Comparison of Hubs in Effective Normal and Tumor Protein Interaction Networks

Mitra Mirzarezaee[1], Babak N. Araabi[2,*], Mehdi Sadeghi[3]

1. Department of Computer Engineering, Islamic Azad University, Science and Research Branch, Tehran, Iran
2. Control and Intelligent Processing Center of Excellence, School of Electrical and Computer Engineering, University of Tehran, Iran; and School of Cognitive Sciences, IPM, Tehran, Iran
3. National Institute of Genetic Engineering and Biotechnology, Tehran, Iran; and School of Computer Sciences, IPM, Tehran, Iran

Article info:
Received: 5 September 2010
First Revision: 10 October 2010
Accepted: 30 October 2010

Key Words:
Cancer,
Protein Interaction,
Network.

ABSTRACT

Introduction: Cancer is caused by genetic abnormalities, such as mutation of ontogenesis or tumor suppressor genes which alter downstream signaling pathways and protein-protein interactions. Comparison of protein interactions in cancerous and normal cells can be of help in mechanisms of disease diagnoses and treatments.

Methods: We constructed protein interaction networks of cancerous and normal cells. These protein interaction networks are constructed using gene-expression profiles measured from different samples of cancerous and normal tissues from four different parts of the body including colon, prostate, lung, and central nervous system. We used pattern recognition techniques to construct these networks. We calculated ten graph related parameters including closeness centrality, graph diameter, index of aggregation, entropy of edge distribution, connectivity, number of edges divided by the number of vertices, entropy, graph centrality, sum of the wiener number, and modified vertex distance numbers for each of the cancerous and normal protein interaction networks. We have also compared number of edges and hubs of the both cancerous and normal resultant protein interaction networks.

Results and Discussion: Our results show that in the studied tissue samples, effective normal protein interaction networks are denser in number of edges and hubs compared with their corresponding effective cancerous protein interaction networks. Number of hubs in effective cancerous protein interaction networks decreases dramatically in comparison with normal tissues. This can be used as a symptom for identification of cancerous tissues.

1. Introduction

Studying the complete set of molecular interactions provides many insights to the regulation of normal and cancerous cells. In tumor progress which is a multi-step process, a normal cell transforms gradually to a malignant one. Cancer is caused by genetic mutations, translocations, amplifications, deletions and viral gene insertions that changes translated proteins. This can affect and disrupt signaling pathways and protein-protein interactions that are necessary for cellular processes.

**** Corresponding Author:***
Mitra Mirzarezaee, PhD
Department of Computer Engineering, Islamic Azad University, Science and Research Branch, Tehran, Iran
Email : mirzarezaee@acm.org

Advances in gene chip technologies have led to quantizing and monitoring the expressions of various genes (Gerhold et al,. 1999). Gene expression profiling is an important tool for diagnosis and classification of diseases. These gene expressions are widely used for identifying genes responsible for various conditions and cancers (Alon U., et al., 1999), (Notterman D.A., et al., 2001), (Golub T.R., et al., 1999), (Perou C.M., et al., 1999), (Perou C.M., et al., 2000). This is done using specialized clustering techniques (Brazma A. & vilo J., 2000), (Eisen M.B., et al., 1998), (Blatt M., et al., 1998). Gene expression profiles can also be used for constructing co- expression gene networks. Since proteins are the end products of genes, various types of protein and gene networks (Maslov S. & Snappen K., 2002) are directly related.

Many research groups have utilized network analysis in gene data sets of cancer. Jonsson and Bates (Jonsson, Bates, 2006) reported that proteins associated with cancer show an increased number of interacting partners, which reflects their increased centrality in the Protein Interaction Network (PIN). The role of genes' molecular interactions differentially regulated in lung cancer has been investigated by Wachi et. al(Wachi et al. 2005). They find increased connectivity among genes, which justifies the findings of Jonsson (Jonsson, Bates, 2006). Platzer and his colleges (Platzer et al., 2007) have studied 22 different graphs related factors for 29 tumor associated gene expression data sets. In the work the differences of the cancerous networks with random networks using their gene-expression and predicted human protein interactions has been studied. They have found that the prevalence of hub proteins was not increased in the presence of the cancer.

In this paper, we analyse the networks of gene co-expressions of colon, prostate, lung, and central nervous system for both the normal and cancerous tissues based on the changes in the most important nodes of the networks called hubs. Hubs play essential roles in the networks. They are the proteins with more than eight interactions in the PINs. Hubs are engaged in many functions of the cell and their disruption results in cell malfunctions. In this study we are interested to know how hubs of PINs are changed from a normal to a cancerous tissue.

