Structural Characteristics of Gene Networks for Colon Cancer

Shudong Wang, Kaikai Li, Xiuzhu Xu

College of Information Science and Engineering

Shandong University of Science and Technology

Qingdao, China

wangshd2008@yahoo.com.cn

Dashun Xu

Department of Mathematics

Southern Illinois University Carbondale

Makanida, USA

Dazhi Meng

College of Applied Science

Beijing University of Technology

Beijing, China

***Abstract*-Genome forms gene networks in terms of complicated interactions to realize its functions. It is believed that the structure of a network dictates its functions. The comparison of the structures of gene networks for tissues with and without cancer can provide better understanding of formation and development of cancer on the molecular level. In this work, mutual information networks of cancer-related genes are constructed based on their expression profiles in colon tissues with and without cancer. The comparison and analysis of these network structures show that characteristic parameters, average degree, maximum ratio of community and modularity can distinguish the networks corresponding to normal tissues and diseased tissues. A method to find key genes is proposed that may have significant impacts on the formation of cancer from the network structure point of view. Here, eleven key genes are given. Our literature review shows that eight of these genes are closely related to the formation and development of colon cancer, leaving the other three genes open. Furthermore, we predict thirteen more genes that may play important roles in the formation of colon cancer according to the change of genes' degrees. These predictions call for empirical studies on these genes.**

Keywords-Systems Biology; Gene Network; Mutual Information; Network Statistic

I. INTRODUCTION

The invasion and metastasis of cancer cells could influence prognosis and even be fatal to patients. In recent years, the study of etiology and pathogenesis of cancers shows that the formation and development of cancers involve complicated multi-step and multi-stage processes, including dynamic expressions of interacting genes on the molecular level. The canceration of tissue cells experiences three stages: initiation, development and diffusion of cancer cells, each of which involves activation of oncogenes and inactivation of cancer suppressor genes. Therefore, finding genes critical to the formation and development of a disease from potentially disease-related genes is of significance to the diagnosis and cure of the disease and drug design. This is an important component in the research of bioinformatics[1-2]. In recent years, scientists have developed many methods to recognize genes closely related to cancers[3-13]. For example, hierarchical clustering, recursive partitioning tree, integrated decision tree, pattern recognition, high-dimensional vector analysis, etc. Alon and his colleagues[3] carried out a cluster analysis on gene expression data in colon cancer cells through the method of *t*-statistics and obtained a correspondence between some gene expression profiles and tumors. Zhang *et al*.[4] applied recursive partitioning tree method to DNA expression profiles of colon cancer and suggested that different genes may coregulate colon cancers. Tibshirani *et al*.[5] devised an approach of nearest shrunken centroid and applied it to finding genes classifying small round blue cell tumors and leukemias. Li *et al*.[6] mined the genes associated with characteristics of colon cancer by the method of integrated decision-making. These studies illustrate that analysis of gene expression profiles can decipher the formation and development of a disease.

In recent years, as a mathematical model of a complex system, complex network has been studied extensively and progress has been made in both theoretical study and applications. There are several statistics[14] that are often used to capture structure characteristics of a network. For example, average degree describes the richness of network edges. Clustering coefficient reflects the extent to which network nodes are in cluster. Degree distribution depicts the heterogeneity of the degree in the network. From the point of view of systems biology, genome, as a large regulatory system, performs its functions by forming a genetic regulatory network. Now, the methods of modeling and analysis of complex network have been broadly applied to the study of biological regulatory systems (e.g., genetic and/or protein regulatory systems, signal transduction systems and metabolism systems. The study of the structures and dynamic behaviors of these networks provides new and better understanding of the corresponding biological systems and biologically relevant predictions[13, 14].

In this research, we construct mutual information networks using gene expression profiles for cancer-related genes in

normal colon tissues and tissues with cancer in stages A, B and C. It is believed that cancer-related genes could perform completely different molecular roles as organism tissues in different situations (with or without a disease, or even with a disease in different disease stages) and hence these differences should be able to be represented and even identifiable by certain statistics that capture the structure characteristics of the corresponding networks. From the analysis of several statistics(average degree, average path length, clustering coefficient, modularity and maximum ratio of community, and the number of non-isolated nodes) for these networks, it follows that modularity and maximum ratio of community can distinguish the networks corresponding only to normal tissues and diseased tissues while average degree can distinguish the networks for all four cases in a broad range of the threshold parameter. Eleven genes are found to contribute the most to the variation of average degree in these four networks. Nine of these genes are predicted to be oncogenes of colon cancer, leaving the others cancer suppressors. Our literature review indicates that eight of these key genes are directly correlated to colon cancer, leaving the other three genes open. These findings call for more empirical study and also might be helpful for clinical diagnosis, treatment and drug design for the disease.

