Prediction of Gene Regulatory Networks using Differential Expression of cDNA Microarray Data

[1]K.C. Shih, [1,2]R.M. Chen, [1]R.M. Hu, [1]F.M. Liu, [2]H.K. Chen, [3]Jeffrey J.P. Tsai
[1]*Institute of Bioinformatics,* [2]*Department of Information and Design*
Taichung Healthcare and Management University, Taiwan
[3]*Department of Computer Science, University of Illinois, Chicago, USA*

E-mail: rmchen@thmu.edu.tw

Abstract

Several of ten thousands functional genes control the growth, genetics, and behavior of living organisms by regulating different gene expressions. The genes in a normal cell control the process of cell growth, differentiation, reproduction, and apoptosis via multiple steps of interactive regulation mechanism. The mechanism of gene regulation is a very important process in human beings. If there is something wrong in the gene regulation mechanism, it may cause some diseases.

It is very difficult to identify the regulatory relations among genes in human genome. Traditional biological research methods consume huge amont of time and man strength. In recent years, with the rapid development of technologies, cDNA can be used to analyze the changes of gene expressions in different cells in a high throughput manner. In this paper, we propose a novel bioinformatics approach to predict the regulatory network of genes based on differential expressions of cDNA microarray databases for tomor and normal tissues. The differences in regulatory networks of genes for tumor and normal tissues reveal the information of finding possible cancer-related genes. The predicted cancerous genes can then be provided to biologists for further verification through biological experiments.

1. Introduction

The gene regulation mechanism is a very complicated process in human body. The expression level affects the process of gene regulation mechanism from the viewpoint of cell biology. The genes in a normal cell control the processes of cell growth, differentiation, reproduction, and apoptosis via multiple steps of interactive regulation mechanism. All the cells in human possess complete genes. Cells with different characteristics produce different gene expressions via different regulation mechanism such that produces different functions. Abnormal gene regulation mechanism results in gene expression which is different from that of normal regulation mechanism. The result is occurrence of disease, such as cancer. We can dig out the reason of arising disease due to abnormal genes from discovering the complex phenomena of gene expression and regulation relationship [1]-[5].

The processes of cancer are multiple steps resulting from instability of DNA. Many environmental factors result in changes of gene regulation mechanism. These factors include activation of oncogenes, inactivation of tumor suppressor gene and DNA repair genes, and disturbance of other cancer related factors, which cause abnormal changes in the process of cell growth, differentiation, reproduction, apoptosis, and cause cancer.

The regulation states of genes can be simplified as two classes: activation or inhibition. Biological signal will be gradually transferred and amplified to downstream genes through one-by-one regulation mechanism. Signal of downstream genes will also be the feedback to upstream genes. Any problems occurred in the gene expression will result in problems of gene regulation in the whole gene network.

Gene regulation mechanism is the basis of all life phenomena. However, it is very difficult to find the genes that cause problems in the gene regulation, especially by using the traditional one-by-one analysis method. Nowadays, high throughput microarray technologies were developed. By using these new technologies, tremendous amount of gene expression data were accumulated in a short time. These data can help us to analyze the complex relationship of gene interactions.

Currently developed approaches for studying gene regulation mechanism include Bayesian network [6]-[9], Boolean network [10], Fourier analysis [11], Clustering [12]-[14], K-mean [15] and Self Organizing Maps (SOM) [16]. In this paper, we propose a bioinformatic approach to analyze the regulatory network of genes based on differential expression of cDNA microarray databases for tumor and normal tissues. The differences

Proceedings of the IEEE Sixth International Symposium on Multimedia Software Engineering (ISMSE'04)
0-7695-2217-3/04 $20.00 © 2004 **IEEE**

in regulatory network of genes for tumor and normal tissues reveal the information of finding possible cancer related genes. A regulation truth table constructed based on the activation and inhibition of gene regulation mechanism is introduced to predict the gene regulation mechanism. Once the gene regulation mechanism is discovered, the biologists can identify the problems in the gene regulation network that cause disease. This new approach can be applied to the analysis of gene mutations, differential gene expression selection, drug discovery, pharmacology research, genotype classify-cation of disease, and gene expression analysis of cell cycle, etc.

