Original Article
Single-cell analysis reveals one cancer-associated fibroblasts subtype linked to metastasis in breast cancer: MXRA5 as a potential novel marker for prognosis

Huifang Deng[1*], Chengxu Wen[1*], Shangxuan Jiang[1], Yuanhang Yu[1], Jianguo Zhao[2#], Bo Zhang[1#]

[1]Department of Thyroid and Breast Surgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, Hubei, China; [2]Department of Thyroid and Breast Surgery, Wuhan No. 1 Hospital, Wuhan 430022, Hubei, China. *Equal contributors. #Co-corresponding author.

Abstract: Cancer-associated fibroblasts (CAFs) are prevalent in the tumor microenvironment of breast cancer, comprising a group of cell subpopulations with spatial, phenotypic, and functional heterogeneity. Due to the lack of specific markers for CAF subpopulations, their specific mechanisms in breast cancer remain unclear. We identified eight distinct CAF phenotypes in breast cancer using multiple single-cell RNA sequencing datasets and determined distinct transcription factors (TFs) of CAFs through SCENIC analysis. Our study highlights one CAF subtype in breast cancer, FN1+CAF2, associated with metastasis and macrophage polarization. We observed elevated FN1 expression in the stromal tissue of breast cancer patients. Furthermore, FN1 knockdown in CAFs reduced the migration ability of breast cancer cells. We identified a regulatory gene, MXRA5, in CAF2, which may play crucial roles in breast cancer. Our results indicated upregulated MXRA5 expression in breast cancer tissues and CAFs from patients with lymph node metastasis in the following experiment. Overall, our study reveals that the FN1+CAF2 subtype is associated with metastasis and suggests that MXRA5 may be a novel marker mediating the effects of CAF2 on breast cancer metastasis. This study enriches our understanding of CAF heterogeneity and offers new insights for treating breast cancer metastasis.

Keywords: MXRA5, cancer-associated fibroblasts, metastasis, single-cell analysis, breast cancer

Introduction

Breast cancer (BC) is the most common cancer worldwide, accounting for approximately 11.7% of all cancer cases [1]. It is the major cause of cancer mortality and morbidity in women, with an estimated 12% of females diagnosed during their lifetimes, posing a substantial health and socioeconomic burden globally in women [2]. e therapy, and molecular targeted therapy [3]. Despite some effectiveness, a significant number of patients succumb to tumor resistance, recurrence, and metastasis [4]. Effective treatment regimens for metastatic breast cancer are lacking due to a limited understanding of the molecular mechanisms involved in tumor metastasis.

Cancer-associated fibroblasts (CAFs), one of the most abundant cell types in the breast can-

cer microenvironment [5], play key roles in regulating tumor progression, metastasis, and immunosuppression [6]. CAFs are highly heterogeneous, potentially endowing them with diverse biological functions during tumorigenesis [7]. A recent pan-cancer single-cell analysis identified six CAF types in 10 solid cancers, highlighting the complexity of CAF subsets [8]. Mechta-Grigoriou et al. discovered four specific CAF subsets by single-cell sequencing, with CAF-S4 mediating epithelial-mesenchymal transition (EMT) through CXCL12 and TGF-β signals, promoting metastasis to lymph nodes and distal organs [9]. Future studies should explore whether CAF subpopulations possess distinct functional properties. Unfortunately, the specific mechanisms in the tumor microenvironment remain elusive due to the absence of specific CAF subpopulation cell line models.

Single-cell analysis of cancer-associated fibroblasts subtypes in breast cancer
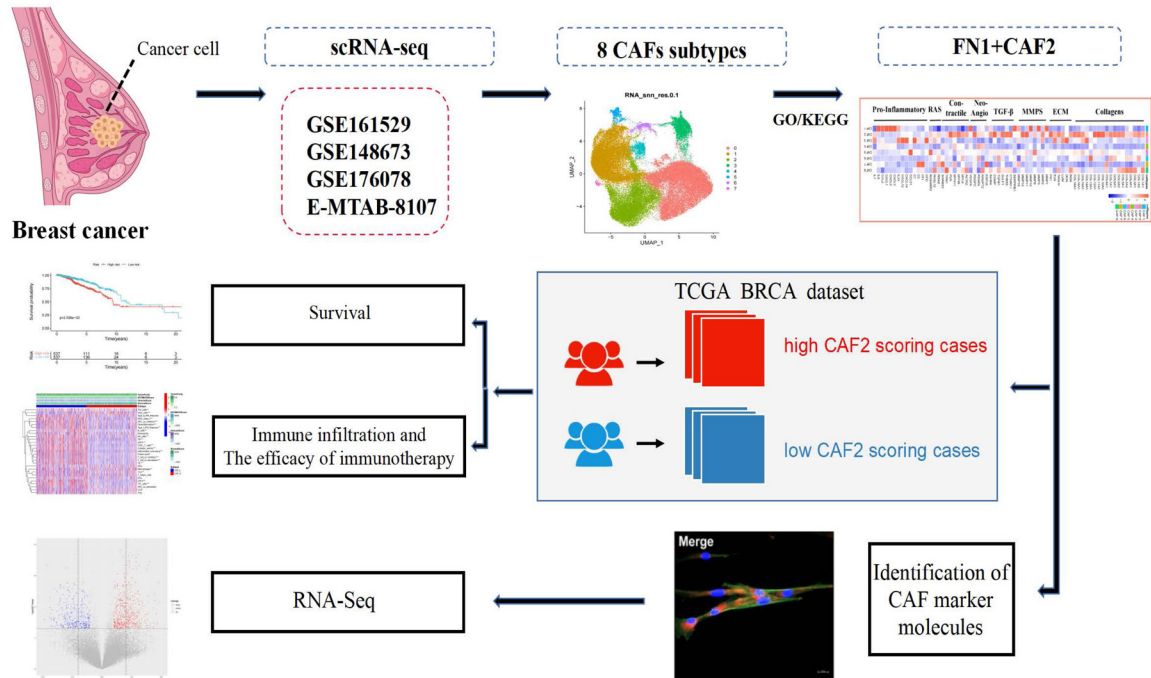
Flow chart. Overall workflow of the present study. Four datasets of single-cell RNA sequencing in breast cancer were collected, and eight CAF subtypes were defined and functionally annotated. The function of FN1+CAF2 subset was analysed in combination with TCGA-BRCA and RNA-seq.

