Original Article
Protein-protein interaction (PPI) network and significant gene analysis of breast cancer

Weihua Ren[1], Yawei Li[2], Shuangting Wu[3], Hongbo Feng[4], Rui Li[5]

[1]Central Laboratory, Luoyang Central Hospital Affiliated to Zhengzhou University, Luoyang, He'nan 471009, China; Departments of [2]Orthopedics, [3]Anesthesiology, [4]Otorhinolaryngology, [5]Neurosurgery, Luoyang Central Hospital Affiliated to Zhengzhou University, Luoyang, He'nan 471009, China

Abstract: The incidence of breast cancer is one of the highest female malignant tumors. While, the early diagnosis and treatment of breast cancer with microarray technology are requisite in breast cancer research. We aimed to identify new potential signaling pathways and key genes in breast cancer. The transcription profile of GSE54002 was downloaded from Gene Expression Omnibus (GEO) database, including 417 breast cancer and 16 healthy samples. The differentially expressed genes (DEGs) between cancer and healthy group were screened with non-paired t-test and analyzed by Cluster 3.0 software. We used the DAVID online tools to enrich the Gene Ontology function and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway of significantly up- and down-regulated genes. After construction of interaction network of proteins encoded by DEGs, the topological properties of networks and function modules were analyzed with Cytoscape. A total of 789 DEGs were identified in breast cancer samples compared to normal tissue samples, including 257 up-regulated and 532 down-regulated genes. In GO terms, the up-regulated genes are mainly related with cell cycle and interaction of extracellular matrix; While in KEGG pathways, up-regulated genes were enriched in cell cycle pathway and ECM-receptor interaction pathway. In addition, the transcription factor FOS and its multiple downstream regulatory factors were highly expressed in cancer tissue. The discovery of the DEGs with high expression in enrichment analysis might help understand the mechanism of breast cancer. Moreover, the key factors we predicted in development of breast cancer could provide references for the diagnosis and treatment of this disease.

Keywords: Breast cancer, differentially expressed gene (DEG), microarray, PPI network

Introduction

The incidence of breast cancer is in the most vicious female malignant tumor, and estimates only for year 2008 revealed 1.38 million new cases and 0.46 million deaths per year, which made it the most common cause of female cancer death both in the developed and developing world [1]. Most death cases were caused by the absence of early diagnosis and systemic treatment [2], especially in developing countries where over two-thirds of women diagnosed with breast cancer die from the disease [3]. But, early detection of breast cancer relies on mammography, which is currently not possible in many developing countries, because of high expense, the relative low incidence of breast cancer. The absence or scarce distribution of radiotherapy and systemic therapy services which are optimal treatment for breast cancer are highlighted in the developing world [4].

Therefore, there is an urgent need to develop the early diagnosis and prognostic and treatment markers of breast cancer for developing a cheaper and simpler diagnosis and therapy method.

Gene expression profiling has been providing a huge pool of candidate targets for cancer diagnosis and therapy [5]. Microarray technology is an effective tool to disclose the global molecular changes occurred at the onset and during the development of cancer [6]. Currently, large-scale microarray studies have been widely used in mining genes which are related with the occurrence, development and prognosis of disease, especially in tumor study, such as in ovarian cancer, colorectal cancer, and renal cell carcinoma, lung cancer [7, 8]. And scientists have succeeded in clarifying breast cancer into 5 molecular subtypes based on gene expression profiles and developing genomic biomarkers for

predicting recurrence in early breast cancer by using microarray technology [9, 10]. But only a small number of these genomic biomarkers have been validated to respond to therapeutic agents or provide prognostic information, which are called traditional biomarkers including estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 (HER2) [9]. Estrogen receptor was widely used to guide therapy. HER2 was identified as a prognostic marker [11]. However, these biomarkers are still deficient in methodologies. Thus, a better understanding of the molecular mechanisms causing breast cancer may provide new opportunities for development of diagnosis and therapy.

In this study, we intended to screen differentially expressed genes (DEGs) between malignant breast tissue and normal breast tissue samples by large-scale microarray technology. Here we highlight the up-regulated genes that are related with cell cycle and extracellular matrix interaction. These genes and the associated signaling pathways are considered as the novel targets for intervention of breast cancer progression.

Materials and methods

Affymetrix microarray data

The transcription profile of GSE54002 [12] was downloaded from GEO database of NCBI (http://www.ncbi.nlm.nih.gov/geo/) which was based on the Affymetrix Human Genome U133 plus 2.0 Array. Profile of GSE54002 contains total 433 breast tissue chips including 417 breast cancer patient samples and 16 samples of healthy people. Due to the cancer samples are far greater than the number of normal samples, 253 cancer samples were randomly selected for analysis.

Data preprocessing

We first converted the probe-level data into expression values by the robust multiarray average (RMA) algorithm with defaulted parameters in R affy package in Bionconductor [13]. Then we matched the probe name into gene name by GPL profiles in GEO. If part probes corresponded to plurality of gene, the expression values of those probes were integrated into the gene expression value. If there were multiple probes corresponded to the same gene,

the expression values of those probes were averaged. Total 21049 genes were selected. Median normalization was taken before difference analysis on sample expression value.

Differentially expressed genes (DEGs) analysis

We used non-paired t-test to identify the differentially expressed genes (DEGs) between cancer group and healthy group, and calculated the fold change. The *p*-value <0.01 and fold change value >1.2 times was used as the cut-off criterion for screening the differentially expressed genes. Cluster 3.0 software was used to analyze the differentially expressed genes, using hierarchical clustering method, and gene expression correlation coefficient as the distance, average connection. Then the Treeview software was used to draw the DEGs cluster map.

Functional enrichment analysis of DEGs

We used the DAVID (The Database for Annotation, Visualization and Integrated Discovery) online tools to enrich the Gene Ontology (GO) function and The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway of significantly up-regulated and down-regulated genes [14]. The FDR (false discovery rate) less than 0.05 was used as the cut-off criterion to screen GO categories and significantly enriched KEGG pathway.

