Analysis for Gene Networks of Colon Cancer Based on Logical Relationships*

Xiuzhu Xu [1], Shudong Wang [1], Kaikai Li [1], Dashun Xu [3], Dazhi Meng [1,2]
[1]*College of Information Science and Engineering, Shandong University of Science and Technology,Qingdao, Shandong 266510, China*
[2]*College of Applied Science, Beijing University of Technology, Beijing 100124, China*
[3]*Department of Mathematics, Southern Illinois University Carbondale, Makanida 62958, USA*
Email: wangshd2008@yahoo.com.cn

Abstract *Genome forms gene networks in terms of complicated interaction to realize its functions. The fundamental principle that "the structure of system determines its function" inspires researchers to contrast the differences between the structure of gene networks for experimental and control groups to get useful information of molecular mechanism of the formation of cancer. In this research, the higher-order logical networks are constructed through reverse network modeling based on the expression data of phylogenetic profiles of cancer-related genes in the normal and diseased stages. Through contrasting of the network structure between experimental and control groups, such obvious differences in the structural components of networks are found as the number of non-isolated nodes and the logical relationships, the distribution of logic motifs of 2-order. Furthermore, the dynamical behaviors of logical networks of the commander genes of experimental and control groups are simulated and analyzed respectively and significant differences in dynamical stability are discovered. When the above method is applied to the expression data of different stages of other cancers, revelatory differences of the variation of reaction modes for the formation and development of cancer can be given, which is useful for the study of the mechanism of the development of cancer.*

Keywords *Systems biology Gene network Logical network Dynamics*

1. Introduction

The invasion and metastasis of colon cancer is an important reason that would influence the prognosis of the patients and lead to death. In recent years, through studying for the etiology and pathogenesis of colon cancer, it is generally accepted by the biomedical scientists that the formation and development of human's colon cancer is a complicated process which involves the alteration of many genes and stages. The process of canceration of the cells contains the stages of initiation, development and diffusion, and every stage involves the activation of oncogenes and the inactivation of the cancer suppressor genes. Hence, finding the functional genome related with disease characteristics from many virulence genes, the reaction modes among genes and the dynamic of oncogenes etc. is of great significance to the diagnosis and cure of the cancer and drug design. This is also an important project in the research of

*Supported by the National Natural Science Foundation of China (Grant Nos. 60874036, 60503002) and SDUST Research Fund

bioinformatics [1-2]. In recent years, the theoretical methods for studying the pathogenesis of cancers are mostly based on the expression data of phylogenetic profiles[3−11], and the computational methods have just been developed to detect functional linkages between proteins or genes, such as Pearson correlation coefficient, Euclidean and Hamming distances, mutual information, the hypergeometric distribution and shortest-path analysis[12−14]. However, traditional pairwise relationships cannot adequately illustrate the complexities that arise in cellular networks because of branching and alternative pathways. For instance, Shikimate 5-dehydrogenase (COG0169) is present if and only if 3-dehydroquinate dehydratase (COG0757) or 3-dehydroquinate dehydratase (COG0710) is present [15]. In order to further reveal the ubiquitous logical relationships mentioned above in biology, a computational approach-logic analysis of phylogenetic profiles (abbreviated as LAPP) was proposed for identifying detailed relationships among proteins on the basis of genomic data, which was applied into 4873 distinct orthologous protein families of 67 fully sequenced organisms, and identified 750, 000 triplets from analysis of the original, unshuffled biological protein profiles [15]. LAPP can reveal the causal or restricted relationships in complex network, help biologists to understand the unknown logical relationships among genes or proteins, and further discover the new biological mechanisms through experiments. As an application of triplet logic analysis to 85 diffuse infiltrating gliomas quantified using oligonucleotide arrays, the meaningful sets of genes that matched clinical outcomes were explored [16]. A Bayesian modeling framework for combining phylogenetic profile data via a likelihood with Rosetta Stone data via a prior probability was described [17]. A three-way gene interaction model was proposed that captured the dynamic of co-expression relationships between two genes [18]. A logical network with 16 active genes of shoot in different external stimuli was constructed, and the dynamics of the network was analyzed [19].

In this work, the logical networks of the normal and diseased stages are constructed through LAPP method based on the expression data of phylogenetic profiles of cancer-related genes. Through contrasting of the network structure between experimental and control groups, such obvious differences in the structural components of networks are found as the number of non-isolated nodes and the logical relationships, the distribution of logic motifs of 2-order. The differences inspire biomedical researchers: the reason that organism is normal in the normal stage might be that the biochemical reaction mode, the extent of gene relationships and most of the genes play normal roles. Furthermore, the dynamical behaviors of logical networks of the commander genes of experimental and control groups are simulated and analyzed respectively and significant differences in dynamical stability are discovered that the number of attractors for control group is far less than the one of the experimental groups. This paper is organized as follows: in the first part, the background of our work and the present researching situation of logical network of genes. In the second part, the method of LAPP is clarified briefly and the data sources and the processing methods of the data are given in the third part. The logical networks corresponding to experimental and control groups are constructed using LAPP method respectively and the structural components of networks are contrasted and analyzed. The dynamical attractors of logical networks of the commander genes in the experimental and control groups are contrasted respectively in the fifth part. The above results are analyzed and discussed, and the problems to be solved further are put forward in the sixth part.