Recent advances in single-cell RNA-sequencing (scRNA-seq) technology and bioinformatics analysis methods have driven interest in studying cellular heterogeneity. scRNA-seq has characterized stromal cellular heterogeneity in various tumor types, including breast cancer [6, 10], lung cancer [11], pancreatic ductal adenocarcinoma [12], prostate cancer [13], head and neck squamous cell carcinoma [14], and melanoma [15] among others. Several studies have highlighted the differential impact of distinct CAF subsets on breast cancer progression. A recent study revealed that CD26+CAFs enhance breast cancer cell invasion by increasing matrix-metalloproteinase activity [10]. Erwei Song et al. confirmed that CD10+GPR77+ CAFs regulate tumorigenesis and chemotherapy resistance in breast cancer by sustaining cancer stemness [16]. The complexity of intercellular communications in the tumor microenvironment and the lack of molecular markers specific for CAF subpopulations make selectively targeting CAFs in cancer treatment challenging [17]. As such, this therapeutic avenue remains largely unexplored. Further investigation into the markers and functions of CAF subsets is essential for accurately targeting subsets that promote tumor progression.

In this study, based on the pivotal roles played by distinct CAF subtypes in the tumor microenvironment, we aimed to identify a subset of CAFs closely associated with breast cancer metastasis and determine whether this subset mediates poor survival outcomes. Utilizing multiple scRNA-seq datasets and the TCGA database, we elucidated the molecular biological characteristics and clinical risk signatures of eight distinct CAF subtypes. Our focus was on the CAF2 subset, characterized by high FN1 expression, which exhibited a strong association with tumor metastasis. We explored the signaling pathways and regulatory genes of this subset which potentially mediates poor prognosis in breast cancer, as shown in the **Flow chart**. These findings contribute to a deeper understanding of CAF subtypes and offer new perspectives for targeted therapy in breast cancer patients.

Methods and materials

Single-cell RNA datasets

Single-cell RNA-seq datasets for breast cancer from GSE161529, GSE148673, GSE176078-GSE and E-MTAB-8107 (https://www.ncbi.nlm.

nih.gov/geo/). Breast cancer transcriptome dataset from TCGA-BRCA (https://portal.gdc.cancer.gov/) and GSE20685.

Patients and tissue samples collection

Human breast tissues were obtained from patients undergoing modified radical mastectomy for invasive breast cancer at Wuhan Union Hospital, Huazhong University of Science and Technology. None of the patients had received neo-adjuvant chemotherapy. All patients provided informed consent prior to tissue collection, and all procedures involving human participants were conducted in accordance with the Declaration of Helsinki. Primary fibroblasts were extracted from the collected specimens, with patient clinical information available in Table S1. Additionally, we constructed a breast tissue microarray (Wuhan Biossci Biotechnology Technology) using 47 breast cancer specimens and paired para-cancerous tissue samples. Detailed clinical information of the breast cancer patients is provided in Table S2.

CAFs isolation and cell culture

CAFs were isolated from fresh breast cancer tissues as previously described [16]. Briefly, the breast cancer tissues were minced into 1 mm^3 pieces and digested with DMEM/F12 medium containing 2 mg/ml type I collagenase (Sigma Aldrich, USA) and 1 mg/ml DNAase I (BioFroxx, DE) for 6-8 hours at 37°C. Undigested tissue fragments were filtered through a 40 μm mesh. The cell suspension was centrifuged at 1800 rpm for 6 minutes to collect the fibroblasts. Subsequently, primary CAFs were cultured in DMEM/F12 with 10% FBS (Gibco, USA, 10099-141) at 37°C in 5% CO$_2$. Fibroblast identification was performed through immunofluorescent staining using specific markers (FAP and α-SMA).

Breast cancer cell lines (MDA-MB-231) were cultured in the recommended media (DMEM, High glucose) supplemented with 10% FBS (Vivacell, C04001, China) at 37°C. For the co-culture model within a 12-well transwell plate with a 0.4 μm pore membrane (Corning, NY, USA) was used. Cancer cells were seeded in the lower chamber, while CAFs were seeded in the upper chamber at a 1:2 ratio. After a 3-day intervention with serum-free culture medium, the cancer cells were harvested for subsequent experiments.

Cell transfection

CAFs in the logarithmic growth phase were seeded into a 6-well plate at a density of 2 × 10^5 cells. At 70-80% confluence, cells were transfected with 50 nM si-FN1 and 50 nM negative control siRNA using Lipofectamine® 3000 (Invitrogen, L3000015, USA), according to the manufacturer's protocol. siRNA targeting FN1 (si-FN1) and the corresponding negative control (si-NC) were designed and synthesized by RiboBio (Guangzhou, China). Cells were collected for subsequent experiments 48 hours post-transfection.

Cell migration assay

Cell migration was evaluated using 8.0 μm pore polycarbonate membrane inserts in 24-well Transwell plates (Corning, NY, USA). Briefly, 800 μl of complete medium was added to the lower chambers, and cancer cells (4 × 10^4) in 200 μl serum-free medium from both co-culture and non-co-culture conditions were added to the upper chambers. After 24 hours at 37°C in 5% CO$_2$, migratory cells were stained with 0.1% crystal violet solution. Cells remaining in the upper chambers were gently removed with a moistened cotton swab. Migratory cells were counted in five randomly selected fields under a light microscope (× 200 magnification).

