Identifying Drug Target Candidates from a Cancer-Related Disease Network

Jea Woon Ryu[1], Chom Kyu Chong[1], Jae Soo Yoo[2], Hak Yong Kim[*1]

[1]Department of Biochemistry, Chungbuk National University, Cheongju 361-763, Republic of Korea

[2]Department of Information and Communication, Chungbuk National University, Cheongju 361-763, Republic of Korea

jwryu84@cbnu.ac.kr; ckchong01@hotmail.kcom; jsy@cbnu.ac.kr; hykim@cbnu.ac.kr

***Abstract*- Experimental methods such as high-throughput screening have been widely used for the discovery of new drug targets. Genes or proteins close to one another in a network can govern similar functions and diseases. In an effort to predict new disease-related genes or proteins, we hypothesized that a network-neighbor of a disease-related gene or protein could cause the same or similar diseases. First, we constructed a cancer-related human disease network based on the known cancer-related genes and proteins. We obtained 60 diseasomal proteins that appeared to be involved at least cancers and/or other diseases, and chose them as potential drug targets. We then compared the cancer-related target proteins with a cancer-related drug target dataset obtained from DrugBank. Excluding 17 diseasomal proteins that have been already targeted by approved drugs, we identified 43 proteins as putative drug candidates for controlling cancer and/or cancer-related diseases.**

Keywords*- *Co-morbidity; Diseasome; Disease Network; Drug Target Proteome; Protein Network

I. INTRODUCTION

Several decades of experimental and clinical studies on individual diseases, genes, and proteins have led to the identification of numerous disease-related genes and proteins. Recently, new computational approaches for studying human diseases have allowed research to analyze single genes and/or proteins for the elucidation of complex diseases on the genome- and/or proteome-wide scale [1-4]. These studies were based on analyzing disease genes datasets, such as those found in the Online Mendelian Inheritance in Man (OMIM) database [5] and the Genome Wide Association Study (GWAS) database [6]. Human disease networks and disease gene networks derived from OMIM were the first to demonstrate the existence of "diseasome," i.e., a set of two or more diseases that share a disease-causing gene [3], and to extract large-scale co-morbidity patterns of disease pairs that share given genes [7]. Similarly, GWAS-based networks have been used to explore the shared genetic architectures of genes involved with complex human diseases [8]. Today, network biology is considered as a general means for understanding the relationships between diseases and genes/proteins and identifying disease mechanisms.

Network-based approaches have been used to elucidate the network properties of various disease-related genes and proteins [9, 10]. For example, neighbors of a disease-related gene or protein in a network have been shown to be more likely to cause a similar disease [9]. Furthermore, in some cases, two or more diseases are linked to each other by a single disease-causing gene, forming the basis for co-morbidity [7, 11]. Most cellular components do not function alone, instead interacting with each other to form functional complexes or pathways. Similarly, some disease-causing genes are functionally related to one another because their products interact in complexes or biological pathways [12].

For diseases that are caused by genetic abnormalities, the encoded disease proteins always directly, engage in the pathological events. Therefore, candidate drug targets can potentially be identified by using network studies to compare disease-related genes and proteins.

The aim of this study is to construct three different cancer-related complex networks from cancer-related gene information for comparing the results of gene based and/or protein based networks to identify disease-related new drug target proteins from tripartite networks that consist of three different cancer-related nodes, proteins, diseases, and related drugs. In addition, the study is to inform the attribution of 95 and 60 drug target genes and proteins, respectively.

II. METHODS

A. Cancer-related Data Information

The human cancer-related gene data used in this study was obtained from MorbidMap of OMIM (http://www.ncbi.nlm.nih.gov/omim/ getmorbid.cgi) [5]. This information was used to construct the disease and gene datasets. Goh *et al*. [3 previously reported the disease class information used based on the physiological system affected. The human protein-protein interaction dataset was obtained from Human Protein Reference Database (HPRD, http://hprd.org). Drug information was obtained from DrugBank (http://www.drugbank.ca).

B. Construction of Disease-related Complex Networks

We constructed the following six cancer-related complex networks: a gene-gene interaction network (GGN), a protein-

DOI: 10.5963/BER0203001

protein interaction network (PPN), a disease-disease interaction network based on cancer-causing genes (DDNG), a disease-disease interaction network based on cancer-performing proteins (DDNP), a drug-disease-gene tripartite network (DDGN), and a drug-disease-protein tripartite network (DDPN) [Fig. 1].