DEGs network construction and network module mining

The STRING (Search Tool for the Retrieval of Interacting Genes) database provides both experimental and predicted interaction information [15]. We mapped the DEGs to the STRING database, and selected these interactions of reliability scoring more than 0.7 to get the interaction relationship of the DEGs. Then we drew the DEGs interactions network by using Cytoscape software, marked up-regulated and down-regulated genes [16]. Network Analyzer plug-in was used to analyze the topological properties of protein interaction network, including the distribution of node degree, clustering coefficient, and the shortest path and connecting centrality etc. Lastly, MINE plug-in was used to decompose the network into modules, and select the modules with the minimum module enrichment *P* value for functional analysis [17].
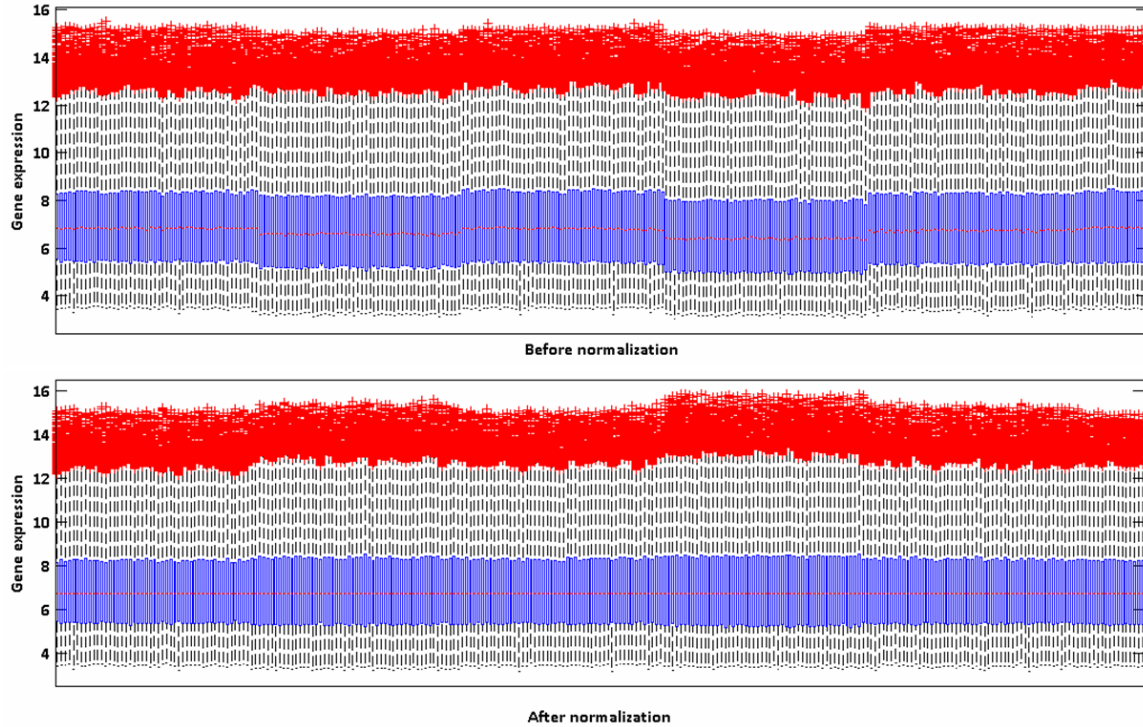
Figure 1. Box-plot of 269 sample expression values before and after normalization. Each vertical bar in the graph corresponds to a sample, the upper line of blue region is the 1/4 quartile of all gene expression values, red line represents the median, the bottom of blue region represents 3/4 quartile. Red star at the top of the box represents the specific expression value. After normalization, each sample has the same median that is shown as a red line in the box plot.

Table 1. The top 10 significant up-regulation and down-regulation genes

Up-regulated gene	P-value	Fold-change	Down-regulated gene	P-value	Fold-change
CTHRC1	9.44×10^{-38}	2.11	PAK7	6.79×10^{-98}	0.62
PTTG3P	5.37×10^{-32}	1.42	TSHZ2	2.16×10^{-86}	0.55
KIAA0101	3.04×10^{-27}	1.30	LOC400128	6.96×10^{-84}	0.63
FN1	3.09×10^{-26}	1.41	NPY2R	4.73×10^{-82}	0.70
DTL	9.52×10^{-26}	1.52	LPPR1	5.40×10^{-78}	0.66
UBE2C	2.66×10^{-25}	1.45	NPCDR1	1.12×10^{-77}	0.68
UHRF1	2.77×10^{-25}	1.44	KCNJ16	2.66×10^{-77}	0.61
SPC24	1.09×10^{-24}	1.38	CNTNAP3B	1.11×10^{-73}	0.67
TOP2A	2.62×10^{-24}	1.58	SDPR	2.61×10^{-73}	0.67
TPX2	2.88×10^{-24}	1.65	TNS4	8.73×10^{-73}	0.75

The significant is measured by the P-value of t-test, the smaller P-value, the more significant. Fold-change is the ratio of the cancer group sample mean value and normal group sample mean value.

Pathway enrichment analysis of interaction network

We downloaded the gene transcription signal data and transcription regulation data from Path PPI database (http://proteomeview.hupo. org.cn/PathPPI/PathPPI.html). The gene transcription regulation and signal transduction data in Path PPI database came from 7 manual collection pathways databases including PID and Biocarta, which had higher credibility than traditional protein-protein interaction (PPI). Then we mapped the DEGs into Path PPI interaction, constructed the signal transduction and transcriptional regulatory network. Then the important modules in this network were analyzed by Cytoscape [16].

