The construction and analysis of ceRNA networks in invasive breast cancer: a study based on The Cancer Genome Atlas

Chundi Gao1,∗
Huayao Li1,∗
Jing Zhuang2,3
HongXiu Zhang4
Kejia Wang5
Jing Yang2
Cun Liu6
Lijuan Liu2,3
Chao Zhou2,3
Changgang Sun2,3

1College of First Clinical Medicine, Shandong University of Traditional Chinese Medicine, Jinan 250014, People’s Republic of China; 2Department of Oncology, Weifang Traditional Chinese Hospital, Weifang 261041, People’s Republic of China; 3Department of Oncology, Affiliated Hospital of Weifang Medical University, Weifang 261031, People’s Republic of China; 4Institute of Virology, Jinan Center for Disease Control and Prevention, Jinan 250021, People’s Republic of China; 5College of Basic Medicine, Qingdao University, Qingdao 266071, People’s Republic of China; 6College of Traditional Chinese Medicine, Shandong University of Traditional Chinese Medicine, Jinan 250014, People’s Republic of China

∗These authors contributed equally to this work

Background: Studies have shown that long noncoding RNAs (lncRNAs) make up the major proportion of the ceRNA network and can regulate gene expression by competitively binding to miRNAs. This reveals the existence of an RNA-miRNA regulatory pathway and is of great biological significance. CeRNAs, as competitive endogenous RNAs, have revealed a new mechanism of interaction between RNAs. Until now, the role of lncRNA-mediated CeRNAs in breast cancer and their regulatory mechanisms have been elucidated to some extent.

Purpose: In this study, comprehensive analysis of large-scale invasive breast cancer samples in TCGA were conducted to further explore the developmental mechanism of invasive breast cancer and the potential predictive markers for invasive breast cancer prognosis in the ceRNA network.

Methods: Abnormal expression profiles of invasive breast cancer associated mRNAs, lncRNAs and miRNAs were obtained from the TCGA database. Through further alignment and prediction of target genes, an abnormal lncRNA-miRNA-mRNA ceRNA network was constructed for invasive breast cancer. Through the overall survival analysis, Identification diagnostic biomarkers for invasive breast cancer patients. In addition, we used Cytoscape plugin in BinGo for the different mRNA performance functional cluster analysis.

Results: Differential analysis revealed that 1059 lncRNAs, 86 miRNAs, and 2138 mRNAs were significantly different in invasive breast cancer samples versus normal samples. Then we construct an abnormal lncRNA-miRNA-mRNA ceRNA network for invasive breast cancer, consisting of 90 DeIncRNAs, 18 DEmiRNAs and 26 DEmRNAs. Further, 4 out of 90 IncRNAs, 3 out of 26 mRNAs, and 2 out of 18 miRNAs were useful as prognostic biomarkers for invasive breast cancer patients (P value < 0.05). It is worth noting that based on the ceRNA network, we found the LINC00466-Hsa-mir-204-5p NTRK2 LINC00466-Hsa-mir-204-5p NTRK2 axis was present in 9 RNAs associated with the prognosis of invasive breast cancer.

Conclusion: This study provides an effective bioinformatics basis for further understanding of the molecular mechanism of invasive breast cancer and for predicting outcomes, which can guide the use of invasive breast cancer drugs and subsequent related research.

Keywords: invasive breast cancer, cancer genome atlas, lncRNA-miRNA-mRNA ceRNA network, bioinformatics, diagnosis and prognosis biomarkers

Introduction

Breast cancer, as the most frequent cancer affecting women worldwide, has strong invasiveness and metastatic qualities,1 and the incidence and mortality rate of breast cancer continue to increase.2 Currently, chemotherapy and surgery are the main treatment methods for breast cancer.1 However, because of the high rate of tumor recurrence and resistance to chemotherapy, the therapeutic effect and prognosis of breast cancer...
have always failed to achieve satisfactory results. Therefore, an in-depth understanding of the molecular pathogenesis of breast cancer and the identification of new candidate therapeutic targets and biomarkers is urgently needed for breast cancer treatment.

