Bioinformatics analysis of differentially expressed miRNA-related mRNAs and their prognostic value in breast carcinoma

GUO-MING ZHANG1,2, HEMANT GOYAL3* and LEI-LEI SONG4*

1Department of Laboratory Medicine, Shuyang People’s Hospital; 2Department of Laboratory Medicine, Shuyang Affiliated Hospital of Xuzhou Medical University, Shuyang, Jiangsu 223600, P.R. China; 3Department of Internal Medicine, Mercer University School of Medicine, Macon, GA 31201, USA; 4Department of Laboratory Medicine, The 82th Hospital of the Chinese People’s Liberation Army, Huai’an, Jiangsu 223001, P.R. China

Received November 14, 2017; Accepted April 17, 2018

DOI: 10.3892/or.2018.6393

Correspondence to:
Dr Guo-Ming Zhang, Department of Laboratory Medicine, Shuyang People's Hospital, 9 Yingbin Road, Shuyang, Jiangsu 223600, P.R. China
E-mail: zly52120@163.com

Dr Lei-Lei Song, Department of Laboratory Medicine, The 82th Hospital of the Chinese People’s Liberation Army, 100 Jiankang Dong Road, Huai'an, Jiangsu 223001, P.R. China
E-mail: songleileiyouxiang@163.com

*Contributed equally

Key words: breast carcinoma, differentially expressed genes, protein-protein interaction, microRNA, Gene Expression Omnibus, Kaplan-Meier plotter

Abstract. Breast carcinoma is one of the most common types of malignant neoplasms, and is associated with high rates of morbidity and mortality. Altered gene expression is critical in the development of breast cancer. To identify the important differentially expressed genes and microRNAs in breast carcinoma, mRNA (GSE26910, GSE42568, and GSE89116) and microRNA (GSE35412) microarray datasets were downloaded from the Gene Expression Omnibus database. The differentially expressed microRNA expression data were extracted with GEO2R online software. The DAVID online database was used to perform a function and pathway enrichment analysis of the key identified differentially expressed genes. A protein-protein interaction (PPI) network was constructed using the STRING online database, and visualized in Cytoscape software. The effect of the expression level of the key identified genes on overall survival (OS) time was analyzed by using the Kaplan-Meier Plotter online database. Furthermore, the online miRNA databases TargetScan, microT-CDS, and TarBase were used to identify the target genes of the differentially expressed miRNAs. A total of 254 differentially expressed genes were identified, which were enriched in cell adhesion, polysaccharide binding, extracellular region part and ECM-receptor interactions. The PPI network contained 250 nodes and 375 edges. Five differentially expressed genes were found to be significantly negatively correlated with the differentially expressed miRNAs, which were potentially also target genes for miRNAs. Four of the five genes, including AKAP12, SOPB, TCF7L2, COL12A1 and TXNIP were downregulated, and were associated with the OS of patients with breast carcinoma. In addition, a total of 130 differentially expressed miRNAs were identified. In conclusion, these results constitute a novel model for miRNA-mRNA differential expression patterns, and further studies may provide potential targets for diagnosing and understanding the mechanisms of breast carcinoma.

Introduction

Breast cancer is one of the most common types of malignancy among women, and is associated with high mortality. Globally, more than 1.7 million individuals are diagnosed with breast cancer annually, and approximately 521,000 individuals succumb to the disease (1). In recent decades, the incidence of breast cancer has increased, with almost one-tenth of all newly diagnosed cancers worldwide originating in the breast (2). Changes in reproductive and lifestyle characteristics are contributing to the increased morbidity and mortality rates of breast carcinoma. However, the exact molecular mechanisms underlying breast carcinoma are not fully understood.

At present, three major protein markers: estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor (EGF) receptor 2 (HER2), are used for determining the classification, treatment and prognosis of breast carcinoma (3). However, there are no identified protein markers for the early diagnosis and treatment of breast carcinoma (4). Therefore, it is necessary to further investigate the molecular regulatory mechanisms of breast carcinoma, and identify molecular markers that can be used for early diagnosis and monitoring.

In recent years, DNA microarray analysis has been developed as a rapid, high-throughput detection technology to simultane-
ously monitor the differential expression of numerous genes or miRNAs in oncology research, including in the field of breast cancer (5-8). miRNAs are small ~21-nt RNAs involved in post-transcriptional gene regulation. It has been demonstrated that miRNAs pair with the mRNA 3'-untranslated region (UTR) of target genes to regulate their expression, including in cancer. In breast cancer, multiple miRNAs have been identified to regulate the expression of target genes, including miRNA-155, miRNA-675, miRNA-519a and miRNA-31 (9-12), among others.

