Article

Changes in protein interaction networks between normal and cancer conditions: Total chaos or ordered disorder?

K. M. Taufiqur Rahman, Md. Fahmid Islam, Rajat Suvra Banik, Ummay Honi, Farhana Sharmin Diba, Sharmin Sultana Sumi, Shah Md. Tamim Kabir, Md. Shamim Akhter
Biotechnology and Genetic Engineering Discipline, Khulna University, Bangladesh
E-mail: fahmid_islam9891@yahoo.com

Received 15 November 2012; *Accepted 18 December 2012*; *Published online 1 March 2013*
IAEES

Abstract
New insights to understand the dynamics of enormous modifications during cancer in comparison to healthy condition have made the ground for the emergence of sophisticated systemic approaches like Network Systems Biology in the twenty first century which is potentially effective to model different biological phenomena such as regulation of gene-expression and protein-protein interaction. In the current study, the construction and computational analysis of protein interaction networks (PINs) based on expression data of proteins involved in 10 major cancer signal transduction pathways were done in case of five different tissues e.g. bone, breast, colon, kidney and liver for both normal and cancer conditions. Differential expression database GeneHubs-Gepis, and protein-protein interaction prediction tools PIPs and STRING were applied for primary data retrieval. Upregulation and downregulation of proteins in various cancers were analyzed to identify patterns in PINs during cancer signaling. Different network parameters were evaluated and comparisons were made among normal and cancer networks for each tissue and for different cancer based on Cytoscape software package. The networks for cancer show notable differences and fluctuations from normal ones for various network parameters. A cluster of 34 upregulated proteins with 76 relevant interactions was also found to be conserved in all five cancerous tissues.

Keywords cancer; network systems biology; signal transduction pathways; protein interaction network.

1 Introduction

Cancer being an abnormal manifestation of the inherent subtleness of biological organization can be viewed as a result of defective organogenesis which acts as an association of multiple diseases and characterized through the process of tumorigenesis (Goldthwaite, 2006; Reya et al., 2001). In the complex enigma of cancer progression cells accumulate mutations in oncogenes or tumor suppressor genes that allow chromosomal aberrations, genomic and proteomic instabilities, and ultimately result into abnormal proliferation and differentiation (Hanahan and Weinberg, 2000). Various approaches like classical clonal genetic model (Arends, 2000; Fearon and Vogelstein, 1990), epigenetic model (Esteller, 2008; Tysnes, 2010) and cancer stem cell model (Ye et al., 2008; Goll et al., 2005) have been proposed so far to understand cancer initiation and metastasis and all these models are based on local alterations of genomic and proteomic status of the cells leading to cancer conditions. But recent understandings have made it plausible that cancer might act as an

exceptionally unusual 'whole' (like organs) in the complex fractal hierarchy of 'wholeness' functioning in our body system (cell\organ\organism). Due to the massive alterations both in genome and proteome, cancer initiation is more likely to be stochastic while it demands more comprehensive systemic approach to endow the non-locality and non-linearity underlying the process of cancer development (Mamun et al., 2011).

Biological research for over the last century has been dominated by the reductionist philosophy and a wealth of knowledge has been generated about structural and functional attributes at cellular level (Kitano, 2002). Despite huge achievement of reductionism, it is gradually becoming clearer that discrete biological functions can rarely be ascribed to individual molecules. Instead, most biological properties emerge from highly interactive complexity gained from functional integrity of cell's numerous constituents (Oltvai and Barabasi, 2002). Therefore, understanding the structural and functional dynamics of the intricate web of interactions at cellular level has been a key challenge for biology in the twenty-first century (Barabasi and Oltvai, 2004).

In cancer condition, genomic alterations result in modifications in downstream signal transduction pathways and protein-protein interactions. Studying the molecular interactions entirely is a must to have an insightful understanding of the comparative regulatory patterns of normal and cancerous cells (Mirzarezaee et al., 2010) and Network Systems Biology has prospective usefulness to model various biological phenomena such as regulation of gene expression and protein-protein interaction (Zhou et al., 2012). Probably the most commonly studied type of biological networks is protein interaction networks (PINs) which can provide some more realistic interpretations about cancer complexity in terms of network properties based on the graph theoretical approaches (Platzer et al., 2007; Huang and Zhang, 2012; Zhang, 2012). The whole array of development is highly regulated with extreme sophistication through controlled proliferation while in cancer things get skewed up and genomic and proteomic instability occurs. But still tumorigenesis follows some fundamental rules of development and hence the term alternative form of life has been used to address cancer (Davies, 2004; Shrodinger, 1958). So very simply there has to be some signs or notifications representing the subtle orderliness of the highly disordered phenomenon of cancer progression. Interestingly we found a small protein interaction network (PIN) cluster which is conserved in different cancerous tissues in accordance with this current approach.

The main aim of this study was to construct and visualize differential PINs in five tissues e.g. bone, breast, colon, kidney and liver for both normal and cancer conditions based on gene expression data for the proteins involved in ten major cancer signal transduction pathways. A comparative analysis of different network parameters among normal and cancer conditions for each of the five tissues was done. Remarkable differences were observed in the network parameters among the networks for normal and cancerous tissues.

2 Materials and Methods

2.1 Construction and analysis of differential networks

The protein molecules involved in cancer signal transduction pathways were listed from Cancer Cell Map Database (http://cancer.cellmap.org/cellmap/) (Memorial Sloan-Kettering Cancer Center, 2006). The ten cancer signal transduction pathways from the database were considered e.g. Alpha-6-Beta-4-Integrin, Androgen Receptor, Kit Receptor, EGFR1, Hedgehog, Wnt, ID, NOTCH, TGFBR and TNF Alpha/NF-kB. The total number of signaling protein molecules was 737. Possible protein-protein interactions were studied for the signaling proteins via PIPs (a database human protein-protein interaction prediction; http://www.compbio.dundee.ac.uk/www-pips/) (McDowall et al., 2009; Scott and Barton, 2007) and STRING (a database of known and predicted protein interactions; http://string.embl.de/) (Auguste et al., 2007; Caldieri and Buccione, 2010). Protein-protein interaction data were available for 722 signaling proteins out of 737

molecules. Here the interactions were considered among the 722 signaling proteins, other predicted interacting proteins were excluded. Thus 609 proteins were found to show total 8359 possible interactions among them. Differential expressions of the signaling protein molecules in normal and cancer conditions for five human tissues e.g. bone, breast, colon, kidney, liver were accumulated and studied using GeneHub-Gepis (an online bioinformatics tool for inferring gene expression patterns in a large panel of normal and cancer tissues; http://research-public.gene.com/Research/genentech/genehubgepis/index.html) (Zhang et al., 2007). The expression data were represented in digital expression unit (DEU). Expression data were available for 598 proteins out of the 609 molecules and total 8245 possible interactions were found to exist among them. A PIN representing all the possible interaction among the proteins was constructed (Fig. 1) and the network properties for this network was listed (Table 1). As the expressions of different signaling proteins differentiate in normal and cancer conditions of various tissues, a fraction of the total possible interactions is manifested in different tissues with normal or cancer conditions. PINs for normal and cancer conditions of the five tissues were constructed based the expression data. The expressed proteins were assigned values 1 and the unexpressed proteins were assigned values 0. As the unexpressed proteins have no chance to interact with other proteins, only the proteins having the value 1 show the possibility to form interactions with other proteins. Thus each pair of proteins having assigned expression values 1 for both proteins of the pair was assumed to have a valid interaction between the proteins. Such sorting of valid interactions was conducted using codes based on JAVA programming language. The binary calculation was utilized for this purpose (only 1+1=1 denotes to valid interaction and 1+0=0, 0+1=0, 0+0=0 denote to invalid interaction). TextPad 4.42 version was used for the coding purpose (http://www.textpad.com/) (Helios Software Solutions, 2012). PINs were established for normal and cancer conditions of five tissues exploiting the valid interactions based on the expression data. Cytoscape 2.8.3 version was used for all the network construction purposes (Smoot et al., 2011; Cline et al., 2007; Shannon et al., 2003). The Network Analysis Plugin was used to determine the network parameters of each network. The parameters considered in the study were clustering coefficient, connected components, network diameter, network radius, network centralization, shortest paths, characteristic path length, average number of neighbors, multi-edge node pairs, number of edges, network density, network heterogeneity, isolated nodes, number of self-loops and number of nodes.