This paper is organized as follows: the materials that are used for this study and the proposed network construction and analysis methods in both the normal and cancerous tissues are explained briefly in the next section. In the Results and Discussion section, this study's results has been represented and finally the paper is concluded.

2. Methods

We studied the differences of hubs in both normal and tumor gene expression networks of four different tissues of the body including colon, prostate, lung, and central nervous system.

2.1. Gene Expression Profiles

Gene expression profiles from different samples of normal and cancerous tissues are downloaded from Kent Ridge Dataset (http://datam.i2r.a-star.edu.sg/datasets/krbd/). From that database, we exclude datasets that containing different stages of cancer or the ones which have preconditions. Specifications of data sets which are used in this study are shown in Table 1. As it can be seen in Table 1, number of studied genes in each case differs. That is because just effective genes in each cancer type are being studied.

These expression data is available in the form of a matrix having N rows and D columns. The Columns represent the tissue in special condition and the rows represent the gene profiles. The gene expression profiles are normalized in a z-score fashion such that the average expression ratio of one profile is 0 and the standard deviation is 1.

2.2. Network Construction Method

Using the gene expression data, we constructed a sparse co-expression network using k mutual nearest neighbor criterion (Agrawal et al., 2002). In this gene co-expression network construction method as explained by Agrawal et. al (Agrawal et al., 2002) for every gene expression profile a list of k nearest neighbor profiles is produced. The nearest neighbor of one expression profile is defined as the most similar profile measured. Different similarity measures are introduced in pattern recognition books. In this study, we used the Euclidean distance measure. The Euclidean similarity measure in an n dimensional space is defined as follows:

$$d(x,y)=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2} \qquad (1),$$

Where xi, and yi are the corresponding x and y values in an n dimentional space. In this case, n is the number of samples from cancerous or normal tissues.

This way, a list of k nearest neighbours of each gene (protein) of the network are calculated. Two different nodes in gene co-expression networks are connected if there are on each others' list. The optimal k is assumed 15, as discussed by Agrawal and Domany (Agrawal et. al, 2003).

2.3. Graph Related Parameters

We calculated ten different graph related parameters that Platzer (Platzer et al., 2007) introduced in his paper including closeness, graph diameter, index of aggregation, entropy of edge distribution, connectivity, number of edges divided by the number of vertices, entropy, graph centrality, sum of the wiener number, and modified vertex distance number. We have calculated the above parameters to analyse their values and differences in different cancerous and normal PINs.

2.4. Number of Edges and Hubs

We counted the edges and hubs of both the normal and cancerous tissues. Hubs are defined as the nodes with more than eight connections in the network. According to the network construction methods we have used, the maximum number of interaction a gene/protein has, is at most 15. So the nodes with less than or equal to eight interactions are assumed to be as non-hubs and the nodes with greater than eight connections are identified as hubs. We have also determined the common hubs of the normal and cancerous tissues to see if hubs are changing significantly from normal to cancer in the studied tissues. The common hubs are the ones which are so in both normal and cancerous PINs.

3. Results

The present paper provides a systematic analysis of protein (gene) networks in normal and cancerous tissues. For this purpose, we have extracted gene-expression data of normal and cancerous tissues from the free online Kent Ridge database. The data downloaded are for four tissues of the body including colon, prostate, lung, and central nervous system for both normal and cancerous tissues. There exist other data sets as for ovarian. But because they have some preconditions on collecting the data from samples that have the possibility of getting ovarian cancer by inheritance, we did not use those data sets. The samples from stages of the cancer were not used here, also.

In the data sets available from Kent Ridge data base, number of studied genes from one tissue to another number of differs. That is because from all the available genes, they have selected the effective genes to the type of studied cancer. Therefore we have constructed the Protein Interaction Networks (PINs) of the effective genes and their variants from normal to cancerous tissues has been studied.

We have constructed the effective PINs for both normal and cancerous tissues as explained in materials and methods. The effective protein (gene) interaction networks of normal and cancerous tissue for central nervous system, using network construction, is shown in Fig. 1.

Although the numbers of available samples for each selected normal and cancerous tissues are not equal, it does not affect the networks having been constructed because the network construction method explained in the materials and methods section is independent of the number of samples used.