2. The Method of LAPP

In order to understand and predict the functions of biological systems, we must first identify the structure of the system. For example, in order to show the regulatory relationships among the genes, we must identify all the components, the functions of every component, their interrelations and all the parameters related in the system. Now, there are two main methods to

build the model of biological system structure: forward and reverse network modelings. Reverse logical network modeling is a method of reverse network modeling and can determine the logical relationships and types among elements of systems. In this work, we identify the logical structure of cancer-related genome using LAPP method based on the data of gene expression profile. We entitle the sample data of gene obtained in many different experiments as its expression profile. For the convenience of writing and computing, we also denote the sample data by the genes. For example, the logic of genes A and B indicates the logical relationships of expression data of genes A and B. LAPP method is determining the logical relationships of 1-, 2-, and 3-order by calculating the uncertainty coefficient (abbreviated as U) of logical functions of every order among the genes based on the expression profiles of genes. The 1-order logical relationship between genes A and B is determined as follows:

$$U(B \mid f_1(A)) = \frac{H(B) + H(f_1(A)) - H(B, f_1(A))}{H(B)}, \tag{1}$$

where H refers to the entropy of the individual or joint distributions. $U(Y|X)$ denotes the uncertainty coefficient of the influence of X on Y. The size of U value denotes the statistical possibility of the uncertainty logic of 1-order of X to Y. f_1 is one of the proper functions of 1-order logic of A to B. The proper functions of 1-order logic are divided into two types (Table 1): $B = A$, namely, the presence of A leads to the presence of B, which is called the synchronization (equivalent with $\neg B = \neg A$); and $B = \neg A$, namely, the presence of A leads to the absence of B, which is called the asynchronization (equivalent with $\neg B = A$).

Table 1 The list of the proper functions and logic types of 1-, 2-order. Obviously, the proper functions of 1-order and logic types are the same.

order	The proper function and its serial number	The logic type and its serial number
1-order	1. $B = A$	1. $B = A$
	2. $B = \neg A$	2. $B = \neg A$
2-order	1. $C = A \wedge B$	1. $C = A \wedge B$
	2. $C = \neg(A \wedge B)$	2. $C = \neg(A \wedge B)$
	3. $C = A \vee B$	3. $C = A \vee B$
	4. $C = \neg(A \vee B)$	4. $C = \neg(A \vee B)$
	5. $C = A \wedge \neg B$ 6. $C = \neg A \wedge B$	5. $C = A \wedge \neg B$ $C = \neg A \wedge B$
	7. $C = A \vee \neg B$ 8. $C = \neg A \vee B$	6. $C = A \vee \neg B$ $C = \neg A \vee B$
	9. $C = \neg(A \leftrightarrow B)$	7. $C = \neg(A \leftrightarrow B)$
	10. $C = A \leftrightarrow B$	8. $C = A \leftrightarrow B$

Similarly, the expressions of 2-, 3-order logics are as follows:

$$U(C \mid f_2(A,B)) = \frac{H(C) + H(f_2(A,B)) - H(C, f_2(A,B))}{H(C)}, \tag{2}$$

$$U(D \mid f_3(A,B,C)) = \frac{H(D) + H(f_3(A,B,C)) - H(C, f_3(A,B,C))}{H(D)}, \tag{3}$$

where f_2, f_3 are one of the proper functions of 2-, 3-order logics respectively. The logics of

2-order among A, B and C contain 10 proper functions and 8 logic types (Table 1). The logics of 3-order among A, B, C and D contain 218 proper functions and 68 logic types.

Of course, the higher order logics can be calculated, but the complexities are higher. In this research, we only consider 1-, 2-, and 3-order Boolean logics among genes. The detailed can be found in references [15]-[17].

3. Data Source and Processing

3.1 Data Source

National Center for Biotechnology Information (NCBI) and the European Bioinformatics Institute (EMBL-EBI) provide the database generally used in field of bioinformations. For description convenience, we refer to the sample data of stage A, B and C of colon cancer as experimental groups, normal stage as control group. The sample data is selected from GPL570 in NCBI. The sample numbers of stage A, B and C are 39, 103, and 92 respectively. They are all from GSE2109 in GPL570. The 53 samples of normal stage consist of 10 samples of GDS2609, 11 samples of GSE10715 and 32 samples of GSE8671. The data mentioned above include p_value and P-M-A, where P, A and M indicate presence, absence and margin respectively. We represent P with 1, A and M with 0. Every sample database includes p_values and 0-1 expression profiles of 20827 genes of human (corresponding to 54676 probes). The four databases (normal, stage A, B and C) mentioned above form the original database used in this work.