Next, a simple introduction to some reverse modeling methods and structural statistics used in this article will be given. The sources of raw gene expression data and the manipulations of the data are presented in Section 3. Section 4 shows the mutual information gene networks constructed from the processed databases and our main results derived from these networks. A conclusion and discussion section comes to the end of the paper with some open problems.

II. REVERSE NETWORK MODELING AND STRUCTURAL PARAMETERS OF NETWORK

To build a network model for a biological system and make biologically relevant predictions on the functions of the system, it is necessary to identify the system's structure. For example, in order to describe regulatory relationships in a set of genes, we have to identify all relevant components, their functions, their interrelationships and values of all relevant parameters. Now, there are two ways to build a network model for a biological system, forward and reverse network modelings. The former is based on studies of individual components of the system and, hence, is appropriate to a system for which there exist lots of experiment data in literature. However, reverse network modeling is based on the analysis of high-throughput data (from DNA chips and some other recently developed techniques) to mine regulatory mechanisms among the components of the system. Recently, reverse network modeling has been broadly applied to the studies of various biological systems. Werhli *et al*.[15] built Raf signal transduction network based on phylogenetic expression profiles through the methods of Gaussian model and Bayesian model. Using the method of reverse network modeling, Perkins *et al*.[16] investigated gene regulatory network of drosophila and explained the activation of genes. Wang *et al*.[17] constructed gene logical network for Arabidopsis subject to external stimuli, and further analyzed dynamical behaviors of the logical networks.

In this work, we consider regulatory networks of cancer-related genes in colon tissues with and without cancer cells. Through reconstructing the corresponding networks and studying the characteristics of network structures, we hope to better understand and find some unknown relevant functions and regulatory mechanisms of the formation and pejoration of colon cancer on the molecular level.

Now, let us briefly introduce some statistics used in this work (see [14], [18] for detailed explanations). Set $G=(V,E)$ be a complex network with node-set $V=\{1,2,\cdots,N\}$ and edge-set E.

1)Average path length(L): The distance between nodes i and j is defined as the smallest number of edges that connect nodes i to j, denoted by d_{ij}. Average path length of network is defined as $L=\dfrac{1}{0.5(N-1)}\sum_{i>j}d_{ij}$.

2) Average clustering coefficient (C): The clustering coefficient of node i, denoted by C_i, is equal to the proportion of the edges among its adjacent nodes in the possible edges. The average clustering coefficient is the average value of clustering coefficient of all the nodes.

*3) Average degree (K):*The degree of a node is the total number of nodes adjacent to it. The average degree is the average value of degrees of all nodes.

4) Modularity(Q): In a complex network, the concept of community is a good tool to describe network structures and provides better understanding of network functions. Newman[18] proposed modularity to measure the probability of a network having communities. Suppose that network G contains k communities $G_1,G_2,\cdots,G_k$. Define symmetric matrix $H=\left(h_{ij}\right)_{k\times k}$, where the element h_{ij} represents the ratio of the number of edges between two communities G_i,G_j to the total number of edges of the network. The modularity is defined as: $Q=\sum_i Q_i=\sum_i\left(h_{ii}-a_i^2\right)$, where a_i is the sum of all elements in the i^{th} row of H, representing the ratio of the number of edges adjacent to community G_i and that of all edges.

*5) Maximum ratio of community (M):*Suppose that a complex network contains m communities $G_1,G_2,\cdots,G_m$. The ratio, M_i, of the number of the interior edges of community G_i to that of the edges adjacent to G_i is called the interior-exterior ratio of community G_i. The maximum ratio $M=\max\{M_i\,|\,i=1,2,\cdots,m\}$, abbreviated as the maximum ratio of community here. If $M_i>1$, community G_i is a true community and may be responsible to certain function of the focal system. The community with the maximum ratio M as a component of the network is critical to network functions.