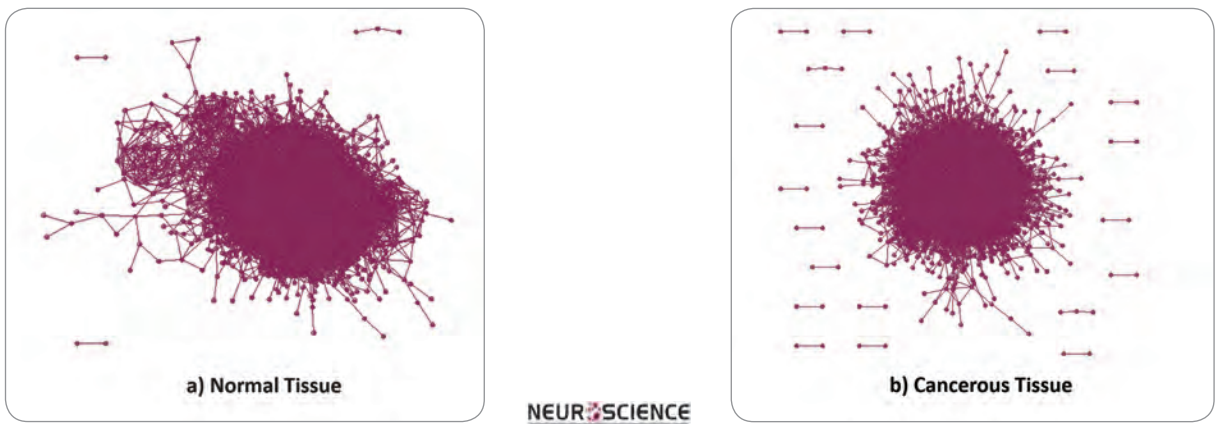

Fig. 1. Protein/Gene Interaction Networks of Central Nervous System
a) Normal Tissue b) Cancerous Tissue

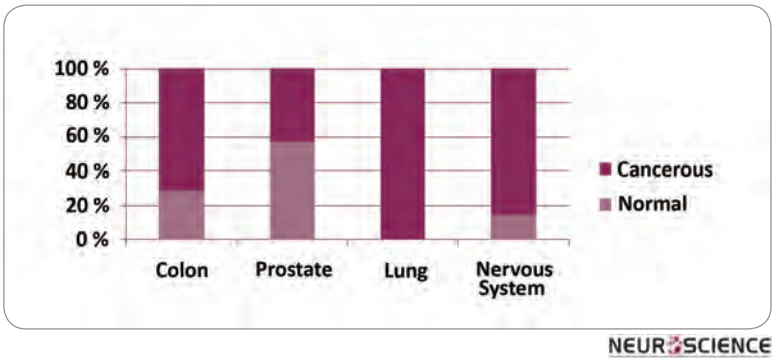

Fig. 2. Closeness Parameter in Studied Normal and Cancerous Tissues

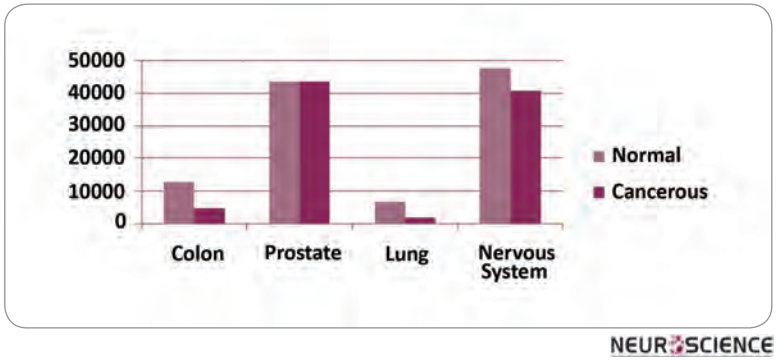

Fig. 3. Number of edges in Studied Normal and Cancerous Tissues

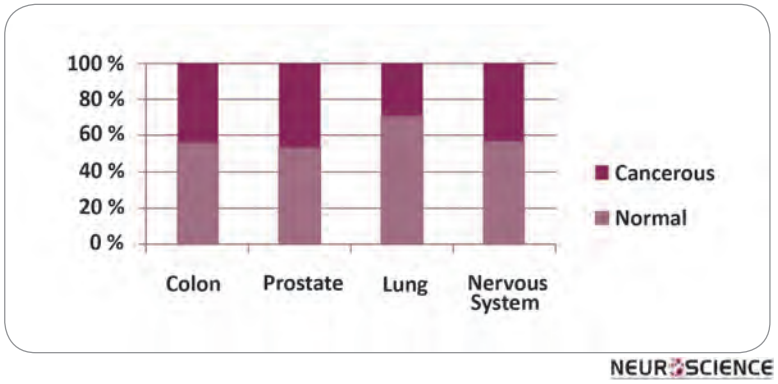

Fig. 4. Hubs in Studied Normal and Cancerous Tissues

We have calculated ten graph related parameters of closeness, graph diameter, index of aggregation, entropy of distribution of edges, connectivity, number of edges divided by the number of vertices, entropy, graph centrality, sum of the wiener number, and modified vertex distance number to see how they change in a normal and cancerous PIN.

