Comprehensive analysis of negatively correlated miRNA-mRNA regulatory pairs associated with breast cancer diagnosis

Tiansong Xia (xiatsswms@163.com)
The First Affifiliated Hospital with Nanjing Medical University

Jiaying Li
The First Affifiliated Hospital with Nanjing Medical University

Xingchen Fan
First Affiliated Hospital of Nanjing Medical University

Xuan Zou
Fudan University Shanghai Cancer Center

Cheng Liu
First Affiliated Hospital of Nanjing Medical University

Tongshan Wang
First Affiliated Hospital of Nanjing Medical University

Shuang Peng
First Affiliated Hospital of Nanjing Medical University

Wei Zhu
First Affiliated Hospital of Nanjing Medical University

Research Article

Keywords: miRNA, miRNA-mRNA regulation pairs, breast cancer

Posted Date: April 13th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1512841/v1

Abstract

Purpose MicroRNAs (miRNAs) represent small noncoding RNAs controlling mRNA expression. MiRNA and mRNA amounts in matching specimens were utilized for identifying miRNA–mRNA interactions. Transcriptome, immunophenotype, methylation, mutation, and survival data analyses were integrated to examine the profiles of miRNAs and target mRNAs and their associations with breast cancer (BC) diagnosis.

Methods: Based on the mRNA/miRNA expression microarray and RNA-Seq data sets in the GEO database and RNA-Seq data in TCGA, we analyzed differential expression profiles through GEO2R and the R language. Differentially expressed miRNAs and targeted mRNAs were screened from experimentally verified miRNA-target interactions databases by Pearson's correlation analysis. qRT-PCR was used for verification in breast cancer and benign disease samples, and logistic regression analysis was utilized to establish a diagnostic model based on miRNAs and target mRNAs. ROC curve analysis was performed to test the ability to recognize miRNA-mRNA pairs. Then, complex interactions of miRNA-mRNA regulatory pairs and phenotypic hallmarks, including tumor-infiltrating lymphocytes, tumor microenvironment, tumor mutation burden and global methylation, were described.

Result: We first identified 27 dysregulated miRNAs and 135 dysregulated mRNAs based on five gene expression and 21 miRNA expression data sets from GEO and RNA-Seq data from TCGA. Utilizing Pearson's correlation analysis, ten negative miRNA-mRNA regulatory pairs were identified after screening in TarBase and miRTarBase databases, and related miRNA and target mRNA amounts were assessed in 40 breast cancer tissue and 40 breast benign disease tissue. Finally, 2 key regulatory pairs (miR-205-5p/HMGB3 and miR-96-5p/FOXO1) were selected to establish a diagnostic model. We next analyzed the functions of these four markers in breast cancer.

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

1. Introduction

The most recent cancer report published by the International Agency for Research on Cancer (IARC) of the World Health Organization indicated that there are as many as 2.26 million incident breast cancer (BS) cases globally each year(1). Data related to the breast imaging report and data system (BI-RADS) reveal 4 types of lesions based on ultrasonographic and mammographic findings. BI-RADS category 4 indicates masses with potential abnormalities that require biopsy, of which three are substandard types (4a, 4b, and 4c) with elevated suspicion. By definition, such lesions do not show BC’s morphological features, 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 BC is an essential factor in the reduction of related mortality. Growing evidence suggests mRNAs and miRNAs play important roles in carcinogenesis, regulating cell division and differentiation, apoptosis, epithelial-to-mesenchymal transition and chemotherapy resistance(2, 3). Bioinformatics could provide novel insights into biomarkers for diagnosis and prognosis based on gene expression profiles. MiRNAs are abnormally expressed in breast cancer, gastric cancer, leukemia and other diseases, and can control gene expression post-transcriptionally via base-pairing with complementary sequences within mRNAs(4-6). In BC research, some differentially expressed miRNAs can target various molecularly regulated mRNAs. For instance, miR-145-5p regulating SOX2 is considered a potential marker for predicting BC stemness(7); in addition, miR-483-3p targeting METTL3 participates in BC treatment(8).

Furthermore, miR-590-5p was found directly targets YAP1 and inhibits tumorigenesis in CRC cells both in vitro and in vivo xenograft model (9). A comprehensive meta-analysis of microRNAs for predicting women's cancer by Milad Bastami et al. demonstrated multiple miRNAs appear to be more favorable than single miRNAs via analysis of 126 studies from 69 articles with a total of 48844 BC and GC (gynecological cancer) patients and 68477 healthy individuals(10). The 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 BC.

Through comprehensive assessment of miRNA-mRNA expression profiles in BC and benign breast diseases, we determined a miRNA-mRNA regulatory network and its complex role in the pathogenesis of BC. *Figure 1* shows the workflow. For identifying essential differentially expressed mRNAs and miRNAs in breast cancer, mRNA and miRNA microarray datasets were obtained from The Gene Expression Omnibus (GEO) database and The Cancer Genome Atlas (TCGA). We successively carried out differential expression analysis, and TarBase and miRTarBase target gene screening to identify critical miRNA-mRNA regulatory pairs and summarized the experimentally confirmed miRNA-mRNA pairs. Then, miRNA and target mRNA amounts were assessed in formalin-fixed paraffin-embedded (FFPE) specimens by real-time quantitative reverse transcription polymerase chain reaction (qRT-PCR). Pearson’s correlation validation was performed to establish a logistic regression model to assess diagnostic significance. This research conducted an extensive analysis of miRNA-mRNA regulation in BC and breast benign disease tissues to further explore the mechanisms of breast cancer. The combination of bioinformatics and qRT-PCR makes miRNA-mRNA dysregulation clearer, further determining the early diagnostic mechanism of breast cancer.

2. Materials And Methods

2.1 Acquisition and processing of miRNA and mRNA expression profiles

The Cancer Genome Atlas (TCGA) Breast Cancer (BC) miRNA and mRNA sequencing profiles and associated clinical and pathological data were obtained from the GDC data portal of the National Cancer Institute (https://portal.gdc.cancer.gov/). These data from TCGA were processed using Perl. A total of

1077 tumor tissue samples and 99 standard samples from TCGA were included in this study. Then, we searched breast cancer-related gene microarray expression datasets and high-throughput sequencing expression datasets in the Gene Expression Omnibus (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", "Non-coding RNA profiling by array", "Non-coding RNA profiling by high-throughput sequencing" and "Homo sapiens". RNA-Seq data were analyzed with the "edgeR" R package. The network analysis tool GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r/) was utilized to compare sample groups through the GEO DataSets queries and limma R packages in the GEO database. Differentially expressed mRNAs (DE-mRNAs) and differentially expressed miRNAs (DE-miRNAs) were obtained from the microarray expression profile using GEO2R.

