GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Screening for key genes associated with invasive ductal carcinoma of the breast via microarray data analysis

J.L. Huan[1], X. Gao[2], L. Xing[3], X.J. Qin[1], H.X. Qian[1], Q. Zhou[1] and L. Zhu[1]

[1]Department of General Surgery, The Eighth People's Hospital of Shanghai, Shanghai, China
[2]Department of Neurosurgery, Yidu Center Hospital of Weifang City, Weifang, China
[3]Department of Gynaecology and Obstetrics, The Eighth People's Hospital of Shanghai, Shanghai, China

Corresponding author: J.L. Huan
E-mail: jinlianghuan@hotmail.com

Genet. Mol. Res. 13 (3): 7919-7925 (2014)
Received July 31, 2013
Accepted August 7, 2014
Published September 29, 2014
DOI http://dx.doi.org/10.4238/2014.September.29.5

ABSTRACT. The aim of this study was to identify key genes related to invasive ductal carcinoma (IDC) of the breast by analyzing gene expression data with bioinformatic tools. Microarray data set GSE31138 was downloaded from Gene Expression Omnibus, including 3 breast cancer tissue samples and 3 normal controls. Differentially expressed genes (DEGs) between breast cancer and normal control were screened out (FDR < 0.05 and |logFC| > 2). Coexpression between genes was examined with String, and a network was then constructed. Relevant pathways and diseases were retrieved with KOBAS. A total of 56 DEGs were obtained in the IDC of the breast compared with normal controls. A gene coexpression network including 27 pairs of genes was constructed and all the genes in the network were upregulated. Further study indicated that most of the genes in the coexpression network were enriched in ECM-receptor interaction (COL4A2, FN1, and HMMR) and nucleotide excision repair

(CETN2 and PCNA) pathways, and that the most significantly related disease was autoimmune lymphoproliferative syndromes. A number of DEGs were acquired through comparative analysis of gene expression data. These findings are beneficial in promoting the understanding of the molecular mechanisms in breast cancer. More importantly, some key genes were revealed via gene coexpression network analysis, which could be potential biomarkers for IDC of the breast.

Key words: Breast cancer; Differentially expressed genes; Gene coexpression network analysis; Pathway analysis

INTRODUCTION

Breast cancer poses a great threat to human health, especially for women. It comprises 22.9% of all cancers (excluding non-melanoma skin cancers) in women (Buchholz, 2009) and caused 13.7% of cancer deaths in 2008 (Saracci, 2008). It can be divided into ductal carcinomas and lobular carcinomas (Florescu et al., 2011), where the former is more common.

Since early diagnosis and treatment are of great importance in reducing the mortality of breast cancer, people have been trying to identify more biomarkers (Weigel and Dowsett, 2010; Misek and Kim, 2011). The most important finding has been human epidermal growth factor receptor as a biomarker (Slamon et al., 1987), and now, targeted therapy has been successfully developed (Slamon et al., 2001; Vogel et al., 2002). Various new biomarkers have been reported, such as kallikrein gene 14 (Borgoño et al., 2003) and BAG-1 (Turner et al., 2001). Zehentner et al. (2004) indicated that mammaglobin can serve as a diagnostic tool for breast cancer. Dunning et al. (2003) found that a transforming growth factor β1 signal peptide variant was associated with increased incidence of invasive breast cancer.

Although considerable achievements have been made, more studies are needed to gain more knowledge about breast cancer and to discover effective biomarkers for diagnosis or treatment. Microarray technology is an effective tool to disclose the global changes in the incidence and development of cancer (DeRisi et al., 1996). Therefore, in the present study, we tried to identify potential biomarkers for invasive ductal carcinoma (IDC) of the breast via differential analysis and gene coexpression network analysis of gene expression data.

MATERIAL AND METHODS

Microarray data

The gene chip dataset GSE31138 was downloaded from Gene Expression Omnibus, containing 3 breast cancer samples and 3 normal controls. All samples were donated by the London BARTS Cancer Institute Laboratory. The platform was Affymetrix Human Genome U133 Plus 2.0 Array.