Table 1 Graph related parameters of the network of 8245 interactions

Parameters Name	Parameters Value
Clustering Coefficient	0.254
Connected Components	1
Network Diameter	7
Network Radius	4
Network Centralization	0.146
Shortest Paths	467172 (100%)
Characteristic Path length	2.966
Avg. Number of Neighbors	16.655
Number of Nodes	684
Number of Edges	8202
Network Density	0.024
Network Heterogeneity	1.149
Isolated Nodes	0
Number of Self-Loops	3
Multi-edge node pairs	2503

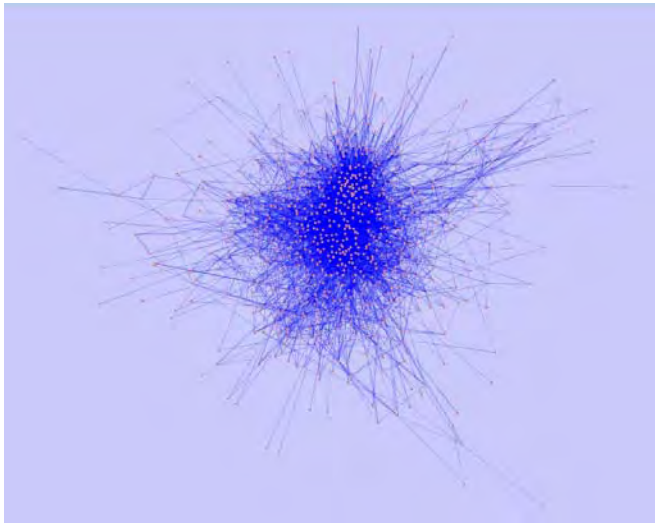


Fig. 1.1 BioLayout of 8245 interactions

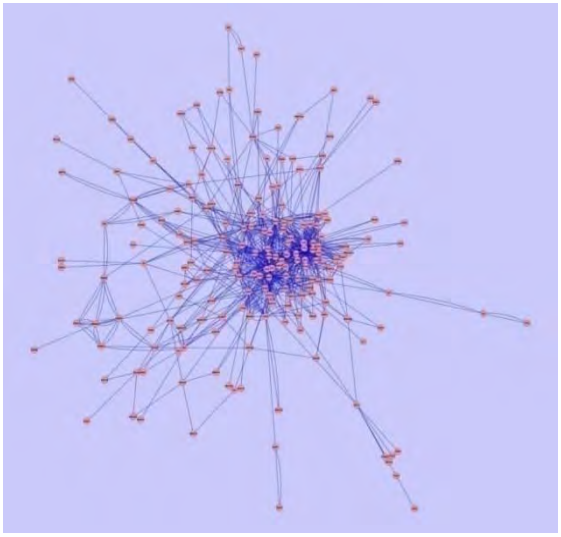


Fig. 1.2 BioLayout of PIN for bone (normal)

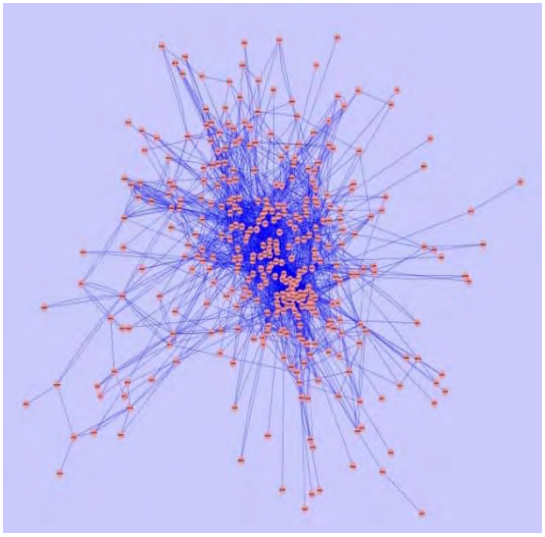


Fig. 1.3 BioLayout of PIN for bone (cancer)

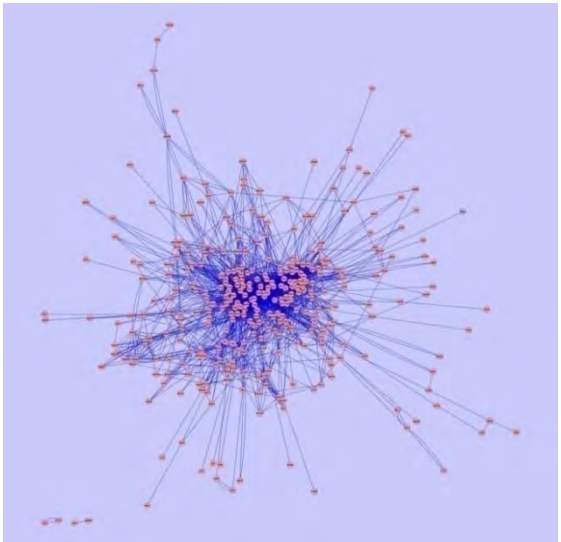


Fig. 1.4 BioLayout of PIN for breast (normal)

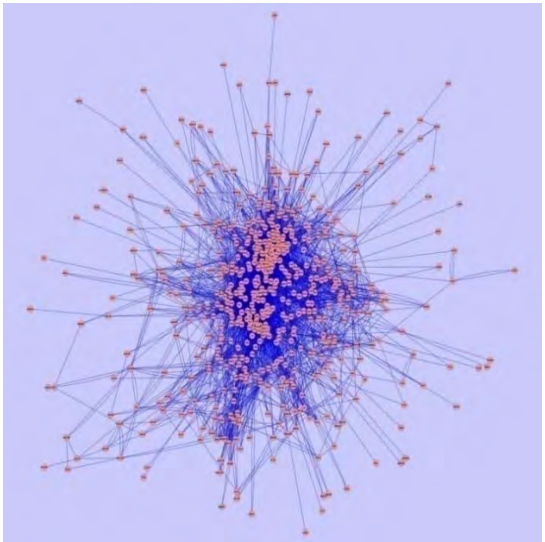


Fig. 1.5 BioLayout of PIN for breast (cancer)

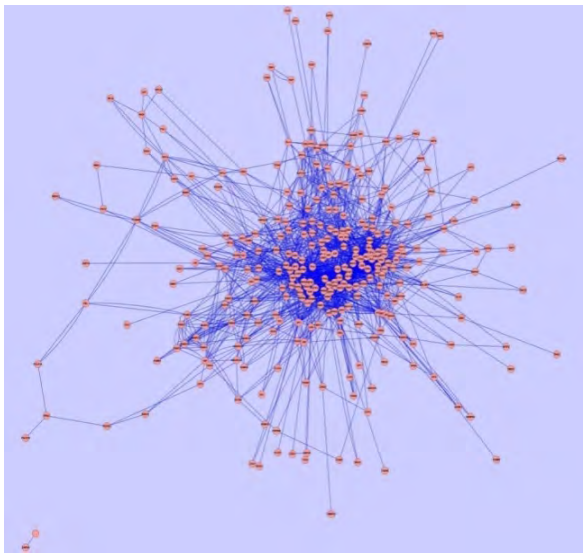


Fig. 1.6 BioLayout of PIN for colon (normal)