This paper is organized as follows. In Section 2, we propose a new method for prediction of gene network. In Section 3, we discuss the pre-processing of microarray gene expression data. In Section 4, we illustrate the effectiveness of our method by comparing with P53 signaling pathway and applying this method to the analysis of stomach tumor and normal tissues data . Finally, we have the conclusion in Section 5.

2. Methods

In this section, we present an approach based on correlation analysis and differential expression of genes for the prediction of gene regulatory relations. The predicted gene regulatory pathway is then applied to investigate the differences of gene regulation mechanisms that were predicted based on microarray gene expression data of stomach primary tumor and non-tumor tissues.

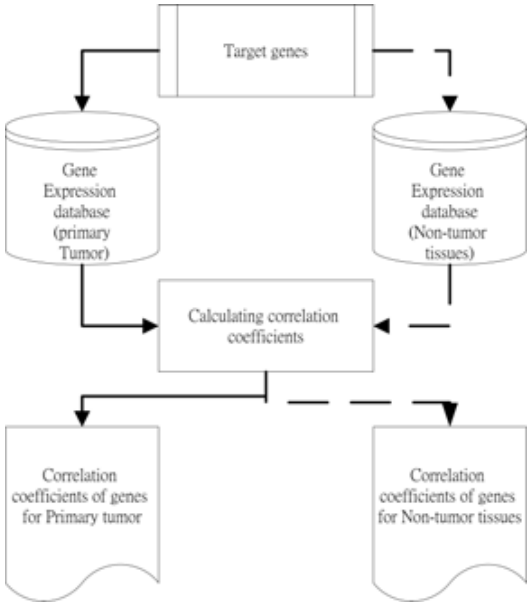

Figure 1. Flowchart for calculating correlation coefficients of microarray gene expression data

2.1. Correlation

Correlation is a technique for investigating the relations between two quantitative variables. The relations of genes can be constructed by computing their correlations. The correlations could be obtained by computing Pearson's correlation coefficient. Pearson's correlation coefficient is a measure of the strength of the association between two sets of gene expression data.

In Table 1, we show an example of the gene expression data of genes denoted by X and Y. Only three pairs of gene expression data in this table can be used to compute the Pearson's correlation coefficient since only three tissues recorded expression data in both genes.

The value of Pearson's correlation coefficient is between -1 and 1. If the value of correlation coefficient is between -1 and -0.5, the regulatory relation of genes is defined as a state of negative regulation. If the value of correlation coefficient is between 0.5 and 1, the regulatory relation of genes is defined as a state of positive regulation. If the value of correlation coefficient is between -0.5 and 0.5, the regulatory relation of genes is defined as a state of low correlation. The correlations have to be computed for each pair of genes to recognize the relations and regulatory states between them. Figure 1 illustrates the flowchart for calculating correlation coefficients of microarray gene expression data established from stomach primary tumor and non-tumor tissues. However, the correlation of genes could not be used to predict the regulation flow or direction.

2.2. Regulation truth table

The gene expression type in microarray data can be classified as over-expression, under-expression and NA (not available). Over-expression denotes over-activation of the gene. Under-expression denotes inadequate expression of the gene. NA in gene expression denotes unknown expression of the gene. The expression of a gene would possibly be regulated by the other genes. A gene may be activated or inhibited by the other genes, and vice versa. Figure 2 illustrates a basic gene regulation type where symbol → denotes activation and ⊣ denotes inhibition. The regulatory relations between genes X, Y and hidden gene are explained as follows.

(1) X → Y means that gene X activates gene Y. If the expressivity of gene X increases, the expressivity of gene Y increases (X↑=>Y↑). If the expressivity of gene X decreases, the expressivity of gene Y decreases (X↓=>Y↓).