This in-depth study on the molecular pathogenesis of tumors based on bioinformatics analysis exploits an important method for tumor research. It can not only explore the molecular pathogenesis of tumors in depth but also identify novel biomarkers for the onset and prognosis of tumors. The screening of biomarkers for the diagnosis and prognosis of breast cancer has a crucial role in the early diagnosis of breast cancer, in improving the prognosis of breast cancer and reducing breast cancer mortality.

In recent years, sufficient research evidence has shown that IncRNAs play key roles in the physiological and pathological processes of various organisms, such as embryonic development, organogenesis, and tumorigenesis. IncRNAs function as ceRNAs to communicate with mRNAs by competing for shared miRNAs and many IncRNAs serve as master regulators that affect the expression levels of dozens or even hundreds of target genes. A large number of studies have shown that the abnormal expression of ceRNAs is closely related to the occurrence, development, and prognosis of tumors, including breast cancer. For example, Zheng et al found that activation of the STARD13-associated ceRNA network attenuates YAP/TAZ nuclear accumulation and transcriptional activity in breast cancer by co-modulating Hippo and Rho-GTPase/F-actin signaling and proposed the possibility of promoting the inhibition of breast cancer stem cell traits via the STARD13-associated ceRNA network. Li et al. found that growth arrest-specific transcript 5 (GAS5), as a ceRNA-antagonizing tumor promoter of miR-196a-5p-expressing triple-negative breast cancer (TNBC) cells, can suppress TNBC progression by competitively binding miR-196a-5p and, therefore, GAS5 may be a prognostic biomarker of TNBC.

ceRNA networks can link the function of protein-coding mRNAs with the function of ncRNAs, such as miRNAs and IncRNAs. In this study, we first obtained meaningful differentially expressed IncRNAs, miRNAs, and mRNAs based on The Cancer Genome Atlas (TCGA) database (https://cancergenome.nih.gov/). Subsequently, by comparing, predicting, and integrating differentially expressed RNAs, we constructed a differentially expressed IncRNA–miRNA–mRNA ceRNA network for invasive breast cancer. Finally, to bring the results closer to the clinical treatment of invasive breast cancer, we performed an overall survival analysis of IncRNAs, miRNAs, and mRNAs in the ceRNA network to identify invasive breast cancer-related prognostic biomarkers. Through this study, we can not only further understand the molecular mechanism of invasive breast cancer but also provide potential IncRNA, miRNA and mRNA diagnostic biomarkers for invasive breast cancer, which will contribute to the early diagnosis, treatment, and prognosis of invasive breast cancer.

Methods

Data processing and differential expression analysis

TCGA database provides the analysis of high-throughput data for various genomic changes, including the expression data of mRNA, IncRNA, and miRNA. By combining bioinformatics analysis with the clinical information of patients, it lays the foundation for improving cancer prevention and the early detection and treatment of diseases. In this study, from TCGA database, we downloaded expression data on mRNA, IncRNA, and miRNA from invasive breast cancer. Among them, mRNA and IncRNA expression data from 1,208 samples, including 112 normal samples and 1,096 invasive breast cancer samples, and miRNA expression data from 1,193 samples, including 103 normal samples and 1,090 invasive breast cancer samples, were acquired. Based on the edger software package, the downloaded data were normalized and analyzed for differences to obtain differentially expressed IncRNA, miRNA, and mRNA molecules. The relevant data provided by TCGA are publicly available and open-ended and, therefore, do not require the approval of the local ethics committee.