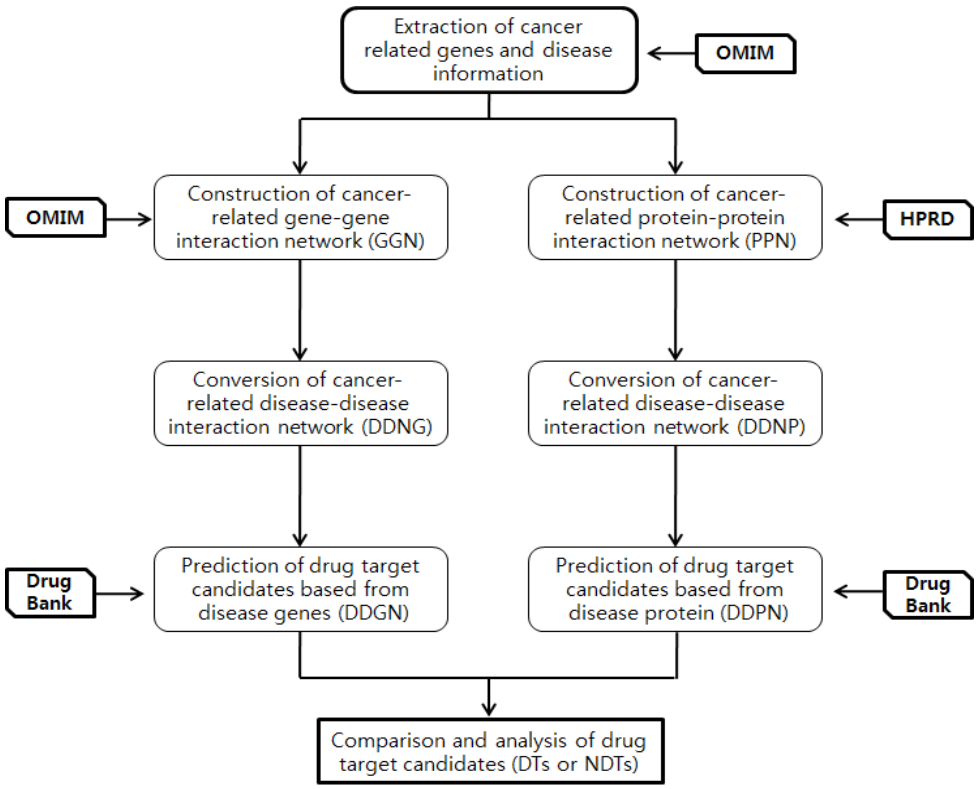

Fig. 1 Flow chart for identifying drug target candidates from cancer-related networks

In the GGN, each node represented a cancer-causing gene, and two genes were linked if the same disease was associated with both genes. This can prove insight into the relationship between cancer-causing genes and other genes, and important disease-related genes that have relationships with various diseases will have numerous links in the network. The nodes and links in the DDNG represented different diseases and the relationships between them, respectively. As a tripartite network, the DDGN also clearly identified the connections among genes, diseases, and drugs.

Although a disease may originate from an abnormality of a gene, the disease manifests because of an abnormality of the protein transcribed by the gene. Therefore, we constructed three different cancer-related networks based on proteins. The PPN was constructed from information on cancer-performing protein-protein interactions obtained from HPRD; in this network, the nodes represented proteins and the links represented protein-protein interactions. The DDNP and DDPN were constructed using the same methods that used to construct the DDNG and DDGN, respectively.

In basic principle, if two given components interacted, the link was designated as one. In contrast, if a component failed to interact with any other component, the link was set to zero. The network was then constructed as an adjacent matrix, wherein the rows and columns were labeled by graph nodes with 1 or 0 in component (i, j), according to whether i and j were adjacent or not. All networks were represented in graph form by using Cytoscape (http://www.cytoscape.org).

C. Identification of Diseasomal Genes and/or Proteins and New Drug Target Candidates

We identified a set of diseasomal genes (those connected with two or more diseases) from the DDGN. We also identified a set of diseasomal proteins (those connected with two or more diseases) from the DDPN. These diseasomal genes or proteins were designated as 'drug target candidates.' After excluding 17 proteins that have been already targeted by approved drugs, we identified a set of 43 novel candidate drug targets.

