

Plasma Extracellular Vesicle Long RNA Profiles in the Diagnosis and Prediction of Treatment Response for Breast Cancer

Yonghui Su

Fudan University Shanghai Cancer Center <https://orcid.org/0000-0003-0865-8862>

Jingjing Zhao

Fudan University Shanghai Cancer Center

Rong Guo

Fudan University Shanghai Cancer Center

Hongyan Lai

Fudan University Shanghai Cancer Center

Weiru Chi

Fudan University Shanghai Cancer Center

Yuchen Li

Fudan University Shanghai Cancer Center

Jia Wang

Fudan University Shanghai Cancer Center

Zhen Wang

Fudan University Shanghai Cancer Center

Lun Li

Fudan University Shanghai Cancer Center

Yuting Sang

Fudan University Shanghai Cancer Center

Jianjing Hou

Fudan University Shanghai Cancer Center

Jingyan Xue

Fudan University Shanghai Cancer Center

Zhimin Shao

Fudan University Shanghai Cancer Center

Yayun Chi

Fudan University Shanghai Cancer Center

Shenglin Huang

Fudan University Shanghai Cancer Center

Jiong Wu (✉ wujiong_lab@163.com)

Research

Keywords: Breast cancer; extracellular vesicle; diagnosis; treatment response

Posted Date: November 20th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-111007/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at npj Breast Cancer on December 1st, 2021. See the published version at <https://doi.org/10.1038/s41523-021-00356-z>.

Abstract

Background: The utility of extracellular vesicle long RNAs (exLRs) as noninvasive biomarkers in breast cancer remains elusive. The purpose of this study was to explore the potential of exLRs as clinically actionable biomarkers for breast cancer diagnosis, classification, and neoadjuvant therapy efficacy prediction.

Methods: One hundred and seventy-two participants, including 112 breast cancer patients, 19 benign patients and 41 healthy controls, were enrolled in this case-control study. The exLR profile of the plasma samples was analyzed by exLR sequencing. The d-signature was identified using a support vector machine algorithm with a training cohort (n=120) and was validated using an internal validation cohort (n=52). Treatment efficacy prediction was conducted with 48 patients who received neoadjuvant chemotherapy.

Results: We constructed a breast cancer diagnostic signature that showed high accuracy with an area under the curve (AUC) of 0.960 in the training cohort and 0.900 in the validation cohort. The signature was able to identify early stage BC (I/II) with an AUC of 0.940. Integrating the signature could increase the diagnosis accuracy by up to 91.9% for breast cancer patients with the corresponding predictive results based on the Breast Imaging Reporting and Data System classification of 4 or 5. Moreover, the exLRs could provide a strong indication of the breast cancer subtypes, and exMSMO1 is employable as a predictive biomarker in response to neoadjuvant chemotherapy.

Conclusions: This study demonstrated the value of exLR profiling to provide potential biomarkers for early detection and treatment efficacy prediction of breast cancer.

Background

Breast cancer (BC) is the most common malignancy in women worldwide. According to the American Cancer Society, the estimated new cases and deaths of BC in the US in 2019 are 271,270 and 42,260, respectively[1]. Early detection and intervention significantly decrease the mortality rate of BC.

Mammography and ultrasound are the optimal methods for BC screening and are recommended by different clinical guidelines. However, their sensitivity and specificity are not consistent among different studies, and their false negatives hamper early diagnosis. In addition, their false positives lead to the overdiagnosis of BC[2–6]. Although magnetic resonance imaging (MRI) has been used to elevate the diagnosis accuracy of BC, the relatively high false positive rate and high cost hinder its routine utilization in BC diagnosis[7]. Carbohydrate antigen 15 – 3 (CA153) and carcino-embryonic antigen (CEA) are the biomarkers that are currently used for BC screening and treatment response monitoring; however, their sensitivity and specificity remain poor[8]. Hence, new efficient diagnostic methods and risk stratification approaches for BC must be developed.

The advent of liquid biopsy, especially extracellular vesicle (EV), has generated great enthusiasm for potential applications in cancer detection and treatment. EVs (exosomes and microvesicles) are lipid

bilayer-enclosed structures that contain various cargoes, including a large number proteins, lipids and nuclear acids[9]. Long RNA species, mainly messenger RNA, long non-coding RNA (lncRNA), pseudogene, and circular RNA (circRNA), have been recently found in human plasma EVs[10]. Certain long RNAs are enriched and stable in body fluids, and they could reflect their tissue origins and are suitable as liquid biopsy markers for cancer diagnosis and treatment efficacy monitoring[11, 12]. Koldemir et al. showed that the cellular expression of lncRNA GAS5 in BC cells leads to its exosomal enrichment, which is considered to be a marker of apoptotic induction[13]. Exosomal MALAT1 secreted by BC cells regulates cancer progression through enhanced cell proliferation of recipient cells in tumor microenvironment[14]. Moreover, exosomal lncRNA-SNHG14 promotes trastuzumab resistance in HER2-positive BC[15]. However, the role of EV long RNAs (exLRs) profiling in BC remains unknown.

In this study, we performed exLR-seq on plasma samples collected from 172 subjects, including BC patients, breast benign disease patients and healthy individuals who were receiving routine healthcare with the aim of exploring the potential of exLR-based signature as a clinically actionable biomarker for diagnosis, molecular classification, and treatment monitoring of BC patients who were receiving neoadjuvant chemotherapy.

Methods

Patients and clinical features

One hundred and seventy-two participants, including patients with BC (n = 112), benign patients (n = 19) and healthy controls receiving routine healthcare (healthy donor, n = 41), were enrolled in this study. All of the enrolled patients were suspicious for malignancy based on clinical or radiological evidence, and they were diagnosed with BC or benign by pathological examination. All of the participants were recruited from Fudan University Shanghai Cancer Center between 1 July 2017 and 30 December 2018. This study is conducted in accordance with Declaration of Helsinki. Informed written consent was obtained from each subject, and the study was approved by Institutional Review Board of Fudan University Shanghai Cancer Center, China.

Plasma sample collection

Blood samples were collected with 10 mL EDTA-coated Vacutainer tubes from all participants. Blood samples were collected before surgery from early stage BC patients and before chemotherapy from metastatic or locally advanced breast cancer (LABC) patients. Plasma was separated by centrifugation at $800 \times g$ (~ 3000 rpm) for 10 min at room temperature (25 °C) within 2 hours after blood collection. This step was followed by a second 10 min centrifugation at $16\,000 \times g$ (~ 13 000 rpm) at 4 °C to remove the cellular debris. Plasma samples were aliquoted and stored at - 80 °C until use.

Isolation of EVs and EV RNAs

For each patient, 1 mL of fresh or once-frozen thawed plasma was used, and EVs were isolated by affinity-based binding to spin columns using an exoRNeasy Serum/Plasma Kit (Qiagen, Hilden, Germany)

following the manufacturer's instructions. Briefly, thawed plasma was mixed with binding buffer and added to the exoEasy membrane affinity spin column. For transmission electron microscopy (TEM), size distribution measurement and Western blotting, the EVs were eluted with 400 μ L of XE elution buffer. To concentrate the EVs, samples were subjected to ultrafiltration using the Amicon Ultra-0.5 Centrifugal Filter 10 kDa (Merck Millipore, Germany). For the EV RNA isolation, EVs were lysed on the column using QIAzol (Qiagen), and the total RNA was then eluted and purified.

Transmission electron microscopy (TEM)

EVs were identified by negative staining with phosphotungstic acid. Fifty microliters of resuspended EVs were placed on a Parafilm membrane. A copper mesh with a formvar supporting membrane was covered with the EVs suspension and floated for 3–10 min to allow sample absorption into the supporting membrane. Next, 50 μ L of 2% phosphotungstic acid was dropped onto the Parafilm membrane. The fluid was then absorbed from the edges of the copper mesh with filter paper. The copper mesh absorbing the sample was covered with 2% phosphotungstic acid and floated for 3 min. Then, the sample was dried for 10 min under incandescent light after the staining solution was absorbed with filter paper. The copper mesh was imaged with a TEM (Phillips CM120, Tokyo, Japan).

Size distribution measurement

Size distribution analysis of the EVs was performed with a Flow NanoAnalyzer (NanoFCM Inc., Xiamen, China) according to the manufacturer's instructions. A series of monodisperse silica nanoparticles (SiNPs) were synthesized and used as size reference standards. Then, the side scattering (SSC) distribution histogram of the mixture was obtained. The SSC intensity of every vesicle was converted into its corresponding vesicle size. One hundred mL phosphate-buffered saline (PBS) resuspended EV samples and 100 mL PBS (blank control) were analyzed using the same instrument settings. The EV data were analyzed and used to construct a size distribution histogram.

Western blotting of EVs

Peripheral blood mononuclear cells (PBMC) were isolated by Lymphoprep (STEMCELL Technologies, USA) according to the manufacturer's instructions. PBMC and the concentrated EVs were lysed in RIPA buffer (1% NP40, 0.5% deoxycholate, 0.1% sodium dodecyl sulfate [SDS] in Tris-buffered saline) with complete protease inhibitors on ice for 30 min. Equal amounts of protein from EVs and PBMC were separated on 10% SDS-polyacrylamide gels and then transferred to nitrocellulose membranes (Bio-Rad, Hercules, CA, USA). Membranes were blocked with 5% non-fat milk and incubated with anti-CD63 (ab92726, 1: 1,000) (Abcam, Cambridge, USA), anti-TSG101 (sc-13611, 1:500) (Santa Cruz Biotechnology, Dallas, TX, USA), anti-Grp94 (14700-1-AP, 1: 1,000) and anti-Calnexin (10427-2-AP, 1: 1,000) (Proteintech, Rosemont, IL, USA) overnight. Target proteins were detected using an enhanced chemiluminescence kit (Amersham Pharmacia Biotech, Uppsala, Sweden).

RNA-seq analysis

Total EV RNA isolated from 1 mL of plasma was treated with DNase I (NEB, Ipswich, Massachusetts, USA) to remove DNA. Strand-specific RNA-seq libraries were prepared using the SMARTer Stranded Total RNA-Seq Kit—Pico Input Mammalian (Clontech, Palo Alto, California, USA). The library quality was analyzed using a Qubit fluorometer (Thermo Scientific, Waltham, Massachusetts, USA) and Qsep100 (BiOptic, New Taipei City, Taiwan). EV RNA-seq libraries could be prepared from 1 mL of plasma on average. ExLR-seq was performed on an Illumina sequencing platform (San Diego, California, USA) with 150 bp paired-end run metrics.

Raw reads were filtered using FastQC and aligned to the GRCh38 human genome assembly using STAR. Annotations of mRNA and lncRNA in the human genome were retrieved from the GENCODE (V.25). The circRNAs were discovered by the Assembling Splice Junctions Analysis (ASJA)[16], and the normalized method was the same as in a previous study[11]. The length and backsplicing ratio of the circRNAs were calculated on the basis of previous study[16]. Gene expression levels were calculated in transcripts per kilobase million (TPM). Differentially regulated exLRs were annotated gene IDs and were assessed for Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment using DAVID (<https://david.ncifcrf.gov/>).

Data and statistical analyses

RNA-seq raw read counts were converted to TPM values to scale all of the comparable variates and were normalized across all samples. Variates with frequencies of < 10% (i.e., expressed in less than 10% of the entire samples) were omitted, and the remaining markers were used for subsequent statistical analyses.

We used the Random Forest to find markers that could distinguish BC and control (benign + healthy) samples. The steps were as followed: (1) ExLR-seq TPM expression profiles (n = 172) were randomly distributed in training (n = 143) and validation cohorts (n = 29). (2) In the training cohort, the Mann-Whitney U test was used to assess the differential expression of exLRs in BC and control cohorts, and the p value of each marker was adjusted by the Benjamini-Hochberg method to control the false discovery rate (FDR). (3) The RandomForestClassifier method of the sklearn package and 0.01 threshold of SelectFromModel were used to define candidate makers and accuracy score. (4) Repeat (1) to (3) steps for 3000 times. Finally, we selected the best candidate makers based on accuracy score and then calculated the out-of-bag (OOB) error to determine markers.

Eleven exLRs evaluated by the mentioned algorithms and annotations were selected to construct a support vector machine (SVM) model for BC prediction. For binary (BC vs benign + healthy) sample classification, the SVM algorithm was executed using the 'LinearSVC' package in python software. In principle, the SVM algorithm determines the location of all samples in a high-dimensional space, in which each axis represents an exLR and the expression level of particular exLR in a sample determines its location on the axis. We divided sample randomly into training (n = 120) and validation (n = 52) cohorts with ratio of 7:3. During the training process, the SVM algorithm draws a hyperplane that best separates the two classes based on the distance between the closest sample of each class and the hyperplane. The different sample classes are positioned at each side of the hyperplane. Moreover, to assess the predictive

value of the SVM algorithm in an independent data set, which is not typically included in the SVM training process, the algorithm was trained with the training data set, all SVM parameters were fixed and samples in the internal validation and external validation cohorts were then evaluated. The internal training performance of the SVM algorithm could be improved by enabling the SVM tuning function, which implies optimal determination of parameters of the SVM algorithm (C, gamma, kernel) by randomly subsampling the data set used for the algorithm training ('fivefold internal cross-validation').

The d-signature score was computed from the predictive strength of the SVM classifier output. To assess the samples' probability of being predicted as BC, we used the R function 'predict' to evaluate the prediction strength in quantitative terms on the internal validation and external validation cohorts. The prediction strength of the SVM classifier output was used to establish the exLR d-signature. The diagnostic efficacy of the d-signature was evaluated by receiver operating characteristic (ROC) curve analysis for the training, internal validation and external validation cohorts. The comparison between the areas under the curve (AUCs) of the different classifiers was evaluated by the bootstrap method with 3000 iterations. Youden's index was determined to identify the optimal cut-off point for calculating the exact diagnostic indices. The d-signature distribution in the different patient groups was tested by the Wilcoxon rank-sum test and Student's t-test after using the Shapiro-Wilk test to determine the data normality.

For most of the experiments, independent sample t-tests were used to calculate the p values. Disease-free survival (DFS) was derived from the Kaplan-Meier estimate and compared by the log-rank test.

All statistical analyses were two-sided, and a p value < 0.05 was considered to be statistically significant. R software packages ('varSelRF' and 'pROC') and python packages ('RandomForestClassifier', 'train_test_split', 'SelectFromModel', 'accuracy_score', 'StratifiedKFold' and 'LinearSVC') were used in this study.

Results

Patient characteristics

One hundred and seventy-two individuals were included in this study; the participants consisted of 112 BC patients, 19 benign patients and 41 healthy donors (Table 1). Among the 112 BC patients, 28 were at stage I, 35 were at stage II, 15 were at stage III, and 34 were at stage IV (Table 1). Forty-nine BC patients with stage II or III received neoadjuvant chemotherapy in our cohort. The benign patients included 10 adenosis, 4 fibroadenomas, 4 mastitis and 1 intraductal papilloma. Other clinical features, including the age, the Breast Imaging Reporting and Data System classification (BI-RADS) of ultrasound or mammography, plasma CA15-3 and CEA level, are shown in Table 1.

EV isolation and exLR-seq

The isolated vesicles were rounded, cup-shaped, and membrane-enclosed, as observed by TEM (Fig. 1A). Flow cytometry revealed a heterogeneous population of spherical nanoparticles, with abundant peaks of less than 200 nm and a mean diameter of 92.5 nm (Fig. 1B). Western blot analysis revealed that the exosomal characteristic markers CD63 and TSG101 were enriched in the isolated vesicles but not in PBMC, whereas Grp94 and calnexin, which were both not expected to be enriched in EVs and considered to be negative markers, were observed in PBMC but not in the isolated vesicles (Fig. 1C). These findings indicate that the isolated EVs consisted mostly of exosomes.

