Research article

Value of negatively correlated miR-205-5p/*HMGB3* and miR-96-5p/*FOXO1* on the diagnosis of breast cancer and benign breast diseases

Jiaying Li [a,b,1], Shuang Peng [c,1], Xuan Zou [d], Xiangnan Geng [e], Tongshan Wang [c], Wei Zhu [c,**], Tiansong Xia [a,*]

[a] Department of Breast Surgery, The First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu 210029, China
[b] Department of Anesthesiology, Jinling Hospital, Medical School of Nanjing University, Nanjing, Jiangsu 210002, China
[c] Department of Oncology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu 210029, China
[d] Department of Pancreatic Surgery, Fudan University Shanghai Cancer Center, Shanghai 200032 China
[e] Department of Clinical Engineering, The First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu 210029, China

HIGHLIGHTS

- A diagnostic model of negatively correlated microRNA (miRNA)/mRNA pairs was developed to distinguish breast cancer from benign breast disease.
- Combined biometric research and experiments can significantly improve diagnostic accuracy.
- The immunological characteristics of breast cancer tumors were analyzed using miR-205-5p/High mobility group box 3 (*HMGB3*) and miR-96-5p/Forkhead Box O1 (*FOXO1*).

GRAPHICAL ABSTRACT
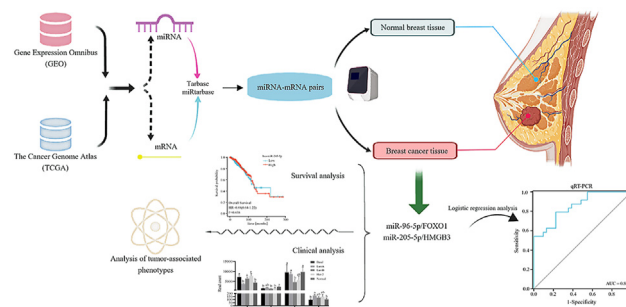
A diagnostic model incorporating negatively correlated microRNA (miRNA)/mRNA pairs was established through biometric studies and experiments, and further tumor-related analyses were performed using this model.

ARTICLE INFO

ABSTRACT

Background: MicroRNA (miRNA) and mRNA levels in matching specimens were used to identify miRNA–mRNA interactions. We aimed to integrate transcriptome, immunophenotype, methylation, mutation, and survival data analyses to examine the profiles of miRNAs and target mRNAs and their associations with breast cancer (BC) diagnosis.

Methods: Based on the Gene Expression Omnibus (GEO) database and The Cancer Genome Atlas (TCGA), differentially expressed miRNAs and targeted mRNAs were screened from experimentally verified miRNA-target interaction databases using Pearson's correlation analysis. We used real-time quantitative reverse transcription polymerase chain reaction to verify BC and benign disease samples, and logistic regression analysis was used to establish a diagnostic model based on miRNAs and target mRNAs. Receiver operating characteristic curve analysis was performed to test the ability to recognize the miRNA-mRNA pairs. Next, we investigated the complex interactions between miRNA-mRNA regulatory pairs and phenotypic hallmarks.

* Corresponding author: Department of Breast Surgery, The First Affiliated Hospital of Nanjing Medical University, 300 Guangzhou Road, Nanjing, Jiangsu 210029, China.
** Corresponding author: Department of Oncology, The First Affiliated Hospital of Nanjing Medical University, 300 Guangzhou Road, Nanjing, Jiangsu 210029, China.
E-mail addresses: zhuwei@njmu.edu.cn (W. Zhu), xiatsswms@163.com (T. Xia).
[1] Jiaying Li and Shuang Peng contributed equally to this work.

Results: We identified 27 and 359 dysregulated miRNAs and mRNAs, respectively, based on the GEO and TCGA databases. Using Pearson's correlation analysis, 10 negative miRNA-mRNA regulatory pairs were identified after screening both databases, and the related miRNA and target mRNA levels were assessed in 40 BC tissues and 40 benign breast disease tissues. Two key regulatory pairs (miR-205-5p/High mobility group box 3 (*HMGB3*) and miR-96-5p/Forkhead Box O1 (*FOXO1*)) were selected to establish the diagnostic model. They also had utility in survival and clinical analyses.

Conclusions: A diagnostic model including two miRNAs and their respective target mRNAs was established to distinguish between BC and benign breast diseases. These markers play essential roles in BC pathogenesis.

Introduction

The most recent cancer report published by the International Agency for Research on Cancer of the World Health Organization indicated that there are as many as 2.26 million incident breast cancer cases globally annually.[1] Data from the Breast Imaging Reporting and Data System (BI-RADS) revealed four types of lesions based on ultrasonographic and mammographic findings. BI-RADS category 4 represents masses with potential abnormalities that require biopsy, of which three are substandard types (4a, 4b, and 4c) with an elevated suspicion.[2] Such lesions do not show the morphological features of breast cancer but have a risk of malignancy over time, ranging from 2 to 95%. This variability may result in unnecessary biopsies and overtreatment of nonmalignant tumors.

Early diagnosis of invasive breast carcinoma (BRCA) is essential for reducing related mortality. Growing evidence suggests that mRNAs and microRNA (miRNA) play important roles in carcinogenesis by regulating cell division and differentiation, apoptosis, epithelial-to-mesenchymal transition, and chemotherapy resistance.[3,4] Bioinformatics can provide novel insights into biomarkers for diagnosis and prognosis based on mRNA expression profiles. miRNAs are abnormally expressed in breast cancer, gastric cancer, leukemia, and other diseases and can control mRNA expression post-transcriptionally via base-pairing with complementary sequences within mRNAs.[5-7] Many studies have emphasized the essential role of miRNA-mRNA axes in breast cancer. For instance, miR-145-5p, which regulates SRY-Box Transcription Factor 2 (*SOX2*), is considered a potential marker for predicting breast cancer stemness,[8] and miR-483-3p, which targets Methyltransferase Like 3 (*METTL3*), is involved in breast cancer treatment.[9]

Furthermore, miR-590-5p directly targeted Yes1 Associated Transcriptional Regulator *(YAP1)* in both *in vitro* and *in vivo* xenograft models, inhibiting tumorigenesis in colorectal cancer cells.[10] A comprehensive meta-analysis of miRNAs for predicting women's cancer by Bastami et al. that analyzed 126 studies from 69 articles, with a total of 48,844 breast and gynecological cancer patients and 68,477 healthy individuals, by comparing the data accuracy of a single miRNA versus multiple miRNAs, demonstrated that a combination of multiple miRNAs appeared to be more favorable than a single miRNA.[11] The same study also revealed that it is adequate to select suitable candidate miRNA-mRNA pairs as a combination of new molecular biomarkers after screening public databases and performing molecular verification in BRCA.

In this study, through a comprehensive assessment of miRNA-mRNA expression profiles in BRCA and benign breast diseases, we determined the miRNA-mRNA regulatory network and its complex role in the pathogenesis of BRCA. To identify essential differentially expressed mRNAs and miRNAs in breast cancer, mRNA and miRNA microarray datasets and RNA-sequencing data were separately obtained from the Gene Expression Omnibus (GEO) database and The Cancer Genome Atlas (TCGA). We employed TarBase and MiRTarBase for target mRNA screening to identify critical miRNA-mRNA regulatory pairs and summarized experimentally confirmed miRNA-mRNA pairs. miRNAs and target mRNAs axes were assessed in formalin-fixed paraffin-embedded (FFPE) specimens using reverse transcription and RT-qPCR. Pearson's correlation validation was performed to establish a logistic regression model to assess diagnostic significance. This study conducted an extensive analysis of miRNA-mRNA regulation in BRCA and benign breast disease tissues to further explore the mechanisms of breast cancer. The combination of bioinformatics and RT-qPCR clarified miRNA-mRNA dysregulation and further determined the early diagnostic mechanism of breast cancer.