Results

Microarray data analysis

The original data were preprocessed by Affymetrix package. We obtained 54,676 probe expression values and integrated them to 21,049 genes. The gene expression value of
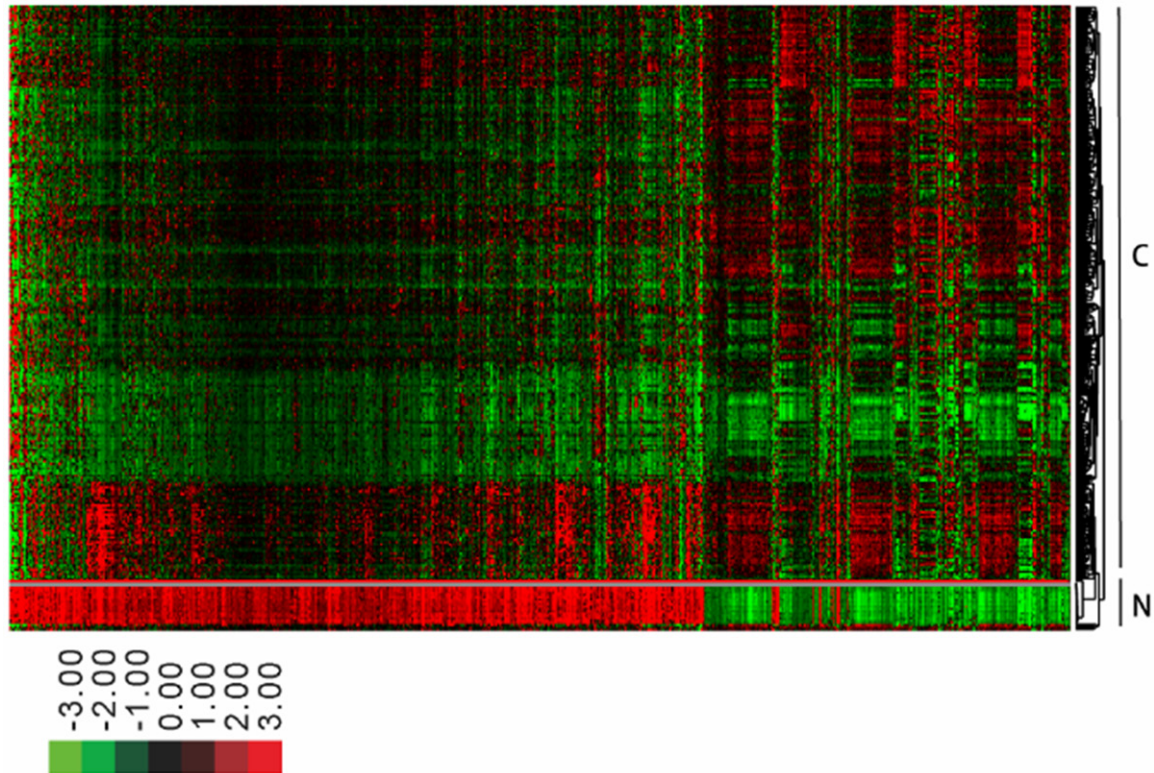
Figure 2. Bilateral cluster analysis of expression values of DEGs. The horizontal axis is the 789 differentially expressed genes, the vertical axis is 253 cancer samples and 16 normal tissue samples. Hierarchical cluster in Cluster 3.0 was used to obtain bilateral cluster. In hierarchical cluster, the gene expression correlation represents the distance between samples. Colors indicate the difference between grouped gene expression value and total gene expression mean value. The red indicates up-regulation, and the green indicates down-regulation. The cancer samples and normal samples form into two clusters. N is the normal tissue sample. C is the cancer tissue samples.

each sample had the same median after median normalization. The sample expression value scale errors were eliminated (**Figure 1**). Then we could compare the normal and cancer samples.

Differentially expressed genes (DEGs) analysis

Non paired t-test was used to identify the differentially expressed genes (DEGs) between cancer group and healthy group. Finally, 789 DEGs were obtained, including 257 up-regulated and 532 down-regulated genes in cancer samples compared to normal control. The top 10 significant up-regulated and down-regulated genes were shown in **Table 1**. Through bilateral cluster analysis (**Figure 2**), the selected DEGs could well identify the cancer sample with the healthy tissue sample, supporting that those identified genes are reliable DEGs in breast cancer. In the cluster figure, we also found that the 16 normal tissue samples had more stable gene expression levels than cancer samples.

Gene ontology (GO) and KEGG pathway enrichment of DEGs

A total of 257 up-regulated genes were mainly in cell cycle related GO function enrichment, including GO: 0007049 (cell cycle, $P = 2.30 \times 10^{-29}$, FDR = 3.78×10^{-26}) and GO: 0000279 (M phase, $P = 4.10 \times 10^{-35}$, FDR = 6.74×10^{-3}). Among 532 down-regulated genes, the majority of them were enriched in GO: 0007398 (ectoderm development, $P = 7.04 \times 10^{-7}$, FDR = 1.2×10^{-2}). With higher significances in GO type statistics according to the *P*-values, up-regulated genes might play more important roles in breast cancer development.