(2) X ← Y means that gene Y activates gene X. If the expressivity of gene Y increases, the expressivity of

gene X increases (Y↑=>X↑). If the expressivity of gene Y decreases, the expressivity of gene X decreases (Y↓=> X↓).

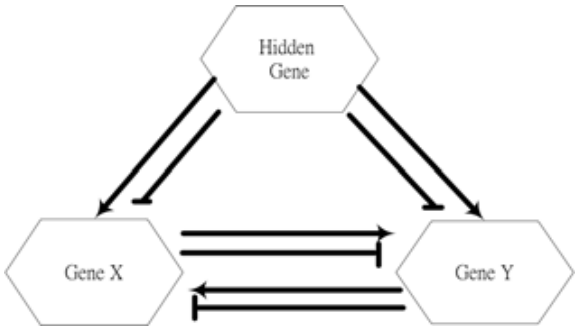

Figure 2. A basic gene regulation type

(3) X ⊣ Y means that gene Y is inhibited by gene X. If the expressivity of gene X increases, the expressivity of gene Y decreases (X↑=>Y↓). If the expressivity of gene X decreases, the expressivity of gene Y increases (X↓=>Y↑).

(4) X ⊢ Y means that gene X is inhibited by gene Y. If the expressivity of gene Y increases, the expressivity of gene X decreases (Y↑=> X↓). If the expressivity of gene Y decreases, the expressivity of gene Y increases (Y↓=> X↑).

(5) Two possible regulatory relations may result from the activation or inhibition of hidden gene on genes X and Y. This would cause the expressivities of gene X and Y increase (XY↑) or decrease (XY↓) in the same time.

The relations among genes X, Y and hidden gene were analyzed above. Ten possible gene regulation states were derived. Over-expression or under-expression of the parent genes will affect the expression of child genes in the gene regulatory network. The gene expression data in microarray provided information of gene expressions in a numerical form. If we use the microarray gene expression data to analyze the gene regulatory relations directly, the results will be influenced by the values of expression data. Hence, we classify the gene expression data as three gene expression codes: +1, -1 and 0 for the analysis of differential gene expressions. If the value of expression data is greater than 1, it is encoded as +1. If the value of expression data is smaller than -1, it is encoded as -1. If the value of expression data is not available, it is classified as 0. According the above rules and ten possible gene regulation states, we establish the regulation truth table (RTT) of genes X and Y as shown in Table 2. A letter 'T' in the cell of table denotes that the gene expression code (GEC) in corresponding row is matched with the regulation state in corresponding column of the table. A letter 'F' in the cell of table denotes that the gene expression code in corresponding row is unmatched with the regulation state in corresponding column of the table. The symbol '–' in the cell of table denotes that the gene expression code in corresponding row is possibly matched with the regulation state in corresponding column of the table.

Table 1. Example of gene expression data for genes X and Y

Gene Symbol	GC(HKG65N	GC(HKG69N	GC(HKG3N)	GC(HKG55N	GC(HKG56N	GC(HKG36N)
X	-1.36233	-1.72833		-1.70333		1.916
Y		1.706		-1.84633	1.991	-0.81033

Table 2. The regulation truth table of genes X and Y

GEC \ Regula	X ↑ =>Y ↑	X ↑ =>Y ↓	X ↓ =>Y ↑	Y ↑ =>X ↑	Y ↑ =>X ↓
X : -1 Y : -1	–	F	F	–	F
X : -1 Y : +1	F	–	T	F	T
X : +1 Y : -1	F	T	–	F	–
X : +1 Y : +1	T	F	F	T	F
X : -1 Y : 0	–	–	T	–	T
X : -1 Y : 0	T	T	–	T	–
X : 0 Y : -1	–	T	–	–	–
X : 0 Y : +1	T	–	T	T	T
X : 0 Y : 0	–	–	–	–	–

