Comparison of gene regulatory networks of benign and malignant breast cancer samples with normal samples

D.B. Chen and H.J. Yang

Breast Surgery, Zhejiang Cancer Hospital, Hangzhou, Zhejiang Province, China

Corresponding author: D.B. Chen
E-mail: daobaochen001@hotmail.com

ABSTRACT. The aim of this study was to explain the pathogenesis and deterioration process of breast cancer. Breast cancer expression profile data GSE27567 was downloaded from the Gene Expression Omnibus (GEO) database, and breast cancer-related genes were extracted from databases, including Cancer-Resource and Online Mendelian Inheritance In Man (OMIM). Next, h17 transcription factor data were obtained from the University of California, Santa Cruz. Database for Annotation, Visualization, and Integrated Discovery (DAVID)-enrichment analysis was applied and gene-regulatory networks were constructed by double-two-way t-tests in 3 states, including normal, benign, and malignant. Furthermore, network topological properties were compared between 2 states, and breast cancer-related bub genes were ranked according to their different degrees between each of the two states. A total of 2380 breast cancer-related genes and 215 transcription factors were screened by exploring databases; the genes were mainly enriched in their functions, such as cell apoptosis and proliferation, and pathways, such as p53 signaling and apoptosis, which were related with carcinogenesis. In addition,

gene-regulatory networks in the 3 conditions were constructed. By comparing their network topological properties, we found that there is a larger transition of differences between malignant and benign breast cancer. Moreover, 8 hub genes (*YBX1*, *ZFP36*, *YY1*, *XRCC5*, *XRCC4*, *ZFHX3*, *ZMAT3*, and *XPC*) were identified in the top 10 genes ranked by different degrees. Through comparative analysis of gene-regulation networks, we identified the link between related genes and the pathogenesis of breast cancer. However, further experiments are needed to confirm our results.

Key words: Gene expression profile; Gene-regulatory network; Network topological properties; Benign and malignant breast cancer

INTRODUCTION

Breast cancer is the most common malignant tumor in females, with an incidence varying largely by race/ethnicity, socioeconomic status, and geographic region (Ma and Jemal, 2013). In general, cancer is categorized as non-invasive or benign when the cancer cells are confined to their place of origin, do not threaten life, and do not spread outside of the breast, and as invasive or malignant when the cancer cells break through the duct into the surrounding fatty and connective tissues and may lead to death if not detected or cured (Jemal et al., 2005). Public health data indicate that the global burden of breast cancer in women, measured by its incidence, mortality, and economic costs, is substantial, and is currently increasing (Coughlin and Ekwueme, 2009). Therefore, the research and treatment of breast cancer is very significant for human health.

Recently, a large number of studies have been conducted to examine breast cancer, which involves many genes and pathways. Tumor suppressor genes (*TP53*, *ATM*, *PTEN*, *CHK2*, *BRCA1*, and *BRCA2*) play major roles in familial human breast cancer, while other genes that are involved in a large number of human cancers, including breast cancer, are either tumor suppressors (*RB1* and *RAD51*) or oncogenes (*cMYC*) (Fritz et al., 2013). In addition, the Notch, Hedgehog, nuclear factor-κB, and PIK3CA pathways are involved in breast cancer development (Jiao et al., 2012). However, the relationships between breast cancer-related genes and the pathogenesis of breast cancer remain unclear. Understanding these mechanisms would facilitate the search for new therapeutics for breast cancer.

Many studies have examined the pathogenesis of diseases by using gene-regulatory networks (Araki et al., 2013). Moreover, gene expression profile data have been widely used for the screening of breast cancer-related genes (Abba et al., 2010), classification of breast cancer subtypes (Morrison et al., 2012), and for the diagnosis, treatment, and prognosis of breast cancer (Aleskandarany et al., 2012). We collected breast cancer-related genes from databases, such as the Gene Expression Omnibus (GEO), Cancer-Resource, and Online Mendelian Inheritance In Man (OMIM), constructed gene-regulatory networks in three states, including normal, benign, and malignant, and then compared network topological properties between 2 states. Finally, we selected all hub genes in the 3 networks, and then ranked all of the genes according to differences in genetic degrees between two conditions.