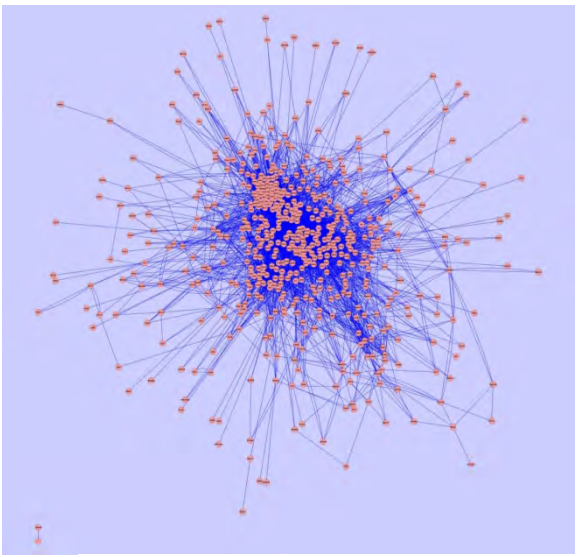


Fig. 1.7 BioLayout of PIN for colon (cancer)

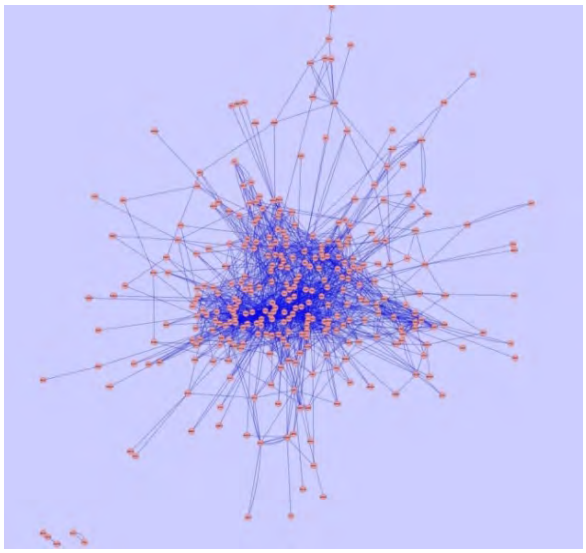


Fig. 1.8 BioLayout of PIN for kidney (normal)

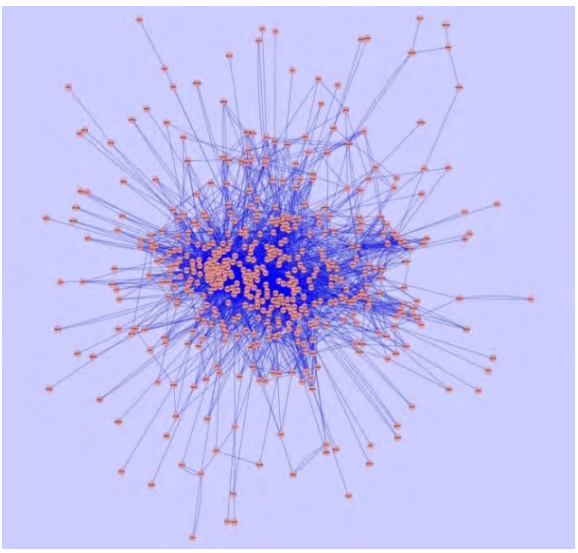


Fig. 1.9 BioLayout of PIN for kidney (cancer)

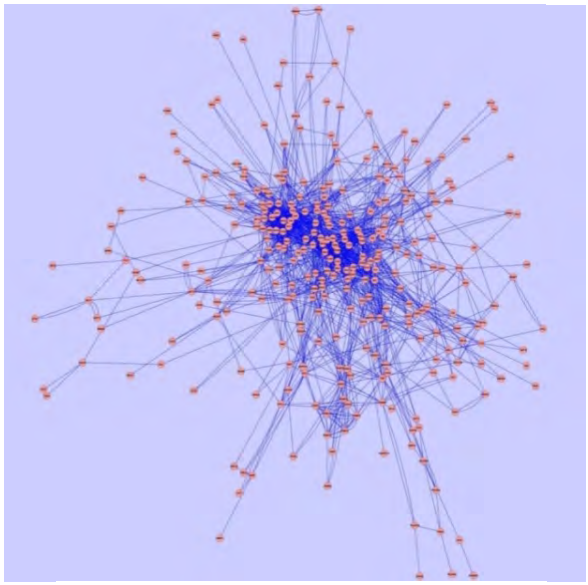


Fig. 1.10 BioLayout of PIN for liver (normal)

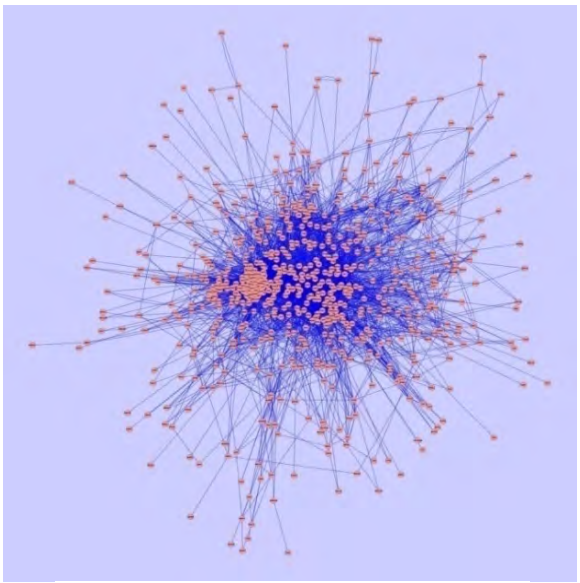


Fig. 1.11 BioLayout of PIN for liver (cancer)

2.2 Identification of conserved cluster of protein-protein interactions

Set of signaling proteins which are upregulated or downregulated during transformation from normal condition to cancer condition were sorted out. The commonly upregulated or downregulated signaling proteins in five tissues in case of cancer conditions were identified. The relevant interactions of the upregulated proteins were only considered (as the single downregulated protein has less interaction importance). TextPad 4.42 version (http://www.textpad.com/) (Helios Software Solutions, 2012) was used to code in JAVA programming language for identifying upregulated signaling proteins (the normal expression of all proteins were subtracted from their cancer expression and the proteins having the positive values were sorted out with the relevant interactions). Networks for interactions of commonly upregulated proteins for five tissues were constructed via Cytoscape 2.8.3 version (Smoot et al., 2011; Cline et al., 2007; Shannon et al., 2003). The largest clusters were identified for the cancer conditions in five tissues, and the proteins and the relevant interactions of the largest clusters were listed. The common set of proteins and interactions for five tissues during cancer conditions was identified and used to construct a PIN. This was further analyzed as conserved cluster of protein-protein interactions of upregulated signaling proteins during cancer conditions via Cytoscape Network Analysis Plugin. It is mentionable that all networks considered here were undirected networks. The differential networks were represented with edge-weighted force-directed (BioLayout) layout and the clusters were represented with degree sorted circle layout.