GEC	Regula	Y↓ =>X↑	X↓ =>Y↓	Y↓ =>X↓	XY ↑	XY ↓
X：-1	Y：-1	F	T	T	–	T
X：-1	Y：+1	–	F	F	–	–
X：+1	Y：-1	T	F	F	–	–
X：+1	Y：+1	F	–	–	T	–
X：-1	Y： 0	–	T	T	–	T
X：-1	Y： 0	T	–	–	T	
X： 0	Y：-1	T	T	T	–	T
X： 0	Y：+1	–	–	–	T	–
X： 0	Y： 0	–	–	–	–	–

Before analyzing the regulation relations of genes by regulation truth table, we have to encode the gene expression data in microarray databases to three gene expression codes and count case numbers of GEC for each gene. An example of encoding gene expression data for genes X and Y is shown in Table 3. The flowchart for predicting gene regulation network based on RTT is depicted in Figure 3. The RTT for genes X and Y with count of GEC is shown in Table 4 as an example of analyzing gene regulatory relations based on RTT. In this example, the matched gene regulation state is XY↑ only. The gene regulation state that is matched with the largest count numbers is selected as the final predicted gene regulatory state.

Table 3. Example of gene expression code for genes X and Y

Gene Symbol	GC(HKG65N	GC(HKG69N	GC(HKG3N)	GC(HKG55N	GC(HKG56N	GC(HKG36N
X	-1	-1	0	-1	0	1
Y	0	1	0	-1	1	-1

Table 4. RTT for genes ATM and MDM2 with count of GEC

Count	GEC	Regulation	X↑ =>Y↑	X↑ =>Y↓	X↓ =>Y↑	Y↑ =>X↑	Y↑ =>X↓
0	X：-1	Y：-1	–	F	F	–	F
0	X：-1	Y：+1	F	–	T	F	T
5	X：+1	Y：-1	F	T	–	F	–
2	X：+1	Y：+1	T	F	F	T	F
6	X：-1	Y： 0	–	–	T	–	T
2	X：-1	Y： 0	T	T	–	T	–
7	X： 0	Y：-1	–	T	–	–	–
3	X： 0	Y：+1	T	–	T	T	T
4	X： 0	Y： 0	–	–	–	–	–

Count	GEC	Regulation	Y↓ =>X↑	X↓ =>Y↓	Y↓ =>X↓	XY ↑	XY ↓
0	X：-1	Y：-1	F	T	T	–	T
0	X：-1	Y：+1	–	F	F	–	–
5	X：+1	Y：-1	T	F	F	–	–
2	X：+1	Y：+1	F	–	–	T	–
6	X：-1	Y： 0	–	T	T	–	T
2	X：-1	Y： 0	T	–	–	T	
7	X： 0	Y：-1	T	T	T	–	T
3	X： 0	Y：+1	–	–	–	T	–
4	X： 0	Y： 0	–	–	–	–	–