Wound healing assay

This assay was used to investigate the effect of FN1 in CAFs on cell migration. Cancer cells (1 × 10^5) were seeded in the lower chamber of 12-well Transwell plates (0.4 μm pore size), and CAFs (2 × 10^5) with si-NC or si-FN1 were added to the upper chamber. Once the cancer cells reached a confluent monolayer, a wound was created by manually scraping the cell monolayer with a 200 μl pipette tip. After removing cell debris with PBS, the cells were cultured in 2% FBS medium. Cancer cells in the lower chamber were photographed (× 40 magnification) at 0 and 24 hours, and the wounded area was analyzed using Image J software.

scRNA-seq analysis for CAFs

Single-cell sequencing data underwent conventional downstream processing using the Seurat R package. Cells with mitochondrial content exceeding 30% were excluded. After filtering, 560,452 cells were retained for downstream

Table 1. Primer sequences

Genes	Forward primer (5'-3')	Reverse primer (5'-3')
GAPDH	TCAAGAAGGTGGTGAAGCAGG	TCAAAGGTGGAGGAGTGGGT
MXRA5	CCTTGTGCCTGCTACGTCC	TTGGTCAGTCCTGCAAATGAG
COL6A3	GATGGGCCAGCAAGTAACCT	ATCTTGGGAAGCGACTCTGC
GJB2	CCGACGCAGAGCAAACCG	TTTGCAGCCACAACGAGGAT

analysis. Cell subsets were annotated using information from the GSE161529 dataset. This involved a software package that annotates cell types by comparing the similarity of unknown datasets to reference datasets with known cell type labels. Fine annotation of CAF cell subsets was performed manually. Results were visualized using uniform manifold approximation and projection (UMAP) downscaling. For CAF subpopulations, GO and KEGG analyses were conducted using clusterProfiler, and transcription factors were predicted using SCENIC.

Model construction

Differential gene expression analyses ($P <$ 0.05) for cells within this specific CAF isoform and all other cells using the FindMarkers function in Seurat. Marker genes were identified for this CAF subtype as signature genes during integrated clustering analysis. Expression levels of eight signature genes unique to different CAF subtypes were obtained from the TCGA-BRCA and GSE20685 RNA-seq datasets in breast cancer after normalizing FPKM values. CAF subtype-based marker genes were used to construct prognostic models through univariate and lasso screening of prognosis-related variables. Survival curves for high CAF subtypes were compared using survival data. Immune cell infiltration in the high and low CAF2 score groups in TCGA-BRCA was assessed using the CIBERSORT tool. The ESTIMATE algorithm was applied to the normalized expression matrix to estimate stromal and immune scores for each breast cancer sample. The area under the curve (AUC) was calculated to assess the performance of the logistic regression model, and the "timeROC" R package was used to compute AUC values for each model.

Cellular communication

The cellchat R package, designed for analyzing cell-cell communication, was used to explore diverse cellular communication networks. This analysis examined receptor-ligand interactions among various cell types or subtypes, enabling deduction of their internally regulated signaling pathways. Assessment of ligand-receptor gene interactions among different cell types was conducted by integrating the L2 paradigm number of ligand-receptor interactions and the activity fraction of downstream transcription factors, calculated using the built-in Gene Set Enrichment Analysis (GSEA) algorithm.

RNA-sequencing of CAFs

For RNA-seq analysis, total RNA from primary CAFs was extracted from breast cancer tissues with (LN+CAF group, n = 3) and without (LN-CAF group, n = 3) lymph node metastasis. A cDNA library was constructed by Novogene (Tianjin, China), and RNA-seq was conducted using the Illumina NovaSeq 6000 platform (Illumina, USA). RNA-seq data were analyzed using the DESeq R package.

Quantitative real-time PCR

Total RNA extraction was performed using TriZol Reagent (Invitrogen, Carlsbad, CA, USA, 15596-018). The extracted RNA was reverse transcribed to cDNA using HiScript III RT SuperMix (Vazyme, Nanjing, China, R312-01), following the manufacturer's instructions. Polymerase chain reaction (PCR) was conducted using cDNA, primers, and SYBR Green Master Mix (Vazyme, Nanjing, China, Q311-01), following the Bio-Rad iCycler guidelines (CFX96, California, USA). Relative mRNA levels of target genes were determined using the 2-ΔΔCt method, with the expression of GAPDH serving as a reference for normalization. Primer sequences are provided in **Table 1**.

Tissue microarray and Immunohistochemistry

A tissue microarray with 47 paired breast cancer tissue samples and adjacent cancer tissue was used to evaluate FN1 protein expression by immunohistochemistry (Biossci, Wuhan, China). Following deparaffinization and antigen retrieval, the microarrays were incubated sequentially with primary antibodies targeting FN1 (ProteinTech, 66042-1-Ig, 1:500) and MXRA5 (ProteinTech, 25472-1-ap, 1:300), fol-

lowed by HRP-conjugated secondary antibodies. Images were acquired using NDP.view software (NanoZoomer®S360, Hamamatsu, Japan), and Integrated Optical Density (IOD) mean was analyzed using Image Pro Plus 6.0 for quantitative histochemical staining statistics. Five cases that received neoadjuvant chemotherapy were excluded from this analysis.

Statistical analysis

All bioinformatics statistical analyses were performed using R software (version 4.3.0). The Wilcoxon test was used to compare two groups. Survival differences were assessed using Kaplan-Meier (K-M) curves and the Log-rank test. Experimental data were analyzed using either paired or unpaired Student's t-test to assess the significant differences in measured data. Each experimental group was replicated in three or more independent experiments. Statistical analyses and visualization were conducted using R package and GraphPad Prism 9.0, with statistical significance being defined as $P < 0.05$ (****$P < 0.0001$, ***$P < 0.001$, **$P < 0.01$ and *$P < 0.05$).

Results

Identification of CAF subtypes based on single cell RNA-seq analysis

Using single-cell RNA-seq (scRNA-seq) analysis, we classified cells into distinct clusters. This classification employed datasets from GSE161529, GSE148673, GSE176078, and E-MTAB-8107. During principal component analysis (PCA), we determined the optimal number of principal components to effectively reduce the dimensionality of the cell groupings. The integration of cancer-associated fibroblasts from four distinct single-cell breast cancer datasets was conducted using the Seurat computational framework. This analysis revealed eight distinct CAF subtypes through high-resolution clustering of the combined data (**Figure 1A**). An examination of CAF cellular composition across these four breast cancer single-cell datasets showed notable variations in the prevalence of different cell types. Notably, CAF1 and CAF2 constituted the majority (**Figure 1B**). The identified eight subclasses consisted of a total of 40 marker genes, with each CAF subtype (CAF1-8) characterized by five specific marker genes (**Figure 1C**).