ExLR-seq was conducted using plasma samples from 172 healthy individuals and patients. Approximately 15000 annotated genes, including mRNAs, lncRNAs, pseudogenes and circRNAs, were reliably detected in each sample. More mRNAs, lncRNAs, pseudogenes and circRNAs were identified in benign and BC groups than healthy group (Fig. 1D). We identified 1552 exLRs that were differentially expressed in BC samples compared with controls (benign + healthy) by the Mann-Whitney U test ($FDR < 0.01$, $|\text{fold change (FC)}| > 1.5$). Most different exLRs were up-regulated in the BC group. Unsupervised hierarchical clustering revealed a clear separation of the BC and control group (Fig. 1E). KEGG pathway analysis revealed that differentially expressed exLRs were enriched for some pathways involved in cancer, such as the TGF-beta signaling pathway, pathways in cancer, pancreatic cancer, and p53 signaling pathway (Fig. 1F).

Blood exLRs may reflect the relative fractions of different cell types

Because blood EVs are derived from a variety of tissues, we used the xCell circosplot (xCell) tool (<http://xcell.ucsf.edu>) to characterize the cell type proportions of the exLRs. xCell is a web tool that performs cell type enrichment analysis from gene expression data for 64 immune and stroma cell types. The relative proportion of the different cell types in BC and control samples are shown in Fig. 2A. We then adjusted the cell type enrichment scores to cell type proportions; 19 normalized average xCell scores were significantly different between BC and control groups (Fig. 2B). Blood cells (common lymphoid progenitors (CLP), common myeloid progenitor cell (CMP), erythrocytes and granulocyte macrophage progenitor (GMP)), immune cells (CD8 + T-cells, naïve B-cells, plasma cells, Th1 and Th2 cells), and stroma/epithelial cells (preadipocytes and smooth muscle cells) were significantly enriched in BC group compared with control (Fig. 2B).

Establishment of an exLR d-signature for BC

The different exLRs profiling between the BC and control groups implies that the exLRs have potential as biomarkers for the detection of BC. We then explored an exLR d-signature for the diagnosis of BC. ExLRs ($n = 1511$) that were upregulated in BC patients compared with controls were selected using a training cohort of 43 control individuals and 77 BC patients. The selected exLR markers were analyzed using the random forest algorithm and the LASSO method to shrink the number of variables. Finally, eleven exLR markers (BEX2, AC104843.1, AL136981.2, KRT19, NPM1P25, CTSG, CBR3, HOXB7, AL691447.3,

RNA5SP141 and circRNA chr13_42953948_42970670_-) were selected and used to construct a BC classifier (Table S1). Using the SVM algorithm, we established a diagnostic model and generated an exLR d-signature for BC. The exLR d-signature comprising the eleven exLRs distinguished the BC patients from controls with an AUC of 0.96 (95% confidence interval (CI): 0.93 to 0.99, standard deviation (SD): 0.01), a precision of 0.93, a recall of 0.93 and a f1-score of 0.92 in the training cohort (Fig. 3A and 3C). The diagnostic accuracy was 92.5%. The exLR d-signature was then applied to the validation cohort; BC was detected with an AUC of 0.90 (95% CI: 0.81 to 0.98. SD: 0.04), a precision of 0.92, a recall of 0.92 and a f1-score of 0.92 (Fig. 3B and 3D). The diagnostic accuracy was 92.3%. Unsupervised hierarchical clustering using the eleven exLRs effectively distinguished BC from controls with high specificity and sensitivity (Fig. 3E and 3F).

The exLR d-signature detects early BC

Early stage BC diagnosis allows immediate surgery without prior chemotherapy or radiation therapy and has a favorable prognosis. We found that BC exhibited a high median exLR d-signature score when compared with benign disease (0.809 vs 0.444; Mann-Whitney U test, $p < 0.001$) and healthy (0.809 vs 0.254; Mann-Whitney U test, $p < 0.001$, Fig. 4A). We also observed no correlation between the d-signature scores and tumor stages (Fig. 4B), which suggests that the diagnostic performance of the d-signature was independent of the tumor burden, which would make it an optimal diagnostic tool for the detection of BC. Therefore, we next confirmed the diagnostic performance of the d-signature in early stage of BC. The d-signature can identify early stage (I/II) BC from controls with an AUC of 0.94 (95% CI: 0.90 to 0.98), a precision of 0.89, a recall of 0.89 and a f1-score of 0.88 (Fig. 4C). Furthermore, the d-signature can identify early stage BC from healthy and benign groups with an AUC of 0.96 (95% CI: 0.93 to 0.99, Fig. 4D) and 0.88 (95% CI: 0.80 to 0.95, Fig. 4E), respectively. These results demonstrated that the exLR d-signature could be used for high-accuracy diagnosis of BC, even for early BC.

CA15-3 and CEA are the biomarkers that are currently used for BC screening and recurrence monitoring. The d-signature scores were significantly higher in the BC patients with CA15-3 positive compared to negative patients (Fig. 4F). However, there is no correlation between the d-signature score and CEA status (Fig. 4G).

The exLR d-signature has improved diagnostic performance for BC detection

The ability to complement the limitations of the current imaging examination in the detection of BC would add value to a biomarker for the diagnosis of BC. Clinically, in general, the probability of a malignant tumor for patients with imaging BI-RADS 4 (including 4a, 4b and 4c) is estimated to be within a range of 2–95%. We combined the exLR d-signature with the corresponding BI-RADS scores for predicting the presence of cancer. The diagnosis accuracy was approximately 91.9%, with a precision of 0.92, recall of 0.92, f1-score of 0.92, and AUC of 0.90 (95% CI: 0.83 to 0.97, SD: 0.04) (Fig. 5B) for patients with BI-RADS $\geq 4a$ (including 4a, 4b, 4c and 5). If only the patients with BI-RADS 4a or 4b were

considered, the integrated predictive value for predicting the presence of cancer in these patients was approximately 91.3%, with a precision of 0.92, recall of 0.91, f1-score of 0.91, and AUC of 0.90 (95% CI: 0.80 to 0.98, SD: 0.05) (Fig. 5C). These results indicate that exLRs combined with the BI-RADS system could be utilized as a more accurate biomarker for differential diagnosis of early-stage BC compared with the BI-RADS system only.

ExLR profiles in different BC subtypes

To further evaluate the potential role of plasma exLRs in BC subtypes classification, we performed a SVM algorithm with all of the BC patients' exLRs. We observed that 687 exLRs could be separated from each other by the specific exLRs (Fig. 6A).

We also investigated the different cell types from exLR-seq data of these four types of BC using the xCell tool. From about 64 immune and stromal cell types, 8 could be evaluated, and their proportions varied considerably (Fig. 6B). Astrocytes were significantly enriched in hormone receptor (HR) positive and HER2 positive BC, whereas plasma cells were enriched in triple negative breast cancer. Furthermore, the HR positive and HER2 negative subtype has the highest immune and microenvironment score (Fig. 6B).

Plasma exMSM01 as predictive biomarker for neoadjuvant chemotherapy of BC

We next investigated the ability of exLR profiling to predict treatment responses in BC patients who received neoadjuvant chemotherapy (NACT). Of the 112 BC patients, 48 received NACT (paclitaxel and/or doxorubicin-base regimens, plus trastuzumab if HER2 positive), and 14 achieved a pathological complete response (pCR) after NACT. We identified 552 exLRs that were differentially expressed in the pCR group compared with the non-pCR group by the Mann-Whitney U test ($p < 0.05$, $|FC| > 2$). Different expressed exLRs are shown in Fig. 7A. DAVID GO analysis revealed that the increased exLRs were enriched for biological processes such as RNA processing and splicing, while decreased exLRs were strongly involved in intracellular signal transduction and cell proliferation in non-pCR group (Fig. 7B). Gene set enrichment analysis (GSEA) analyses revealed that the steroid biosynthesis pathway was one of the most upregulated biological processes in the non-pCR group (Fig. 7C).

Methylsterol monooxygenase 1 (MSM01), an intermediate enzyme in the cholesterol biosynthetic pathway, was identified based on its relative high enrichment in the non-pCR group (Fig. 7D). ExMSM01 distinguished non-pCR from pCR patients with an AUC of 0.79 (95% CI: 0.62 to 0.97, SD: 0.09) (Fig. 7E). MSM01 was highly expressed in tumor tissue of BC compared with adjacent normal tissue in the cancer genome atlas (TCGA) dataset (Fig. 7F). We also evaluated the expression of MSM01 by the qRT-PCR assay in an independent BC cohort from FUSCC (Table S2), and we found that high MSM01 expression was significantly associated with poor DFS ($p = 0.0008$, Fig. 7G). Multivariate analysis also demonstrated that high MSM01 expression was independent of unfavorable prognostic factors for DFS in BC patients (Table S3; hazard ratio (HR) = 2.683; 95% CI: 1.571–4.583; $p < 0.001$). We performed a functional study using a small interference RNAs (siRNAs) pool and found that the silencing of MSM01 could significantly

enhance the sensitivity of MDA-MB-231 cells to the chemotherapy drug paclitaxel (PAX) and doxorubicin (DOX) (Fig. 7H). Moreover, the inhibition of MSMO1 expression increased the apoptosis rates of MDA-MB-231 to DOX (1.0 μ M) and PAX (0.1 μ M) (Fig. 7I). To better understand how MSMO1 promotes drug resistance in breast cancer cells, RNA-seq was performed to analyze the gene expression profile affected by MSMO1 knockdown. Gene set enrichment analysis showed that MSMO1 knockdown affected multiple signaling pathways, such as the mTORC1 signaling pathway (Fig. 7J). Further analysis showed that silencing MSMO1 reduced the phosphorylation of AKT and mTOR (Fig. 7K). Considering that mTORC1 signaling activation has been shown to promote therapy resistance in BC[17, 18], these data suggested that MSMO1 could promote drug resistance through modulating the mTORC1 signaling pathways in BC cells.

Discussion

The EV, circulating tumor cell (CTC) and circulating tumor DNA (ctDNA) are ideal biosources in blood-based liquid biopsies. They are used to detect, diagnose, and monitor cancer, and they could enable the early diagnosis of cancer, low costs by avoiding complex invasive procedures, tailoring molecular targeted treatments, improving convenience for cancer patients, and ultimately supplementing clinical oncological decision-making. However, the low detection rates of CTC and ctDNA and the suboptimal sensitivity for localized cancer patients hamper their implementation for early BC detection. Here, we report that molecular interrogation of blood exosomal long RNA can offer valuable diagnostics information for BC patients. In our study, we obtained exLR-seq expression profiles from 172 human plasma EV samples. Differences in the exLR levels were then compared between patients with BC, patients with breast benign disease and healthy participants, and a diagnostic signature for BC was finally established. In addition, the exLRs can provide a strong indication of the tumor subtype and can be employable as a predictive biosource upon response to NACT for LABC.

Blood-based liquid biopsy is an ideal method for cancer diagnosis. Tumor protein biomarkers, such as alpha-feto protein (AFP), prostate specific antigen (PSA), CEA, and CA 19 – 9, have been utilized for decades to detect early stage cancer, disease recurrence, and progression after therapy[19–21]. Although the utility of these tumor markers has been well established in clinical practice, they are not entirely specific and remain elevated in non-malignant conditions. In addition, they do not reflect the details of tumor biology, nor do they provide any predictive information on the response to therapy. However, the latest blood-based biomarkers include CTCs, ctDNAs, circulating miRNAs and tumor-derived EVs, and they have the potential of assessing real-time tumor responses to therapy as well as identifying dynamically resistant clones. EVs contain many bioactive molecules, such as proteins and nucleic acids, from parent cells, thereby playing vital roles in intercellular communication [22–25]. Recent studies have suggested that EVs represent an appealing source of diagnostic biomarkers[23, 25]. The analysis of EVs encompasses several advantages over CTCs and ctDNA due to their higher abundancies and stability in the bloodstream, as well as their functionality in supporting tumor-host cross talk or tumorigenesis.

Until Valadi et al. demonstrated that variable RNAs can be transported between cells by EVs in 2007, EVs have begun to attract the attention of scientists[26]. Recent studies have suggested that EVs contain large numbers of long RNAs[11]. Certain functional exLRs were involved in intercellular communication and were valuable for the prediction of treatment responses[27–29]. Moreover, some exLRs have been shown to be differentially expressed between cancer and healthy controls and could have potential for cancer diagnosis[30, 31]. To determine the differences in the plasma exLR profiles among BC, breast benign disease and healthy, we performed exLR profiling of plasma samples from all 172 participants using an optimized exLR-seq strategy we recently developed[11]. We identified 1552 exLRs that were differentially expressed in the BC samples compared with controls (benign + healthy) by the Mann-Whitney U test. We then used the xCell tool and found that these two groups had different cell type proportions. All of these previous studies suggested that exLRs were of great valuable and could serve as a non-invasive diagnostic biomarker.

To establish a plasma exLR d-signature for BC detection, we integrated the random forest algorithm and the LASSO method to shrink the number of variables. Finally, eleven exLR markers were selected and used to construct a BC classifier. For the 5 mRNAs that were used in the final d-signature, all were upregulated in both BC tissues and plasma EV when compared with normal breast tissues or plasma EV from healthy or benign controls, respectively. BEX2, KRT19, CTSG, and HOXB7 have been previously shown to be upregulated and to possess a tumor-promoting function in BC[32–35]. CBR3 is a member of the short-chain reductase family, and it has close correlation with radiotherapy sensitivity and chemotherapy sensitivity in esophageal cancer and BC[36, 37]. Moreover, Wu et al. demonstrated that CBR3 could play an important role in the development and malignancy of BC[38]. Little is known about the 5 pseudogene markers (AC104843.1, AL136981.2, NPM1P25, AL691447.3 and RNA5SP141) in previous cancer studies. The candidate circRNA (chr13: 42953948–42970670 -) is derived from the protein coding gene EPSTI1. EPSTI1 is a recently identified stromal fibroblast-induced gene that is highly overexpressed and critical for BC invasion and metastasis[39]. Previous study has demonstrated that circEPSTI1 (chr13: 43528083–43544806) sponges miR-4753 and miR-6809 to regulate BCL11A expression and affects cell proliferation and apoptosis in triple-negative BC[40]. Collectively, these results suggest that plasma EVs contain a considerable number of exLRs that could serve as potential biomarkers for BC detection.

The d-signature can distinguish BC from controls with high diagnostic accuracy in both training and validation cohorts (92.5% and 92.3%, respectively). Moreover, it is not influenced by the tumor burden and can act as an effective classifier to identify early stage (I/II) BC from controls (AUC = 0.94). The diagnosis of early stage BC is critical in clinical practice, which could enable immediate surgery, thereby improving the prognosis of BC.

BI-RADS based on mammography or ultrasound is routinely used for BC screening. However, there are disadvantages associated with these two forms of screening, including a high rate of false positive tests, frequent false negative results, and considerable implications for public health spending. For example, clinical doctors routinely choose invasive methods, either core needle biopsy or excision biopsy, to further

examine patients who have a BI-RADS 4a or higher finding. If all patients with BI-RADS 4a findings receive invasive procedures, more than 90% of the lesions will be benign, which indicates unnecessary overtreatment for these patients. When integrating the diagnosis exLRs signature with BI-RADS 4a or higher findings, the diagnosis accuracy was approximately 91.9%. If we considered only the patients with BI-RADS 4a or 4b, the integrated predictive value was approximately 91.3%. Therefore, by integrating the exLRs signature and film findings, the possibility of false positive and the overtreatment of patients would be greatly decreased.