Methods

Acquisition and processing of miRNA and mRNA expression profiles

Figure 1 illustrates the workflow of the current study. RNA-sequencing profiles and associated clinical information from TCGA Breast Invasive Carcinoma (TCGA-BRCA) database were obtained from the Global Data Consortium (GDC) data portal of the National Cancer Institute (https://portal.gdc.cancer.gov/). We then searched for breast cancer-related gene microarray and high-throughput sequencing expression datasets in the GEO database (http://www.ncbi.nlm.nih.gov/geo/) using the keyword "breast cancer". Filters were set to "series" and "expression profiling by array", "expression profiling by high-throughput sequencing", "noncoding RNA profiling by array", "noncoding RNA profiling by high-throughput sequencing", and "*Homo sapiens*". RNA-Seq data were analyzed using the "edgeR" R package. The network analysis tool GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r/) was used to conduct the Wilcoxon rank-sum test between normal and cancerous specimens based on GEO dataset queries and limma R packages in the GEO database. The fold-change (FC) of differentially expressed mRNAs (DE-mRNAs) and miRNAs (DE-miRNAs) was calculated. The cut-off criteria were $P < 0.05$ and $|log_2FC| \geq 1$.

Predicting the target mRNAs of miRNAs in BRCA

To screen the target mRNAs of the miRNAs, we used TarBase and miRTarBase, which contain experimentally verified miRNA-mRNA interactions. Pearson's correlation analysis based on TCGA-BRCA RNA-sequencing data was performed to verify the reverse regulatory relationship between miRNAs and potential target mRNAs in BRCA. $P < 0.05$ and $r < 0$ were considered as cut-off criteria.

Functional analysis for DE-miRNAs and DE-mRNAs in BRCA

Gene ontology (GO) functional and Kyoto Genome Encyclopedia (KEGG) pathway analyses of the DE-mRNAs in the network were performed using HiPlot (https://hiplot.com.cn). DAVID-mirPath (a miRNA path analysis network server)[12] was used for the analysis, considering the cluster profile tool in HiPlot. $P < 0.05$ and counts >2 indicated statistically significant differences in enriched GO/KEGG terms.

Sample collection and RNA isolation

We obtained 40 FFPE breast cancer tissue samples, and benign breast tumor samples were obtained from patients surgically treated at the First Affiliated Hospital of Nanjing Medical University. Table 1 shows the clinical characteristics of the 40 patients with BC. An RNAprep Pure FFPE Kit (TIANGEN Biotech, Beijing, China) was used for total RNA extraction from the FFPE tissues, according to the manufacturer's instructions. The RNA concentration was assessed using a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).
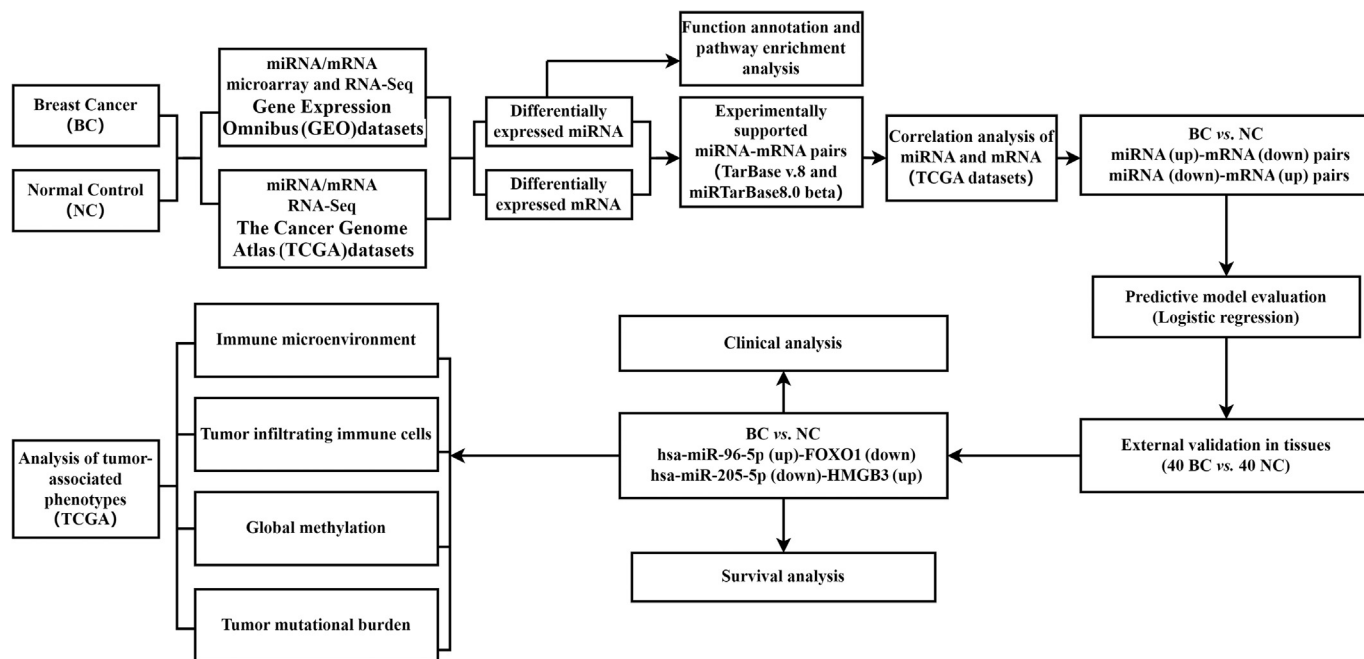
Figure 1. Flow chart for identifying microRNA (miRNAs)-mRNA regulatory pairs and comprehensive analysis of regulatory pair roles in breast cancer (BC).

Quantitative reverse transcription polymerase chain reaction

After adding poly(A) tails to all RNA samples, a poly(A) polymerase Kit (Takara Bio USA, Inc., San Jose, CA, USA), PrimeScript RT Kit, and SYBR Premix Ex Taq II (Takara Bio USA Inc.) were used for quantitative reverse transcription polymerase chain reaction (RT-qPCR) to verify the selected DE-mRNAs and DE-miRNAs, as directed by the manufacturer. PCR amplification was performed on a qTOWER^3 84 Real-Time Thermal Cycler (Analytik Jena, Jena, Germany) in 10-μL reactions at 95 °C (20 s), followed by 40 cycles at 95 °C (10 s) and 60 °C (20 s). Supplementary Table 1 shows the sequences of the PCR primers used. RUN6B (U6) and 18S rRNA served as reference genes, and Livak et al.'s comparative cycle threshold ($2^{-\triangle\triangle Ct}$) method was used to analyze miRNA and mRNA expression data.[13]

Survival analysis

Due to the lack of survival data in GEO datasets, the overall survival (OS) provided by the TCGA-BRCA database was applied in the survival analysis. Kaplan–Meier curves were drawn using the Kaplan–Meier plotter to determine the impact of miRNA-mRNA axes on BRCA prognosis (http://kmplot.com/analysis/index.php?p=service&cancer=breast).[14] This online tool collects transcriptomic datasets, follow-up data, and clinical information. Factors affecting relapse-free survival were assessed using Cox proportional hazards regression analysis. Kaplan–Meier plots were generated to visualize the prognostic value of the candidate miRNAs and mRNAs. The sample size used to calculate OS was 1880; the best-performing threshold was selected as the cut-off, and the maximum follow-up time was 240 months. Hazard ratios (HR) and 95% confidence intervals (CI) were determined. $P < 0.05$ indicated statistical significance.