Although DNA microarray application in oncology research has been widely recognized, the test has high variability. Therefore, in our research, we identified differentially expressed genes and microRNAs by analyzing three breast cancer mRNA microarray datasets, and one microRNA dataset. We then aimed to identify the key genes in breast cancer with survival, mRNA-microRNA interaction, ontology enrichment and network analyses.

Materials and methods

Microarray data. The GEO (https://www.ncbi.nlm.nih.gov/gds. Accessed Jan. 26, 2018) is a free international public repository of high-throughput functional genomics data, including microarray and next-generation sequencing data. In this study, we used three gene expression profiles (GSE26910, GSE42568 and GSE89116) and one miRNA expression profile (GSE35412) from GEO.

The GSE26910 dataset was comprised of six breast cancer and six para-carcinoma tissue sample mRNA expression profiles (13); GSE42568 included 104 breast cancer and 17 para-carcinoma tissue sample mRNA expression profiles (14); GSE89116 contained 30 breast cancer and nine para-carcinoma tissue sample mRNA expression profiles (15); and GSE35412 was comprised of 29 breast cancer and 21 para-carcinoma tissue sample mRNA expression profiles (16). All datasets were downloaded in a processed and normalized format.

Data processing. The GEO2R (https://www.ncbi.nlm.nih.gov/geo/geo2r. Accessed Jan. 26, 2018) is an online data analysis tool that can be utilized to analyze GEO data series obtained under the same experimental conditions (17). In this study, GEO2R was used to identify the differentially expressed miRNAs and genes between the breast cancer and para-carcinoma tissue sample expression profiles. Adjusted P-values (adj.p) were calculated using the Benjamini and Hochberg false discovery rate method to correct for the likelihood of false positive results. An adj.P<0.01 and |logFC| >1 were set as the cut-off criteria for differential expression.

Functional and pathway enrichment analysis of the differentially expressed genes. The DAVID (https://david-d.ncifcrf.gov/summary.jsp. Accessed Jan. 26, 2018) online database is an online program that provides comprehensive gene annotation tools (18). The database was used to perform gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses. P<0.05 was applied as the cut-off criterion.

Protein-protein interaction (PPI) network construction and module selection for the differentially expressed genes. The construction of a network of interactions between proteins can establish a framework for the study of molecular mechanisms. In this study, a PPI network of the differentially expressed genes was constructed using the Search Tool for the Retrieval of Interacting Genes and Proteins (STRING) database (https://string-db.org. Accessed Jan. 26, 2018) (19), followed by visualization using Cytoscape software (20). The confidence score ≥0.7 was set as the cut-off criterion. The PPI network module selection criteria included a degree cut-off=2, node score cut-off=0.2, k-core=2 and maximum depth=100 (21).


Analysis of the effect of the differentially expressed genes on overall survival (OS). In the online database, Kaplan-Meier (KM) Plotter (http://www.kmplot.com. Accessed Jan. 26, 2018), the impact of 54,675 genes on the survival time of cancer patients was evaluated by analyzing the data from 10,188 cancer samples, including 4,142 breast, 1,648 ovarian, 2,437 lung and 1,065 gastric cancer sample microarray expression profiles (22). In the present study, we divided breast carcinoma patients into two groups depending on the expression of specific genes (high vs. low expression). The KM Plotter database was utilized to analyze the OS of breast cancer patients. Hazard ratios (HR) with 95% confidence intervals (CI) and a log-rank P-value were calculated and displayed.

Results

Identification of differentially expressed genes in three GEO datasets. The total number of differentially expressed genes was 1,293, 4,251 and 1,130 from GSE26910, GSE42568 and GSE89116, respectively. A total of 254 genes showed the same expression trend in all three data sets (Fig. 1). This
included 44 upregulated and 210 downregulated differentially expressed genes.

Construction of a PPI network of the differentially expressed genes. The PPI network of the differentially expressed genes consisted of 250 nodes and 375 edges, including 20 upregulated and 139 downregulated genes (Fig. 2).

Function and pathway enrichment analysis. We used the DAVID database to identify enriched functions and pathways of the differentially expressed genes, in order to further understand their function. The differentially expressed genes were the most significantly enriched in cell adhesion (biological process category), polysaccharide binding (molecular function category) and extracellular region part (cellular component category). Function and pathway enrichment analysis. We used the DAVID database to identify enriched functions and pathways of the differentially expressed genes, in order to further understand their function. The differentially expressed genes were the most significantly enriched in cell adhesion (biological process category), polysaccharide binding (molecular function category) and extracellular region part (cellular component category).
In addition, these genes were significantly associated with 10 KEGG pathways, including ECM-receptor interactions, complement and coagulation cascades and adhesion spots (Fig. 3).