To further explore the role of exLRs in predicting the neoadjuvant chemotherapy treatment response in BC, we divided 48 patients who received NACT into pathological complete response (pCR) (n = 14) and non-pCR (n = 34) groups based on the post-surgical pathology. Different exLR profiles were found between these two groups. Through GSEA analysis, the steroid biosynthesis pathway is one of the most upregulated pathways in the non-pCR group. Many studies have implicated the function of the sterol synthesis pathway in tumor growth and response to treatment. For example, the sterol composition of the membrane has been shown to regulate EGFR signaling[41] and the sensitivity of head and neck cancer cells to apoptosis[42]. ExMSMO1 enriched in the non-pCR group and could distinguish non-pCR from pCR with an AUC of 0.79 (95% CI: 0.62 to 0.97, SD: 0.09) (Fig. 7D, E). MSMO1 is an intermediate enzyme in the cholesterol biosynthetic pathway. Previous study has indicated that the inactivation of MSMO1 markedly sensitized tumor cells to therapeutic anti-EGFR antibody via increased EGF receptor degradation[43]. In this study, we demonstrated that MSMO1 was overexpressed in BC and correlated with poor survival. Moreover, MSMO1 can decrease the sensitivity of BC cells to PAX and DOX by regulating the mTORC1 signaling pathway. In other words, these findings demonstrate that exMSMO1 can act as a predictive biomarker for neoadjuvant treatment efficacy of BC.

Several limitations in this study warrant mention. First, as a single-center study, it is uncertain whether this diagnosis signature is applicable to other populations composed of different cultures and ethnic groups. Future work with larger cohorts from multiple centers is still needed to externally validate our results. Second, the roles of exLRs in BC subtype distinguishing and treatment efficient predictions are still not fully evaluated in this study. Further work must be done on this issue.

Conclusions

Our study evaluated the extracellular vehicle long RNA profiles among BC, benign, and healthy samples and developed a stable and valuable SVM classifier model based on exLRs to distinguish BC and controls. We believe that the development of an effective, noninvasive diagnostic and risk stratification model could be helpful for clinicians.

Abbreviations

ASJA: Assembling Splice Junctions Analysis; AUC: areas under the curve (AUC); BC: Breast cancer; BI-RADS: Breast Imaging Reporting and Data System classification; CA153: Carbohydrate antigen 15 – 3;

CEA: carcino-embryonic antigen; CI: confidence interval; circRNA: circular RNA; CTC: circulating tumor cell; ctDNA: circulating tumor DNA; DFS: Disease-free survival; DOX: doxorubicin; EV: extracellular vehicle; exLR: extracellular vehicle long RNA; FC: fold change; FDR: false discovery rate; GSEA: Gene set enrichment analysis; HR: hormone receptor; KEGG: Kyoto Encyclopedia of Genes and Genomes; LABC: locally advanced breast cancer; LASSO: least absolute shrinkage and selection operator; lncRNA: long non-coding RNA; MRI: magnetic resonance imaging; MSMO1: Methylsterol monooxygenase 1; NACT: neoadjuvant chemotherapy; OOB: out-of-bag; PAX: paclitaxel; PBMC: Peripheral blood mononuclear cells; PBS: phosphate-buffered saline; pCR: pathological complete response; ROC: receiver operating characteristic; SD: standard deviation; SVM: support vector machine; TCGA: the cancer genome atlas; TEM: transmission electron microscopy; TPM: transcripts per kilobase million.

Declarations

Ethics approval and consent to participate: This study is conducted in accordance with Declaration of Helsinki. Informed written consent was obtained from each subject, and the study was approved by Institutional Review Board of Fudan University Shanghai Cancer Center, China.

Consent for publication: Not applicable.

Availability of data and materials: Datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests: The authors declare that they have no competing interests.

Funding: This work was supported by the National Key R&D Program of China (2017YFC1311004).

Authors' contributions: JW, SH, YC and YS contributed to conception and design; YS, RG, WC, JiaW, YuS, LL, JH, JX and ZS contributed to provision of study materials or patients; YS, JZ, RG, WC, YL, JiaW, ZW, YuS, LL and JH collected and assembled the data; JW, SH, YS, JZ, HL, RG, and YL contributed to data analysis and interpretation; JW, SH, YC and YS wrote the manuscript. All authors finally approved the manuscript.

Acknowledgements: We sincerely appreciate the support from the patients and their families. We are grateful to Dr. Bingqiu Xiu, Dr. Qi Zhang and Prof. Xianghuo He for critical reading of the manuscript.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin.* 2019;69:7–34.
2. Chan CH, Coopey SB, Freer PE, Hughes KS. False-negative rate of combined mammography and ultrasound for women with palpable breast masses. *Breast Cancer Res Treat.* 2015;153:699–702.
3. Harding C, Pompei F, Burmistrov D, Welch HG, Abebe R, Wilson R. Breast Cancer Screening, Incidence, and Mortality Across US Counties. *JAMA Intern Med.* 2015;175:1483–9.

4. Narod S. Breast cancer: The importance of overdiagnosis in breast-cancer screening. *Nat Rev Clin Oncol.* 2016;13:5–6.
5. Ohuchi N, Suzuki A, Sobue T, Kawai M, Yamamoto S, Zheng YF, Shiono YN, Saito H, Kuriyama S, Tohno E, et al. Sensitivity and specificity of mammography and adjunctive ultrasonography to screen for breast cancer in the Japan Strategic Anti-cancer Randomized Trial (J-START): a randomised controlled trial. *Lancet.* 2016;387:341–8.
6. Ya-jie J, Wei-jun P, Cai C, Jian-hui D, Wei Z, Min C, Guang-yu L. Application of breast ultrasound in a mammography-based Chinese breast screening study. *Cell Biochem Biophys.* 2013;65:37–41.
7. Sun H, Li H, Si S, Qi S, Zhang W, Ma H, Liu S, Yingxue L, Qian W. Performance evaluation of breast cancer diagnosis with mammography, ultrasonography and magnetic resonance imaging. *J Xray Sci Technol.* 2018;26:805–13.
8. Mirabelli P, Incoronato M. Usefulness of traditional serum biomarkers for management of breast cancer patients. *Biomed Res Int.* 2013;2013:685641.
9. Colombo M, Raposo G, Thery C. Biogenesis, secretion, and intercellular interactions of exosomes and other extracellular vesicles. *Annu Rev Cell Dev Biol.* 2014;30:255–89.
10. van Niel G, D'Angelo G, Raposo G. Shedding light on the cell biology of extracellular vesicles. *Nat Rev Mol Cell Biol.* 2018;19:213–28.
11. Li Y, Zhao J, Yu S, Wang Z, He X, Su Y, Guo T, Sheng H, Chen J, Zheng Q, et al. Extracellular Vesicles Long RNA Sequencing Reveals Abundant mRNA, circRNA, and lncRNA in Human Blood as Potential Biomarkers for Cancer Diagnosis. *Clin Chem.* 2019;65:798–808.
12. Li Y, Zheng Q, Bao C, Li S, Guo W, Zhao J, Chen D, Gu J, He X, Huang S. Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis. *Cell Res.* 2015;25:981–4.
13. Koldemir O, Ozgur E, Gezer U. Accumulation of GAS5 in exosomes is a marker of apoptosis induction. *Biomed Rep.* 2017;6:358–62.
14. Zhang P, Zhou H, Lu K, Lu Y, Wang Y, Feng T. Exosome-mediated delivery of MALAT1 induces cell proliferation in breast cancer. *Onco Targets Ther.* 2018;11:291–9.
15. Dong H, Wang W, Chen R, Zhang Y, Zou K, Ye M, He X, Zhang F, Han J. Exosome-mediated transfer of lncRNASNHG14 promotes trastuzumab chemoresistance in breast cancer. *Int J Oncol.* 2018;53:1013–26.
16. Zhao J, Li Q, Li Y, He X, Zheng Q, Huang S. ASJA: A Program for Assembling Splice Junctions Analysis. *Comput Struct Biotechnol J.* 2019;17:1143–50.
17. Jiang BH, Liu LZ. Role of mTOR in anticancer drug resistance: perspectives for improved drug treatment. *Drug Resist Updat.* 2008;11:63–76.
18. Butt G, Shahwar D, Qureshi MZ, Attar R, Akram M, Birinci Y, Karatoprak GS, Gasparri ML, Farooqi AA. Role of mTORC1 and mTORC2 in Breast Cancer: Therapeutic Targeting of mTOR and Its Partners to Overcome Metastasis and Drug Resistance. *Adv Exp Med Biol.* 2019;1152:283–92.

19. Jia J, Zhang P, Gou M, Yang F, Qian N, Dai G. The Role of Serum CEA and CA19-9 in Efficacy Evaluations and Progression-Free Survival Predictions for Patients Treated with Cetuximab Combined with FOLFOX4 or FOLFIRI as a First-Line Treatment for Advanced Colorectal Cancer. *Dis Markers*. 2019;2019:6812045.
20. Kearns JT, Holt SK, Wright JL, Lin DW, Lange PH, Gore JL. PSA screening, prostate biopsy, and treatment of prostate cancer in the years surrounding the USPSTF recommendation against prostate cancer screening. *Cancer*. 2018;124:2733–9.
21. Mok TS, Yeo W, Yu S, Lai P, Chan HL, Chan AT, Lau JW, Wong H, Leung N, Hui EP, et al. An intensive surveillance program detected a high incidence of hepatocellular carcinoma among hepatitis B virus carriers with abnormal alpha-fetoprotein levels or abdominal ultrasonography results. *J Clin Oncol*. 2005;23:8041–7.
22. Manier S, Liu CJ, Avet-Loiseau H, Park J, Shi J, Campigotto F, Salem KZ, Huynh D, Glavey SV, Rivotto B, et al. Prognostic role of circulating exosomal miRNAs in multiple myeloma. *Blood*. 2017;129:2429–36.
23. Melo SA, Luecke LB, Kahlert C, Fernandez AF, Gammon ST, Kaye J, LeBleu VS, Mittendorf EA, Weitz J, Rahbari N, et al. Glypican-1 identifies cancer exosomes and detects early pancreatic cancer. *Nature*. 2015;523:177–82.
24. Severino V, Dumonceau JM, Delhaye M, Moll S, Annessi-Ramseyer I, Robin X, Frossard JL, Farina A. Extracellular Vesicles in Bile as Markers of Malignant Biliary Stenoses. *Gastroenterology*. 2017;153:495–504 e498.
25. Yang KS, Im H, Hong S, Pergolini I, Del Castillo AF, Wang R, Clardy S, Huang CH, Pille C, Ferrone S, et al: **Multiparametric plasma EV profiling facilitates diagnosis of pancreatic malignancy**. *Sci Transl Med* 2017, 9.
26. Valadi H, Ekstrom K, Bossios A, Sjostrand M, Lee JJ, Lotvall JO. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat Cell Biol*. 2007;9:654–9.
27. Del Re M, Biasco E, Crucitta S, Derosa L, Rofi E, Orlandini C, Miccoli M, Galli L, Falcone A, Jenster GW, et al. The Detection of Androgen Receptor Splice Variant 7 in Plasma-derived Exosomal RNA Strongly Predicts Resistance to Hormonal Therapy in Metastatic Prostate Cancer Patients. *Eur Urol*. 2017;71:680–7.
28. Del Re M, Marconcini R, Pasquini G, Rofi E, Vivaldi C, Bloise F, Restante G, Arrigoni E, Caparello C, Bianco MG, et al. PD-L1 mRNA expression in plasma-derived exosomes is associated with response to anti-PD-1 antibodies in melanoma and NSCLC. *Br J Cancer*. 2018;118:820–4.
29. Qu L, Ding J, Chen C, Wu ZJ, Liu B, Gao Y, Chen W, Liu F, Sun W, Li XF, et al. Exosome-Transmitted IncARSR Promotes Sunitinib Resistance in Renal Cancer by Acting as a Competing Endogenous RNA. *Cancer Cell*. 2016;29:653–68.
30. Goldvaser H, Gutkin A, Beery E, Edel Y, Nordenberg J, Wolach O, Rabizadeh E, Uziel O, Lahav M. Characterisation of blood-derived exosomal hTERT mRNA secretion in cancer patients: a potential

- pan-cancer marker. *Br J Cancer*. 2017;117:353–7.
31. Yu S, Li Y, Liao Z, Wang Z, Wang Z, Li Y, Qian L, Zhao J, Zong H, Kang B, et al. Plasma extracellular vesicle long RNA profiling identifies a diagnostic signature for the detection of pancreatic ductal adenocarcinoma. *Gut*. 2020;69:540–50.
 32. Naderi A. Molecular functions of brain expressed X-linked 2 (BEX2) in malignancies. *Exp Cell Res*. 2019;376:221–6.
 33. Liu S, Jin K, Hui Y, Fu J, Jie C, Feng S, Reisman D, Wang Q, Fan D, Sukumar S, Chen H. HOXB7 promotes malignant progression by activating the TGFbeta signaling pathway. *Cancer Res*. 2015;75:709–19.
 34. Morimoto-Kamata R, Yui S. Insulin-like growth factor-1 signaling is responsible for cathepsin G-induced aggregation of breast cancer MCF-7 cells. *Cancer Sci*. 2017;108:1574–83.
 35. Saha SK, Choi HY, Kim BW, Dayem AA, Yang GM, Kim KS, Yin YF, Cho SG. KRT19 directly interacts with beta-catenin/RAC1 complex to regulate NUMB-dependent NOTCH signaling pathway and breast cancer properties. *Oncogene*. 2017;36:332–49.
 36. Lal S, Sandanaraj E, Wong ZW, Ang PC, Wong NS, Lee EJ, Chowbay B. CBR1 and CBR3 pharmacogenetics and their influence on doxorubicin disposition in Asian breast cancer patients. *Cancer Sci*. 2008;99:2045–54.
 37. Ogawa R, Ishiguro H, Kuwabara Y, Kimura M, Mitsui A, Mori Y, Mori R, Tomoda K, Katada T, Harada K, Fujii Y. Identification of candidate genes involved in the radiosensitivity of esophageal cancer cells by microarray analysis. *Dis Esophagus*. 2008;21:288–97.
 38. Wu Y, Liu F, Luo S, Yin X, He D, Liu J, Yue Z, Song J. **Co-expression of key gene modules and pathways of human breast cancer cell lines.** *Biosci Rep* 2019, 39.
 39. Li T, Lu H, Shen C, Lahiri SK, Wason MS, Mukherjee D, Yu L, Zhao J. Identification of epithelial stromal interaction 1 as a novel effector downstream of Kruppel-like factor 8 in breast cancer invasion and metastasis. *Oncogene*. 2014;33:4746–55.
 40. Chen B, Wei W, Huang X, Xie X, Kong Y, Dai D, Yang L, Wang J, Tang H, Xie X. circEPSTI1 as a Prognostic Marker and Mediator of Triple-Negative Breast Cancer Progression. *Theranostics*. 2018;8:4003–15.
 41. Sigismund S, Argenzio E, Tosoni D, Cavallaro E, Polo S, Di Fiore PP. Clathrin-mediated internalization is essential for sustained EGFR signaling but dispensable for degradation. *Dev Cell*. 2008;15:209–19.
 42. Bionda C, Athias A, Poncet D, Alphonse G, Guezguez A, Gambert P, Rodriguez-Lafrasse C, Ardail D. Differential regulation of cell death in head and neck cell carcinoma through alteration of cholesterol levels in lipid rafts microdomains. *Biochem Pharmacol*. 2008;75:761–72.
 43. Sukhanova A, Gorin A, Serebriiskii IG, Gabitova L, Zheng H, Restifo D, Egleston BL, Cunningham D, Bagnyukova T, Liu H, et al. Targeting C4-demethylating genes in the cholesterol pathway sensitizes cancer cells to EGF receptor inhibitors via increased EGF receptor degradation. *Cancer Discov*. 2013;3:96–111.