We then performed functional analyses on the significantly expressed marker genes in these 8 CAF subtypes. The GO enrichment analysis in CAF2 identified functions such as extracellular matrix structural components, actin binding, collagen binding, integrin binding, and others conferring tensile strength (**Figure 1D**). Furthermore, we investigated the phenotypic and functional diversity of pan-CAFs and their various isoforms in breast cancer, following the methodology previously outlined by Galbo PM Jr et al. [18]. Our results indicated that, while collagen gene expression levels were low in CAF1, pro-inflammatory molecules such as CXCL1-3, CCL2, and IL-6 were highly expressed in CAF2. The CAF2 subtype was observed to express distinct patterns of various collagens (COL10A1, COL11A1, etc.), angiogenesis-related genes (ANGPT2, PDGFA, PDGFC), stromal matrix metalloproteinases (MMP2, MMP9), transforming growth factor beta (TGFβ) signaling molecules, and extracellular matrix (ECM)-related genes (**Figure 1E**).

Next, we utilized the SCENIC algorithm to identify distinct isoform transcription factors (TFs) of CAFs and their corresponding target gene regulatory networks in breast cancer (Figure S1A). It was found that CAF2 was enriched for SATB1, SKI, HOXC6, among other factors [19]. Notably, SATB1 was particularly abundant in the CAF2 subpopulation, suggesting its potential role in promoting breast cancer metastasis in this subtype (Figure S1B). These findings offer insights into the primary transcription factors driving or sustaining gene expression patterns in CAF subgroups and support the idea that distinct CAF subpopulations fulfill diverse biological roles.

Assessment of distinct CAF subtype in breast cancer prognosis

We compiled gene signatures related to various CAF cell subtypes in breast cancer and utilized them to assess the prevalence of CAF1-8 in the TCGA-BRCA and GSE20685 datasets. This analysis enabled the identification of cohorts with high and low CAF scores, as outlined in the methods section. We observed that high expression of genes characteristic of CAFs correlated with poor breast cancer prognosis, suggesting that CAFs may promote tumor growth in the breast (**Figures 2A-H**, S2A-H). Our study primarily examined CAF1 and CAF2 cell subtypes, with Area Under the Curve (AUC) values
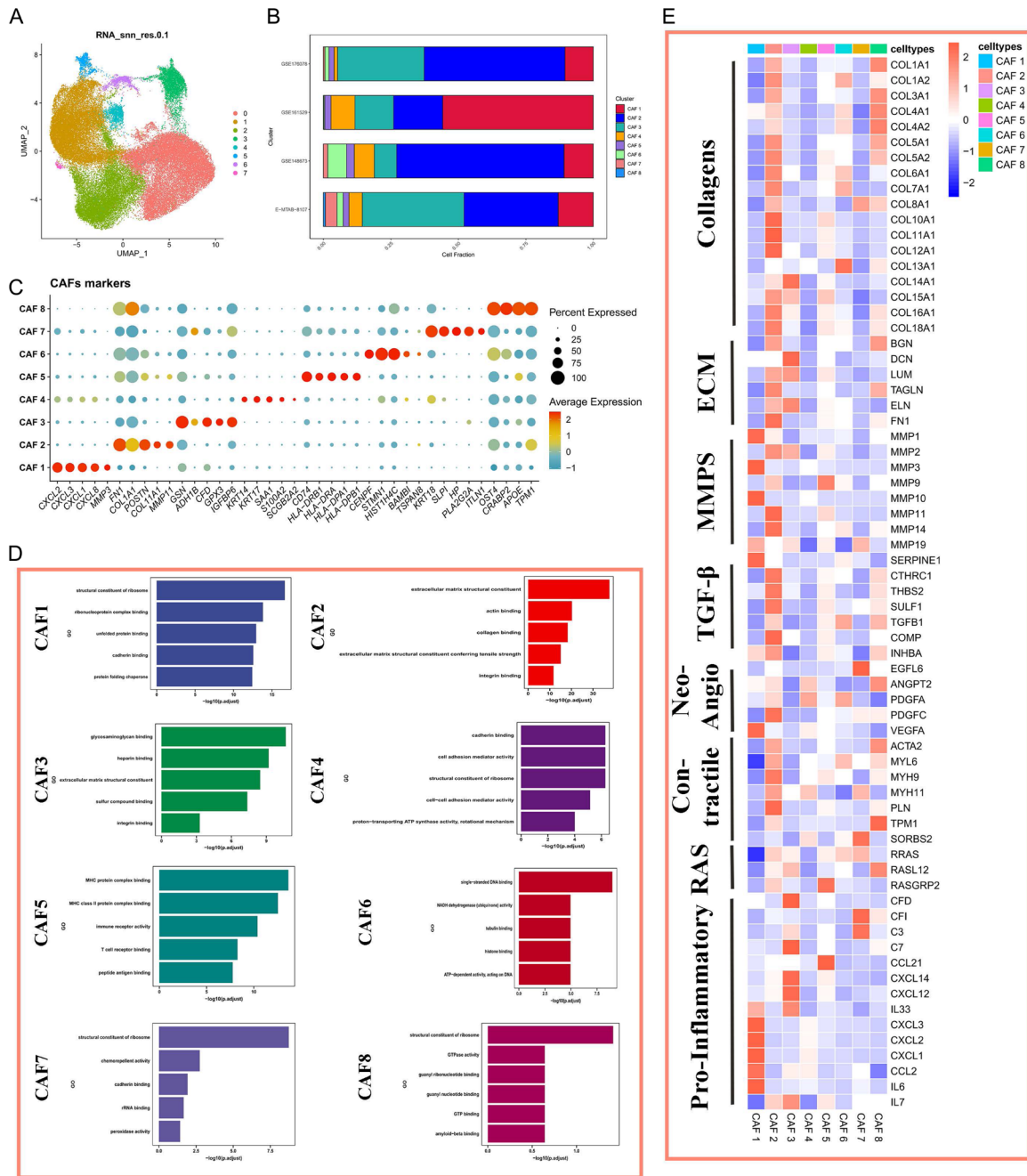
Figure 1. Identification of CAF subtypes based on single-cell RNA-seq analysis. A. UMAP describes subtypes, which are divided into a total of eight subtypes. B. Proportion of different CAF subtypes in breast cancer in datasets: GSE161529, GSE148673, GSE176078 and E-MTAB-8107. C. Dot plot showing expression of caf subtype marker genes in breast cancer. D. GO functional enrichment of different CAF subtypes in breast cancer. E. Heatmap of gene expression associated with selected functions of Collagens, ECM, MMPs, TGF-β, Neo-Angio, Contractile, RAS and Pro-Inflammatory.

