model in equation (6) could match the regulatory function extracted from microarray data. In the figure, the gene expressions are expressed in linear scale.

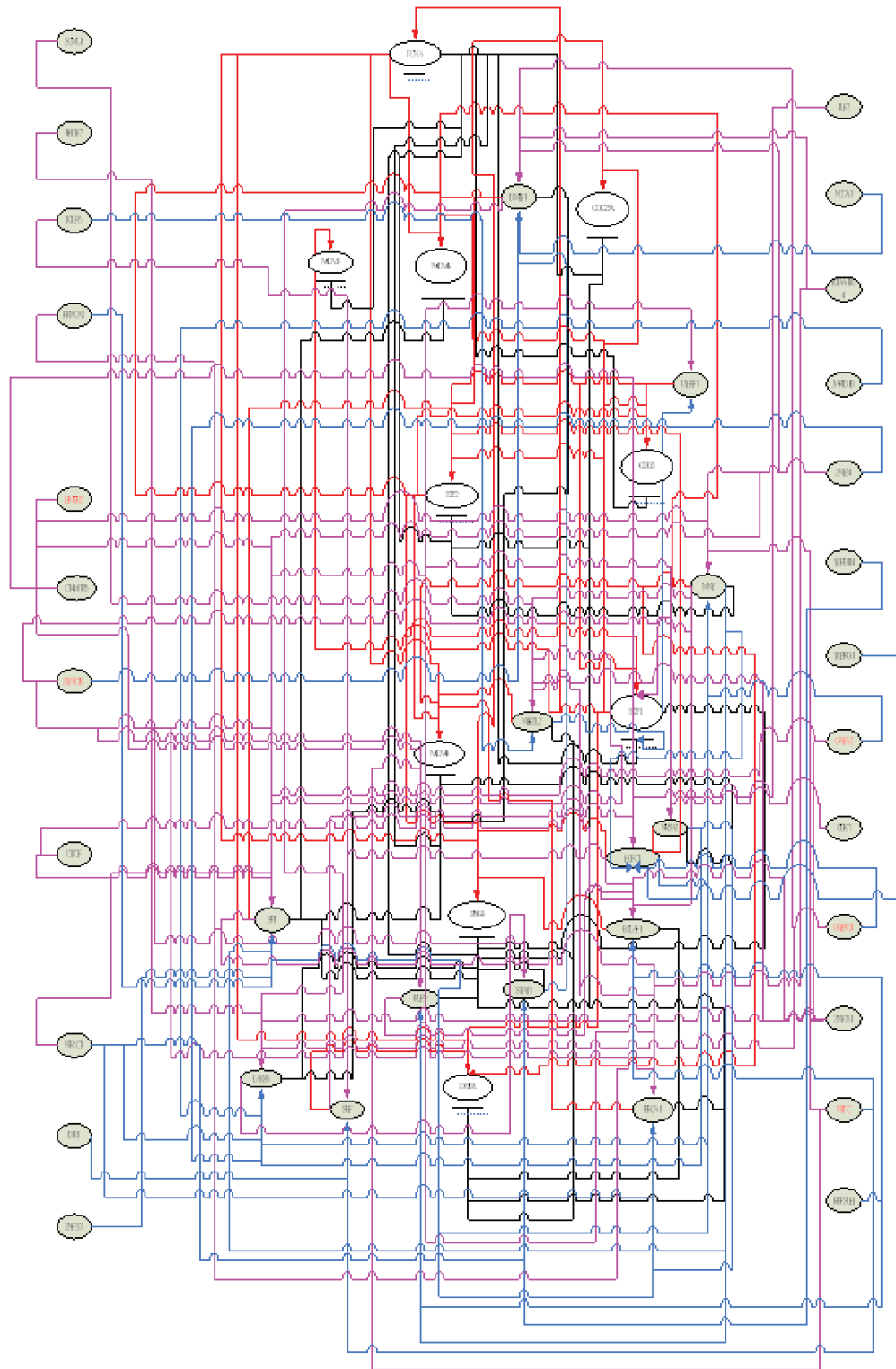


**Figure 3. The regulatory pathways of E2F target genes in cancer cell cycle based on the dynamic regulatory modeling in Table 1.** The related genes are represented as gray ovals and the regulatory abilities of the related regulators implying different degree of influence are converted into red-colored lines as positive regulations (activations) and black-colored lines as negative regulations (inhibitions) for each target gene. These regulatory pathways could be integrated as the regulatory network of cancer cell cycle in Figure 4.

found in many breast, head and neck cancers (Wu et al. 1998). CDC6 plays a critical role in the regulation of the onset of DNA replication in eukaryotic cells and Cdc6 expression is down-regulated in prostate cancer (Robles et al. 2002). From the results of Table 3, we could infer that E2F directly regulates MCM4, MCM5, CDC6, CDC25A, UNG and E2F2 in HeLa cell cycle progression. In Fig. 5, we represent some E2F target genes and show the importance of E2F transcription factor in HeLa cell cycle progression. In Fig. 7, we also predict the probable transcription regulations in some target genes, which express in G2/M phase of human cell cycle as shown in the regulatory network of Fig. 8. Further, we can construct not only E2F related

regulatory network but also the whole HeLa cell cycle network if we have genome-wide microarray data and CHIP data. However, at present, we still need to get enough evidence and CHIP experiments in the same cellular system to confirm our gene regulatory network of HeLa cell cycle (Bracken et al. 2004).