Data pre-processing and differential analysis

The original data were converted into recognizable expression data. The missing values were then filled in with the KNN method (Troyanskaya et al., 2001). The standardization

was performed with the median method (Fujita et al., 2006), followed by differential expression analysis with the package Limma (Kerr, 2003) of *R*. The Benjamini-Hochberg method (Benjamini et al., 2001) was used for multiple testing correction and the differentially expressed genes (DEGs) were selected out according to the criteria: $P < 0.05$, $FDR < 0.05$ and $|logFC| > 1$.

Gene coexpression network analysis

String (Szklarczyk et al., 2011) was chosen for gene co-expression network analysis of the DEGs. The coexpression coefficients between DEGs were calculated on the basis of the characteristics of gene sequences and spatial structures. The coexpression pairs with coexpression >0.5 were selected out and used for gene coexpression network construction.

Retrieval of relevant pathways and diseases

Relevant pathways and diseases were retrieved for DEGs with KOBAS (KEGG Orthology Based Annotation System) (Xie et al., 2011). KOBAS was the first software to identify significantly enriched pathways using a hypergeometric test. It has been successfully used in pathway analysis for plants, animals and bacteria. Its purpose is to identify significantly enriched pathways and diseases for a set of genes or proteins, using pathway and disease information from multiple databases. In present study, $P < 0.05$ was set as the cut-off to screen out pathways related to IDC of the breast.

RESULTS

Differentially expressed genes

A good performance of data standardization was acquired. A total of 56 DEGs were then identified for breast cancer, 51 upregulated and 5 downregulated. It was obvious that most DEGs played roles in IDC via overexpression.

Gene coexpression network

The coexpression pairs with coexpression >0.5 were selected out and included in the network. A total of 27 pairs of coexpressed gene pairs were revealed (Figure 1), such as collagen 4A2 (COL4A2)-fibronectin 1 (FN1), COL4A2-TOP2A and neuroblastoma RAS viral (v-ras) oncogene homolog (NRAS)-WASL.

Relevant pathways and diseases

KOBAS (Xie et al., 2011) was chosen for pathway and disease analysis for DEGs and $P < 0.05$ was set as the cut-off. Finally, 2 relevant pathways [extracellular matrix (ECM)-receptor interaction and nucleotide excision repair] and 5 diseases [autoimmune lymphoproliferative syndromes (ALPS), malignant melanoma, adrenal carcinoma, oral cancer, and multiple myeloma] were obtained (Tables 1 and 2).

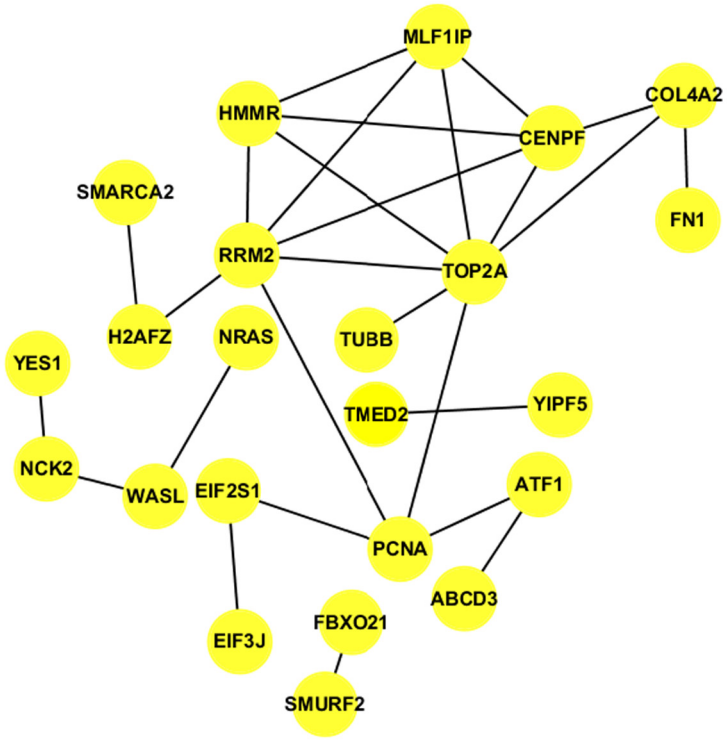

Figure 1. The coexpression network for differentially expressed genes.