MATERIAL AND METHODS

Expression profile data

We searched the GEO database (http://www.ncbi.nlm.nih.gov/geo/) (Barrett et al., 2005) and downloaded breast cancer expression profile data GSE27567 (LaBreche et al., 2011), which contains 31 sample chips from normal tissues, 37 from benign breast tissues, and 57 from malignant breast tissues. Through corresponding relationships between genes and probes in chip platform GPL570, the probe number was converted into a corresponding gene symbol. As one gene corresponds to multiple probes and consists of multiple expression values, we used the average value as the sole representative.

Breast cancer gene data

Cancer-Resource (http://bioinf-data.charite.de/cancerresource/) (Ahmed et al., 2011) is an integrated database of interactions between cancer-associated proteins and complexes. OMIM (http://www.ncbi.nlm.nih.gov/omim) (van Triest et al., 2011) is a human Mendelian genetic disease database. COSMIC (http://www.sanger.ac.uk/genetics/CGP/cosmic/) (Bamford et al., 2004) is a cancer-related gene database that originated in the Cold Spring Harbor Laboratory, and GeneRIF (http://www.ncbi.nlm.nih.gov/gene/about-generif/) stores corresponding literature for the genes. We searched for breast cancer-related genes in the databases described above. Furthermore, to verify these collected genes, Gene Ontology (GO) (Harris et al., 2004) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa, 2002) enrichments were conducted using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (Dennis Jr. et al., 2003). A false discovery rate (FDR) <0.05 was used as a significance cut-off criterion.

Transcription factor data

We downloaded h17 transcription factor data from UCSC (http://genome.ucsc.edu/) (Karolchik et al., 2003).

Gene-regulatory network construction

We constructed gene-regulatory networks under specific conditions using the double-two-way t-test, which was proposed in 2012 by Qi and Michoel. This method is superior to many existing methods such as CLR, Pearson, GENIE3, Spearman, Inferelator, and LeMoNe. The first step of the double-two-way t-test is to define the key control of each gene; the second is to calculate the differentially expressed value using a simple t algorithm. Using a set of candidate transcription factors and key controls of all target genes, we next conducted the t-test for transcription factors according to defined key controls of genes, and obtained P values that were considered to represent the regulation strength of transcription factors to genes. Finally, potential regulatory relationships were screened according to the P value (<0.05). Gene-regulatory networks were constructed based on regulatory relationships.

Comparison of network topological properties

Transcription factors and target genes are nodes in the gene-regulatory network, while regulation relationships between them are the edges in the network. Degree, which is defined as the number of edges connecting particular nodes, is the simplest measure of network nodes (Boccaletti et al., 2006). For a directed graph, degree is divided into out-degree and in-degree, where out-degree is the number of edges from one node to other nodes and in-degree is the number of edges from other nodes to one node. Betweenness measures the probability that a node appears between the other nodes along the shortest path (Joy et al., 2005). The larger the value of betweenness, the larger the bottleneck effect of this node. Once this node is broken, the entire connectivity of the network may be largely disordered. Closeness refers to the average distance from one node to all other nodes in the network (Ma and Zeng, 2003). A lower value for the closeness of a node indicates higher information transmission efficiency of this node. The cluster coefficient refers to closeness between one node and its neighboring nodes. The greater the clustering coefficient, the better the modularity of the node (Lin et al., 2011).

We calculated the out-degree, in-degree, betweenness, cluster coefficient, and closeness of the gene-regulatory networks using the igraph package (Csardi and Nepusz, 2006) under 3 conditions, and then compared the differences of these 5 attributes between 2 states using the rank sum test (Thomas et al., 2001).

Gene ranking

We selected all hub genes (level ≥10) in 3 networks, and then ranked all of the genes according to different degrees between each of the 2 conditions.

RESULTS

Breast cancer-related gene collection

We extracted breast cancer expression profile data GSE27567 from GEO and retrieved a total of 21,609 genes after processing. Next, we searched Cancer-Resource, OMIM, COSMIC, and GeneRIF, and obtained 631, 94, 19, and 1972 breast cancer-related genes, respectively. After integrating the 4 databases, 2380 non-redundant breast cancer-related genes were identified. Finally, we downloaded h17 transcription factor data from UCSC and obtained a total of 215 transcription factors, of which 59 were breast cancer-related genes with gene expression profile data.