Based on the dynamic regulatory modeling, the 152 E2F target genes are found and shown in Table 3. These target genes may be regulated by E2F directly or indirectly. In these 152 E2F target genes, 6 genes match the E2F target genes found by Elkouf et al. (2003) from 124 E2F target genes with E2F promoter edited by Ren et al. (2002) and 17 genes match the E2F target genes found by



**Figure 4. The network of E2F target genes in cancer cell cycle based on the dynamic regulatory modeling.**

The regulatory pathway of the  $i$ th target gene is constructed by tracing back its upstream regulatory genes. Then, these upstream regulatory genes are considered as target genes again to trace back their upstream regulatory upstream regulatory genes. Iteratively, we can construct the whole gene regulatory network of cancer cell cycle globally. In the first layer of the upstream regulators, red solid lines represent positive regulation and black solid lines represent negative regulation, and in the second layer of the upstream regulators, pink solid lines represent positive regulation and blue solid lines represent negative regulation. Based on this procedure, this E2F regulatory network is the interaction of regulatory pathways in Fig. 3.

**The enlarged version of Figure 4 and the more clear figures of this paper are provided in <http://www.ee.nthu.edu.tw/bschen/Li.htm>.**

**Table 3. The 152 E2F target genes in the gene regulatory network based on the dynamic regulatory modeling.**

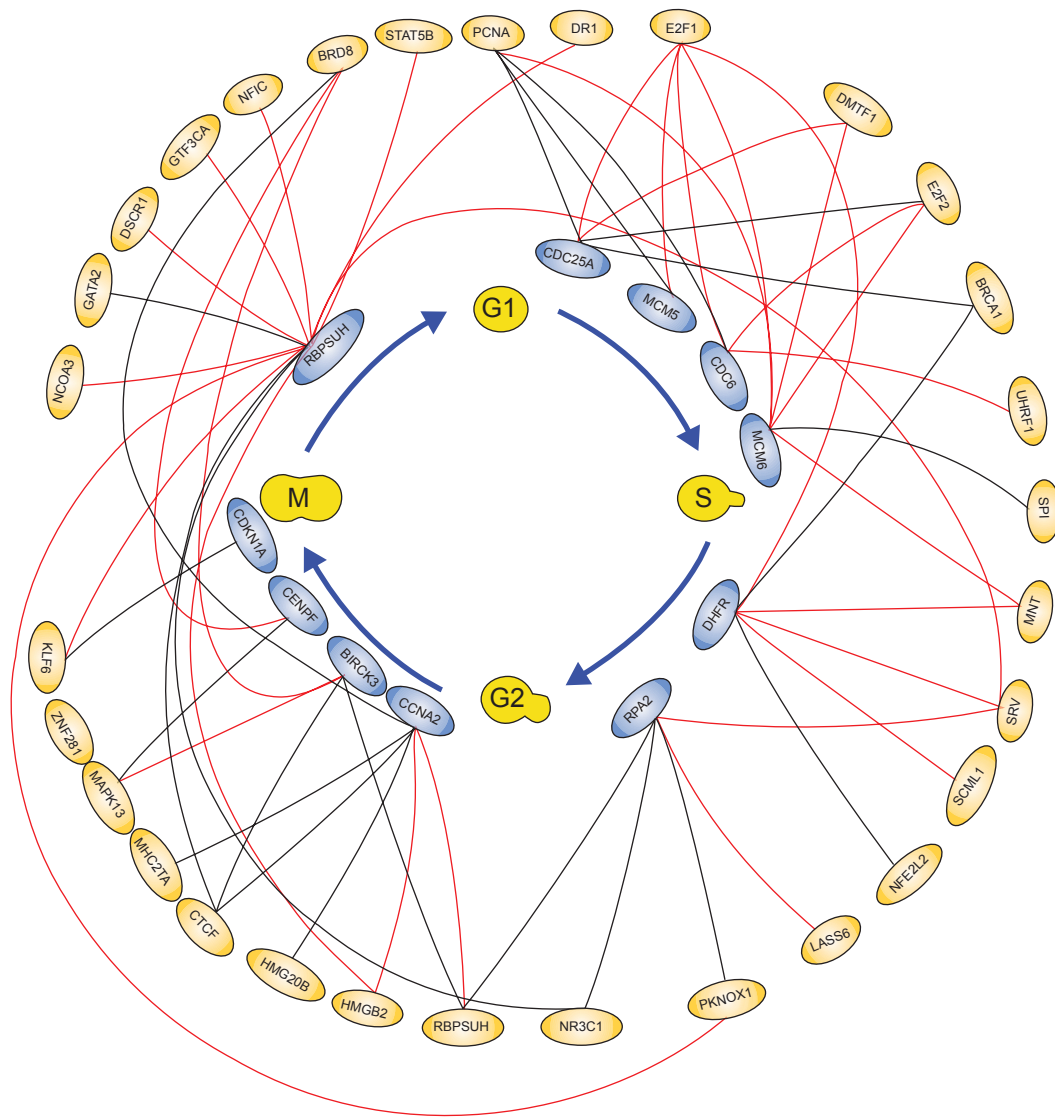
The marked signs are the matched E2F target genes found by Elkon et al. (2003).