In this work, we study the structure characteristics of networks consisting of cancer-related genes. A gene

expression profile is a vector whose components are its expressions in many different experiments. The mutual information of genes A and B means that of their expression profiles. The idea of mutual information stems from information theory. It measures dependence degree of two stochastic variables. Let A and B be two genes (regarded as two stochastic variables). Their mutual information $I(A;B)$ is given by $I(A;B)=H(A)+H(B)-H(A,B)$, where $H(X)=-\sum_{x\in X} p(x)\log_2 p(x)$ is the Shannon entropy of vector X. $H(A,B)$ is the joint entropy of genes A and B. The larger the value of $I(A;B)$ is, the closer the interrelation between genes' is.

III. DATA SOURCE AND PROCESSING

A .Data Source

The gene expression data we work on are all from GPL570 in NCBI. For description convenience, we refer to the sample data sets for stages A, B and C of colon cancer as experimental groups and that for normal tissues as control group. The networks corresponding to experimental and control groups are called diseased networks and normal network respectively. The sample data sets for stage A, B and C includes 39, 103, 92 samples respectively. These data are all from GSE2109. Our data set for control group contains 10 samples in GDS2609, 11 samples in GSE10715 and 32 samples of GSE8671. Each of these data sets includes p_values and P-M-A (P, A and M stand for presence, absence and margin respectively) for 20827 genes, corresponding to 54676 probes. We represent P with 1, A and M with 0 in the data sets. Our work is based on these four databases (corresponding to normal stage, stages A, B and C)[*].

B. Selection of Cancer-Related Genes

It is out of our scope to construct and analyze the mutual information networks for all genes in the databases (each of the four databases contains expressions of more than 20,000 genes). Our main focus is on cancer-related genes[19]. From the databases we chose 286 genes that are known to be linked with cancers. These genes correspond to 801 probes. In the cases where several probes correspond to one gene, the highest expression value is chosen to form the gene's expression profile. We are interested in the structures of gene networks. Genes with expression profiles consisting of almost completely 0 or 1 contribute little to the structure of a network. In our study, if less than 15% or more than 90% of the total components of a gene's profile are equal to 1, we exclude the gene from our databases. There are 91 genes remaining in the database for control group, 79 in the database for stage A of colon cancer, 70 for stage B and 60 for stage C. Therefore, we focus on mutual information networks of these remaining genes in each database[†].

C. Data Processing

To calculate mutual information between genes, we discretize p_values in each database as follows. (i) Select the interval [Min, Max] for p_values and divide it into 20 portions such that each portion contains almost the same number of p_values. Rank the expression values as 1^{st}, 2^{nd}, … , 20^{th}. (ii) Replace the p-values in an interval by its labeling value. Obviously, the granularity of our discretization is finer than that of 0-1 discretization and hence our discretization loses less information than that contained in the 0-1 discretization. Therefore, it is reasonable to believe that the mutual information networks based on our finer discretization better reflect the nature of the gene regulatory system.

D. Numerical Experiments, Methods and Results

1) Comparison of network statistics

For each of the discretizated databases, we can calculate mutual information values and hence obtain a complete network of all genes in the database with mutual information values as edge weights. Note that the ranges for mutual information values for our four databases are different. For the purpose of comparison of networks, we normalize the mutual information values for each database as $x'=\frac{x-min}{max-min}$, where x and x' represent the original and normalized mutual information values, max and min are the maximum and the minimum values of the mutual information values respectively.

In order to highlight the structural characteristics of the networks so that valuable biological conclusions can be drawn, it is necessary to choose a threshold value to carry out coarse graining on normalized mutual information. For a chosen threshold value, we can build four gene networks (with normalized mutual information values as edge weights), corresponding to the cases of control group, stages A, B and C of colon cancer. In this work, we focus on the influence of structural differences of networks on their functional differences, and the statistics of a network can describe the structure of mutual information networks. Therefore, for each network derived from a threshold value, we compute six statistics: average degree K, average path length L, maximum ratio M of community, modularity Q, average clustering coefficient C and the number R of non-isolated nodes. These statistics are plotted versus the threshold values in Fig. 1. Comparing these statistics of the four networks, one can see that M, Q and R can distinguish normal and diseased networks clearly in a broad range of the threshold variation while there exist little changes in these statistics for stages A, B and C. However, the differences among average degrees K for control group and different disease stages are pretty clear. Hence, parameters M, Q and R can be used to describe the difference of normal and diseased networks, while K could be an index to identify the networks for control group and disease stages A, B and C. Fig. 2 illustrates the four networks where the threshold is 0.45.