2.2 Functional assessment of miRNA-mRNA pairs

Firstly, to screen the target DE-mRNAs corresponding to potential DE-miRNAs, we used TarBase and miRTarBase, which contain experimentally verified miRNA-mRNA pairs. Pearson correlation analysis was carried out to evaluate expression correlations for the negatively correlated miRNA-mRNA pairs identified from the abovementioned database. Cutoff criteria were $p<0.05$ and $|log^2FC|\geq 1$. The Cytoscape software was utilized to visualize the miRNA-mRNA negative regulatory network (v3.8.0)(11). Interaction relationships between the identified miRNAs and mRNAs were used for uploading to Cytoscape.

Next,, gene ontology (GO) function and Kyoto Genome Encyclopedia (KEGG) pathway analyses of DE-miRNAs and DE-mRNAs in the network were performed. DIANA TOOLS - mirPath v.3 (a miRNA path analysis network server)(12) and XianTao (a data retrieval website) (https://www.xiantao.love)were utilized for analysis, considering the cluster profile tool in Hiplot(https://hiplot.com.cn). $P<0.05$ indicated statistically significant differences in enriched GO/KEGG terms.

2.3 Survival analysis

Overall survival (OS) rates based on DE-miRNAs and DE-mRNAs were analyzed. Used Kaplan-Meier Plotter (http://kmplot.com/analysis/index.php?p=service&cancer=breast) (13) to make Kaplan-Meier curve analysis to utilized further determine the impact on BC prognosis. The website collected transcriptomic datasets and included follow-up and patient data. Factors affecting relapse-free survival were assessed by Cox proportional hazards regression analysis. False discovery rates were determined for multiple hypothesis tests. Kaplan-Meier plots were generated for visualizing best-performing genes. The sample size used to calculate OS was 1880; the best performing threshold was chosen as a cutoff, and the follow up time reached 240 months. Hazard ratio (HR) and 95% confidence interval (CI) were determined. $P<0.05$ indicated a significant cutoff.

2.4 Assessment of interactions of miRNA-mRNA pairs and tumor-related phenotypes

The fractions of 22 infiltrating immune cell types were calculated with CIBERSORT, a gene-based deconvolution algorithm (https://cibersort.stanford.edu/index.php/). Differences in these immune cells between TCGA-BCs and normal controls were compared via the Wilcoxon rank-sum test. The UCSC Xena platform (https://xena.ucsc.edu/) was searched by the Illumina Infinium HumanMethylation450 BeadChips platform. The methylation level of CpG sites in TCGA-BRCA samples was obtained from this platform, and these data were utilized to calculate the overall DNA methylation level. Tumor mutational burden (TMB) was used to measure the genome's total number of somatic variants/megabase (MB).

2.5 Sample collection and RNA isolation

We obtained 80 formalin-fixed paraffin-embedded (FFPE) BC and benign breast tumor samples of patients surgically treated in the First Affiliated Hospital of Nanjing Medical University. Each patient provided signed informed consent for sample collection. This trial had approval from the institutional ethics committee (ID: 2016-SRFA-148). *Table 1* shows the clinical characteristics of the 40 breast cancer patients. A RNAprep Pure FFPE Kit (TIANGEN) was utilized for total RNA extraction from FFPE tissues as directed by the manufacturer. RNA amounts were assessed on a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific).

Table 1

Clinicopathological and molecular features of breast cancer patients.

	Breast cancer	Rate (%)
	(n=40)	
Age(year)		
mean (SD)	50.8 (12.1)	
median [min, max]	49 (31,75)	
Grade		
	4	10
	18	45
	18	45
TNM stage		
In situ	1	2.5
	17	42.5
	12	30
	10	25
Epithelial subtype		
Luminal	17	42.5
HER2-enriched	8	20
Triple-negative	4	10
In situ	1	2.5

2.6 Quantitative reverse transcription PCR (qRT-PCR)

Upon addition of a poly(A) tail to the RNA with the Poly(A) polymerase kit (Takara), PrimeScript RT Kit and SYBR Premix Ex Taq II (Takara) were utilized for qRT-PCR for verifying the selected DEGs and DEMs, as directed by the manufacturer. PCR amplification was performed on a qTOWER³ 84 (Analytik Jena) in 10-μL reactions at 95°C (20 s) followed by 40 cycles of 95°C (10 s) and 60°C (20 s). Table S1 shows the sequences of the applied PCR primers. RUN6B (U6) and 18S rRNA served as reference genes, and d the

Livak KJ *et al*.'s comparative cycle threshold ($2^{-\triangle\triangle Ct}$) method was performed for analyzing miRNA and mRNA expression(14).

2.7 Statistical analysis

R v3.6.3 (https://cran.r-project.org/), GraphPad Prism and SPSS v.26 (IBM, USA) were utilized to analyze the data. Continuous data were expressed as mean and standard deviation (SD), and compared by the Student's t-test. MiRNAs and mRNAs with $|log^2FC|>0.58$ and $p<0.05$ were considered differentially expressed with statistical significance. The associations of miRNA and mRNA expression in BC tissues were determined by Pearson correlation analysis. ROC curves were established based on the indicated miRNA-mRNA pairs, and areas under the curves (AUCs) were assessed for evaluating the diagnostic accuracy of these indicators. Kaplan-Meier curves were compared by the log-rank test, with $p<0.05$ indicating statistical significance. The time interval from surgery to death was defined as OS. Cox regression analysis was carried out for statistically significant DE-miRNAs or DE-mRNAs ($p<0.05$), and hazard ratios (HRs) and 95% confidence intervals (CIs) were assessed to determine whether these indicators are survival related. R v3.6.3, GraphPad Prism and Hiplot were utilized to draw plots.

3. Results

3.1 DE-miRNAs and DE-mRNAs in BC

This study included 21 miRNA expression data sets, including 4 RNA-Seq data sets with three from tissue (GSE131599, GSE117452 and GSE68085), one from peripheral blood (GSE72080), and 17 microarray data sets from tissue, peripheral blood, serum and plasma. There were three gene expression RNA-Seq datasets, including two from tissue (GSE52194 and GSE99680) and one from peripheral blood (GSE41245), five gene expression microarray datasets from tissue samples, shown in *Figure 2A*. *Table 2* shows the information of GEO datasets in this article. Upregulated and downregulated DEMs/DEGs in BCs vs. controls were identified based on log^2FC (BC *vs*. normal). Totally 27 DE-miRNAs and 135 DE-mRNAs were the intersections of TarBase and miRTarBase datasets shown in *Figure 2B*.

Table 2

Information pertaining to the selected GEO datasets for breast cancer.