	TOP3A	ACYP1	BIVM	ADCK2	OGT	C20orf55
BAIAP2	MED31	CDC6**	FLJ10154	FANCG	PILRB	MET
SSR3	NASP	ADCY6	MGC15716	APEX2	FLJ11021	NR5A2
SLBP#	CDC25A**	ANKRD10	ABCA7	LOC197336	ASF1B	CDKN2D
MNT	UNG**	BARD1	FLJ13912	RBBP8	SFRS5#	CENPE
LOC221955	MGC3207	E2F2*	Pfs2	MCM8	C22orf18	USP16
LOC90110	UBQLN2	MCM5**	EGFL5	C20orf111	LOC58486	ARHGAP8
FBXL20	RAMP	CCRK	NKTR	ABCC5	MAP3K2	PSEN1
ANKRD25	MCM6#	TREX1	KIAA0092	MLF1IP	EZH2	C6
KIAA1529	DMTF1#	RAB23#	LOC200895	FANCA	CENTB5	USP6NL
PTOP	KIAA0738	FLJ10618	HSPB8	DNAJC6	HIST3H2A	LOC51334
TRIM26	PNN	OACT1	C14orf130	RAD18	BRCA1	AKAP13
ZMYND19	MAP2K6	ORC1L#	KIAA1586	SP1	FLJ35740	AGPAT3
CDC47	GMNIN	MCM4#	HELLS	USP1	HSPC150	ASAM
CASP8AP2#	LOC283596	MSH2	CLSPN	MYCBP2	ORC3L	RAB3A
PLCXD1	CASP2	LCHN	RFC4	TOPBP1#	DNAJB4	HIST2H2BE
ZNF367	C9orf42	FLJ20280	DONSON	DHFR	RRM1	HIST2H2AA
RNP	E2F1#	PCNA	AP4B1	DKFZP434B168#	RHOBTB3	H2AFY
USP53	CCNE2#	MBD4	MGC2610	DLNB14	KIAA0841	H3F3A
SKP2	SCML1	TLOC1	ATAD2	RFC2	NFE2L2	HIST1H3D
PDXP	SERPINB3	SDC1	RPA2#	NSUN3	BBS2	
FLJ20530	UHRF1	PRPS2	FEN1	FKBP6	FANCD2	
PANK2						

\*E2F target genes found by Elkon et al. (2003) from 124 E2F target genes edited by Ren et al. (2002).

#E2F target genes found by Elkon et al. (2003) from 872 periodic genes edited by Whitfield et al. (2002).





**Figure 5. The miniature cancer cell cycle network.**

In this figure we just list some E2F target genes and show the role of E2F transcription factors in cancer cell cycle (Muller and Helin 2000). Red solid lines represent positive regulations and black solid lines represent negative regulations.

Elkon et al. (2003) from 872 periodic genes in HeLa cell cycle edited by Whitfield et al. (2002).

Finally, in order to validate the proposed approach, an independent validation is also given by randomly reshuffling the time order of microarray experiment but with the same choices of target gene and regulatory genes, to confirm the reliability of the proposed method as shown in Fig.6. As previous statement, BRCA1 plays as a transcriptional repressor in cell cycle progress (Kennedy et al. 2005) (shown in Fig. 5). From the shuffling results shown in Fig. 6, BRCA1 becomes a transcriptional activator. It is clearly seen that the proposed HeLa cell cycle regulatory pathway

in Fig. 5 is destroyed by reshuffling the experimental data.

## Methods

### Maximum likelihood Estimation of $a_i$ , $b_i$ , $\alpha_n$ and $\beta_n$

The dynamic equation (3) must match the expression profile at all time points and then is arranged in a vector difference form. Consequently, the vector dynamic form of this equation is applied to  $m$  time points of expression profile in order to make the dynamic model work.

$$X_{ik} = -a_i X_{ik-1} - b_i X_{ik-2} + \sum_{n=0}^N [\alpha_n \cos(n) + \beta_n \sin(n)] + \varepsilon_{ik} \quad (M.1)$$

$$X_{im} = \begin{bmatrix} X_i(3) \\ X_i(4) \\ \vdots \\ X_i(m) \end{bmatrix}, \quad X_{im-1} = \begin{bmatrix} X_i(2) \\ X_i(3) \\ \vdots \\ X_i(m-1) \end{bmatrix},$$

$$X_{im-2} = \begin{bmatrix} X_i(1) \\ X_i(2) \\ \vdots \\ X_i(m-2) \end{bmatrix},$$

where

$$\cos(n) = \begin{bmatrix} \cos(n3) \\ \cos(n4) \\ \vdots \\ \cos(nm) \end{bmatrix}, \quad \sin(n) = \begin{bmatrix} \sin(n3) \\ \sin(n4) \\ \vdots \\ \sin(nm) \end{bmatrix},$$

$$\varepsilon_{im} = \begin{bmatrix} \varepsilon_i(3) \\ \varepsilon_i(4) \\ \vdots \\ \varepsilon_i(m) \end{bmatrix}.$$

Next, we translate equation (M.1) into a matrix form,

$$Y_i = A_i \Phi_i + E_i \quad (M.2)$$

where  $\Phi_i = [a_i \ b_i \ \alpha_0 \ \beta_0 \ \dots \ \alpha_N \ \beta_N]^T$  and  $E_i = \varepsilon_{im}$  are in vector forms, and  $A_i = [-X_{im-1} \ -X_{im-2} \ \cos(0) \ \sin(0) \ \dots \ \cos(N) \ \sin(N)]^T$  is a matrix.

Then we use the maximum likelihood estimation to derive the optimal parameters estimation of  $\hat{\Phi}_i$  (Johansson, 1993).