In the KEGG pathway analysis, down-regulated genes were not enriched in a particular pathway. While up-regulated genes were enriched in cell cycle pathway ($P = 1.48 \times 10^{-7}$, FDR = 1.44×10^{-4}) and ECM-receptor interaction pathway ($P = 1.99 \times 10^{-8}$, FDR = 1.93×10^{-5}), which functions in cell-cell surface interaction. These two pathways covered CDK1, CDC6, PKMYT1,
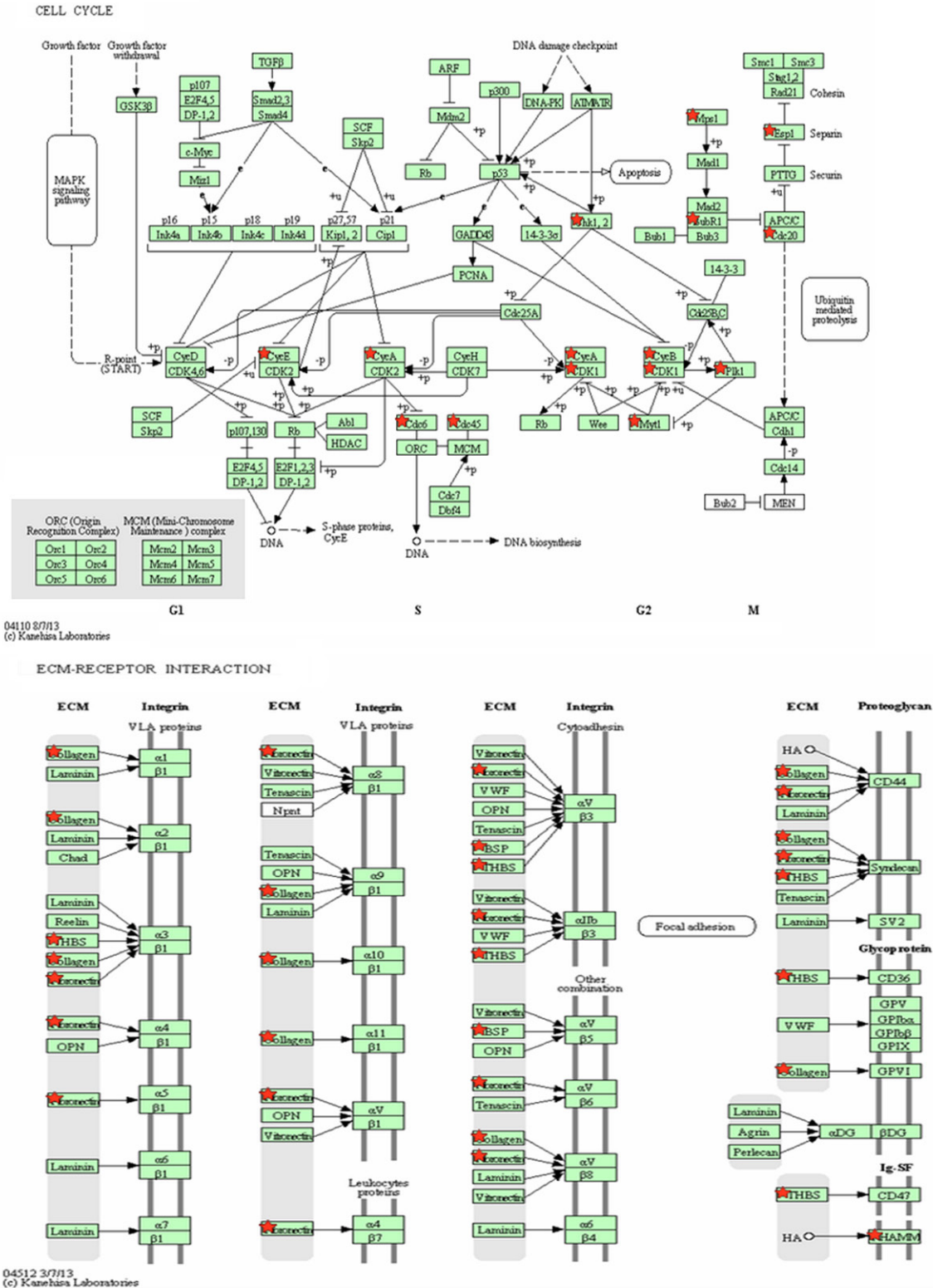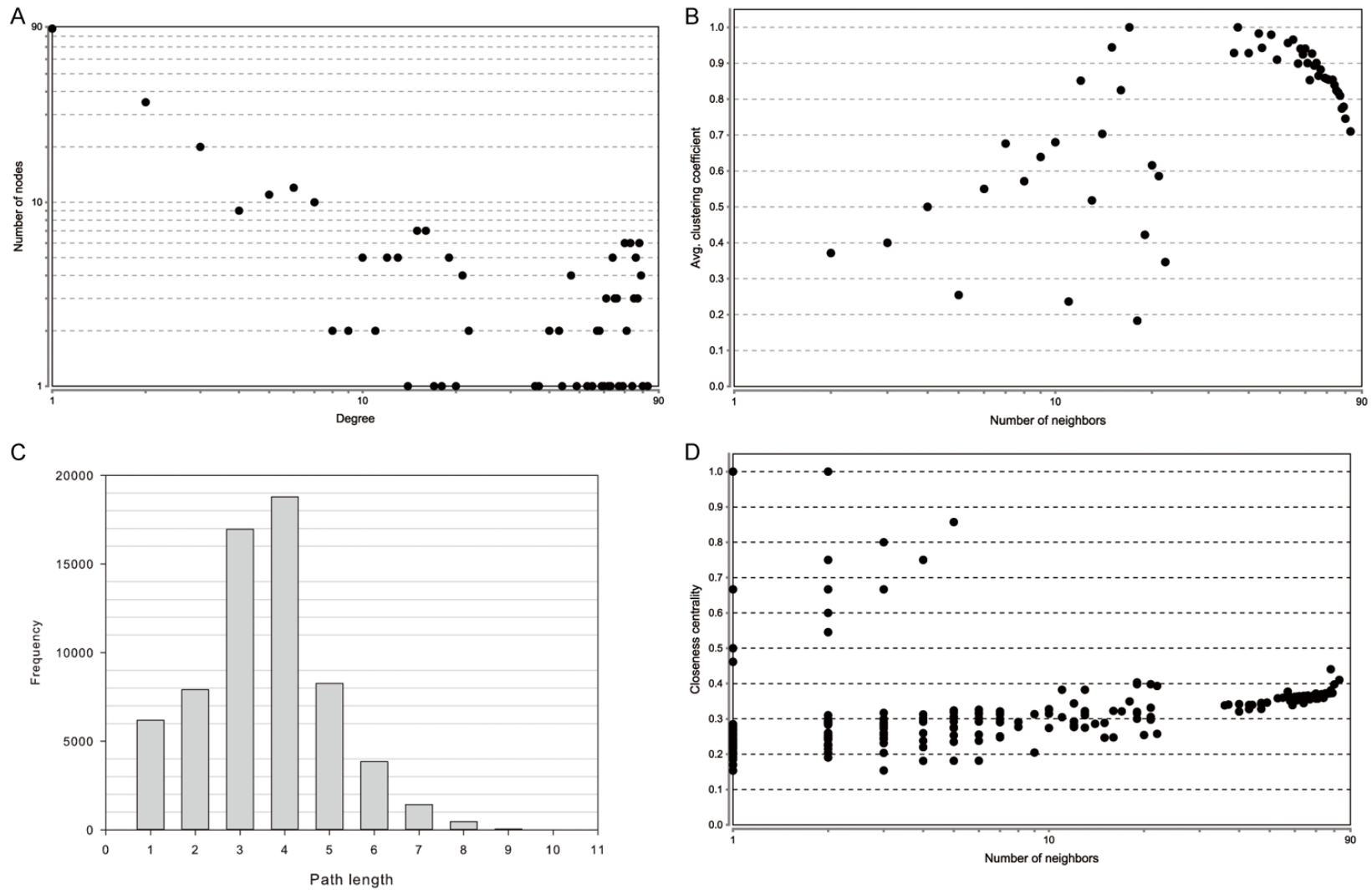
Figure 3. Two Pathways of Up-regulated Genes in KEGG pathway Enrichment. Cell cycle pathway and ECM-receptor interaction pathway in KEGG database. Red stars indicate the genes from the list of up-regulated genes.

TTK, CDC20, ESPL1, CHEK1, CCNB1, CCNE2, PLK1, BUB1B and CCNA (**Figure 3**), indicating these DEGs might be potential early diagnostic markers for breast cancer.

Figure 4. Topological Properties of DEGs Network. The node degree of network of DEGs obeys power-law distribution, average aggregation coefficient shows the characteristics that the bigger the degree of connection the higher clustering coefficient. The shortest pathway is concentrated in 3-4. The closeness centrality of node collection is proportional to node degree, but the proportion coefficient is very low.

Table 2. 10 modules with more than 5 nodes

Module	Score	Protein Number	Interactions Number
1	59.05	87	2,539
2	15.85	17	127
3	9.69	14	63
4	7.40	11	37
5	6.33	7	19
6	5.00	5	10
7	5.00	7	15
8	4.67	7	14
9	4.40	6	11
10	3.50	5	7

Score = density of module connection ×protein number in module.