Assessment of miRNA-mRNA pair and tumor-related phenotype interactions

The clinical information of the patients collected from the TCGA-BRCA database was acquired from the GDC data portal of the National Cancer Institute. According to Thorsson, TCGA-BRCA samples were classified into five groups using the Pam50 method: luminal A, luminal B, Her-2 positive, basal-like, and normal-like.[15] The fractions of 22 infiltrating immune cell types were calculated using CIBERSORT, a gene-based deconvolution algorithm (https://cibersort.stanford.edu/index.php/). Differences in immune cells between TCGA-BRCA patients and normal controls were compared using the Wilcoxon rank-sum test. The UCSC Xena platform (https://xena.ucsc.edu/) was searched using the Illumina Infinium Human-Methylation450 BeadChip platform (Illumina, San Diego, CA, USA). The methylation levels of CpG sites in TCGA-BRCA samples were obtained using this platform, and these data were used to calculate the overall DNA methylation level. Tumor mutational burden was used to measure the total number of somatic variants and megabases.

Statistical analysis

R v3.6.3 (https://cran.r-project.org/; R Foundation for Statistical Computing, Vienna, Austria), GraphPad Prism (GraphPad Software, San Diego, CA, USA), and SPSS v.26 (SPSS Inc. Chicago, IL, USA) were used to analyze the data. Continuous data are expressed as mean ± SD and were compared using the Student's *t*-test. Differentially expressed miRNAs and mRNAs with $|log^2FC| > 0.58$ and $P < 0.05$ were considered statistically significant. The association between miRNA and mRNA expression in

Table 1
Clinicopathological and molecular features of breast cancer patients.

Parameters	Breast cancer (*n*=40)	Rate (%)
Age (year)		
Mean (standard deviation)	50.8 (12.1)	
Median (min, max)	49 (31,75)	
Grade, *n*		
I	4	10
II	18	45
III	18	45
TNM stage *in situ, n*	1	2.5
I	17	42.5
II	12	30
III	10	25
Epithelial subtype, *n*		
Luminal	17	42.5
HER2-enriched	8	20
Triple-negative	4	10
In situ	1	2.5

HER2: Human epidermal growth factor receptor 2; TNM: Tumor-node-metastasis.

BRCA tissues was determined using Pearson's correlation analysis. Receiver operating characteristic (ROC) curves were established based on the indicated miRNA-mRNA pairs, and the areas under the curves (AUC) were assessed to evaluate the diagnostic accuracy of these indicators. Kaplan–Meier curves were compared using the log-rank test, with $P < 0.05$ indicating statistical significance. The time interval from diagnosis to death was defined as the OS. Cox regression analysis was conducted for statistically significant DE-miRNAs or DE-mRNAs ($P < 0.05$). HR and 95% CIs were assessed to determine whether these indicators are survival related. R v3.6.3, GraphPad Prism, and HiPlot generated the plots.

Results

Identification of DE-miRNAs and DE-mRNAs in BRCA

This study included 21 miRNA expression datasets from GEO, including four RNA-Seq datasets (3 from tissue, 1 from peripheral blood) and 17 microarray datasets from tissue, peripheral blood, serum, and plasma. Five mRNA expression microarray datasets were obtained from tissue samples. In addition, three mRNA expression RNA-Seq datasets, including two from tissues and one from peripheral blood, were included. Table 2 presents detailed information on the GEO datasets used in this study. A total of 27 DE-miRNAs and 359 DE-mRNAs were at the intersection of TCGA and GEO datasets.

Functional enrichment and pathway analyses

Using DAVID-mirPath, the enrichment analysis of five upregulated and five downregulated miRNAs was performed. The targets of the five upregulated miRNAs (hsa-miR-21-5p, hsa-miR-182-5p, hsa-miR-96-5p, hsa-miR-200c-3p, and hsa-miR-342-3p) were enriched in 70 KEGG pathways, 208 GO biological processes, 11 GO cellular components, and 20 GO molecular functions, and the top 15 annotations ordered by -$\log_{10}p$-value are shown on the left of Supplementary Figure 1. The targets of the five downregulated miRNAs (hsa-miR-195-5p, hsa-miR-139-5p, hsa-miR-205-5p, hsa-miR-125b-5p, and hsa-miR-145-5p) were enriched in 44 KEGG pathways, 208 GO biological processes, 20 GO cellular components, and 20 GO molecular functions, and the top 15 annotations ordered by -$\log_{10}p$-value are shown on the right of Supplementary Figure 1. Next, we analyzed the 359 differentially expressed mRNAs using the Hiplot platform with a cluster profile, as shown in Supplementary Figure 2. The results disclosed that the enriched KEGG pathways of dysregulated miRNAs were frequently associated with signal transduction pathways such as the *p53, FoxO, TGF-beta, Hippo,* and *ErbB* pathways. Tumorigenesis-related events such as proteoglycans in cancer, fatty acid metabolism, fatty acid biosynthesis, and pancreatic cancer were also involved.

Negatively regulated miRNA/mRNA pairs related to BRCA

First, miRTarBase and TarBase were used to select experimentally verified target mRNAs associated with DE-miRNAs. As shown in Figure 2B, negatively regulated miRNA-mRNA pairs were identified based on the intersections between the 359 DE-mRNAs and the two databases. Next, the Benjamini–Hochberg method was used to filter out 10 miRNA-mRNA pairs with significant negative correlations (adjusted $P < 0.05$) in TCGA. Upregulated miRNA/downregulated mRNA pairs included miR-21-5p/Phosphoinositide-3-Kinase Regulatory Subunit 1 (*PIK3R1*), miR-182-5p/Arrestin domain-containing 3 (*ARRDC3*), miR-96-5p/Forkhead Box O1 (*FOXO1*), miR-200c-3p/Fibulin 5 (*FBLN5*), and miR-342-3p/Transforming Acidic Coiled-Coil Containing Protein 1 (*TACC1*), and downregulated miRNA/upregulated mRNA pairs included miR-195-5p/Rac GTPase-

Table 2
Information on the selected GEO datasets for breast cancer.

| Parameters	Experiment type	Source name	GEO accession	Platform	Tumor group, n			Control group, n
					Total	Stage (0,I,II,III,IV)	Classification (Luminal A, Luminal B, Her2, Triple negative)	
microRNA expression	Array	Tissue	GSE144463	GPL15468	40	0, 13, 15, 12, 0	Not given	10
			GSE58606	GPL18838	122	Not given	31, 33, 27, 31	11
			GSE38167	GPL14943	44	Not given	0, 0, 0, 44	23
			GSE48088	GPL14613	33	Not given	8, 10, 11, 4	3
			GSE44124	GPL14767	50	Not given	Not given	3
			GSE32922	GPL7723	22	Not given	Not given	15
			GSE45666	GPL14767	101	Not given	Not given	15
			GSE42072	GPL16249	7	Not given	3, 0, 0, 4	7
			GSE26659	GPL8227	77	2, 20, 47, 5, 3	Not given	17
			GSE7842	GPL5173	93	0, 2, 66, 22, 9	Not given	5
		Serum	GSE98181	GPL21572	24	Not given	Not given	24
		Plasma	GSE118782	GPL8786	30	Not given	Not given	10
			GSE41526	GPL8179	40	Not given	Not given	20
			GSE22981	GPL8179	20	Early stage	Not given	20
		Peripheral blood	GSE83270	GPL22003	6	Not given	Not given	6
			GSE53179	GPL16550	11	Early stage	11, 0, 0, 0	5
			GSE31309	GPL14132	48	0, 41, 7, 0, 0	Not given	57
	Sequencing	Tissue	GSE131599	GPL18573	189	Not given	40,55,44,50	2
			GSE117452	GPL16791	58	Not given	Not given	10
			GSE68085	GPL10999	103	Not given	62,2,10,29	11
		Peripheral blood	GSE72080	GPL11154	14	Not given	Not given	18
Gene expression	Array	Tissue	GSE50428	GPL13648	21	Not given	5,5,5,11	10
			GSE59246	GPL13607	86	Not given	20,33,17,16	19
			GSE71053	GPL570	6	Not given	Not given	12
			GSE115275	GPL21827	6	Not given	0,0,0,6	6
			GSE64790	GPL19612	3	Not given	0,0,0,3	3
		Tissue	GSE52194	GPL11154	17	Not given	Not given	3
	Sequencing		GSE99680	GPL18573	14	Not given	Not given	19
		Peripheral blood	GSE41245	GPL14761	10	Not given	Not given	20