3 Results and Discussion

According to the objectives of our study we primarily constructed and visualized the differential PINs for five tissues e.g. bone, breast, colon, kidney and liver in normal and cancer conditions (Fig. 1.2-Fig. 1.11). The network parameters of the differential PINs were analyzed afterwards (Table 2). The graphical representations were done to compare the parameters (Fig. 2 (a)-2(n)). The network of 8245 interactions is found to have 684 nodes and 8202 edges. The differential networks based on the expression data show fluctuations from this network. The parameters for differential networks vary between normal and cancer conditions. Number of nodes, number of edges, multi-edge node pairs, average number of neighbors increase in cancer conditions for all five tissues. Network density and characteristic path length decrease in cancer conditions for all five tissues. Clustering coefficient increases in cancer conditions in the tissues under study except colon. Network diameter increases in kidney and liver, decreases in breast and colon and remains constant in bone during cancer conditions. Network radius increases in breast, kidney, and liver and remains constant in bone and colon. Network centralization increases in cancer conditions in the tissues under study except liver. Connected components decreases in breast and kidney and remains constant in bone, colon and liver. Shortest paths increases in breast, colon and kidney and remains constant in bone and liver. Network heterogeneity increases in colon, kidney and liver and decreases in bone and breast. Number of self-loops increases in kidney and liver and remains constant in bone, breast and colon. Isolated nodes number is zero for all five tissues.

Differential upregulation and downregulation of proteins in bone, breast, colon, kidney and liver in cancer conditions are presented (Fig. 3.1- Fig. 3.10). 64 proteins are found to be commonly upregulated (Fig. 4.1) in five tissues during cancer conditions and only one protein is found commonly downregulated. Interactions among the upregulated proteins show a large cluster and some discrete interactions in each tissue (Fig. 4.2 (a)-4.2 (e)). These large clusters from each of the different tissues has a common set of 34 proteins with 76 relevant interactions (Fig. 5.1 and 5.2(a)-5.2(e)) and this cluster of PIN remains conserved in all five tissues with respect to various network attributes (Table 3).

From the above study it is obviously evident to summarize that the intracellular biomolecular dynamics is quantitatively different in regard of PINs while at the same time our results suggest that qualitative fluctuations

might hold the underlying mechanism of observable disorders which could be subjected to an orderly control that remains subtle mostly. And the conservancy of PIN cluster eventually stands as a support for this interpretation. It is also found that a PIN cluster of interactions of 34 proteins remain conserved in the five cancerous tissues. It can be assumed that the conserved cluster play a non-trivial role at the very fundamental level of cancer and metastasis. Though we know that cancer is a result of chromosomal instability and random genetic mutations, PIN conservation points toward to a non-genetic regulation in cancer progression and also directs us to a new window of understanding the cell molecular biology.

Table 2 Graph Related Parameters for both the Normal and Cancerous Tissues

Tissue / Paramete	Bone		Breast		Colon		Kidney		Liver	
	Normal	Cancer	Normal	Cancer	Normal	Cancer	Normal	Cancer	Normal	Cancer
Clustering Coefficient	0.210	0.229	0.228	0.245	0.259	0.248	0.232	0.234	0.217	0.263
Connected Components	1	1	3	1	2	2	3	1	1	1
Network Diameter	7	7	9	8	8	7	7	8	8	9
Network Radius	4	4	1	4	1	1	1	5	4	5
Network Centralization	0.178	0.160	0.170	0.156	0.176	0.152	0.162	0.160	0.153	0.159
Shortest Paths	36672 (100%)	122850 (100%)	106606 (97%)	292140 (100%)	91508 (98%)	300854 (99%)	95798 (96%)	240590 (100%)	90902 (100%)	397530 (100%)
Characteristic Path length	3.214	3.082	3.060	2.976	3.037	2.967	3.110	3.017	3.243	3.035
Avg. Number of Neighbors	6.448	10.160	10.242	14.233	9.757	14.577	5.559	12.782	8.126	14.187
Number of Nodes	192	351	331	541	305	551	315	491	302	631
Number of Edges	899	2530	2537	5631	2109	5830	1895	4505	1752	5834
Network Density	0.034	0.029	0.031	0.026	0.032	0.027	0.027	0.026	0.027	0.023
Network Heterogeneity	1.069	1.046	1.117	1.095	1.061	1.124	1.075	1.134	1.079	1.232
Isolated Nodes	0	0	0	0	0	0	0	0	0	0
Number of Self-Loops	0	0	1	1	2	2	1	2	1	2
Multi-edge node pairs	280	747	741	1780	619	1812	546	1365	524	1365

Table 3 Graph Related Parameters for 34 conserved proteins

Parameters Name	Parameters Value
Clustering Coefficient	0.286
Connected Components	1
Network Diameter	9
Network Radius	5
Network Centralization	0.133
Shortest Paths	1122 (100%)
Characteristic Path length	3.982
Avg. Number of Neighbors	2.882
Number of Nodes	34
Number of Edges	76
Network Density	0.087
Network Heterogeneity	0.531
Isolated Nodes	0
Number of Self- Loops	0
Multi-edge node pairs	27