As the results in Table 2 show, none of the parameters have significant differences in the same normal and cancerous tissues except for closeness. Closeness parameter values for both normal and cancerous tissues are shown in Fig. 2 in percentage. Closeness parameter changes in different studied tissues. In some of the tissues this value is higher in the normal cell While in other ones it's high for cancerous cell. The point is value changes significantly from a normal tissue to a cancerous one and if such a difference in value can be found in two tissues of the same kind that can be a symptom of a cancer. If this value is not known for the normal tissue, identification of cancerous tissues would not be easy using this parameter.

Table 1. Types of tissues and number of samples

Tissue	Number of samples Normal/Cancer	Number of Genes
Colon	22/40	2000
Prostate	54/65	12600
Lung	10/86	7129
Nervous System	21/39	7129

NEUROSCIENCE

Table 1. Graph Related Parameters calculated for both the Normal and Cancerous Tissues

Parameter \ Tissue Name	Colon		Prostate		Lung		Central Nervous System	
	N*	C+	N	C	N	C	N	C
Closeness Centrality (CC)	15.62	39.29	756.35	575.56	2.13	12.78	2.13	12.78
Graph Diameter (GD)	0.0077	0.0117	0.0035	0.0032	0.0027	0.0023	0.0027	0.0023
Index of Aggregation (IoA)	0.98	0.94	0.77	0.83	0.98	0.96	0.98	0.96
Entropy of edge distribution	2.64	2.56	2.00	2.06	2.57	2.50	2.57	2.50
Connectivity	0.004	0.003	2.73e-4	2.79e-4	9.31e-4	7.99e-4	9.31e-4	7.99e-4
Number of edges divided by the number of vertices	3.76	3.38	2.24	2.12	2.37	2.95	3.37	2.95
Entropy	8.11e+4	6.71e+4	2.50e+5	2.54e+5	3.06e+5	2.52e+5	3.07e+5	2.51e+5
Graph centrality (GC)	0.089	0.083	0.127	0.1014	0.0755	0.0846	0.0755	0.0846
Sum of the Wiener Number	6.6e+3	6.42e+3	3.83e+4	4.54e+4	2.34e+4	2.29e+4	2.34e+4	2.30e+4
Modified Vertex Distance Number (mVD)	1.25e+5	9.56e+4	7.68e+5	1.056e+6	1.40e+6	1.22e+6	1.40e+6	1.22e+6

NEUROSCIENCE

Table 3. Edges Differences in Normal and Cancerous Tissues

Type of Tissue	Common Edges (11)	Common non-edges(00)	Edges Visited only in Normal Tissue(10)	Edges Visited only in Cancerous Tissue(01)
Colon	1048	1988364	6295	5293
Prostate	691	79343115	21012	21482
Lung	381	25357746	39938	16820
Central Nervous System	565	25371495	23080	19745

NEUROSCIENCE

The other parameters we have studied within the PINs of the normal and cancerous tissues were the number of edges and hubs. We counted number of edges in both normal and cancerous tissues. The results are shown in Fig. 3. As we have seen in four different cancer types especially in the nervous system, the number of edges is higher in an effective normal PIN in comparison with its corresponding cancerous PIN. This means some of the edges have been disappeared in the cancerous tissues and the new edges have been replaced.

Table 4. Hub differences in Normal and Cancerous Tissues

Tissue Type	No. Hubs Normal/Cancerous	Common Hubs	Exclusive Normal Hubs	Exclusive Tumor Hubs
Colon	944/744	469	475	275
Prostate	1865/1653	648	1217	1005
Lung	6671/1792	1670	5001	127
Central Nervous System	2718/2070	994	1724	1076

NEUROSCIENCE

Table 5. Mean and Variance of Common Hubs in Normal and Cancerous Tissues

Tissue Type	Mean and Var. of the differences in Common Hubs	Mean and Var. of Common Hubs in Normal Tissues	Mean and Var. of Common Hubs in Cancerous Tissues
Colon	2.3113/1.8049	2.2372/11.4563	2.3875/11.4136
Prostate	2.3009/1.7671	2.2708/11.3781	2.2630/11.0278
Lung	2.5413/1.8124	1.9915/11.7976	2.3470/10.9293
Central Nervous System	2.2223/2.8124	2.2212/10.9557	2.2943/10.7093