GEO: Gene Expression Omnibus.
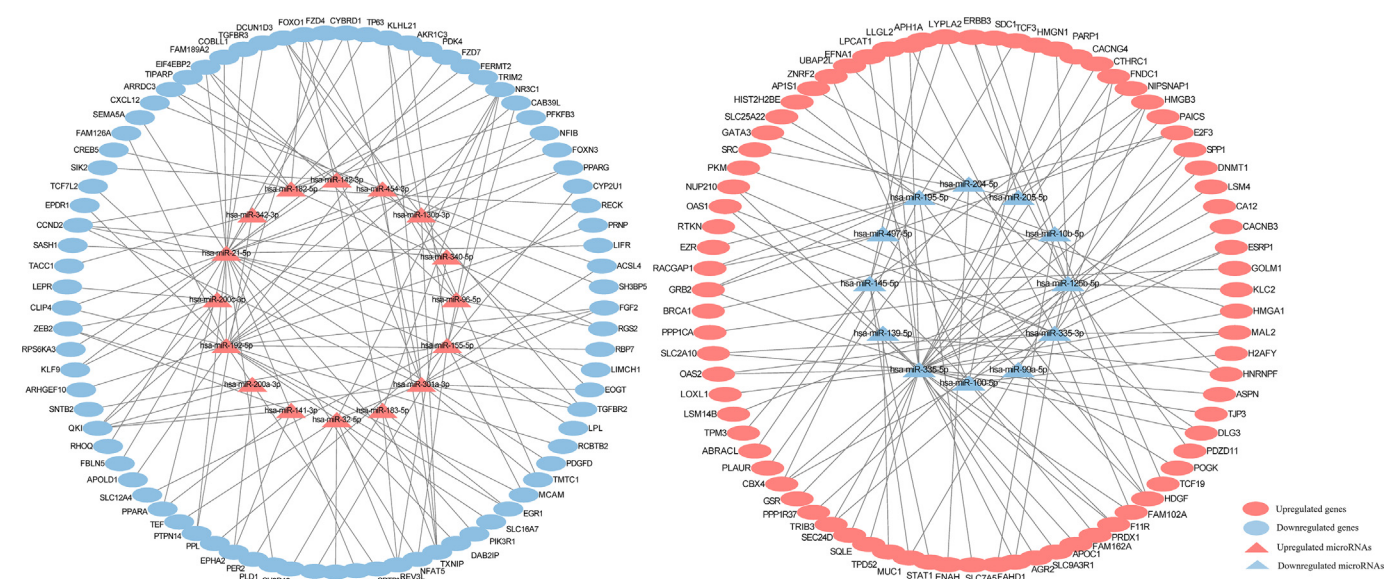
Figure 2. The 27 microRNA (miRNAs) and 359 mRNA network visualized by Cytoscape. There were 177 miRNA (up)-mRNA (down) pairs and 247 miRNA (down)-mRNA (up) pairs screened out by miRtarbase and Tarbase, which contain experimentally validated miRNA-mRNA regulatory pairs. Red represents the upregulated miRNAs/mRNAs in BC versus NCs, while blue represents the downregulated miRNAs/mRNAs in BC versus NCs. Ovals represent mRNAs, and triangles represent miRNAs. BC: Breast cancer; NCs: Negative controls.

activating protein 1 (RACGAP1), miR-139-5p/Tumor protein D52 (*TPD52*), miR-205-5p/High mobility group box 3 (*HMGB3*), miR-125b-5p/ Poly (ADP-ribose) polymerase 1 (*PARP1*), and miR-145-5p/Tropomyosin 3 (*TPM3*) [Table 3]. Through database screening, we identified 10 negatively correlated miRNA/mRNA pairs for experimental verification.

Verification of miRNA and mRNA amounts in the BRCA tissue

Poly(A) RT-qPCR was performed to analyze the expression levels of the 10 DE-miRNAs and 10 DE-mRNAs in 40 tumors and 40 benign breast disease tissue samples. The results showed that among miRNAs, miRNA-205-5p, miRNA-139-5p, miR-145-5p, and miR-96-5p met the standard. The test results for the mRNAs targeted by these four miRNAs showed that TPD52 (*P* = 0.021, *FC* = 1.287), HMGB3 (*P* = 0.015, *FC* = 1.329), PARP1 (*P* = 0.004, *FC* = 2.349), and TPM3 (*P* = 0.029, *FC* = 3.993) met the standards. The complete RT-qPCR data are shown in Figure 3A. Pearson's correlation analysis was performed to examine the interactions between DE-miRNAs and DE-mRNAs. Among the four miRNA/mRNA pairs, miR-205-5p/*HMGB3* (*P* = 0.008, *r* = −0.350) and miR-96-5p/ *FOXO1* (*P* = 0.028, *r* = −0.290) showed a significant negative correlation

Table 3
Pearson's correlation analysis of miRNA-mRNA pairs in breast cancers in TCGA.

Parameters	miRNA	Parameters	mRNA	*p*-value	*r*-value
Up	miR-21-5p	Down	*PIK3R1*	8.02E-09	−0.170
	miR-182-5p		*ARRDC3*	2.70E-08	−0.164
	miR-96-5p		*FOXO1*	2.33E-20	−0.269
	miR-200c-3p		*FBLN5*	3.26E-05	−0.123
	miR-342-3p		*TACC1*	1.02E-05	−0.130
Down	miR-195-5p	Up	*RACGAP1*	1.46E-18	−0.256
	miR-139-5p		*TPD52*	4.58E-15	−0.229
	miR-205-5p		*HMGB3*	1.58E-03	−0.093
	miR-125b-5p		*PARP1*	4.83E-16	−0.237
	miR-145-5p		*TPM3*	8.87E-13	−0.209

ARRDC3: Arrestin domain-containing 3; *FBLN5*: Fibulin 5; *FOXO1*: Forkhead Box O1; *RACGAP1:* Rac GTPase-activating protein 1; *HMGB3*: High mobility group box 3; *PARP1*: Poly(ADP-ribose) polymerase 1; *PIK3R1*: Phosphoinositide-3-Kinase Regulatory Subunit 1; *TACC1:* Transforming Acidic Coiled-Coil Containing Protein 1; TCGA: The Cancer Genome Atlas; *TPD52*: Tumor protein D52; *TPM3*: Tropomyosin 3.

[Figure 3B and Table 4]. Immunohistochemistry images in the HPA database revealed elevated *HMGB3* levels in BRCA cells compared with those in normal breast cells, whereas the expression of *FOXO1* in BRCA cells was lower than that in normal breast cells [Supplementary Figure 3].

Predictive value of miRNA-mRNA regulator pairs in BRCA

Logistic regression analysis was performed to evaluate the predictive value of the miR-205-5p/HMGB3 and miR-96-5p/FOXO1 panels by including two miRNA-mRNA pairs in an RT-qPCR validation cohort containing 40 BRCA tissues. The ROC curve (logistic regression model = −0.2286 + 0.0082*miR-96-5p + −0.8506*FOXO1 + −0.2136*miR-205-5p + 0.2818*HMGB3) analysis [Figure 4A] confirmed that this model had good diagnostic value (AUC = 0.856, 95% CI: 0.759–0.953). ROC curves were generated for the two DE-miRNAs and two DE-mRNAs in TCGA [Figure 4B], further confirming that these four indicators have diagnostic value [Figures 4C and D]. According to the ROC curve of the four combined indicators (AUC = 0.999), their diagnostic advantage was significantly better than that of a single indicator.