We assume that each element in the error vectors,  $\varepsilon_i(k), k = \{3, \dots, m\}$ , is an independent random variable with a normal distribution with zero mean and variance  $\sigma^2$ , and we will estimate the parameter  $\hat{\Phi}_i$ , by the maximum likelihood method.

$$p(Y_i | \Phi_i, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{[Y_i - A_i \Phi_i]^T [Y_i - A_i \Phi_i]}{2\sigma^2} \right\} \quad (M.3)$$

The log-likelihood function for given  $m$  data points is then described by-

$$\log L(\Phi_i, \sigma^2) = -\frac{m-2}{2} \ln[2\pi\sigma^2] - \frac{1}{2\sigma^2} \sum_{k=1}^m [Y_i - A_i \Phi_i]^T [Y_i - A_i \Phi_i] \quad (M.4)$$

Here we expect the log-likelihood function to have the maximum at  $\Phi = \hat{\Phi}$  and  $\sigma^2 = \hat{\sigma}^2$ . The necessary condition for maximum likelihood estimates  $\hat{\Phi}$  and  $\hat{\sigma}^2$  as follows (Johansson, 1993).

$$\left. \frac{\partial \log L(\Phi, \sigma^2)}{\partial \Phi} \right|_{\Phi=\hat{\Phi}} = 0$$

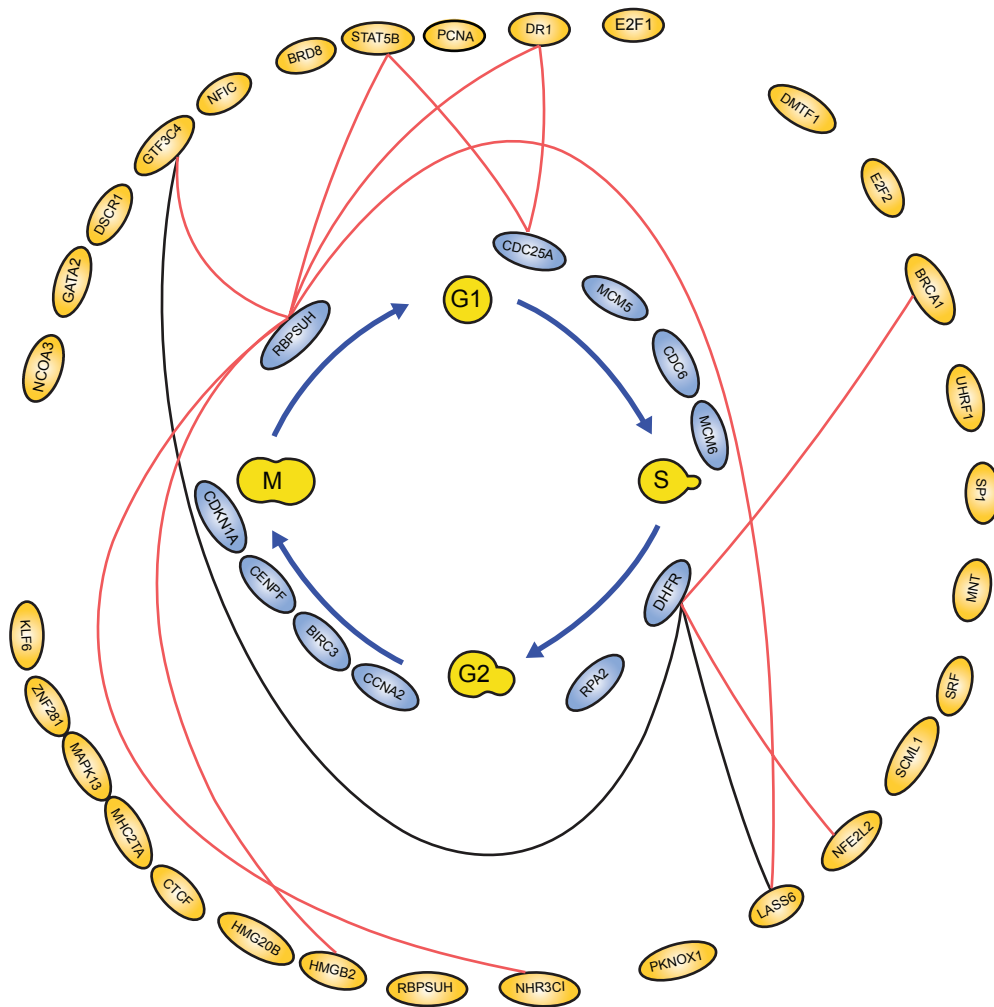
$$\left. \frac{\partial \log L(\Phi, \sigma^2)}{\partial \sigma^2} \right|_{\sigma^2=\hat{\sigma}^2} = 0 \quad (M.5)$$

The estimated parameters  $\hat{\Phi}$  and  $\hat{\sigma}^2$  are shown below.

$$\hat{\Phi}_i = (A_i^T A_i)^{-1} A_i^T Y_i \quad (M.6)$$

$$\hat{\sigma}^2 = \frac{1}{m-2} (Y_i - A_i \hat{\Phi}_i)^T (Y_i - A_i \hat{\Phi}_i) \quad (M.7)$$

Theoretically,  $E_i$  is just the noise of the gene expression profile of the microarray chips, but some modeling errors and approximation errors in equation (2) are also involved in  $E_i$ . So that taking the modeling error and approximation error in our consideration makes our dynamic model equation more approach the actual situation. The number of Fourier series  $N$  is determined by tradeoff between the computational complexity of parameter estimation in equation (M.6) and the accuracy of approximation in equation (2). According to our expression data, we choose  $N = 16$  to make the synthesis of these harmonics be the best approximation to the expression data.