3.2 Selection of Cancer-related Genes

In the original database mentioned above, every database includes the sample data of more than 20,000 genes of human genome. To construct and analyze the logical networks of more than 20,000 genes, the computational complexity is so high that it is beyond our computational ability. In order to simplify the computation, take the genes corresponding to 801 probes into account according to the known cancer-related genes [21]. If several probes correspond to the same gene, then we choose the data expressing the most in probes as the expression profile data of the gene. Thus, 286 cancer-related genes in the original database are obtained. Because we are concerned about the structure of the gene network, among these 286 cancer-related genes, the data almost 0 or 1 contribute very little to the difference of the network structure. Hence, we delete these genes in whose expression profile the numbers of 1 are less than 15% or more than 90%. And then, the numbers of the remaining cancer-related genes of these four databases is normal: 91, stage A: 79, B: 70, C: 60 respectively. After the above processing, there are some genes with the same 0-1 expression profiles, which contribute equally the construction of network. Hence, we preserve one of them, and delete other genes with the same expression profiles. Thus, the number of genes in normal database is reduced to 79. After simplifying the samples treated as above, the obtained databases are the working databases.

4. Numerical Experiments, Methods and Results

4.1 Determination of Every Order Logics

Based on the above working databases, all the 1-order U values between genes are calculated respectively by the formula (1). Obviously, there are 1-order U values of two directions between any two genes, denoted by $U(\mathrm{B}|\mathrm{A})$ and $U(\mathrm{A}|\mathrm{B})$ respectively. If the absolute value of the ratio of the difference between the two directions and the average value is less than

certain threshold d, namely, $\frac{2\left|U\left(\mathrm{B}|\mathrm{A}\right)-U\left(\mathrm{A}|\mathrm{B}\right)\right|}{U\left(\mathrm{B}|\mathrm{A}\right)+U\left(\mathrm{A}|\mathrm{B}\right)}<d$, then by the certainty of logical relationship, there is not 1-order logic between the two genes. Only when $\frac{2\left|U\left(\mathrm{B}|\mathrm{A}\right)-U\left(\mathrm{A}|\mathrm{B}\right)\right|}{U\left(\mathrm{B}|\mathrm{A}\right)+U\left(\mathrm{A}|\mathrm{B}\right)}>d$ and the U value is greater than 1-order threshold, there is 1-order logic between genes A and B. If $U\left(\mathrm{B}|\mathrm{A}\right)>U\left(\mathrm{A}|\mathrm{B}\right)$, then we consider that A regulates B, otherwise, B regulates A. By the formula (1) in the algorithm of U value, $U\left(\mathrm{B}|\mathrm{A}\right)=U\left(\mathrm{B}|\neg\mathrm{A}\right)$, which means that the logic type between A and B might be $\mathrm{A}\to\mathrm{B}$ or $\neg\mathrm{A}\to\mathrm{B}$. Then we determine the logic type between A and B according to the support of $\mathrm{A}\to\mathrm{B}$ and $\neg\mathrm{A}\to\mathrm{B}$, namely, the logic type with greater support is the one between A and B. Here, the support refers the ratio of the number of the two states being 1 in the expression profile data of the genes and the total number of samples. For instance, suppose the expression profiles of the genes *M* and *N* are $(0,0,0,1,1)$ and $(1,0,0,1,1)$ respectively, then the supports of logic type $\mathrm{M}=\mathrm{N}$ and $\mathrm{M}=\neg\mathrm{N}$ are $\frac{2}{5}$ and $\frac{1}{5}$ respectively. Take the threshold of 1-order $u_1=0.3$, then $U(\mathrm{N}|\mathrm{M})=U(\mathrm{N}|\neg\mathrm{M})=0.43>u_1$, the support $S_{\mathrm{M}=\mathrm{N}}=\frac{2}{5}>S_{\mathrm{M}=\neg\mathrm{N}}=\frac{1}{5}$, thus we think the logic type of genes M and N is $\mathrm{M}=\mathrm{N}$.

The proper functions of 2-, 3-order among genes in every database are determined by the formula (2), (3) and the support. That is to say, for any three genes A, B and C, calculating the U values corresponding to the 10 proper functions, then, the function f_2 corresponding to the maximum U value is the proper function of A, B to C. Similarly, for any four genes A, B, C and D, calculating the U values corresponding to the 218 proper functions, then, the function f_3 corresponding to the maximum U value is the proper function of A, B , C and D. If the U values corresponding to several proper functions are the maximum value, then take the proper function with greater support as the one among genes (similarly to determining the logic type (or proper function) of 1-order).

By the above numerical calculation, the complete directed logical networks corresponding to four databases of experimental and control groups, namely, each network contains all the proper functions of 1-, 2-, 3-order of its own database.

4.2 Determination of Threshold of Every Order

In order to highlight the characteristics of the network and obtain the valuable biological information, we need to choose a threshold to carry out coarse graining to the complete directed logical networks mentioned above. The threshold is the expression of coarse-grained level and the structural difference between logical network structures should be contrasted in the same coarse-grained level. In order to make the logical network structures for experimental and control groups comparable, we normalize the U values of 1-, 2-, 3-order of the four networks, namely: logic U values of every order of each database are replaced with $\frac{\mathrm{U}}{\max \mathrm{U}}$, where $\max \mathrm{U}$ is the maximum U value of each order in the own database. For simplicity, the normalized values are also denoted by the U values.