Table 1. Relevant pathways of coexpressed genes.

#Term	Database	Id	P value	Genes
ECM-receptor interaction	KEGG PATHWAY	hsa04512	0.017655	COL4A2, FN1, HMMR
Nucleotide excision repair	KEGG PATHWAY	hsa03420	0.035558	CETN2, PCNA

Table 2. Diseases related to the coexpressed genes in the network.

#Term	Database	Id	P value	Genes
Autoimmune lymphoproliferative syndromes	KEGG DISEASE	H00108	0.025546096	NRAS
Malignant melanoma	KEGG DISEASE	H00038	0.035587324	NRAS
Adrenal carcinoma	KEGG DISEASE	H00033	0.035587324	NRAS
Oral cancer	KEGG DISEASE	H00016	0.040570449	NRAS
Multiple myeloma	KEGG DISEASE	H00010	0.045528742	NRAS

DISCUSSION

In the present study, 51 upregulated and 5 downregulated genes were obtained in the IDC of the breast. Gene coexpression network analysis was performed for all the DEGs, and relevant pathways and disease were then retrieved with KOBAS. Finally, 2 pathways (ECM-

receptor interaction and nucleotide excision repair) and 5 diseases (ALPS, malignant melanoma, adrenal carcinoma, oral cancer and multiple myeloma) were revealed.

COL4A2, FN1 and HMMR (hyaluronan-mediated motility receptor) are components of ECM. The expression of ECM genes is associated with the prognosis of patients with lymph node-negative breast cancer as well as clinical benefit from endocrine treatment (Insalaco et al., 2012). FN1 expression is found to be upregulated during epithelial to mesenchymal transition, which is an early event in malignant transformation accompanied by a reduced adhesion of the tumor cells (Helleman et al., 2008). Several studies suggest that FN1 is related to tumor invasion and metastasis (Landstrom et al., 1992) by playing a key role in the tissue remodeling and cell migration events that occur during normal development. In particular, FN1 is a major constituent of the cell surface of many cultured cells, and it is either eliminated or reduced on the surface of oncogenically transformed cells (Horii et al., 2006). Many reports have suggested that there is a correlation between the loss of cell surface FN1 and the ability of a cell to metastasize (Caraglia et al., 2004). The changes in the cytoskeletal components such as production and organization of FN1, actin and collagen have been implicated in eliciting the transition from dormancy to metastatic growth (Calvo et al., 2008). This was in accordance with our finding since there was a coexpression between FN1 and COL4A2, which suggested that the interaction between the two proteins could be a good cut-in point to modulate the mobility of tumor cells. HMMR has also been reported to mediate migration, transformation, and metastatic spread of cancer cells (Du et al., 2011; Veiseh and Turley, 2011). Of course, further studies are needed to determine the use of these ECM genes in decisions regarding treatment and whether they can serve as targets for therapy (Vargas et al., 2012).

Centrin EF-hand protein 2 (CETN2) is a structural component of the centrosome, and proliferating cell nuclear antigen (PCNA) is a cofactor of DNA polymerase delta, both of which are involved in nucleotide excision repair. Nucleotide excision repair is linked with cancer risk (Lockett et al., 2005; Barry et al., 2012), and therefore, both proteins are worthy of further research to fully elucidate their roles in IDC.

According to Table 2, NRAS was associated with ALPS as well as several kinds of cancers, such as melanoma and oral cancer. ALPS is characterized by nonmalignant lymphadenopathy, splenomegaly, and autoimmune cytopenias (Krueger et al., 2002). Defective lymphocyte apoptosis secondary to mutations in the FAS gene is identified as a molecular basis for ALPS (Fisher et al., 1995). It was the first disease known to be caused by a primary defect in programmed cell death and was the first autoimmune disease with a defined genetic basis (John et al., 2008). Besides, previous studies have confirmed that NRAS is one of the specific and sensitive indices for breast cancer diagnosis and prognosis (Zhu et al., 2004). It was very interesting to find that breast cancer was associated with ALPS to a certain degree since it indicates a possibility to ameliorate both ALPS and breast cancer syndrome by targeted therapy on NRAS gene expression.

