BMC Cancer

RESEARCH

Open Access

A machine learning model revealed that exosome small RNAs may participate in the development of breast cancer through the chemokine signaling pathway

Jun-luan Mo[1,2,3,4], Xi Li[1], Lin Lei[2], Ji Peng[2], Xiong-shun Liang[2], Hong-hao Zhou[1], Zhao-qian Liu[1], Wen-xu Hong[2*] and Ji-ye Yin[1,3,4*]

Abstract

Background Exosome small RNAs are believed to be involved in the pathogenesis of cancer, but their role in breast cancer is still unclear. This study utilized machine learning models to screen for key exosome small RNAs and analyzed and validated them.

Method Peripheral blood samples from breast cancer screening positive and negative people were used for small RNA sequencing of plasma exosomes. The differences in the expression of small RNAs between the two groups were compared. We used machine learning algorithms to analyze small RNAs with significant differences between the two groups, fit the model through training sets, and optimize the model through testing sets. We recruited new research subjects as validation samples and used PCR-based quantitative detection to validate the key small RNAs screened by the machine learning model. Finally, target gene prediction and functional enrichment analysis were performed on these key RNAs.

Results The machine learning model incorporates six small RNAs: piR-36,340, piR-33,161, miR-484, miR-548ah-5p, miR-4282, and miR-6853-3p. The area under the ROC curve (AUC) of the machine learning model in the training set was 0.985 (95% CI = 0.948-1), while the AUC in the test set was 0.972 (95% CI = 0.882–0.995). RT-qPCR was used to detect the expression levels of these key small RNAs in the validation samples, and the results revealed that their expression levels were significantly different between the two groups ($P < 0.05$). Through target gene prediction and functional enrichment analysis, it was found that the functions of the target genes were enriched mainly in the chemokine signaling pathway.

Conclusion The combination of six plasma exosome small RNAs has good prognostic value for women with positive breast cancer by imaging screening. The chemokine signaling pathway may be involved in the early stage of breast cancer. It is worth further exploring whether small RNAs mediate chemokine signaling pathways in the pathogenesis of breast cancer through the delivery of exosomes.

Keywords Breast cancer, Breast nodules, Exosomes, miRNA, piRNA, Small RNAs

*Correspondence:
Wen-xu Hong
szbloodcenter@hotmail.com
Ji-ye Yin
yinjiye@csu.edu.cn
Full list of author information is available at the end of the article

Introduction

Female breast cancer is the second leading cause of global cancer incidence in 2022, with an estimated 2.3 million new cases, comprising 11.6% of all cancer cases [1]. There are many factors influencing breast cancer, so it is important to select appropriate molecular markers for the prevention of breast cancer, and some new molecular markers have gradually attracted the attention of researchers.

Exosomes are derived from late-stage endosomes and are one of the three main subgroups of extracellular vesicles (EVs) released by mammalian cells (exosomes, microbubbles, and apoptotic vesicles), with diameters ranging from 50 to 150 nm [2, 3]. Exosomes contain proteins, RNA, and lipids, which can participate in regulating important cellular physiological activities [3]. They have been reported to be involved in processes such as the immune response, apoptosis, angiogenesis, and inflammatory response and are believed to be potential therapeutic targets, early diagnostic markers, and targeted drug carriers for various diseases [4].

Small RNAs mainly include microRNAs (miRNAs), PIWI-interacting RNAs (P-element-induced wimpy test-interacting RNAs, piRNAs), transfer RNA(tsRNAs), and small nucleolar RNAs (snoRNAs). Small RNAs participate in important biological processes, such as gene expression and regulation, RNA processing and splicing, protein translation, genetic "invasion" inhibition, and gametogenesis, and are indispensable regulatory factors in life activities. MiRNAs are the main component of small RNAs and the main RNA in exosomes. Some miRNAs in exosomes have been proven by cell experiments to be related to the formation of the microenvironment of breast cancer cells, the proliferation and differentiation of breast cancer cells, and the migration of breast cancer cells [5–11]. piRNAs are believed to participate in the pathogenesis of tumors by regulating key signaling pathways at the transcriptional or posttranscriptional level [12]. It is believed that piRNAs participate in the development of breast cancer through DNA methylation [13–15]. The relationship between piRNAs in exosomes and breast cancer has not been confirmed, but the role of piRNAs in exosomes in breast cancer is worth exploring. Transfer RNAs are crucial for protein synthesis. They are considered not only diagnostic and prognostic indicators but also targets for cancer treatment [16].

To explore the role of small RNAs in exosomes in the development of breast cancer and screen new molecular markers in the early stage of breast cancer, we used blood samples from a breast cancer screening population, sequenced small RNAs in plasma exosomes (including miRNAs, piRNAs and tRNAs), screened key small RNAs in exosomes and used validation samples for qPCR verification. Our research focused on new molecular markers when early breast lesions occur. Breast nodules are one of the early symptoms of breast cancer. The selection of breast cancer screening positive population as the study object is to catch the early stage of breast lesions. Therefore, the results of this study will help to further understand the molecular function of small RNAs in exosomes, and explore a new way for the accurate prevention and treatment of breast cancer.

Materials and methods

Study population

The sample of the study population is from the breast cancer screening project in Shenzhen, a southern city in China. This is a free public health project. Led by the National Cancer Center of China, this study was carried out in multiple cities. This project recruits permanent urban female residents aged between 40 and 74 years to voluntarily participate in breast imaging examinations. Whole blood samples were taken for routine blood analysis, and the remaining plasma samples were frozen for use. All participants signed an informed consent form before participating in the public health project. The project was approved by the Ethics Review Committee of the National Cancer Center of China. In this study, frozen plasma samples were used for testing, and there was no harm to the health or interests of the study population. This study followed the Helsinki Declaration and was approved by the Ethics Committee of the Shenzhen Chronic Disease Control Center.

Imaging examination

Women aged 45 and older underwent ultrasound examination and X-ray radiography; women under the age of 45 were examined by ultrasound, and X-ray radiography was added if suspicious or positive results were found. Diagnosis by ultrasound and imaging was made by physicians with more than 5 years of working experience. If there were positive lesions, a review by a senior physician was needed. The inspection equipment used was a color ultrasound diagnostic apparatus equipped with a high-frequency linear array probe and a mammography camera with a high-resolution digital acquisition imaging system. Each patient was classified according to the diagnostic criteria established by the American College of Radiology Breast Imaging Reporting and Data System (BI-RADS) [17]. The process of this study is shown in Fig. 1.

Extraction and purification of exosomes

Venous blood was collected from the study population using a 5 ml EDTA blood collection vessel. Blood cells and plasma were centrifuged and stored at -80 °C until

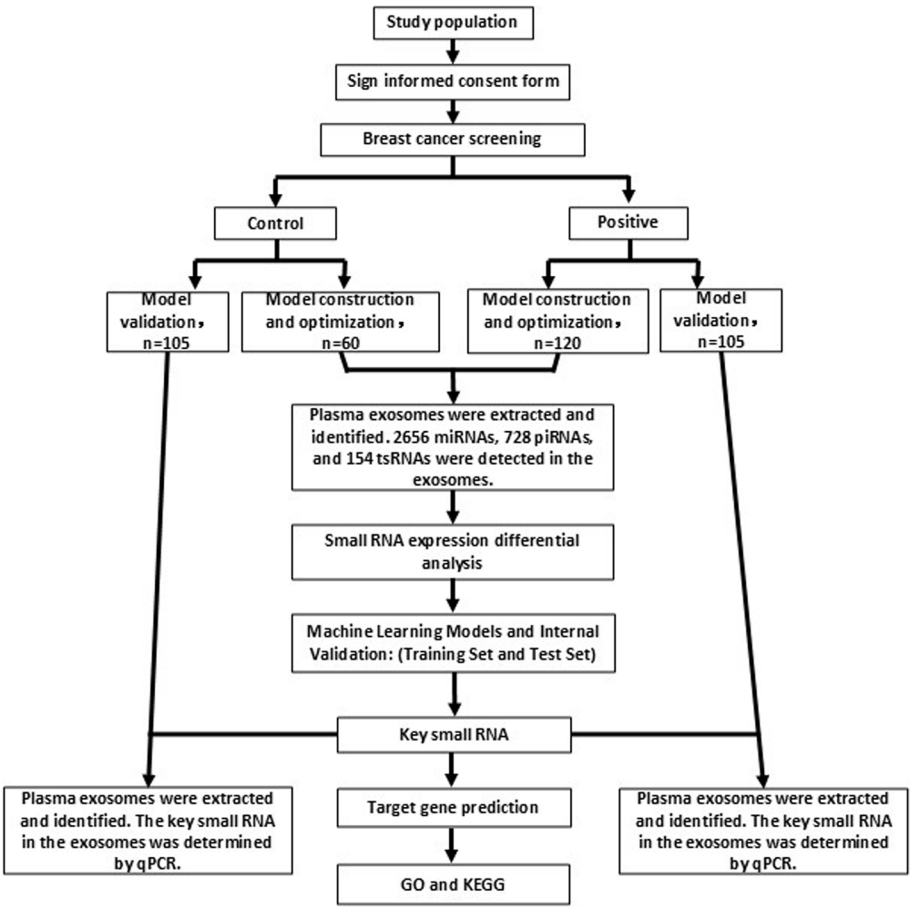

Fig. 1 Research flowchart

use. After the plasma samples were returned to room temperature, the exosomes were separated using a kit. The exoEasy Maxi Kit (QIAGEN Company) was used for the operation. Briefly, a syringe filter was used to filter the plasma to remove more than 0.8 μM particles. One volume of buffer solution (XBP) was added to 1 volume of sample, the test tube was gently inverted 5 times, the mixture was mixed evenly, and the mixture was allowed to rise to room temperature. The sample XBP mixture was added to an exoEasy centrifuge column and centrifuged at 500 × g for 1 min. The effluent was discarded, and the column was placed back in the same collection tube. Next, 10 ml of buffer solution XWP was added, and the mixture was centrifuged at 5000 × g for 5 min to remove the remaining buffer solution from the column. The flow solution was discarded together with the collection tube, and the centrifuge column was transferred to a new collection tube. Then, 400 μL of buffer XE was added to the membrane, which was subsequently incubated for 1 min and centrifuged at 500 × g for 5 min to collect the eluent. The eluent was subsequently transferred to an exoEasy rotating column membrane, which was subsequently incubated for 1 min and centrifuged at 5000 × g for 5 min to collect the eluent, which was subsequently transferred to an EP tube for storage.

Identification of exosomes

The isolated exosomes were selected for identification, including NTA (Nanoparticle Tracking Analysis), TEM (transmission electron microscope), and flow cytometry detection of exosome surface proteins. The nanoparticle tracking analysis instrument (NanoSight ns300, Malvern, UK) was used to detect the concentration and particle size of exosomes. TEM (HT-7700 transmission electron microscope, Hitachi Company, Japan) was used to observe the exosomes. Flow cytometry (Accuri C6 flow cytometry instrument, BD company, USA) was used to detect surface proteins in the exosomes. The procedure of flow cytometry was briefly described as follows [18, 19]: exosomes precipitated by extraction and separation were redissolved by filtered PBS, divided into 100μL tubes and stored in sealed ice. Non-stained exosomes

were used as negative controls and labeled as NC; The samples were stained with CD63 (CD63-Antibody-FIT, BD, No.557288) and CD81 (CD81-Antibody-PE, BD, No.555676) and labeled as CD63 and CD81 [20–22]. All operations were performed in accordance with the standard operating procedures of the equipment.

Total RNA extraction and small RNA sequencing of exosomes

The total RNA in the exosomes was extracted using a kit (QIAGEN exoRNeasy Serum/Plasma Maxi Kit, QIAGEN Corporation). The library was constructed using the proprietary digital NGS technology of QIAseq NGS panels based on the library building strategy of the QIAseq miRNA Library Kit. Sequencing was performed using the NextSeq CN500 sequencing platform in SE75 mode. In addition, 2656 miRNAs, 728 piRNAs, and 154 tsRNAs were analyzed. The expression level of small RNA was expressed using a standardized TPM value.

Quality Control: The raw sequencing data, including the quality distribution of nucleotides, position-specific sequencing quality, GC content, proportion of PCR duplications, kmer frequency, etc., were also evaluated by FAST-QC.

Bioinformatics analysis

Differential expression and clustering

We used software to perform differential analysis on the expression levels of two groups of genes. The R package "DESeq2" was used to analyze the differences in gene expression between two groups of samples. The GraphPad Prism software tool (version 6.01, https://www.graphpad.com//) was used to construct volcano plots of gene expression. The cutoff values of differentially expressed RNA in volcano plots were |log2(FoldChange)| > 1 and adj$P$ < 0.05.The MeV software tool (V4.9.0, https://mev-tm4.sourceforge.net/svnroot/mev-tm4/trunk) was used to construct clustering heatmaps. The cutoff values of differentially expressed RNA in the cluster heat map were |log2(FoldChange)| > 1.584 and adj$P$ < 0.05.

Machine learning model construction

A total of 180 samples were randomly divided into training sets and test sets at a 1:1 ratio via the random number method. Training sets are used to fit model parameters and build models, while test sets are used for testing and optimization to obtain the best performance indicators of the model. The small RNAs of the exosomes incorporated into the machine learning model were further analyzed and verified by qPCR as key small RNAs.