Prediction of IncRNA–miRNA and miRNA–mRNA interactions

miRcode (http://www.mircode.org/) provides “whole transcriptome” human miRNA target predictions based on the comprehensive GENCODE gene annotation. In this study, miRcode online software was used to predict the interaction between IncRNAs and miRNAs and DEIncRNAs and DEMiRNAs that could be compared successfully were obtained. In addition, the target mRNAs of miRNAs were predicted using TargetScan (http://www.targetscan.org/), miRDB (http://www.mirdb.org/miRDB/), and miRTarBase (http://miRTarBase.mbc.nctu.edu.tw) online analysis tools. To further improve the reliability of bioinformatics analysis, a Venny diagram was used to obtain the portion of the
target mRNA that overlaps with the differentially expressed mRNA in invasive breast cancer and was further analyzed as DEmRNA. Finally, we established matched DElncRNA–DEmiRNA and DEmiRNA–DEmRNA pairs.

Construction of the DElncRNA–DEmiRNA–DEmRNA ceRNA network

Based on the relationships between DElncRNAs, DEmiRNAs, and DEmRNAs and the obtained DElncRNA–DEmiRNA and DEmiRNA–DEmRNA pairs, we used Cytoscape v3.017 to construct and visualize the DElncRNA–DEmiRNA–DEmRNA ceRNA network. Cytoscape is an open source software platform for visualizing molecular interaction networks and biological pathways and integrating these networks with annotations, gene expression profiles, and other state data. Cytoscape plug-in (BiNGO)18 provides a comprehensive set of functional annotation tools for investigators to understand the biological meaning behind Gene Ontology (GO) and the Kyoto Encyclopedia list of genes. To further understand the function of these differentially expressed genes, functional enrichment analysis was performed using the Cytoscape plug-in BiNGO, which involves the following three parts: molecular function (MF), biological process (BP), and cellular component (CC). P<0.05 was selected as the cutoff condition when GO conditions were selected.

Associations of DElncRNAs, DEmiRNAs, and DEmRNAs and patient prognosis

Each sample in TCGA was independent of each other, and it contained all sample data, such as gene expression, prognosis, and survival time. From TCGA database, we obtained clinical information about invasive breast cancer and combined the expression data of DElncRNAs, DEmiRNAs, and DEmRNAs with the clinical data. Cox regression analysis of survival packages and Kaplan–Meier curves was used to assess the relationship between the expression of DElncRNAs, DEmiRNAs, and DEmRNAs and invasive breast cancer patients’ overall survival. In addition, DElncRNAs, DEmiRNAs, and DEmRNAs that were significantly associated with overall survival were identified as prognostic biomarkers.

Results

TCGA-related data analysis and differential expression analysis

We calculated mRNA, miRNA, and lncRNA expression profiles between invasive breast cancer samples and non-malignant samples obtained from TCGA. Based on the relevant limma software package, data on differential expression were extracted and analyzed. Using log 2-fold change ≥2.0 and P-value <0.01 as screening cutoff conditions, 2,138 differentially expressed mRNAs, 1,059 differentially expressed lncRNAs, and 86 differentially expressed miRNAs were screened out (Supplementary Tables 1–3). Among them, the 2,138 differentially expressed mRNAs included 1,375 upregulated mRNAs and 763 downregulated mRNAs and the 1,059 differentially expressed lncRNAs included 842 upregulated lncRNAs and 217 downregulated lncRNAs. In addition, 86 differentially expressed miRNAs contained 67 upregulated miRNAs and 19 downregulated miRNAs. Cluster analysis of RNA heat maps and volcano diagrams are shown in Figures 1–3.

Matching of DElncRNA–DEmiRNA and prediction of DEmiRNA–DEmRNA interactions

Based on the miRcode online software, we predicted and compared DElncRNA–DEmiRNA pairs and found successful comparisons of 90 DElncRNAs and 18 DEmiRNAs. The target genes of 18 DEmiRNAs were predicted using TargetScan, miRDB, and miRTarBase target gene prediction software. The target genes predicted by the three online software packages were simultaneously used as prediction results for target genes of DEmiRNAs, and finally, 506 target genes were obtained. To obtain a differentially expressed miRNA–mRNA interaction pair in invasive breast cancer, the predicted target gene was matched with differentially expressed mRNAs and the results showed that 26 DEmRNAs could be successfully matched with DEmiRNAs. Finally, 323 DElncRNA–DEmiRNA pairs and 28 DEmiRNA–DEmRNA interaction pairs were obtained.