---

[*] http://cise.sdkd.net.cn/institute/isbbc/data/SCGN/database1.rar

[†] http://cise.sdkd.net.cn/institute/isbbc/data/SCGN/database2.rar

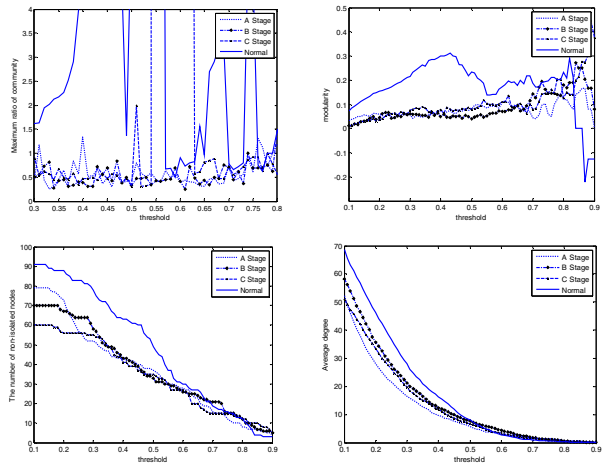
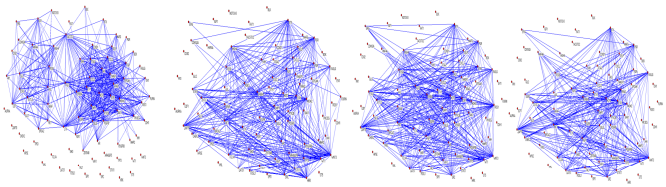


Fig. 1. Plots of four statistics M, Q, R and K versus threshold values.



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

Fig. 2. The mutual information networks of eighty two cancer-related genes in our four databases corresponding to the cases of control group and disease stages A, B and C. Obviously, some isolated nodes (genes) in the network for control group (a) are not isolated in the cases of experimental groups (b, c, d) while some other isolated genes in the cases of experimental groups (b, c, d) are not isolated in control group (a), and the isolated nodes in (b), (c), (d) are almost same. Here, the threshold value to construct the networks is 0.45.

2) Selection of structural key genes

From the above observation, we conclude that for a fixed threshold, the variation pattern of average degree K in the mutual information networks for the cases of normal stage, stages A, B and C represents the variation pattern of the structure of the genetic regulatory system. The variation pattern of K is called the database pattern. Note that each node (gene) of a network has a degree. The variation pattern of a gene's degree in the networks for the cases of normal stage, stage A, B and C is called the variation pattern of the gene. The average degree of a network describes the abundance of network edges that together with nodes form the network structure. From the point of view that the structure of a system dictates its functions, changes in the structure may cause changes in its functions. Therefore, we compare the variation patterns of genes with the database pattern and identify the genes, called structural key genes (abbreviated as SKGs), which significantly contribute to the database pattern. A SKG may contribute to the database pattern either positively, implying that the variation pattern of the gene is the same as the database pattern, or negatively, implying that the gene's variation pattern is the reversed with the database pattern. For the networks shown in Fig. 2, the variation of average degree K (i.e. the database pattern) in the four networks is as follows, average degree for normal stage>average degree for stage B>average degree for stage C>average degree for stage A. There are nine SKGs positively contributing to the database pattern and two SKGs contributing negatively. These SKGs are listed in Table 1.