RT–qPCR

We randomly selected 210 research subjects as validation subjects. They participated in the imaging examination of breast cancer in the same way as the previous method. We collected whole blood samples for RT-qPCR. For the aforementioned key small RNAs, the tail addition method was used to detect miRNAs, and the stem ring method was used to detect piRNAs. In the RT–qPCR experiments, cel-miR-39-3p was used as the external reference gene, and three replicates were used for each. The main experimental process is briefly summarized as follows. We used a reagent kit (RiboTM Exosome Isolation Reagent Kit, Guangzhou Ruibo Biotechnology Co., Ltd.) to extract exosomes from plasma. A small number of exosome samples were selected for identification using the aforementioned method. A small number of exosome samples were selected for identification using the aforementioned method. Total RNA in the exosomes was extracted using a kit (HiPure Serum/Plasma miRNA Kit, Meiji Biology). Total RNA extracted from exosomes was used as a template for cDNA reverse transcription and quantitative PCR detection. The reaction system, experimental conditions, and operating steps followed the kit instructions (miDETECT A TrackTM miRNA RT–qPCR Starter Kit, Bulge LoopTM miRNA RT–qPCR Starter Kit, Guangzhou Ruibo Biotechnology Co., Ltd.). qPCR was performed using a quantitative PCR instrument (CFX96, Bio-Rad, USA). The conditions for cDNA reverse transcription were as follows. Tailing method (for miRNA): Prepare a "tailing" reaction system according to the operating manual, and react at 37 °C for 60 min to obtain the tailing product. A new reaction system was established using the tail product as a template. The reaction was carried out at 42 °C for 60 min and 72 °C for 10 min, and the product was diluted three times to form a cDNA template. Stem ring method (for piRNA): A reaction system was established using the extracted total RNA as a template at 42 °C for 60 min and 72 °C for 10 min. The product was the cDNA template. qPCR was carried out under the following conditions: initial denaturation at 95 °C for 10 min, followed by 40 amplification cycles of 95 °C denaturation for 15 s, 60 °C annealing for 30 s, and 70 °C extension for 30 s. The gene and amplified primer sequences are shown in Table 1(Gene and primer sequences of RT-qPCR).

Target gene prediction

The target genes for small RNAs were queried in the database (http://www.mirbase.org; http://genome.bioch.virginia.edu/trfdb/search.php; and http://pirnabank.ibab.ac.in/). Miranda (http://www.microrna.org/microrna/home.do) and RNAhybrid (https://bibiserv.cebit

Table 1 Sequence of genes and primers for RT qPCR

Gene Name	MiRNA/piRNA sequence (5′-3′)	qPCR Primer sequence (5′-3′)	RT Primer sequence (5′-3′)
miR-39-3p	UCACCGGGUGUAAAUCAGCUUG	CCGGGTCACCGGGTGTAAATC	GTGCAGGGTCCGAGGTCAGAGCCACCTGGGCAA TTTTTTTTTTCAAGCT
hsa-miR-484	TCAGGCTCAGTCCCCTCCCGAT	TTCGGTCAGGCTCAGTCCCCT	GTGCAGGGTCCGAGGTCAGAGCCACCTGGGCAA TTTTTTTTTTATCGGG
hsa-miR-548ah-5p	AAAAGTGATTGCAGTGTTTG	CCGGGAAAAGTGATTGCAG	GTGCAGGGTCCGAGGTCAGAGCCACCTGGGCAA TTTTTTTTTTCAAACA
hsa-miR-4282	TAAAATTTGCATCCAGGA	CCGGGTAAAATTTGCATCCAGGA	GTGCAGGGTCCGAGGTCAGAGCCACCTGGGCAA TTTTTTTTTTTCCTGG
hsa-miR-6853-3p	TGTTCATTGGAACCCTGCGCAG	GTCGGAGCGTGGGATGTCCATGA	GTGCAGGGTCCGAGGTCAGAGCCACCTGGGCAA TTTTTTTTTTCTGACTTCATGG
pir-36340	GGTCGCTGGTTCGTTTCCGGCTCGAAGGACC	TGGGGTCGCTGGTTCGTTTCCG	GTGCAGGGTCCGAGGTCAGAGCCACCTGGGCAA TTTTTTTTTTGGTCCT
pir-33161	CCGGCTAGCTCAGTCGGTAGAGCATGAGA	CGGCTAGCTCAGTCGGTAGAGC	GTGCAGGGTCCGAGGTCAGAGCCACCTGGGCAA TTTTTTTTTTTCTCAT
miR-Universal-R	-	CAGTGCAGGGTCCGAGGT	-

ec.uni-bielefeld.de/rnahybrid/) software were used to predict targeted regulatory relationships for key small RNAs. The parameter values used for analysis were as follows: energy_miranda < -20, score_miranda > 150, and energy_RNAhybird < -25. A Venn diagram (https://bioinfogp.cnb.csic.es/tools/venny/index.html) was generated, and the results of the intersection of the two prediction software programs were used to construct the final target gene prediction results.

Gene ontology (GO) analysis

We conducted GO analysis to clarify the biological significance of key small RNA-regulated target genes [23]. We downloaded the GO annotations from the NCBI (http://www.ncbi.nlm.nih.gov/), UniProt Corporation (http://www.uniprot.org/), and Gene Ontology (http://www.geneontology.org/). Based on the database, the intersection genes obtained by predicting the target genes of key small RNAs were annotated at three levels: biological process (BP), molecular function (MF), and cellular component (CC). Fisher's test was used to calculate the significance level (*P* value) for each GO, and the FDR was used to correct for the *P* value. Therefore, we screened out the most significant enrichment of GO items ($P < 0.01$). A graph of the GO analysis results is shown.

Pathway analysis

We annotated the intersection genes of key small RNA target genes based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database for pathway annotation [24]. The significance level (*P* value) of Pathway was calculated by Fisher test. The most enriched Pathway Term was selected ($P < 0.05$). A scatter map of pathway enrichment was generated.

Statistical analyses

Statistical analysis was conducted using the SPSS software package (version 26.0, IBM, USA), GraphPad Prism software (version V6, La Jolla, CA, USA), and R 3.63 software package (https://www.r-project.org/). A t test was used for continuous variables, and a chi-square test was used for categorical variables to analyze the differences between the two groups. The relative expression of small RNAs was calculated using the $2^{-\Delta\Delta Cq}$ method. We used the "glmnet" language package in R to perform minimum absolute contraction and selection operator (LASSO) regression to screen the best predictor and used the "rms" language package to incorporate the factors selected by LASSO regression into multiple logistic regression analysis to establish a prediction model. Glmnet is mainly used to fit generalized linear models. Linear models were used for regression, logistic regression for classification, and the Cox model for survival analysis. The "multiROC R" package was used to construct receiver operating characteristic (ROC) curves and calculate the area under the ROC curve (AUC). The miRNA–gene network was used to identify the relationships between candidate miRNAs and their target genes via Cytoscape software (Cytoscape_v3.6.1) [25]. A $P < 0.05$ was considered to indicate statistical significance.

Results

Basic information of the population

A total of 4380 urban female residents completed breast imaging screening. We randomly selected 180 people as the study subjects for exosome small RNA sequencing, and an additional 210 people were randomly selected as the study subjects for qPCR quantitative detection. Of

these 180 people, 60 were classified as BI-RADS1, and 120 were classified as BI-RADS3-5. Defined as a negative group and a positive group, respectively. Their average ages were 52.07 ± 7.6 years and 49.33 ± 6.7 years, respectively (negative group vs. positive group, t = 2.35, $p = 0.021$). Among the 210 validation sample populations, 105 were classified as BI-RADS1, and 105 were classified as BI-RADS3-5. Their average ages were 58.11 ± 7.4 years and 55.76 ± 7.2 years, respectively (negative vs. positive, t = 0.049, $p = 0.020$).

Identification of plasma exosomes

We successfully isolated exosomes from plasma. Three methods were used to identify exosomes: detection of exosome concentration and particle size, transmission electron microscopy, and flow cytometry detection of exosome surface proteins. The complete morphology and structure of exosomes can be observed via transmission electron microscopy (TEM). The morphology of exosomes was intact under electron microscope (Fig. 2A, B). The results of nanoparticle tracking analysis (NTA) showed that the main peak of the particle size was at 108 nm, the average diameter of the exosomes was 156.5 ± 4.8 nm, and the concentration of the exosomes in the original solution was $1.07 \times 10^9 \pm 2.12 \times 10^6$ particles/mL (Fig. 2C). The diameter of the particles was mainly distributed between 40 and 200 nm (Fig. 2D). The expression of two surface proteins, CD63 and CD81, in the sample was analyzed using a BD Accuri C6 flow cytometer. The expression levels of CD63 and CD81 were higher than those of negative control (Fig. 2E-I).

Differential analysis of gene expression

Gene expression difference analysis revealed that the expression of 49 small RNAs in plasma exosomes was significantly downregulated (fold change < -2, $P < 0.05$, pad j/FDR < 0.05), and the expression of 276 small RNAs was significantly upregulated (fold change > 2, $P < 0.05$, pad j/FDR < 0.05). Among them, 40 miRNAs were significantly downregulated, and 269 miRNAs were significantly upregulated. The expression of 8 piRNAs was significantly downregulated, while the expression of 5 piRNAs was significantly upregulated. The expression of one tsRNA was significantly downregulated, while the expression of two tsRNAs was significantly upregulated (Table 2, Gene list-those with significant expression differences between the two groups). The volcano plot reflects the expression of all genes in both groups (Fig. 3A to D).

The clustering heatmap results of genes with significant differences in expression showed that the samples were significantly clustered into 2 clusters, with cluster 1 being the negative group sample and cluster 2 being the positive group sample. The genes corresponding to the sample cluster are clustered into different clusters. Among them, one gene cluster exhibited the most obvious change among the two sample clusters. The gene expression of this cluster was upregulated in sample cluster 1 and downregulated in sample cluster 2 (Fig. 3E, gene cluster 1).

Machine learning model and RT-qPCR validation

According to the analysis of gene expression differences, genes with significant differences in expression were used to construct machine learning models. The selection criteria for differential gene expression were |log2(FoldChange)| > 0.585, adj$P < 0.05$. A total of 180 samples were randomly divided into training and testing sets at a 1:1 ratio. A training set model was constructed and validated on the test set. The model is a linear model with an intercept of 1.2779, and the variables and coeff values are hsa-miR-548ah-5p (-0.0038), pir-33,161 (-0.0135), pir-36,340 (-0.0196), hsa-miR-484 (-0.003), hsa-miR-4282 (-0.0056), and hsa-miR-6853-3p (-0.0915). In the training set, the area under the ROC curve (AUC) of the model was 0.985 (95% CI = 0.948-1). In the test set, the AUC was 0.972 (95% CI = 0.882–0.995). As shown in Fig. 4A.We selected people who recently participated in breast cancer screening for external validation of the model. qPCR was used to detect key small RNAs (hsa-miR-548ah-5p, pir-33161, pir-36340, hsa-miR-484, hsa-miR-4282, and hsa-miR-6853-3p) in the model. We completed qPCR testing on a total of 210 validation samples. Among them, 105 were in the positive group, and 105 were in the control group. The relative quantitative results revealed significant differences in the

(See figure on next page.)

Fig. 2 Identification of exosomes. **A**, **B** Exosome particles captured by TEM. **C** NTA was used to detect exosome concentration and particle size. The main peak of the particle size is at 108 nm. The average diameter was 156.5 ± 4.8 nm. The concentration of exosomes in the original solution was 1.07×10^9 particles/mL. **D** The scatter plot of the luminosity and particle diameter detected by NTA shows that the particle diameter is mainly distributed between 40 and 200 nm. **E**-**I** Flow cytometry analysis. **E** No gating, dotted line is Gate P1; **F** The positive proportion of negative control (NC) was 0.0%, and the negative proportion was 100%; **G** The positive proportion of CD63 + was 61.8%, and the negative proportion was 38.2%. **H** The positive proportion of negative control (NC) was 0.0% and the negative proportion was 100%. **I** The positive proportion of CD81 + was 77.1%, and the negative proportion was 22.9%

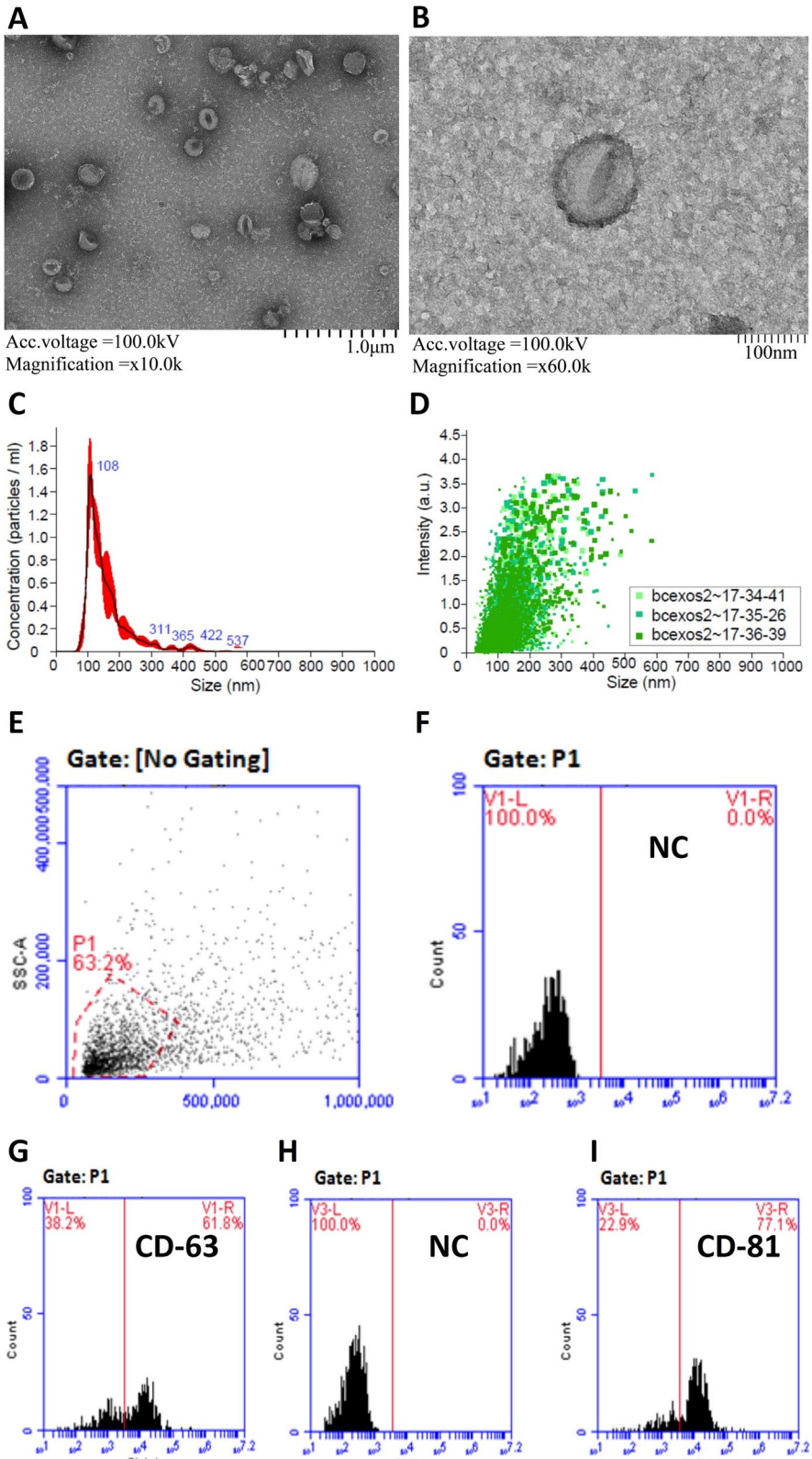

Fig. 2 (See legend on previous page.)