of 0.735 and 0.661 for 3-year survival in TCGA-BRCA, and 0.839 and 0.693 for 3-year survival in GSE20685, respectively. The AUC values range from 0 to 1, with values greater than 0.5 indicating better prediction performance of the model. There is no specific threshold for what is considered a good AUC score. Therefore, we concluded that CAF1 and CAF2 subtypes are associated with worse clinical prognosis in breast cancer. Further analysis revealed that
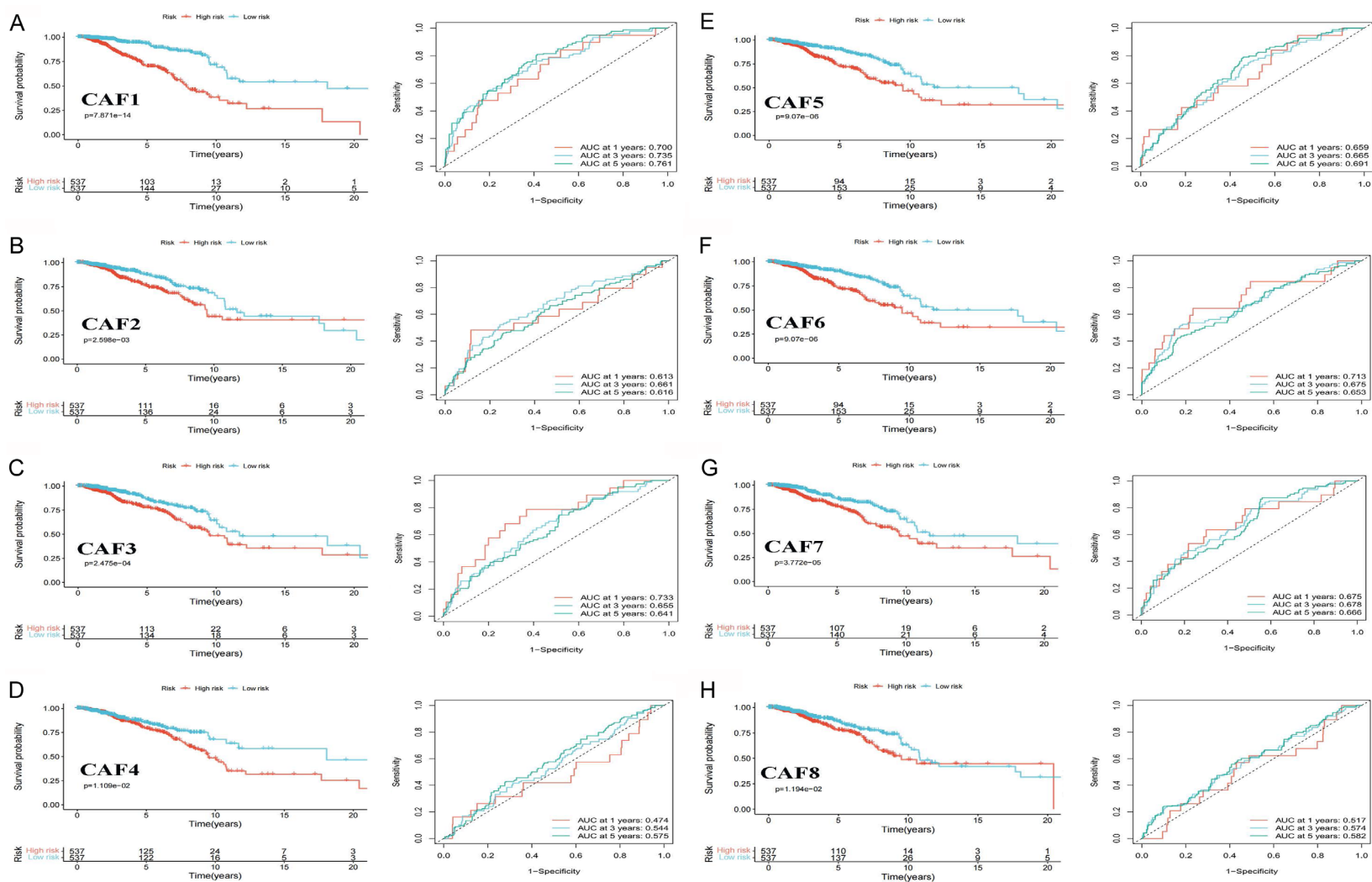
Figure 2. Assessment of distinct CAF subtypes in breast cancer prognosis. A-H. The KM survival curves in TCGA-BRCA showed that among the different CAF subtypes of breast cancer, the group subtypes with higher score had a worse prognosis.

multiple collagen, angiogenesis-related genes, matrix metalloproteinases, transforming growth factor β (TGFβ) signaling molecules, and extracellular matrix (ECM)-related genes were expressed in the high CAF2 group in both datasets. This expression pattern, combined with the poor prognosis of the high CAF2 group, suggests a potential promotional role in breast cancer (Figure S2I).