Tables

Table 1. Patient characteristics			
	Breast cancer	Breast benign disease	Healthy cohort
Total	112	19	41
Age, years			
Median	51.5	46	54
Range	31–75	22–61	40–85
Mammography (BI-RADS)*			
1–3	18 (20.7%)	3 (21.4%)	/
4	48 (55.2%)	11 (78.6%)	/
5	21 (24.1%)	0 (0%)	/
Ultrasound (BI-RADS)			
1–3	1 (1%)	7 (43.8%)	/
4	50 (50.5%)	9 (56.2%)	/
5	48 (49.5%)	0 (0%)	/
Serum CA15-3, U/ml*			
≤ 25	58 (70.7%)	/	/
> 25	24 (29.3%)	/	/
Serum CEA, ng/ml*			
≤ 5.2	59 (72.0%)	/	/
> 5.2	23 (28.0%)	/	/
Cancer stage			
I	28 (25.0%)	/	/
II	35 (31.3%)	/	/
III	15 (13.4%)	/	/
IV	34 (30.4%)	/	/
Hormone receptor			
Positive	63 (56.3%)	/	/
Negative	49 (43.8%)	/	/
HER2			

Table 1. Patient characteristics			
Positive	64 (57.7%)	/	/
Negative	47 (42.3%)	/	/
Abbreviations: BI-RADS, breast imaging reporting and data system; CA15-3, carbohydrate antigen 15 - 3; CEA, carcino-embryonic antigen; HER2, human epidermal growth factor receptor 2.			
*Excluded the unknown category.			

Figures

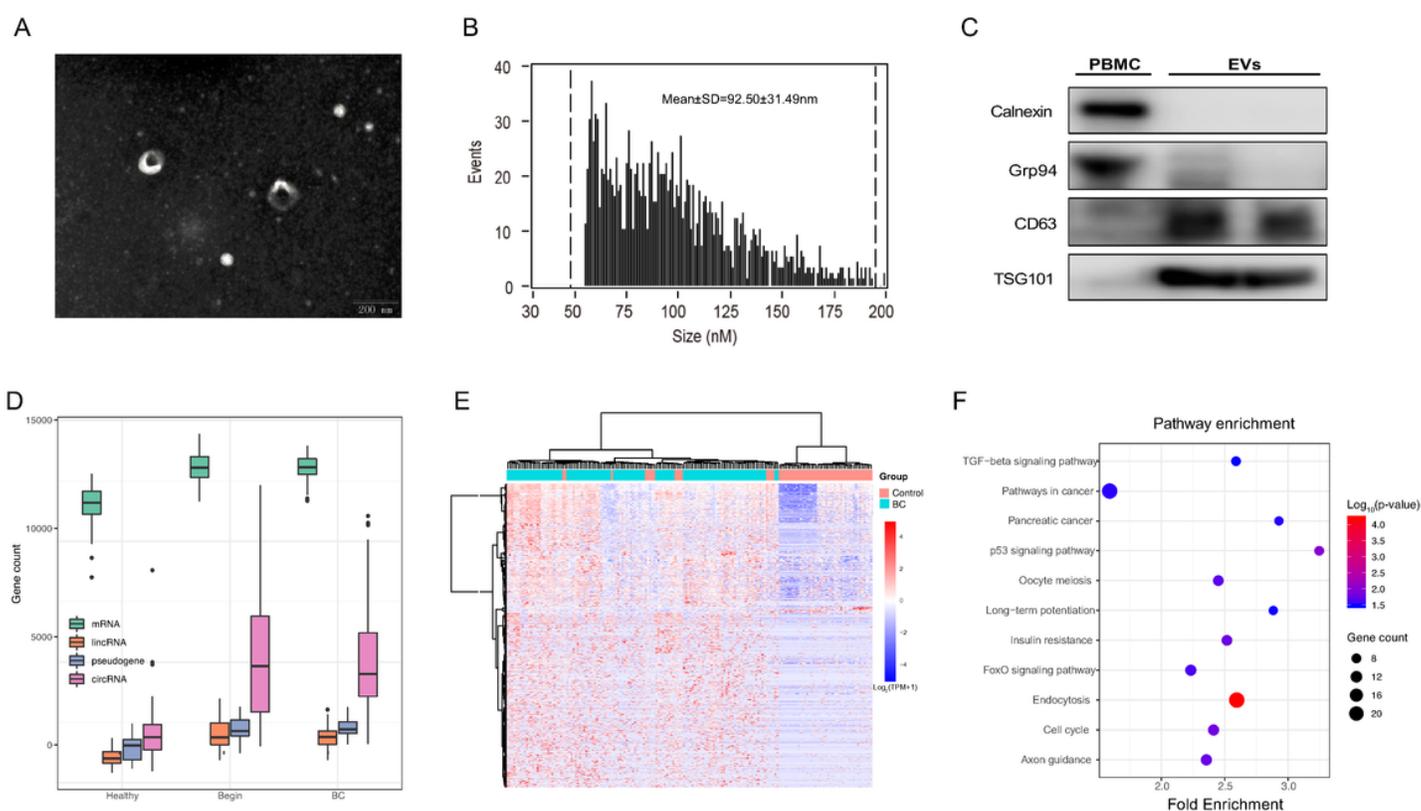


Figure 1

ExLR profiles of the cohort. (A, B) EVs were detected by transmission electron microscopy (A) and flow cytometry (B). Scale bar, 200 nm. (C) Western blots of EV markers TSG101, CD63 Grp94 and calnexin expression in peripheral blood mononuclear cells (PBMC) and isolated vesicles. (D) The distribution of exLRs per sample among BC, benign and healthy patients. (E) Heatmap of unsupervised hierarchical clustering of the exLRs that were differentially expressed between BC patients and controls (healthy+CP). Each column represents an individual sample, and each row represents an exLR. The scale represents the expression values. (F) KEGG pathway enrichment analysis for the differentially expressed exLRs.

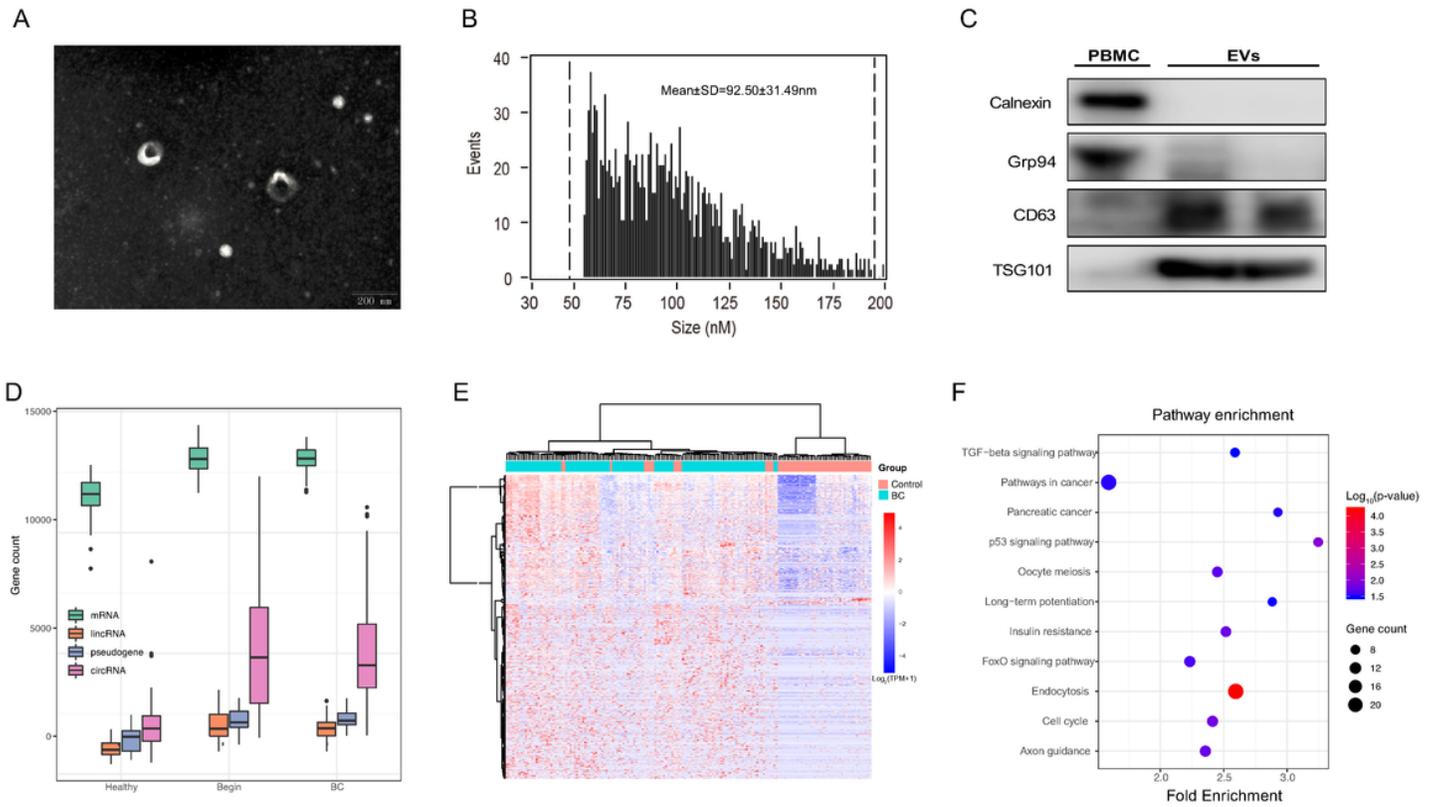


Figure 1

ExLR profiles of the cohort. (A, B) EVs were detected by transmission electron microscopy (A) and flow cytometry (B). Scale bar, 200 nm. (C) Western blots of EV markers TSG101, CD63 Grp94 and calnexin expression in peripheral blood mononuclear cells (PBMC) and isolated vesicles. (D) The distribution of exLRs per sample among BC, benign and healthy patients. (E) Heatmap of unsupervised hierarchical clustering of the exLRs that were differentially expressed between BC patients and controls (healthy+CP). Each column represents an individual sample, and each row represents an exLR. The scale represents the expression values. (F) KEGG pathway enrichment analysis for the differentially expressed exLRs.

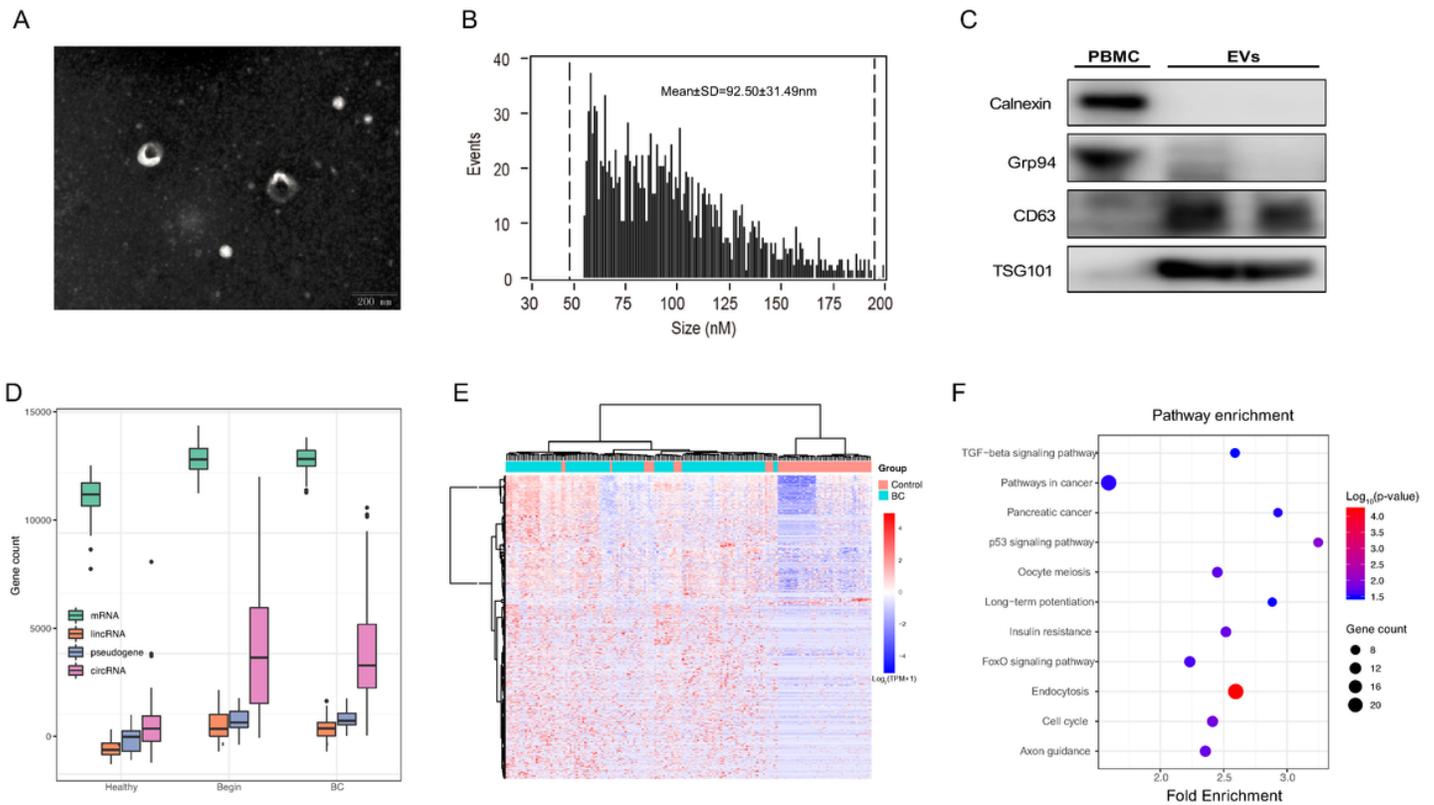
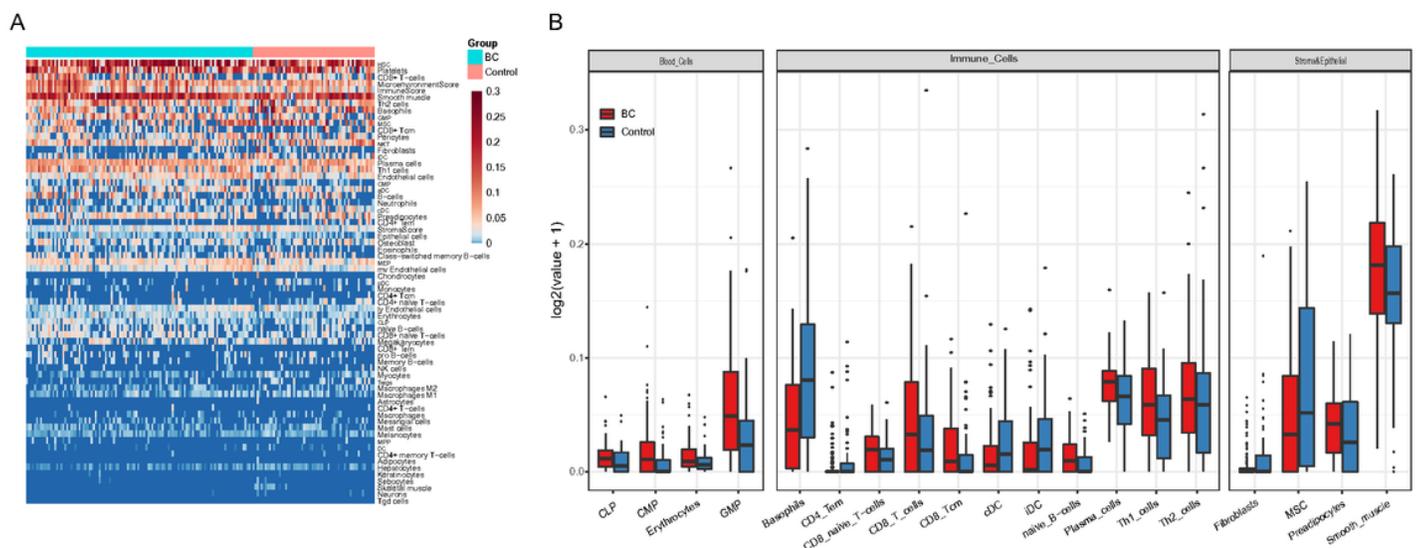


Figure 1

ExLR profiles of the cohort. (A, B) EVs were detected by transmission electron microscopy (A) and flow cytometry (B). Scale bar, 200 nm. (C) Western blots of EV markers TSG101, CD63 Grp94 and calnexin expression in peripheral blood mononuclear cells (PBMC) and isolated vesicles. (D) The distribution of exLRs per sample among BC, benign and healthy patients. (E) Heatmap of unsupervised hierarchical clustering of the exLRs that were differentially expressed between BC patients and controls (healthy+CP). Each column represents an individual sample, and each row represents an exLR. The scale represents the expression values. (F) KEGG pathway enrichment analysis for the differentially expressed exLRs.