Table 2 Gene list-those with significant expression differences between the two groups

GENE ID	baseMean	log2FoldChange	lfcSE	stat	*p* value	padj
pir-43768@piRNAcluster	5.273922	-4.06666	0.442546	-9.18923	3.96E-20	1.17E-17
pir-36340@tRNA	6.034089	-3.75914	0.374987	-10.0247	1.19E-23	4.11E-21
hsa-miR-3189-3p	16.50177	-3.72703	0.420095	-8.87187	7.19E-19	1.66E-16
pir-33123@piRNAcluster	3.734723	-2.8601	0.431424	-6.62943	3.37E-11	1.84E-09
hsa-miR-6790-3p	65.24721	-2.83862	0.360109	-7.88267	3.20E-15	3.92E-13
pir-36339@tRNA	12.64644	-2.47089	0.335068	-7.37431	1.65E-13	1.49E-11
hsa-miR-4634	69.17661	-2.36656	0.271033	-8.73165	2.51E-18	4.35E-16
hsa-miR-762	56.99161	-2.00153	0.2339	-8.55721	1.16E-17	1.85E-15
hsa-miR-484	225.1639	-1.92533	0.166614	-11.5557	6.91E-31	4.79E-28
hsa-miR-663a	14.78654	-1.92185	0.252239	-7.61915	2.55E-14	2.79E-12
hsa-miR-602	11.94308	-1.84384	0.331742	-5.55807	2.73E-08	6.75E-07
hsa-miR-4782-3p	3.206305	-1.76321	0.40544	-4.34888	1.37E-05	0.000131
hsa-miR-6853-3p	2.722824	-1.7519	0.38386	-4.5639	5.02E-06	5.84E-05
hsa-miR-6516-3p	12.26504	-1.68565	0.271719	-6.20364	5.52E-10	2.05E-08
hsa-miR-3130-3p	55.42746	-1.65168	0.241425	-6.84137	7.84E-12	4.94E-10
hsa-miR-10401-3p	7.720431	-1.63142	0.308082	-5.29541	1.19E-07	2.35E-06
hsa-miR-4492	23.98364	-1.5929	0.274903	-5.7944	6.86E-09	1.93E-07
hsa-miR-3130-5p	56.93336	-1.55454	0.236628	-6.56952	5.05E-11	2.56E-09
hsa-miR-3651	4.681523	-1.52806	0.324837	-4.70409	2.55E-06	3.31E-05
hsa-miR-601	34.95517	-1.51554	0.28377	-5.34073	9.26E-08	1.96E-06
hsa-miR-6883-5p	52.61562	-1.4356	0.330165	-4.34814	1.37E-05	0.000131
hsa-miR-6131	5988.582	-1.40098	0.273102	-5.12988	2.90E-07	4.98E-06
hsa-miR-4466	67.09178	-1.35756	0.244457	-5.55338	2.80E-08	6.85E-07
tsrna-5011a-proagg/protgg/procgg	4.94054	-1.28152	0.234001	-5.47656	4.34E-08	1.02E-06
hsa-miR-4282	3.530108	-1.27822	0.308512	-4.14318	3.43E-05	0.000282
pir-33161@piRNAcluster	7.507297	-1.27485	0.203495	-6.26478	3.73E-10	1.44E-08
hsa-miR-8061	6.501053	-1.24114	0.420317	-2.95287	0.003148	0.012955
hsa-miR-519e-5p	6.734538	-1.20611	0.314765	-3.83178	0.000127	0.00085
hsa-miR-548ah-5p	4.588421	-1.19387	0.307233	-3.88587	0.000102	0.000702
pir-36037@piRNAcluster	12.04192	-1.14818	0.293136	-3.91689	8.97E-05	0.000628
pir-36742@tRNA	9.669324	-1.14589	0.187618	-6.10758	1.01E-09	3.61E-08
hsa-miR-548ar-5p	2.417538	-1.13658	0.294579	-3.85832	0.000114	0.00077
hsa-miR-4795-3p	2.434752	-1.11023	0.459949	-2.4138	0.015787	0.047066
hsa-miR-206	64.03445	-1.09601	0.331554	-3.30568	0.000947	0.004654
hsa-miR-5585-5p	96.10953	-1.09467	0.249482	-4.38775	1.15E-05	0.000113
hsa-miR-4319	3.346996	-1.09435	0.371739	-2.94387	0.003241	0.013285
hsa-miR-450b-5p	5.880137	-1.07624	0.280854	-3.83203	0.000127	0.00085
hsa-miR-652-3p	77.41713	-1.06306	0.175972	-6.04107	1.53E-09	4.97E-08
hsa-miR-92a-3p	712.0584	-1.06078	0.140693	-7.53968	4.71E-14	4.45E-12
hsa-miR-320a-3p	1069.457	-1.06051	0.117297	-9.04126	1.55E-19	4.02E-17
hsa-miR-576-5p	16.58124	-1.05772	0.20887	-5.06402	4.11E-07	6.61E-06
hsa-miR-675-5p	18.74678	-1.05166	0.33257	-3.16221	0.001566	0.007135
pir-43604@piRNAcluster	5.822169	-1.03926	0.213208	-4.87438	1.09E-06	1.51E-05
hsa-miR-8058	2.669253	-1.03576	0.410629	-2.52237	0.011657	0.036813
hsa-miR-219a-1-3p	27.40213	-1.03106	0.259004	-3.98089	6.87E-05	0.000503
hsa-miR-493-5p	18.97664	-1.02962	0.232389	-4.43058	9.40E-06	9.57E-05
hsa-miR-4529-3p	14.16044	-1.02096	0.296001	-3.44917	0.000562	0.002958
hsa-miR-4741	3.969526	-1.0209	0.384488	-2.65522	0.007926	0.027403
hsa-miR-4305	8.901559	-1.01817	0.338903	-3.00432	0.002662	0.011219

Table 2 (continued)

GENE ID	baseMean	log2FoldChange	lfcSE	stat	*p* value	padj
hsa-miR-6812-5p	15.19577	1.001935	0.281884	3.55443	0.000379	0.002133
hsa-miR-29c-3p	1164.239	1.005001	0.138658	7.248043	4.23E-13	3.66E-11
hsa-miR-4270	4.300679	1.014802	0.400094	2.536413	0.011199	0.035804
hsa-miR-6758-5p	12.48407	1.015534	0.314264	3.231464	0.001232	0.005777
hsa-miR-5193	45.86903	1.017913	0.19662	5.177068	2.25E-07	4.07E-06
hsa-miR-152-5p	6.157649	1.022675	0.355897	2.873513	0.004059	0.015767
hsa-miR-204-5p	11.40335	1.023067	0.367882	2.780963	0.00542	0.020111
hsa-miR-10400-5p	22.42546	1.023468	0.232775	4.396805	1.10E-05	0.00011
hsa-miR-6780b-3p	45.4451	1.023713	0.267398	3.828432	0.000129	0.000858
hsa-miR-6511a-3p	8.637987	1.024062	0.286526	3.574059	0.000351	0.002029
hsa-miR-8085	49.57246	1.027942	0.181148	5.674588	1.39E-08	3.66E-07
hsa-miR-3165	3.673989	1.031082	0.365429	2.821564	0.004779	0.018155
hsa-miR-619-5p	82.03324	1.03334	0.258805	3.992743	6.53E-05	0.000483
hsa-miR-4271	20.71882	1.035391	0.239119	4.330016	1.49E-05	0.00014
hsa-miR-4685-3p	38.93117	1.035991	0.249993	4.144074	3.41E-05	0.000282
hsa-miR-5739	10.29803	1.039595	0.293141	3.546403	0.000391	0.002181
hsa-miR-3065-3p	12.94636	1.040422	0.243731	4.268735	1.97E-05	0.000175
hsa-miR-3529-3p	61.78114	1.040988	0.152351	6.832831	8.33E-12	5.09E-10
hsa-let-7c-5p	3123.368	1.041179	0.130741	7.963692	1.67E-15	2.17E-13
hsa-miR-6867-5p	42.91252	1.043958	0.221825	4.706213	2.52E-06	3.30E-05
hsa-miR-4509	18.02722	1.045982	0.29226	3.57894	0.000345	0.001997
hsa-miR-4496	4.972344	1.046232	0.346395	3.020341	0.002525	0.010795
hsa-miR-4279	169.8719	1.046285	0.233754	4.476018	7.60E-06	8.02E-05
hsa-miR-12116	22.1185	1.04758	0.265028	3.952714	7.73E-05	0.000552
hsa-miR-320e	23.70333	1.048723	0.255311	4.107632	4.00E-05	0.000318
hsa-miR-6760-5p	31.87524	1.049963	0.266394	3.941387	8.10E-05	0.000575
hsa-miR-335-5p	627.7201	1.054713	0.187608	5.621902	1.89E-08	4.85E-07
hsa-miR-6858-5p	3.747282	1.055385	0.402909	2.619413	0.008808	0.029713
hsa-miR-4763-5p	10.73812	1.059821	0.334984	3.163793	0.001557	0.007112
hsa-miR-136-5p	18.43608	1.068841	0.264946	4.034179	5.48E-05	0.000416
hsa-miR-4728-5p	4.338289	1.069886	0.427763	2.50112	0.01238	0.038512
hsa-miR-4676-3p	4.29253	1.073184	0.417222	2.572215	0.010105	0.03312
hsa-miR-6747-5p	13.96205	1.074185	0.31437	3.416943	0.000633	0.003282
hsa-miR-6769a-3p	3.289954	1.07604	0.409037	2.630668	0.008522	0.02903
hsa-miR-6845-3p	7.253508	1.076678	0.318013	3.385637	0.00071	0.003626
hsa-miR-3668	8.718136	1.077706	0.396371	2.718932	0.006549	0.023304
hsa-miR-6809-3p	205.3226	1.07778	0.214965	5.013753	5.34E-07	8.34E-06
hsa-miR-6870-3p	8.881645	1.081087	0.350799	3.081782	0.002058	0.009117
hsa-miR-455-3p	3.125794	1.087061	0.435321	2.49715	0.01252	0.038888
hsa-miR-4701-3p	3.835101	1.089926	0.392156	2.77932	0.005447	0.020141
hsa-miR-6890-5p	5.101537	1.092628	0.405965	2.691431	0.007115	0.024889
hsa-miR-6796-5p	9.492646	1.093146	0.310516	3.52042	0.000431	0.002362
hsa-miR-4632-5p	4.380368	1.093896	0.3186	3.433446	0.000596	0.003104
pir-33382@rRNA	6.467728	1.094969	0.331613	3.301944	0.00096	0.004695
hsa-miR-7110-3p	3716.172	1.098611	0.222294	4.942151	7.73E-07	1.15E-05
hsa-miR-6756-3p	328.0147	1.098988	0.182695	6.01543	1.79E-09	5.59E-08
hsa-miR-152-3p	320.9129	1.099632	0.180107	6.105443	1.03E-09	3.61E-08
hsa-miR-6852-3p	3.676198	1.099888	0.422412	2.603829	0.009219	0.030734
hsa-miR-4516	104.8834	1.102198	0.203332	5.420678	5.94E-08	1.36E-06

Table 2 (continued)