Topological properties of DEGs network

We mapped the DEGs to the STRING database and obtained an interaction network of DEGs composing of 313 DEGs and 3093 sides. The node degree of the network followed the power-law distribution (**Figure 4A**). There were a larger proportion of high degree nodes, especially connectivity in the range of 70-80. This big variance might due to the formation of larger high connection density module. The aggregation coefficients of high connectivity protein were bigger than 0.7 (**Figure 4B**) which indicated high node aggregation. The peak of the shortest pathway concentrated in 3-4 (**Figure 4C**) and the closeness centrality of node connection distributed in the 0.2-0.4 (**Figure 4D**) indicated the good node connectivity and close function relation of DEGs.

Module mining and function analysis for DEGs

We used MINE plugin in Cytoscape software to select the interaction network of DEGs in which there were more than 5 gene sets and obtained 10 modules (**Table 2**).

The first module contained 87 proteins, which formed the biggest sub-network (Figure S1) and were enriched in cell cycle related pathway (**Figure 3**). The second module containing 17 DEGs (Figure S2) was related to chemokine signal pathway. The third module covered 14 DEGs (Figure S3) was involved in extracellular mechanism (**Table 3**). The other modules were not mined pathway enrichment in the KEGG database.

Signal transduction and transcriptional regulatory network of DEGs

We mapped the DEGs into Path PPI interaction data to construct the signal transduction network and transcriptional regulatory network. This transcriptional regulatory network contained 35 DEGs and 37 interaction relationships. The signal transduction network could be divided into two large sub-networks (**Figure 5A** and **5B**), and the transcriptional regulatory network included two sub-network (**Figure 5C** and **5D**). Genes in **Figure 5A** were mainly related to cytokines. Genes in **Figure 5D** were up-regulated by transcriptional regulatory factor FOS. The FOS is the cancer gene, and the expression level of c-FOS is considered an independent prognostic marker of breast cancer [18], so the up-regulated factor (KRT5, KRT17, GRIA2, PTGS2, FIGF and CXCL2) in downstream regulated by the FOS are worthy of emphatically researching.

Discussion

In this study, we identified 257 up-regulated genes and 532 down-regulated genes in 253 breast cancer samples compared to 16 normal tissue samples. And the 257 up-regulated genes were mainly enriched in cell cycle related GO function, and in ECM-receptor interaction pathway and Cell cycle in KEGG database. The ECM-receptor interaction pathway indicated the abnormal expression of extracellular matrix receptor in cancer tissue which might be relative to cancer metastasis. Through analysis, we found that the expression of cell cycle related genes in cancer tissue changed most, which was consistent with the rapid propagation characteristics of cancer.

The abnormal expression genes which were related with extracellular matrix interaction function indicated the universality of risk of cancer metastasis in breast cancer. The abnormal expression of chemokine signaling pathway might be associated with inflammatory response. Since inflammation had either promotive or inhibitive effect on different stages of the cancer [19, 20], it was not strange that inflammation-related DEGs accounted for a high percentage [21]. Therefore, genes related with the inflammation process were worthy of further research, including CDK1, CDC6, PKMYT1, TTK, CDC20, ESPL1, CHEK1, CCNB1,

Table 3. Pathway enrichment of the first 3 modules in KEGG database

Module	KEGG pathway	Gene number	Gene	P-value	FDR
Module 1	hsa04110 Cell cycle	14	CDK1,CDC6, CDC14A, PKMYT1, TTK, CDC20, ESPL1, CHEK1, CCNB1, CCNE2, CDC45, PLK1, BUB1B, CCNA2	2.59×10^{-16}	1.55×10^{-13}
Module 2	hsa04062 Chemokine signaling pathway	8	CXCL13, CXCL3, CXCL2, CX3CR1, CXCL9, CXCL11, CCL28, CXCL10	5.52×10^{-8}	3.44×10^{-5}
Module 3	hsa04512 ECM-receptor interaction	8	COL6A3, COL3A1, COL1A2, COL1A1, COL5A2, THBS2, COL11A1, COL5A1	2.61×10^{-13}	8.21×10^{-11}
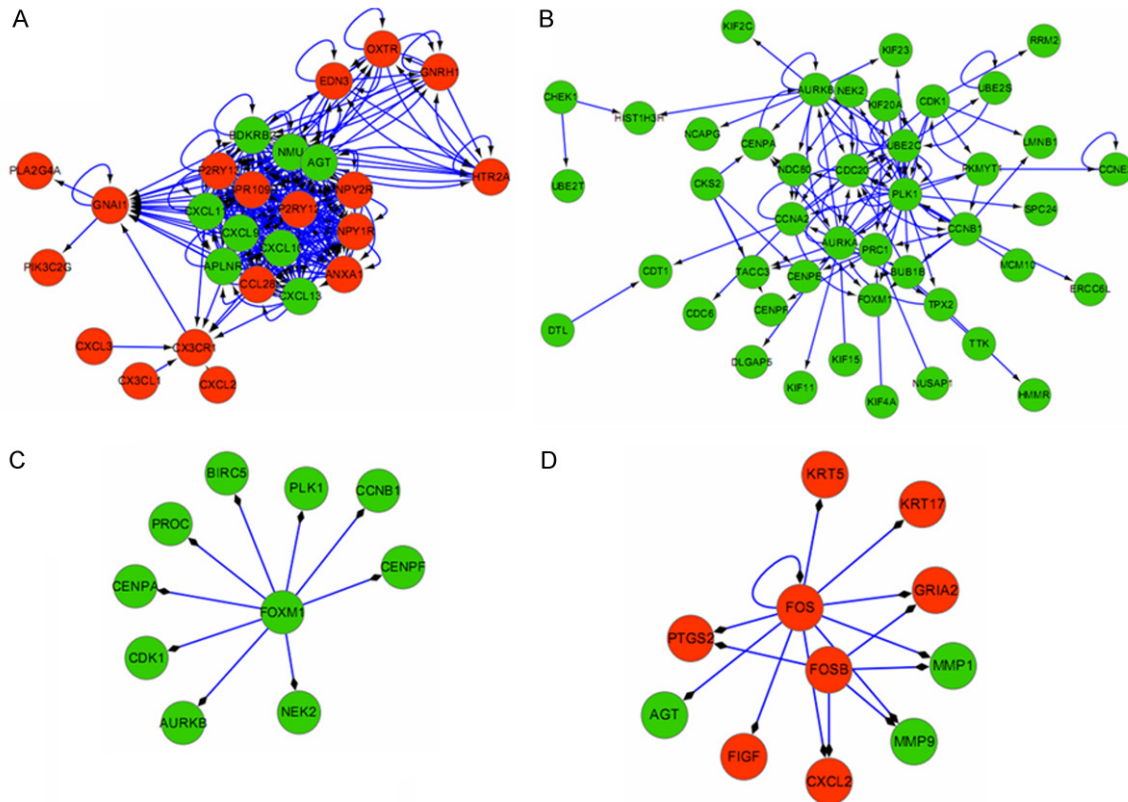
Figure 5. The main module of signal transduction and transcriptional regulation. A, B are from signal transduction network of DEGs. C, D are from transcriptional regulatory network of DEGs. Red represents up-regulation. Green shows down-regulation. The arrow indicates the direction of gene regulation. Transcription regulation and signal transduction are distinguished by two different arrows. Gene in the network of signal transduction show high aggregation. Transcriptional regulation shows star shaped regulation; the transcription factors regulate downstream genes.