Interaction between CAF2 and other cellular subpopulations in breast cancer

To comprehensively examine the relationships between CAFs and other cellular subpopulations in breast cancer, we analyzed potential interactions between CAF1-8 and various cell subpopulations (**Figure 3A**). We discovered that the CAF2 subpopulation specifically exhibited increased communication with both tumor cells (Epithelial subpopulation) and myeloid cells compared to other non-CAF subpopulations (**Figures 3B** and S3). Dividing the transcriptome data of TCGA-BRCA into two groups based on their CAF2 scores (high and low), we found that the CAF2 low-hit subgroup showed increased infiltration of innate and adaptive immune cells, including macrophages, compared to the CAF2 high-hit subgroup (**Figure 3C**). The ESTIMATE score algorithm indicated higher scores for the CAF2 high-hit subgroup in our cohort (**Figure 3D**). Using CIBERSORT, we assessed immune cell infiltration differences between the CAF2 high and low-hit subgroups. We observed upregulated M0 and M2 macrophages and downregulated M1 macrophages in the CAF2 high-hit subgroup (**Figure 3E**). These findings suggest that the CAF2 cell subtype might contribute to breast cancer progression by promoting macrophage polarization.

Signaling, intercellular communication, and specific gene signatures in CAF2 subset

To elucidate the signaling pathways involved in intricate intercellular communication in breast cancer, we calculated each signaling system's output and input interaction strengths (**Figure 4A**). Collagen signaling emerged as the predominant signal in fibroblasts. Additionally, along with collagen signaling, pathways including fibronectin 1 (FN1) and midkine (MK) appeared similarly important in facilitating fibroblast interactions with other cells. Based on the findings of our study, it is concluded that

CAF2 holds significant prognostic value in breast cancer. Consequently, our investigation focused on elucidating the signaling pathways involved in CAF2 interactions. The five signaling pathways that are highly likely to interact and significantly contribute to CAF2 interaction are CD46, CD99, COLLAGEN, GAS, and ncWNT, as seen in **Figure 4B-F**. Subsequently, we conducted an analysis of the impact of marker gene interactions within CAF2-associated signaling pathways on the communication between intercellular subpopulations. The findings of the study indicated that there were significant COL6A3-CD44 interactions seen in CAF2-Epithelial and CAF2-Myeloid interactions, and found that MIF-(CD74+CXCR4) interactions had the most pronounced effect on CAF2-B cell contacts (**Figure 4G**). In summary, our findings suggest that FN1 plays a crucial role in maintaining relationships between CAF2 subtypes and other cellular entities. Furthermore, it is plausible that COL6A3 plays a pivotal role in facilitating intercellular communication between subtype 2 breast cancer fibroblasts, tumor cells, and myeloid cells.

Validation of the expression and function of FN1+CAF2 in breast cancer

CAFs were isolated from six breast cancer patients in order to investigate and understand the expression and significance of the CAF2 subgroup in human breast cancer. Subsequently, immunofluorescence analysis was conducted to determine the presence of alpha-smooth muscle actin (α-SMA) and fibroblast activation protein (FAP), both of which serve as universal markers for CAFs (**Figure 5A**). The findings revealed significant upregulation of FN1, COL1A1, and POSTN in the CAF2 subtype, identifying FN1 as a distinguishing marker for this subtype. We conducted a comparative analysis of FN1 expression in 47 paired human breast cancer and normal breast tissue specimens using tissue microarray. This revealed notably higher FN1 expression in stromal tissue of tumor patients (**Figure 5B**, **5C**). Following this, we conducted transcriptome sequencing on primary CAFs derived from breast cancer patients. Single-cell transcriptomic data was categorized into three groups with high CAF2 scores and three groups with low CAF2 scores. We compared differential genes between these groups (**Figure 5D**). Interestingly, all cases with high CAF2 scores were positive for lymph node
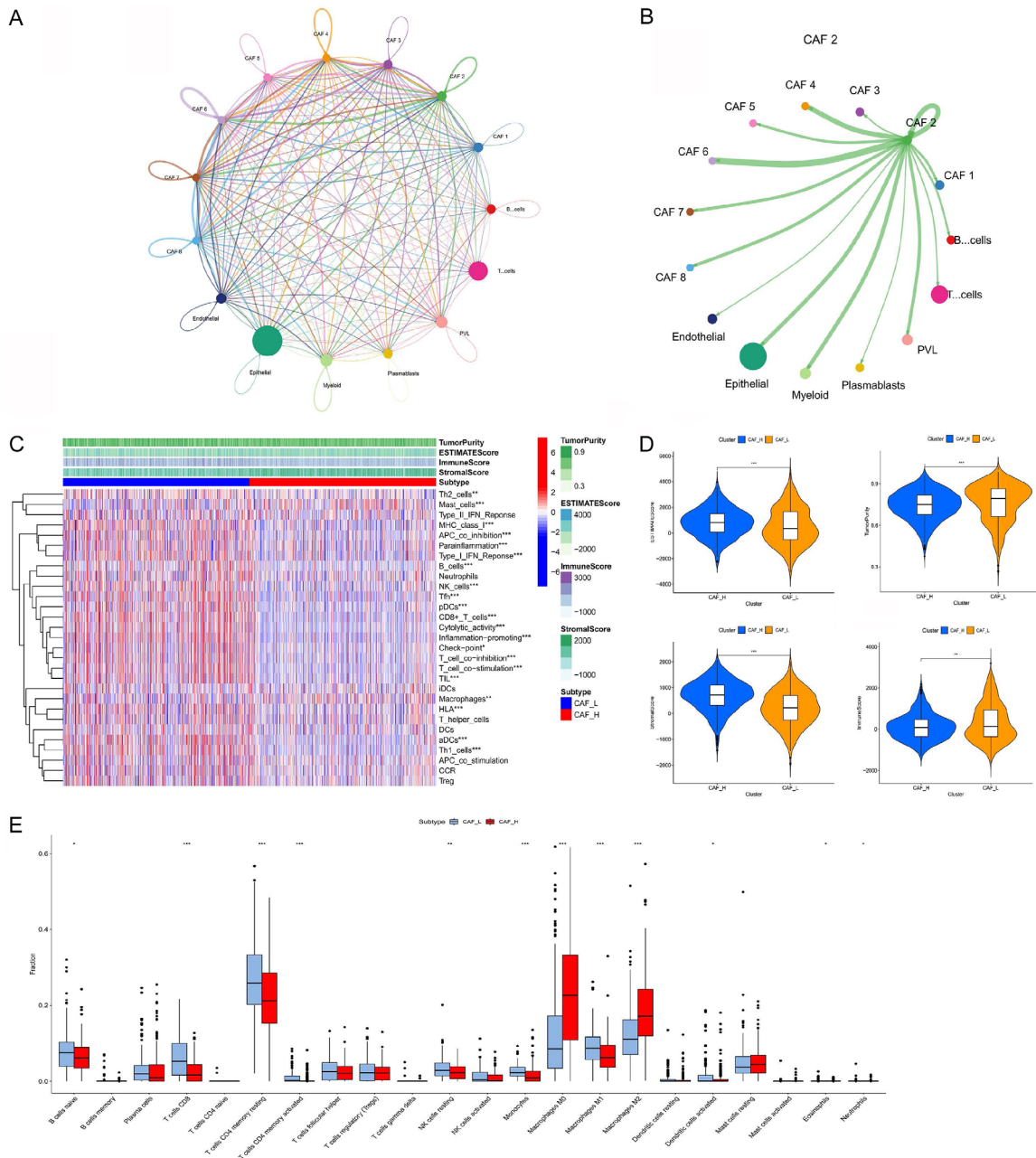
Figure 3. Interaction between CAF2 and other cellular subpopulations in breast cancer. A. Network analysis of the frequency and weight strength of interactions between subsets of 15 units. B. The strength of interactions between CAF2 and other subclass cell populations. C. Microenvironmental phenotypic landscapes of CAF2 subtypes based on estimated number of microenvironmental phenotypic clusters of 24 microenvironmental cell subpopulations calculated by GSEA. D. Scores of stromal, immune, tumor purity, and ESTIMATE between high and low CAF2 scoring groups. E. Box plots comparing differences in immune cell scores between high and low CAF2 scoring groups.