Let the thresholds of 1-, 2-, 3-order be u_1,u_2,u_3 respectively. The 2-order logic of A , B to C is considered only when the 1-order logics of A to C and B to C do not exist, and claim that

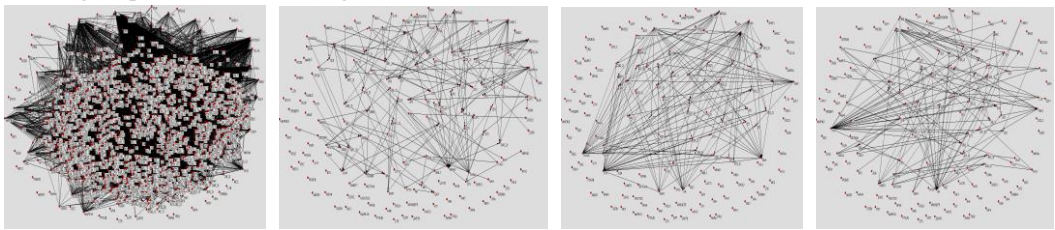
u_2 is greater than u_1:

$$\begin{cases} U(C|f_2(A,B)) > u_2 \\ U(C|f_1(A)) < u_1 \\ U(C|f_1(B)) < u_1 \\ u_2 > u_1 \end{cases}, \quad (4)$$

The 3-order logic of A, B, C to D is considered only when the 1-order logics of A to D, B to D, C to D, and the 2-order logics of any two of A, B, C to D do not exist, and u_3 is not less than u_2:

$$\begin{cases} U(D|f_3(A,B,C)) > u_3 \\ U(D|f_2(A,B)) < u_2 \,\&\, U(D|f_2(A,C)) < u_2 \,\&\, U(D|f_2(B,C)) < u_2 \\ U(D|f_1(A)) < u_1 \,\&\, U(D|f_1(B)) < u_1 \,\&\, U(D|f_1(C)) < u_1 \\ u_3 \geq u_2 \end{cases}, \quad (5)$$

Through the above numerical calculation, the four logical networks for experimental and control groups are obtained (Fig. 1). The detailed situations can be seen in Table 2.



(a) Normal (b) Stage A (c) Stage B (d) Stage C

Fig. 1 Taking 0.42, 0.75 and 0.85 as the threshold of 1-, 2-, 3-order respectively, the logical networks of 111 genes of the four databases are located, where (a), (b), (c) and (d) represent the logical networks corresponding to normal, stage A, B and C respectively. The nodes in the networks are divided into two types: the node labeling *i* denotes gene *i*, the node labeling "j(. . .)" (j = 2, 3) denotes the intermediate node of *j*-order logical relationships. For example, "2(5_1)" denotes logic type 5_1 of 2-order. The out-degree and in-degree of the intermediate nodes are fixed, i.e. when j =2, the in-degree and out-degree of the node is 2 and 1, respectively; when j = 3, the in-degree and out-degree of the node is 3 and 1, respectively. For instance, the node "2(3)" in Fig. 1(b) denotes that gene CXCL and MMP2 regulate gene EXTL3 through logic type 3 of 2-order.

Table 2 The detailed list of the logical networks for control and three experimental groups

Stage	No. of 1-order logics	No. of 2-order logics	No. of of 3-order logics	No. of all the logics	No. of genes	No. of the non-isolated nodes
Normal	90	1614	39	1743	79	74
Stage A	22	33	18	73	79	60
Stage B	37	51	2	90	70	55
Stage C	4	32	9	45	60	35

4.3 Distribution of Logic Motifs of 2-Order

The subgraph with higher frequency in complex network is called the network motif [22]. The identification to the motifs of complex network helps to identify the typical local connections

of network. Previous studies have showed that: the motif in complex biological network has a direct biological significance. For example: the motif in the protein interaction networks of yeast highly evolves to protect the components, the transcription regulatory networks of different species have an evolutionary trend to the same motif [23]. In the logical networks, each of 8 logic types of 2-order is corresponding to a motif, called the logic motif. Each type of logic motifs is corresponding to a typical mode of biochemical reactions. In order to explore the differences of reaction modes in the internal mechanism of normal and diseased organisms, we contrast and analyze the logic motifs of 2-order of the logical networks for experimental and control groups. Fig. 2 and Table 3 are the distribution histogram and the detailed list about logic motifs of 2-order of the four logical networks for experimental and control groups respectively. From Fig. 2 and Table 3, we can discover that the logic motifs of 2-order in control network cover six logic types 1, 2, 3, 4, 5, 6. However, in the networks for experimental groups, there only contain three logic types 1, 3, 6 (Table 3) except that the network corresponding to stage A contains logic types 1，2，3，5，6，7 of 2-order. Moreover, three logic motifs of types 1, 3 and 6 appear more in the network for control group and the corresponding proper functions are $C = A \wedge B$, $C = A \vee B$ and $C = \neg A \vee B$ or $C = A \vee \neg B$, respectively; One logic motif of type 1 appears more in the network for stage A and the corresponding proper function is $C = A \wedge B$; Two logic motifs of types 1 and 3 appear more in the networks for stage B and C and the corresponding proper functions are $C = A \wedge B$, respectively.