GENE ID	baseMean	log2FoldChange	lfcSE	stat	*p* value	padj
hsa-miR-664b-5p	33.65601	1.103032	0.188804	5.842192	5.15E-09	1.51E-07
hsa-miR-6775-5p	59.72879	1.10318	0.290861	3.792802	0.000149	0.00097
hsa-miR-6767-5p	42.37173	1.10458	0.170696	6.471034	9.73E-11	4.70E-09
hsa-miR-483-3p	139.6355	1.105226	0.20537	5.381644	7.38E-08	1.63E-06
hsa-miR-384	6.403695	1.105891	0.462277	2.392268	0.016745	0.049289
hsa-miR-6766-3p	39.53218	1.106652	0.23409	4.727473	2.27E-06	3.01E-05
hsa-miR-6740-5p	25.62507	1.110597	0.259968	4.27206	1.94E-05	0.000174
tsrna-3017a-valtac	18.50822	1.112356	0.214375	5.18884	2.12E-07	3.86E-06
hsa-miR-6739-5p	27.89208	1.113968	0.240546	4.630995	3.64E-06	4.53E-05
hsa-miR-3679-3p	31.20019	1.115121	0.272078	4.098528	4.16E-05	0.000329
hsa-miR-548m	2.238835	1.117903	0.458107	2.440268	0.014676	0.044522
hsa-miR-6880-5p	14.99494	1.118306	0.245596	4.553433	5.28E-06	5.98E-05
hsa-miR-6832-3p	15.71233	1.120393	0.318188	3.521166	0.00043	0.002362
hsa-miR-212-5p	6.039485	1.121005	0.353325	3.172731	0.00151	0.006912
hsa-miR-299-3p	11.93318	1.121083	0.361106	3.10458	0.001905	0.008543
hsa-miR-3652	8.154903	1.122505	0.386348	2.905423	0.003668	0.014611
tsrna-5004c-glygcc/glyccc	2797.279	1.123204	0.17467	6.430426	1.27E-10	5.88E-09
hsa-miR-6511b-5p	2.407739	1.123572	0.404883	2.775055	0.005519	0.020335
hsa-miR-6070	6.178348	1.134898	0.375272	3.0242	0.002493	0.010703
hsa-miR-4443	20.01355	1.137522	0.213647	5.324318	1.01E-07	2.06E-06
hsa-miR-765	33.1189	1.137608	0.22346	5.090887	3.56E-07	5.88E-06
hsa-miR-3135a	15.96451	1.138659	0.315349	3.61079	0.000305	0.001797
hsa-miR-6870-5p	6.437924	1.141498	0.40764	2.800256	0.005106	0.019118
hsa-miR-6746-5p	14.26383	1.145761	0.251065	4.563601	5.03E-06	5.84E-05
hsa-miR-7107-5p	67.71994	1.148486	0.199708	5.750828	8.88E-09	2.43E-07
hsa-miR-608	6.523094	1.149816	0.354839	3.240391	0.001194	0.005663
hsa-miR-4327	5.762245	1.150165	0.43674	2.633523	0.00845	0.028834
hsa-miR-7851-3p	31.40837	1.15103	0.255023	4.513429	6.38E-06	6.87E-05
hsa-miR-6802-3p	6.540652	1.153505	0.420845	2.740923	0.006127	0.022119
hsa-miR-615-3p	13.45157	1.153668	0.28401	4.062063	4.86E-05	0.000379
hsa-miR-2909	3.64493	1.15583	0.439729	2.628509	0.008576	0.029167
hsa-miR-5003-3p	18.65078	1.156265	0.296519	3.899457	9.64E-05	0.000668
hsa-miR-6780a-3p	24.96015	1.157989	0.265228	4.36602	1.27E-05	0.000122
hsa-miR-4769-3p	35.60592	1.161701	0.221264	5.250284	1.52E-07	2.90E-06
hsa-miR-3196	35.08339	1.163274	0.213803	5.440873	5.30E-08	1.22E-06
hsa-miR-627-3p	90.61546	1.165471	0.226009	5.156743	2.51E-07	4.46E-06
hsa-miR-411-5p	130.6615	1.168592	0.183838	6.356642	2.06E-10	9.03E-09
hsa-miR-3934-5p	12.79118	1.173356	0.330556	3.54964	0.000386	0.002161
hsa-miR-494-5p	4.730812	1.174257	0.453787	2.587683	0.009662	0.03196
hsa-miR-6738-5p	3.246645	1.174325	0.39285	2.989249	0.002797	0.01167
hsa-miR-187-3p	22.18659	1.174999	0.283609	4.143022	3.43E-05	0.000282
hsa-miR-6795-3p	18.04341	1.179051	0.285305	4.132596	3.59E-05	0.000291
hsa-miR-8077	44.36149	1.181471	0.223498	5.286267	1.25E-07	2.42E-06
hsa-miR-4640-3p	23.12602	1.182017	0.266723	4.431631	9.35E-06	9.57E-05
hsa-miR-6799-5p	5.471566	1.183646	0.333269	3.551623	0.000383	0.00215
hsa-miR-367-5p	7.993674	1.185029	0.362522	3.268846	0.00108	0.005182
hsa-miR-3940-5p	12.60028	1.185534	0.268937	4.40823	1.04E-05	0.000105
hsa-miR-6754-3p	5.697167	1.18609	0.366673	3.234732	0.001218	0.005724
hsa-miR-6876-5p	8.75441	1.187619	0.35371	3.357611	0.000786	0.003963

Table 2 (continued)

GENE ID	baseMean	log2FoldChange	lfcSE	stat	*p* value	padj
hsa-miR-4667-3p	17.02965	1.189304	0.236781	5.022794	5.09E-07	8.02E-06
hsa-miR-6892-3p	66.05888	1.189562	0.21466	5.541601	3.00E-08	7.24E-07
hsa-miR-301a-3p	50.20079	1.191282	0.1812	6.57439	4.89E-11	2.56E-09
hsa-miR-4505	3.14565	1.193904	0.494951	2.412168	0.015858	0.04721
hsa-miR-153-3p	9.264027	1.19473	0.360631	3.312889	0.000923	0.004569
hsa-miR-3173-3p	17.16387	1.196625	0.287726	4.158905	3.20E-05	0.000268
hsa-miR-7515	21.39156	1.200374	0.278704	4.306979	1.65E-05	0.000153
hsa-miR-7111-3p	138.933	1.200419	0.234432	5.120533	3.05E-07	5.19E-06
hsa-miR-4513	9.212644	1.20219	0.296491	4.054723	5.02E-05	0.000389
hsa-miR-6877-3p	4.155689	1.204262	0.407293	2.956744	0.003109	0.012819
hsa-miR-1233-5p	3.516214	1.207238	0.453428	2.662471	0.007757	0.026958
hsa-miR-6507-3p	14.50699	1.213101	0.331538	3.659011	0.000253	0.001543
hsa-miR-1238-3p	3.233356	1.214256	0.4966	2.445139	0.01448	0.044118
hsa-miR-7155-5p	6.298462	1.218344	0.414955	2.936086	0.003324	0.013495
hsa-miR-6084	16.28812	1.221418	0.248533	4.914506	8.90E-07	1.29E-05
hsa-miR-5196-5p	16.81292	1.221882	0.272497	4.484021	7.32E-06	7.81E-05
hsa-miR-7161-3p	16.654	1.222935	0.294297	4.15545	3.25E-05	0.000271
hsa-miR-3911	24.55191	1.223167	0.229513	5.329413	9.85E-08	2.03E-06
hsa-miR-6507-5p	27.96597	1.226534	0.27107	4.524793	6.05E-06	6.58E-05
hsa-miR-29a-3p	3982.383	1.230048	0.138877	8.857098	8.21E-19	1.71E-16
hsa-miR-641	32.55563	1.233482	0.242155	5.093764	3.51E-07	5.84E-06
hsa-miR-6891-3p	41.02567	1.234865	0.254776	4.846862	1.25E-06	1.70E-05
hsa-miR-6734-3p	62.49775	1.235453	0.269755	4.579912	4.65E-06	5.58E-05
hsa-miR-432-3p	5.779256	1.235489	0.366775	3.368523	0.000756	0.003849
hsa-miR-6895-3p	4.264528	1.236274	0.441873	2.797805	0.005145	0.019229
hsa-miR-4778-5p	257.7488	1.236613	0.220166	5.616718	1.95E-08	4.93E-07
pir-35284@tRNA	899.3577	1.237676	0.178495	6.933971	4.09E-12	2.83E-10
hsa-miR-9898	7.790836	1.241259	0.379844	3.267815	0.001084	0.005189
hsa-miR-6124	229.071	1.253723	0.233077	5.379003	7.49E-08	1.64E-06
hsa-miR-5585-3p	30.15565	1.255611	0.332348	3.777995	0.000158	0.00102
hsa-miR-1224-3p	46.28261	1.256799	0.222985	5.636248	1.74E-08	4.51E-07
hsa-miR-3934-3p	7.247214	1.260648	0.362211	3.480425	0.000501	0.002674
hsa-miR-4800-3p	21.16399	1.262144	0.285619	4.418979	9.92E-06	0.000101
hsa-miR-34b-3p	5.573271	1.262416	0.491148	2.570335	0.01016	0.033248
hsa-miR-514a-3p	3.089709	1.263711	0.453224	2.788273	0.005299	0.019734
hsa-miR-5698	7.565769	1.263736	0.273928	4.61338	3.96E-06	4.87E-05
hsa-miR-6878-5p	3.0982	1.265209	0.389589	3.247546	0.001164	0.005548
hsa-miR-4290	55.39154	1.270363	0.255287	4.976206	6.48E-07	1.01E-05
hsa-miR-6720-5p	12.33975	1.271448	0.327396	3.883522	0.000103	0.000706
hsa-miR-510-3p	8.268897	1.271725	0.356782	3.564437	0.000365	0.002082
hsa-miR-6858-3p	5.602649	1.276972	0.375458	3.401107	0.000671	0.003452
hsa-miR-4739	335.8693	1.278951	0.201355	6.351708	2.13E-10	9.03E-09
hsa-miR-7110-5p	105.9698	1.279172	0.254087	5.034387	4.79E-07	7.60E-06
hsa-miR-5191	46.02098	1.281654	0.249729	5.132187	2.86E-07	4.96E-06
hsa-miR-6783-3p	10.79338	1.282327	0.315174	4.068633	4.73E-05	0.000369
hsa-miR-10398-5p	15.41961	1.282431	0.293815	4.364756	1.27E-05	0.000122
hsa-miR-512-3p	4.310554	1.282458	0.465697	2.753843	0.00589	0.021364
hsa-miR-1911-5p	18.15194	1.284439	0.24168	5.314633	1.07E-07	2.16E-06
hsa-miR-5088-3p	59.18921	1.289557	0.254742	5.062209	4.14E-07	6.62E-06

Table 2 (continued)

GENE ID	baseMean	log2FoldChange	lfcSE	stat	*p* value	padj
hsa-miR-6758-3p	5.704943	1.293244	0.348458	3.711338	0.000206	0.001283
hsa-miR-9-3p	15.0062	1.296744	0.302457	4.287363	1.81E-05	0.000165
hsa-miR-497-3p	5.209294	1.298128	0.364539	3.561012	0.000369	0.002097
hsa-miR-4668-5p	40.69811	1.300014	0.255878	5.080612	3.76E-07	6.11E-06
hsa-miR-596	3.886692	1.304459	0.430161	3.032488	0.002425	0.010435
hsa-miR-7108-3p	27.10026	1.305591	0.253732	5.145548	2.67E-07	4.66E-06
hsa-miR-6833-3p	53.19394	1.307131	0.287493	4.546648	5.45E-06	6.09E-05
hsa-miR-605-3p	5.769574	1.307647	0.427198	3.060987	0.002206	0.00959
pir-31068@tRNA	1537.993	1.307764	0.172962	7.561005	4.00E-14	3.96E-12
pir-35982@tRNA	1567.356	1.309043	0.170947	7.657616	1.89E-14	2.19E-12
hsa-miR-5187-3p	4.874782	1.312833	0.375243	3.498625	0.000468	0.002544
hsa-miR-524-3p	3.49263	1.314948	0.479889	2.740107	0.006142	0.022119
hsa-miR-4502	5.774322	1.319822	0.428726	3.078477	0.002081	0.00916
hsa-miR-8485	122.0877	1.322776	0.22321	5.926147	3.10E-09	9.48E-08
hsa-miR-3059-3p	11.27161	1.328764	0.329352	4.034487	5.47E-05	0.000416
hsa-miR-6880-3p	26.88382	1.334729	0.274191	4.867889	1.13E-06	1.54E-05
hsa-miR-6727-3p	7.643389	1.335868	0.360073	3.709989	0.000207	0.001286
hsa-miR-193b-3p	3.281447	1.339592	0.510195	2.625647	0.008648	0.029329
hsa-miR-6728-3p	7.231938	1.343479	0.40886	3.285914	0.001017	0.004924
hsa-miR-6828-3p	4.427921	1.348088	0.474585	2.840565	0.004503	0.01733
hsa-miR-2682-5p	12.51355	1.348118	0.477017	2.826145	0.004711	0.017963
hsa-miR-6847-5p	12.43217	1.348759	0.276425	4.879292	1.06E-06	1.48E-05
hsa-miR-3916	1600.508	1.351035	0.222278	6.078123	1.22E-09	4.14E-08
hsa-miR-138-5p	14.35189	1.353084	0.298636	4.530874	5.87E-06	6.43E-05
hsa-miR-6798-3p	15.46678	1.358756	0.265685	5.114159	3.15E-07	5.32E-06
hsa-miR-198	5.486667	1.360558	0.427212	3.184733	0.001449	0.006691
hsa-miR-5588-5p	492.9436	1.366224	0.233213	5.858267	4.68E-09	1.39E-07
hsa-miR-518a-3p	5.266761	1.367954	0.356809	3.83386	0.000126	0.000848
hsa-miR-1322	2.309011	1.370363	0.544727	2.51569	0.01188	0.037347
hsa-miR-6881-3p	16.22719	1.373691	0.303201	4.530624	5.88E-06	6.43E-05
hsa-miR-199a-5p	530.2113	1.375253	0.164797	8.345159	7.11E-17	1.06E-14
hsa-miR-3147	26.13932	1.377223	0.329703	4.17716	2.95E-05	0.000248
hsa-miR-210-3p	30.18561	1.381657	0.247511	5.582201	2.37E-08	5.95E-07
hsa-miR-424-5p	37.19066	1.383048	0.228249	6.05937	1.37E-09	4.58E-08
hsa-miR-345-3p	12.10461	1.383407	0.281924	4.907016	9.25E-07	1.33E-05
hsa-miR-3139	4.221183	1.385765	0.506232	2.737413	0.006192	0.022263
hsa-miR-6777-5p	17.59924	1.38689	0.271371	5.110684	3.21E-07	5.38E-06
hsa-miR-7106-5p	8.465483	1.388214	0.370933	3.742495	0.000182	0.001153
hsa-miR-5006-3p	36.9788	1.38928	0.279898	4.963518	6.92E-07	1.05E-05
hsa-miR-6739-3p	6.508085	1.390838	0.478843	2.904584	0.003677	0.014611
hsa-miR-7154-3p	56.14248	1.391787	0.2808	4.956512	7.18E-07	1.07E-05
hsa-miR-4515	11.86712	1.401099	0.339544	4.126409	3.68E-05	0.000294
hsa-miR-4311	43.11074	1.405281	0.230817	6.0883	1.14E-09	3.95E-08
hsa-miR-378j	5.582816	1.416563	0.437094	3.240865	0.001192	0.005663
hsa-miR-5708	12.1545	1.417009	0.338418	4.18715	2.82E-05	0.000239
hsa-miR-455-5p	16.61306	1.42558	0.267294	5.333377	9.64E-08	2.02E-06
hsa-miR-6769b-3p	6.047018	1.427217	0.362804	3.933848	8.36E-05	0.000589
hsa-miR-3686	6.845929	1.43008	0.363913	3.929727	8.50E-05	0.000597
hsa-miR-6762-3p	32.43652	1.433591	0.25126	5.705606	1.16E-08	3.09E-07