III. RESULTS AND DISCUSSION

We herein constructed six different cancer-related complex networks using human diseases, genes, proteins, and/or drugs as nodes and their interactions as links. The utilized datasets were obtained from OMIM, HPRD, and DrugBank. The GGN comprised 194 cancer-related genes and 1,644 interactions; its largest network consisted of 168 genes participating in 1,623 interactions [Fig. 2A]. PPN comprised 105 cancer-related proteins and 246 interactions; its largest network consisted of 103 proteins participating in 245 interactions [Fig. 2B].

DOI: 10.5963/BER0203001

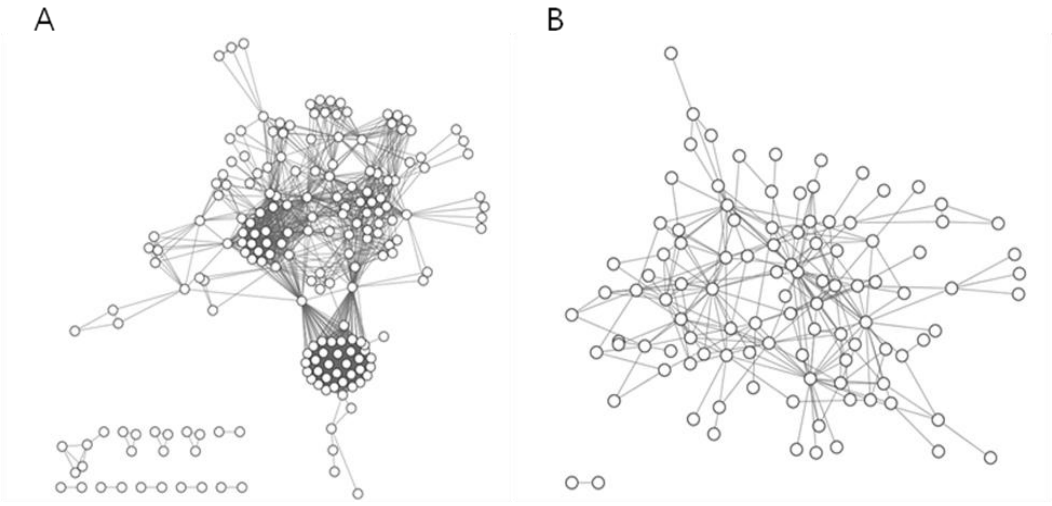

Fig. 2 Cancer-related gene-gene interaction network (GGN) and protein-protein interaction network (PPN). The GGN is composed of 194 genes and 1,644 interactions (A), while the PPN is composed of 105 proteins and 246 interactions (B).

Just as functions may be interrelated within a cell, diseases may be connected with one another. Although our networks were composed of cancer-causing genes and cancer-performing proteins, we explored cancer could be closely linked with various other diseases. Therefore, we converted the GGN and PPN to the DDNG and DDNP, respectively. The DDNG comprised 190 diseases and 541 interactions; its largest network consisted of 154 genes participating in 494 interactions [Fig. 3A]. The DDNP contained only a single large network, composed of 134 diseases and 1,191 interactions [Fig. 3B]. The nodes of the DDNP (134 proteins) were smaller than those of the DDNG (154 genes), but the links of the DDNP (1,191) were larger than those of the DDNG (494). Although genes always cause the diseases, proteins always directly govern the disease state; thus the number of links in the DDNP was, as expected, larger than that in the DDNG.