Prediction of the target genes of the differentially expressed miRNAs. A total of 130 differentially expressed miRNAs were detected in the GSE35412 dataset, including 17 upregulated and 113 downregulated miRNAs. miR-183-5p was the most significantly upregulated miRNA, whereas miR-129-1-3p was the most significantly downregulated (Fig. 4). Subsequently, target genes of the differentially expressed miRNAs were obtained from online databases (TargetScan, microT-CDS and Tarbase). We determined that AKAP12, SOBP, TCF7L2, COL12A1 and TXNIP were the target genes for multiple differentially expressed miRNAs (Fig. 5).

Survival analysis of key identified differentially expressed genes. The prognostic values of five key genes in the PPI network were assessed from KMplots. The OS rate for breast cancer patients was analyzed based on the low and high expression of the key genes. The results showed that high AKAP12 mRNA expression [HR 0.67 (95% CI, 0.58-0.79), P=6.1e-07] was associated with an improved OS for breast carcinoma patients (Fig. 6), as was high expression of TXNIP [HR 0.66 (95% CI, 0.59-0.74), P=1e-13], SOBP [HR 0.83 (95% CI, 0.74-0.92), P=0.00059] and TCF7L2 [HR 0.76 (95% CI, 0.68-0.85), P=6e-07] (Fig. 6).

Discussion
Although our understanding of the pathogenesis and clinical treatment of breast cancer has made significant progress,
the overall mortality rate for breast cancer has not improved significantly, which can be attributed to the lack of molecular markers for effective diagnosis and treatment. Therefore, it is important to explore the molecular markers of breast carcinoma to improve the survival rate and prevention of patients.

Microarray technology has been applied to detect the genetic changes associated with the progression of various types of malignancy. Microarray technology has also been widely used to select molecular markers for determining the diagnosis, treatment and prognosis of tumors. In this study, we screened a total of 254 differentially expressed genes through analyzing three mRNA datasets, which included 44 upregulated genes and 210 downregulated genes. These differentially expressed genes were significantly enriched in cell adhesion, polysaccharide binding, extracellular region part and ECM-receptor interaction, which are all associated with the pathogenesis of carcinomas.

We also analyzed a miRNA dataset, in which we identified 130 differentially expressed miRNAs, including 17 upregulated and 113 downregulated miRNAs, in breast carcinoma. miR-183-5p was the most significantly upregulated miRNA, while miR-129-3p was the most markedly downregulated miRNA. miRNAs are small, non-coding RNAs of ~22 nucleotides in length, which regulate gene expression by targeting the 3’UTR of target mRNAs, resulting in their degradation or the inhibition of translation. Previous research has suggested that the dysregulation of miRNAs is involved in the pathogenesis of many types of cancer, including breast carcinoma. For example, it has been demonstrated that miR-21, miR-210 and miR-221 are upregulated in triple-negative primary breast carcinomas (23). In addition, the overexpression of miR-301 is considered a negative prognostic index for lymph node-negative invasive ductal breast carcinoma. Shi et al identified that miR-301 is a crucial oncogene in breast carcinoma that promotes nodal and distant relapse via multiple pathways and mechanisms (24).

As miRNAs negatively regulate the expression of their target genes, we analyzed the correlation of upregulated genes with downregulated miRNAs, downregulated genes and upregulated miRNAs. Notably, we identified that AKAP12, SOBP, TCF7L2 and TXNIP were potentially the common targets of hsa-miRNA-183-5p, hsa-miRNA-454-3p and hsa-let-7g-5p among the downregulated genes. COL12A1 was potentially the common target of hsa-miRNA-139-3p and hsa-miRNA-654-3p among the upregulated genes. Subsequently, survival analysis of the relationship between the postoperative survival of patients and the expression of these genes suggested that four genes were closely associated with improved OS of breast cancer.

Figure 6. Prognostic value of four genes in breast carcinoma patients. Plots of the prognostic value of (A) AKAP12, (B) TXNIP, (C) SOBP and (D) TCF7L2 genes were obtained from www.kmplot.com. The corresponding Affymetrix IDs were 227529_s_at (AKAP12), 201010_s_at (TXNIP), 218974_at (SOBP) and 216037_x_at (TCF7L2).
TCF7L2 is a part of the Wnt/β-catenin signaling pathway, which plays an important role in cell proliferation, apoptosis, and angiogenesis (25). Studies have also shown that AKAP12 is involved in multiple signaling pathways through regulating protein kinases A and C. AKAP12 is a tumor metastasis inhibitory factor, which is associated with carcinoma susceptibility and tumor cell behavior in a variety of tumor types, including breast carcinoma (26-28). AKAP12 has also been associated with miRNAs in various types of disease. In liver cirrhosis, the interaction of miRNA-183 or miRNA-186 can downregulate the expression of AKAP12 (29). In addition, Xia et al. indicated that the overexpression of miR-103 can promote cell proliferation and inhibit apoptosis by downregulating AKAP12 expression in hepatocellular carcinoma cell lines (30).