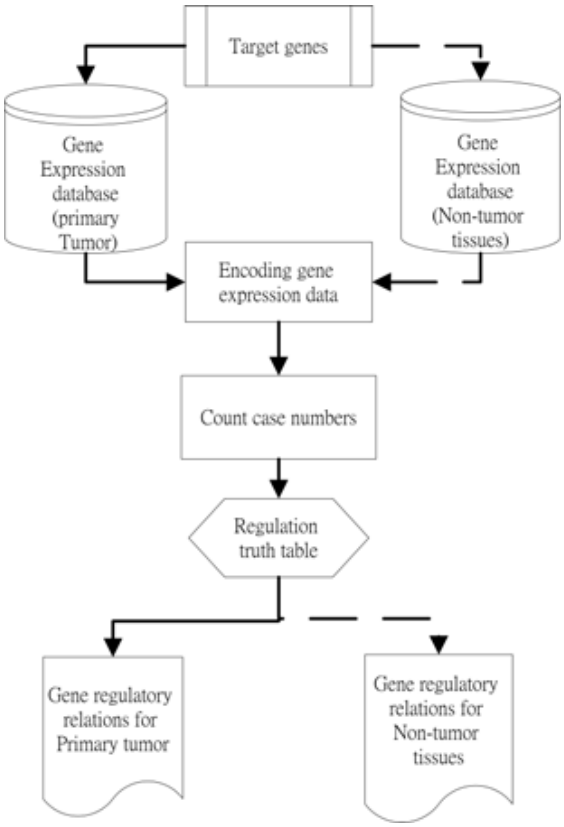

Figure 3. Flowchart for predicting gene regulation network based on RTT

3. Data pre-processing

Microarray technologies provided analysis of thousands of genes simultaneously, tremendous amount of gene expression data were accumulated in a short time. However, errors in data may occur due to artificial or environmental factors. For instance, some genes have multiple expression data with different expression types (over-expression or under-expression). And some genes have multiple expression data with different expression levels of the same type. Moreover, some genes have no expression data available or have expression data but without gene symbols. We call the last two cases as "bad expression data" cases. Hence, we have to pre-process gene expression data before proceeding to analyzing the differential expression and correlation coefficients of genes. The flowchart of data pre-processing is shown in Figure 4. The steps for acquiring data and data pre-processing are stated and explained as follows.

Step 1. The microarray raw data of stomach tumor tissues was collected from Stanford microarray database [17]. Every case was saved as an Excel file format. In this paper, we acquire 90 cases of expression data for primary tumor stomach tissues and 22 cases for non-tumor stomach tissues.

Step 2. Delete the "bad expression data" cases as mentioned above. These data cases would cause incorrect analysis results if we did not delete them from the data set.

Step 3. The microarray raw data of stomach tumor in Stanford microarray database had been normalized using GenePix. Hence, we select the data listed in column LOG_RAT2N_MEAN as the gene expression data.

Step 4. All the microarray raw data had the same format. Every data table includes fixed column names as Name, Clone ID, Gene Symbol, Gene Name, Cluster ID, and Accession. We combine the gene expression data (LOG_RAT2N_MEAN) of all cases into single table to analyze differential expressions of genes.

Step 5. One gene may have multiple experiments in microarray analysis. Hence, one gene may possess different signs of expression values (positive or negative) or have multiple expressions with different expression levels of the same numerical sign. For the first case, we have to give up the data since that the expression data of one gene should have only one kind of numerical sign theoretically. For the second case, these expression data of that gene are averaged. The averaged expression data is set as the expression data of that gene.

Step 6. Finally, the expression data is integrated as two databases: gene expression database for primary tumor and gene expression database for non-tumor tissue.

4. Results

In this section, we apply the proposed method to study the related genes in P53 signaling pathway. The gene regulatory relations in non-tumor and tumor tissues are compared to investigate the differences of gene regulatory relations in both cases. There are thirteen genes in our microarray gene expression database that are related to P53 signaling pathway (BCL2, ATM, TP53, TIMP3, CDKN1A, PCNA, RB1, E2F1, MDM2, GADD45A, CDK4, CCND1 and CDK2).

According to the RTT-based prediction flowchart depicted in previous section, we compute first the correlation coefficients and then analyze the regulation relations for each pair of genes via regulation truth table. The gene regulatory relations in P53 signaling pathway for non-tumor tissues are illustrated in Table 5. Ten possible regulation types are listed in the first column of the table denotes. The second row is the correlation coefficient of genes X & Y. This value could be used to recognize the relations of genes X & Y. A symbol * in

the table denotes that the genes (X&Y) in corresponding column possesses the regulation indicated in the corresponding row. The gene regulatory relations in P53 signaling pathway for primary tumor tissues are illustrated in Table 6. Note that only partial results are shown in Table 5 and Table 6 for illustration due to large amount of data.