**Figure 6. The miniature cancer cell cycle network in Fig.5 is repeated as independent validation by randomly reshuffling the time order of microarray experiment but with the same choices of target and regulatory genes.**

Obviously, the proposed regulatory network in Fig.5 is destroyed by the reshuffling of experimental data. Red solid lines represent positive regulations and black solid lines represent negative regulations. For example, BRCA1 in Fig. 5 plays as a transcriptional repressor in cell cycle progress (Kennedy et al. 2005). From Fig. 6, BRCA1 becomes a transcriptional activator.

### Parameter Estimation of $c_{i0}$ and $c_{ij}$

To estimate the pathway kinetic parameters  $c_{ij}$  in equation (6) by  $\hat{G}_i(k)$  and  $\tilde{X}_j(k)$  with  $m$  time points of upstream regulatory expression profile, equation (6) is represented with algebraic form as follows,

$$\hat{G}_i = B_i \Omega_i + V_i \tag{M.8}$$

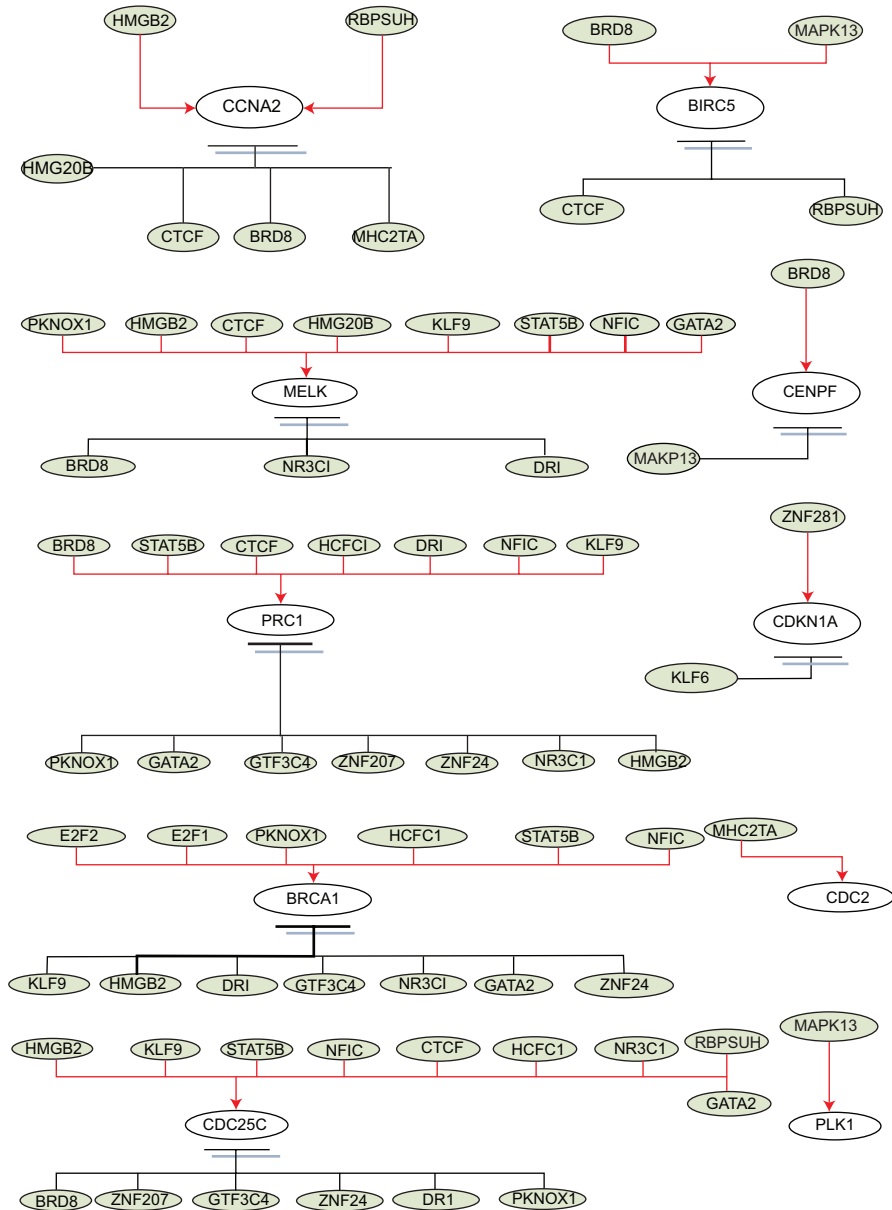
$$\hat{G}_i = \begin{bmatrix} \hat{G}_i(3) \\ \hat{G}_i(4) \\ \vdots \\ \hat{G}_i(m) \end{bmatrix} B_i = \begin{bmatrix} 1 & \tilde{X}_1(3) & \dots & \dots & \tilde{X}_{R_i}(3) \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ 1 & \tilde{X}_1(m) & \dots & \dots & \tilde{X}_{R_i}(m) \end{bmatrix}$$

where

$$\Omega_i = \begin{bmatrix} c_{i0} \\ c_{i1} \\ \vdots \\ c_{iR_i} \end{bmatrix}, \text{ and } V_i = \begin{bmatrix} e_i(3) \\ e_i(4) \\ \vdots \\ e_i(m) \end{bmatrix}$$