DAVID-enrichment analysis

Functions and pathways were enriched to verify the reliability of the collected breast cancer-related genes. DAVID-enrichment analysis showed that these genes were mainly enriched in biological processes associated with carcinogenesis, such as cell death, cell proliferation, and apoptosis (Table 1) and pathways, such as the p53 signaling pathway, pathways in cancer, and apoptotic pathways (Table 2).

Table 1. DAVID function-enrichment analysis.

Category	Term	FDR
GOTERM_BP_FAT	GO:0042127-regulation of cell proliferation	1.51E-107
GOTERM_BP_FAT	GO:0010941-regulation of cell death	8.09E-101
GOTERM_BP_FAT	GO:0043067-regulation of programmed cell death	1.07E-100
GOTERM_BP_FAT	GO:0042981-regulation of apoptosis	1.25E-98
GOTERM_BP_FAT	GO:0010033-response to organic substance	6.95E-67
GOTERM_BP_FAT	GO:0043069-negative regulation of programmed cell death	2.69E-64
GOTERM_BP_FAT	GO:0060548-negative regulation of cell death	5.03E-64
GOTERM_BP_FAT	GO:0043066-negative regulation of apoptosis	5.90E-64
GOTERM_BP_FAT	GO:0008284-positive regulation of cell proliferation	2.26E-56
GOTERM_BP_FAT	GO:0010604-positive regulation of macromolecule metabolic process	1.26E-53

BP = biological process; FDR = false-discovery rate.

Table 2. DAVID pathway-enrichment analysis.

Category	Term	FDR
KEGG_PATHWAY	hsa05200:Pathways in cancer	4.89E-42
KEGG_PATHWAY	hsa04115:p53 signaling pathway	6.84E-18
KEGG_PATHWAY	hsa05212:Pancreatic cancer	5.73E-17
KEGG_PATHWAY	hsa05215:Prostate cancer	2.33E-16
KEGG_PATHWAY	hsa05222:Small cell lung cancer	8.51E-15
KEGG_PATHWAY	hsa04210:Apoptosis	1.47E-14
KEGG_PATHWAY	hsa04510:Focal adhesion	1.53E-14
KEGG_PATHWAY	hsa05220:Chronic myeloid leukemia	1.33E-13
KEGG_PATHWAY	hsa05210:Colorectal cancer	4.11E-13
KEGG_PATHWAY	hsa04012:ErbB signaling pathway	6.88E-13

FDR = false-discovery rate.

Gene-regulatory network construction

According to the 2252 identified breast cancer-related genes and 59 transcription factors with known expression values, we constructed the gene-regulatory networks under 3 conditions including normal, benign, and malignant. The gene-regulatory network under the normal condition consisted of 59 transcription factors, 1618 target genes, and 2512 regulation pairs (Figure 1). The regulatory network under the benign condition contained a total of 59 transcription factors, 1687 target genes, and 2307 regulation pairs (Figure 2). The regulatory network under the malignant condition consisted of 59 transcription factors, 1766 target genes, and 2419 regulation pairs in all (Figure 3).

Comparison of network topological properties

As shown in Table 3, the difference in gene-regulatory networks between the normal state and malignant breast cancer was the most significant, and the difference between normal and benign breast cancer was less significant than that between benign and malignant breast cancer.

Gene ranking

We selected 53 genes at the hub (level ≥10) location for the 3 conditions (normal, benign, malignant). Moreover, in the gene ranking conducted according to different degrees