TABLE I LIST OF SKGS

SKG	Normal	A	B	C	Relationship with database pattern
ALK	29	19	22	0	same
AR	16	15	18	14	same
ERG	23	2	8	4	same
ESR1	30	22	30	27	same
FGF2	36	10	26	14	same
GLI2	29	0	4	0	same
ZAP70	25	23	28	26	same
TP63	35	8	10	0	same
LATS1	19	10	22	20	same
PDGFB	0	7	0	1	reversed
STK11	0	14	1	10	reversed

IV. CONCLUSIONS AND DISCUSSIONS

Based on the expression of cancer-related genes in normal tissues and diseased tissues in stages A, B and C of colon cancer, we construct the corresponding networks of mutual information. From the comparison of network structures, we observe that the characteristic parameters of network structure: average degree, maximum ratio of community, modularity and the number of non-isolated nodes can distinguish these networks. Furthermore, we find eleven SKGs that significantly contribute to the database pattern for average degree K (see Table 1). Based on the idea that the structure of a system dictates its functions, we predict these genes are closely related to the formation and development of colon cancer. Nine of the eleven genes having the same variation pattern as the database pattern play a stimulating role in the processes of formation, development and even the pejoration of colon cancer on the molecular level. Their functions and/or involved bioprocesses may be inducing appearance and division of cancer cells and hence these genes are possibly the oncogenes of colon cancer. On the contrary, the other two genes (PDGFB and STK11) having the reversed database pattern might inhibit formation and development of colon cancer and hence possibly are cancer suppressor genes. Empirical studies on colon cancer show that eight genes (ALK, AR, ESR1, FGF2, GLI2, TP63, PDGFB, STK11) are closely related to the formation and development of colon cancer [20-26]. For example, ALK hydrolyzes the substance in the intestine and the sphingomyelin in epithelial cells of the intestinal villus and indirectly regulates cell proliferation, differentiation and apoptosis by regulating the production of Cer, sphingosine and 1-phosphate sphingosine. In ulcerative colitis and colon tumor, if the enzyme activity of ALK is down-regulated, then some mutations occur to gene ALK-Smase. The concentration of AR in cells is believed to be correlated to the degree of differentiation of a tumor and the formation of colon cancer. Gene STK11 is a cancer suppressor gene and plays a role in the formation and development of sporadic colorectal cancer. Our literature review does not show that the other three genes (ERG, ZAP70, LATS1) are directly related to colon cancer. In fact, In tumors like colorectal cancer, cutaneous angiosarcoma

and hepatocellular carcinoma, the expression of ETS1 can be found in stromal cells of tumors that have been infiltrated and the expression is going to be up-regulated with the invasion and metastasis. And ERG is a subfamily of ETS transcription factor family and ETS is correlated with transcriptional activation of enzymes that degrade the extracellular matrix in the process of metastasis and angiogenesis of tumor [27, 28]. Based on our analysis and the facts mentioned above for these genes, we predict that these genes are oncogenes of colon cancer. Our predictions call for more empirical studies on the functions of these genes in colon cancer.

Let us look at the networks shown in Fig. 2 again. The degrees of some nodes (genes) in normal network are far greater than those in diseased networks. For example, the degree of gene BLK in normal network is 37, while it is 0 in diseased networks. That is, these genes are very active in normal cells but exceptionally silent in cancer cells and, hence, they are probably cancer suppressor genes. However, the degrees of some other genes in normal network are far less than those in diseased networks. For example, the degree of gene ERBB4 in normal network is 0, while they are 25, 21 and 16 in the networks for disease stage A, B and C respectively. Hence they might be oncogenes of colon cancer. Nineteen of these genes together with their degrees are listed in Table 2. It has been shown that six (IGF1, CD82, PDGFRB, PIK3CG, ERBB4, WT1) of the nineteen genes in Table 2 are closely related to the formation and development of colon cancer [29-34]. For example, CD82 is a type of membrane surface glycoprotein and plays an inhibiting role in the metastasis of colon cancer, affecting cancer cells' invasion to surrounding tissues and metastasis to lymphaden and distant organs through interactions between cells, and cells and extracellular matrix. ERBB4 is a gene regulating the growth of colon cancer. Over-expression of ERBB4 has an important relation with clinical pathology of colon cancer. Assume that the more significantly the degree of a gene varies with tissues' states, the closer its expression is related to the states. We predict that the left thirteen genes (marked with "*" in Table 2) are closely correlated with the formation of colon cancer.