metastasis, while those with low scores showed no such metastasis. Functional enrichment analysis of the data correlated the high CAF2 score group with pathways including peroxisome, gap junction, and ribosome (**Figure 5E, 5F**). To assess the impact of FN1 downregula-

tion in CAFs on breast cancer cell migration, wound healing and Transwell assays were performed. Small interfering RNA (siRNA) significantly reduced FN1 expression in CAFs by qRT-PCR (**Figure 6A**, n = 3). The co-culture model of CAFs and breast cancer cells (**Figure 6B**) dem-
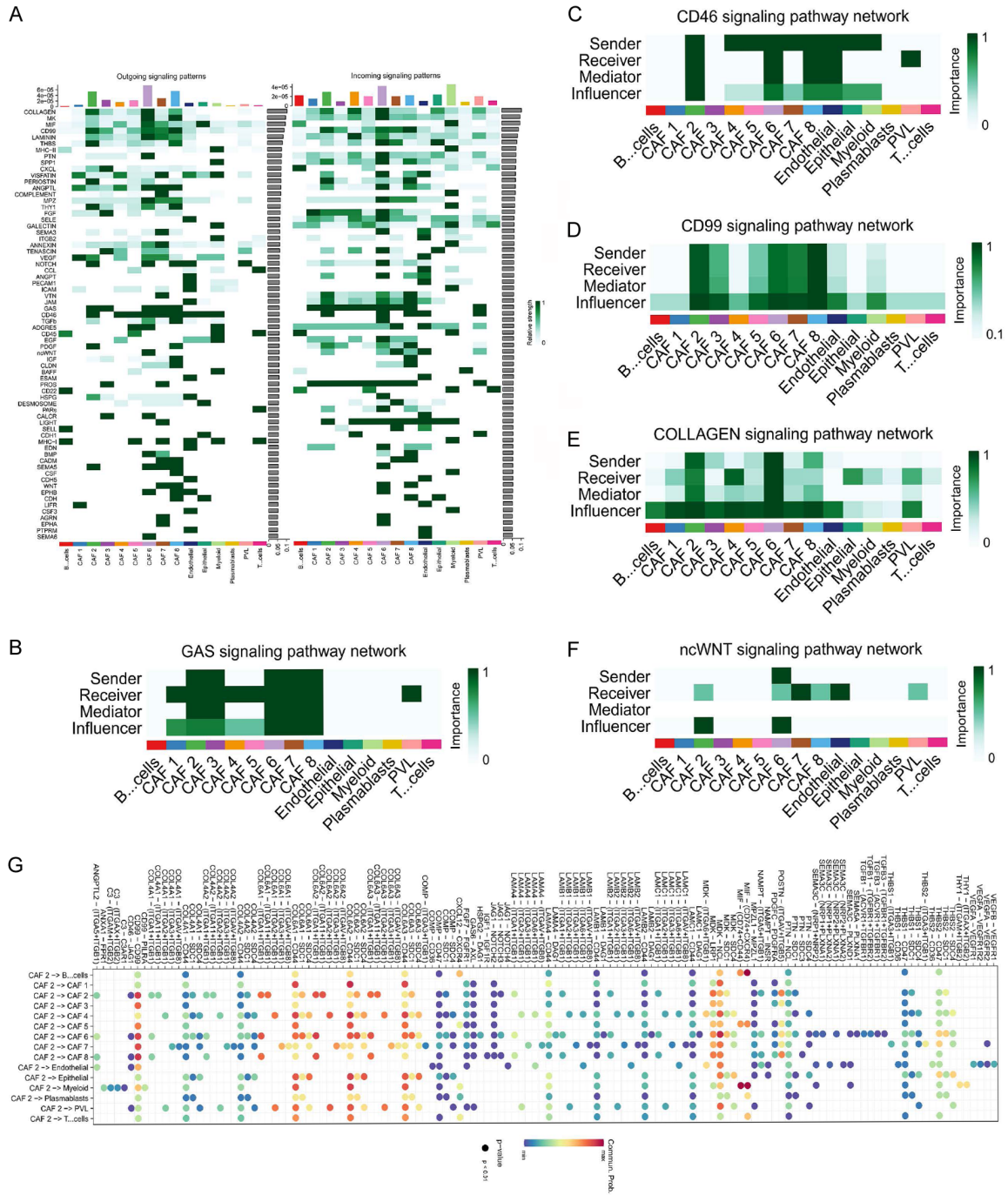
Figure 4. Signalling intercellular communication and specific gene signature in CAF2 subset. A. Input and output signalling pathways for each cell type. B-F. The CD46 signalling pathway, CD99 signalling pathway, COLLAGEN signalling pathway, GAS signalling pathway and ncWNT signalling pathway in CAF2 communicate more closely with the microenvironment. G. Significantly relevant ligand-receptor interactions between CAF2 and most communicating cell types.