Construction of the ceRNA network and functional enrichment analysis

To further explore the mechanism of invasive breast cancer development, a ceRNA network was constructed based on the interactions of 323 DElncRNA–DEmiRNA pairs and 28 DEmiRNA–DEmRNA pairs between 90 DElncRNAs, 18 DEmiRNAs, and 26 DEmRNAs. We used Cytoscape to visualize the network (Figure 4). To further understand the functional role of DEmRNAs in invasive breast cancer, we used the BiNGO plugin for functional enrichment analysis of these DEmRNAs. The results showed that the DEmRNAs were particularly rich in the MF, BP, and cell components’ classifications. In the BP group, these genes were mainly enriched...
in the regulation of DNA-dependent transcription, the RNA metabolic process, and gene expression, among others (Figure 5A). The MF was mainly enriched in multiple binding, such as chromatin DNA binding, DNA regulatory region binding, transcription regulator activity, and transcription factor binding (Figure 5B). In addition, CC terms were mainly involved in the cytoplasm and nucleoplasm (Figure 5C).

**Prognostic overall survival assessment of lncRNAs, miRNAs, and mRNAs**

Survival analysis based on a survival package found that using \( P<0.05 \) as the screening criteria, 4 lncRNAs (AC007731.1, AL356479.1, LINC00460, and LINC00466) among 90 DElncRNAs were closely associated with the overall survival of invasive breast cancer patients. High expression of LINC00460 and LINC00466 was associated with high survival rates. For AC007731.1, however, low expression was associated with a high survival rate. The effect of AL356479.1 on the survival rate of patients changed over time. High expression during the early stage was associated with a higher survival rate. Two of the 18 DEMiRNAs (hsa-mir-204 and hsa-mir-301b) were associated with prognosis. High expression of hsa-mir-204 suggested a long survival time, while high expression of hsa-mir-301b suggested that the survival time was relatively short. In addition, among the 26 DEMiRNAs, 3 mRNAs (KPNA2, NTRK2, and SFRP1) were associated with overall survival. High expression of NTRK2 and SFRP1 and low expression of KPNA2 showed an association with a higher survival rate (Figure 6). It is worth noting that based on the ceRNA network, we found that the LINC00466–hsa-mir-204–NTRK2 axis was linked to nine RNAs associated with the prognosis of invasive breast cancer.

**Discussion**

As one of the most invasive malignant tumors in women, it is very urgent and meaningful to deeply explore the molecular pathogenesis of breast cancer. In this study, we used bioinformatics analysis of the invasive breast cancer-associated lncRNA–miRNA–mRNA ceRNA network to study the molecular pathogenesis of invasive breast cancer. We obtained a total of 90 DElncRNAs, 18 DEMiRNAs, and 26 DEMRNAs, and by analyzing overall survival, 4 lncRNAs (AC007731, AL356479, LINC00460, and LINC00466), 2 miRNAs (hsa-mir-204 and hsa-mir-301b), and 3 mRNAs (KPNA2, NTRK2, and SFRP1) were found to be associated with overall survival in invasive breast cancer patients.
To further investigate the cellular mechanisms involved in invasive breast cancer, GO analysis of 26 DEMRNAs was performed. GO analysis indicated that DEMRNAs were mainly enriched in single-cell proliferation and apoptosis, including DNA-dependent transcription, the processes of RNA metabolism, chromatin DNA binding, DNA regulatory region binding, transcriptional regulation activity, and transcription factor binding, and cytoplasm and nucleoplasm were associated with gene mutations and the malignant proliferation of invasive breast cancer cells, which are consistent with the findings of current invasive breast cancer research.