Several genes associated with breast cancer have been utilized as diagnostic or prognostic markers, or even therapeutic targets, such as NF-κb (Biswas et al., 2001, 2004) and HER-2 (Helms et al., 2010; Perez et al., 2010). However, currently available biomarkers for the early diagnosis and therapy of breast cancer are far from enough. Therefore, in the present study, we analyzed gene expression data to identify biomarkers for breast cancer. Since ECM-related COL4A2, FN1 and HMMR as well as CETN2 and PCNA are associated with breast cancer, they may be potential biomarkers. Meanwhile, NRAS, which showed close as-

sociations with several kinds of tumors and ALPS, could also be a promising biomarker for breast cancer.

REFERENCES

Barry KH, Koutros S, Andreotti G, Sandler DP, et al. (2012). Genetic variation in nucleotide excision repair pathway genes, pesticide exposure and prostate cancer risk. *Carcinogenesis* 33: 331-337.

Benjamini Y, Drai D, Elmer G, Kafkafi N, et al. (2001). Controlling the false discovery rate in behavior genetics research. *Behav. Brain Res.* 125: 279-284.

Biswas DK, Dai SC, Cruz A, Weiser B, et al. (2001). The nuclear factor kappa B (NF-kappa B): a potential therapeutic target for estrogen receptor negative breast cancers. *Proc. Natl. Acad. Sci. U. S. A.* 98: 10386-10391.

Biswas DK, Shi Q, Baily S, Strickland I, et al. (2004). NF-kappa B activation in human breast cancer specimens and its role in cell proliferation and apoptosis. *Proc. Natl. Acad. Sci. U. S. A.* 101: 10137-10142.

Borgoño CA, Grass L, Soosaipillai A, Yousef GM, et al. (2003). Human kallikrein 14: a new potential biomarker for ovarian and breast cancer. *Cancer Res.* 63: 9032-9041.

Buchholz TA (2009). Radiation therapy for early-stage breast cancer after breast-conserving surgery. *N. Engl. J. Med.* 360: 63-70.

Calvo A, Catena R, Noble MS, Carbott D, et al. (2008). Identification of VEGF-regulated genes associated with increased lung metastatic potential: functional involvement of tenascin-C in tumor growth and lung metastasis. *Oncogene* 27: 5373-5384.

Caraglia M, D'Alessandro AM, Marra M, Giuberti G, et al. (2004). The farnesyl transferase inhibitor R115777 (Zarnestra) synergistically enhances growth inhibition and apoptosis induced on epidermoid cancer cells by Zoledronic acid (Zometa) and Pamidronate. *Oncogene* 23: 6900-6913.

DeRisi J, Penland L, Brown PO, Bittner ML, et al. (1996). Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.* 14: 457-460.

Du YC, Chou CK, Klimstra DS and Varmus H (2011). Receptor for hyaluronan-mediated motility isoform B promotes liver metastasis in a mouse model of multistep tumorigenesis and a tail vein assay for metastasis. *Proc. Natl. Acad. Sci. U. S. A.* 108: 16753-16758.

Dunning AM, Ellis PD, McBride S, Kirschenlohr HL, et al. (2003). A transforming growth factorbeta1 signal peptide variant increases secretion *in vitro* and is associated with increased incidence of invasive breast cancer. *Cancer Res.* 63: 2610-2615.

Fisher GH, Rosenberg FJ, Straus SE, Dale JK, et al. (1995). Dominant interfering Fas gene mutations impair apoptosis in a human autoimmune lymphoproliferative syndrome. *Cell* 81: 935-946.

Florescu A, Amir E, Bouganim N and Clemons M (2011). Immune therapy for breast cancer in 2010 - hype or hope? *Curr. Oncol.* 18: e9-e18.

Fujita A, Sato JR, Rodrigues LO, Ferreira CE, et al. (2006). Evaluating different methods of microarray data normalization. *BMC Bioinformatics* 7: 469.

Helleman J, Jansen MP, Ruigrok-Ritstier K, van Staveren IL, et al. (2008). Association of an extracellular matrix gene cluster with breast cancer prognosis and endocrine therapy response. *Clin. Cancer Res.* 14: 5555-5564.