of genes between 2 states, we found that the degrees of hub genes were clearly changed, and there were 8 hub genes in the top 10 genes with highly different degrees (Table 4, bold).
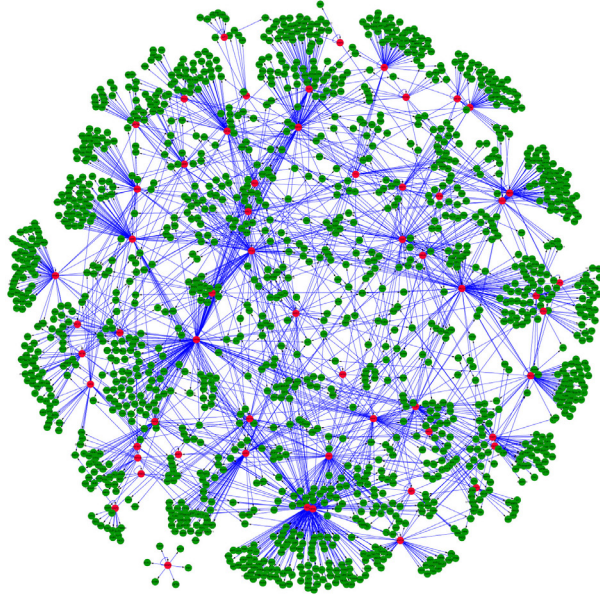
Figure 1. Gene-regulatory network in normal condition. Red nodes are transcription factors, green nodes are target genes.
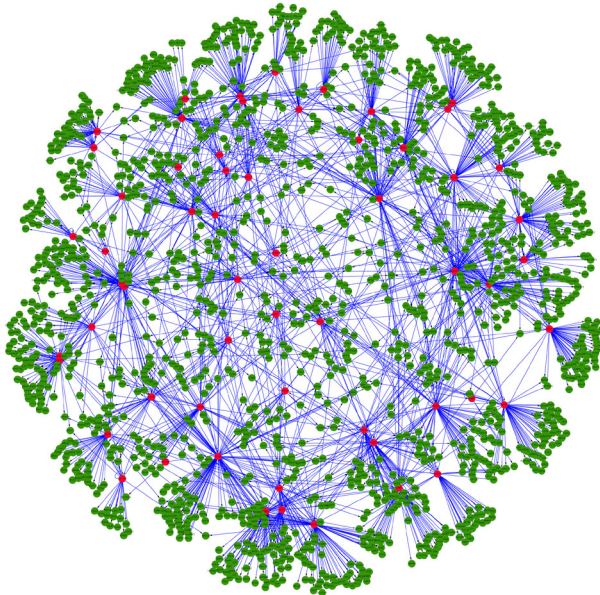
Figure 2. Gene-regulatory network in benign breast cancer samples. Red nodes are transcription factors, green nodes are target genes.
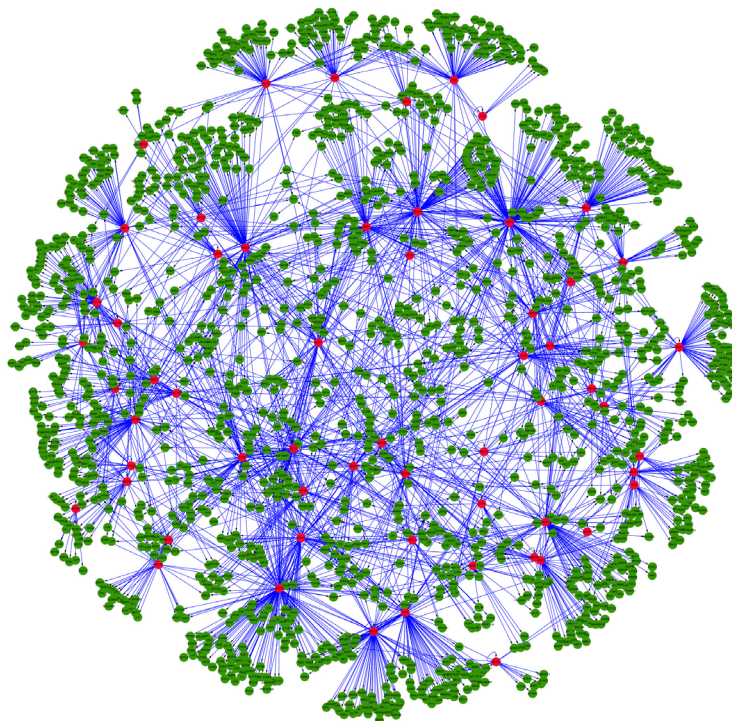
Figure 3. Gene-regulatory network in malignant breast cancer samples. Red nodes are transcription factors, green nodes are target genes.

Table 3. Rank sum test results of network topological properties.

	In-degree	Out-degree	Betweenness	Cluster coefficient	Clossness
Normal *vs* benign	0.6	0.8	0.04	0.03	<2.2e-16
Normal *vs* malignant	1.46E-07	0.64	0.19	0.01	<2.2e-16
Benign *vs* malignant	1.17E-06	0.8	0.38	0.7	<2.2e-16

Table 4. Top 10 hub genes by degree ranking.