NEUROSCIENCE

As the collaboration of the proteins in performing a function of the cell is shown as edges between proteins involved, If some of the edges would be dismissed or replaced by the others- except for the noises that may arise from the gene expressions and network construction methods- it's a representation of a change of function in the specified cell. That may be the source of malfunction and the cause of cancer. Edge differences in these four tissues for effective normal and cancerous PINs are shown in Table 3.

We have also identified the hubs of normal and cancerous PINs. A node with more than eight interactions is called a hub. Numbers of hubs in both PINs are being counted. As it seen in Fig. 4, the percentages of hubs in normal tissues are higher than their corresponding cancerous tissues.

We have also studied the differences of normal and cancerous hubs to identify how hubs are transformed from normal situations to cancer ones. We have studied the number of common and exclusive hubs of the normal and cancerous tissues. The result of this study is shown in Table 4.

The mean and variance of changes in hub partners of common hubs from normal and cancerous tissues have also been tested, as shown in Table 5. According to the mean and variances of that data, these hubs are almost the same with the same mean and variance, which could prove them to be of a same distribution.

In this study we have tried to test if the hubs are changed from a normal to a cancerous cell and the results show that the common hubs, the ones that do not change from normal to cancerous are almost the same which could be good news. So that what caused the cell to malfunction is not related to common hubs because they do not change their functions and it can be understood that the proteins probably involved in cancer are among the other exclusive hubs of the cancer network or due to the absence of Hubs existing in the normal tissue..

4. Discussion

We have studied the differences in the edges and the hubs of effective Protein Interaction Networks (PINs) of both normal and cancerous tissues from different parts of the body including colon, prostate, lung, and central nervous system. The effective PINs are constructed using the gene expression profiles of normal and cancerous tissues. We have compared the edges and the hubs of the both PINs. As it can be inferred from the results

of this study, normal tissues of the four studied samples are denser in the effective interaction networks. That means their effective PINs contains more edges and number of hubs is higher in comparison with the cancerous networks of the same tissue. The other studied graph related parameter that shows discrimination from normal to cancerous tissue is closeness. The value of this parameter changes significantly from normal to the cancerous tissue of the same type and that may be a good parameter and an easy one to calculate for discriminating normal networks from cancerous ones.

References

Agrawal, H. and Domany, E., (2003). Potts ferromagnets on coexpressed gene networks: identifying maximally stable partitions. Phys. Rev. Lett., 90, 158, 102-106.

Agrawal, H., (2002). Extreme self-organization in networks constructed from gene expression data. Phys. Rev. Lett., 89, 268, 702-706.

Alon U., et al.,(1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc. Natl acad. Sci. USA, 96,6745-6750.

Blatt M., Wisemann S., and Domany E. (1998). Superparamagnetic Clustering of Data, Phys. Rev. Lett. 76,3251.

Brazma A. and Vilo J. (2000). Gene Expression data analysis, FEBS Lett. 480, 17.

Eisen M.B., et al. (1998). cluster analysis and display of genome wide expression patterns,Proc. Natl. Acad. Sci. U.S.A. 95, 14863-14868.

Gerhold D.,Rushmore T., and Caskey C. T., (1999). DNA chips: promising toys have become powerful tools, Trends in Biochemical Sciences, 24(5), 168-173.

Golub T.R., et al. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286, 531.

Jonsson PF, (2006). Bates PA: Global topological features of cancer proteins in the human interactome. Bioinformatics, 22:2291-2297.

Maslov S. and Sneppen K.(2002). Specificity and stability in topology of protein networks , Science 296, 910-913.

Notterman DA, Alon U, Sierk AJ, Levine AJ. (2001). Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. Cancer Res. 61, 3124-3130.

Perou C. M, et al., (2000). Nature (London) 406, 747–752, pmid:10963602.

Perou C. M., et al. (1999). Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. Proc. Natl. Acad. Sci. U.S.A. 96, 9212-9217.

Platzer A., Perco P., Lukas A., et al. (2007). Characterization of protein-interaction networks in tumors, BMC Bioinformatics, 8:224,doi:10.1186/1471-2105-8-224.

Wachi S, Yoneda K, Wu R (2005). Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. Bioinformatics, 21, 4205-4208.