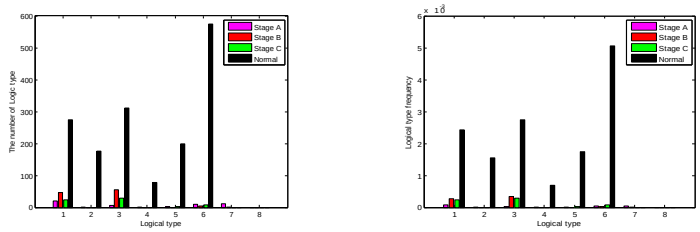


Fig. 2 The distribution histogram of logic types of 2-order in the logical networks for control and three experimental groups

Table 3 The list of logic types of 2-order in the logical networks for control and three experimental groups

Logic type	1	2	3	4	5	6	7	8
Normal	275	177	311	78	199	574	0	0
Stage A	13	1	5	0	2	7	5	0
Stage B	21	0	28	0	0	1	1	0
Stage C	12	0	17	0	0	3	0	0

4.4 Dynamical Analysis of Logical Network

Although the study and analysis of topological structure of the above logical networks can obtain some biological principles and information, the structure of actual biological system is a changing network over time and the dynamical behaviors cannot be reflected by topological structure in most situations. Therefore, we research the evolution of gene regulatory logical network from the dynamical point of view.

Because the number of the non-isolated nodes are too great in the logical networks corresponding to experimental and control groups to analyse the dynamical transfer laws, “the structural key genes” must be selected appropriately in every stages. The logical network is directed and every node in the network has out-degrees and in-degrees. If node i has an out-arc to j, then node i has a logical regulatory relationship to j. On the other hand, if node i has an in-arc, then some node has a logical regulatory relationship to i. The degree-difference of node is defined as the difference of out-degree and in-degree. The nodes

whose degrees and degree-differences are relatively great are called "the commander genes", and the changes of states of these genes can influence most of genes of the network. In this work, the method of discovering the commander genes is as follows: (1) Calculate the maximum degree and the maximum degree-difference of network; (2) Normalize the the maximum degree and the maximum degree-difference of the non-isolated nodes respectively, namely: the degree and degree-difference of each non-isolated node divide the maximum degree and the maximum degree-difference of the network respectively. (3) Set the parameters $c_1 = 0.05$ and $c_2 = 0.15$. Take these genes with degree c_1 times greater than the maximum degree and degree-difference c_2 times greater than the maximum degree-difference as the commander genes. When taking the parameters c_1 and c_2, the number of commander genes must be appropriate, because some key genes might be lost if the commander genes are too few, and we cannot calculate if the commander genes are too many. The detailed situations of the commander genes of control and three experimental groups can be seen in Table 4 and 5.

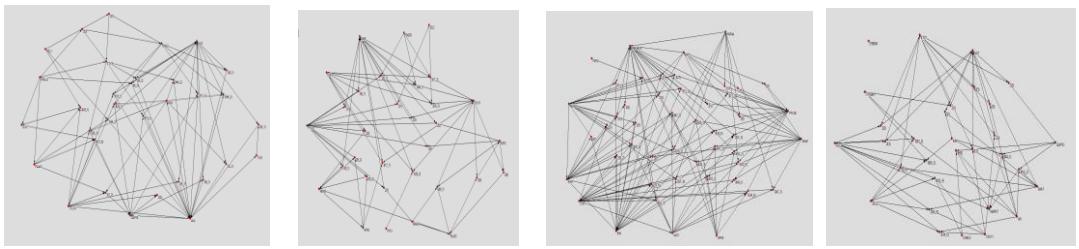
Table 4 Taking the parameters $c_1 = 0.05$ and $c_2 = 0.15$, the detailed list of the commander genes of control and three experimental groups

Stage	The maximum degree	The maximum degree-difference	No. of the commander genes	No. of the non-isolated nodes
Normal	855	109	12	74
Stage A	30	15	12	60
Stage B	42	6	12	55
Stage C	32	5	13	35

Table 5 The list of the commander genes of every stages

Normal	Stage A	Stage B	Stage C
FASLG	ARAF	ARAF	AR
AXL	BOK	BAK1	BAK1
ETV3	ESR2	ESR1	ELK1
GLI3	PTK2B	PTK2B	ERBB4
IL1A	GLI2	FGF2	ETV1
PLAG1	MAFG	HNF4A	FGF2
RBL1	MDM2	IHH	IHH
TAL1	PDGFB	ABCB1	MAPK7
VAV1	PLAG1	SHH	SELE
ZAP70	STK11	VAV1	STK11
TCL1A	TP73	ARHGEF5	TP53BP1
BAG2	HPSE	HPSE	ZAP70
			CHEK2

Using the method of constructing the logical networks in (1) (2) of Section 4, we can obtain the networks of the commander genes of control and three experimental groups (Fig. 3)



(a) Normal (b) Stage A (c) Stage B (d) Stage C

Fig. 3 The logical networks of the commander genes of control and three experimental groups.