Table 2 (continued)

GENE ID	baseMean	log2FoldChange	lfcSE	stat	*p* value	padj
hsa-miR-379-5p	240.4924	1.436865	0.215494	6.667783	2.60E-11	1.46E-09
hsa-miR-3127-5p	37.61088	1.440635	0.263277	5.471947	4.45E-08	1.04E-06
hsa-miR-4430	25.08197	1.445465	0.315804	4.577095	4.71E-06	5.60E-05
hsa-miR-6747-3p	12.22378	1.446292	0.335875	4.306042	1.66E-05	0.000153
hsa-miR-2467-3p	7.857414	1.446423	0.380879	3.797589	0.000146	0.000955
hsa-miR-542-3p	30.57727	1.446519	0.248401	5.823316	5.77E-09	1.67E-07
hsa-miR-6873-3p	129.4148	1.449515	0.230424	6.290629	3.16E-10	1.24E-08
hsa-miR-3162-5p	11.20337	1.450855	0.280764	5.167517	2.37E-07	4.25E-06
hsa-miR-1282	9.394666	1.451542	0.33785	4.296414	1.74E-05	0.000159
hsa-miR-4428	4.276177	1.458907	0.520027	2.805447	0.005025	0.018984
hsa-miR-3183	3.095016	1.45933	0.491723	2.967791	0.002999	0.012441
hsa-miR-4512	6.000103	1.46255	0.522237	2.800548	0.005102	0.019118
hsa-miR-4534	10.78325	1.477748	0.322737	4.578803	4.68E-06	5.58E-05
hsa-miR-6813-5p	22.89356	1.486301	0.284006	5.233342	1.66E-07	3.14E-06
hsa-miR-147a	3.621202	1.49903	0.487363	3.075796	0.002099	0.009184
hsa-miR-127-3p	477.2779	1.501043	0.198066	7.578516	3.50E-14	3.63E-12
hsa-miR-5194	9.890997	1.504619	0.346976	4.336373	1.45E-05	0.000137
hsa-miR-6861-3p	5.20444	1.508275	0.437769	3.445369	0.00057	0.002993
hsa-miR-6886-3p	187.0129	1.510457	0.304313	4.963499	6.92E-07	1.05E-05
hsa-miR-511-5p	3.353722	1.515016	0.489762	3.093369	0.001979	0.008806
hsa-miR-3059-5p	18.07627	1.525932	0.385616	3.957128	7.59E-05	0.000544
hsa-miR-4644	86.78652	1.530262	0.232897	6.570544	5.01E-11	2.56E-09
hsa-miR-4307	50.57032	1.531367	0.297505	5.147358	2.64E-07	4.65E-06
hsa-miR-124-5p	8.609369	1.540106	0.34622	4.448345	8.65E-06	9.04E-05
hsa-miR-6767-3p	2.787504	1.541755	0.510019	3.022939	0.002503	0.010726
hsa-miR-5196-3p	8.654876	1.542461	0.293167	5.261372	1.43E-07	2.75E-06
hsa-miR-202-3p	3.432079	1.555386	0.462119	3.365769	0.000763	0.003869
hsa-miR-1227-3p	4.428663	1.563336	0.492704	3.172969	0.001509	0.006912
hsa-miR-6743-5p	3.690666	1.567789	0.449384	3.488749	0.000485	0.002609
hsa-miR-6773-5p	4.033865	1.589988	0.408095	3.896118	9.77E-05	0.000675
hsa-miR-589-3p	33.51418	1.616727	0.311389	5.191987	2.08E-07	3.83E-06
hsa-miR-5092	4.859055	1.622597	0.426577	3.803761	0.000143	0.000934
hsa-miR-3156-5p	3.190753	1.650101	0.552336	2.987491	0.002813	0.011713
hsa-miR-6791-3p	4.769538	1.653496	0.448131	3.689757	0.000224	0.001388
hsa-miR-218-2-3p	9.679162	1.654458	0.417834	3.959607	7.51E-05	0.00054
hsa-miR-6749-3p	6.394308	1.670049	0.394208	4.236469	2.27E-05	0.000197
hsa-miR-12117	2.913759	1.687835	0.632612	2.668041	0.00763	0.026601
hsa-miR-4723-3p	3.201022	1.688363	0.503594	3.352626	0.0008	0.004018
hsa-miR-31-5p	31.48515	1.690555	0.260292	6.494851	8.31E-11	4.11E-09
hsa-miR-3189-5p	3.783917	1.690844	0.605525	2.792362	0.005232	0.019521
hsa-miR-522-3p	4.968542	1.690892	0.436276	3.875739	0.000106	0.000724
hsa-miR-29b-3p	1009.676	1.702567	0.169283	10.05753	8.51E-24	3.54E-21
hsa-miR-127-5p	39.35896	1.712527	0.273668	6.257675	3.91E-10	1.48E-08
hsa-miR-149-5p	5.851922	1.729144	0.379426	4.557261	5.18E-06	5.95E-05
hsa-miR-124-3p	12.95018	1.743358	0.322161	5.411455	6.25E-08	1.40E-06
hsa-miR-6769b-5p	8.57052	1.74815	0.409942	4.264381	2.00E-05	0.000177
hsa-miR-6788-3p	12.5679	1.766022	0.36032	4.901257	9.52E-07	1.36E-05
hsa-miR-587	5.390748	1.770033	0.463001	3.822954	0.000132	0.000867
hsa-miR-6752-5p	8.782962	1.788351	0.388964	4.597727	4.27E-06	5.19E-05

Table 2 (continued)

GENE ID	baseMean	log2FoldChange	lfcSE	stat	*p* value	padj
hsa-miR-299-5p	5.178998	1.79171	0.489477	3.660458	0.000252	0.001539
hsa-miR-6827-3p	3.808142	1.79738	0.535163	3.358564	0.000783	0.003961
hsa-miR-7157-5p	4.127762	1.801984	0.535263	3.366539	0.000761	0.003867
hsa-miR-376c-5p	5.857519	1.803492	0.427206	4.221599	2.43E-05	0.00021
hsa-miR-4999-3p	3.864624	1.804901	0.663216	2.721439	0.0065	0.023207
hsa-miR-6788-5p	6.81227	1.805488	0.37703	4.788711	1.68E-06	2.25E-05
hsa-miR-34a-5p	123.3922	1.818667	0.254935	7.133851	9.76E-13	8.11E-11
hsa-miR-6514-3p	2.395409	1.865027	0.735289	2.536455	0.011198	0.035804
hsa-miR-6798-5p	2.269192	1.877263	0.65084	2.884368	0.003922	0.015319
hsa-miR-6726-3p	3.473671	1.877684	0.600343	3.127686	0.001762	0.007994
hsa-miR-4733-5p	2.649108	1.897048	0.667249	2.843088	0.004468	0.017225
hsa-miR-3117-5p	2.993637	1.944939	0.590292	3.294876	0.000985	0.004781
hsa-miR-137-3p	27.49126	1.986392	0.314283	6.320392	2.61E-10	1.06E-08
pir-39538@piRNAcluster	2.480167	1.993348	0.765007	2.605658	0.00917	0.030635
hsa-miR-12135	5.245336	2.089928	0.458476	4.558425	5.15E-06	5.95E-05
hsa-miR-136-3p	62.86165	2.107252	0.256335	8.220684	2.02E-16	2.80E-14
hsa-miR-376b-5p	5.570574	2.123875	0.408853	5.194721	2.05E-07	3.80E-06
hsa-miR-6830-5p	4.443684	2.125405	0.581979	3.652031	0.00026	0.001572
hsa-miR-6882-3p	17.77538	2.146727	0.373675	5.744901	9.20E-09	2.48E-07
hsa-miR-196a-5p	113.4209	2.353928	0.266661	8.827406	1.07E-18	2.02E-16
hsa-miR-2682-3p	22.13441	2.358546	0.333433	7.07353	1.51E-12	1.16E-10
hsa-miR-1255b-2-3p	3.393024	2.439419	0.579125	4.212253	2.53E-05	0.000216
hsa-miR-34c-5p	14.96931	2.560805	0.383255	6.681733	2.36E-11	1.36E-09
hsa-miR-31-3p	7.834028	2.626093	0.434614	6.042353	1.52E-09	4.97E-08
hsa-miR-708-5p	9.47941	2.701465	0.393973	6.856977	7.03E-12	4.57E-10
hsa-miR-129-5p	12.76503	2.750368	0.387631	7.095322	1.29E-12	1.03E-10
hsa-miR-376a-5p	4.366374	2.795203	0.608708	4.592028	4.39E-06	5.30E-05
hsa-miR-218-5p	71.5591	2.806262	0.260961	10.75358	5.70E-27	2.96E-24
hsa-miR-125b-5p	751.1633	3.039909	0.235812	12.89124	5.04E-38	5.24E-35
hsa-miR-4455	9.562778	4.195321	0.774577	5.416274	6.09E-08	1.37E-06
hsa-miR-9-5p	940.2017	4.540712	0.318006	14.27869	2.97E-46	6.17E-43

relative expression levels of key small RNAs between the two groups: hsa-miR-548ah-5p (control group vs. positive group = 1.00 ± 0.1712 vs. 0.3362 ± 0.04602, $p = 0.0002$), pir-33,161 (0.9998 ± 0.1777 vs. 0.4232 ± 0.06473, $p = 0.0026$), pir-36,340 (1.000 ± 0.1616 vs. 0.3621 ± 0.04354, $p = 0.0002$), hsa-miR-484 (1.000 ± 0.1135 vs. 0.3644 ± 0.03818, $P < 0.0001$), hsa-miR-4282 (1.000 ± 0.1233 vs. 0.3864 ± 0.04173, $P < 0.0001$), and hsa-miR-6853-3p (0.9997 ± 0.1558 vs. 0.3448 ± 0.03808, $P\ p < 0.0001$). Compared with those in the control group, the relative expression levels of six key genes in the positive control group were significantly lower (Fig. 4B-G). This finding is consistent with the gene sequencing results.

GO and KEGG analysis

The software predicts target genes for six key small RNAs and conducts GO analysis. These 6 key small RNAs (hsa-miR-548ah-5p, pir-33161, pir-36340, hsa-miR-484, hsa-miR-4282, and hsa-miR-6853-3p) target gene numbers were 570, 9111, 38, 1683, 163 and 496, respectively. (A list of key small RNAs and their target genes is shown in supplementary material 1)The GO analysis revealed that the biological processes with the most significant functional clustering of these target genes included "positive regulation of chemotaxis", "positive regulation of chemotaxis", "positive regulation of receptor activity", and "T-cell-mediated immune regulation". The cellular components included