We also analyzed the expression levels of the two DE-mRNAs in miRNA-mRNA pairs based on clinicopathological features. As shown in Figures 5A and B, the expression levels of *FOXO1* and miR-205-5p were associated with the age at diagnosis and breast cancer stage. The expression levels of the miR-205-5p/*HMGB3* pair differed between the luminal A and B groups [Figure 5C]. The same difference was observed between the luminal B and normal-like groups.

Overall survival data

Because the clinical tissue samples were limited and the GEO database had no clinical data, survival analysis was conducted based on TCGA data. The HRs of various clinical parameters in the TCGA testing set (*n* = 1880) were estimated using univariate and multivariate Cox regression analyses. As is depicted in Supplementary Figure 4, *HMGB3* level significantly correlated with OS (HR = 1.39, 95% CI: 1.01–1.91; *P* = 0.044) in TCGA. This demonstrates that *HMGB3* influences the prognosis of breast cancer and provides novel insights into breast cancer treatment. Unfortunately, we did not have additional survival data to estimate the prognostic model, and more research is required regarding the predictive value of the two miRNA-mRNA regulatory pairs in BRCA.
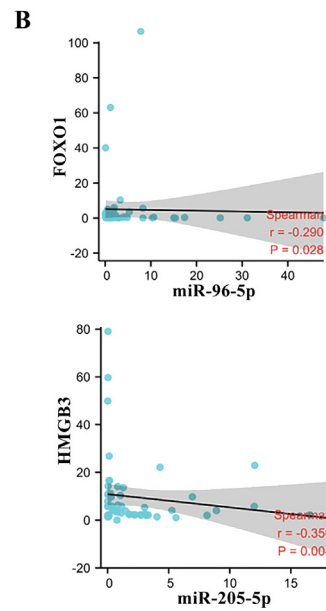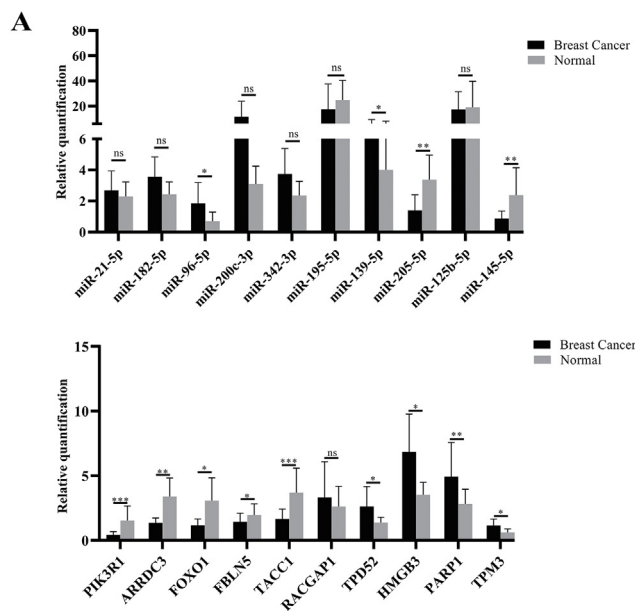
Figure 3. Validation of the expression of 10 differentially expressed microRNA (miRNAs) and 10 differentially expressed mRNAs by RT-qPCR. (A) The miRNA expression levels of miR-96-5p were upregulated in BC, whereas those of miR-139-5p, miR-205-5p, and miR-145-5p were downregulated. The mRNA expression levels of *PIK3R1, ARRDC3, FOXO1, FBLN5,* and *TACC1* were downregulated in BC, whereas those of *TPD52, HMGB3, PARP1,* and *TPM3* were upregulated. (B) Pearson's correlation analysis of miRNA-mRNA regulatory pairs in 80 samples. Two negatively correlated miRNA-mRNA regulatory pairs were plotted. Data are presented as mean ± SEM. *$p < 0.05$, **$p < 0.01$ and ***$p < 0.001$ (Student's *t*-test). *p*-values are listed in Table 4. *ARRDC3*: Arrestin domain-containing 3; BC: Breast cancer; FBLN5: Fibulin 5; *FOXO1*: Forkhead Box O1; *HMGB3*: High mobility group box 3; *PARP1*: Poly (ADP-ribose) polymerase 1; *PIK3R1*: Phosphoinositide-3-Kinase Regulatory Subunit 1; RT-qPCR: Real-time quantitative reverse transcription PCR; SEM: Standard error of the mean; *TACC1:* Transforming Acidic Coiled-Coil Containing Protein 1; *TPD52*: Tumor protein D52; TPM3: Tropomyosin 3.

Table 4
Pearson's correlation analysis of miRNA-mRNA pairs in FFPE breast cancer samples.

Parameters	miRNA	*p*-value	Fold-change (2-ΔΔCT)	Parameters	mRNA	*p*-value	Fold-change (2-ΔΔCT)	Pearson's correlation	
								p-value	*r*-value
Up	miR-21-5p	0.823	1.177	Down	*PIK3R1*	<0.0001	0.131	0.087	0.514
	miR-182-5p	0.133	1.461		*ARRDC3*	0.005	0.198	0.016	0.310
	miR-96-5p	0.025	3.415		*FOXO1*	0.033	0.239	0.028	−0.290
	miR-200c-3p	0.251	1.395		*FBLN5*	0.029	0.450	0.021	0.300
	miR-342-3p	0.197	1.193		*TACC1*	<0.0001	0.228	<0.0001	0.480
Down	miR-195-5p	0.196	0.664	Up	*RACGAP1*	0.204	1.996	0.016	−0.310
	miR-139-5p	0.019	0.261		*TPD52*	0.021	1.287	0.004	0.550
	miR-205-5p	0.009	0.338		*HMGB3*	0.015	1.329	0.008	−0.350
	miR-125b-5p	0.081	0.235		*PARP1*	0.004	2.349	0.095	0.230
	miR-145-5p	0.002	0.195		*TPM3*	0.029	3.993	0.494	0.091

ARRDC3: Arrestin domain-containing 3; *FBLN5*: Fibulin 5; *FOXO1*: Forkhead Box O1; FFPE: Formalin-fixed paraffin-embedded; *RACGAP1:* Rac GTPase-activating protein 1; *HMGB3*: High mobility group box 3; *PARP1*: Poly(ADP-ribose) polymerase 1; *PIK3R1*: Phosphoinositide-3-Kinase Regulatory Subunit 1; RACGAP1: Rac GTPase-activating protein 1; *TACC1:* Transforming Acidic Coiled-Coil Containing Protein 1; *TPD52*: Tumor protein D52; *TPM3*: Tropomyosin 3.

Tumor-related phenotypes associated with signatures

We used a computational method (CIBERSORT) to analyze multiple gene expression profiles in breast cancer to infer the proportions of the 22 immune cell subsets. Twelve immune cell types exhibited differential expression in the BRCA and adjacent normal samples, as shown in Figure 6A [all results are listed in Supplementary Table 2]. We further investigated the association between each cell type and miRNA and target mRNA expression. As shown in Figure 6B, miR-205-5p and its target gene *HMGB3* were associated with activated dendritic cells, whereas miR-96-5p and its target gene *FOXO1* were associated with M0 macrophages. Analysis of the tumor revealed that miR-96-5p/*FOXO1* interacts with DNA methylation, tumor immunity, and inflammation in the tumor microenvironment [Figure 6C].

Discussion

miRNAs are endogenous regulators of gene expression that affect cell proliferation, invasion, and metastasis, all of which are vital for cancer progression.[16,17] miRNA-mRNA interaction networks may act as new biomarkers for early diagnosis and long-term prognosis of breast cancer and may be potential targets for breast cancer treatment.[18] It is of far-reaching significance to identify miRNA-mRNA regulatory networks and clarify their complex roles in immunity, tumorigenesis, and molecular mechanisms.