TABLE II LIAST OF GENES WHOSE DEGREES VARY GREATLY BETWEEN CONTROL GROUP AND EXPERIMENTAL GROUPS

(a)

Gene	Normal	A	B	C
BLK* EMP1*	37	0	0	0
ETV1*	20	0	0	0
CXCL3*	27	1	0	0
ZBTB48*	33	0	0	0
IGF1	23	0	0	0
CD82	22	0	0	0
LTK*	26	0	0	0
EPCAM*	25	0	0	0
NOTCH3*	20	0	0	0
PDGFRB	32	0	0	0
PIK3CG	43	0	0	0
SPI1*	31	0	0	0
	40	0	0	0

(b)

Gene	Normal	A	B	C
ERBB4F	0	25	21	16
OSB*	0	13	29	10
RET*	0	23	25	29
SELE*	0	23	26	14
WNT3*	0	35	29	29
WT1	0	29	33	15

Fig. 2 illustrates the location maps of cancer-related genes in our four databases. In the right side of the location map in Fig. 2a, there exists a community in normal network (the maximum ratio of community is equal to 4.8169). The community disappears in diseased networks (the maximum ratio of community is less than 1). In the network for a complex biological system, a community of system components is often in charge of some specific biological functions. Therefore, we believe that for control group, the genes in the community of the network together might play a role of inhibiting colon cancer in an ordered manner, and correspondingly the tissue is in the normal state.

ACKNOWLEDGMENTS

The authors would like to thank the National Natural Science Foundation of China (Nos: 60874036, 60503002 and 60970065) and SDUST Research Foundation of China for the support to this work.

