Screening of the prognostic targets for breast cancer based co-expression modules analysis

HUIJUAN LIU¹ and HUI YE²

Departments of ¹Breast Surgery and ²Thoracic Surgery, Shanxi Tumor Hospital, Taiyuan, Shanxi 030013, P.R. China

Received September 14, 2016; Accepted May 23, 2017

DOI: 10.3892/mmr.2017.7063

Abstract. The purpose of the present study was to screen the prognostic targets for breast cancer based on a co-expression modules analysis. The microarray dataset GSE73383 was downloaded from the Gene Expression Omnibus (GEO) database, including 15 breast cancer samples with good prognosis and 9 breast cancer samples with poor prognosis. The differentially expressed genes (DEGs) were identified with the limma package. The Database for Annotation, Visualization and Integrated Discovery was used to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. Furthermore, the co-expression analysis of DEGs was conducted with weighted correlation analysis. The interaction associations were analyzed with the Human Protein Reference Database and BioGRID. The protein-protein interactions (PPI) network was constructed and visualized by Cytoscape software. A total of 491 DEGs were identified in breast cancer samples with poor prognosis compared with those with good prognosis, and they were enriched in 85 GO terms and 4 KEGG pathways. 368 DEGs were co-expressed with others, and they were clustered into 10 modules. Module 6 was the most relevant to the clinical features, and 21 genes and 273 interaction pairs were selected out. Abnormal expression levels of required for meiotic nuclear division 5 homolog A (RMND5A) and angiopoietin-like protein 1 (ANGPTLI) were associated with a poor prognosis. It was indicated that SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily D, member 1, SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily D, member 1, dihydropyrimidinase-like 2, RMND5A and ANGPTLI were potential prognostic markers in breast cancer, and the cell cycle may be involved in the regulation of breast cancer.

Introduction

Breast cancer is the uncontrolled growth of breast cells, which develops from breast tissue. It is one of the most common cancers among the female malignant tumors in worldwide, that affects about 12% of global women (1). In North America and Western Europe, breast cancer is the most common malignancy and the second most common cause of cancer-associated death for women (2). It was estimated that there were 246,660 new cases and 40,450 mortalities from breast cancer in the US in 2016 (3). Morbidity and mortality increase year by year, particularly in Southeast Asia (4). Patients with breast cancer experience various symptoms, including a lump in the breast, fluid coming from the nipple, dimpling of the skin and a change in breast shape (5). Breast cancer is usually treated with surgery, which may be followed by chemotherapy or radiotherapy, and a multidisciplinary approach is preferable, such as surgery, radiotherapy, chemotherapy and hormone therapy (6). Breast cancers that are estrogen receptor positive (ER+) and human epidermal growth factor receptor-2 positive (HER2+) can be treated with agents that target these receptors (such as tamoxifen, toremifene, and fulvestrant for ER+ breast cancers and trastuzumab, pertuzumab and lapatinib for HER2+ breast cancers) (7). Unfortunately, the prognosis is still unsatisfactory, especially in advanced patients. Prognostic factors of breast cancer mainly depend on stage (i.e., tumor size, location, and metastasis), grade, recurrence, age and health of the patient. The stage of breast cancer has a greater effect on the prognosis than the other factors, and it is considered the most important component of traditional classification methods (8).

Breast cancer is a complex systemic disease with multi-genes involved, and a series of abnormal changes of genome which participate in the regulation of cell growth, differentiation and apoptosis, may affect the prognosis. Some prognostic markers have been reported, such as, INPP4b, KLF4, NANOG, OXYTAM and EZH2 (9,10). However, the ability to predict the metastatic behavior in breast cancer is still limited, and the clinical outcome of breast cancer remains to be improved. In the present study, breast cancer samples with different prognosis were analyzed via microarray analysis, in order to find more prognostic markers and provide some clues for the metastatic behavior in breast cancer.

Correspondence to: Dr Hui Ye, Department of Thoracic Surgery, Shanxi Tumor Hospital, 3 Zhigongxinjie Street, Taiyuan, Shanxi 030013, P.R. China
E-mail: 397944545@qq.com

Key words: breast cancer, prognosis, co-expression modules analysis
Materials and methods

Microarray data. The expression microarray dataset GSE73383 was downloaded from the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo) database. In this profile, there was a subset of 24 breast cancer samples comprising 15 samples that remained free of disease after surgery (good prognosis) and 9 samples that developed metastasis (poor prognosis). In this study, the 24 samples were used to identify candidate biomarkers associated with the prognosis of breast cancer, which were detected with the GPL11010 CodeLink Human Whole Genome Bioarray platform (Applied Microarrays, Inc., Tempe, AZ, USA).