We assume that each element in the error vectors,  $e_i(k)$ ,  $k = \{3, \dots, m\}$ , is an independent random variable with a normal distribution with zero mean and variance  $\sigma_{e_i}^2$ . Then by the similar procedure, the estimated parameters  $\hat{\Omega}$  and  $\sigma_{e_i}^2$  are shown below.

$$\hat{\Omega} = (B_i^T B_i)^{-1} B_i^T \hat{G}_i \tag{M.9}$$



**Figure 7. The regulatory pathways of target genes, which expressed in G2/M phase of human cell cycle, are based on the dynamic regulatory modeling.**

The related genes are represented as gray ovals and the regulatory abilities of the related regulators implying different degree of influence are converted into red-colored lines as positive regulations (activations) and black-colored lines as negative regulations (inhibitions) for each target gene.



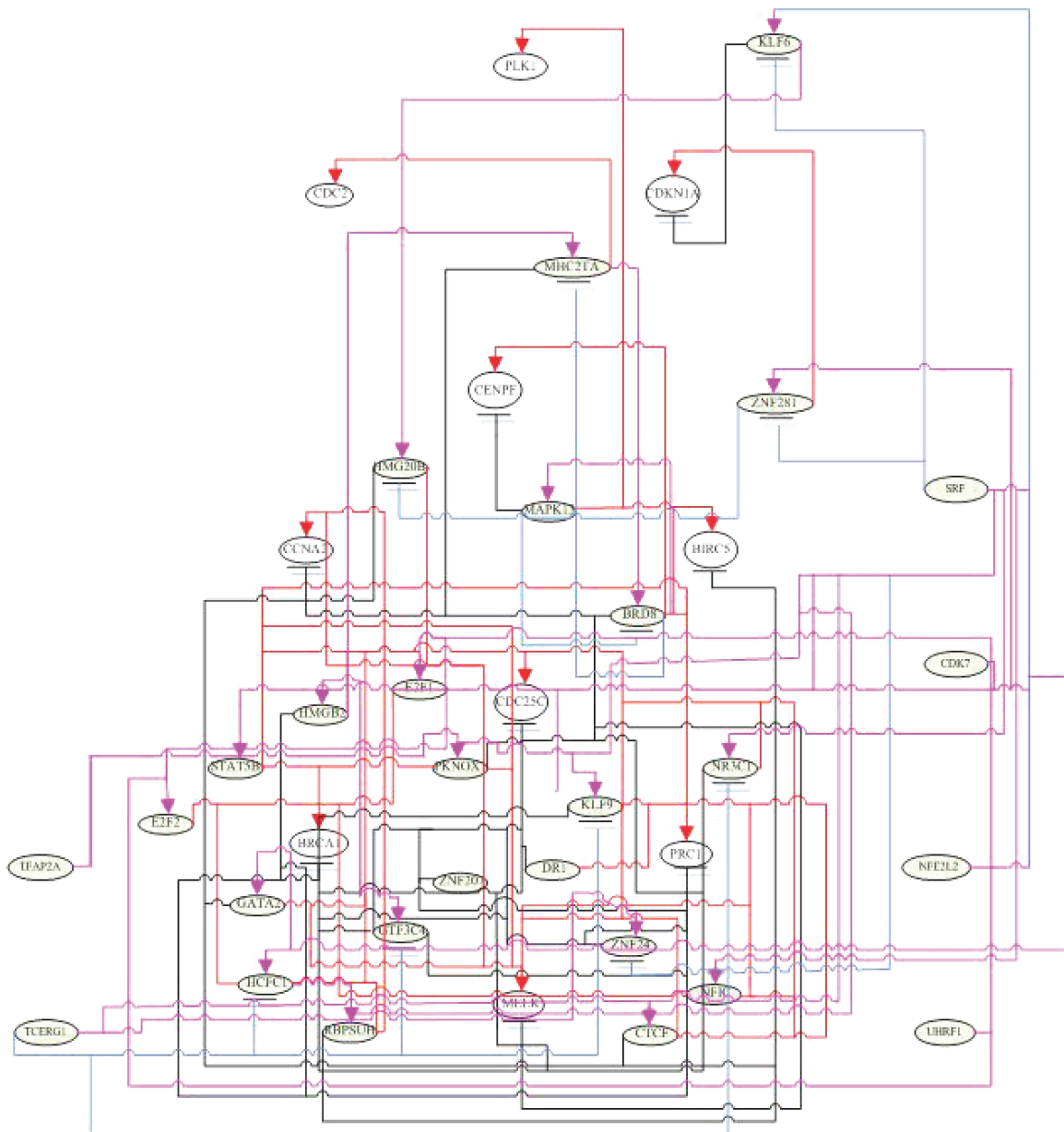


Figure 8. The gene regulatory network of in G2/M phase of cancer cell cycle based on the dynamic regulatory modeling and the interactions of regulatory pathways in Fig. 7.

$$\sigma_{e_i}^2 = \frac{1}{m-2} (\hat{G}_i - B_i \Omega_i)^T (\hat{G}_i - B_i \Omega_i) \quad (\text{M.10})$$

## Acknowledgements

We thank the National Science Council, Taiwan for grants NSC 95-2627-B-007-011.