CCNE2, CDC45, PLK1, BUB1B and CCNA2. In studies on cyclin dependent kinase (CDK1), approximately 75 targets of CDK1 have been identified to control critical cell cycle events, such as DNA replication and segregation, transcriptional programs and cell morphogenesis [22]. Cell division cycle 6 (CDC6), CDC20, CDC45 were recruited to take part in regulation of the cell cycle [23]. Overexpression of CDC6, leaded to re-replication, a form of replication stress, fuelling genomic instability, and promot-

ing malignant behavior [24-27]. CDC20 was a highly conserved activator of the anaphase-promoting complex (APC), promoting cell cycle regulated ubiquitination and proteolysis [28]. CDC45 was an essential factor required for the establishment and progression of the DNA replication fork in cycling cells and was more abundant in proliferating cells [29]. CHEK1 was composed of the ATR-CHEK1 DNA damage response pathway which was the key for maintenance of genome stability [30]. CCNB1-dependent Cdc2

kinase triggered the progression of cells in the G2 phase to M phase during a normal cell cycle [31]. And using nocodazole for treatment of MCF-7 human breast cancer cells could strongly up-regulate cyclin B1 and Cdc2 levels [32]. Polo-like kinase 1 (PLK1) was a regulator of many cell cycle-related events, had a close correlation with carcinogenesis and have been proposed as a novel diagnostic marker for cancer [33]. BUB1B was a critical mitotic checkpoint kinase. It has been validated that BUB1B overexpression might be a new immunohistochemical biomarker of malignancy in histologically normal breast tissues [34]. Interestingly, in this study the top ten up-regulated genes including CTHRC1, KIAA0101, FN1, DTL, UBE2C, UHRF1, SPC24, TOP2A and TPX2 were reported to play key roles in cancer development. The first one of top ten up-regulation gene CTHRC1 has been reported in cancers of gastrointestinal tract, lung, breast, thyroid, ovarian, cervix, liver, and the pancreas which was associated with cancer tissue invasion and metastasis [35]. KIAA0101 was involved in the regulation of DNA repair and cell proliferation, cell cycle progression, and migration. Aberrant expression of KIAA0101 in breast cancer was able to protect cell from UV-induced cell death [36]. The KEGG pathway analysis discovered that FN1 were significantly enriched in focal adhesion, extracellular matrix (ECM)-receptor interaction, and pathways in cancer. It has been shown that down-regulation of FN1 was able to suppress the migration and invasion [37]. UBE2C was associated with the selected metastasis-related genes VEGF, CXCL-4, CCL5, NEDD9 and RHoC, so UBE2C might be involved in breast cancer metastasis and been considered a potential biomarker candidate or therapeutic target for early breast cancer [38]. For diagnosing and therapeutic application, these candidates required further study to confirm their function.

Moreover, we mapped the DEGs into PathPPI interaction database to construct the signal transduction network and transcriptional regulatory network. Up-regulated genes in signal transduction network were mainly related to cytokines. Genes which were regulated by transcriptional regulatory factor FOS showed upregulation in transcriptional regulatory network. The FOS was a known oncogene, and the expression level of c-FOS was considered as an independent prognostic marker of breast cancer. Similarly, we found the FOS transcriptional factor and its downstream up-regulated factors (KRT5, KRT17, GRIA2, PTGS2, FIGF, CXCL2 etc.) over-expressed in breast cancer tissues. Specifically, the KRT5 positive cell induced by progesterone which played a deleterious role in the onset of breast cancer possessed the characteristics of stem cells in breast cancer [39], indicating the KRT5 could be the cell marker of antineoplastic drug. PTGS2 involved in the biosynthesis of prostaglandins in inflammation and hyperplasia process of breast cancer [40]. The expression pattern of PTGS2 was associated with the differential degree and prognosis of ovarian cancer and not related with the pathogenesis of the new model [41]. And aspirin and ibuprofen, nonselective PEGS2 inhibitors, used on a regular basis, significantly reduced the risk of human breast cancer [42]. FIGF (also called VEGF) was produced in blood vessel and grown actively in lymphatic endothelial cell, so the FIGF had pretty much concern with the cancer metastasis [43]. And simultaneous inhibition of FIGF pathways via the dual targeting agent EVRi might benefit select subsets of ovarian cancer tumors [44]. CXCL2 was produced by endothelial cells, macrophages, epithelial cells and tumor cells [45]. The interaction of CXCL2, with endothelial cell expressed CXCR2, caused angiogenesis increasing which link to tumor growth [46]. The high expression of CXCL2 in inflammatory tissue could inhibit the proliferation of hematopoietic cells and invasiveness of breast cancer cells [47, 48]. Therefore, we could focus on the downstream genes of FOS regulation which were associated with tumor, inflammation and angiogenesis etc. for follow-up research.

Conclusions

In conclusion, in breast cancer tissue, the high expression of genes related with cell cycle and extracellular matrix interaction lead to abnormal breast cancer cycle and accompanied by cancer metastasis. The transcription factor FOS and its multiple downstream regulatory factors which significantly higher express in cancer tissue were the key factor in development of breast cancer.

Disclosure of conflict of interest

None.

Address correspondence to: Rui Li, Department of Neurosurgery, Luoyang Central Hospital Affiliated to Zhengzhou University, Zhongzhou Middle Road No. 288, Luoyang, He'nan 471009, China. Tel: 860371118114; Fax: 860371118114; E-mail: medbio1@163.com

References

[1] Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 2010; 127: 2893-917.

[2] Howell A. The emerging breast cancer epidemic: early diagnosis and treatment. Breast Cancer Res 2010; 12 Suppl 4: S10.

[3] Rastogi T, Hildesheim A and Sinha R. Opportunities for cancer epidemiology in developing countries. Nat Rev Cancer 2004; 4: 909-17.