Relative fractions of different cell types by exLR-seq in BC (A) ExLRs reflect the relative proportions of different cell types using xCell. Each column represents an individual sample, and each row represents a cell type. The scale represents the relative fractions. (B) Comparison of different cell types from the exLR-seq data between BC and controls. Only significant differences between these two groups are shown.

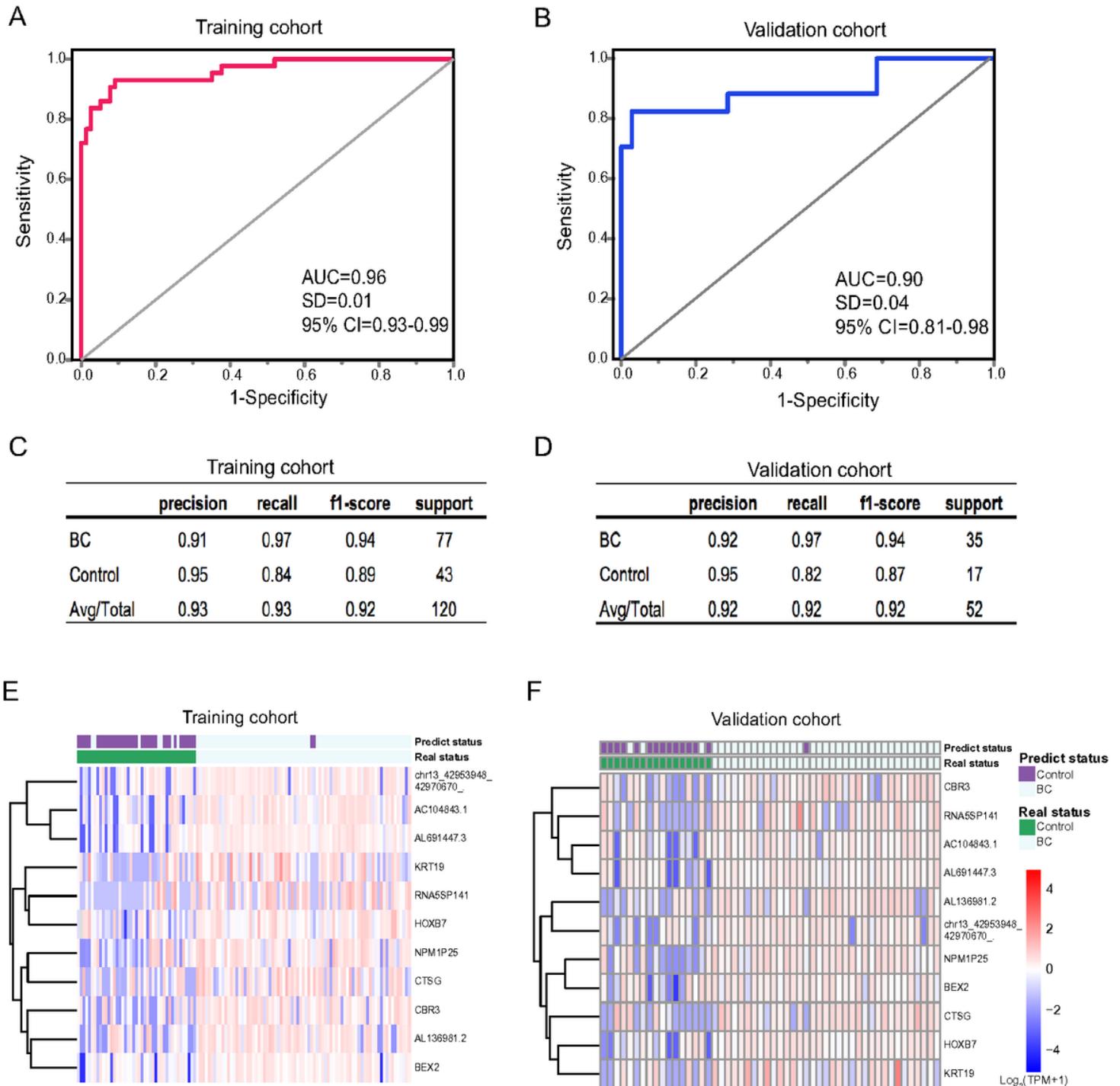


Figure 3

Blood exLR profiles can distinguish patients with BC from controls. (A, B) ROC for the performance of the exLR d-signature in the training (n=120, A) and validation (n=52, B) cohorts. (C, D) The diagnosis effects

of the d-signature in the training (C) and validation (D) cohorts. (E, F) Unsupervised hierarchical clustering of eleven exLRs selected for use in the d- signature in the training (E) and validation (F) cohorts. Each column represents an individual sample, and each row represents an exLR. The scale represents the expression values. ROC, receiver operating characteristic; AUC, area under the curve; SD, standard deviation; CI, confidence interval.

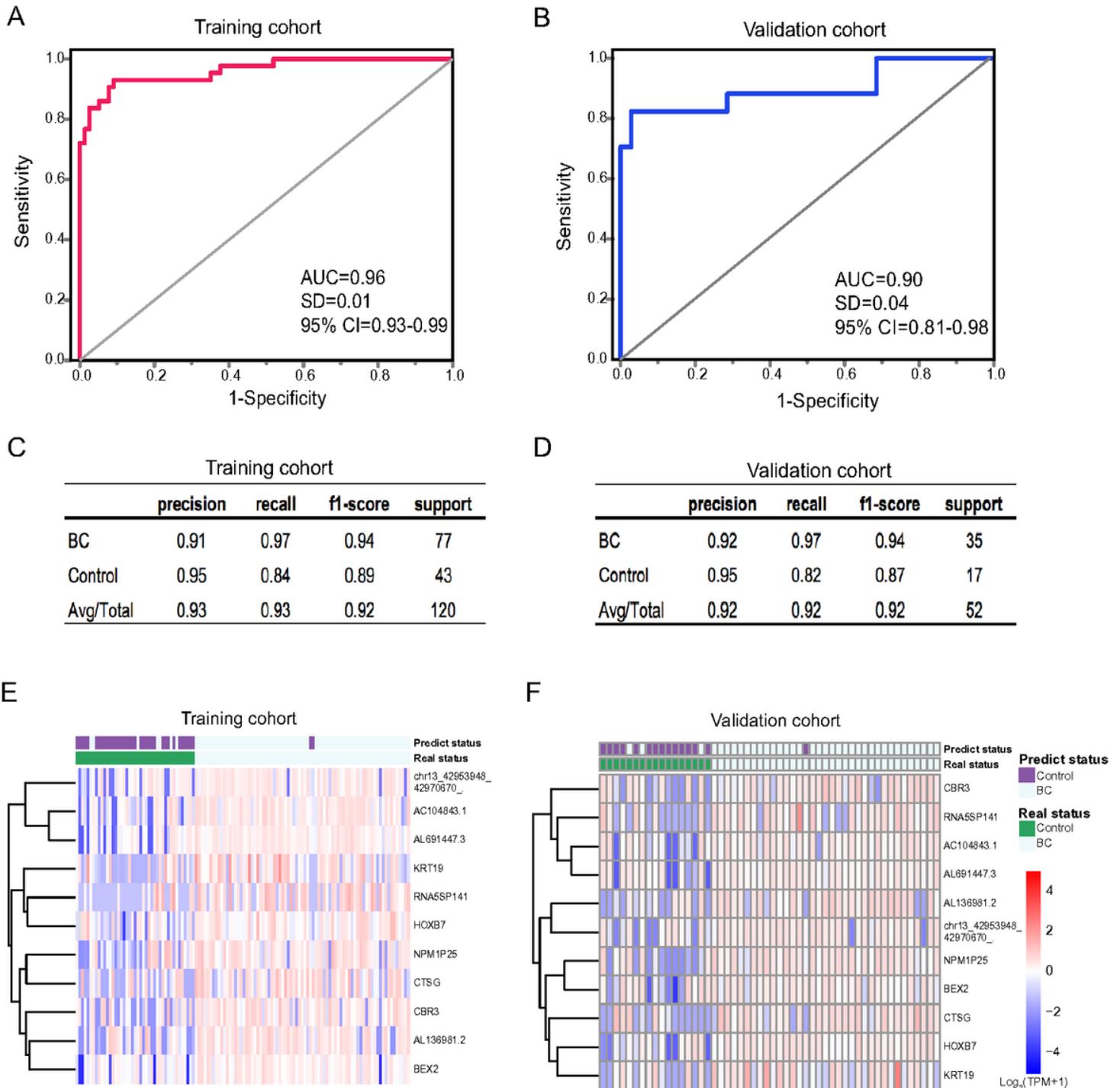


Figure 3

Blood exLR profiles can distinguish patients with BC from controls. (A, B) ROC for the performance of the exLR d-signature in the training (n=120, A) and validation (n=52, B) cohorts. (C, D) The diagnosis effects of the d-signature in the training (C) and validation (D) cohorts. (E, F) Unsupervised hierarchical clustering of eleven exLRs selected for use in the d-signature in the training (E) and validation (F) cohorts. Each column represents an individual sample, and each row represents an exLR. The scale represents the expression values. ROC, receiver operating characteristic; AUC, area under the curve; SD, standard deviation; CI, confidence interval.

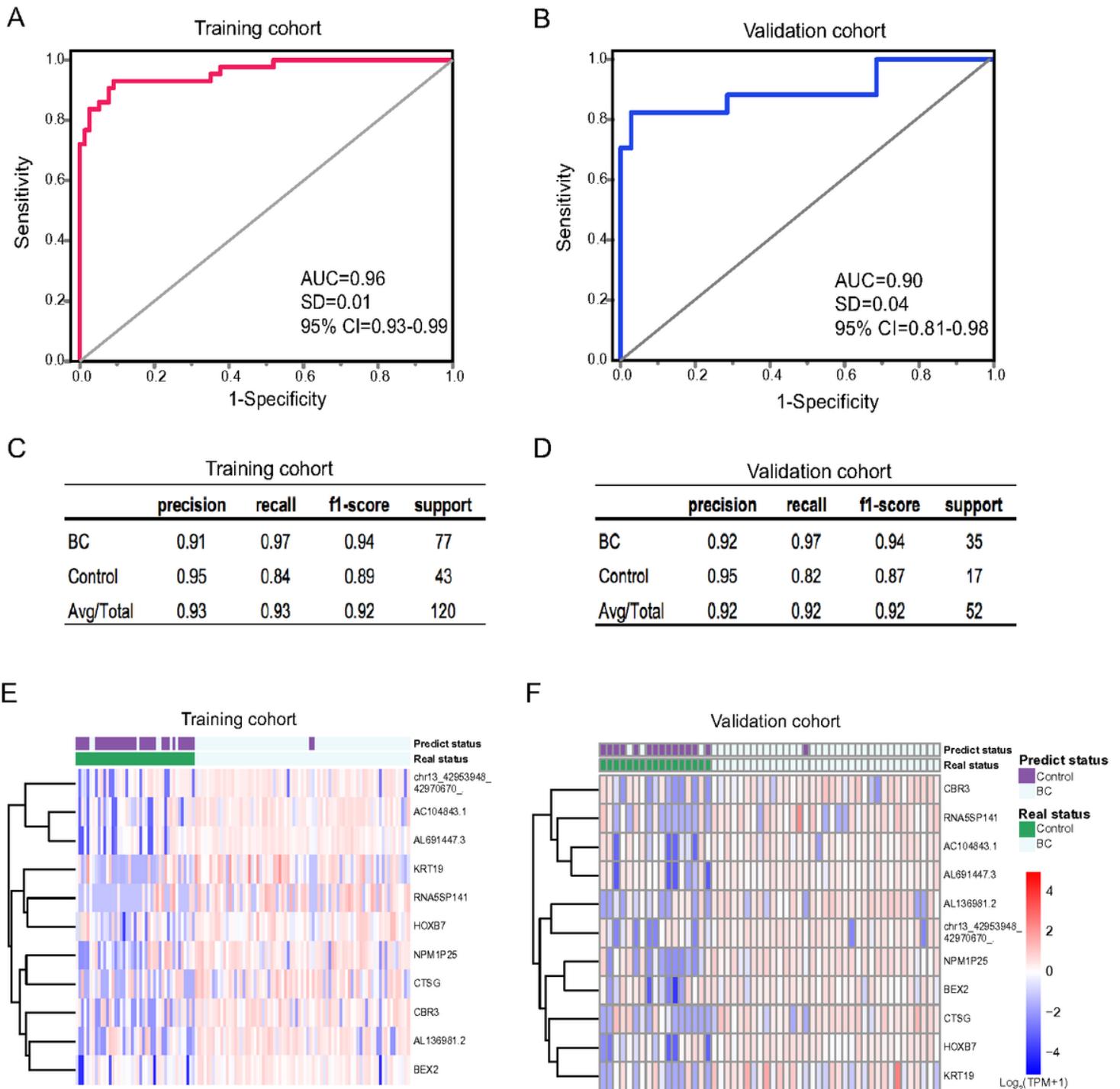


Figure 3

Blood exLR profiles can distinguish patients with BC from controls. (A, B) ROC for the performance of the exLR d-signature in the training (n=120, A) and validation (n=52, B) cohorts. (C, D) The diagnosis effects of the d-signature in the training (C) and validation (D) cohorts. (E, F) Unsupervised hierarchical clustering of eleven exLRs selected for use in the d- signature in the training (E) and validation (F) cohorts. Each column represents an individual sample, and each row represents an exLR. The scale represents the expression values. ROC, receiver operating characteristic; AUC, area under the curve; SD, standard deviation; CI, confidence interval.