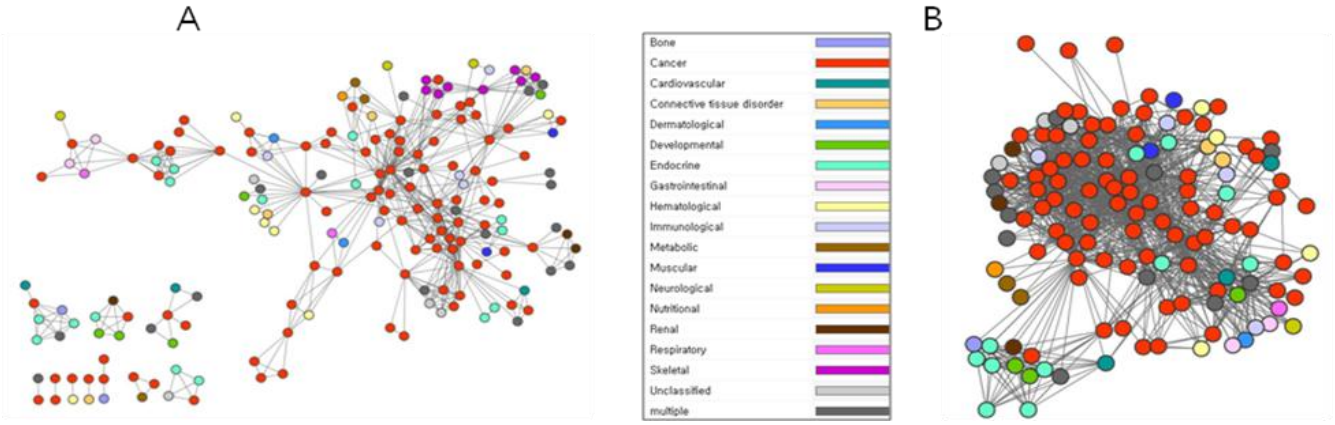

Fig. 3 Disease-disease interaction networks. The disease-disease interaction network (DDNG) derived from the GGN is composed of 190 diseases and 541 interactions (A), while the disease-disease interaction network (DDNP) derived from the PPN is composed of 134 diseases and 1,191 interactions (B).

By combining drug information with the DDNG and DDNP, we constructed the tripartite networks, DDGN and DDPN. The DDGN contained two tripartite networks (one large network and one very small network) and 19 relatively tiny bipartite networks that were not linked with any drugs [Fig. 4A]. The largest tripartite network contained 386 nodes (176 genes, 156 diseases, and 54 drugs) and 460 interactions; in it, all of the drugs related with diseasomal genes were connected with two or more diseases. The smaller tripartite network contained two diseasomal genes, nine diseases, nine drugs, and their interactions [Fig. 4A, bottom right].

The DDPN contained two tripartite networks (one large network and one very small network) and 10 relatively tiny bipartite networks [Fig. 4B]. The largest tripartite network contained 250 nodes (90 proteins, 110 diseases, and 50 drugs) and 302 interactions. As in the DDGN, all of the drugs related with diseasomal proteins were connected with two or more diseases. The smaller tripartite network contained only one diseasomal protein, five diseases, five drugs, and their interactions [Fig. 4B, bottom].

We identified 95 diseasomal genes and 60 diseasomal proteins in the DDGN and DDPN, respectively. The encoding genes of all 60 diseasomal proteins were found within the dataset containing the 95 diseasomal genes. The rest 35 genes have only gene information but not protein one in datasets is used in this study. In DDGN and DDPN, all of the connected drugs targeted diseasomal genes or proteins, indicating that the drug developers had aimed at such targets despite lacking identified

DOI: 10.5963/BER0203001

diseasomal information, and further suggesting that identified diseasomal genes and proteins could be considered candidate drug target.

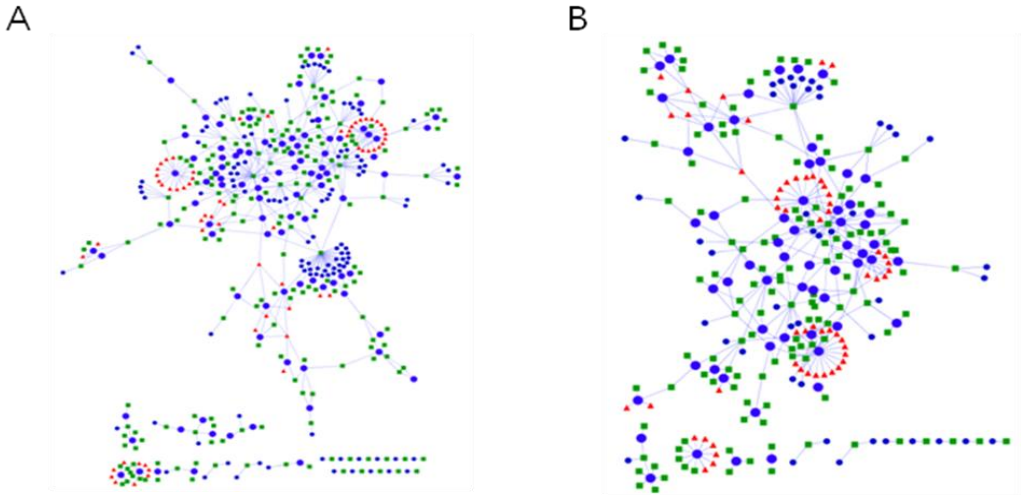

Fig. 4 Drug-disease-gene tripartite networks. The drug-disease-gene network (DDGN) is composed of 63 drugs, 201 diseases, 209 genes, and 528 interactions (A), while the drug-disease-protein network (DDPN) is composed of 56 drugs, 134 diseases, 105 proteins, and 336 interactions (B). Drugs, diseases, genes/proteins, and diseasomal genes/proteins are indicated with red triangles, green squares, small blue circles, and large blue circles, respectively.