## References

- Badawi, R.A., Birns, J. and Watson, T. et al. 2005. Growth factors and their relationship to neoplastic and paraneoplastic disease. *Eur. J. Intern. Med.*, 16:83–94.
- Bracken, A.P., Ciro, M. and Cocito, A. et al. 2004. E2F target genes: unravelling the biology. *Trends Biochem. Sci.*, 29:409–17.
- Banerjee-Basu, S. and Baxevasis, A.D. 2001. Molecular evolution of the homeodomain family of transcription factors. *Nucleic Acids Res.*, 29(15):3258–69.
- Cam, H. and Dynlacht, B.D. 2003. Emerging roles for E2F: beyond the G1/S transition and DNA replication. *Cancer Cell*, 3:311–6.
- Chen, H.C., Lee, H.C. and Lin, T.Y. et al. 2004. Quantitative characterization of the transcriptional regulatory network in the yeast cell cycle. *Bioinformatics*, 20:1914–27.
- Chen, H., Rossier, C. and Nakamura, Y. et al. 1997. Cloning of a novel homeobox-containing gene, PKNOX1, and mapping to human chromosome 21q22.3. *Genomics*, 41(2):193–200.
- Cho, R.J., Campbell, M.J. and Winzler, E.A. et al. 1998. A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell.*, 2:65–73.
- Cho, R.J., Huang, M. and Campbell, M.J. et al. 2001. Transcriptional regulation and function during the human cell cycle. *Nat. Genet.*, 27:48–54.
- Di Stefano, L., Jensen, M.R. and Helin, K. 2003. E2F7, a novel E2F featuring DP-independent repression of a subset of E2F-regulated genes. *EMBO J.*, 22:6289–98.
- Dominguez, A., Ramos-Morales, F. and Romero, F. et al. 1998. hpttg, a human homologue of rat pttg, is overexpressed in hematopoietic neoplasms. Evidence for a transcriptional activation function of hPTTG. *Oncogene*, 17(17):2187–93.
- Filippova, G.N., Fagerlie, S. and Klenova, E.M. et al. 1996. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol. Cell. Biol.*, 16(6):2802–13.
- Fuentes, J.J., Pritchard, M.A. and Planas, A.M. et al. 1995. A new human gene from the Down syndrome critical region encodes a proline-rich protein highly expressed in fetal brain and heart. *Hum. Mol. Genet.*, 4(10):1935–44.
- Hayashi, R., Goto, Y. and Haga, A. et al. 2006. Comparative genomics on MCM8 orthologous genes reveals the transcriptional regulation by transcription factor E2F. *Gene*, 367:126–34.
- Hinchcliffe, E.H. and Sluder, G. 2001. “It takes two to tango”: understanding how centrosome duplication is regulated throughout the cell cycle. *Genes Dev.*, 15:1167–81.
- Hsieh, J.J., Zhou, S., Chen, L., Young, D.B. and Hayward, S.D. 1999. CIR, a corepressor linking the DNA binding factor CBF1 to the histone deacetylase complex. *Proc. Natl. Acad. Sci. U.S.A.*, 96(1):23–8.0
- Imataka, H., Sogawa, K. and Yasumoto, K. et al. 1992. Two regulatory proteins that bind to the basic transcription element (BTE), a GC box sequence in the promoter region of the rat P-4501A1 gene. *EMBO J.*, 11(10):3663–71.
- Ishida, S., Huang, E. and Zuzan, H. et al. 2001. Role for E2F in control of both DNA replication and mitotic functions as revealed from DNA microarray analysis. *Mol. Cell. Biol.*, 21:4684–99.
- Ivey-Hoyle, M., Conroy, R. and Huber, H.E. et al. 1993. Cloning and characterization of E2F-2, a novel protein with the biochemical properties of transcription factor E2F. *Mol. Cell. Biol.*, 13:7802–12.
- Johansson, R. 1993. System modeling and identification. Prentice-Hall. p. 113–9.
- Kennedy, R.D., Gorski, J.J. and Quinn, J.E. et al. 2005. BRCA1 and c-Myc associate to transcriptionally repress psoriasis, a DNA damage-inducible gene. *Cancer Res.*, 65(22):10265–72.
- Klipp, E., Herwig, R. and Kowald, A., et al. 2005. Systems Biology in Practice: Concepts, Implementation and Application. Wiley-VCH.
- Kreyszig, R. 1993. Advanced Engineering Mathematics. 7th ed. John Wiley, New York.
- Knebel, A., Morrice, N. and Cohen, P. 2001. A novel method to identify protein kinase substrates: eEF2 kinase is phosphorylated and inhibited by SAPK4/p38delta. *EMBO J.*, 20(16):4360–9.
- Lang, S.E., McMahon, S.B. and Cole, M.D. et al. 2001. E2F transcriptional activation requires TRRAP and GCN5 cofactors. *J. Biol. Chem.*, 276:32627–34.
- Law, D.J., Du, M. and Law, G.L. et al. 1999. ZBP-99 defines a conserved family of transcription factors and regulates ornithine decarboxylase gene expression. *Biochem. Biophys. Res. Commun.*, 262(1):113–20.
- Liu ET. 2005. Systems Biology, Integrative Biology, Predictive Biology. *Cell.*, 121:505–6.
- Meroni, G., Reymond, A. and Alcalay, M. et al. 1997. Rox, a novel bHLHZip protein expressed in quiescent cells that heterodimerizes with Max, binds a non-canonical E box and acts as a transcriptional repressor. *EMBO J.*, 16(10):2892–906. Erratum in: *EMBO J.*, 16(19):6055.
- Monden, T., Wondisford, F.E. and Hollenberg, A.N. 1997. Isolation and characterization of a novel ligand-dependent thyroid hormone receptor-coactivating protein. *J. Biol. Chem.*, 272(47):29834–41.
- Muller, H. and Helin, K. 2000. The E2F transcription factors: key regulators of cell proliferation. *Biochim. Biophys. Acta.*, 1470:M1–12.
- Muller, H., Bracken, A.P. and Vernell, R. et al. 2001. E2Fs regulate the expression of genes involved in differentiation, development, proliferation, and apoptosis. *Genes Dev.*, 15:267–85.
- Nevins, J.R. 2001. The Rb/E2F pathway and cancer. *Hum. Mol. Genet.*, 10:699–703.
- Nurse, P. 2000. A long twentieth century of the cell cycle and beyond. *Cell.*, 100:71–8.
- Ortega, S., Malumbres, M. and Barbacid, M. 2002. Cyclin D-dependent kinases, INK4 inhibitors and cancer. *Biochim. Biophys. Acta.*, 1602:73–87.
- Pahl, P.M., Hodges, Y.K. and Meltesen, L. et al. 1998. ZNF207, a ubiquitously expressed zinc finger gene on chromosome 6p21.3. *Genomics*, 53(3):410–2.
- Polager, S., Kalma, Y. and Berkovich, E. et al. 2002. E2Fs up-regulate expression of genes involved in DNA replication, DNA repair and mitosis. *Oncogene*, 21:437–446.
- Ratziu, V., Lalazar, A. and Wong, L. et al. 1998. Zf9, a Kruppel-like transcription factor up-regulated in vivo during early hepatic fibrosis. *Proc. Natl. Acad. Sci. U.S.A.*, 95(16):9500–5.
- Reichman, T.W., Muniz, L.C. and Mathews, M.B. 2002. The RNA binding protein nuclear factor 90 functions as both a positive and negative regulator of gene expression in mammalian cells. *Mol. Cell. Biol.*, 22(1):343–56.
- Ren, B., Cam, H. and Takahashi, Y. et al. 2002. E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. *Genes Dev.*, 16:245–56.
- Robles, L.D., Frost, A.R. and Davila, M. et al. 2002. Down-regulation of Cdc6, a cell cycle regulatory gene, in prostate cancer. *J. Biol. Chem.*, 277:25431–8.
- Shah, J.V. and Cleveland, D.W. 2000. Waiting for anaphase: Mad2 and the spindle assembly checkpoint. *Cell.*, 103:997–1000.
- Sherlock, G. and Hernandez-Boussard, T. and Kasarskis, A. et al. 2001. The Stanford Microarray Database. *Nucleic Acids Res.*, 29:152–5.