We conducted an in-depth analysis of miRNA-mRNA pairs in breast cancer tissues and benign controls. We found 28 eligible GEO datasets and used GEO2R, "R-limma", and "R-edgeR" to determine the expression profiles of miRNAs and mRNAs. TCGA data were then combined to
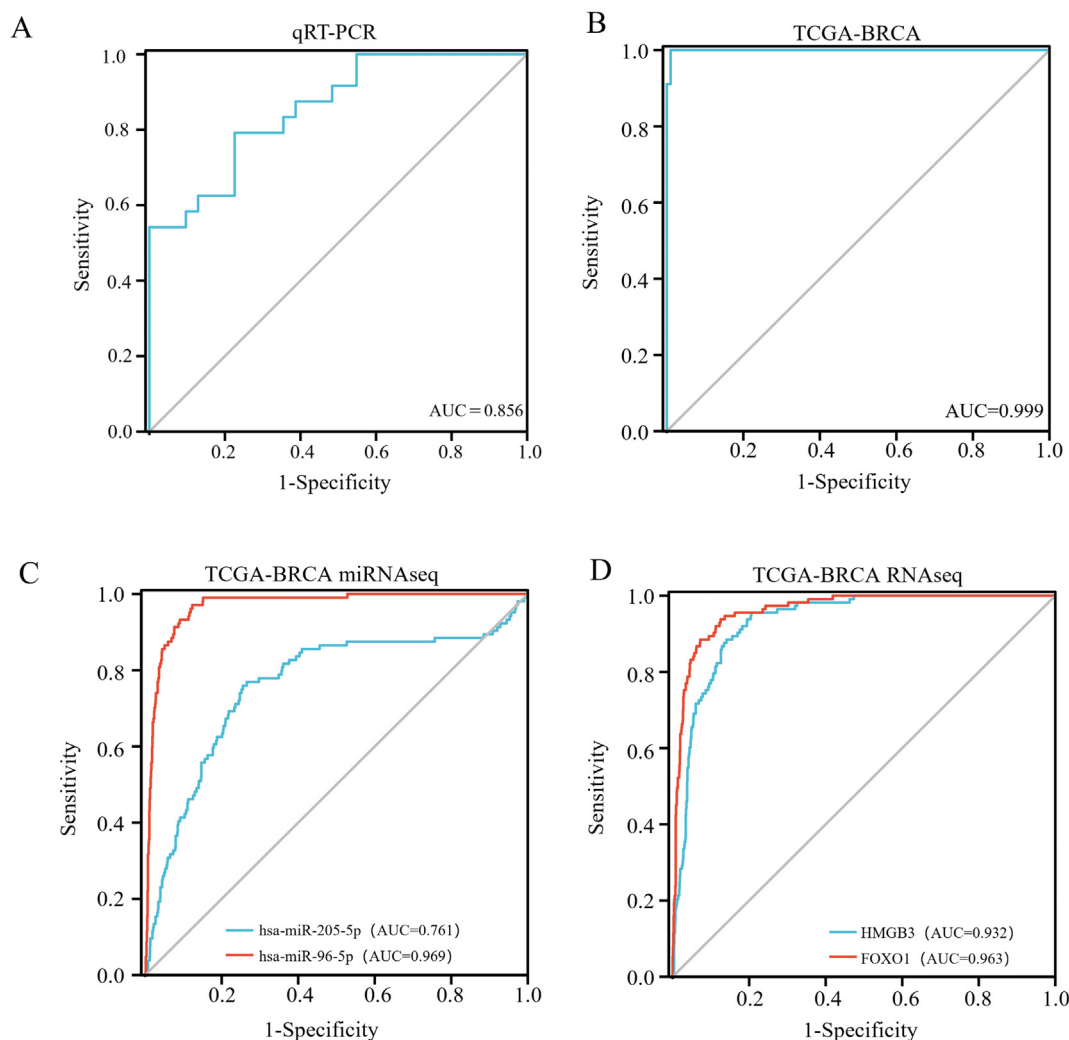
Figure 4. Receiver operating characteristic (ROC) curves of the complex predictive model include four signatures (miR-205-5p, HMGB3, miR-96-5p, and FOXO1) to distinguish BC from control samples. (A) ROC curves of the complex predictive model in the external validation cohort. (B) ROC curves of the complex predictive model in TCGA. (C) ROC curves of each DE-miRNA (miR-205-5p and miR-96-5p) in TCGA. (D) ROC curves of each DE-mRNA (HMGB3 and FOXO1) in TCGA. De-mRNAs: Differentially expressed mRNAs; De-miRNAs: Differentially expressed miRNAs; miRNA: microRNA; FOXO1: Forkhead Box O1; HMGB3: High mobility group box 3; TCGA: The Cancer Genome Atlas.

identify candidate DE-miRNAs and DE-mRNAs. A multistep method identified 27 differentially expressed miRNAs and 359 differentially expressed mRNAs.

Functional analysis of candidate DE-miRNAs and DE-mRNAs selected from TCGA and GEO databases revealed that they were related to the classic p53, FoxO, TGF-beta, and hippo pathways, consistent with previous studies.[19–22] KEGG pathway analysis of proteoglycans in cancer has attracted increasing attention. Proteoglycans affect malignant cells and the tumor microenvironment in both solid and hematopoietic cancers.[23] In the present study, we adopted strict criteria to identify vital miRNA-mRNA regulatory pairs. DE-miRNAs and their target mRNAs showed negatively correlated differential expression, as determined by Pearson's correlation analysis, which was then tested in miRTarbase and Tarbase, and the selected pairs were further examined.

After cross-examining the 27 DE-miRNAs and 359 DE-mRNAs between the 2 databases, 259 miRNA (up)-mRNA (down) pairs and 177 miRNA (down)-mRNA (up) pairs were screened. Pearson's correlation analysis was used to screen the differentially expressed miRNAs and their target mRNAs from experimentally verified miRNA-target interaction databases (miR-TarBase and TarBase). We filtered out 10 significantly negatively correlated miRNA-mRNA pairs. Using poly(A) RT-qPCR, these 10 miRNA-mRNA pairs were examined for expression in 40 BRCA and 40 benign

breast disease specimens. Finally, two negatively correlated regulatory pairs, miR-205-5p/*HMGB3*, and miR-96-5p/*FOXO1* were included in the binary logistic regression model. Subsequent analyses supported the predictive value of these miRNA-mRNA pairs. Compared with single indicators, the combination logistic regression model (y = − 0.2286 + 0.0082 × miR-96-5p − 0.8506 × *FOXO1* − 0.2136 × miR-205-5p + 0.2818 × *HMGB3)* had higher performance, with significance in diagnosis.