Gene symbol	Normal degree	Benign degree	Malignant degree	*Diff degree*
YBX1	156	56	178	312
ZFP36	135	111	146	292
YY1	26	75	33	98
XRCC5	58	25	68	86
WISP3	5	45	6	80
XRCC4	94	73	111	76
ZFHX3	80	49	87	76
WNT5A	38	5	42	74
ZMAT3	57	94	71	74
XPC	55	41	77	72

$Diff\_\deg ree = \sum_{ij} abs(xi - xj), i, j \in (normal, benign, malignat)$. Bold = hub genes with highly different degrees.

DISCUSSION

Breast cancer is characterized by a distinct metastatic pattern involving regional lymph nodes, as well as the bone marrow, lung, and liver (Müller et al., 2001) and confers a considerable disease burden to individuals as well as an economic burden to society (Parkin and Fernández, 2006). Therefore, there is an urgent need to explore the pathogenesis of breast cancer for the development of an effective prevention strategy. Gene-regulatory network analysis allows for the systematic study of global network dynamics and permits the quantification of the relative influence and sensitivity of genes in their interactions with other genes (Shmulevich et al., 2002). A protein-network-based approach is used to identify markers involved in breast cancer metastasis, and screened markers are thought to be more reproducible than individual maker genes selected without network information; they also achieve higher accuracy in the classification of metastatic versus non-metastatic tumors (Chuang et al., 2007). In their study on microRNAs (miRNAs) and their target gene networks, O'Day and Lal (2010) described an association between alternating miRNA signatures and breast cancer tumorigenesis and metastasis. However, there have been no studies comparing the gene expression and gene-regulatory networks under the normal state and benign and malignant breast cancers.

In this study, breast cancer-related genes were first screened and then verified by using function and pathway annotation. Next, the gene-regulatory networks for the 3 conditions, including normal, benign breast cancer, and malignant breast cancer, were constructed based on related genes and transcription factors. Network topological properties were compared to explore more subtle discrepancies. Finally, related genes were ranked according to different degrees between 2 disease states, and 8 genes were found to play an important role in the entire process of breast cancer.

Based on DAVID-enrichment analysis, breast cancer-related genes were mainly enriched in biological processes associated with carcinogenesis, such as apoptosis, and in the p53 signaling pathway. Apoptosis repressor with caspase recruitment domain (ARC) promotes breast tumorigenesis, metastasis, and chemoresistance (Medina-Ramirez et al., 2011). The p53 pathway reduces cancer initiation by inducing apoptosis or cell cycle arrest in response to a variety of stress signals (Carson and Lois, 1995).

According to gene-regulatory networks constructed for the 3 conditions, including normal and benign and malignant cancer, there were 59 transcription factors, but different corresponding target genes and regulation relationships among them. For example, the transcription factor Wnt-induced-secreted-protein-1 (WISP1) regulated the *RUNX2*, *PIM1*, *MAST1*, *LMTK3*, *KLK12,* and *KCNB1* genes under normal conditions, while WISP1 regulated the *PDE5A* and *BHMT* genes under malignant conditions. However, in the benign condition, WISP1 regulated many different genes, including *TSG101*, *TIAM1*, and *SCGB2A2*. WISP-1 is a cysteine-rich, secreted factor belonging to the CCN family of growth factors (Soon et al., 2003), and its expression in some cells results in transformation and tumorigenesis (Su et al., 2002).

By conducting a comparative analysis of the network topological properties of gene-regulatory networks for the 3 conditions, differences in gene-regulatory networks between the normal state and the malignant breast cancer state were more significant than those between the normal state and the benign breast cancer state. This indicates that a large transition occurs from the benign to malignant breast cancer condition, which is consistent with the medical definition. We can infer that the double-two-way *t*-test used for the gene-regulatory construction was reliable.