Table 5. Gene regulatory relations in P53 signaling pathway for non-tumor tissues

Genes(X&Y)	BCL2&ATM	BCL2&TP53	CL2&CDKN1A	BCL2&RB1	BCL2&GADD45A	ATM&TIMP3	ATM&CDKN1A	ATM&RB1	ATM&MDM2	ATM&CCND1
Corrleation	0.657	-0.636	1	0.522	0.501	-0.703	-1	-0.635	0.609	-0.534
X↑=>Y↑			*							
X↑=>Y↓		*								
X↓=>Y↑						*		*		*
Y↑=>X↑			*	*	*					
Y↑=>X↓							*	*		
Y↓=>X↑		*								
X↓=>Y↓			*	*	*					
Y↓=>X↓	*		*						*	
XY ↑										
XY ↓									*	

Genes(X&Y)	TP53&TIMP3	TP53&RB1	TP53&E2F1	TP53&CDK4	TP53&CCND1	TIMP3&E2F1	TIMP3&MDM2	CDKN1A&RB1	CDKN1A&E2F1	CDKN1A&GADD45A
Corrleation	-0.507	-0.609	-0.536	0.681	0.507	0.758	0.63	-0.893	0.605	-0.547
X↑=>Y↑										
X↑=>Y↓										
X↓=>Y↑	*							*		*
Y↑=>X↑					*					
Y↑=>X↓	*	*								
Y↓=>X↑			*					*		*
X↓=>Y↓				*						
Y↓=>X↓						*	*		*	
XY ↑										
XY ↓						*	*		*	

Genes(X&Y)	PCNA&RB1	PCNA&MDM2	PCNA&CDK4	RB1&MDM2	RB1&CCND1	E2F1&MDM2	MDM2&CCND1	GADD45A&CDK4	CDK4&CCND1
Corrleation	0.635	0.534	0.634	-0.823	0.725	0.639	-0.633	0.705	0.581
X↑=>Y↑			*		*			*	*
X↑=>Y↓									
X↓=>Y↑							*		
Y↑=>X↑	*		*					*	
Y↑=>X↓							*		
Y↓=>X↑				*					
X↓=>Y↓	*		*			*			
Y↓=>X↓		*	*			*			
XY ↑					*				*
XY ↓		*				*			

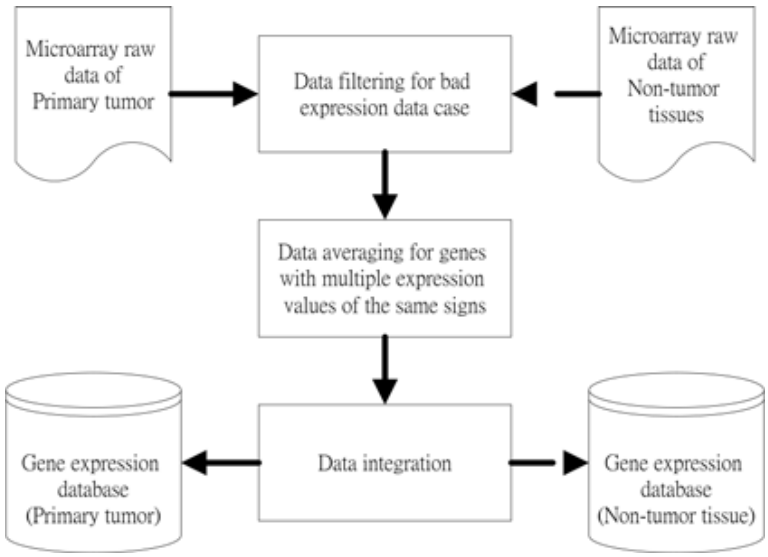

Figure 4. Flowchart of data pre-processing for gene expression data

Table 6. Gene regulatory relations in P53 signaling pathway for primary tumor tissues

Genes(X&Y)	BCL2&ATM	ATM&PCNA	ATM&MDM2	TP53&TIMP3	TP53&PCNA	TP53&RB1	TIMP3&PCNA	TIMP3&RB1	TIMP3&CCND1	CDKN1A&RB1
Corrleation	0.822	0.471	-0.723	0.342	0.34	0.481	-0.352	0.424	-0.31	0.243
X↑ =>Y↑					*			*		*
X↑ =>Y↓			*						*	
X↓ =>Y↑										
Y↑ =>X↑	*					*		*		*
Y↑ =>X↓									*	
Y↓ =>X↑							*			
X↓ =>Y↓	*									
Y↓ =>X↓										
XY ↑				*				*		*
XY ↓		*								