MiR-205-5p inhibits breast cancer cell proliferation, migration, and invasion and induces cell apoptosis in association with the long non-coding RNA FGF14-AS2.[24] Silencing miR-205-5p in BRCA decreased tumor growth and metastatic spread in a mouse model.[25] In the present study, miR-205-5p expression in BRCA tissue specimens was significantly downregulated compared with that in non-cancerous breast tissues, which is consistent with the elevated miR-205-5p expression levels that inhibited breast cancer development in the aforementioned studies. For miR-96-5p analysis, many reports have demonstrated that BRCA cell migration is promoted by activating mitogen-activated protein kinase kinase (MEK)/extracellular signal-regulated kinase (ERK) signal transduction,[26] whereas the long noncoding RNA CASC2 inhibits growth and metastasis in BRCA by regulating the miR-96-5p/Synoviolin 1 (SYVN1) pathway.[27] Studies have shown that miRNA-96-5p may negatively regulate the tumor suppressor gene *FOXO3,* promote cell growth, and
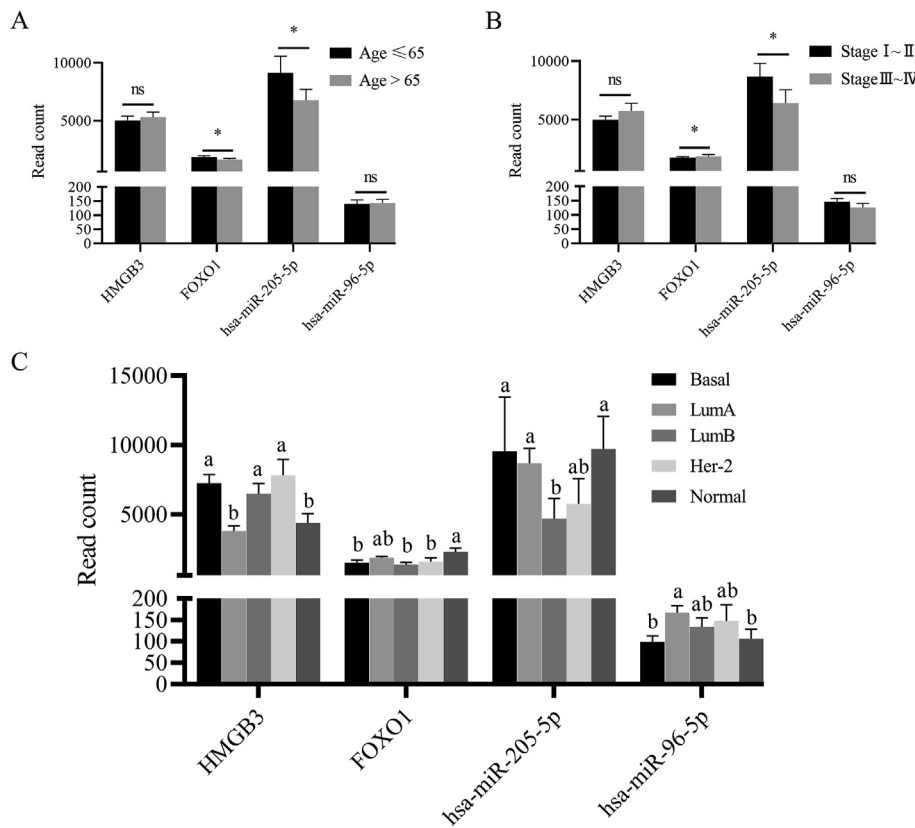
Figure 5. Expression levels of microRNA (miRNAs)-mRNA axes were compared in subgroups divided by clinical features of *BRCA* patients obtained from TCGA. Expression levels of FOXO1 and miR-205-5p were associated with the age at diagnosis and tumor stage of *BRCA* patients. *$p < 0.05$, both a and b indicate statistically significant differences, and those with a or b alone indicate no statistical difference. BRCA: Breast carcinoma; *FOXO1*: Forkhead Box O1; TCGA: The Cancer Genome Atlas.
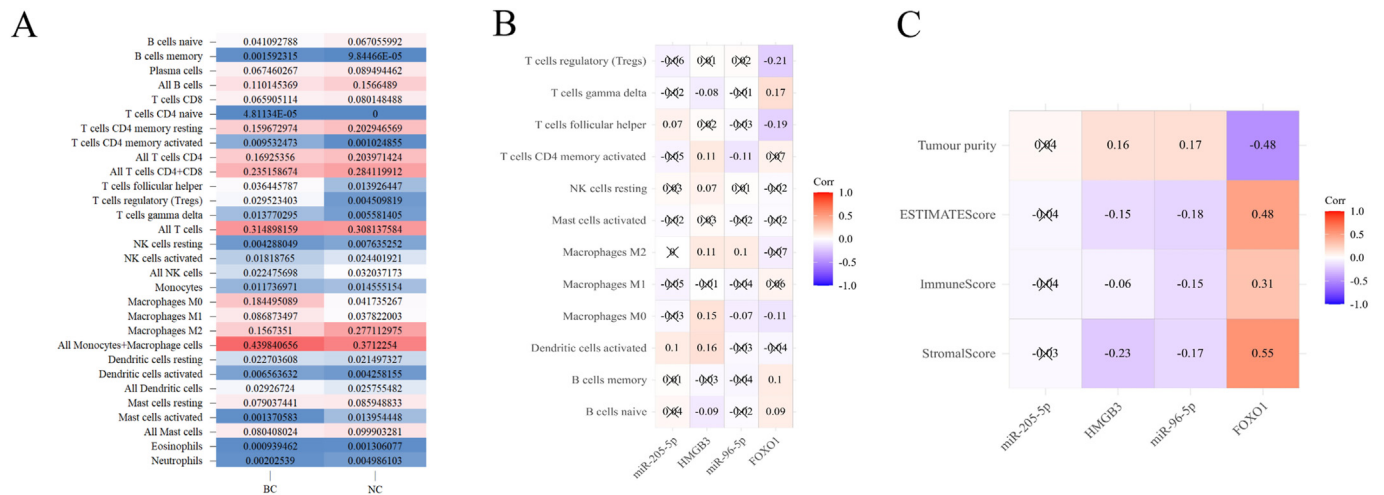

Figure 6. The association between immune-related phenotypes and microRNA (miRNA)/target mRNA expression levels in BC. (A) Heatmap showing the immune cell subset proportions in BC and NCs. Data are presented as the mean. (B) Association between expression of miRNAs/target mRNAs and the proportions of the 22 immune cell subsets by Pearson's correlation. (C) Association between miRNA/target mRNA expression and global methylation, tumor mutation burden, and four tumor microenvironment factors (stromal score, immune score, tumor purity, and ESTIMATE score). "×" was placed through the cell when the *p*-value >0.05. BC: Breast cancer; NCs: Negative controls.

enhance malignancy.[28] In the current study, miR-96-5p targeting of *FOXO1* affected the occurrence of breast cancer, suggesting the need for further research on the relationship between miR-96-5p and the *FOXO* family of proteins.

In OS analysis, *HMGB3* expression significantly correlated with OS (HR = 1.39, 95% CI: 1.01–1.91; P = 0.044) in TCGA. *HMGB3* is absorbed in a breast cancer prognostic model.[29] Although no correlation was detected between OS and miR-96-5p expression levels in tumor tissues, circ_miR-96-5p is involved in evaluating breast cancer prognosis.[30]

To predict breast cancer using miRNA-mRNA pairs, we used a combination of bioinformatics methods and experiments that are more reliable than a single biological information analysis or a single essential investigation. Experimental methods can be used to verify the authenticity of the identified miRNA-mRNA regulatory pairs, including photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP), high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation (HITS-CLIP, also known as CLIP-Seq), crosslinking ligation and sequencing of hybrids (CLASH), biotin microarray, and

western blotting. Bioinformatics has an absolute advantage over large-scale data analyses.[31] The combination of biometric research and experiments can significantly improve diagnostic accuracy. However, with the limited sample size, the current data require validation in large trials. Although the clinical application of these findings remains distant, they provide supplementary or alternative tools for routine diagnostic approaches to *BRCA*.

Funding

Author contributions

Study conception and design: Wei Zhu, Xuan Zou, and Liu Cheng; Data collection: Xuan Zou, Tiansong Xia, and Xiangnan Geng; Data analysis and interpretation: Jiaying Li, Shuang Peng, and Tongshan Wang; Draft manuscript preparation: Jiaying Li and Shuang Peng; Supervision: Wei Zhu. All authors have approved the final version of the manuscript.

Ethics statement

This trial was approved by the Institutional Ethics Committee of the First Affiliated Hospital of Nanjing Medical University (ID: 2016-SRFA-148). Each patient provided signed informed consent for sample collection.

Data availability statement

The datasets analyzed in this study are available from the corresponding author upon reasonable request.