Now we begin to analyse the dynamics of the above logical networks of the commander genes. Assume the network have n genes. Let the state of gene i at moment t be $x_i(t) \in \{0,1\}$, and the state-vector of gene i at moment $t+1$ be $s(t+1) = (x_1(t+1), x_2(t+1), ..., x_n(t+1))$. Without loss of generality, the triple set comprised of the U values, the corresponding proper functions of 1-order and the genes affecting gene i at moment t is $Z_1 = \{(k, U_{k\to i}, f_k^1) \mid k \in \{1,2,...,n\}, k \neq i\}$, which denotes that gene k regulates gene i by the proper function f_k^1 and the uncertainty coefficient is $U_{k\to i}$; the triple set consisting of the U values, the corresponding proper functions of 2-order and the genes affecting gene i is $Z_2 = \{((k_1,k_2), U_{k_1,k_2\to i}, f_{k_1,k_2}^2) \mid k_1,k_2 \in \{1,2,...,n\}, k_1 \neq i, k_2 \neq i, k_1 \neq k_2\}$, which denotes that gene k_1,k_2 regulate gene i by the proper function f_{k_1,k_2}^2 and the uncertainty coefficient is $U_{k_1,k_2\to i}$; the triple set consisting of the U values, the corresponding proper functions of 3-order and the genes affecting gene i is

$Z_3 = \{((k_1,k_2,k_3), U_{k_1,k_2,k_3\to i}, f_{k_1,k_2,k_3}^3) \mid k_1,k_2,k_3 \in \{1,2,...,n\}, k_1 \neq i, k_2 \neq i, k_3 \neq i, k_1 \neq k_2, k_1 \neq k_3, k_2 \neq k_3\}$, which denotes that gene k_1,k_2,k_3 regulate gene i by the proper function f_{k_1,k_2,k_3}^3 and the uncertainty coefficient is $U_{k_1,k_2,k_3\to i}$. In this research, we consider that the state of gene i at moment $t+1$ is related to the states of itself and its adjacent nodes. At moment t, let genes $k_1,k_2,\cdots,k_j$ regulate gene i by proper function $f_{k_1,k_2,\cdots,k_j}^j$ ($j \in \{1,2,3\}$). If the truth-value of proper function $f_{k_1,k_2,\cdots,k_j}^j$ with independent variable being the state of genes $k_1,k_2,\cdots,k_j$ at moment t is 1, then these j genes $k_1,k_2,\cdots,k_j$ activate the expression of gene i. Let $\underset{k_1,k_2,\cdots,k_j\to i}{\text{Sign}}(t) = 1$. Otherwise, these j genes $k_1,k_2,\cdots,k_j$ repress the expression of gene i. Let $\underset{k_1,k_2,\cdots,k_j\to i}{\text{Sign}}(t) = -1$. At moment t, for gene i, let

$$P_i(t) = x_i(t) + \sum_{Z_1} U_{k\to i} \underset{k\to i}{\text{Sign}}(t) + \sum_{Z_2} U_{k_1,k_2\to i} \underset{k_1,k_2\to i}{\text{Sign}}(t) + \sum_{Z_3} U_{k_1,k_2,k_3\to i} \underset{k_1,k_2,k_3\to i}{\text{Sign}}(t). \quad (6)$$

Then, the transition rule is denoted by

$$x_i(t+1)=\begin{cases}1, P_i(t)\geq 0.5,\\ 0, P_i(t)<0.5,\end{cases} \quad i=1,2,\ldots,n. \tag{7}$$

According to the above transition rules, the state transition configuration network for control and three experimental groups are obtained. The detailed situations are listed in Table 6, Fig. 4 and Fig. 5. In the networks corresponding to control and three experimental groups, the numbers of nodes (genes) have certain differences. In order to eliminate the differences caused by the number of attractors, the stability of system is measured by the frequency of attractors instead of the number of attractors.

Table 6 The situations of attractors of control and three experimental groups

Stage	No. of genes	No. (frequency) of 1-periodic attractor	No. (frequency) of other attractor	No. (frequency) of all the attractors
Normal	12	262（6.40%）	22（0.54%）	284(6.94%)
Stage A	12	594（14.50%）	6（0.15%）	600(14.65%)
Stage B	12	407（9.94%）	0（0%）	407(9.94%)
Stage C	13	1812（22.12%）	116（1.42%）	1928(23.54%)

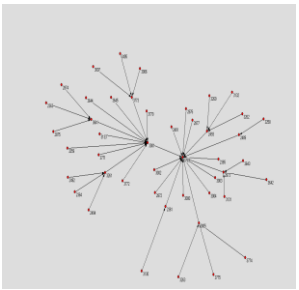
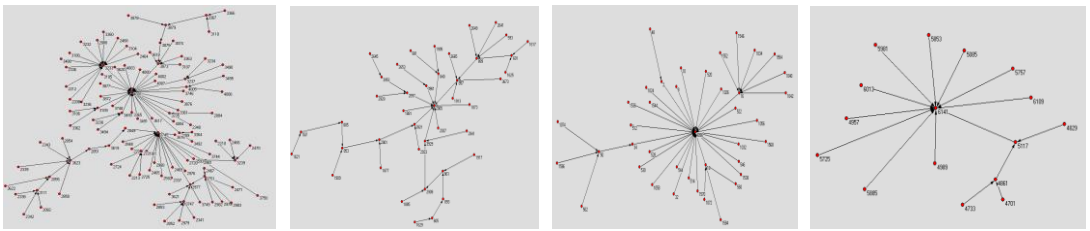


Fig. 4 2-periodic domain of attraction with the most nodes (46 nodes) of control group, where the node is labeled by the decimal number corresponding to the binary state vertor.