	Experiment Type	Source name	GEO Accession	Platform	Group	
					Tumor	Control
microRNA expression	Array	Tissue	GSE144463	GPL15468	40	10
			GSE58606	GPL18838	122	11
			GSE38167	GPL14943	44	23
			GSE48088	GPL14613	33	3
			GSE44124	GPL14767	50	3
			GSE32922	GPL7723	22	15
			GSE45666	GPL14767	93	15
			GSE42072	GPL16249	7	7
			GSE26659	GPL8227	77	17
			GSE7842	GPL5173	45	26
		Serum	GSE98181	GPL21572	24	24
		Plasma	GSE118782	GPL8786	30	12
			GSE41526	GPL8179	40	20
			GSE22981	GPL8179	20	20
		Peripheral blood	GSE83270	GPL22003	6	6
			GSE53179	GPL16550	11	5
			GSE31309	GPL14132	48	57
	Sequencing	Tissue	GSE131599	GPL18573	189	2
			GSE117452	GPL16791	58	10
			GSE68085	GPL10999	103	11
		Peripheral blood	GSE72080	GPL11154	14	18
Gene expression	Array	Tissue	GSE50428	GPL13648	21	10
			GSE59246	GPL13607	86	19
			GSE71053	GPL570	6	12
			GSE115275	GPL21827	6	6
			GSE64790	GPL19612	3	3

Sequencing	Tissue	GSE52194	GPL11154	17	3
		GSE99680	GPL18573	14	19
	Peripheral blood	GSE41245	GPL14761	10	20

3.2 Function enrichment and pathway analyses

For the target miRNAs and mRNAs screened in the database, KEGG signaling pathway analysis and GO functional annotation analysis were performed. According to the analysis, these 27 DE-miRNAs are associated with various solid tumors including breast cancer, such as glioma, prostate cancer, bladder cancer, colorectal cancer, thyroid cancer, pancreatic cancer, lung cancer and melanoma, etc. Mainly involved in cellular nitrogen compound metabolism, biosynthesis, gene expression and cellular protein modification, etc. Mainly enriched in cellular components such as organelles, cell movement components, cytosol, nucleoplasts, and protein complexes. Involved in ion binding, enzyme binding, enzyme activity regulation, protein binding transcription factor activity regulation, nucleic acid binding transcription factor regulation and other molecular functions. *(Shows in Figure S1-S4)*

Among the 135 DE-mRNAs, the up-regulated mRNAs are mainly involved in biological processes such as DNA recombination, regulation of nuclease activity, and regulation of chromosome segregation; they are mainly enriched in nuclear chromosomes, MCM complexes, mitochondrial spindles and other cellular components; significantly enriched in proteins Molecular functions such as C-terminal binding, Cadherin binding, S100 protein binding, and cell adhesion collection. The down-regulated mRNAs are mainly involved in focal adhesion, regulation of actin cytoskeleton, leukocyte transendothelial migration, AMP signaling and other signaling pathways; they are involved in biological processes such as cell-matrix adhesion, cell behavior changes, and actin filament action; they are mainly enriched in Collagen-containing extracellular matrix, cell-matrix junction, cell surface and other cellular components; significantly enriched in molecular functions such as growth factor binding, kinesin binding, cytokine binding, and integrin binding. According to the results of functional enrichment and pathway analysis, the down-regulated DE-mRNAs are closely related to actin filaments. *(Shows in Figure S5)*

3.3 Negative miRNA/mRNA regulatory pairs related to BC

miRTarBase and TarBase were utilized to select experimentally verified target mRNAs associated with differentially expressed miRNAs. Negative miRNA-mRNA pairs were obtained based on intersections between 135 DE-mRNAs and two databases. Then, the Pearson's correlation analysis method was utilized to filter out ten miRNA-mRNA pairs with significant negative correlations (adjusted p<0.05) in TCGA. Upregulated miRNA/downregulated mRNA pairs were miR-21-5p/PIK3R1, miR-182-5p/ARRDC3, miR-96-5p/FOXO1, miR-200c-3p/FBLN5 and miR-342-3p/TACC1, and downregulated miRNA/upregulated mRNA pairs were miR-195-5p/RACGAP1, miR-139-5p/TPD52, miR-205-5p/HMGB3, miR-125b-5p/PARP1 and

miR-145-5p/TPM3. They are listed in *Table 3*. Through the screening of databases, we identified these ten pairs of negatively correlated miRNA/mRNA pairs for experimental verification.

Table 3

Pearson's correlation analysis of miRNA-mRNA pairs in breast cancer.

miRNA(up)	***mRNA(down)***	***p-value***	***r-value***
21-5p	PIK3R1	8.02E-09	-0.16956
182-5p	ARRDC3	2.70E-08	-0.163531
96-5p	FOXO1	2.33E-20	-0.268737
200c-3p	FBLN5	3.25947E-05	-0.122556
342-3p	TACC1	1.02468E-05	-0.130085
miRNA(down)	***mRNA(up)***	***p-value***	***r-value***
195-5p	RACGAP1	1.46E-18	-0.256018
139-5p	TPD52	4.58E-15	-0.228997
205-5p	HMGB3	0.001584085	-0.093331
125b-5p	PARP1	4.83E-16	-0.236886
145-5p	TPM3	8.87E-13	-0.209282

3.4 Verification of miRNA and mRNA amounts in the BC tissue

To investigate whether the 10 DE-miRNAs and 10 DE-mRNAs are differentially expressed between BC and benign breast disease tissues, ploy(A) qRT-PCR was performed to analyze their expression in 40 tumor 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 mRNAs targeted by these 4 miRNAs showed TPD52 (p=0.021, *FC*=1.287), HMGB3 (p=0.015, *FC*=1.329), TPM3 (p=0.029, *FC*=3.993) and FOXO1(p=0.033, *FC*=0.239)also met the standard. The complete qRT-PCR data are shown in *Figure 3A*. Next, Pearson correlation analysis was performed to examine the interactions between DE-miRNAs and DE-mRNAs. Among the 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 (*Figure 3B* and *Table 4*).

Table 4

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

miRNA (up)	*p-value*	*Fold change (2-ΔΔCT)*	*mRNA (down)*	*p-value*	*Fold change (2-ΔΔCT)*	Pearson's correlation	
						p-value	*r-value*
21-5p	0.823	1.177	PIK3R1	< 0.0001	0.131	0.087	0.514
182-5p	0.133	1.461	ARRDC3	0.005	0.198	0.016	0.31
96-5p	**0.025**	**3.415**	**FOXO1**	**0.033**	**0.239**	**0.028**	**-0.29**
200c-3p	0.251	1.395	FBLN5	0.029	0.45	0.021	0.3
342-3p	0.197	1.193	TACC1	< 0.0001	0.228	< 0.0001	0.48
miRNA (down)			*mRNA (up)*			*p-value*	*r-value*
195-5p	0.196	0.664	RACGAP1	0.204	1.996	0.016	-0.31
139-5p	0.019	0.261	TPD52	0.021	1.287	0.004	0.55
205-5p	**0.009**	**0.338**	**HMGB3**	**0.015**	**1.329**	**0.008**	**-0.35**
125b-5p	0.081	0.235	PARP1	0.004	2.349	0.095	0.23
145-5p	0.002	0.195	TPM3	0.029	3.993	0.494	0.091

IHC images in the HPA database demonstrated elevated HMGB3 amounts in BC cells compared with normal breast cells, while the expression of FOXO1 in BC cells was lower than that of normal breast cells (*Figure 4*).