The lncRNAs are ncRNA molecules with lengths >200 nucleotides. They do not participate in the coding of proteins, but they play an important role in the physiological and pathological processes of the human body. Studies have found that the main functions of lncRNAs occur through their interactions with miRNAs, mainly through miRNA binding to silence the miRNA, thereby indirectly regulating mRNA, DNA, and protein expressions. In humans, lncRNAs are involved in chromosomal rearrangements, gene expression, cell development, proliferation, apoptosis, and other physiological and pathological processes. In recent years, many studies have found that lncRNAs play a crucial role in the molecular pathogenesis of cancer and that the molecular mechanisms of action of lncRNA in breast cancer and other cancers, chromatin modification, and DNA methylation are key epigenetic events that are fundamentally disturbed during the development of cancer.

In this study, we analyzed the overall survival rate of patients associated with DElncRNAs in the ceRNA network. We screened four lncRNAs that correlated with the prognosis of patients with invasive breast cancer. It is noteworthy that among the four lncRNAs, LINC00460 has been confirmed to be involved in the molecular pathogenesis of many kinds of tumors and has a strong relationship with the occurrence, development, proliferation, metastasis, and prognosis of
cancer. Xing et al., using luciferase reporter gene assays and Western blot analysis, revealed that LINC00460 promotes MMP-9 expression by targeting miR-539, thereby acting as an oncogenic RNA in malignant meningiomas and accelerating the proliferation and metastasis of meningiomas. Using real-time quantitative polymerase chain reactions, Transwell assays, Western blots, and immunofluorescence assays, the researchers found that the interference of LINC00460 expression could inhibit the proliferation, invasion, and migration of non-small-cell cancers and interference with the expression of LINC00460 could affect epithelial–mesenchymal transition (EMT)-related proteins. The expression changes were used to confirm that LINC00460 could affect tumor invasion and migration by affecting EMT, so LINC00460’s expression had potential in the prognosis and treatment of non-small-cell lung cancer patients. In addition, LINC00460 has been proven to have roles in esophageal squamous cell carcinoma, laryngeal squamous cell carcinoma, nasopharyngeal carcinoma, and other malignant tumors. It can be used as a new biomarker for diagnosis and prognosis and may have great value in guiding clinical treatment. However, although we identified four IncRNAs that were closely related to the diagnosis and prognosis of invasive breast cancers using the bioinformatics analysis, the complex molecular regulation mechanisms among them require further clinical trials.

The important role of miRNAs in posttranscriptional gene regulation reveals another fine-tuning step in the expression of genetic information. The cellular pathways involved in the regulation of miRNAs play an important role in the physiological and pathological processes of the body’s development. At the molecular level, miRNAs inhibit protein production by affecting the stability of their target mRNAs and/or by downregulating their translation. Although the number of miRNAs is very small, miRNAs can participate in the regulation of cell development, proliferation, apoptosis, differentiation, and metabolism. Increasing numbers of studies have shown that abnormal miRNA expression can participate in many types of cancer. Breast cancer is one of these cancers, and miRNA can either act as an oncogene or tumor suppressor gene in breast cancer. Therefore, as a key factor in signal cascades, miRNAs are potentially useful biomarkers in clinical diagnosis and may become targets and tools for cancer treatment development.

In this study, we have identified two miRNAs related to the diagnosis and prognosis of invasive breast cancer, including hsa-mir-204 and hsa-mir-301b. Shen et al., using
real-time qPCR, MTT assay, flow cytometry, and Western blot analysis, found that miR-204 regulates the biological behavior of breast cancer cells through direct targeting of FOXA1, including cell proliferation, invasion, metastasis, and apoptosis. In addition, hsa-mir-204 can be used to predict the outcomes of patients with clear cell renal cell carcinoma, lung adenocarcinoma, and other cancers.32,36 Fort et al37 found that the correlation between hsa-miR-301b and candidate target genes supports it having a role in the inhibition of tumor suppressor genes in prostate cancer, indicating the oncogenic role of miR-301b in prostate cancer. In addition, these researchers also found that hsa-miR-301b is involved in the pathogenesis of cancer in both pancreatic cancer and thyroid papillary carcinoma.38,39