Genes(X&Y)	CDKN1A&E2F1	CDKN1A&CCND1	PCNA&MDM2	PCNA&CDK4	RB1&MDM2	RB1&CCND1	E2F1&CCND1	MDM2&CCND1
Corrleation	-0.358	0.262	0.155	0.651	-1	1	0.269	-0.391
X↑ =>Y↑						*		
X↑ =>Y↓					*			*
X↓ =>Y↑								
Y↑ =>X↑		*				*	*	
Y↑ =>X↓								*
Y↓ =>X↑	*							
X↓ =>Y↓		*		*				
Y↓ =>X↓								
XY ↑						*		
XY ↓			*					

To illustrate the effectiveness of our method, we compare the predicted gene regulatory relations with the regulatory relations of genes in P53 signaling pathway found from BioCarta [18]. For example, it could be found that genes BCL2 and TIMP3 were suppressed by P53 in P53 signaling pathway. The same regulatory relations could also be found in Table 5, i.e., genes BCL2 and TIMP3 decrease expression when TP53 increases expression, and vice versa. We could find that TP53 and BCL2 is negative correlated (correlation coefficients is -0.636) and so is TP53 and TIMP3 (correlation coefficients is -0.507). The other example is the regulatory relation between genes MDM2 and RB1 (MDM2↓=> RB1↓), whose correlation coefficient is -0.823. The third example is the regulatory relations of genes GADD45A and PCNA to gene CDK4. It could be found from P53 signaling pathway that genes GADD45A and PCNA activate gene CDK4. The same results could also be found in Table 5 (GADD45A↑=>CDK4↑ and PCNA↑=>CDK4↑). The correlation coefficient of GADD45A and CDK4 is 0.705 and 0.634 for PCNA and CDK4.

On the other hand, we investigate the differences of gene regulatory relations for cases of non-tumor and tumor tissues. By comparing the regulatory relations in Tables 5 and 6, it could be found that a lot of genes change their original regulation types. The gene correlations in tumor case are significantly reduced as compared to those of non-tumor case. For example, genes TP53 and RB1 are negative related with correlation coefficient -0.609 in non-tumor case; however, they are positive related in tumor case with correlation coefficient 0.481. Another example is that gene RB1 is suppressed by MDM2 in non-tumor case, but MDM2 is suppressed by RB1 in tumor case. Finally, it could also be easily found that some genes are unregulated one another in non-tumor case but have regulatory relations in tumor case. These changes in gene regulatory relations are very important in investigating the differences of gene regulatory networks for non-tumor and tumor cases. These differences could provide the biologists to acquire possible cancerous genes in a short time.

5. Conclusion

This paper considered the problem of predicting gene regulatory network using differential expressions of cDNA microarray data. A regulation truth table constructed based on the activation and inhibition of gene regulation mechanism was proposed to predict the gene regulation mechanism. We obtained consistent gene regulatory relations with those found from P53 signaling pathway. This approach had also been applied to investigate the differences of gene regulation mechanism that was predicted based on microarray gene expression data for primary tumor and non-tumor stomach tissues. The biologists could recognize the possible cancerous genes from comparing these two gene networks constructed above.

Resulting from artificial or environmental factors in the experiment, there are still many uncertain errors or noise in microarray gene expression data. These errors would cause an incorrect prediction of gene regulatory

network. For instance, by analyzing the regulation truth table, it was found that the same set of genes possessed both of active and inhibited relations. Hence, the gene expression data used in this paper was pre-filtered before proceeding to analyzing the differential expression and correlation coefficients of genes. This filtering process could enhance the reliability of microarray data and hence the correctness of predicted gene regulatory network.