SopB (also known as SigD) is an effector of the Salmonella typhimurium Type III secretion system that acts on phospholipids in the host cell membrane (31,32). SopB may induce epithelial-mesenchymal transition (EMT), which is also associated with malignant disease. SopB plays a central role in activating Wnt/β-catenin signaling, which can induce cell transformation and Wnt/β-linked regulatory signaling transduction. SopB-dependent activation of Akt kinases can lead to the inhibitory phosphorylation of GSK3β, which further induces cytosolic β-catenin and Wnt/β-catenin-mediated EMT (33).

The transcription factor 7-like 2 (TCF7L2) gene is located on the long arm of chromosome 10q25.2 (previously called TCF-4). TCF7L2 is a part of the Wnt/β-catenin signaling pathway, which plays an important role in the regulation of cell development and growth (34,35). In addition, epidemiological studies have shown that TCF7L2 gene polymorphisms are associated with increased susceptibility to carcinomas, including the breast (35,36). Additionally, TCF7L2 can increase the expression of genes involved in the proliferation, apoptosis, invasion and metastasis of tumor cells (37,38). TCF7L2 synthesis is directly regulated by miR-21 transcription (39).

In cervical carcinoma, the expression levels of miR-328 and TCF7L2 are negatively correlated. Furthermore, miR-328 can reduce the expression of TCF7L2 to affect the treatment of cervical carcinoma (40). Cervical carcinoma metastasis and progression may also be inhibited by miR-212, through its direct targeting of TCF7L2 expression (41). In addition, miR-181a-5p may regulate the Wnt signaling pathway through the direct targeting of TCF7L2, promoting 3T3-L1 preadipocyte differentiation and adipogenesis (42).

TXNIP (also known as VDUP-1 or TBP-2) is proapoptotic, and inhibits growth and metastasis (43). TXNIP has a variety of functions, including an important regulatory role in the redox equilibrium, and can increase the production of reactive oxygen species (ROS) to induce apoptosis through oxidative stress. TXNIP is a major tumor suppressor gene that is downregulated in a variety of solid tumors, including breast carcinoma (44-46). There is a correlation between the expression of TXNIP and the metastasis and survival prognosis of breast carcinoma (45). TXNIP also plays a critical role in the treatment of HER-1/HER-2-positive tumors, and is a potential prognostic indicator in breast carcinoma (47). TXNIP has been associated with a variety of miRNAs and has been demonstrated as a target of miR-342 (48), miR-135a (49) and miR-20a (50). The miR-373 expression is negatively correlated with the expression of TXNIP, and activation of the miR-373-TXNIP signal transduction axis is associated with a poor outcome in breast carcinoma (51). This may be mediated through an effect on the invasion and migration of breast carcinoma cells, which is associated with the prognosis of breast carcinoma patients (52).

In summary, the present study intended to identify the differentially expressed genes in breast carcinoma and thus find the potential biomarkers for predicting disease progression using comprehensive bioinformatic analyses. In this study, a total of 254 differentially expressed genes and 130 differentially expressed miRNAs were screened; AKAP12, SOPB, TCF7L2 and TXNIP, and several miRNAs, including miR-183-5p, let-7g-5p and miR-454-3p, may be key breast carcinoma-associated genes. Our results suggested that data mining and integration is a useful tool to predict the progression of breast carcinoma, and to identify the mechanisms of the occurrence and development of tumors. To apply these gene expression profiles in clinical practice, it is necessary to improve the reliability and reproducibility of this model within dependent datasets in the future. Nevertheless, our research may provide new information for the diagnosis and treatment of breast carcinoma patients.

Acknowledgements

Not applicable.

Funding

No funding was received.

Availability of data and materials

The datasets used during the present study are available from the corresponding author upon reasonable request.

Authors’ contributions

GMZ and LLS conceived and designed the study. HG was the corresponding author upon reasonable request.

Availability of data and materials

The datasets used during the present study are available from the corresponding author upon reasonable request.

GMZ and LLS read and approved the manuscript and agree to be accountable for all aspects of the research in ensuring that the accuracy or integrity of any part of the work is appropriately investigated and resolved.

Ethics approval and consent to participate

The study was approved by the Institutional Review Board of the Department of Shuyang People's Hospital (Shuyang, China).

Consent for publication

Not applicable.


Competing interests

The authors declare that they have no competing interests.

References