[4] Barton MB, Frommer M and Shafiq J. Role of radiotherapy in cancer control in low-income and middle-income countries. Lancet Oncol 2006; 7: 584-95.

[5] Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang CH, Angelo M, Ladd C, Reich M, Latulippe E, Mesirov JP, Poggio T, Gerald W, Loda M, Lander ES, Golub TR. Multiclass cancer diagnosis using tumor gene expression signatures. Proc Natl Acad Sci U S A 2001; 98: 15149-54.

[6] DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen Y, Su YA, Trent JM. Use of a cDNA microarray to analyse gene expression patterns in human cancer. Nat Genet 1996; 14: 457-60.

[7] Ying H, Lv J, Ying T, Li J, Yang Q. Screening of feature genes of the ovarian cancer epithelia with DNA microarray. J Ovarian Res 2013; 6: 39.

[8] Lee JM, Yoo JK, Yoo H, Jung HY, Lee DR, Jeong HC, Oh SH, Chung HM, Kim JK. The novel miR-7515 decreases the proliferation and migration of human lung cancer cells by targeting c-Met. Mol Cancer Res 2013; 11: 43-53.

[9] Dowsett M and Dunbier AK. Emerging biomarkers and new understanding of traditional markers in personalized therapy for breast cancer. Clin Cancer Res 2008; 14: 8019-26.

[10] Weigelt B, Baehner FL and Reis-Filho JS. The contribution of gene expression profiling to breast cancer classification, prognostication and prediction: a retrospective of the last decade. J Pathol 2010; 220: 263-80.

[11] Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science 1987; 235: 177-82.

[12] Tan TZ, Miow QH, Miki Y, Noda T, Mori S, Huang RY, Thiery JP. Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. EMBO Mol Med 2014; 6: 1279-93.

[13] Gautier L, Cope L, Bolstad BM, Irizarry RA. affy-analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 2004; 20: 307-15.

[14] Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol 2003; 4: P3.

[15] Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ, von Mering C. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res 2011; 39: D561-8.

[16] Kohl M, Wiese S and Warscheid B. Cytoscape: software for visualization and analysis of biological networks. Methods Mol Biol 2011; 696: 291-303.

[17] Rhrissorrakrai K and Gunsalus KC. MINE: Module Identification in Networks. BMC Bioinformat 2011; 12: 192.

[18] Mahner S, Baasch C, Schwarz J, Hein S, Wölber L, Jänicke F, Milde-Langosch K. C-Fos expression is a molecular predictor of progression and survival in epithelial ovarian carcinoma. Br J Cancer 2008; 99: 1269-75.

[19] Hanahan D and Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011; 144: 646-74.

[20] Moss SF and Blaser MJ. Mechanisms of disease: inflammation and the origins of cancer. Nat Clin Pract Oncol 2005; 2: 90-7; quiz 1 p following 113.

[21] Jiang X and Shapiro DJ. The immune system and inflammation in breast cancer. Mol Cell Endocrinol 2014; 382: 673-82.

[22] Enserink JM and Kolodner RD. An overview of Cdk1-controlled targets and processes. Cell Div 2010; 5: 11.

[23] Sideridou M, Zakopoulou R, Evangelou K, Liontos M, Kotsinas A, Rampakakis E, Gagos S, Kahata K, Grabusic K, Gkouskou K, Trougakos IP, Kolettas E, Georgakilas AG, Volarevic S, Eliopoulos AG, Zannis-Hadjopoulos M, Moustakas A, Gorgoulis VG. Cdc6 expression represses E-cadherin transcription and activates adjacent replication origins. J Cell Biol 2011; 195: 1123-40.

[24] Blow JJ and Gillespie PJ. Replication licensing and cancer--a fatal entanglement? Nat Rev Cancer 2008; 8: 799-806.

[25] Liontos M, Koutsami M, Sideridou M, Evangelou K, Kletsas D, Levy B, Kotsinas A, Nahum O, Zoumpourlis V, Kouloukoussa M,

Lygerou Z, Taraviras S, Kittas C, Bartkova J, Papavassiliou AG, Bartek J, Halazonetis TD, Gorgoulis VG. Deregulated overexpression of hCdt1 and hCdc6 promotes malignant behavior. Cancer Res 2007; 67: 10899-909.

[26] Halazonetis TD, Gorgoulis VG and Bartek J. An oncogene-induced DNA damage model for cancer development. Science 2008; 319: 1352-5.

[27] Negrini S, Gorgoulis VG and Halazonetis TD. Genomic instability--an evolving hallmark of cancer. Nat Rev Mol Cell Biol 2010; 11: 220-8.

[28] Robbins JA and Cross FR. Regulated degradation of the APC coactivator Cdc20. Cell Div 2010; 5: 23.

[29] Pollok S, Bauerschmidt C, Sänger J, Nasheuer HP, Grosse F. Human Cdc45 is a proliferation-associated antigen. FEBS J 2007; 274: 3669-84.

[30] Enders GH. Expanded roles for Chk1 in genome maintenance. J Biol Chem 2008; 283: 17749-52.

[31] Ohi R and Gould KL. Regulating the onset of mitosis. Curr Opin Cell Biol 1999; 11: 267-73.

[32] Choi HJ, Fukui M and Zhu BT. Role of cyclin B1/Cdc2 up-regulation in the development of mitotic prometaphase arrest in human breast cancer cells treated with nocodazole. PLoS One 2011; 6: e24312.

[33] Li Z, Li J, Bi P, Lu Y, Burcham G, Elzey BD, Ratliff T, Konieczny SF, Ahmad N, Kuang S, Liu X. Plk1 phosphorylation of PTEN causes a tumor-promoting metabolic state. Mol Cell Biol 2014; 34: 3642-61.

[34] Nasir A, Chen DT, Gruidl M, Henderson-Jackson EB, Venkataramu C, McCarthy SM, McBride HL, Harris E, Khakpour N, Yeatman TJ. Novel molecular markers of malignancy in histologically normal and benign breast. Patholog Res Int 2011; 2011: 489064.

[35] Tang L, Dai DL, Su M, Martinka M, Li G, Zhou Y. Aberrant expression of collagen triple helix repeat containing 1 in human solid cancers. Clin Cancer Res 2006; 12: 3716-22.

[36] Yu P, Huang B, Shen M, Lau C, Chan E, Michel J, Xiong Y, Payan DG, Luo Y. p15(PAF), a novel PCNA associated factor with increased expression in tumor tissues. Oncogene 2001; 20: 484-9.