Helms MW, Prescher JA, Cao YA, Schaffert S, et al. (2010). IL-12 enhances efficacy and shortens enrichment time in cytokine-induced killer cell immunotherapy. *Cancer Immunol. Immunother.* 59: 1325-1334.

Horii Y, Takei H, Koibuchi Y, Horiguchi J, et al. (2006). The regulatory effect of tamoxifen on fibronectin expression in estrogen-dependent MCF-7 breast carcinoma cells. *Oncol. Rep.* 15: 1191-1195.

Insalaco L, Di Gaudio F, Terrasi M, Amodeo V, et al. (2012). Analysis of molecular mechanisms and anti-tumoural effects of zoledronic acid in breast cancer cells. *J. Cell Mol. Med.* 16: 2186-2195.

John MJ, Rajasekhar R and Mathews V (2008). Autoimmune lymphoproliferative syndrome (ALPS): a rare cause of immune cytopenia. *Indian Pediatr.* 45: 148-150.

Kerr MK (2003). Linear models for microarray data analysis: hidden similarities and differences. *J. Comput. Biol.* 10: 891-901.

Krueger GR, Brandt ME, Wang G, Berthold F, et al. (2002). A computational analysis of Canale-Smith syndrome: chronic lymphadenopathy simulating malignant lymphoma. *Anticancer Res.* 22: 2365-2371.

Landstrom M, Bergh A, Thornell LE, Virtanen I, et al. (1992). Estrogen treatment of Dunning tumors in castrated rats: qualitative and quantitative morphology. *Prostate* 20: 199-211.

Lockett KL, Snowhite IV and Hu JJ (2005). Nucleotide-excision repair and prostate cancer risk. *Cancer Lett.* 220: 125-135.

Misek DE and Kim EH (2011). Protein biomarkers for the early detection of breast cancer. *Int. J. Proteomics* 2011: 343582.

Perez EA, Reinholz MM, Hillman DW, Tenner KS, et al. (2010). HER2 and chromosome 17 effect on patient outcome in the N9831 adjuvant trastuzumab trial. *J. Clin. Oncol.* 28: 4307-4315.

Saracci R (2008). International Agency for Research on Cancer at a crossroads. *Lancet* 372: 291-292.

Slamon DJ, Clark GM, Wong SG, Levin WJ, et al. (1987). Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 235: 177-182.

Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, et al. (2001). Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N. Engl. J. Med.* 344: 783-792.

Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, et al. (2011). The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* 39: D561-D568.

Troyanskaya O, Cantor M, Sherlock G, Brown P, et al. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics* 17: 520-525.

Turner BC, Krajewski S, Krajewska M, Takayama S, et al. (2001). BAG-1: a novel biomarker predicting long-term survival in early-stage breast cancer. *J. Clin. Oncol.* 19: 992-1000.

Vargas AC, McCart Reed AE, Waddell N, Lane A, et al. (2012). Gene expression profiling of tumour epithelial and stromal compartments during breast cancer progression. *Breast Cancer Res. Treat.* 135: 153-165.

Veiseh M and Turley EA (2011). Hyaluronan metabolism in remodeling extracellular matrix: probes for imaging and therapy of breast cancer. *Integr. Biol.* 3: 304-315.

Vogel CL, Cobleigh MA, Tripathy D, Gutheil JC, et al. (2002). Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. *J. Clin. Oncol.* 20: 719-726.

Weigel MT and Dowsett M (2010). Current and emerging biomarkers in breast cancer: prognosis and prediction. *Endocr. Relat. Cancer* 17: R245-R262.

Xie C, Mao X, Huang J, Ding Y, et al. (2011). KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 39: W316-W322.

Zehentner BK, Persing DH, Deme A, Toure P, et al. (2004). Mammaglobin as a novel breast cancer biomarker: multigene reverse transcription-PCR assay and sandwich ELISA. *Clin. Chem.* 50: 2069-2076.

Zhu L, Lam CK and Chow LW (2004). Sentinel lymph node biopsy or detection of micrometastasis in bone marrow: which might be an alternative to axillary lymph node dissection in breast cancer patients? *Asian J. Surg.* 27: 279-283.