(a) Normal (b) Stage A (c) Stage B (d) Stage C

Fig. 5 In the state transition configuration networks for control and experimental groups, the 2-periodic domain of attraction with the most nodes, where the transition networks for normal, stage A, B and C contain 111, 42, 40, 16 nodes respectively.

5. Conclusions and Discussions

In this work, we construct the logical networks using LAPP method based on the expression data of phylogenetic profiles of cancer-related genes in normal, stage A, B and C of colon cancer . Through contrasting and analyzing of the structures of four networks corresponding to

experimental and control groups, we find the significant differences in the structural components (Fig. 1 and Table 2). (1) The significant differences in the number of non-isolated nodes: in the network for control group, the number of non-isolated nodes is 74; in stage A, B and C of colon cancer, the numbers of non-isolated nodes are reduced obviously, moreover, the fewer the numbers of non-isolated nodes are, the more cancer deteriorates. (2) The significant differences in the the number of logical relationship: the number of logics of the network for control group is far more than the one of the three experimental groups. In the complex networks, the number of non-isolated nodes reflects the intensity of correlation relationship of nodes and the number of the relationships of nodes indicates the richness of the relationship of nodes of network. In the complex networks, most of genes in the normal stage play cancer suppression roles with other genes cooperating perfectly. For some (internal and external) reasons, the regulatory relationships of several genes with some other genes are lost completely or the regulatory pathways malfunction seriously. The normal tissue appears abnormal; the organism suffers from the colon cancer. With the intensifying of the failure of regulation and broadening of the failure propagation, the illness increase gradually. (3) The significant differences in the distribution of logic motifs of 2-order (see Fig. 2 and Table 3): in the network for control group, there are three logic motifs: type 1, 3 and 6; in the network corresponding to stage A, there are one logic motif: type 1; in the networks for stage B and C, there are two logic motifs: type 1 and 3. We find from the above results that the logic type 1 appears in the networks for control and experimental groups, yet the logic type 6 appears only in the network for control group. In this research, we consider that logic types (motifs) of 2-order are corresponding to typical modes of biochemical reactions in life actions. The above phenomena might show that logic type 1 ($C = A \wedge B$) is the basic mode maintaining human existence and physiological metabolism reactions, yet logic type 6 ($C = A \vee \neg B$ or $C = \neg A \vee B$) is the reaction mode suppressing the normal organism not suffering from the colon cancer. The materials[24] have showed that logic type 1 does appear frequently in the complex biochemical reactions of the human body. For example: the coupling protein X and Y plays a regulatory role in the expression of protein Z, where X, Y and Z are the specific proteins. In conclusion, the logical networks for control and experimental groups are the models of structural mechanism of the normal organism suffering from the colon cancer gradually. The organism is normal in the normal stage because the biochemical reaction mode (type 6), the extent of genes relationship (the number of logical relationship) and most of the genes (the number of non-isolated nodes) play normal roles. With the change of the reaction mode of the organism and the decreasing of the number of acting genes and their relationships, many intrinsic biological functions are weaken or lost gradually, thus the organism suffers from the colon cancer. However, the significant differences between the networks for control and experimental groups mentioned above are the causes or the consequences of the organism suffering from the colon cancer, which will also need to be verified by the clinical experiments and annlysis of biomedical researchers.

Table 5 shows the commander genes of control and three experimental groups. We can find that: three genes PLAG1, VAV1 and ZAP70 appearing in control group are also active in the experimental groups, and the three genes might regulate the basic survival activities of human. The material shows: gene VAV1 as a kind of signal transduction molecule can happen quickly tyrosine phosphorylation with IFN-α, IL-3, GM-CSF, growth factor or antigen receptors after the activation and plays an important role in the process of cell differentiation, proliferation[21]. Gene ZAP70 plays a very important role of transfer signal in the process of activation of T-cell mediated by CD3ζ and/or CD3 chain [21]. Other genes FASLG, AXL, ETV3, GLI3, IL1A, RBL1, TAL1, TCL1A, BAG2 of control group do not appear in the experimental groups and they might be the suppressor genes of colon cancer in human genome. Furthermore, genes ARAF，BAK1，PTK2B，FGF2，IHH，STK11，HPSE appear in at least two stages of the

experimental groups and not in control group, thus the seven genes might be the oncogenes of colon cancer. The facticity and reliability predicted in the above will also need to be verified by biological experiments.

In the process of the dynamic analysis of logical network of the commander genes, we find from the numerical experiments that the number of attractors of control and three experimental groups has obvious difference (Table 6). The former is far less than the latter. According to the principle and theory of dynamics: the less the number of attractors is, the more stable the system is. From the view of biology, the construction should be more stable in the normal organism than the stages of colon cancer. The views from two points are just coincident. Thus, we conclude that the instability of the commander genes of organisms might be an important reason or result in the process of the colon cancer disease.