Data pre-processing and identification of differentially expressed genes. The original data were converted into recognizable format in R, and the preprocess Core package was used for the normalization. Afterwards, the limma (11) package of R was used to identify the differentially expressed genes (DEGs) in the 9 samples with poor prognosis compared with the 15 samples with good prognosis. Furthermore, the DEGs were selected out according to the criteria: \(|\log (\text{fold change})| > 1\) and \(P<0.05\).

Functional and pathway enrichment analysis. The Database for Annotation, Visualization and Integrated Discovery (david.ncifcrf.gov) (12) is a widely-used web-based tool for functional and pathway enrichment analysis. Here, it was used to perform Gene Ontology (GO; www.geneontology.org) and Kyoto Encyclopedia of Genes and Genomes (KEGG; www.genome.jp/kegg) pathway enrichment analysis of DEGs. GO terms and KEGG pathways were selected out with a \(P\)-value <0.05.

Module analysis. Weighted correlation analysis (WGCNA) version 1.13 (www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/Rpackages/WGCNA), an R software package, is a comprehensive collection of R functions for performing various aspects of weighted correlation network analysis (13). In the present study, the co-expression analysis of DEGs was conducted with WGCNA. Afterwards, the co-expression modules were obtained, and the relationships between every module and clinical features were calculated.

Construction of the protein-protein interactions network. The Human Protein Reference Database (HPRD) is an object database that integrates a wealth of information relevant to the function of human proteins in health and disease (14). The Biological General Repository for Interaction Datasets (BioGRID) is a curated biological database of protein-protein interactions, genetic interactions, chemical interactions and post-translational modifications (15). The module which had the closest correlation with the prognosis was analyzed with HPRD version 9 (www.hprd.org) and BioGRID version 2.0 (thebiogrid.org) software. In addition, genes in the above module and their associated interaction pairs were selected out. In addition, the protein-protein interactions (PPI) network was constructed and visualized by Cytoscape version 3.0.1 (16) software. Nodes were screened out in the PPI network with degree \(\geq 1\), and ‘degree’ represented the connections with other nodes. In addition, the associations between certain nodes and the prognostic of breast cancer were analyzed using the online tool kmplot, version 1.2.0 (edu.kde.org/kmplot), which is a mathematical function plotter for the KDE-Desktop, and can be used to plot different functions simultaneously and combine their function terms to build new functions.

Results

DEGs. A total of 491 DEGs (316 up- and 175 downregulated) were identified in breast cancer samples with poor prognosis compared with those with good prognosis. The top 30 most significant DEGs are presented in Table I.

GO terms and KEGG pathways. The above DEGs were enriched in 85 GO terms and 4 KEGG pathways. The top 15 GO terms were presented in Table II, including cell cycle phase, chromosome segregation and nuclear division. Furthermore, the 4 KEGG pathways were cell cycle, neuroactive ligand-receptor interaction, oocyte meiosis and progesterone-mediated oocyte maturation.

Modules. A total of 368 DEGs were co-expressed with others and they were clustered into 10 modules (recorded as Module 1-10). The cluster graph of samples and genes are presented in Fig. 1. From the sample cluster graph, GSM1892326 was identified as an outlier; thus, it was excluded in the further module analysis. The associations between every module and clinical features (Her 2, Ki67, metastasis and prognosis) are presented in Fig. 2. Module 6 was the most relevant to the clinical features (Fig. 2). There were 26 DEGs (23 over-expressed and 3 low-expressed) in Module 6, and the associations between their expression and the prognosis of breast cancer are presented in Fig. 3 (‘Pgood’ represents good prognosis and ‘Ppoor’ represents poor prognosis).

PPI network. Module 6 was analyzed with HPRD and BioGRID, and 21 genes and 273 interaction pairs were selected out. Afterwards, the PPI network was established based on the above interactions, including 266 nodes, and is presented in Fig. 4. The top 20 nodes are presented in Table III. One of the over-expressed gene in Module 6, required for meiotic nuclear division 5 homolog A (RMND5A), and the low-expressed gene angiopoietin-like protein 1 (ANGPTL1) were screened out, and the association between them and prognosis were analyzed with the online tool of kmplot. Survival curves are presented in Fig. 5. It revealed that abnormal expression levels (low- and over-expression) of RMND5A and ANGPTL1 were associated with a poor prognosis.