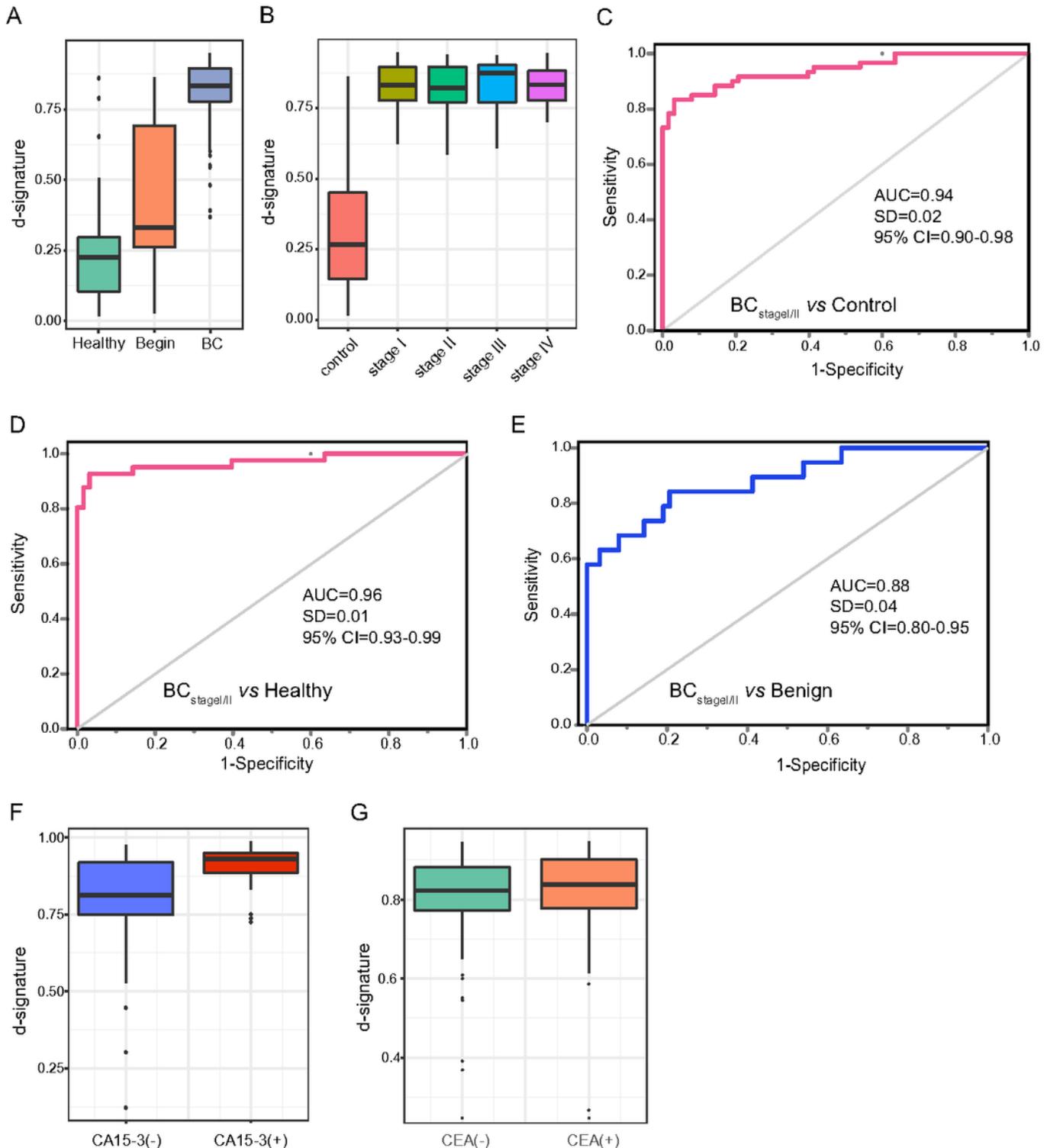


Figure 4

ExLR-based d-signature for the diagnosis of early stage BC (A) ExLR d-signature scores in healthy (n=41), benign (n=19) and BC (n=112). (B) ExLR d-signature scores in BC patients with stage I (n=28), II (n=35), III (n=15) and IV (n=34). (C-E) ROC for the performance of the exLR d-signature in BC with early stages (Stage I/II) compared to control (C), healthy (D) and (benign). (F, G) ExLR d-signature scores in BC patients with different serum CA15-3 (F) and CEA (G) statuses. AUC, area under the curve. SD, standard deviation. CI, confidence interval.

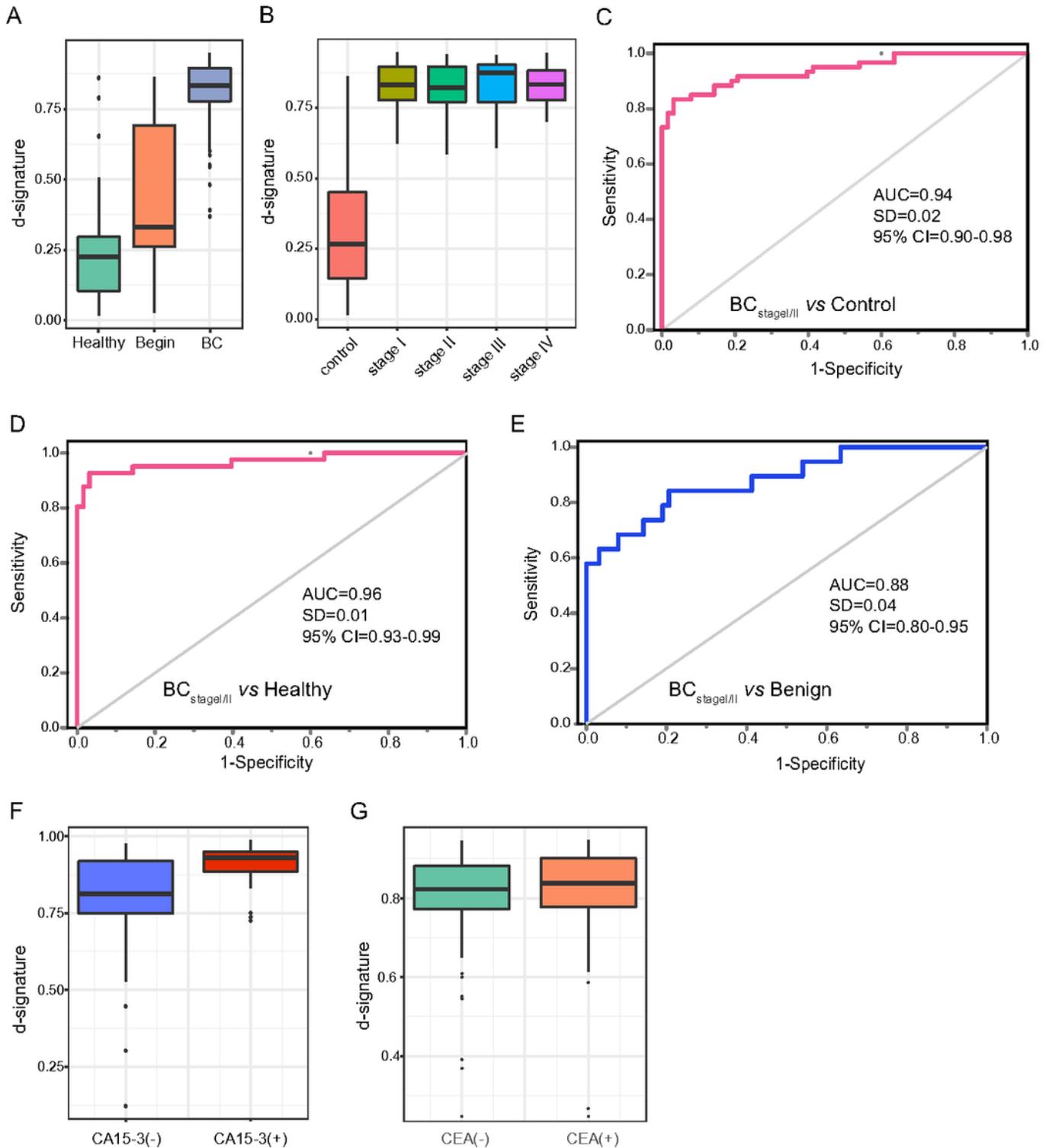


Figure 4

ExLR-based d-signature for the diagnosis of early stage BC (A) ExLR d-signature scores in healthy (n=41), benign (n=19) and BC (n=112). (B) ExLR d-signature scores in BC patients with stage I (n=28), II (n=35), III (n=15) and IV (n=34). (C-E) ROC for the performance of the exLR d-signature in BC with early stages (Stage I/II) compared to control (C), healthy (D) and (benign). (F, G) ExLR d-signature scores in BC patients with different serum CA15-3 (F) and CEA (G) statuses. AUC, area under the curve. SD, standard deviation. CI, confidence interval.

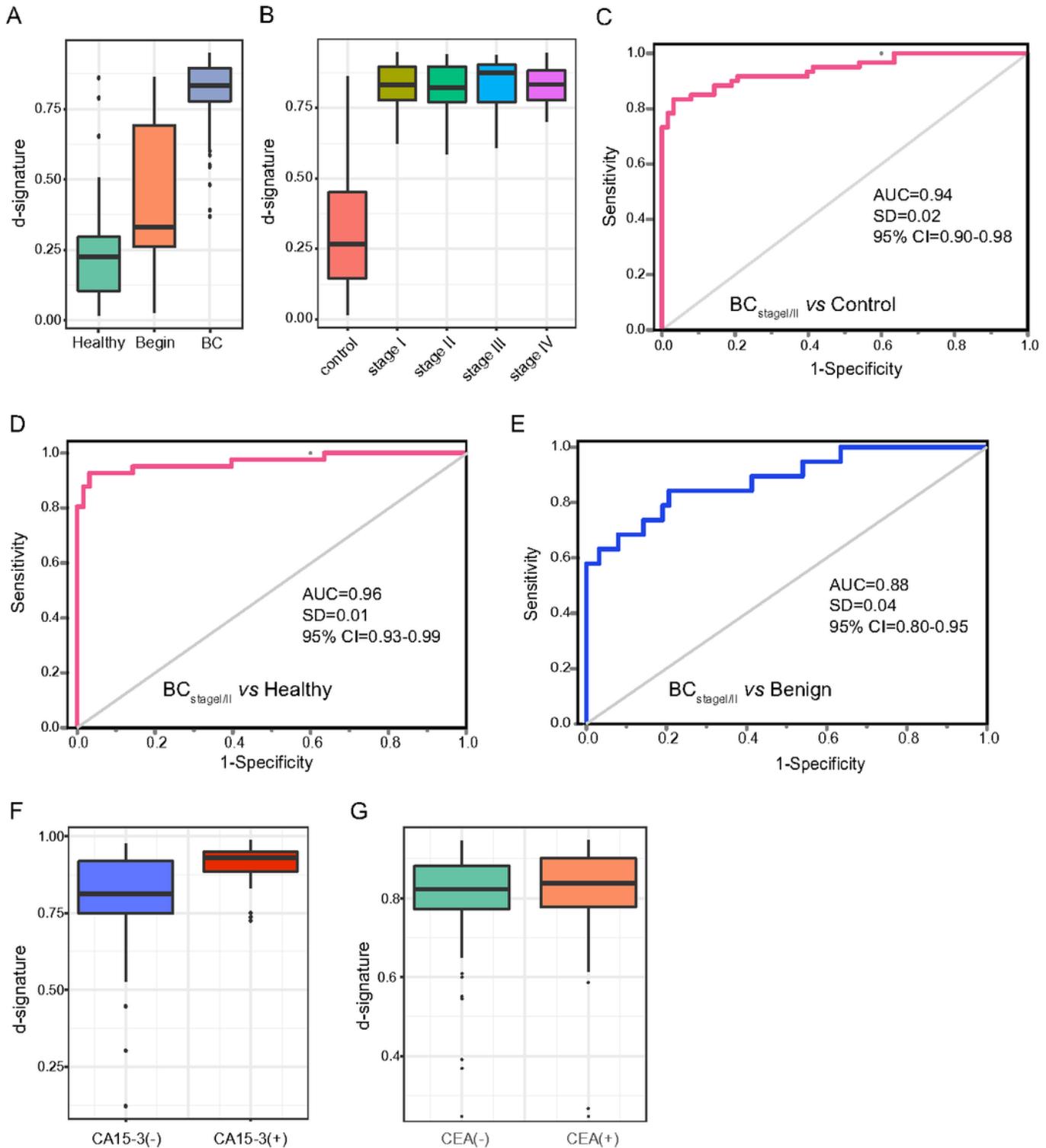


Figure 4

ExLR-based d-signature for the diagnosis of early stage BC (A) ExLR d-signature scores in healthy (n=41), benign (n=19) and BC (n=112). (B) ExLR d-signature scores in BC patients with stage I (n=28), II (n=35), III (n=15) and IV (n=34). (C-E) ROC for the performance of the exLR d-signature in BC with early stages (Stage I/II) compared to control (C), healthy (D) and (benign). (F, G) ExLR d-signature scores in BC patients with different serum CA15-3 (F) and CEA (G) statuses. AUC, area under the curve. SD, standard deviation. CI, confidence interval.

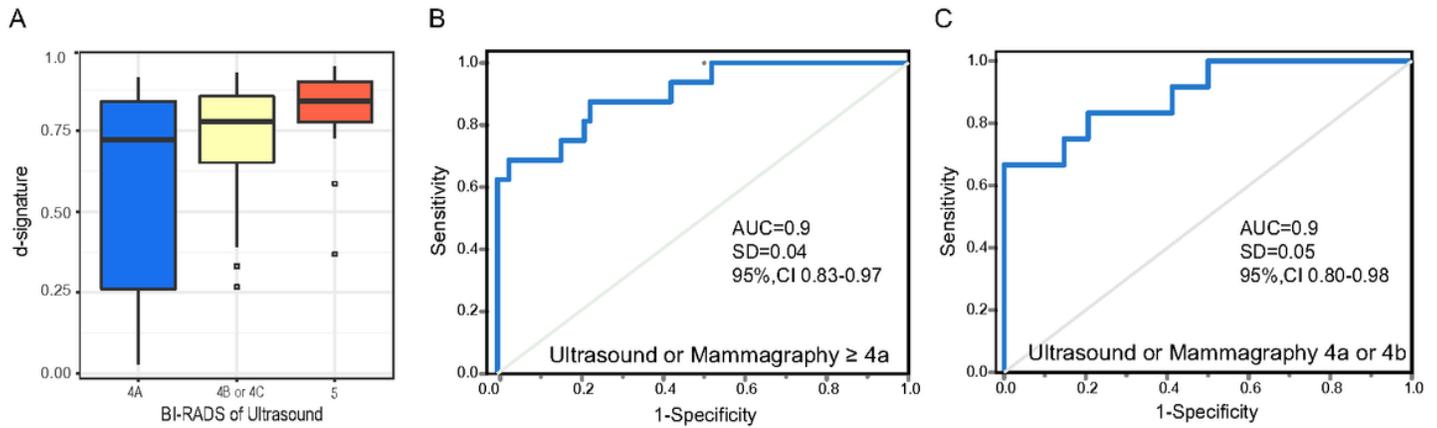


Figure 5

Combined imaging results with exLR-based d-signature for BC diagnosis. (A) ExLR d-signature scores in BC patients with different BI-RADS scores. (B, C) ROC for the performance of exLR d-signatures in BC with ultrasound or mammagrophy ≥ 4a (B) and 4a or 4b (C).