3.5 Predictive value of miRNA-mRNA regulator pairs in BC

Logistic regression analysis was performed for evaluating the predictive value of the miR-205-5p/HMGB3 and miR-96-5p/FOXO1 panel by including 2 miRNA-mRNA pairs in the qRT-PCR validation cohort containing 40 BC tissues. 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 5A*) confirmed this model had a good diagnostic value (AUC=0.856, CI 0.759-0.953). ROC curves were also generated for the 2 DE-miRNAs and 2 DE-mRNAs in TCGA (*Figure 5B*), the predicted value of the independent indicator is lower than the predicted value of the combination(AUC=0.999) (*Figure 5C and Figure 5D*).

We also analyzed the expression levels of the 2 DE-mRNAs in miRNA-mRNA pairs based on cancer stage. As shown in *Figure 6*, we found that the expression levels of the 2 DE-mRNAs in various breast cancer

stages had no significant difference, indicating they can differentiate non-malignant breast tumors from breast cancer of any stage, they all had diagnostic and predictive significance.

3.6 Overall survival data

Since the clinical tissue samples were limited, with the GEO database having 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 by univariable and multivariable cox regression analyses. As depicted in *Figure 7*, HMGB3 level had a significant correlation with OS (HR=1.39, 95% CI 1.01 to 1.91; p=0.044) in TCGA. This demonstrated HMGB3 has a particular influence on the prognosis of breast cancer, and provided novel insights into breast cancer treatment. We did not have extra survival data to estimate the prognostic model, and more research is required about the predictive value of the 2 miRNA-mRNA regulatory pairs in BC.

3.7 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 22 immune cell subsets. There were 12 types of immune cells with differential amounts in BC and control samples, as shown in *Figure 8A* (all results are listed in *Table S2*). We further investigated the associations of each cell type with miRNA/target mRNA expression. As shown in *Figure 8B*, miR-205-5p and its target gene HMGB3 were related to activated dendritic cells; miR-96-5p and its target gene FOXO1 were related to M0 macrophages. In the analysis of the tumor microenvironment, miR-96-5p/FOXO1 could interact with DNA methylation, tumor immunity and inflammation in the tumor microenvironment (*Figure 8C*).

4. Discussion

MiRNAs are endogenous regulators of gene expression and control cell proliferation, cell invasion and tumor metastasis. They are vital to the progression of cancer. Combining miRNAs and mRNAs has a potential clinical value in BC diagnosis, prognosis and treatment. 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 21 eligible GEO datasets and used GEO2R, "R-limma," and "R-edgeR" to determine the expression profiles of miRNAs and mRNAs. Then, TCGA data were combined for identifying candidate DE-miRNAs and DE-mRNAs. A multi-step method was used for identifying 27 miRNAs and 135 mRNAs with differential expression.

Functional analysis of candidate DE-miRNAs and DE-mRNAs selected from the TCGA and GEO databases revealed that they were related to the classic p53, FoxO, TGF-beta and hippo pathways. The KEGG pathway of proteoglycans in cancer attracts growing attention. It was confirmed that proteoglycans affect malignant cells and the tumor microenvironment both in solid legions and hematopoietic cancers(15). In the present study, we adopted strict criteria to identify vital miRNA-mRNA regulatory pairs. Simply put, DE-miRNAs and their targeted mRNAs should show negatively correlated differential expressions, as determined by Pearson correlation analysis, and then tested in miRTarbase and Tarbase; the selected pairs were further examined.

First, after 27 DE-miRNAs and 135 DE-mRNAs between the two databases were cross-examined, 247 miRNA (upregulation)-mRNA (downregulation) pairs and 177 miRNA (downregulation)-mRNA (upregulation) pairs were screened. Then, Pearson's correlation analysis was utilized to screen differentially expressed miRNAs and their target mRNAs from experimentally verified miRNA-target interaction databases (miRTarBase and TarBase). We filtered out ten significantly negatively correlated miRNA-mRNA pairs. Through poly(A) qRT-PCR, these ten miRNA-mRNA pairs were examined for expression in 40 BC and 40 benign breast disease specimens. Finally, we included in the logistic regression model the two negatively correlated regulatory pairs miR-205-5p/HMGB3 and miR-96-5p/FOXO1. Subsequent analysis supported the predictive values 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 a 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(16); silencing miR-205-5p in BC decreased tumor growth and metastatic spread in a mouse model(17). In this study, miR-205-5p amounts in BC tissue specimens were significantly downregulated in comparison with noncancerous breast tissues, which is consistent with elevated miR-205-5p amounts inhibiting the development of breast cancer in studies mentioned above. For miR-96-5p analysis, many reports have demonstrated BC cell migration is promoted by activating MEK/ERK signal transduction(18), while the long noncoding RNA CASC2 inhibits growth and metastasis in BC by regulating the miR-96-5p/SYVN1 pathway (19). Studies have shown that miRNA-96-5p may negatively regulate the tumor suppressor gene FOXO3 and promote cell growth and enhance malignancy(20). In this research, miR-96-5p targeting FOXO1 affected BC occurrence, which is suggestive for further research on the relationship between miR-96-5p and the FOXO family of proteins.

In overall survival (OS) analysis, HMGB3 expression had a significant correlation with OS (HR=1.39, 95% CI 1.01 to 1.91; p=0.044) in TCGA. No previous studies have examined HMGB3 expression in relation to OS, suggesting this could be a new research direction.

For predicting breast cancer using miRNA-mRNA pairs, we used a combination of bioinformatic methods and experiments, which is 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-Microarrays and Western blot. Bioinformatics has an absolute advantage over large data analysis(21). The combination of biometric research and experiments can significantly improve the accuracy of diagnosis. However, with a 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 to routine diagnostic approaches in BC.

Declarations

Funding section

This work was supported by the Graduate Research and Practice innovation Plan of Graduate Education Innovation Project in Jiangsu Province[Grant number: JX10213728].

Ethical approval

The study was conducted in accordance with the guidelines of the Hospital Ethics Committee and approved by the Institutional Review Boards of the First Affiliated Hospital of Nanjing Medical University(ID: 2016-SRFA-148).

This study was conducted in accordance with the Declaration of Helsinki.

Consent for publication

Informed consent was signed by each participant in advance and written informed consent for publication was obtained.

Authors' Contributions

Study conception and design: Jiaying Li and Xingchen Fan;

data collection: Xuan Zou;

data analysis and interpretation: Jiaying Li. Cheng Liu and Tongshan Wang;

draft manuscript preparation: Jiaying Li, Xingchen Fan and Shuang Peng.

All authors have approved the final version of the manuscript.