Figure 5 shows the relationships between our identified diseasomal genes and proteins and the known, previously developed drugs. Of 95 diseasomal genes, only 22 were found to be targeted by developed drug. Thus, the remaining 73 could be considered target genes for future drug development. At the protein level, we identified 43 candidates that have not yet been targeted by drug development [Fig. 5]. Although a disease may originate from an abnormality of a gene, the disease reveals abnormal action of the protein transcribed by the gene. The latter might be more useful, given that proteins directly shape the disease phenotype. Thus, we herein propose 43 proteins as putative drug candidates for controlling cancer and cancer-related diseases.

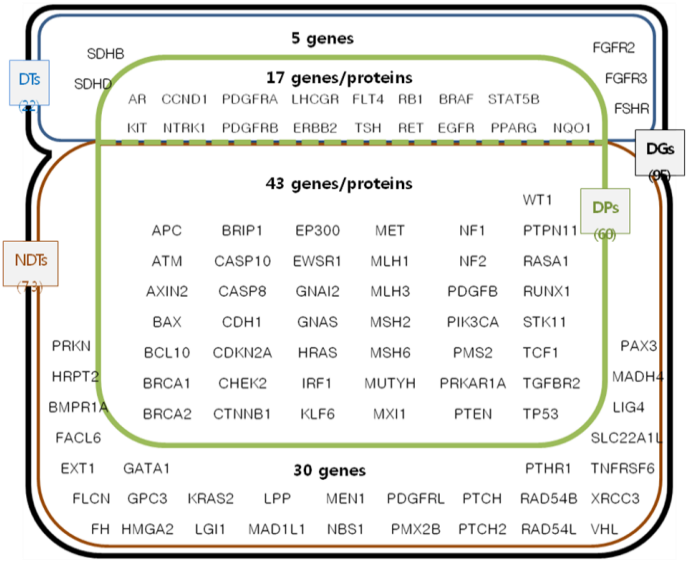

Fig. 5 Identification of drug target candidates. The 95 diseasomal genes (DGs) and 60 diseasomal proteins (DPs) obtained from the DDGN and DDPN, respectively, were identified as candidate drug targets. Their presence or absence in DrugBank is indicated as drug targets (DTs) or non-drug targets (NDTs), respectively.

IV. CONCLUSIONS

In the years since the first report of a human disease network [3], researchers have extensively studied disease-gene networks and disease-protein networks in the hope of identifying new network properties and/or novel potential drug targets [7, 10, 11]. Here, we constructed two different networks containing disease-causing genes and disease-performing proteins, and then combined the different approaches to obtain drug target information. Although a protein-interaction network in tumors was previously reported and characterized with respect to its topological and informational aspects [13], we herein employed a more biological approach based on cancer-related networks.

To consider diseases at the systems level, it is important to obtain diseasome information from related networks. Here, we

DOI: 10.5963/BER0203001

derived 95 diseasomal genes and 60 diseasomal proteins from the DDGN and DDPN, respectively [Fig. 3]. These are useful systems-level drug candidates because most of the identified genes/proteins are associated with multiple different diseases, forming a huge network. Of the 95 diseasomal genes, 22 were found to be targeted by commercialized drugs (In case of diseasomal proteins, 17 were found). Most of these drugs are small molecules that act as enzyme inhibitors, antagonists, anti-angiogenesis agents, and/or pharmacological reagents. Two, RB1 and TSHR, are targeted by protein drugs for suppressing cancer and/or related diseases (recombinant insulin and thyrotropin α, respectively)

When considering candidate drug targets, disease-performing proteins are better than disease-causing genes because proteins directly generate the disease phenotype. Of the 60 diseasomal proteins identified herein, 53% (32) are involved in signal transduction, and 25% (15) are involved in transcriptional regulation. Many cancers are due to abnormal cell proliferation, which is closely related to signal transduction and transcriptional regulation [14]; thus, these could be strong putative targets for cancer therapy. Effective targeting of the 43 new potential drug targets identified herein might allow us to suppress cancers via enzyme inhibition, antagonists, anti-angiogenesis agents, anti-cancer reagents, and related protein drugs. Although we suggested 43 proteins for new potential drug targets, it is still far away from a clinical application. Generally, action mechanism study for a drug candidate has to be performed before applying clinical approaches.