Finally, the amount and quality of microarray data would largely affect the prediction results of gene regulatory network. Hence, further research will be focused on the collection of cancer microarray gene expression data and data pre-processing in the future.

References

[1] W.C. Hahn and R.A. Weinberg, "Modeling the Molecular Circuitry of Cancer", *Nat Rev Cancer*, Vol. 2, 2002, pp. 331-41.

[2] A. Kramer , K. Neben, and A.D. Ho, "Centrosome Replication, Genomic Instability and Cancer", *Leukemia*, Vol. 16, 2002, pp. 767-75.

[3] S.E. Andrew and A.C. Peters, "DNA Instability and Human Disease", *Am J. Pharmacogenomics*, Vol. 1, 2002, pp. 21-8.

[4] I.M. Frayling, "Methods of Molecular Analysis: Mutation Detection in Solid Tumors", *Mol Pathol*., Vol. 55, 2002, pp. 73-9.

[5] C. Bernstein , H. Bernstein , C.M. Payne, and H. Garewal, "DNA Repair/Pro-apoptotic Dual-role Proteins in Five Major DNA Repair Pathways: Fail-safe Protection Against Carcinogenesis", *Mutat Res*., Vol. 511, 2002, pp. 145-78.

[6] H. Dirk, "Sensitivity and Specificity of Inferring Genetic Regulatory Interaction from Microarray Experiments with Dynamic Bayesian Networks", *Bioinformatics*, Vol. 19, 2003, pp. 2271-82.

[7] P. Baldi and A.D. Long, "A Bayesian Framework for the Analysis of Microarray Expression Data: Regularized-test and Statistical Inferences of Gene Changes", *Bioinformatics*, Vol. 17, 2001, pp. 509-519.

[8] N. Friedman, "Using Bayesian Networks to Analyze Expression Data", *RECOMB*, Vol. 127, 2000.

[9] C. Yoo, V. Thorsson, and G. F. Cooper, "Discovery of Causal Relationship in a Gene-regulation Pathway from a Mixture of Experimental and Observational DNA Microarray data", *Pac Symposium Biocomputing*, 2002, pp. 498-509.

[10] L. Shmulevich, Edward R. Douherty, S. Kim and W. Zhang, "Probabilistic Boolean Network: a Rule-based Uncertainly Model for Gene Regulatory Networks", *Bioinformatics*, Vol. 18, 2002, pp. 261-274.

[11] P.T. Spellman, "Comprehensive Identification of Cell Cycle Regulated Genes of the Yeast Saccharomyces Cerevisiae by Microarray Hybridization", *Mol. Biol*. Cell, Vol. 9, 1998, pp. 3273-97.

[12] K.S. Pollard and M.J. van der Laan, "Statistical Inference for Simultaneous Clustering of Gene Expression Data", *Mathematical Biosciences*, Vol. 176, 2002, pp. 99-121.

[13] M.B. Eisen, "Clusters Analysis and Display of Genome Wide Expression Patterns", *PNAS*, Vol. 95, 1998, pp. 14863–14868.

[14] G.J. Mclachlan, R.W. Bean and D. Peel, "A Mixture Model-based Approach to the Clustering of Microarray Expression Data", *Bioinformatics*, Vol. 18, 2002, pp. 413-422.

[15] C. Arima, T. Hanai, and M. Okamoto, "Gene Expression Analysis Using Fuzzy K-Means Clustering", *Genome Informatics*, Vol. 14, 2003, pp. 334-335.

[16] A. Wallqvist, Alfred A. Rabow, Robert H. Shoemaker, Edward A. Sausville and David G. Covell, "Linking the Growth Inhibition Response from the National Cancer Institute's Anticancer Screen to Gene Expression Levels and Other Molecular Target Data", *Bioinformatics*, Vol. 19, 2003, pp. 2212-2224.

[17] http://genome-www.stanford.edu

[18] http://www.biocarta.com