Disclosure

The author reports no conflicts of interest in this work.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgements section

All authors acknowledge anyone who contributed to the article who does not meet the criteria for authorship including anyone who provided professional writing services or materials. All authors has permission from everyone mentioned in the acknowledgments section.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, 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(3):209-49.
2. Garzon R, Calin GA, Croce CM. MicroRNAs in Cancer. Annu Rev Med. 2009;60:167-79.
3. Zhang X, Gao S, Li Z, Wang W, Liu G. Identification and Analysis of Estrogen Receptor alpha Promoting Tamoxifen Resistance-Related lncRNAs. Biomed Res Int. 2020;2020:9031723.
4. Farazi TA, Hoell JI, Morozov P, Tuschl T. MicroRNAs in human cancer. Adv Exp Med Biol. 2013;774:1-20.
5. Markopoulos GS, Roupakia E, Tokamani M, Chavdoula E, Hatziapostolou M, Polytarchou C, et al. A step-by-step microRNA guide to cancer development and metastasis. Cell Oncol (Dordr). 2017;40(4):303-39.
6. Sato-Kuwabara Y, Melo SA, Soares FA, Calin GA. The fusion of two worlds: non-coding RNAs and extracellular vesicles–diagnostic and therapeutic implications (Review). Int J Oncol. 2015;46(1):17-27.
7. Rajarajan D, Kaur B, Penta D, Natesh J, Meeran SM. miR-145-5p as a predictive biomarker for breast cancer stemness by computational clinical investigation. Comput Biol Med. 2021;135:104601.
8. Cheng L, Zhang X, Huang YZ, Zhu YL, Xu LY, Li Z, et al. Metformin exhibits antiproliferation activity in breast cancer via miR-483-3p/METTL3/m(6)A/p21 pathway.Oncogenesis. 2021;10(1):7.
9. Ou C, Sun Z, Li X, Li X, Ren W, Qin Z, et al.MiR-590-5p, a density-sensitive microRNA, inhibits tumorigenesis by targeting YAP1 in colorectal cancer. Cancer Lett. 2017;399:53-63.
10. Bastami M, Choupani J, Saadatian Z, Zununi Vahed S, Ouladsahebmadarek E, Mansoori Y, 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(20).
11. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498-504.

12. Vlachos IS, Zagganas K, Paraskevopoulou MD, Georgakilas G, Karagkouni D, Vergoulis T, et al. DIANA-miRPath v3.0: deciphering microRNA function with experimental support. Nucleic Acids Res. 2015;43(W1):W460-6.
13. Gyorffy B. Survival analysis across the entire transcriptome identifies biomarkers with the highest prognostic power in breast cancer. Comput Struct Biotechnol J. 2021;19:4101-9.
14. 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(4):402-8.
15. Espinoza-Sanchez NA, Gotte M. Role of cell surface proteoglycans in cancer immunotherapy. Semin Cancer Biol. 2020;62:48-67.
16. 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(16):6971-82.
17. De Cola A, Lamolinara A, Lanuti P, Rossi C, Iezzi M, Marchisio M, et al.MiR-205-5p inhibition by locked nucleic acids impairs metastatic potential of breast cancer cells. Cell Death Dis. 2018;9(8):821.
18. 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(8):e3188.
19. Gao Z, Wang H, Li H, Li M, Wang J, Zhang W, 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(5):2081-90.
20. Yin Z, Wang W, Qu G, Wang L, Wang X, Pan Q. MiRNA-96-5p impacts the progression of breast cancer through targeting FOXO3. Thorac Cancer. 2020;11(4):956-63.
21. Shao YC, Nie XC, Song GQ, Wei Y, Xia P, Xu XY. Prognostic value of DKK2 from the Dickkopf family in human breast cancer. Int J Oncol. 2018;53(6):2555-65.

Figures

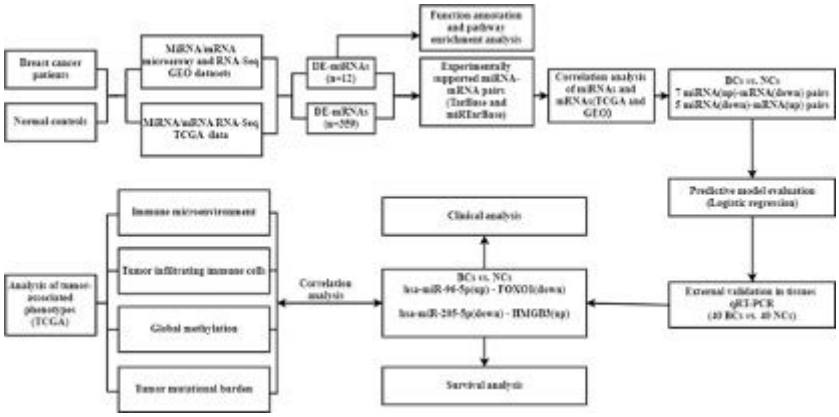

Figure 1

Flow chart for identifying the miRNA-mRNA regulatory pairs and the comprehensive analysis of regulatory pairs role in breast cancer (BC).

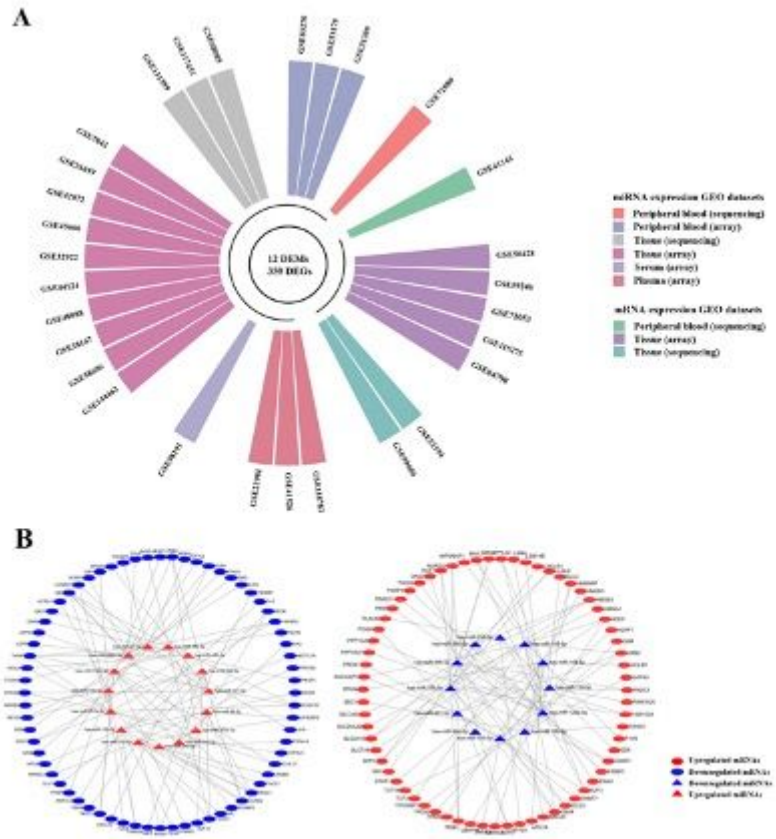

Figure 2

(A) *The circular-barplot shows GEO datasets' basic information (GEO accession, source name and experiment type).****(B)*** *27 microRNAs (miRNAs) to 135 mRNAs network visualized by Cytoscape. There were 177 miRNA (up) - mRNA (down) pairs and 247 miRNA (down) - mRNA (up) pairs screened out by Tarbase and miRTarBase, which contain experimentally validated miRNA-mRNA regulatory pairs. The red mark represents the upregulated miRNAs/mRNAs in BCs versus NCs, while the blue mark represents the downregulated miRNAs/mRNAs in BCs versus NCs. The circle represents mRNAs and the triangle represents miRNAs.*