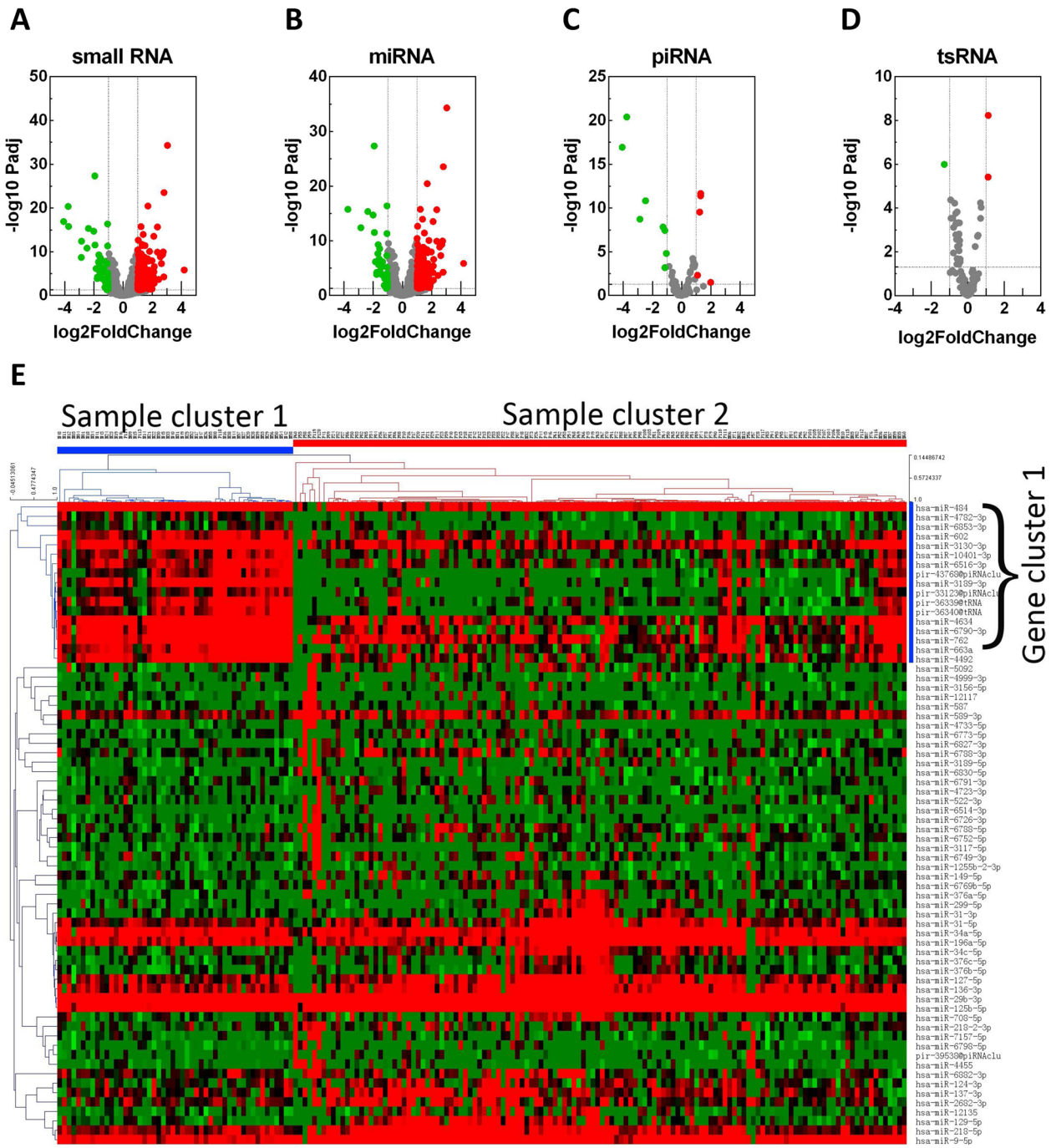

Fig. 3 Volcano plot and cluster heatmaps for gene expression differential analysis. **A**-**D** Volcano plot showing the difference in gene expression between the positive group and the positive group. Red: Gene expression upregulated; Green: Gene expression downregulated. **A** All small RNAs, **B** miRNAs, **C** piRNAs, **D** tsRNAs. **E** The cluster heatmap of genes with significant differences in expression. Red: genes whose expression was upregulated ; green: genes whose expression was downregulated. The samples were significantly clustered into 2 clusters. Genes cluster into different clusters. The expression of gene cluster 1 was up-regulated in sample cluster 1 and down-regulated in sample cluster 2

the "cis-Golgi network", "membrane region", and "membrane microdomain". The molecular functions included "potassium channel activity", "guanyl ribonucleotide binding", and "purine ribonucleoside binding" (Fig. 5A-C).

KEGG signaling pathway analysis revealed that the most significantly enriched signaling pathways included "cytokine cytokine receptor interaction", "chemokine signaling pathway", "apoptosis", and "TGF-β signaling

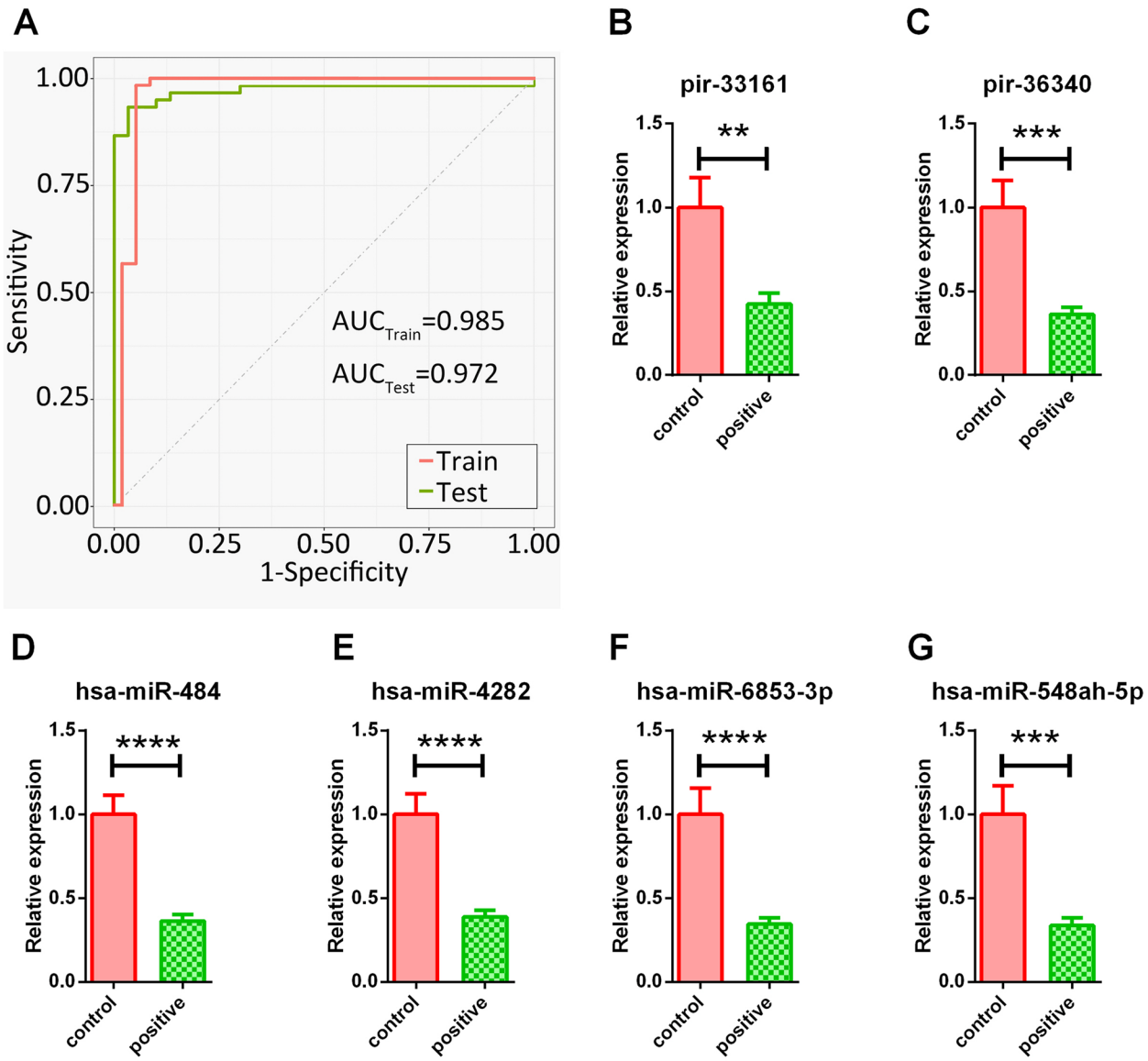

Fig. 4 ROC curves of machine learning models and RT-qPCR results of key genes. **A** ROC curve of the training set and test set. AUC (train) = 0.985 (95% CI = 0.948-1); AUC (test) = 0.972 (95% CI = 0.882–0.995)). **B**-**G** The relative expression levels of small RNAs included in the machine learning model. The qPCR reference gene used was cel-miR-39-3p. Compared to those in the control group (normalized to 1), the relative expression levels of six target genes were decreased: miR-548ah-5p (0.3362 ± 0.04602, $P = 0.0002$), pir-33161 (0.4232 ± 0.06473, $P = 0.0026$), pir-36340 (0.3621 ± 0.04354, $P = 0.0002$), miR-484 (0.3644 ± 0.03818, $P < 0.0001$), miR-4282 (0.3864 ± 0.04173, $P < 0.0001$), and miR-6853-3p (0.3448 ± 0.03808, $P < 0.0001$)

pathway" (Fig. 5D). The top two signaling pathways, "cytokine cytokine receptor interaction" and "chemokine signaling pathway", are closely related to tumor development. Therefore, we plotted their relationships with target genes as a regulatory network diagram (Fig. 6). The regulatory network diagram of small RNAs and their target genes revealed that these key small RNAs regulate genes involved in 84 "cytokine cytokine receptor interaction signaling pathways" and 56 "chemokine signaling pathways". Among these genes, 26 were involved in both the "cytokine cytokine receptor interaction signaling pathway" and the "chemokine signaling pathway". Interestingly, all 26 target genes are chemokines or chemokine receptors.

Discussion

MiRNAs are noncoding RNAs 17–24 nt in length that can mediate posttranscriptional silencing of genes and are associated with cell proliferation, differentiation, migration, disease occurrence, and disease progression.

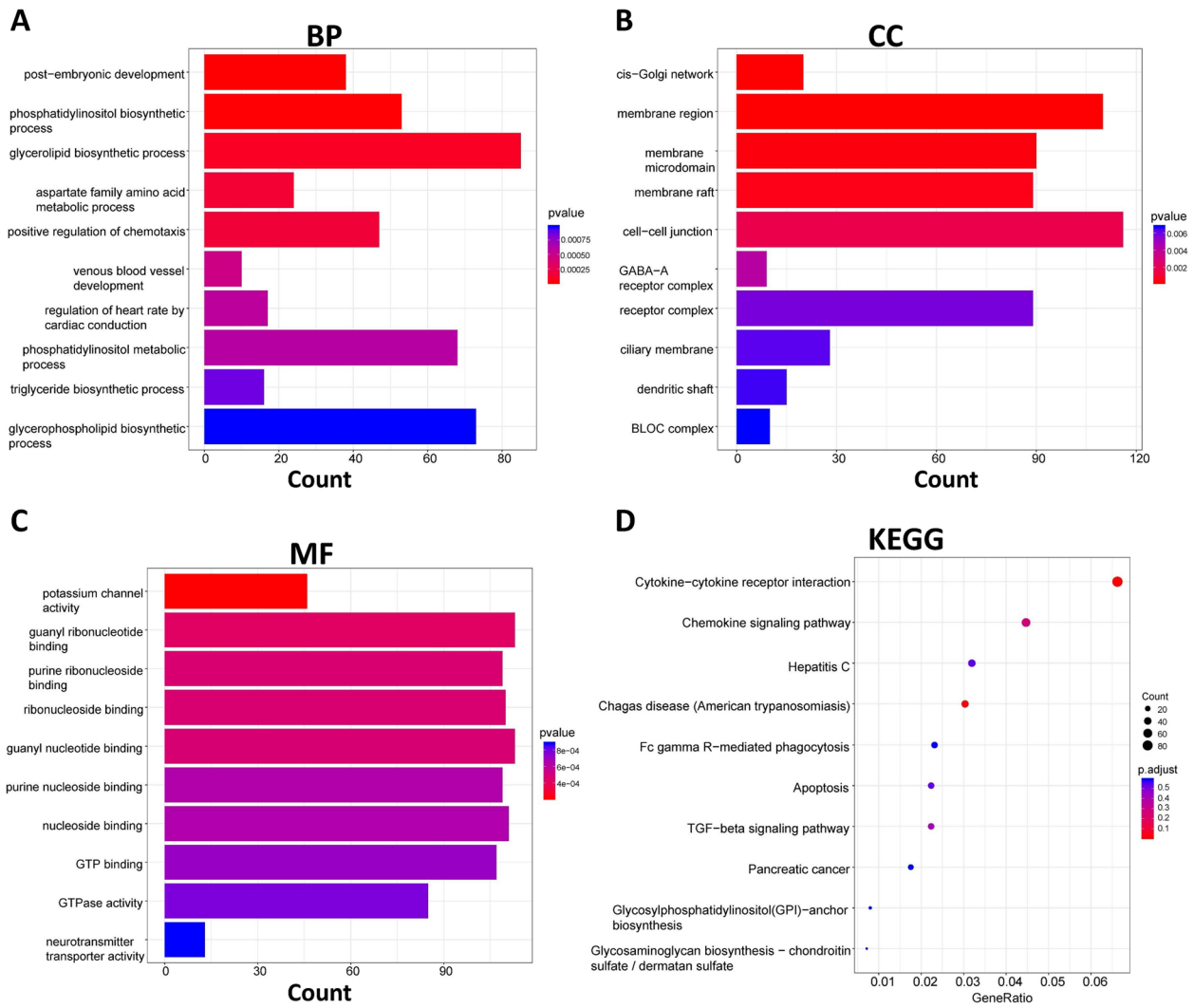

Fig. 5 GO and KEGG analyses of small RNAs showing significant differences between groups. **A**-**C** Gene function analysis (GO). The image shows the top 10 items in the prominent GO term. The ordinate Y-axis represents the GO term entry name, and the coordinate X-axis represents the Count (Number of genes clustered to the item). a: Biological process (BP). The top 20 most significant biological process functional clusters of target genes for the six small RNAs included "positive regulation of chemotaxis", "regulation of positive chemotaxis", "positive regulation of receptor activity", and "T-cell-mediated immune regulation". b: Cellular component (CC). The target gene functions of the six key small RNAs were mainly related to cell components such as the "cis Golgi network", "memory microdomain", and "memory microdomain". c: Molecular function (MF). The target gene functions of the six key small RNAs were mainly clustered in molecular functions such as "potassium channel activity", "guanyl ribonucleotide binding", and "purine ribonucleotide binding". **D** KEGG signaling pathway enrichment analysis. The enrichment degree of each pathway was determined by the GeneRatio (Rich factor), *P* value and number of genes enriched in the pathway. The "GeneRatio (Rich factor)" refers to the ratio of the number of key small RNA-regulated target genes enriched in the pathway to the number of annotated genes. The greater the GeneRatio is, the greater the degree of enrichment. The target gene functions of the six small RNAs were mainly enriched in two signaling pathways: "cytokine cytokine receptor interaction" and "chemokine signaling pathway"

They have been proven to be closely related to various diseases, including cancer, cardiovascular disease, and immune system disease [26, 27]. They participate in the occurrence of breast cancer mainly by regulating cell apoptosis, autophagy and epithelial mesenchymal transformation (EMT) [28].