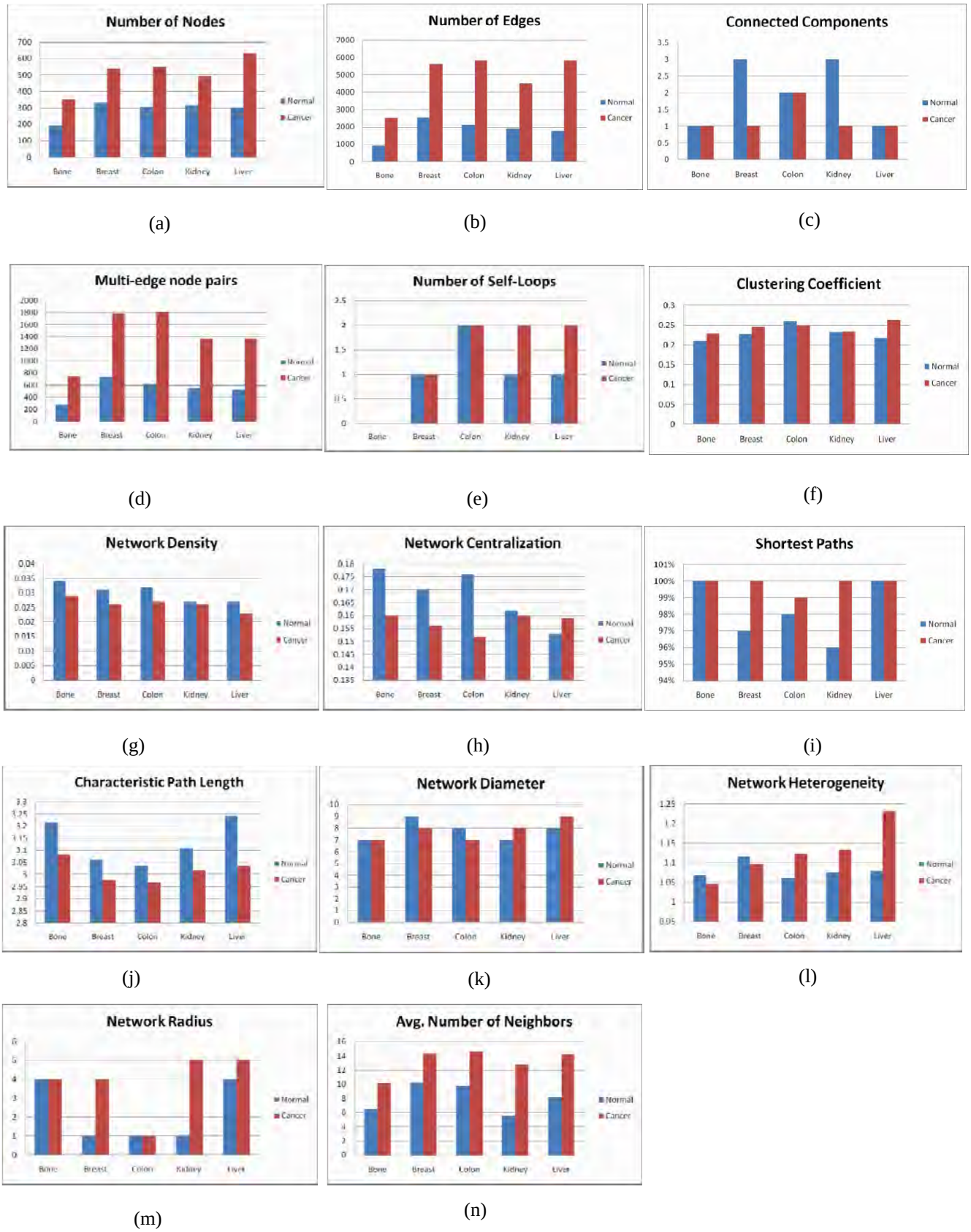


Fig. 2.1 Different network attributes for normal and cancerous Tissues. Number of nodes (a), Number of edges (b), Connected components (c), Multi-edge node pairs (d), Number of self-loops (e), Clustering Coefficient (f), Network density (g), Network centralization (h), Shortest path (i), Characteristic path length (j), Network diameter (k), Network heterogeneity (l), Network radius (m), Avg. number of neighbors (n).