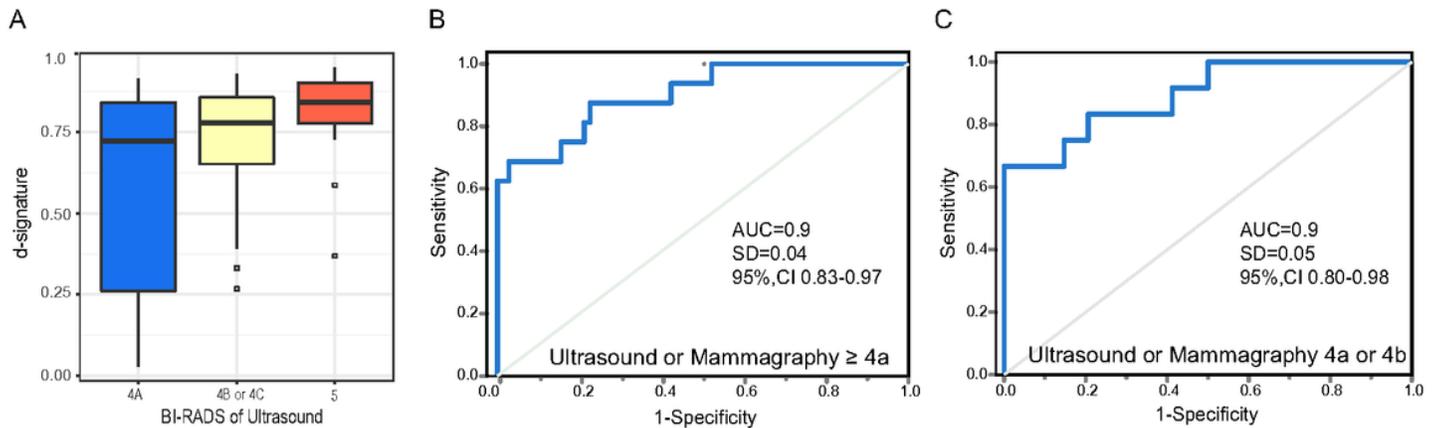


Figure 5

Combined imaging results with exLR-based d-signature for BC diagnosis. (A) ExLR d-signature scores in BC patients with different BI-RADS scores. (B, C) ROC for the performance of exLR d-signatures in BC with ultrasound or mammagrophy ≥ 4a (B) and 4a or 4b (C).

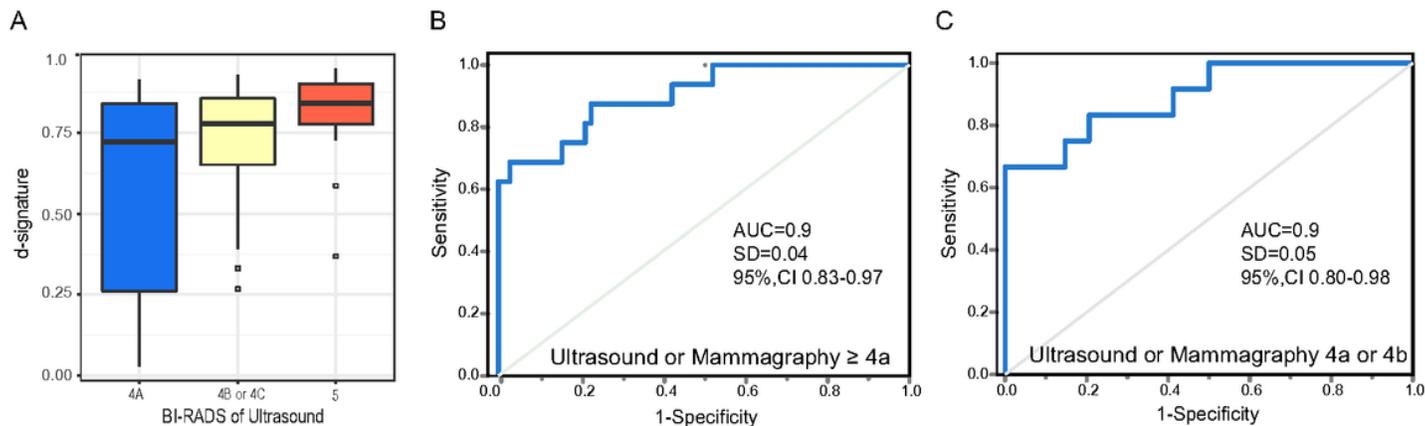


Figure 5

Combined imaging results with exLR-based d-signature for BC diagnosis. (A) ExLR d-signature scores in BC patients with different BI-RADS scores. (B, C) ROC for the performance of exLR d-signatures in BC with ultrasound or mammography $\geq 4a$ (B) and 4a or 4b (C).

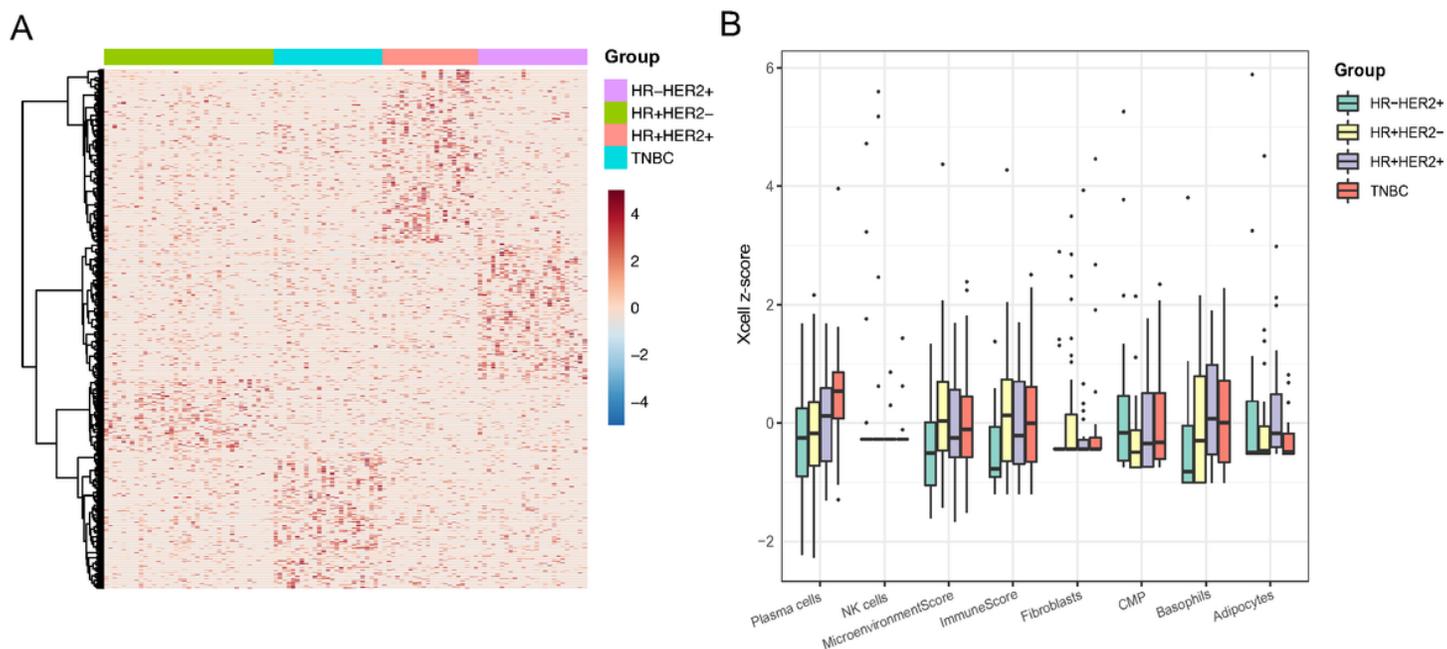


Figure 6

ExLR profiles and immune cell fraction in different BC subtypes (A) Heatmap of the specific exLRs for different BC subtypes. Each column represents an individual sample, and each row represents an exLR. The scale represents the expression values. (B) Comparison of different cell types from the exLR-seq data among different BC subtypes. Only significant differences between these two groups are shown.

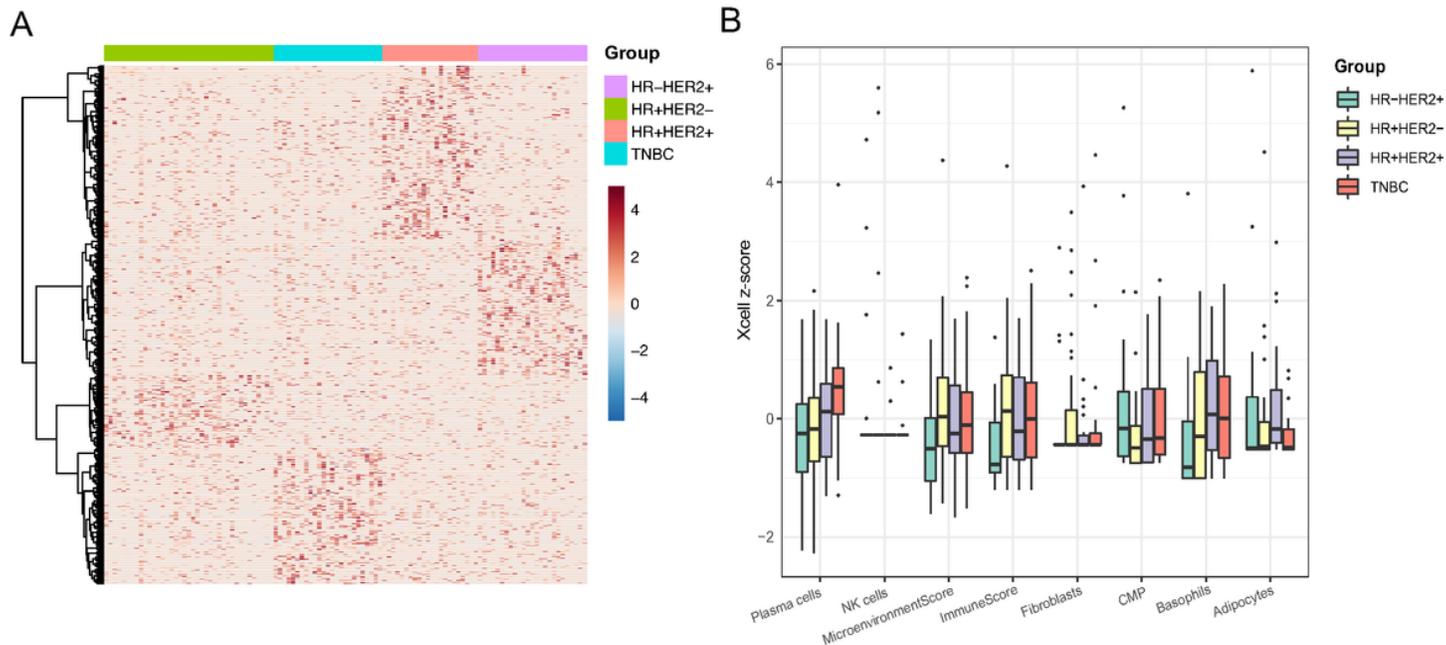


Figure 6

ExLR profiles and immune cell fraction in different BC subtypes (A) Heatmap of the specific exLRs for different BC subtypes. Each column represents an individual sample, and each row represents an exLR. The scale represents the expression values. (B) Comparison of different cell types from the exLR-seq data among different BC subtypes. Only significant differences between these two groups are shown.

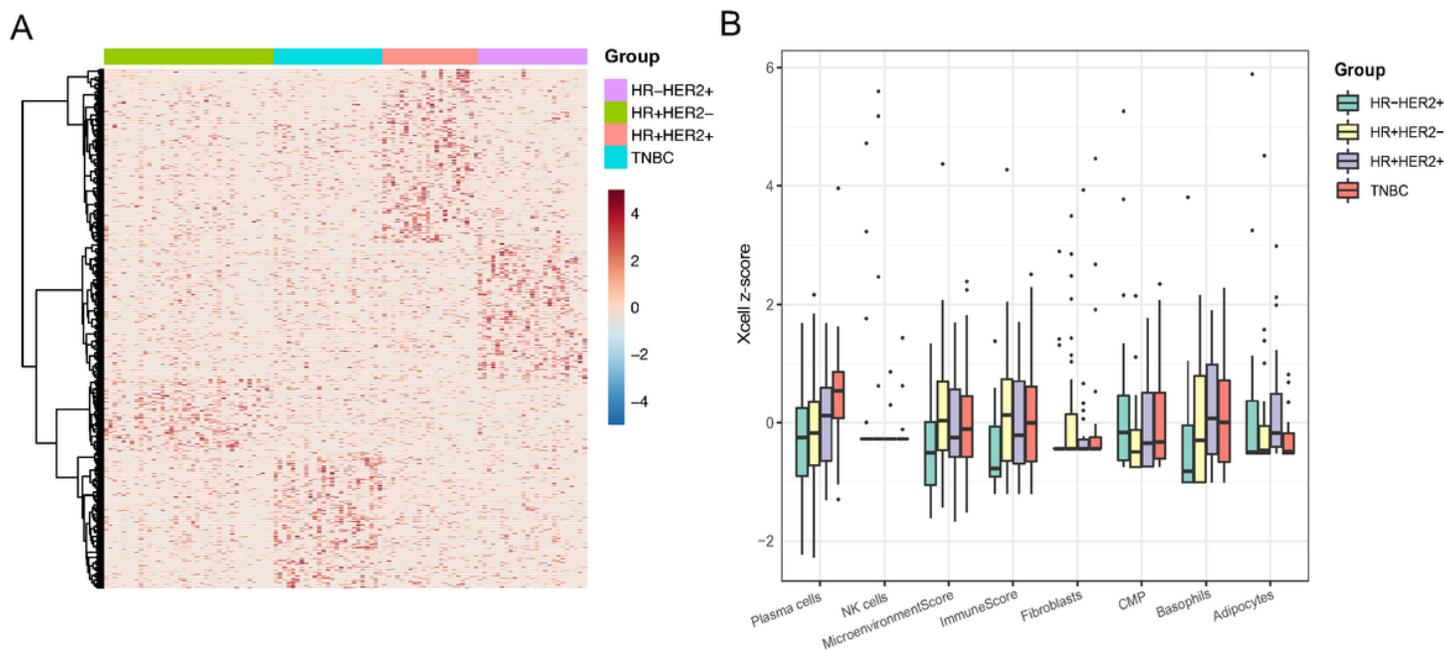


Figure 6

ExLR profiles and immune cell fraction in different BC subtypes (A) Heatmap of the specific exLRs for different BC subtypes. Each column represents an individual sample, and each row represents an exLR.

The scale represents the expression values. (B) Comparison of different cell types from the exLR-seq data among different BC subtypes. Only significant differences between these two groups are shown.