The proportion of miRNAs in exosomes is greater than that in cells that excrete exosomes, indicating that miRNAs are specifically enriched in exosomes [29]. Therefore, the role of miRNAs in exosomes is receiving increasing attention. We extracted exosomes from the plasma of the study subjects and detected 2656 miRNAs and found that 309 miRNAs were significantly

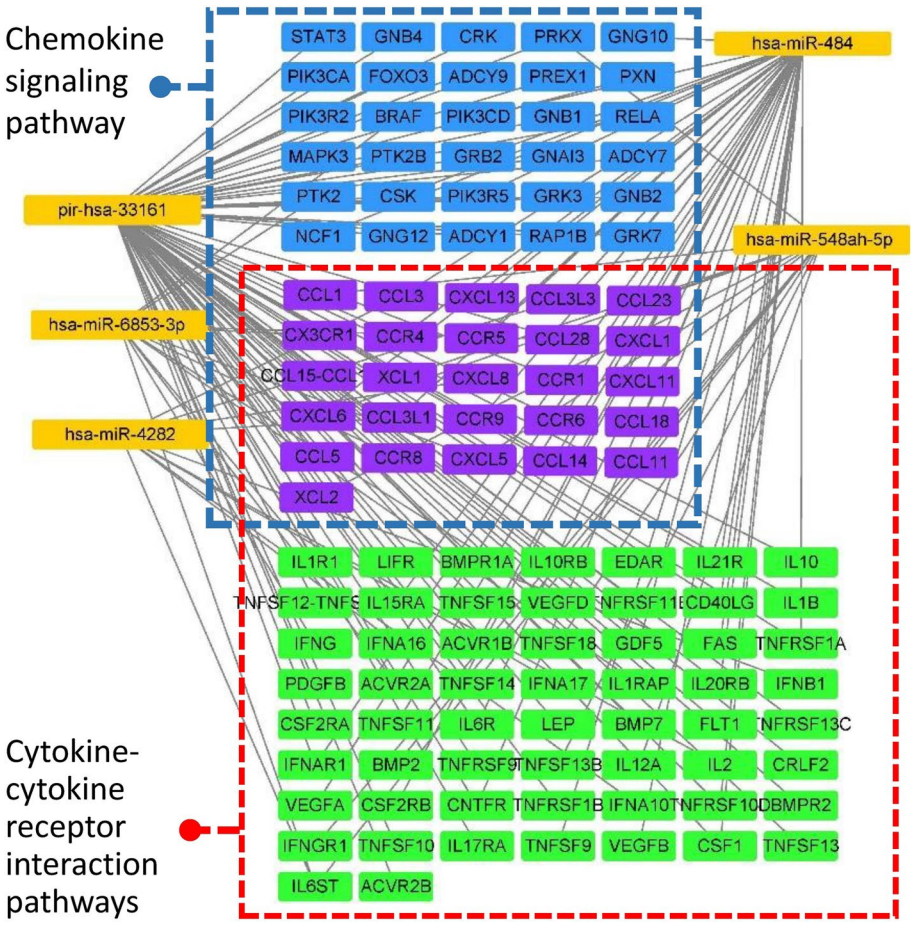

Fig. 6 PirRNA/miRNA–mRNA regulatory network. A regulatory network diagram of pirRNAs/miRNAs and mRNAs was constructed using Cytoscape 3.7.0 software. Interaction networks between small RNAs and genes on the "cytokine - cytokine receptor interaction Signaling pathway" and the "chemokine signaling pathway"

differentially expressed between the groups (fold change > 2, $P < 0.05$, padj/FDR < 0.05). Through machine learning model analysis and qPCR experimental verification, we ultimately identified four key miRNAs (miR-484, miR-548ah-5p, miR-4282, and miR-6853-3p). Among the four key miRNAs, miR-484 has been extensively studied. Some studies suggest that miR-484 is a highly promising biomarker for health and disease, playing an important role in the development and regression of different diseases. According to reports, miR-484 is a key regulatory factor for common cancer and noncancer diseases [30]. It has been found to be involved in the EMT of colorectal cancer [31], as well as in the development of liver cancer [32, 33], and is therefore considered a key regulator of the progression from precancerous lesions to cancer. The studies of miR-484 and breast cancer have mostly focused on drug resistance and prognosis [30]. However, some studies have shown that the expression of miR-484 in the plasma of breast cancer patients is abnormal [34, 35]. At present, there have been no reports of miR-484 in human plasma exosomes.

MiR-4282 is another key miRNA found in our study. There is little research on its relationship with cancer. Two studies have shown that miR-4282 is associated with breast cancer: one study revealed that it inhibits the proliferation, migration and invasion of breast cancer cells by reducing the expression of centromere protein F [36], and another study revealed that miR-4282 is downregulated in human breast cancer tissue and that miR-4282 inhibits the proliferation, invasion and metastasis of human breast cancer cells [37]. There are no reports that miR-4282 from an exocrine source is associated with breast cancer.

MiR-548ah-5p and miR-6853-3p are also the key miRNAs we identified, but their relationships with breast cancer have not yet been reported.

This study also detected 728 piRNAs in plasma exosomes. There were 13 piRNAs whose expression differed significantly between the two groups (fold change > 2, $P < 0.05$, padj/FDR < 0.05). After machine learning model analysis and qPCR experimental verification, two key piRNAs were ultimately selected: piR-36,340 and piR-33,161. PiRNAs are initially expressed in the reproductive system, and their relationship with the occurrence and development of breast cancer has gradually become a research hotspot in recent years. Many studies have shown that the expression of piRNAs is associated with various diseases, including cancer. It participates in the processes of cancer cell proliferation, apoptosis, invasion, and migration [38]. Some studies have shown that piRNAs participate in the occurrence of human breast cancer through epigenetic mechanisms, such as piR-021285 [13], piR-823 [15] and piR-932 [14]. The relationship between piR-36,340 or piR-33,161 (the two exosome piRNAs screened in this study) and breast cancer has not been reported. However, we observed significant differences in piR-36,340 and piR-33,161 in exosomes between the two groups, not only in the fully sequenced samples but also in the qPCR validation samples. Therefore, the role and mechanism of piR-36,340 and piR-33,161 in the development of breast cancer are worthy of further study.

In addition to miRNAs and piRNAs, we also detected 154 tsRNAs. There were three tsRNAs with significant differences between the two groups, namely, tsrna-3017a-valtac, tsrna-5004c-glygcc/glyccc, and tsrna-5011a-progg/progg/procgg (fold change > 2, $P < 0.05$, padj/FDR < 0.05). However, these genes were not included in the machine learning model, so we did not use validation samples for qPCR detection. Plasma small RNA sequencing revealed that a tsRNA (Tsrna-26576) mediates the tumorigenesis of breast cancer [39]. Some studies have also shown that specific TSRNAs in plasma can be used as biomarkers for the diagnosis and prognosis of breast cancer [40].

In this study, we conducted GO and KEGG analyses on six key small RNAs. The results showed that the functions of their target genes were mainly clustered in biological processes such as "chemotaxis regulation" and "receptor activity regulation". The most enriched signaling pathways were "cytokine cytokine receptor interaction" and "chemokine signaling pathway". By aggregating the target genes of these two signaling pathways, it was found that 26 genes are involved in these two signaling pathways, and all of them are chemokines or chemokine receptors.

Chemokines are small-molecule peptides produced by immune cells with a molecular weight of 8–10 kDa and belong to the cytokine superfamily. Chemokine receptors belong to the seven transmembrane G protein-coupled receptor superfamily. When chemokines bind to specific receptors, they can aggregate immune cells in the tumor microenvironment, thereby regulating immune surveillance, angiogenesis, invasion, and metastasis. Chemokines and chemokine receptors participate in the proliferation, differentiation, invasion and metastasis of breast cancer through a variety of mechanisms. Some studies have shown that CCL3 may be secreted by breast cancer cells and attract myeloid suppressor cells (MDSCs) to the tumor microenvironment, thereby activating the PI3K-Akt-mTOR pathway, leading to EMT and promoting the migration and invasion of breast cancer cells [41]. Some studies also suggest that CCR3 interacts with CCL5 and that the CCL5-CCR3 axis promotes the progression of breast cancer [42]. CCL2/CCR2 is the main chemokine pathway related to the growth of breast cancer. They were first known to regulate the progression of breast cancer through macrophage-dependent mechanisms. Subsequent research focused on regulating the recruitment of immune cells. For example, studies have shown that CCR2 signal transduction in breast cancer cells promotes tumor growth and invasion by promoting CCL2 and inhibiting the effect of CD154 on angiogenesis and the immune microenvironment [43]. Recent studies have shown that gene knockout of matrix CCL2 or epithelial CCR2 can significantly inhibit the growth of breast tumors; CCL2/CCR2 signaling mediates the growth of breast cancer cells through the PKC and SRC pathways [44]. In addition, CCL2/CCR2 chemokine signals regulate the growth and migration of breast cancer cells through Smad3 protein- and p42/44 mitogen activated protein kinase (MAPK)-dependent mechanisms [45]. Other studies have shown that tumor-associated macrophages promote the EMT of triple-negative breast cancer through CCL2 signal transduction [46]. CXCR4 is a chemokine that plays an important role in the development of breast cancer. CXCR4 promotes the growth of breast cancer by promoting angiogenesis, stimulating typical signaling pathways related to cell proliferation, and mediating immune cell aggregation [47].

There are few reports of small RNAs acting on breast cancer through chemokines, but there have also been attempts. For example, studies have shown that miR-494 mimics significantly decrease the protein expression level of chemokine (CXC motif) receptor 4 (CXCR4) in breast cancer cells, while miR-494 inhibitors significantly increase the expression level/β-catenin signaling pathway, which targets CXCR4 and inhibits the development of breast cancer in vitro [48]. Another study revealed that miR-139-mediated CXCR4/p-Akt signal inhibition affects the EMT process of breast cells, thus participating in the process of malignant transformation in breast

cancer [49]. Other studies have shown that the miRNA let-7a inhibits the migration and invasion of breast cancer cells by downregulating CCR7 [50]. miRNA-101 inhibits the growth and metastasis of breast cancer by targeting CXR7 [51]. In this study, these key exosome small RNAs formed a regulatory network with genes involved in the chemokine signaling pathway. At present, there have been no reports on the relationships between these key exosome small RNAs and chemokine signaling pathways. Therefore, the role of these key exosome small RNA-mediated chemokine signaling pathways in the pathogenesis of breast cancer deserves further exploration.

Breast nodules or breast masses are often the first symptom of breast cancer. It is of great social value and scientific significance to understand the formation of breast nodules and the mechanism of their malignant transformation, identify molecular markers of early precancerous lesions, establish risk assessment models, predict and screen the risk of high-risk individuals, and develop targeted interventions to achieve accurate prevention and treatment of high-risk groups. We used breast cancer screening population samples to conduct full sequencing of small RNAs in plasma exosomes and used a machine learning algorithm to analyze small RNAs with significant differences in expression between the two groups. Six key small RNAs were ultimately screened, namely, piR-36,340, piR-33,161, miR-484, miR-548ah-5p, miR-4282, and miR-6853-3p. The AUC values of these key small RNAs in both the machine learning model training and testing sets exceeded 0.97. Subsequently, we used validation samples for qPCR-based quantitative detection and detected significant differences in the expression of these genes between the two groups ($P < 0.05$). The samples in this study were obtained from a population with a positive initial screening for breast cancer, which is the key time point of early breast nodule lesion detection, and the identified molecular markers are more valuable for prevention. Through target gene prediction and functional enrichment analysis, it was found that the functions of the target genes of these small RNAs were enriched mainly in the chemokine signaling pathway. These results suggest that the chemokine signaling pathway may be involved in the pathogenesis of breast cancer. However, whether small RNAs mediate the chemokine signaling pathway to play a role in the development of breast cancer and whether this role is accomplished through the delivery of exosomes still need to be confirmed through mechanistic research.

In summary, whether Small RNA mediates chemokine signaling pathway through exosome delivery to play a role in the occurrence of breast cancer still needs further research. The predictive effect of this machine learning model also needs to be verified with more sample sizes and multiple populations. We will continue to study along these lines. In any case, the value of this combination of exosome key Small RNA as a molecular marker for early screening of breast cancer is still worth looking forward to.

Abbreviations

AUC	Area under the ROC curve
BI-RADS	Breast Imaging Reporting and Data System
EMT	Epithelial mesenchymal transformation
Evs	Extracellular vesicles
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
miRNAs	MicroRNAs
piRNAs	PIWI-interacting RNAs, P-element-induced wimpy tests of interacting RNAs
ROC	Receiver operating characteristic
snoRNAs	Small nucleolar RNAs
tRNAs	Transfer RNA

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-024-13173-x.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements
Guangzhou Epibiotek Co. Ltd. provided exosome small RNA sequencing and bioinformatics analysis services for this project. Thank you.

Authors' contributions
Ji Peng, Lin Lei, Junluan Mo, Wenxu Hong and Xiongshun Liang were responsible for the organization and implementation of population surveys, hospital physical examinations and sample collection; Junluan Mo and Xiongshun Liang were responsible for genetic testing of blood samples; Junluan Mo was responsible for data analysis and article writing; Wenxu Hong and Jiye Yin were responsible for the guidance and revision of the article; Xi Li, Zhaoqian Liu, and Honghao Zhou assisted in guiding the data analysis and article writing.