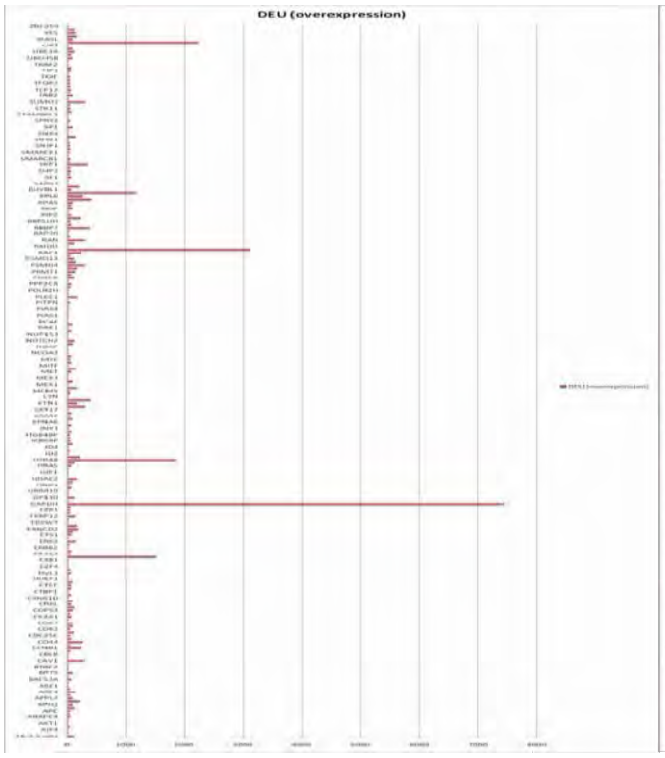


Fig. 3.1 Upregulated proteins in bone

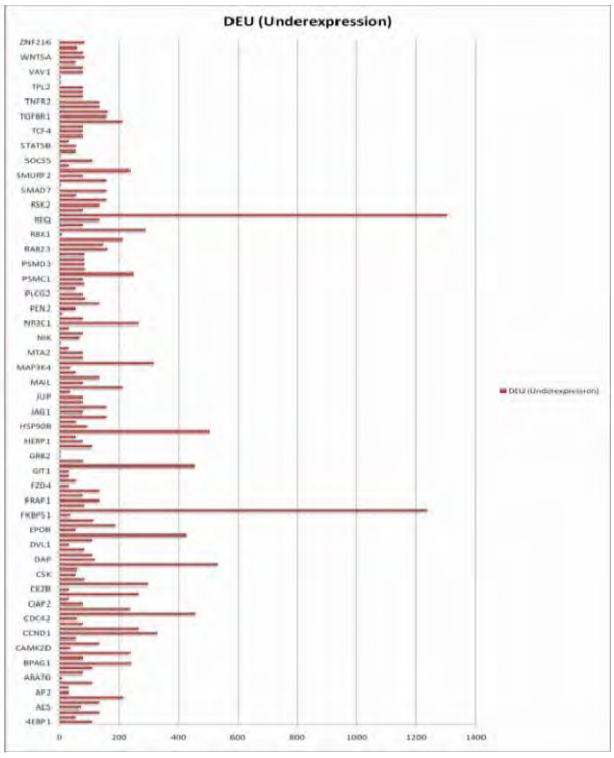


Fig. 3.2 Downregulated proteins in bone

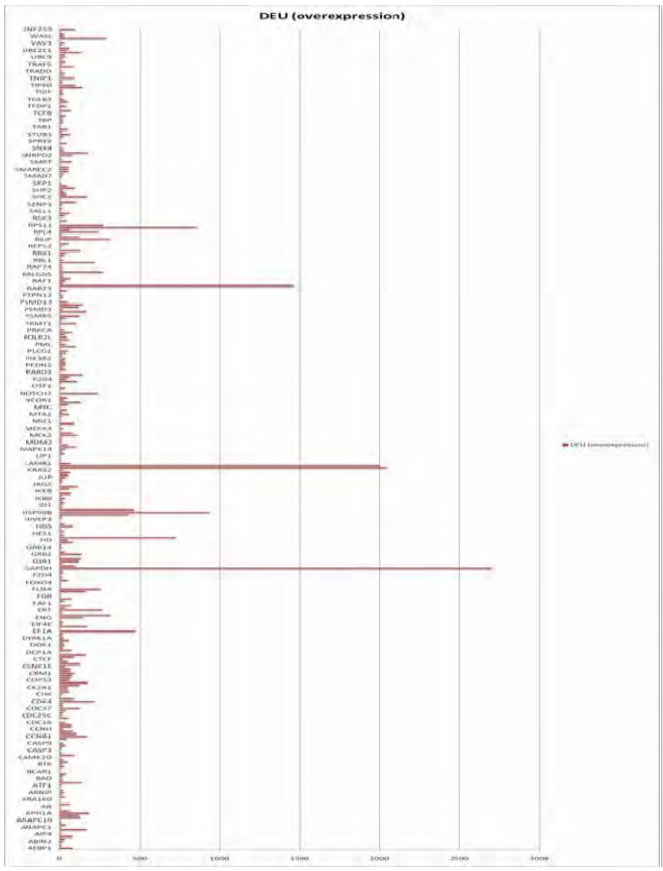


Fig. 3.3 Upregulated proteins in breast

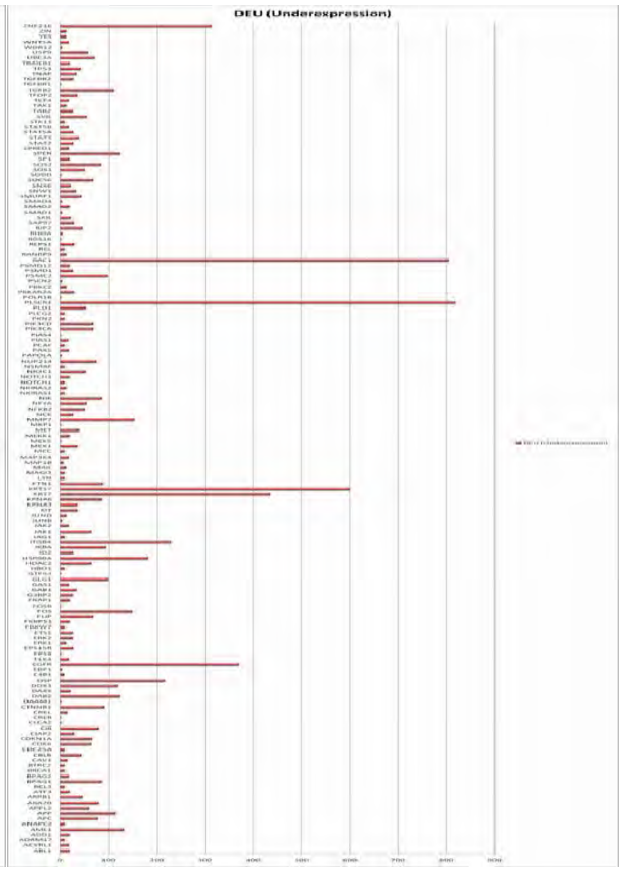


Fig. 3.4 Downregulated proteins in breast

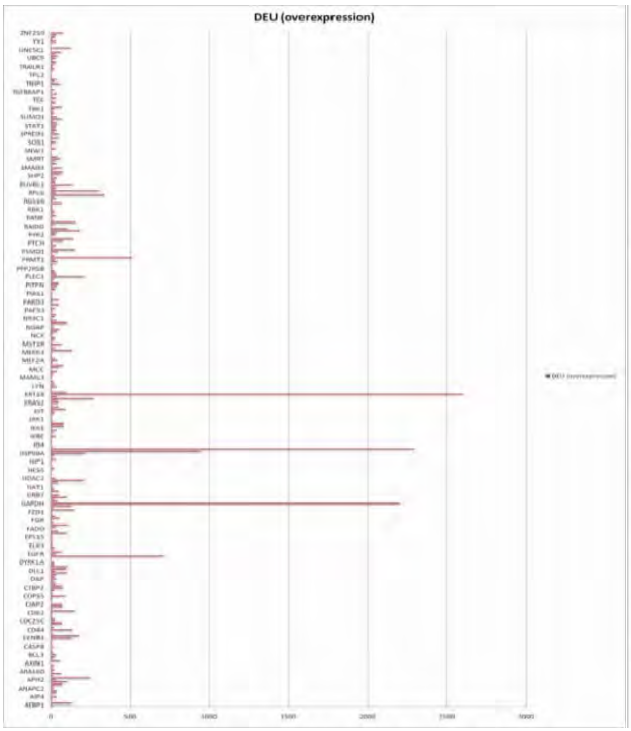


Fig. 3.5 Upregulated proteins in colon

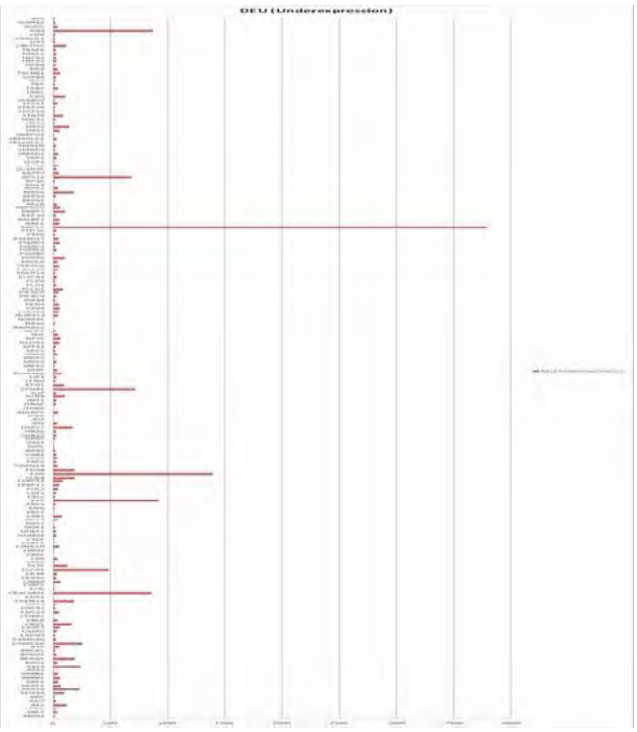


Fig. 3.6 Downregulated proteins in colon

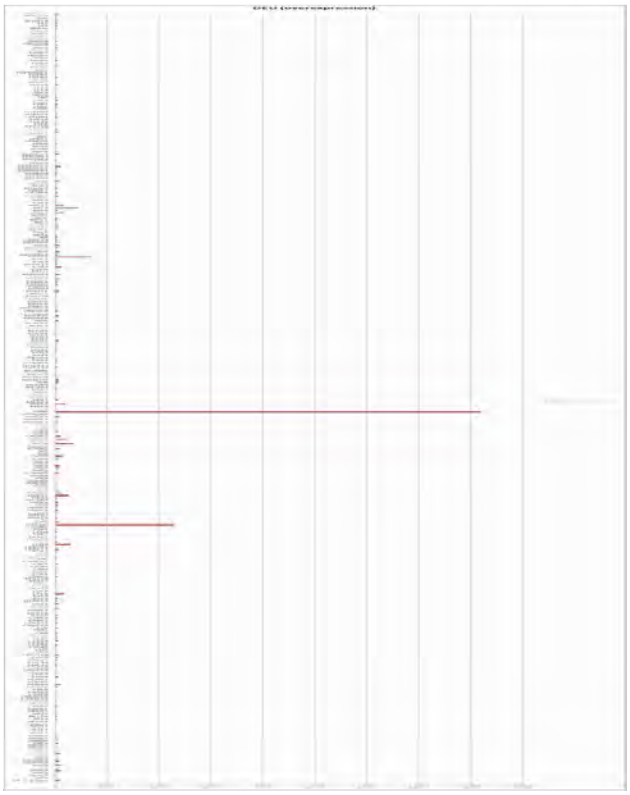


Fig. 3.7 Upregulated proteins in kidney

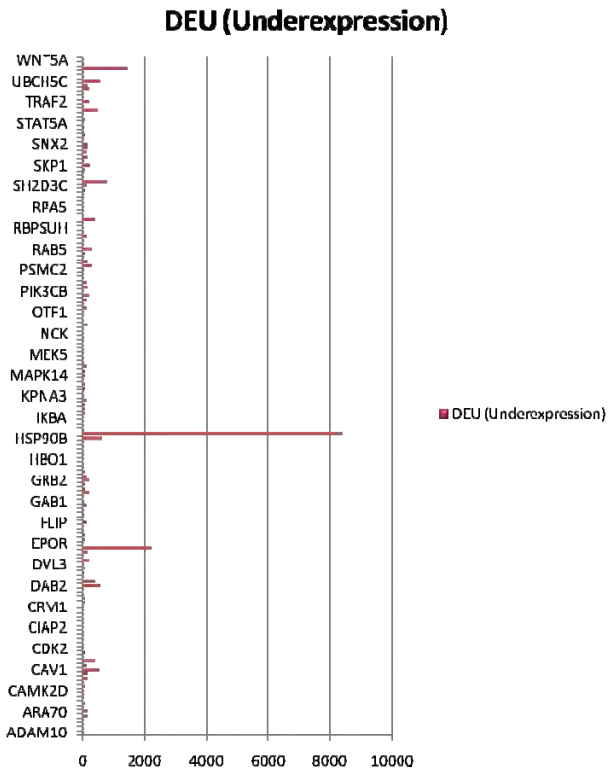


Fig. 3.8 Downregulated proteins in kidney