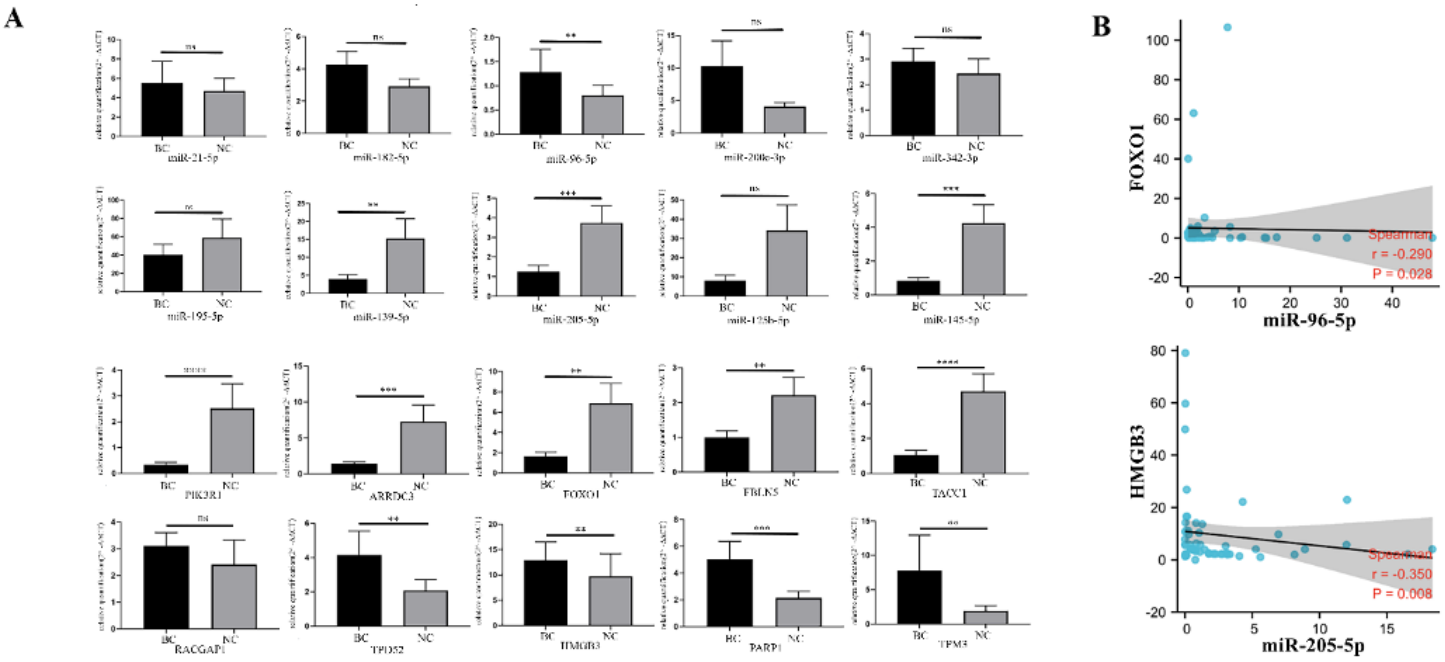

Figure 3

(A) *Validating the expression of 10 differentially expressed miRNAs and ten differentially expressed mRNAs by qRT-PCR.* ***(B)*** *Pearson's correlation analysis of miRNA-mRNA regulatory pairs in 80 samples. Two negative correlated miRNA-mRNA regulatory pairs were plotted.*