In addition, microRNAs (miRNAs), which are small non-coding RNAs that function in the post-transcriptional regulation of gene expression, have recently been reported as strong candidates for the treatment of many diseases, especially cancer [15]. To regulate the 60 diseasomal proteins for controlling cancer and/or related diseases at the mRNA level, it might be possible to design specific miRNAs. In addition, it may also be possible to control the 95 diseasomal genes at the transcriptional level via epigenetic approaches [16].

ACKNOWLEDGMENT

This work was supported by a fund (2013E6200100) by Research of Korea Centers for Disease Control and Prevention and by the Korea Institute of Science and Technology Information (K-13-L01-C02).

REFERENCES

[1] S. Wachi, K. Yoneda, and R. Wu, "Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues," *Bioinformatics,* vol. 21, pp. 4205-4208, Dec. 2005.

[2] P.F. Jonsson and P.A. Bates, "Global topological features of cancer proteins in the human interactome," *Bioinformatics,* vol. 22, pp. 2291-2297, Jul. 2006.

[3] K.I. Goh, M.E. Cusick, D. Valle, B. Childs, M. Vidal, and A.L. Barabasi, "The human disease network," *Proc. Natl. Acad. Sci. USA,* vol. 104, pp. 8685-8690, May 2007.

[4] J. Loscalzo, I. Kohane, and A.L. Barabasi, "Human disease classification in the postgenomic era: a Complex systems approach to human pathobiology," *Mol. Syst. Biol.*, vol. 3, pp. 124, Jul. 2007.

[5] A. Hamosh, A.F. Scott, J.S. Amberger, C.A. Bocchini, and V.A. McKusick, "Online Mendelian inheritance in man (OMIM), a knowledgebase of human genes and genetic disorders," *Nucleic. Acids Res.,* vol. 33, pp. D514-D517, Jan. 2005.

[6] J.G. Smith and C. Newton-Cheh, "Genome-wide association study in humans," *Methods Mol. Biol.*, vol. 573, pp. 231-258, Nov. 2009.

[7] J. Park, D.S. Lee, N.A. Christakis, and A.L. Barabasi, "The impact of cellular networks on disease comorbidity," *Mol. Syst. Biol.*, vol. 5, pp. 262, Apr. 2009.

[8] F. Barrenas, S. Chavali, P. Holme, R. Mobini, and M. Benson, "Network properties of complex human disease genes identified through genome-wide association studies," *PLoS ONE,* vol. 4, pp. e8090, Nov. 2009.

[9] S. Chavali, F. Barrenas, K. Kanduri, and M. Benson, "Network properties of human disease genes with pleiotropic effects," *BMC Syst.* Bio., vol. 4, pp. 78, Jun. 2010.

[10] T. Ideker and R. Sharan, "Protein networks in disease," *Genome Res.*, vol. 18, pp. 644-652, Apr. 2008.

[11] D.S. Lee, J. Park, K.A. Kay, N.A. Christakis, Z.N. Oltvai, and A.L. Barabasi, "The implications of human metabolic network topology for disease comorbidity," *Proc. Natl. Acad. Sci. USA,* vol. 105, pp. 9880-9885, Jul. 2008.

[12] Y. Li and P.A. Agarwal, "A Pathway-based view of human diseases and disease relationships," *PLoS ONE,* vol. 4, pp. e4346, Feb. 2009.

[13] A. Platzer, P. Perco, A. Lukas, and B. Mayer, "Characterization of protein-interaction networks in tumors," *BMC Bioinformatics,* vol. 8, pp. 224, Jun. 2007.

[14] R. Bianco, D. Melisi, F. Ciardiello, and G. Tortora, "Key cancer cell signal transduction pathways as therapeutic targets," *Eur. J.* Cancer, vol. 42, pp. 290-294, Feb. 2006.

[15] M. Lu, Q. Zhang, M. Deng, J. Miao, Y. Guo, W. Gao, and Q. Cui, "An analysis of human microRNA and disease associations," *PLoS ONE,* vol. 3, pp. e3420, Oct. 2008.

[16] R.M. Pollock and V.M. Richon, "Epigenetic approaches to cancer therapy," *Drug Discovery Today,* vol. 6, pp. 71-79, Jul. 2009.

DOI: 10.5963/BER0203001