REFERENCES

1. Lander E S, Weinberg R A. Genomics: journey to the center of biology [J]. 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 [J]. Modern Oncology, 2006, 14(9): 1174-1176.
3. Alon U, Barkain, Notterman D A, et al. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays [J]. Proc Natl Acad Sci USA, 1999, 96: 6745-6750.
4. Zhang H, Yu C Y, Singer B, et al. Recursive partioning for tumor classification with gene expression microarray data [J]. Proc Natl Acad Sci USA, 2001, 98: 6730-6735.
5. Tibshirani R, Hastie T, Narasimhan B, et al. Diagnosis of multiple cancer types by shrunken centroids of gene expression [J]. Proc Natl Acad Sci USA, 2002, 99(10): 6567-6572.
6. 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 applyng the DNA chip data [J]. Science in China, Ser. C-life Science, 2004, 34(2):195-202.
7. Huang DS, Zheng CH. Independent component analysis-based penalized discriminant method for tumor classification using gene expression data [J]. Bioinformatics, 2006, 22 (15): 1855-1862.
8. Ambroise C, McLachlan G J. Selection bias in gene extraction on the basis of microarray gene-expression data [J]. Proc Natl Acad Sci USA, 2002, 99: 6562-6566.
9. Chow M L, Moler E J, Mian I S. Identifying marker genes in transcription profiling data using a mixture of feature relevance experts [J]. Physiol Genomics, 2001, 5: 99-111.
10. Li L, Weinberg C R, Darden T A, et al. Gene selection for sample classification based on gene expression data: study of sensitivity to choice of parameters of the GA/KNN method [J]. Bioinformatics, 2001, 17: 1131-1142.
11. Burke H B. Discovering patterns in microarray data [J]. Mol Diagn, 2000, 5(4): 349-357.
12. 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 [J]. Nucleic Acids Research, 2004, 32 (9):2685-2694.
13. L. Diambra, L. da, F. Costa. Complex networks approach to gene expression driven phenotype imaging [J]. Bioinformatics, 2005, 21(20): 3846-3851.
14. Réka Albert, Albert-László Barabási. Statistical mechanics of complex networks. Reviews of Modern Physics, 2002, 74: 47-97.
15. Adriano V. Werhli, Marco Grzegorczyk, Dirk Husmeier. Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical Gaussian models and bayesian networks [J]. Bioinformatics, 2006, 22(20):2523-2531.
16. Theodore J. Perkins, Johannes Jaeger, John Reinitz, Leon Glass. Reverse engineering the gap gene network of Drosophila melanogaster [J]. Plos Comput Biol., 2006, 2(5): 0417-0428.
17. Shudong Wang, Yan Chen, Qingyun Wang, Eryan Li, Yansen Su, Dazhi Meng. Analysis for gene networks based on logical relationships [J]. Journal of Systems Science and Complexity, 2010, 23(5): 999-1011.
18. M. E. J. Newman, M. Girvan. Finding and evaluating community structure in networks. Phys. Rev. E, 2004, 69: 026113.
19. Lŭsheng Si, Xu Li. Oncogenes, cancer suppressor genes, cancer-related genes [M]. Shanxi Technology and Science Press, Xi'an, 2002.12.
20. S joqvist U, Hertervig E, Nilsson A, et al. Chronic colitis is associated with a reduction of mucosal alkaline sphing omyelinase activity [J]. Inflamm Bowel Dis, 2002, 8(4): 258-263.
21. He Y, Zhou J, Wu JS, Dou KF. Inhibitory effects of EGFR antisense oligodeoxynucleotide in human colorectal cancer cell line [J]. World J Gastroenterol, 2000, 6:747-749.
22. Martha L. Slattery, Carol Sweeney, Maureen Murtaugh, Khe Ni Ma, Roger K. Wolff, John D. Potter, et al. Associations between ERa, ERb, and AR genotypes and colon and rectal cancer [J]. Cancer Epidemiol Biomarkers Prev, 2005, 14(12): 2936-2942.
23. Guillaume Chatel, Corine Ganeff, Naima Boussif, Laurence Delacroix, Alexandra Briquet, Gregory Nolens et al. Hedgehog signaling pathway is inactive in colorectal cancer cell lines [J]. Int J Cancer, 2007, 121(11):2622-2627.
24. Nahor I, Abramovitch S, Engeland K, Werner H. The p53-family members p63 and p73 inhibit insulin-like growth factor-I receptor Gene expression in colon cancer cells [J]. Growth Horm IGF Res, 2005, 15(6):388-396.
25. Nakamura Y, Tanaka F, Yoshikawa Y, Mimori K, Inoue H, et al. PDGF-BB is a novel prognostic factor in colorectal cancer[J]. Ann Surg Oncol, 2008, 15(8):2129-2136.
26. Nicoletta Resta, Cristiano Simone, Cristina Mareni, Mariapina Montera et al. STK11 mutations in peutz-jeghers syndrome and sporadic colon cancer [J]. Cancer Research, 1998, 58: 4799-4801.
27. Naito S, Shimizu K, Nakashima M, Nakayama T, ItoT, Ito M, Yamashita S, Sekine I. Overexpression of Ets-1 transcription factor in angiosarcoma of the skin [J]. Pathol Res Pract, 2000, 196:103-109.
28. Ozaki I, Mizuta T, Zhao G, Yotsumoto H, Hara T, Kajihara S, Hisatomi A, Sakai T, Yamamoto K. Involvement of the Ets-1 gene in overexpression of matrilysin in human hepatocellular carcinoma [J]. Cancer Res, 2000, 60: 6519-6525.
29. Ryo Fukuda, Kiichi Hirota, Fan Fan, Young Do Jung, Lee M. Ellis, Gregg L. Semenza. Insulin-like growth factor 1 induces hypoxia-inducible factor 1-mediated vascular endothelial growth factor expression, which is dependent on MAP Kinase and phosphatidylinositol 3-Kinase signaling in colon cancer cells [J]. The Journal of Biological Chemistry, 2002, 277(41): 38205- 38211.
30. Li Liu, De-Hua Wu, Zu-Guo Li, Guang-Zhi Yang, Yan-Qing Ding. Effects of KAI1/CD82 on biological behavior of human colorectal carcinoma cell line [J]. World Journal of Gastroenterology, 2003, 9(6): 1231-1236.
31. J.-P. Spano, C. Lagorce, D. Atlan, G. Milano, J. Domont, R. Benamouzig, A. Attar, et al. Impact of EGFR expression on colorectal cancer patient prognosis and survival[J]. Annals of Oncology, 2005, 16(1):102-108.
32. Kato S., Iida S., Higuchi T. PIK3CA mutation is predictive of poor survival in patients with coloretal cancer [J]. Int J Cancer, 2007, 121(8):1771-1778.
33. Yiming Yu, Zhong Liu, Zaichun Deng,Yunshan Tan, Yonghua Xu. The expression of erbB4/HER4 in colon cancer [J]. Chinese Journal of Experimental Surgery, 2008, 25(11):1452-1453.
34. Robert Koesters, Michael Linnebacher, Johannes F. Coy, et al. WT1 is a tumor-associated antigen in colon cancer that can be recognized by in Vitro stimulated cytotoxic T cells [J]. Int J Cancer, 2004, 109(3):385-392.