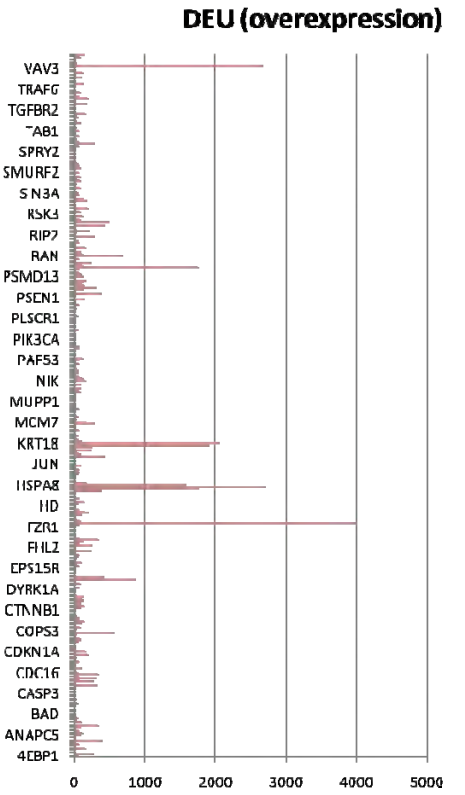


Fig. 3.9 Upregulated proteins in liver

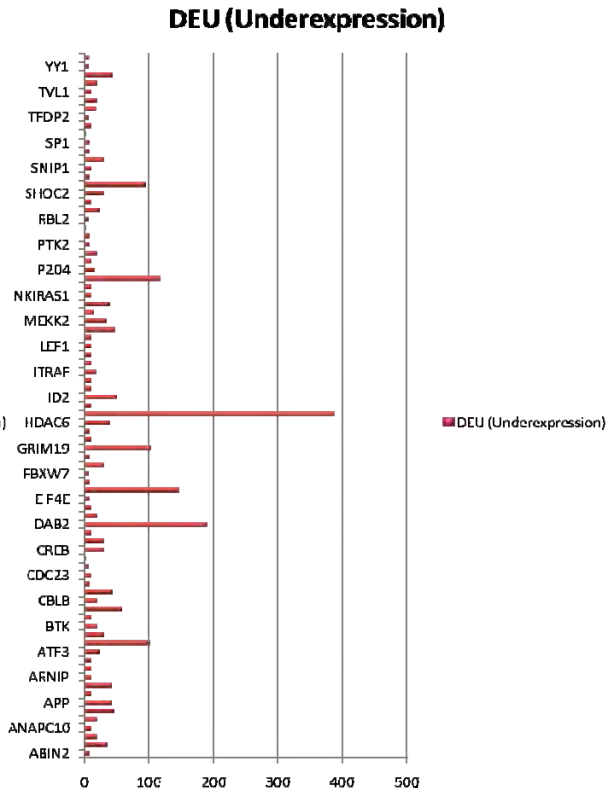


Fig. 3.10 Downregulated proteins in liver

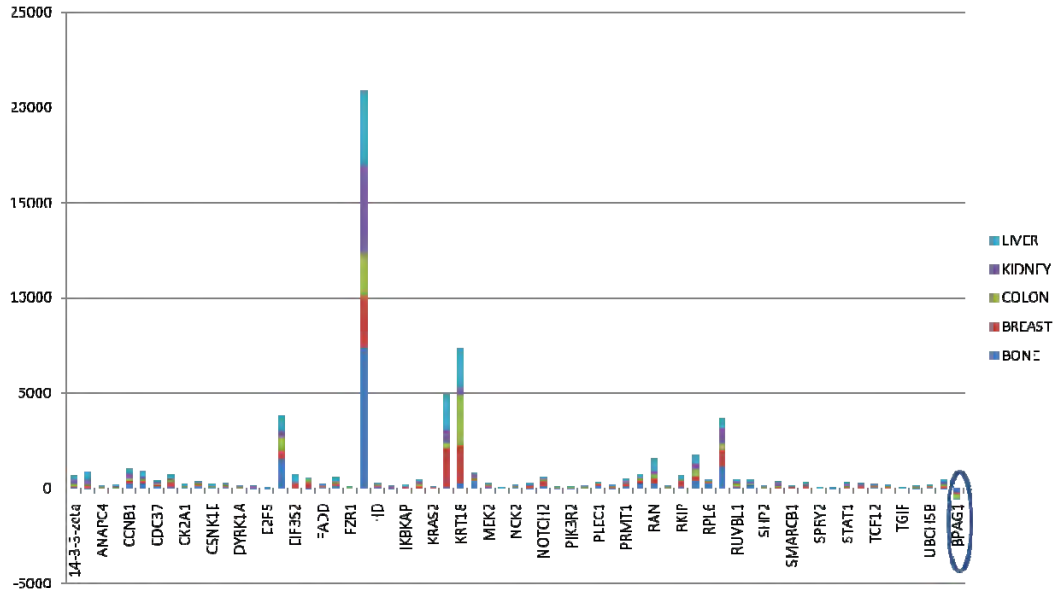
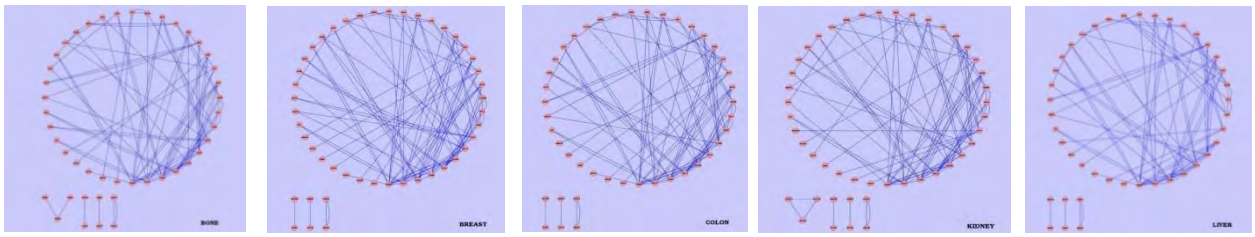


Fig. 4.1 Commonly expressed 64 proteins of all five tissues in cancer conditions.



(a) (b) (c) (d) (e)

Fig. 4.2 Cytoscape layout of PIN of commonly overexpressed proteins during cancer conditions in bone (a), breast (b), colon (c), kidney (d) and liver (e). Degree sorted circle layout is used in representation.

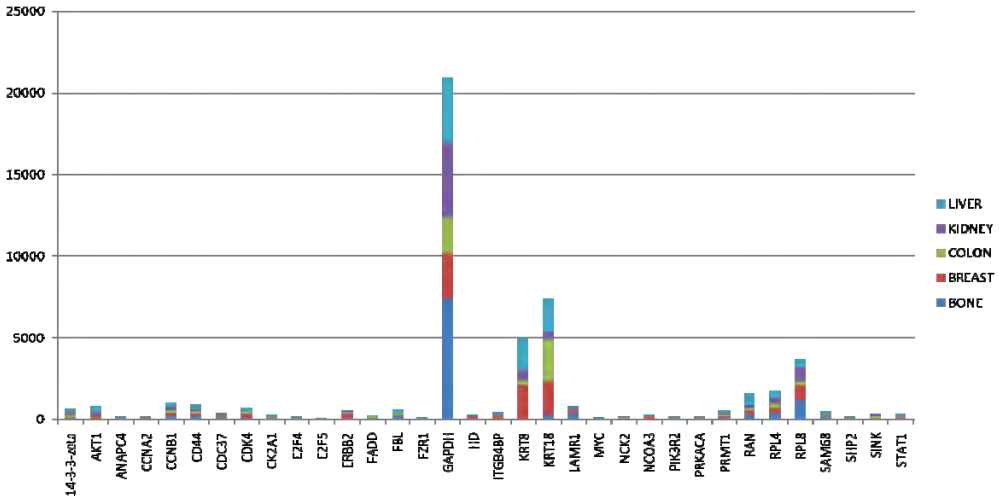
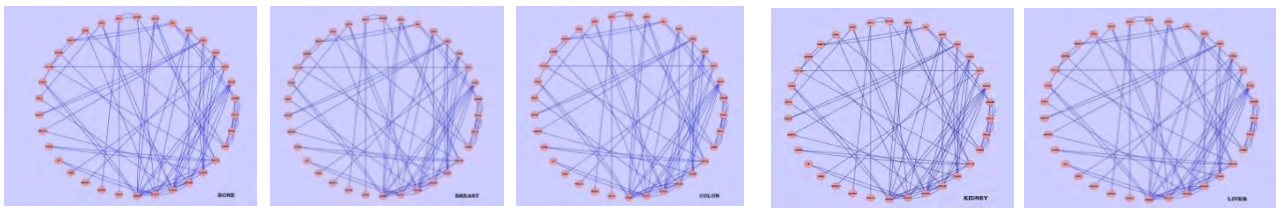