Based on gene ranking analysis, we screened 8 hub genes (*YBX1*, *ZFP36*, *YY1*, *XRCC5*, *XRCC4*, *ZFHX3*, *ZMAT3*, and *XPC*) among the top 10 genes in the change in degree rank, which are thought to play an important role in the entire process of breast cancer. The Y box binding protein-1 gene (*YBX-1*) is an oncogene that is overexpressed in cancers (Dolfini and Mantovani, 2013) and plays a significant role in the processes of proliferation, apoptosis, and control of tumor cell response to toxic agents, including chemotherapy. Its high nuclear expression is associated with poor survival of patients with early-stage breast cancer, indicating its potential as a prognostic factor (Maciejczyk et al., 2012). Zinc finger protein 36 (ZFP36) is a tandem CCCH zinc-finger RNA-binding protein that regulates the stability of certain AU-rich element mRNAs involved in breast cancer-related processes (Al-Souhibani et al., 2010). *ZFP36* gene defects will cause an imbalance in the cellular steady state, leading to over-expression of cancer genes and cancer cell metastasis (Al-Ahmadi et al., 2013). Yin Yang 1 (YY1) is highly expressed in various types of cancers and regulates tumorigenesis through multiple pathways. Depletion of YY1 inhibited the migration, invasion, and tumor formation of breast cancer cells (Wan et al., 2012). Additionally, activation of the YY1 transcription factor is associated with the migration of breast cancer (Siletz et al., 2013). Thus, these genes may be involved in the pathogenesis of breast cancer. In conclusion, our data provide a comprehensive bioinformatic analysis of the genes and networks involved in the progression of breast cancer. Through comparative analysis of gene-regulatory networks, we identified the link between related genes and the pathogenesis of breast cancer. However, further experiments are needed to confirm our results.

REFERENCES

Abba MC, Lacunza E, Butti M and Aldaz CM (2010). Breast cancer biomarker discovery in the functional genomic age: a systematic review of 42 gene expression signatures. *Biomark. Insights* 5: 103-118.

Ahmed J, Meinel T, Dunkel M, Murgueitio MS, et al. (2011). CancerResource: a comprehensive database of cancer-relevant proteins and compound interactions supported by experimental knowledge. *Nucleic Acids Res.* 39: D960-D967.

Al-Ahmadi W, Al-Ghamdi M, Al-Souhibani N and Khabar KS (2013). miR-29a inhibition normalizes HuR over-expression and aberrant AU-rich mRNA stability in invasive cancer. *J. Pathol.* 230: 28-38.

Al-Souhibani N, Al-Ahmadi W, Hesketh JE, Blackshear PJ, et al. (2010). The RNA-binding zinc-finger protein tristetraprolin regulates AU-rich mRNAs involved in breast cancer-related processes. *Oncogene* 29: 4205-4215.

Aleskandarany MA, Negm OH, Rakha EA, Ahmed MA, et al. (2012). TOMM34 expression in early invasive breast cancer: a biomarker associated with poor outcome. *Breast Cancer Res. Treat.* 136: 419-427.

Araki R, Seno S, Takenaka Y and Matsuda H (2013). An estimation method for a cellular-state-specific gene regulatory network along tree-structured gene expression profiles. *Gene* 518: 17-25.

Bamford S, Dawson E, Forbes S, Clements J, et al. (2004). The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br. J. Cancer* 91: 355-358.

Barrett T, Suzek TO, Troup DB, Wilhite SE, et al. (2005). NCBI GEO: mining millions of expression profiles - database and tools. *Nucleic Acids Res.* 33: D562-D566.

Boccaletti S, Latora V, Moreno Y, Chavez M, et al. (2006). Complex networks: structure and dynamics. *Physics Rep.* 424: 175-308.

Carson DA and Lois A (1995). Cancer progression and p53. *Lancet* 346: 1009-1011.

Chuang HY, Lee E, Liu YT, Lee D, et al. (2007). Network-based classification of breast cancer metastasis. *Mol. Syst. Biol.* 3: 140.

Coughlin SS and Ekwueme DU (2009). Breast cancer as a global health concern. *Cancer Epidemiol.* 33: 315-318.

Csardi G and Nepusz T (2006). The Igraph Software Package for Complex Network Research. InterJournal, Complex Systems, 1695.

Dennis G Jr, Sherman BT, Hosack DA, Yang J, et al. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* 4: 3.

Dolfini D and Mantovani R (2013). YB-1 (YBX1) does not bind to Y/CCAAT boxes *in vivo*. *Oncogene* 32: 4189-4190.

Fritz A, Sinha S, Marella N and Berezney R (2013). Alterations in replication timing of cancer-related genes in malignant human breast cancer cells. *J. Cell Biochem.* 114: 1074-1083.