Discussion

DEGs were identified in breast cancer samples with poor prognosis compared with those with good prognosis, and they were mainly enriched certain cell cycle- and mitosis-associated biological processes. At present, research has indicated that the cell cycle is closely associated with the outcome of breast cancer. Certain cell cycle markers (for example cyclin B1) are differentially expressed in benign and
malignant papillary breast lesions (17). Over-expression of a novel cell cycle regulator ecdysoneless predicted a poorer survival in Her2/neu over-expressing breast cancer patients. Oxygenase domain-containing protein 1 is required for breast cancer cell proliferation and is involved in poor prognosis of breast cancer (18). Therefore, it is important to intervene the progression, prognosis and drug efficacy in breast cancer by regulating or modulating some cell cycle phases or cyclin proteins. The transcription factor deleted in esophageal cancer could regulate breast cancer cell proliferation by stabilizing cyclin E protein and delaying the progression of cell cycle S phase (19). Phosphatase and tensin homolog (PTEN) was a well-known tumor suppressor gene and is frequently mutated or lost in breast cancer. Insufficiency of PTEN could modulate ER+ breast cancer cell cycle progression and increase cell growth, and thus is associated with advanced breast cancer and poor prognosis of breast cancer patients (20). Nexrutine inhibits survival and induces G1 cell cycle arrest, which might provide a novel approach for protection against breast cancer (21). Mitosis is the most important factor in cell the

Table I. Top 30 most significant DEGs in breast cancer samples with poor prognosis compared with those with good prognosis.

| Gene     | P-value   | LogFC | Gene     | P-value   | LogFC |
|----------|-----------|-------|----------|-----------|-------|
| CHRDL2   | 5.74E-05  | -2.95172 | CSH1     | 0.003025  | 1.197913 |
| ERP27    | 0.000228  | -1.67607 | CPT1A    | 0.003214  | 1.239627 |
| PRSS21   | 0.000235  | -1.98312 | LOC105370832 | 0.003264 | -1.71612 |
| LOC100129447 | 0.000542 | -3.5659  | IBSP     | 0.003517  | 1.247671 |
| TRIM37   | 0.001246  | 1.758675 | TMEM185B | 0.00362   | 1.212002 |
| TTF2     | 0.001416  | 1.823105 | PMM1D    | 0.003634  | 1.3262  |
| ANKR60   | 0.001461  | -2.87722 | PDP1     | 0.003666  | 2.083321 |
| HMGB3    | 0.001635  | 1.915918 | CDC45    | 0.003698  | 2.059617 |
| NBR1     | 0.001681  | -1.71848 | NEB      | 0.003822  | 1.118614 |
| FKBP1B   | 0.001731  | 1.198288 | PDE1B    | 0.003835  | -1.64844 |
| MGC16275 | 0.002375  | 1.402384 | ZNF264   | 0.004253  | -1.65155 |
| C9orf64  | 0.0025    | 1.277235 | GTSE1    | 0.004292  | 1.209348 |
| OPRK1    | 0.002658  | 1.891967 | HIST1H3B | 0.004466  | 1.522681 |
| LGI4     | 0.002916  | -1.67571 | EMX2     | 0.004667  | -1.25036 |
| HIST1H2BE| 0.003025  | 1.817212 | FAM9C    | 0.004707  | 1.524613 |

DEGs, differentially expressed genes; FC, fold-change.

Table II. Top 15 most significant GO terms of DEGs.

| Category | Term                                           | Count | P-value   |
|----------|------------------------------------------------|-------|-----------|
| BP       | GO:0022403~cell cycle phase                     | 29    | 2.07E-07  |
| BP       | GO:0000087~M phase of mitotic cell cycle        | 20    | 7.28E-07  |
| BP       | GO:000278~mitotic cell cycle                    | 26    | 9.76E-07  |
| BP       | GO:0007059~chromosome segregation               | 12    | 1.84E-06  |
| BP       | GO:0007067~mitosis                              | 19    | 2.41E-06  |
| BP       | GO:000280~nuclear division                      | 19    | 2.41E-06  |
| BP       | GO:0024202~cell cycle process                   | 32    | 4.12E-06  |
| BP       | GO:0048285~organelle fission                    | 19    | 4.25E-06  |
| BP       | GO:0000279~M phase                               | 23    | 5.26E-06  |
| BP       | GO:0007049~cell cycle                           | 38    | 1.18E-05  |
| CC       | GO:0000940~outer kinetochore of condensed chromosome | 5    | 3.64E-05  |
| BP       | GO:0000075~cell cycle checkpoint                | 11    | 3.64E-05  |
| BP       | GO:0051301~cell division                        | 20    | 3.92E-05  |
| BP       | GO:0007093~mitotic cell cycle checkpoint         | 8     | 4.35E-05  |
| CC       | GO:000777~condensed chromosome kinetochore      | 9     | 6.51E-05  |

GO, Gene Ontology; DEGs, differentially expressed genes; BP, biological process; CC, Cellular Component.
cycle, which is a complicated and refined adjusting process. However, relatively few studies have reported that mitosis is a relevant biological process involved in the prognosis of breast cancer. Mitosis counts in histopathological slides serve a crucial role for invasive breast cancer grading (22,23). To summarize, the biological processes of cell cycle and mitosis

Figure 1. Cluster graph of samples and genes.