[37] Xu X, Liu Z, Zhou L, Xie H, Cheng J, Ling Q, Wang J, Guo H, Wei X, Zheng S. Characterization of genome-wide TFCP2 targets in hepatocellular carcinoma: implication of targets FN1 and TJP1 in metastasis. J Exp Clin Cancer Res 2015; 34: 6.

[38] Chou CP, Huang NC, Jhuang SJ, Pan HB, Peng NJ, Cheng JT, Chen CF, Chen JJ, Chang TH. Ubiquitin-conjugating enzyme UBE2C is highly expressed in breast microcalcification lesions. PLoS One 2014; 9: e93934.

[39] Axlund SD, Yoo BH, Rosen RB, Schaack J, Kabos P, Labarbera DV, Sartorius CA. Progesterone-inducible cytokeratin 5-positive cells in luminal breast cancer exhibit progenitor properties. Horm Cancer 2013; 4: 36-49.

[40] Harris RE. Cyclooxygenase-2 (cox-2) and the inflammogenesis of cancer. Subcell Biochem 2007; 42: 93-126.

[41] Magnowska M, Zaborowski M, Surowiak P, Nowak-Markwitz E, Zabel M, Spaczyński M. COX-2 expression pattern is related to ovarian cancer differentiation and prognosis, but is not consistent with new model of pathogenesis. Ginekol Pol 2014; 85: 335-41.

[42] Harris RE, Casto BC and Harris ZM. Cyclooxygenase-2 and the inflammogenesis of breast cancer. World J Clin Oncol 2014; 5: 677-92.

[43] Pommier AJ, Shaw R, Spencer SK, Morgan SR, Hoff PM, Robertson JD, Barry ST, Jürgensmeier JM. Serum protein profiling reveals baseline and pharmacodynamic biomarker signatures associated with clinical outcome in mCRC patients treated with chemotherapy +/- cediranib. Br J Cancer 2014; 111: 1590-604.

[44] Becker MA, Farzan T, Harrington SC, Krempski JW, Weroha SJ, Hou X, Kalli KR, Wong TW, Haluska P. Dual HER/VEGF receptor targeting inhibits in vivo ovarian cancer tumor growth. Mol Cancer Ther 2013; 12: 2909-16.

[45] Baggiolini M, Loetscher P and Moser B. Interleukin-8 and the chemokine family. Int J Immunopharmacol 1995; 17: 103-8.

[46] Strieter RM, Burdick MD, Mestas J, Gomperts B, Keane MP, Belperio JA. Cancer CXC chemokine networks and tumour angiogenesis. Eur J Cancer 2006; 42: 768-78.

[47] Valente AL, Kane JL, Ellsworth DL, Shriver CD, Ellsworth RE. Molecular response of the axillary lymph node microenvironment to metastatic colonization. Clin Exp Metastasis 2014; 31: 565-72.

[48] Yao C, Lin Y, Chua MS, Ye CS, Bi J, Li W, Zhu YF, Wang SM. Interleukin-8 modulates growth and invasiveness of estrogen receptor-negative breast cancer cells. Int J Cancer 2007; 121: 1949-57.
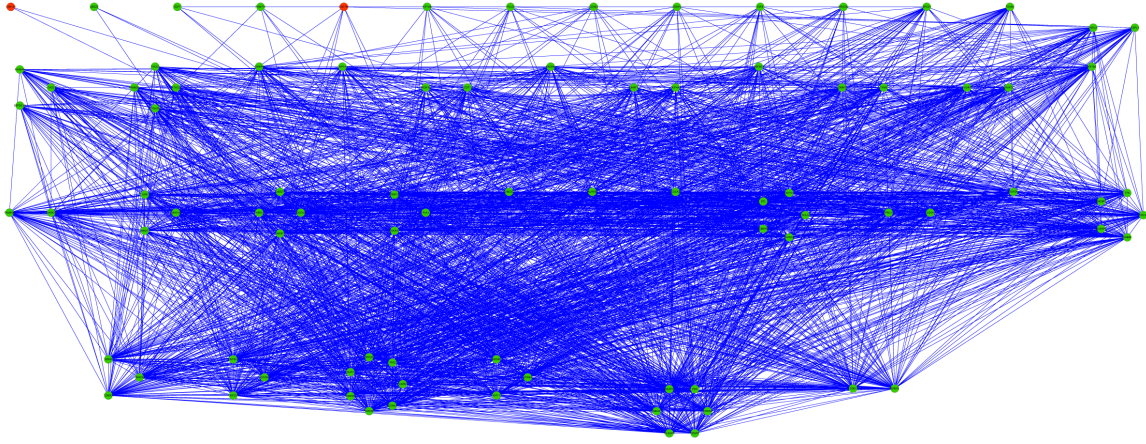
Analysis of DEGs in breast cancer

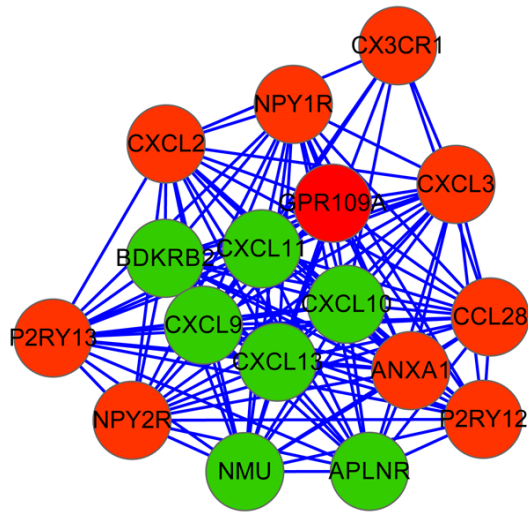
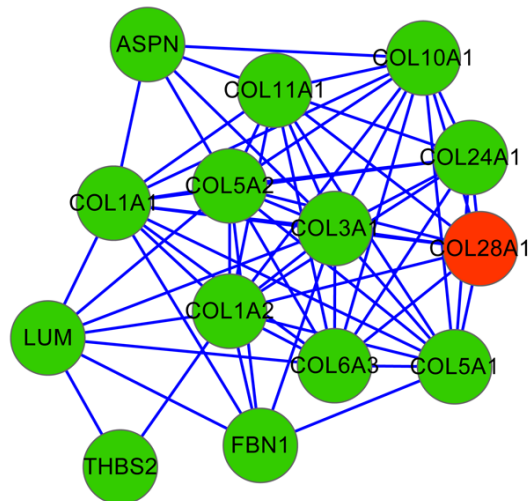**Figure S1.** The biggest sub-network of the first module.

Figure S2. The DEGs in the second module.

Figure S3. The DEGs in the third module.