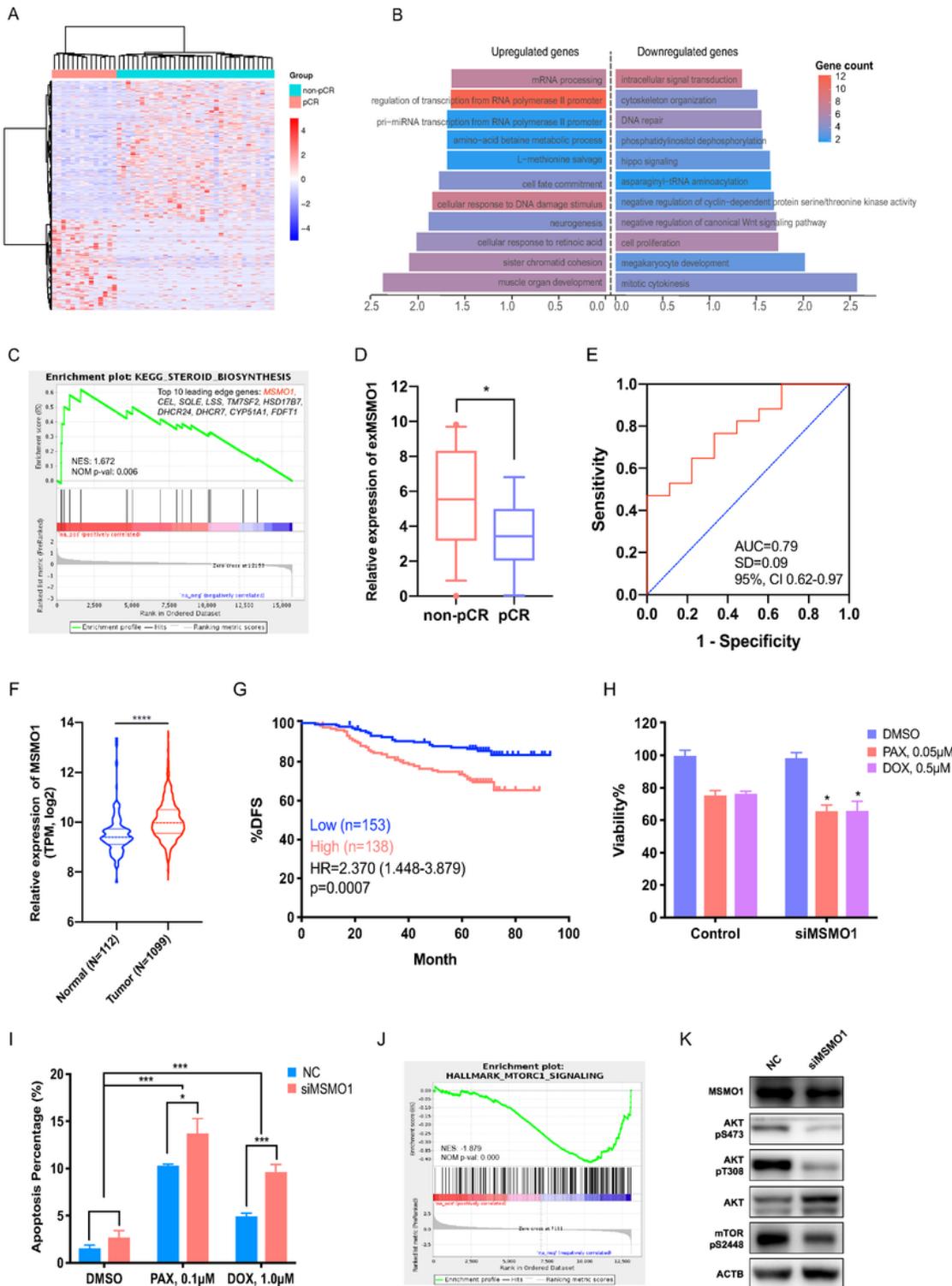


Figure 7

Plasma exMSMO1 as a predictive biomarker for neoadjuvant chemotherapy of BC. (A) Heatmap of different exLRs expressions between pCR (n=14) and non-pCR (n=34) groups. (B) KEGG pathway enrichment analysis for the differentially expressed exLRs of (A). (C) The steroid biosynthesis pathway

was enriched in the non-pCR group by gene set enrichment analysis (GSEA). (D) Comparison of exMSMO1 between pCR and non-pCR groups. (E) ROC for the performance of exMSMO1 in predictive neoadjuvant chemotherapy treatment efficacy of BC. (F) Higher expression of MSMO1 in BC tumor tissue compared to adjacent normal in the TCGA database. (G) Kaplan-Meier survival analysis (log-rank test) of disease free survival of BC patients with low (n = 157) or high (n = 138) MSMO1 expression. (H) MDA-MB-231 cells made deficient in MSMO1 by the siRNAs pool (siMSMO1-1, siMSMO1-2) were treated with the indicated chemotherapy drugs. Viability data from 3 independent experiments were normalized to control-transfected cells. (I) MDA-MB-231 cells were treated with MSMO1 siRNAs pool followed with PAX or DOX for 24 h, and apoptosis was analyzed with the FACS assay. (H, I) Only significant differences were shown. Each column represents averaged results. Bars, SDs. (J) Enrichment plots of the hallmark mTORC1 signaling pathway in MSMO1 deficient cells compared to controls, as identified by GSEA. (K) MDA-MB-231 cells were transfected with the MSMO1 siRNAs pool. Cell lysates were immunoblotted as shown. Abbreviations: ROC, receiver operating characteristic; AUC, area under the curve; SD, standard deviation; CI, confidence interval; DMSO, dimethylsulfoxide; PAX, paclitaxel; DOX, doxorubicin. *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001.

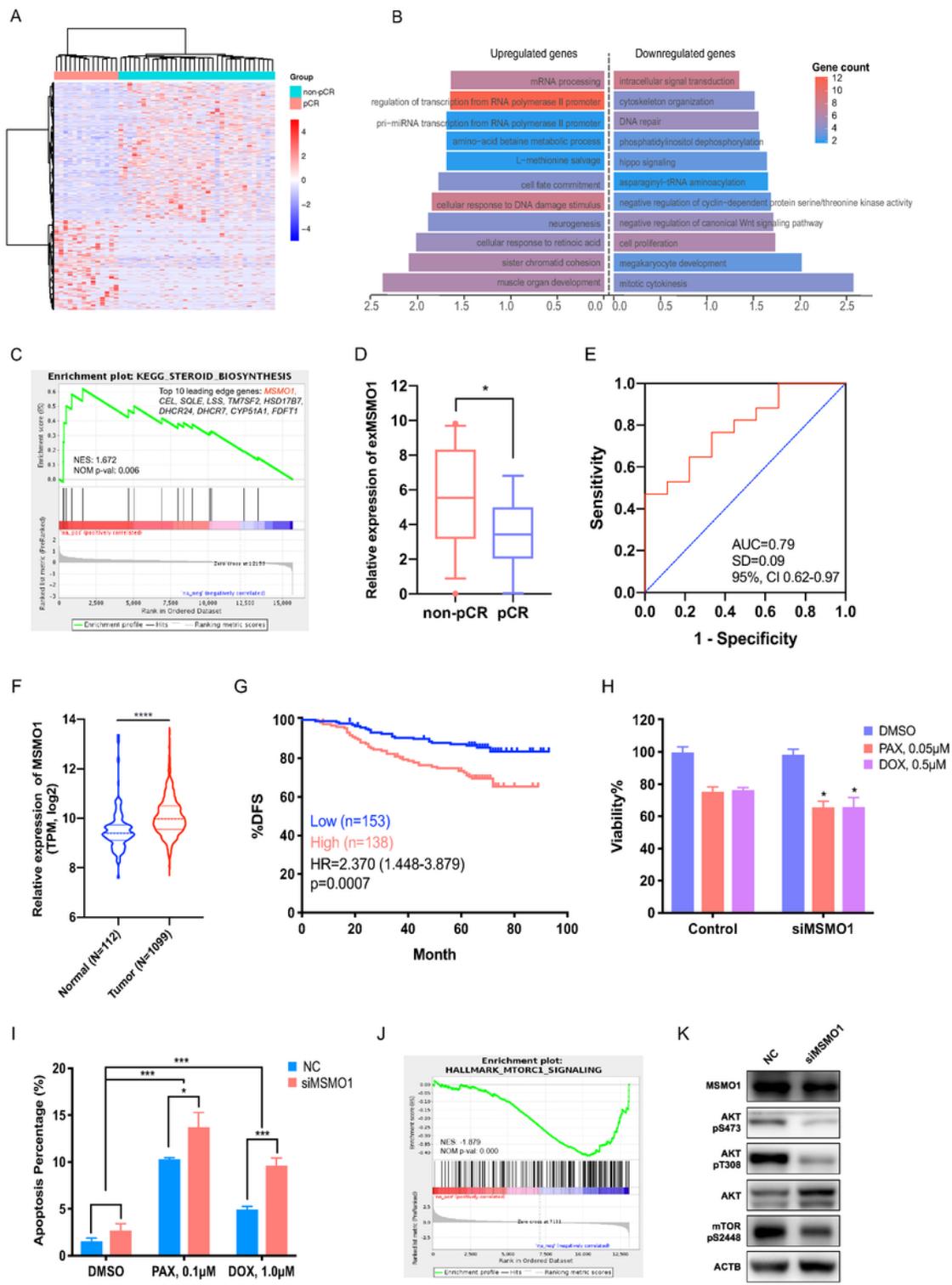


Figure 7

Plasma exMSMO1 as a predictive biomarker for neoadjuvant chemotherapy of BC. (A) Heatmap of different exLRs expressions between pCR (n=14) and non-pCR (n=34) groups. (B) KEGG pathway enrichment analysis for the differentially expressed exLRs of (A). (C) The steroid biosynthesis pathway was enriched in the non-pCR group by gene set enrichment analysis (GSEA). (D) Comparison of exMSMO1 between pCR and non-pCR groups. (E) ROC for the performance of exMSMO1 in predictive

neoadjuvant chemotherapy treatment efficacy of BC. (F) Higher expression of MSMO1 in BC tumor tissue compared to adjacent normal in the TCGA database. (G) Kaplan-Meier survival analysis (log-rank test) of disease free survival of BC patients with low (n = 157) or high (n = 138) MSMO1 expression. (H) MDA-MB-231 cells made deficient in MSMO1 by the siRNAs pool (siMSMO1-1, siMSMO1-2) were treated with the indicated chemotherapy drugs. Viability data from 3 independent experiments were normalized to control-transfected cells. (I) MDA-MB-231 cells were treated with MSMO1 siRNAs pool followed with PAX or DOX for 24 h, and apoptosis was analyzed with the FACS assay. (H, I) Only significant differences were shown. Each column represents averaged results. Bars, SDs. (J) Enrichment plots of the hallmark mTORC1 signaling pathway in MSMO1 deficient cells compared to controls, as identified by GSEA. (K) MDA-MB-231 cells were transfected with the MSMO1 siRNAs pool. Cell lysates were immunoblotted as shown. Abbreviations: ROC, receiver operating characteristic; AUC, area under the curve; SD, standard deviation; CI, confidence interval; DMSO, dimethylsulfoxide; PAX, paclitaxel; DOX, doxorubicin. *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001.

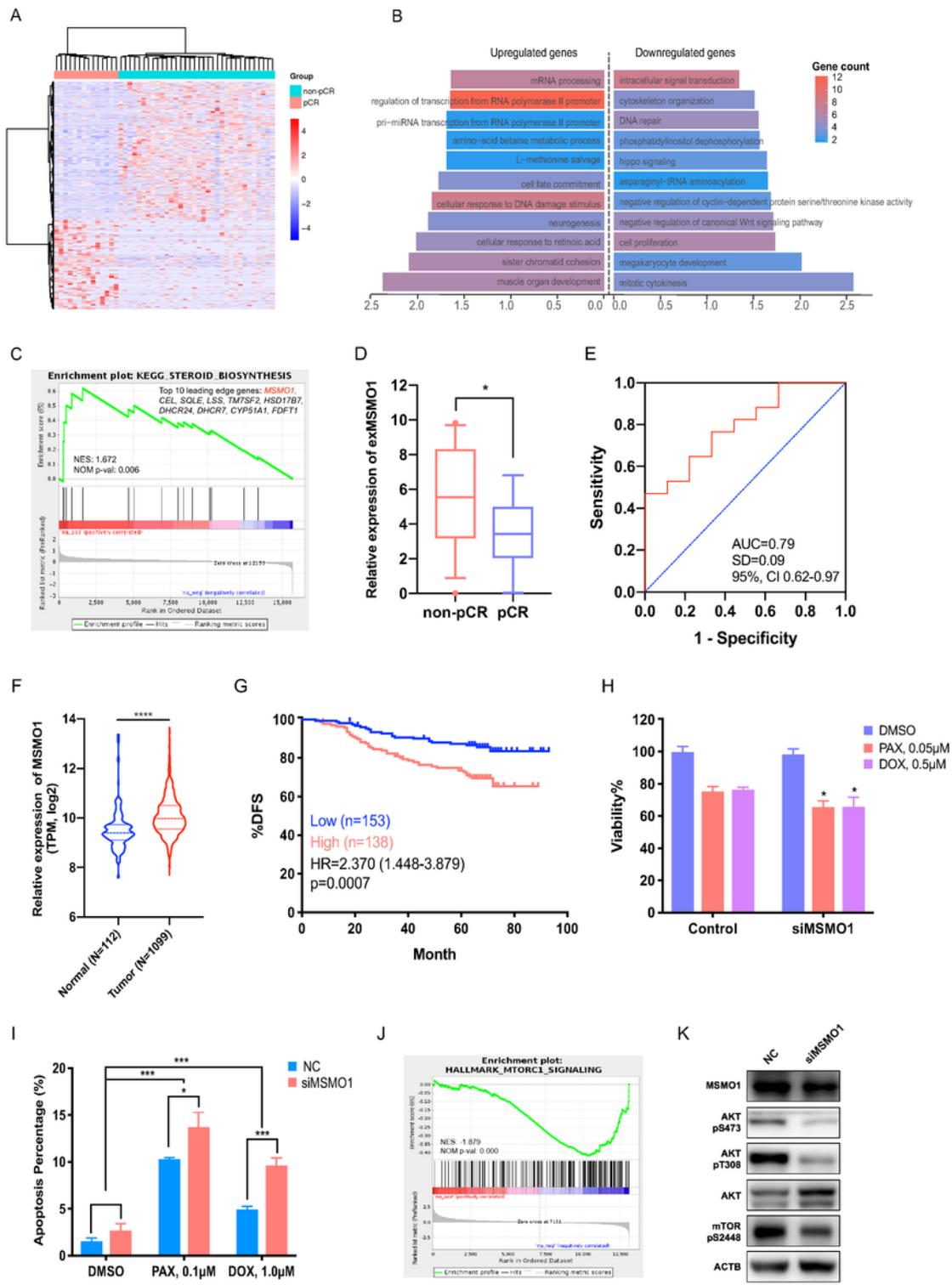


Figure 7

Plasma exMSMO1 as a predictive biomarker for neoadjuvant chemotherapy of BC. (A) Heatmap of different exLRs expressions between pCR (n=14) and non-pCR (n=34) groups. (B) KEGG pathway enrichment analysis for the differentially expressed exLRs of (A). (C) The steroid biosynthesis pathway was enriched in the non-pCR group by gene set enrichment analysis (GSEA). (D) Comparison of exMSMO1 between pCR and non-pCR groups. (E) ROC for the performance of exMSMO1 in predictive

neoadjuvant chemotherapy treatment efficacy of BC. (F) Higher expression of MSMO1 in BC tumor tissue compared to adjacent normal in the TCGA database. (G) Kaplan-Meier survival analysis (log-rank test) of disease free survival of BC patients with low (n = 157) or high (n = 138) MSMO1 expression. (H) MDA-MB-231 cells made deficient in MSMO1 by the siRNAs pool (siMSMO1-1, siMSMO1-2) were treated with the indicated chemotherapy drugs. Viability data from 3 independent experiments were normalized to control-transfected cells. (I) MDA-MB-231 cells were treated with MSMO1 siRNAs pool followed with PAX or DOX for 24 h, and apoptosis was analyzed with the FACS assay. (H, I) Only significant differences were shown. Each column represents averaged results. Bars, SDs. (J) Enrichment plots of the hallmark mTORC1 signaling pathway in MSMO1 deficient cells compared to controls, as identified by GSEA. (K) MDA-MB-231 cells were transfected with the MSMO1 siRNAs pool. Cell lysates were immunoblotted as shown. Abbreviations: ROC, receiver operating characteristic; AUC, area under the curve; SD, standard deviation; CI, confidence interval; DMSO, dimethylsulfoxide; PAX, paclitaxel; DOX, doxorubicin. *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001.

Supplementary Files

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

- [SupplementaryMaterial.docx](#)
- [SupplementaryMaterial.docx](#)
- [SupplementaryMaterial.docx](#)
- [Supplementarytables.docx](#)
- [Supplementarytables.docx](#)
- [Supplementarytables.docx](#)