- Sherr, C.J. and Roberts, J.M. 1999. CDK inhibitors: positive and negative regulators of G1-phase progression. *Genes Dev.*, 13:1501–12.
- Shiekhattar, R., Mermelstein, F. and Fisher, R.P. et al. 1995. Cdk-activating kinase complex is a component of human transcription factor TFIIH. *Nature*, 374(6519):283–7.
- Shirakawa, H., Yoshida, M. 1992. Structure of a gene coding for human HMG2 protein. *J. Biol. Chem.*, 267(10):6641–5.
- Stanelle, J. Stiewe, T. and Theseling, C.C. et al. 2002. Gene expression changes in response to E2F1 activation. *Nucleic Acids Res.*, 30:1859–67.
- Stillman, B. 1996. Cell cycle control of DNA replication. *Science*, 274:1659–1664.
- Stott, F.J., Bates, S. and James, M.C. et al. 1998. The alternative product from the human CDKN2A locus, p14(ARF), participates in a regulatory feedback loop with p53 and MDM2. *EMBO. J.*, 17:5001–5014.
- Takahashi, S., Wakui, H. and Gustafsson, J.A. et al. 2000. Functional interaction of the immunosuppressant mizoribine with the 14-3-3 protein. *Biochem. Biophys. Res. Commun.*, 274(1):87–92.
- Trimarchi, J.M. and Lees, J.A. 2002. Sibling rivalry in the E2F family. *Nat. Rev. Mol. Cell. Biol.*, 3(1):11–20.
- Valsesia-Wittmann, S., Magdeleine, M. and Dupasquier, S. et al. 2004. Oncogenic cooperation between H-Twist and N-Myc overrides failsafe programs in cancer cells. *Cancer Cell.*, 6:625–30.
- Whitfield, M.L., Sherlock, G. and Saldanha, A.J. et al. 2002. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol. Biol. Cell.*, 13:1977–2000.
- Wu, W., Fan, Y.H. and Kemp, B.L. et al. 1998. Overexpression of cdc25A and cdc25B is frequent in primary non-small cell lung cancer but is not associated with overexpression of c-myc. *Cancer Res.*, 58:4082–5.
- Wysocka, J., Myers, M.P. and Laherty, C.D. et al. 2003. Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1. *Genes Dev.*, 17(7):896–911.
- Yuh, C.H. and Davidson, E.H. 1996. Modular *Cis*-Regulatory Organization of *Endo16*, a Gut-Specific Gene of the Sea Urchin Embryo. *Development*, 122:1069–1082.