Harris MA, Clark J, Ireland A, Lomax J, et al. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* 32: D258-D261.

Jemal A, Murray T, Ward E, Samuels A, et al. (2005). Cancer statistics, 2005. *CA Cancer J. Clin.* 55: 10-30.

Jiao X, Wood LD, Lindman M, Jones S, et al. (2012). Somatic mutations in the Notch, NF-KB, PIK3CA, and Hedgehog pathways in human breast cancers. *Genes Chromosomes Cancer* 51: 480-489.

Joy MP, Brock A, Ingber DE and Huang S (2005). High-betweenness proteins in the yeast protein interaction network. *J. Biomed. Biotechnol.* 2005: 96-103.

Kanehisa M (2002). The KEGG database. *Novartis Found. Symp.* 247: 91-101.

Karolchik D, Baertsch R, Diekhans M, Furey TS, et al. (2003). The UCSC Genome Browser Database. *Nucleic Acids Res.* 31: 51-54.

LaBreche HG, Nevins JR and Huang E (2011). Integrating factor analysis and a transgenic mouse model to reveal a peripheral blood predictor of breast tumors. *BMC Med. Genomics* 4: 61.

Lin M, Zhou X, Shen X, Mao C, et al. (2011). The predicted *Arabidopsis* interactome resource and network topology-based systems biology analyses. *Plant Cell* 23: 911-922.

Ma HW and Zeng AP (2003). The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics* 19: 1423-1430.

Ma J and Jemal A (2013). Breast Cancer Statistics. Breast Cancer Metastasis and Drug Resistance. Springer, New York, 1-18.

Maciejczyk A, Szelachowska J, Ekiert M, Matkowski R, et al. (2012). Elevated nuclear YB1 expression is associated with poor survival of patients with early breast cancer. *Anticancer Res.* 32: 3177-3184.

Medina-Ramirez CM, Goswami S, Smirnova T, Bamira D, et al. (2011). Apoptosis inhibitor ARC promotes breast tumorigenesis, metastasis, and chemoresistance. *Cancer Res.* 71: 7705-7715.

Morrison DH, Rahardja D, King E, Peng Y, et al. (2012). Tumour biomarker expression relative to age and molecular subtypes of invasive breast cancer. *Br. J. Cancer* 107: 382-387.

Müller A, Homey B, Soto H, Ge N, et al. (2001). Involvement of chemokine receptors in breast cancer metastasis. *Nature* 410: 50-56.

O'Day E and Lal A (2010). MicroRNAs and their target gene networks in breast cancer. *Breast Cancer Res.* 12: 201.

Parkin DM and Fernández LM (2006). Use of statistics to assess the global burden of breast cancer. *Breast. J.* 12 (Suppl 1): S70-S80.

Qi J and Michoel T (2012). Context-specific transcriptional regulatory network inference from global gene expression maps using double two-way *t*-tests. *Bioinformatics* 28: 2325-2332.

Shmulevich I, Dougherty ER, Kim S and Zhang W (2002). Probabilistic Boolean Networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics* 18: 261-274.

Siletz A, Kniazeva E, Jeruss JS and Shea LD (2013). Transcription factor networks in invasion-promoting breast carcinoma-associated fibroblasts. *Cancer Microenviron.* 6: 91-107.

Soon LL, Yie TA, Shvarts A, Levine AJ, et al. (2003). Overexpression of WISP-1 down-regulated motility and invasion of lung cancer cells through inhibition of Rac activation. *J. Biol. Chem.* 278: 11465-11470.

Su F, Overholtzer M, Besser D and Levine AJ (2002). WISP-1 attenuates p53-mediated apoptosis in response to DNA damage through activation of the Akt kinase. *Genes Dev.* 16: 46-57.

Thomas JG, Olson JM, Tapscott SJ and Zhao LP (2001). An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles. *Genome Res.* 11: 1227-1236.

van Triest HJ, Chen D, Ji X, Qi S, et al. (2011). PhenOMIM: an OMIM-based secondary database purported for phenotypic comparison. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2011: 3589-3592.

Wan M, Huang W, Kute TE, Miller LD, et al. (2012). Yin Yang 1 plays an essential role in breast cancer and negatively regulates p27. *Am. J. Pathol.* 180: 2120-2133.