Conflicts of interest

None.

Acknowledgments

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cpt.2023.04.002.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71:209–249. https://doi.org/10.3322/caac.21660.
2. Farras Roca JA, Tardivon A, Thibault F, et al. Correlation of ultrasound, cytological, and histological features of 110 benign BI-RADS categories 4C and 5 nonpalpable breast lesions. The Institut Curie's experience. *Cancer Cytopathol*. 2021;129:479–488. https://doi.org/10.1002/cncy.22402.
3. Garzon R, Calin GA, Croce CM. MicroRNAs in cancer. *Annu Rev Med*. 2009;60: 167–179. https://doi.org/10.1146/annurev.med.59.053006.104707.
4. Zhang X, Gao S, Li Z, et al. Identification and analysis of estrogen receptor alpha promoting tamoxifen resistance-related lncRNAs. *BioMed Res Int*. 2020;2020: 9031723. https://doi.org/10.1155/2020/9031723.
5. Farazi TA, Hoell JI, Morozov P, et al. MicroRNAs in human cancer. *Adv Exp Med Biol*. 2013;774:1–20. https://doi.org/10.1007/978-94-007-5590-1_1.
6. Markopoulos GS, Roupakia E, Tokamani M, et al. A step-by-step microRNA guide to cancer development and metastasis. *Cell Oncol*. 2017;40:303–339. https://doi.org/10.1007/s13402-017-0341-9.
7. Sato-Kuwabara Y, Melo SA, Soares FA, et al. The fusion of two worlds: non-coding RNAs and extracellular vesicles–diagnostic and therapeutic implications (Review). *Int J Oncol*. 2015;46:17–27. https://doi.org/10.3892/ijo.2014.2712.
8. Rajarajan D, Kaur B, Penta D, et al. miR-145-5p as a predictive biomarker for breast cancer stemness by computational clinical investigation. *Comput Biol Med*. 2021;135: 104601. https://doi.org/10.1016/j.compbiomed.2021.104601.
9. Cheng L, Zhang X, Huang YZ, et al. Metformin exhibits antiproliferation activity in breast cancer via miR-483-3p/METTL3/m(6)A/p21 pathway. *Oncogenesis*. 2021;10: 7. https://doi.org/10.1038/s41389-020-00290-y.
10. Ou C, Sun Z, Li X, et al. MiR-590-5p, a density-sensitive microRNA, inhibits tumorigenesis by targeting YAP1 in colorectal cancer. *Cancer Lett*. 2017;399:53–63. https://doi.org/10.1016/j.canlet.2017.04.011.
11. Bastami M, Choupani J, Saadatian Z, et al. Evidences from a systematic review and meta-analysis unveil the role of MiRNA polymorphisms in the predisposition to female neoplasms. *Int J Mol Sci*. 2019;20:5088. https://doi.org/10.3390/ijms20205088.
12. Vlachos IS, Zagganas K, Paraskevopoulou MD, et al. DIANA-miRPath v3.0: deciphering microRNA function with experimental support. *Nucleic Acids Res*. 2015; 42:W460–W466. https://doi.org/10.1093/nar/gkv403.
13. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods*. 2001;25:402–408. https://doi.org/10.1006/meth.2001.1262.
14. Győrffy B. Survival analysis across the entire transcriptome identifies biomarkers with the highest prognostic power in breast cancer. *Comput Struct Biotechnol J*. 2021; 19:4101–4109. https://doi.org/10.1016/j.csbj.2021.07.014.
15. Thorsson V, Gibbs DL, Brown SD, et al. The immune landscape of cancer. *Immunity*. 2018;48:812–830. https://doi.org/10.1016/j.immuni.2018.03.023.
16. He B, Zhao Z, Cai Q, et al. miRNA-based biomarkers, therapies, and resistance in cancer. *Int J Biol Sci*. 2020;16:2628–2647. https://doi.org/10.7150/ijbs.47203.
17. Petri BJ, Klinge CM. Regulation of breast cancer metastasis signaling by miRNAs. *Cancer Metastasis Rev*. 2020;39:837–886. https://doi.org/10.1007/s10555-020-09905-7.
18. Pham VV, Zhang J, Liu L, et al. Identifying miRNA-mRNA regulatory relationships in breast cancer with invariant causal prediction. *BMC Bioinf*. 2019;20:143. https://doi.org/10.1186/s12859-019-2668-x.
19. Mao C, Wang X, Liu Y, et al. A G3BP1-Interacting lncRNA promotes ferroptosis and apoptosis in cancer via nuclear sequestration of p53. *Cancer Res*. 2018;78: 3484–3496. https://doi.org/10.1158/0008-5472.CAN-17-3454.
20. Gong L, Tang H, Luo Z, et al. Tamoxifen induces fatty liver disease in breast cancer through the MAPK8/FoxO pathway. *Clin Transl Med*. 2020;10:137–150. https://doi.org/10.1002/ctm2.5.
21. Xie F, Zhou X, Su P, et al. Breast cancer cell–derived extracellular vesicles promote CD8(+) T cell exhaustion via TGF-beta type II receptor signaling. *Nat Commun*. 2022; 13:4461. https://doi.org/10.1038/s41467-022-31250-2.
22. Ma S, Wu Z, Yang F, et al. Hippo signalling maintains ER expression and ER(+) breast cancer growth. *Nature*. 2021;591:E1–E10. https://doi.org/10.1038/s41586-020-03131-5.
23. Espinoza–Sanchez NA, Gotte M. Role of cell surface proteoglycans in cancer immunotherapy. *Semin Cancer Biol*. 2020;62:48–67. https://doi.org/10.1016/j.semcancer.2019.07.012.
24. Yang Y, Xun N, Wu JG. Long non-coding RNA FGF14-AS2 represses proliferation, migration, invasion, and induces apoptosis in breast cancer by sponging miR-205-5p. *Eur Rev Med Pharmacol Sci*. 2019;23:6971–6982. https://doi.org/10.26355/eurrev_201908_18737.
25. De Cola A, Lamolinara A, Lanuti P, et al. MiR-205-5p inhibition by locked nucleic acids impairs metastatic potential of breast cancer cells. *Cell Death Dis*. 2018;9:821. https://doi.org/10.1038/s41419-018-0854-9.
26. Qin WY, Feng SC, Sun YQ, Jiang GQ. MiR-96-5p promotes breast cancer migration by activating MEK/ERK signaling. *J Gene Med*. 2020;22:e3188. https://doi.org/10.1002/jgm.3188.
27. Gao Z, Wang H, Li H, et al. Long non-coding RNA CASC2 inhibits breast cancer cell growth and metastasis through the regulation of the miR-96-5p/SYVN1 pathway. *Int J Oncol*. 2018;53:2081–2090. https://doi.org/10.3892/ijo.2018.4522.
28. Yin Z, Wang W, Qu G, et al. MiRNA-96-5p impacts the progression of breast cancer through targeting FOXO3. *Thorac Cancer*. 2020;11:956–963. https://doi.org/10.1111/1759-7714.13348.
29. Yao Y, Kong X, Liu R, et al. Development of a novel immune-related gene prognostic index for breast cancer. *Front Immunol*. 2022;26:845093. https://doi.org/10.3389/fimmu.2022.845093.
30. Pourteimoor V, Paryan M, Mohammadi-Yeganeh S, et al. microRNA as a systemic intervention in the specific breast cancer subtypes with C-MYC impacts; introducing subtype-based appraisal tool. *J Cell Physiol*. 2018;233:5655–5669. https://doi.org/10.1002/jcp.26399.
31. Shao YC, Nie XC, Song GQ, et al. Prognostic value of DKK2 from the Dickkopf family in human breast cancer. *Int J Oncol*. 2018;53:2555–2565. https://doi.org/10.3892/ijo.2018.4588.