Figure 2. Associations between every module and clinical features.
are closely involved in the prognosis of breast cancer, which might provide novel insights into the clinical therapy for breast cancer patients. Furthermore, it was possible to reveal more prognostic biomarkers by exporting primarily involved and regulatory genes.

In the present study, DEGs were enriched in 4 KEGG pathways, namely, cell cycle, neuroactive ligand-receptor interaction, oocyte meiosis and progesterone-mediated oocyte maturation. The neuroactive ligand-receptor interaction pathway mainly serves roles in neurological illness and injury, such as Parkinson's Disease (24). Huan et al (25) identified time courses of genes that were either stimulated or inhibited by estradiol (E2) in human breast cancer cells, and results showed that the changes were mainly enriched in neuroactive ligand-receptor interaction at the 24 h time point. A study of the hub subnetwork in breast cancer revealed that oocyte meiosis pathways were significant subnetworks, in which hub nodes mostly distributed (26). Another bioinformatics analysis reported that the pathogenesis of breast cancer was associated with oocyte meiosis and progesterone-mediated oocyte maturation (27). In this study, it was indicated the prognosis of breast cancer was associated with cell cycle, neuroactive ligand-receptor interaction, oocyte meiosis and progesterone-mediated oocyte maturation. However, further laboratory and clinical researches are required to verify this.

In the present study, SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily D, member 1 (SMARCD1) and dihydropyrimidinase-like 2 (DPYSL2) were the top two nodes in the PPI network. SMARCD1, a regulator gene of embryonic stem cell differentiation, was considered as one of the driver mutations in breast cancer (28,29). Furthermore, the attenuated expression of SMARCD1 could inhibit breast cancer cell proliferation in vitro and in mouse tumor xenografts (30). DPYSLs are a family of proteins developmentally regulated during maturation of the nervous system, and mainly comprising of DPYSL1, DPYSL2 and DPYSL3. DPYSL2 encodes the DPYSL2 protein, which is a target gene of gastric cancer (31). DPYSL2 is identified to increase with selenomethionine treatment in colon cancer (32). In the present study, it was revealed that SMARCD1 and DPYSL2 may be prognostic markers in breast cancer, although further studies are required to confirm this. Furthermore, in the present study, abnormal expression levels of RMND5A and ANGPTL1 were associated with poor prognosis in breast cancer patients. RMND5A is involved in microtubule dynamics, cell migration, nucleokinesis and chromosome segregation (33). However, few reports have revealed the association between RMND5A and breast cancer. ANGPTL1 encodes ANGPTL1 protein, which targets endothelial cells and affects tumor cell behavior. A previous study reported that ANGPTL1 represses lung cancer cell motility (34). The results of the present study revealed that RMND5A and ANGPTL1 are novel potential prognostic markers in breast cancer.

In conclusion, the results of the present study indicated that SMARCD1, DPYSL2, RMND5A and ANGPTL1 are potential prognostic markers in breast cancer. Furthermore, cell cycle, mitosis, neuroactive ligand-receptor interaction, oocyte meiosis and progesterone-mediated oocyte maturation may be involved in the regulation of breast cancer.

### Table III. Top 20 nodes in the PPI network.

| Gene      | Degree | Gene   | Degree |
|-----------|--------|--------|--------|
| SMARCD1   | 78     | PNKD   | 7      |
| DPYSL2    | 32     | CSH1   | 5      |
| HIRIP3    | 32     | GALNT14| 4      |
| ARH1      | 30     | APP    | 3      |
| UBF1      | 27     | ELAVL1 | 3      |
| NEB       | 19     | NRF1   | 3      |
| PARN      | 15     | ANGPTL1| 2      |
| ATP6V0A2  | 10     | CALM1  | 2      |
| RMND5A    | 10     | HSPH1  | 2      |
| UBC       | 9      | IQCB1  | 2      |

PPI, protein-protein interaction.
Figure 4. Protein-protein interaction network.

Figure 5. Survival curve of RMND5A and ANGPTL1. RMND5A, required for meiotic nuclear division 5 homolog A; ANGPTL1, angiopoietin-like protein 1.