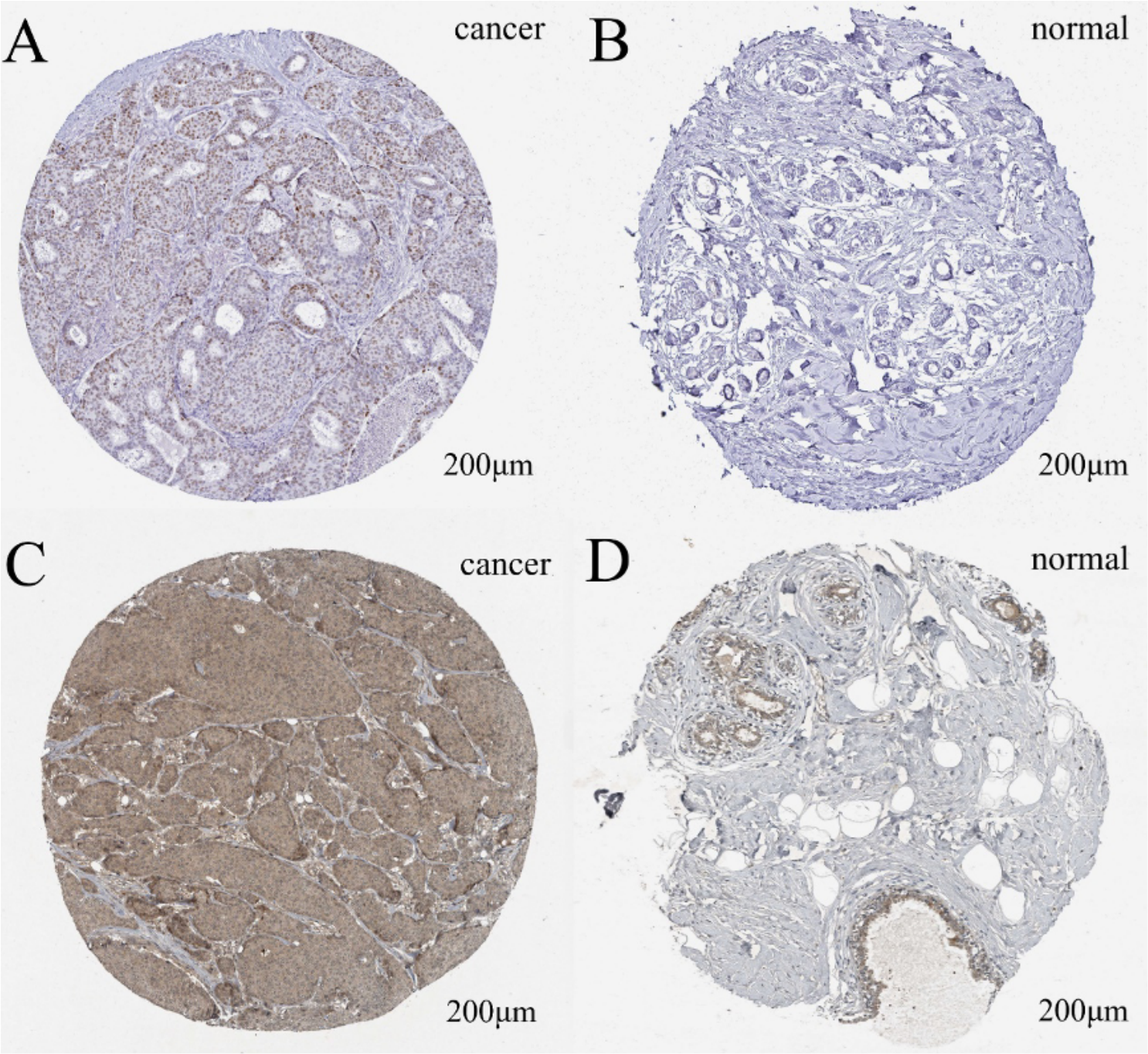

Figure 4

Immunohistochemistry images of HMGB3 and FOXO1 in BCs and NCs from HPA database. ***(A)****Medium immunostaining of HMGB3 in BC cells (antibody HPA062583);* ***(B)****Immunostaining of HMGB3 was not detected in normal breast cells (antibody HPA062583);* ***(C)*** *Low immunostaining of FOXO1 in BC cells (antibody CAB022326);* ***(D)****Medium immunostaining of FOXO1 in normal breast glandular cells (antibody CAB022326).*

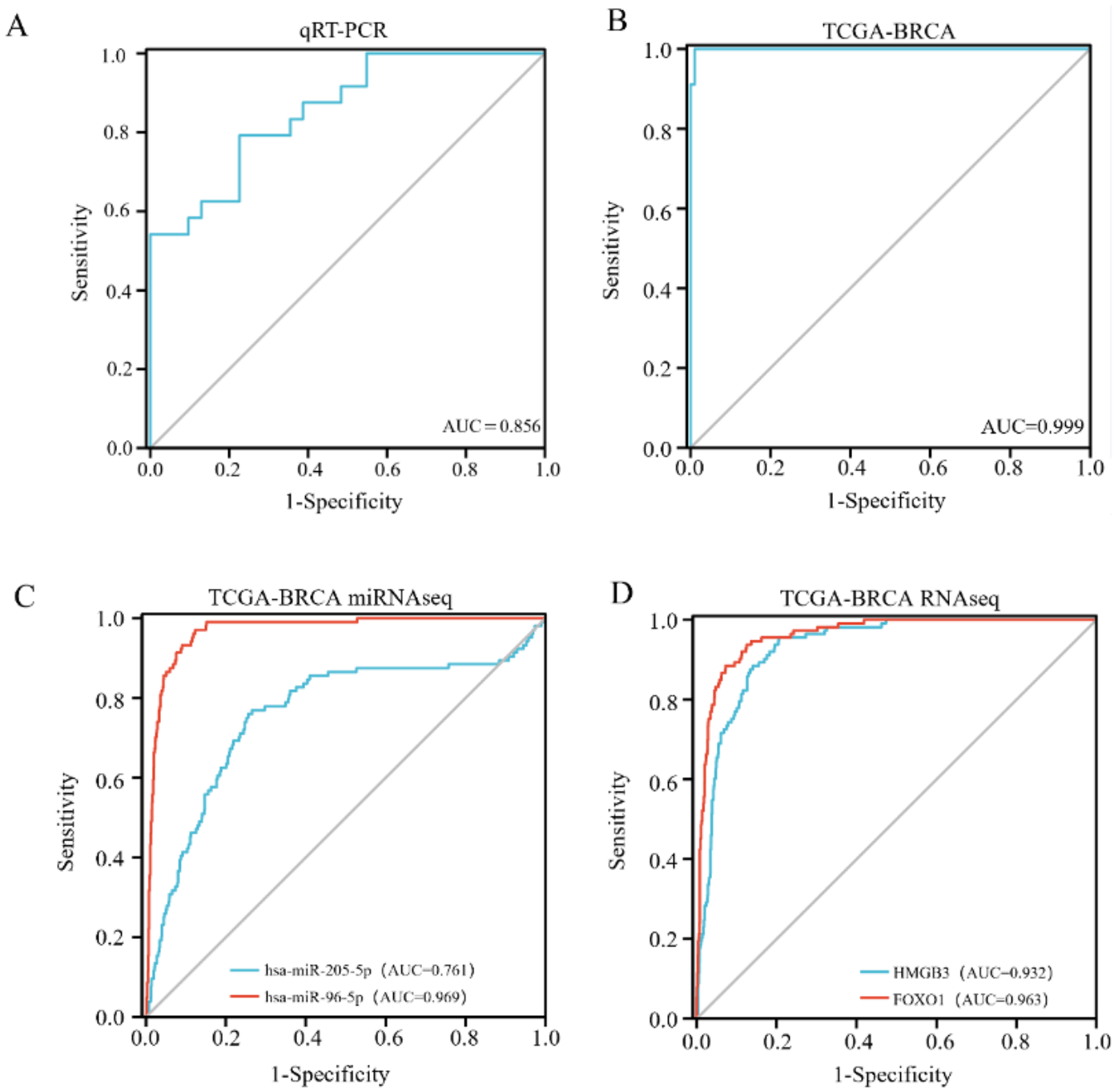

Figure 5

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 samples 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.*