Funding
This study was funded by the National Natural Science Foundation of China (82073943 and 8237131447 to JYY), the Furong Laboratory Science and Technology Projects (2023SK2083-2 to JYY), the Shenzhen Science and Technology Program (No. JCYJ20230807120859030), the Sanming Project of Medicine in Shenzhen (No. SZSM201811057), and the Guangdong Medical Science and Technology Research Fund (No. B2021240).

Data availability
The sncRNA-seq data have been deposited in NCBI's GEO under accession number GSE270497.

Declarations

Ethics approval and consent to participate
This study protocol was reviewed and approved by the Shenzhen Center for Chronic Disease Control, approval number (SZCCC-2020-018-06-PJ). The research subjects are sourced from the Urban Cancer Early Diagnosis and Treatment Project, which was approved by the Ethics Committee of Cancer Institute and Hospital, Chinese Academy of Medical Sciences. The approval number is 15–070/997. All participants signed informed consent forms.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Clinical Pharmacology, Xiangya Hospital, Central South University, Changsha 410078, P. R. China. [2]Shenzhen Center for Chronic Disease Control, Shenzhen 518020, P. R. China. [3]Institute of Clinical Pharmacology, Central South University, Hunan Key Laboratory of Pharmacogenetics, Changsha, P.R. China. [4]National Clinical Research Center for Geriatric Disorders, Changsha, P.R. China.

Received: 3 June 2024 Accepted: 7 November 2024
Published online: 21 November 2024

References

1. Bray F, Laversanne M, Sung H. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Cancer J Clin. 2024;74(3):229–63.
2. Zhang L, Yu D. Exosomes in cancer development, metastasis, and immunity. Biochim Biophys Acta Rev Cancer. 2019;1871(2):455–68.
3. Lu M, Shao W, Xing H, Huang Y. Extracellular vesicle-based nucleic acid delivery. Interdisciplinary Med. 2023;1(2):e20220007.
4. Jayaseelan VP. Emerging role of exosomes as promising diagnostic tool for cancer. Cancer Gene Ther. 2020;27(6):395–8.
5. Pakravan K, Babashah S, Sadeghizadeh M, Mowla SJ, Mossahebi-Mohammadi M, Ataei F, Dana N, Javan M. MicroRNA-100 shuttled by mesenchymal stem cell-derived exosomes suppresses in vitro angiogenesis through modulating the mTOR/HIF-1α/VEGF signaling axis in breast cancer cells. Cell Oncol (Dordr). 2017;40(5):457–70.
6. Lee JK, Park SR, Jung BK, Jeon YK, Lee YS, Kim MK, Kim YG, Jang JY, Kim CW. Exosomes derived from mesenchymal stem cells suppress angiogenesis by down-regulating VEGF expression in breast cancer cells. PLoS ONE. 2013;8(12):e84256.
7. Baroni S, Romero-Cordoba S, Plantamura I, Dugo M, D'Ippolito E, Cataldo A, Cosentino G, Angeloni V, Rossini A, Daidone MG, et al. Exosome-mediated delivery of miR-9 induces cancer-associated fibroblast-like properties in human breast fibroblasts. Cell Death Dis. 2016;7(7):e2312.
8. Gernapudi R, Yao Y, Zhang Y, Wolfson B, Roy S, Duru N, Eades G, Yang P, Zhou Q. Targeting exosomes from preadipocytes inhibits preadipocyte to cancer stem cell signaling in early-stage breast cancer. Breast Cancer Res Treat. 2015;150(3):685–95.
9. Uen Y, Wang JW, Wang C, Jhang Y, Chung JY, Tseng T, Sheu M, Lee S. Mining of potential microRNAs with clinical correlation - regulation of syndecan-1 expression by mir-122-5p altered mobility of breast cancer cells and possible correlation with liver injury. Oncotarget. 2018;9(46):28165–75.
10. Kong X, Zhang J, Li J, Shao J, Fang L. MiR-130a-3p inhibits migration and invasion by regulating RAB5B in human breast cancer stem cell-like cells. Biochem Biophys Res Commun. 2018;501(2):486–93.
11. Fong MY, Zhou W, Liu L, Alontaga AY, Chandra M, Ashby J, Chow A, O'Connor ST, Li S, Chin AR, et al. Breast-cancer-secreted miR-122 reprograms glucose metabolism in premetastatic niche to promote metastasis. Nat Cell Biol. 2015;17(2):183–94.
12. Liu Y, Dou M, Song X, Dong Y, Liu S, Liu H, Tao J, Li W, Yin X, Xu W. The emerging role of the piRNA/piwi complex in cancer. Mol Cancer. 2019;18(1):123.
13. Fu A, Jacobs DI, Hoffman AE, Zheng T, Zhu Y. PIWI-interacting RNA 021285 is involved in breast tumorigenesis possibly by remodeling the cancer epigenome. Carcinogenesis. 2015;36(10):1094–102.
14. Zhang H, Ren Y, Xu H, Pang D, Duan C, Liu C. The expression of stem cell protein Piwil2 and piR-932 in breast cancer. Surg Oncol. 2013;22(4):217–23.
15. Ding X, Li Y, Lü J, Zhao Q, Guo Y, Lu Z, Ma W, Liu P, Pestell RG, Liang C, et al. piRNA-823 is involved in Cancer Stem Cell Regulation through altering DNA methylation in Association with luminal breast Cancer. Front Cell Dev Biol. 2021;9:641052.
16. Wang Y, Weng Q, Ge J, Zhang X, Guo J, Ye G. tRNA-derived small RNAs: mechanisms and potential roles in cancers. Genes Dis. 2022;9(6):1431–42.
17. Mendelson EB. ACR BIRADS® Ultrasound. In: ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System, vol. 2013: American College of Radiology, Reston; 2013.
18. Tian Y, Ma L, Gong M, Su G, Zhu S, Zhang W, Wang S, Li Z, Chen C, Li L, et al. Protein profiling and sizing of Extracellular vesicles from Colorectal Cancer patients via Flow Cytometry. ACS Nano. 2018;12(1):671–80.
19. Tian Y, Gong M, Hu Y, Liu H, Zhang W, Zhang M, Hu X, Aubert D, Zhu S, Wu L, et al. Quality and efficiency assessment of six extracellular vesicle isolation methods by nano-flow cytometry. J Extracell Vesicles. 2020;9(1):1697028.
20. Théry C, Zitvogel L, Amigorena S. Exosomes: composition, biogenesis and function. Nat Rev Immunol. 2002;2(8):569–79.
21. Williams S, Fernandez-Rhodes M, Law A, Peacock B, Lewis MP, Davies OG. Comparison of extracellular vesicle isolation processes for therapeutic applications. J Tissue Eng. 2023;14:20417314231174609.
22. Zhang L, Ma W, Gan X, Chen W, Guo J, Cui Y, Wang T. Adequate enrichment of extracellular vesicles in laboratory medicine. Interdisciplinary Med. 2023;1(3):e20220003.
23. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25–9.
24. Draghici S, Khatri P, Tarca AL, Amin K, Done A, Voichita C, Georgescu C, Romero R. A systems biology approach for pathway level analysis. Genome Res. 2007;17(10):1537–45.
25. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504.
26. Takahashi RU, Miyazaki H, Ochiya T. The roles of MicroRNAs in breast Cancer. Cancers (Basel). 2015;7(2):598–616.
27. Loh HY, Norman BP, Lai KS, Rahman N. The Regulatory role of MicroRNAs in breast Cancer. Int J Mol Sci. 2019;20(19):4940.
28. Ghafouri-Fard S, Khanbabapour Sasi A, Abak A, Shoorei H, Khoshkar A, Taheri M. Contribution of miRNAs in the pathogenesis of breast Cancer. Front Oncol. 2021;11:768949.
29. Goldie BJ, Dun MD, Lin M, Smith ND, Verrills NM, Dayas CV, Cairns MJ. Activity-associated miRNA are packaged in Map1b-enriched exosomes released from depolarized neurons. Nucleic Acids Res. 2014;42(14):9195–208.
30. Jia YZ, Liu J, Wang GQ, Song ZF. miR-484: a potential biomarker in Health and Disease. Front Oncol. 2022;12:830420.
31. Li Y, Wang W, Wu M, Zhu P, Zhou Z, Gong Y, Gu Y. LncRNA LINC01315 silencing modulates cancer stem cell properties and epithelial-to-mesenchymal transition in colorectal cancer via miR-484/DLK1 axis. Cell Cycle (Georgetown Tex). 2022;21(8):851–73.
32. Wang S, Wang W, Han X, Wang Y, Ge Y, Tan Z. Dysregulation of miR484-TUSC5 axis takes part in the progression of hepatocellular carcinoma. J BioChem. 2019;166(3):271–9.
33. Yang Y, Lin X, Lu X, Luo G, Zeng T, Tang J, Jiang F, Li L, Cui X, Huang W, et al. Interferon-microRNA signalling drives liver precancerous lesion formation and hepatocarcinogenesis. Gut. 2016;65(7):1186–201.
34. Holubekova V, Kolkova Z, Grendar M, Brany D, Dvorska D, Stastny I. Pathway analysis of selected circulating miRNAs in plasma of breast Cancer patients: a preliminary study. Int J Mol Sci. 2020;21(19):7288.
35. Zearo S, Kim E, Zhu Y, Zhao JT, Sidhu SB, Robinson BG, Soon P. MicroRNA-484 is more highly expressed in serum of early breast cancer patients compared to healthy volunteers. BMC Cancer. 2014;14:200.
36. Zhou MF, Wang W, Wang L, Tan JD. LINC00536 knockdown inhibits breast cancer cells proliferation, invasion, and migration through regulation of the miR-4282/centromere protein F axis. Kaohsiung J Med Sci. 2022;38(11):1037–47.
37. Zhao J, Jiang GQ. MiR-4282 inhibits proliferation, invasion and metastasis of human breast cancer by targeting Myc. Eur Rev Med Pharmacol Sci. 2018;22(24):8763–71.
38. Yu Y, Xiao J, Hann SS. The emerging roles of PIWI-interacting RNA in human cancers. Cancer Manag Res. 2019;11:5895–909.
39. Zhou J, Wan F, Wang Y, Long J, Zhu X. Small RNA sequencing reveals a novel tsRNA-26576 mediating tumorigenesis of breast cancer. Cancer Manag Res. 2019;11:3945–56.

40. Wang J, Ma G, Ge H, Han X, Mao X, Wang X, Veeramootoo JS, Xia T, Liu X, Wang S. Circulating tRNA-derived small RNAs (tsRNAs) signature for the diagnosis and prognosis of breast cancer. NPJ Breast Cancer. 2021;7(1):4.
41. Luo A, Meng M. Myeloid-derived suppressor cells recruited by chemokine (C-C Motif) Ligand 3 promote the progression of breast Cancer via Phosphoinositide 3-Kinase-protein kinase B-Mammalian target of Rapamycin Signaling. J Breast Cancer. 2020;23(2):141–61.
42. Yamaguchi M, Takagi K, Narita K, Miki Y. Stromal CCL5 promotes breast Cancer progression by interacting with CCR3 in Tumor cells. Int J Mol Sci. 2021;22(4):1918–30.
43. Brummer G, Fang W, Smart C, Zinda B, Alissa N, Berkland C, Miller D, Cheng N. CCR2 signaling in breast carcinoma cells promotes tumor growth and invasion by promoting CCL2 and suppressing CD154 effects on the angiogenic and immune microenvironments. Oncogene. 2020;39(11):2275–89.
44. Yao M, Fang W, Smart C, Hu Q, Huang S, Alvarez N. CCR2 chemokine receptors enhance growth and cell-cycle progression of breast Cancer cells through SRC and PKC activation. Mol Cancer Res. 2019;17(2):604–17.
45. Fang WB, Jokar I, Zou A, Lambert D, Dendukuri P, Cheng N. CCL2/CCR2 chemokine signaling coordinates survival and motility of breast cancer cells through Smad3 protein- and p42/44 mitogen-activated protein kinase (MAPK)-dependent mechanisms. J Biol Chem. 2012;287(43):36593–608.
46. Chen X, Yang M, Yin J, Li P, Zeng S, Zheng G, He Z, Liu H, Wang Q, Zhang F, et al. Tumor-associated macrophages promote epithelial-mesenchymal transition and the cancer stem cell properties in triple-negative breast cancer through CCL2/AKT/β-catenin signaling. Cell Commun Signal. 2022;20(1):92.
47. Xu C, Zhao H, Chen H, Yao Q. CXCR4 in breast cancer: oncogenic role and therapeutic targeting. Drug Des Devel Ther. 2015;9:4953–64.
48. Song L, Liu D, Wang B, He J, Zhang S, Dai Z, Ma X, Wang X. miR-494 suppresses the progression of breast cancer in vitro by targeting CXCR4 through the Wnt/β-catenin signaling pathway. Oncol Rep. 2015;34(1):525–31.
49. Cheng CW, Liao WL, Chen PM, Yu JC, Shiau HP, Hsieh YH. MiR-139 modulates Cancer Stem cell function of human breast Cancer through Targeting CXCR4. Cancers (Basel). 2021;13(11):2582.
50. Kim SJ, Shin JY, Lee KD, Bae YK, Sung KW, Nam SJ, Chun KH. MicroRNA let-7a suppresses breast cancer cell migration and invasion through downregulation of C-C chemokine receptor type 7. Breast Cancer Res. 2012;14(1):R14.
51. Li JT, Jia LT, Liu NN, Zhu XS, Liu QQ, Wang XL, Yu F, Liu YL, Yang AG, Gao CF. MiRNA-101 inhibits breast cancer growth and metastasis by targeting CX chemokine receptor 7. Oncotarget. 2015;6(31):30818–30.

Publisher's Note