In this study, the specific molecular pathogenesis of hsa-miR-301b’s involvement in breast cancer needs further investigation. miRNAs regulate gene expression mainly by inducing targeted mRNA degradation. In this study, we identified a total of three mRNAs (KPNA2, NTRK2, and SFRP1) that have been shown to be associated with the prognosis of invasive breast cancer. KPNA2 has been found to play a role in altering cell phenotype and carcinogenesis. Through immunohistochemical evaluation and long-term follow-up of 1,494 invasive breast cancer patients, it was found that KPNA2 expression was associated with the negative estrogen receptor and triple-negative phenotype. Survival analysis showed that KPNA2 expression was associated with a poor prognosis (P<0.0001).40 Studies have found that miR-200c is essential for maintaining the characteristics, behavior, and sensitivity to chemotherapy in invasive breast cancer epithelial cells.

NTRK2, a direct target of miR-200c, codes for proteins expressed in mesenchymal or neuron-derived cells and may

Figure 4 The dysregulated lncRNA–mRNA–miRNA ceRNA network.

Notes: Diamond denotes lncRNA, square represents miRNA, and round rectangle represents mRNA. All shapes in red and blue colors stand for upregulation and downregulation, respectively.
contribute to the relative risk for symptom burden during and shortly after the period of chemotherapy in women with early-stage breast cancer.\textsuperscript{41-43} Li et al.\textsuperscript{44} using the gene microarray analysis, found that miR-105/93-3p activates Wnt/\(\beta\)-catenin signaling by the downregulation of SFRP1, thereby promoting stem cells, chemoresistance, and TNBC cell metastasis. The has-miR-542-5p is a predictive biomarker and potential target for the treatment of breast cancer patients. Studies have confirmed that has-miR-542-5p functions through a mechanism involving its target gene SFRP1.\textsuperscript{45} Therefore, KPNA2, NTRK2, and SFRP1 are involved in the pathogenesis, infiltration, and metastasis of invasive breast cancer.

It is worth noting that based on the ceRNA network, we compared and identified nine biomarkers related to the prognosis of invasive breast cancer. We obtained a meaningful lncRNA–miRNA–mRNA regulatory axis, namely, LINC00466–hsa-mir-204–NTRK2. This suggests that LINC00466, hsa-mir-204, and NTRK2 are involved in the molecular mechanism of invasive breast cancer, while hsa-mir-204 and NTRK2 are involved in or closely related to the molecular mechanism of invasive breast cancer. Thus, LINC00466 is a new lncRNA that can be used in the prediction of invasive breast cancer and may become a new target for the treatment of invasive breast cancer.

Figure 5 Gene Ontology analysis of differentially expressed mRNAs associated with breast cancer.

Note: (A) The enrichment analysis of biological processes; (B) enrichment analysis of molecular function; and (C) enrichment analysis of cellular component.
discovery of this axis provides new insights into the molecular mechanisms of invasive breast cancer and may provide new approaches to its treatment.

**Conclusion**

In this study, we first constructed an invasive breast cancer-associated ceRNA network consisting of 90 DElncRNAs, 18 DEMiRNAs, and 26 DEMRNAs. Based on the total survival analysis, we obtained a total of nine prognostic biomarkers associated with invasive breast cancer, including AC007731, AL356479, LINC00460, LINC00466, hsa-mir-204, hsa-mir-301b, KPNA2, NTRK2, and SFRP1. We obtained a LINC00466–hsa-mir-204–NTRK2 axis that correlates with the treatment and prognosis of invasive breast cancer through comparative integration. This further demonstrates the contribution of lncRNA, miRNA, and mRNA interactions to the pathogenesis of invasive breast cancer. This study further investigated the molecular pathogenesis of invasive breast cancer and provides new diagnostic and prognostic biomarkers that can be used to guide the treatment and in-depth study of invasive breast cancer.


Acknowledgment
This work was supported by grants from the National Natural Science Foundation of China (81673799) and the National Natural Science Foundation of China Youth Fund (81703915).

Disclosure
The authors report no conflicts of interest in this work.

References
CeRNA networks in invasive breast cancer