Fig. 5.1 Commonly expression 34 proteins of the large protein interaction network cluster for all five tissues in cancer conditions.



(a) (b) (c) (d) (e)

Fig. 5.2 Cytoscape layout of conserved PIN cluster of overexpressed proteins during cancer conditions in bone (a), breast (b), colon (c), kidney (d) and liver (e). Degree sorted circle layout is used in representation.

4 Conclusion and Recommendation

In general this study suggests the requirement of a more holistic understanding of cancer and metastasis and the inherent regulatory pattern of cancer emergence. This study includes only the networks of signaling proteins of cancer signal transduction pathways. But the total proteomic networks of cancer cells would be more convenient. Here only the simple parameters have been considered but more significant parameters like

network complexity, network entropy etc. are also required to be analyzed. This approach is based on some static networks, but dynamic network based studies are needed to bring out more realistic interpretations. Protein interaction network with association of gene regulatory networks would provide more holistic results which are beyond the scope of this study. To overcome the drawbacks of such studies high throughput proteomic study and highly comprehensive computational tools are required to use network systems biology as future tool of understanding cancer related biomolecular alterations more inclusively. A combination of wet lab and dry lab approaches is a must in this regard. Moreover, the evolutionary conservancy among cancer protein networks for different metazoa can be studied to decipher the common nature of cancer evolution which can lead us to a step ahead towards the pattern recognition in tumorigenesis. Also from the therapeutic point of view this type of network analysis can evidently identify important nodes and hubs in cancer PINs which can be used as new drug targets.

Acknowledgements

The authors like to thank Mahbub-E-Sobhani and Md. Shaifur Rahman for their endless inspiration, support and guidance throughout the work. The authors also like to acknowledge Hannan Hossain and Riasat Azim for their help and assistance in coding purposes. The authors are thankful to Mehdi Rahman, Sanjoy Roy, Ahmad Ullah, Md. Tauhid Siddiki Tomal, Saimoon Rahman Imran, Ahmed Ronju for their cordial help and assistance. In this study, KMTR has contributed to idea development and the construction and computational analysis of various networks. MFI has contributed to the development of idea and methodology and interpretation of the results. RSB, UH, FSD, SSS, SMTK have contributed to data retrieval, processing, maintenance and analysis. MSA has supervised the whole work. All the authors have contributed equally to the writing of the paper.

References

Arends JW. 2000. Molecular interactions in the Vogelstein model of colorectal carcinoma. Journal of Pathololology, 190: 412-416

Auguste P, Fallavollita L, Wang N, et al. 2007. The host inflammatory response promotes liver metastasis by increasing tumor cell arrest and extravasation. American Journal of Pathololology, 170(5): 1781-1792

Barabasi AL, Oltvai ZN. 2004. Network biology: understanding the cell's functional organization. Nature Reviews Genetics, 5: 101-113

Caldieri G, Buccione R. 2010. Aiming for invadopodia: orga-nizing polarized delivery at sites of invasion. Trends in Cell Biology, 20(2): 64-70

Cline MS, Smoot M, Cerami E, et al. 2007. Integration of biological networks and gene expression data using Cytoscape. Nature Protocols, 2: 2366- 2382

Davies PCW. 2004. Does quantum mechanics play a non-trivial role in life? BioSystems, 78: 69-79

Esteller M. 2008. Epigenetics in cancer. New England Journal of Medicine, 358(11): 1148-1159

Fearon ER, Vogelstein B. 1990. A genetic model for colorectal tumorigenesis. Cell, 61:759-767

Goldthwaite CA. 2006. Are stem cells involved in cancer? Regenerative Medicine**,** 9: 89-96

Goll MG, Bestor TH. 2005. Eukaryotic cytosine methyltransferase. Annual Review of Biochemistry, 74: 481-514

Hanahan D, Weinberg RA. 2000. The hallmarks of cancer. Cell, 100(1): 57-60

Huang JQ, Zhang WJ. 2012. Analysis on degree distribution of tumor signaling networks. Network Biology,

2(3): 95-109

Kitano H. 2002. Computational systems biology. Nature, 420: 206-210

Mamun MA, Rahman MS, Islam MF, et al. 2011. Molecular biology and the riddle of cancer: the 'Tom and Jerry' show. Oncology Reviews, 5: 215-222

McDowall MD, Scott MS, Barton GJ. 2009. PIPs: Human protein-protein interactions prediction database. Nucleic Acids Research, 37: D651-D656

Mirzarezaee M, Araabi BN, Sadeghi M. 2010. Comparison of hubs in effective normal and tumor protein interaction networks. Basic and Clinical Neuroscience, 2(10): 44-50

Oltvai ZN, Barabasi AL. 2002. Life's complexity pyramid. Science, 298: 763-764

Platzer A, Perco P, Lukas A, et al. 2007. Characterizaztion of protein-interactions networks in tumors. BMC Bioinformatics, 8: 224

Reya T, Morrison SJ, Clarke MF, et al. 2001. Stem cells, cancer, and cancer stem cells. Nature, 414: 105-111

Scott MS, Barton GJ. 2007. Probabilistic prediction and ranking of human protein-protein interactions. BMC Bioinformatics, 8: 239-260

Shannon P, Markiel A, Ozier O, et al. 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Research, 13(11): 2498-2504

Shrodinger E. 1958. What is life? Cambridge University Press, Cambridge

Smoot M, Ono K, Ruscheinski J, et al. 2011. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics, 27(3): 431-432

TextPad version 4.42. 2012. Helios Software Solutions, LONGRIDGE, England. http://www.textpad.com/

The Cancer Cell Map. 2006. Memorial Sloan-Kettering Cancer Center. http://cancer.cellmap.org/cellmap/

Tysnes BB. 2010. Tumor initiating and propagating cells: cells that we would like to identify and control. Neoplasia, 12(7): 506-515

Ye F, Zhou C, Cheng Q, et al. 2008. Stem cell abundant proteins Nanog, Nucleostemin and Musashi 1 are highly expressed in malignant cervical epithelial cells. BMC Cancer, 8: 108

Zhang WJ. 2012. Computational Ecology: Graphs, Networks and Agent-based Modeling. World Scientific, Singapore

Zhang Y, Luoh SM, Hon LS, et al. 2007. GeneHub-GEPIS: digital expression profiling for normal and cancer tissues based on an integrated gene database. Nucleic Acids Research, 35 (Web Server issue): W152–W158

Zhou TT. 2012. Network systems biology for targeted cancer therapies. Chinese Journal of Cancer, 31(3): 134-141