We utilize LAPP to search 1-, 2- and 3-order logical relationships and logic types among the genes or proteins. Using LAPP to determine the logical relationships of genes and proteins, there might appear 'false-positive' and 'false-negative'. That is to say there exist some errors during determining logical relationships. We do not analyze the errors caused by LAPP in this paper. As to how to overcome the shortcomings of LAPP method and seek the better algorithm of U value, we are making further study.

References

[1] Lander E S, Weinberg R A. Genomics: journey to the center of biology. *Science*, 2000, 287(5459): 1777-1782.

[2] Zhenfang Liao, Dean Lei, Wei Shen. The research progress of genes related to the metastasis of colon cancer. *Modern Oncology*, 2006, 14(9): 1174-1176.

[3] Zhang H, Yu C Y, Si Nger B, et al.. Recursive Partioning for tumor classification with gene expression microarray data. *PNAS*, 2001, 98: 6730-6735.

[4] Xia Li, Shaoqi Rao, Tianwen Zhang, Zheng Guo, Qingpu Zhang, K. L. Moser, E. J. Topol. Integrated decision method of mining genes related to complex disease appling the DNA chip data. *Science in China, Ser. C-life Science*, 2004, 34(2):195-202.

[5] Tibsh irani R, Ha s ti e T, Narasim han B, et al.. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *PNAS*, 2002, 99(10): 6567-6572.

[6] Khan, J.,Wei, J., Ringner, M., Saal, L., Ladanyi, M.,Westermann, F., Berthold, F., Schwab, M., Antonescu, C., Peterson, C., et al.. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat. Med.*, 2001, 7(6): 673-679.

[7] ALON U, BARKA IN, NOTTERMAN D A, et al.. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *PNAS*, 1999, 96: 6745-6750.

[8] Huang DS, Zheng CH. Independent component analysis-based penalized discriminant method for tumor classification using gene expression data. *Bioinformatics*, 2006, 22 (15): 1855-1862.

[9] Ambroise, C, McLachlan, G J.. Selection bias in gene extraction on the basis of microarray gene-expression data. *PNAS*, 2002, 99: 6562-6566.

[10] Hastie T, Tibshirani R, Eisen M B, et al.. 'Gene shaving' as a method for identifying distinct sets of genes with similar expression patterns. *Genome Biol.*, 2000, 1: RESEARCH0003.1-21.

[11] Li Xia, Rao Shao qi, Wang Yadong, et al.. Gene mining: a novel and powerful ensemble decision approach to hunting for disease genes using microarray expression profiling. *Nucleic Acids Research*, 2004, 32 (9): 2685-2694.

[12] Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, Eisenberg D. Detecting protein

function and protein–protein interactions from genome sequences. *Science*, 1999, 285: 751-753.

[13] Strong M, Graeber TG, Beeby M, Pellegrini M, Thompson MJ, Yeates TO, Eisenberg D. Visualization and interpretation of protein networks in Mycobacterium tuberculosis based on hierarchical clustering of genome-wide functional linkage maps. *Nucleic Acids Res.*, 2003, 31: 7099-7109.

[14] Zhou X, Kao MC, Wong WH. Transitive functional annotation by shortest-path analysis of gene expression data. *PNAS*, 2002, 102(38): 12783-12788.

[15] Bowers P M, Cokus S J, Eisenberg D, et a1.. Use of logic relationships to decipher protein network organization. *Science*, 2004, 306(5706): 2246-2249.

[16] Bowers P M, O'Connor B D, Cokus S J, et al.. Utilizing logical relationships in genomic data to decipher cellular processes. *The FEBS Journal*, 2005, 272 (20): 5110-5118.

[17] Xin Zhang, Seungchan Kim, Tie Wang and Chitta Baral. Joint learning of logic relationships for studying protein function using phylogenetic profiles and the Rosetta Stone method. *IEEE Transactions on Signal Processing*, 2006, 54 (6): 2427-2435.

[18] Zhang J X, Ji Y, Zhang L. Extracting three-way gene interactions from microarray data. *Bioinformatics*, 2007, 23(21): 2903-2909.

[19] Shudong Wang,Yan Chen, Qingyun Wang, Eryan Li, Yansen Su, Dazhi Meng. Analysis for gene networks based on logical relationships. *Journal of Systems Science and Complexity*, 2010, 23(5): 999-1011.

[20] Qingyun Wang, Hui Yi, Laifu Liu, Dazhi Meng. Progress in gene logic networks. *Progress in biochemistry and biophysics*, 2008, 35(11): 1239-1246.

[21] Lǔsheng Si, Xu Li. *Oncogene, cancer suppressor gene, cancer-related gene.* Shanxi Technology and Science Press, Xi'an, 2002.12.

[22] Milo R, Shen Orr S S, Itzkovitz S, et al.. Network motifs: simple building blocks of complex network. *Science*, 2002, 298: 824-827.

[23] Barábasi A L, Oltvai Z N. Network biology: understanding the cell's functional organization. *Nature Reviews-Genetics*, 2004, 5: 101-114.

[24] Fuchu He, Pengyuan Yang, Yunping Zhu. *Systems biology in practice: concepts, implementation and application.* Fudan Unversity Press, Shanghai, 2007.12.