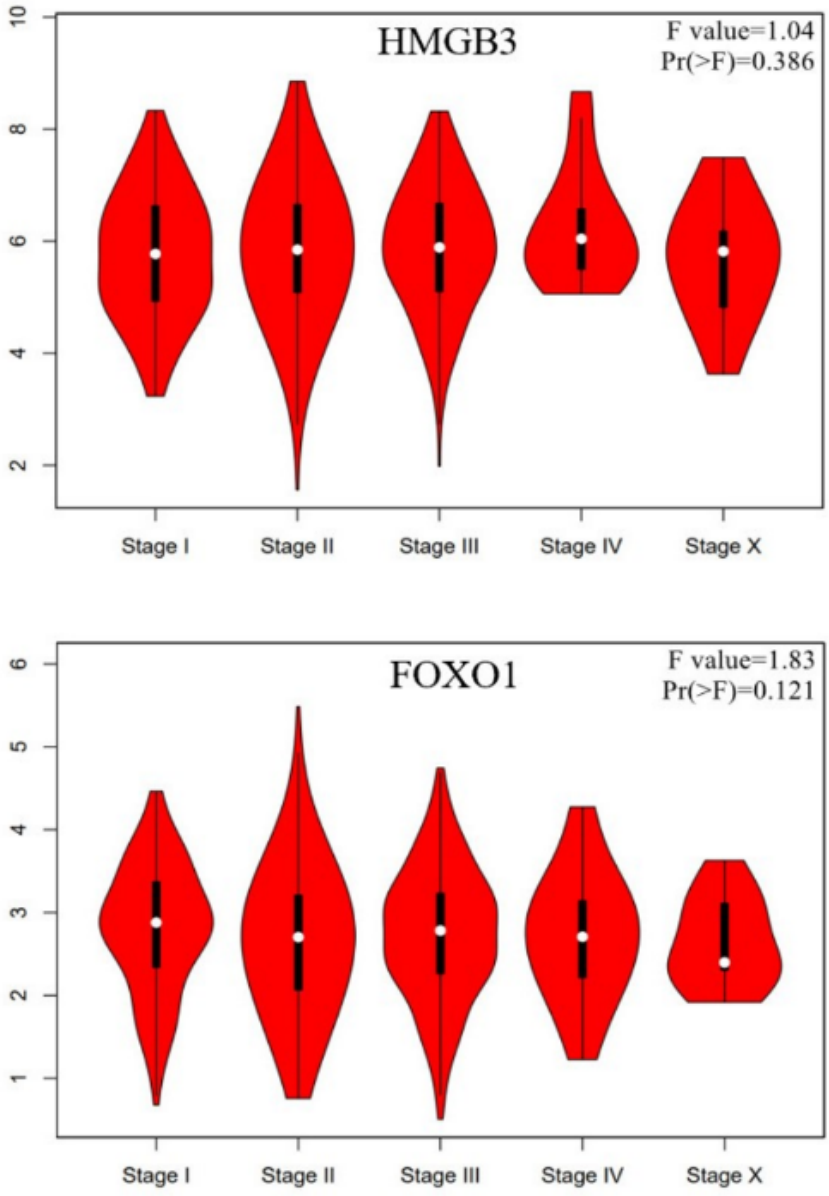

Figure 6

Correlation between HMGB3 and FOXO1 expression and tumor stage in BC (GEPIA). We used log_2(TPM + 1) for the y-axis log-scale; the method for differential gene expression analysis is the one-way analysis of variance, using the pathological stage as a variable for calculating differential expression. The larger F, the better the study fit, and P-value <0.05 as statistically significant (GEPIA Database).

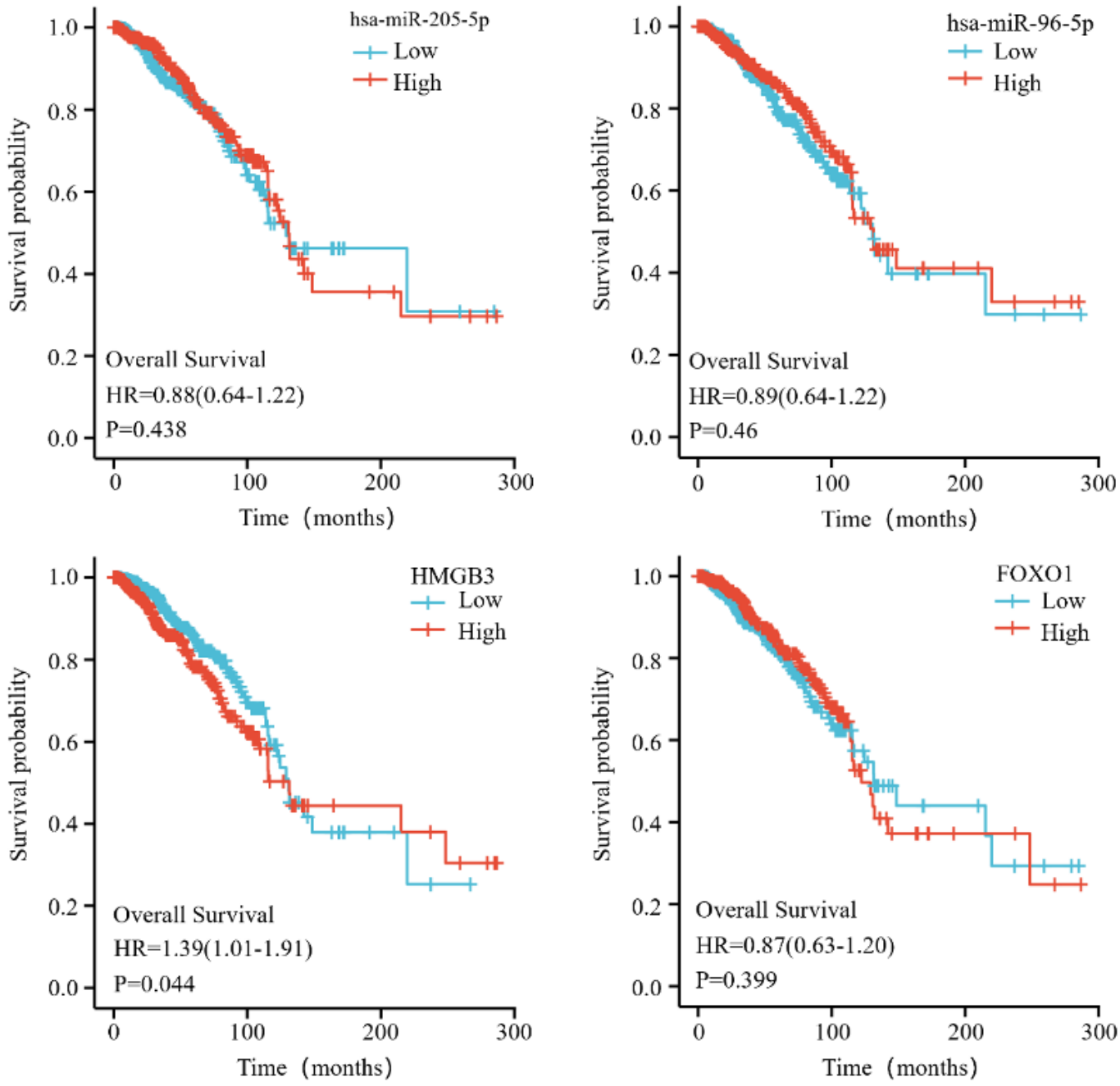

Figure 7

Overall survival of miR-205-5p, miR-96-5p, HMGB3 and FOXO1 expression in BC patients (Kaplan-Meier Plotter).

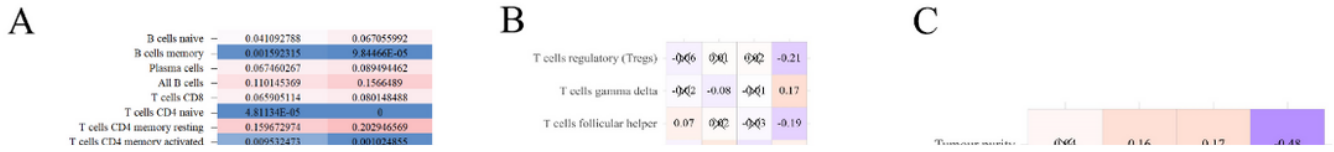

Figure 8

The tumor-related phenotypes of miRNA-mRNA pairs. ***(A)*** *Heatmap showing the immune cell subset proportions in BCs and NCs. Data are presented as mean.* ***(B)*** *Association between expression of miRNA-mRNA pairs and the proportions of 22 subsets of immune cells by Pearson's correlation.* ***(C)*** *Association between expression of miRNA-mRNA pairs 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 p-value > 0.05.*

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- file.xlsx