onstrated that knockdown of FN1 in CAFs significantly decreased the migration ability of breast cancer cells, as illustrated by the Transwell migration (**Figure 6C**, **6D**) and

wound healing assays (**Figure 6E**, **6F**). Collectively, these results revealed that the expression of FN1 (CAF2 marker) was upregulated in breast cancer tissues, and that FN1
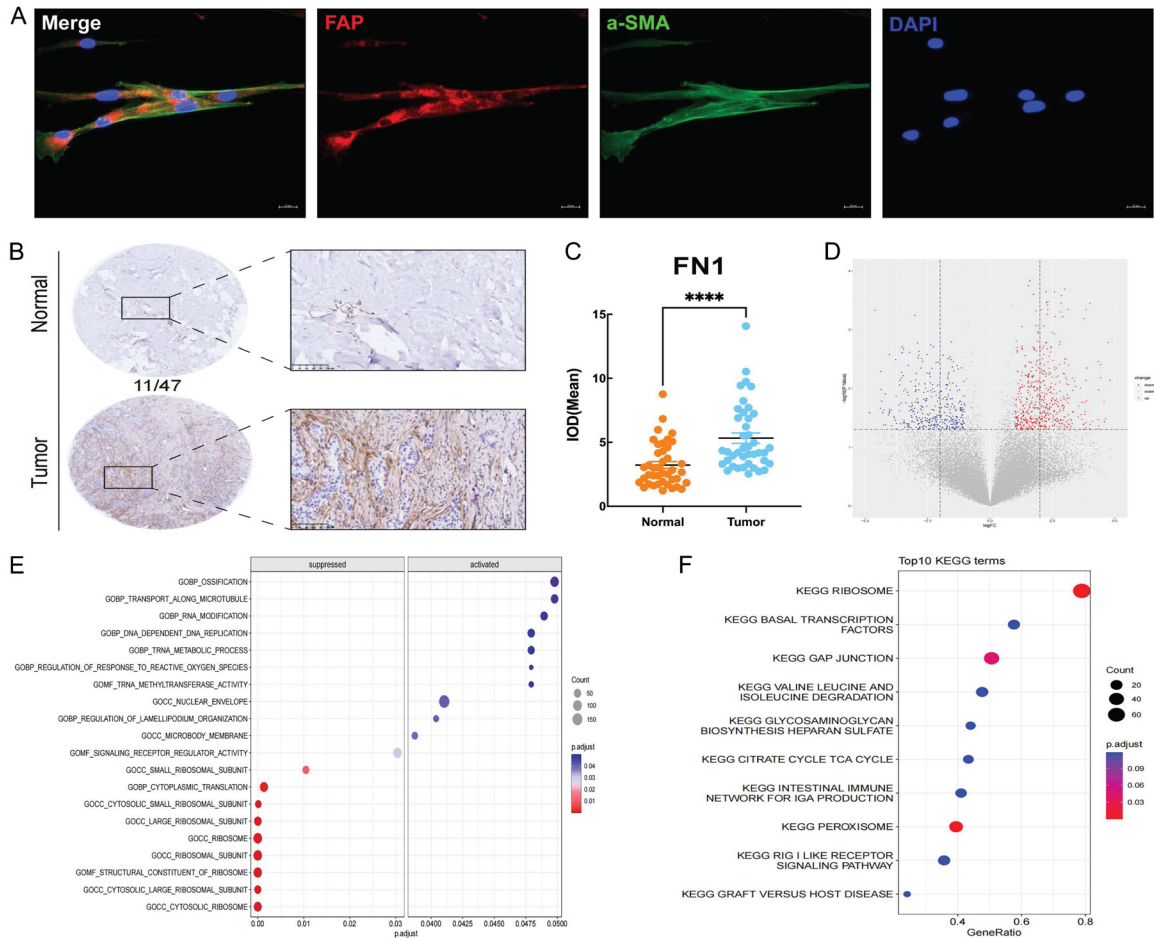
Figure 5. Validation of the expression and function of FN1+CAF2 in breast cancer. A. Cell markers for immunofluorescence identification of CAFs in human breast cancer, bar = 20 µm. B, C. Protein expression of FN1 in 41 human breast cancer samples compared to normal breast cancer samples (****P < 0.0001), bar = 100 µm. D. The transcriptome sequencing data of 6 primary CAF cases were referenced to single-cell CAF2 scoring and divided into 3 high CAF2 scoring cases and 3 low CAF2 scoring cases. Scatter plots indicate differential genes. E, F. GO, KEGG functional enrichment of differential genes obtained from RNA sequencing.

expression may be involved in breast cancer metastasis.

MXRA5 as a novel target in CAF2 subtype mediating tumor metastasis

Focusing on the clinically relevant FN1+ CAF2 subpopulation in breast cancer, we investigated crucial molecular targets for this subpopulation. We used a Venn diagram to assess the overlap between differential genes in high and low CAF2 typing groups in TCGA-BRCA, molecular markers of single-cell CAF2 subpopulations, and differential genes from CAF sequencing data of six human breast cancers. This analysis revealed three key molecules: COL6A3, GJB2, and MXRA5 (**Figure 7A**). Their expres-

sion was verified in TCGA-BRCA, showing high expression in tumors without statistical significance between metastatic and non-metastatic groups, possibly due to limited sample size in the metastatic group (**Figure 7B**). However, mRNA expression of COL6A3, GJB2, and MXRA5 in CAF primary cells was statistically significant between metastatic and non-metastatic groups (**Figure 7C-E**). Kaplan-Meier survival analysis linked elevated MXRA5 expression with poor prognosis (**Figure 7F**).

We then compared the protein expression levels of COL6A3, GJB2, and MXRA5 in both normal breast tissue and breast cancer tissue (**Figure 7H**). Additionally, we analyzed MXRA5 expression using RNA samples from a cohort of
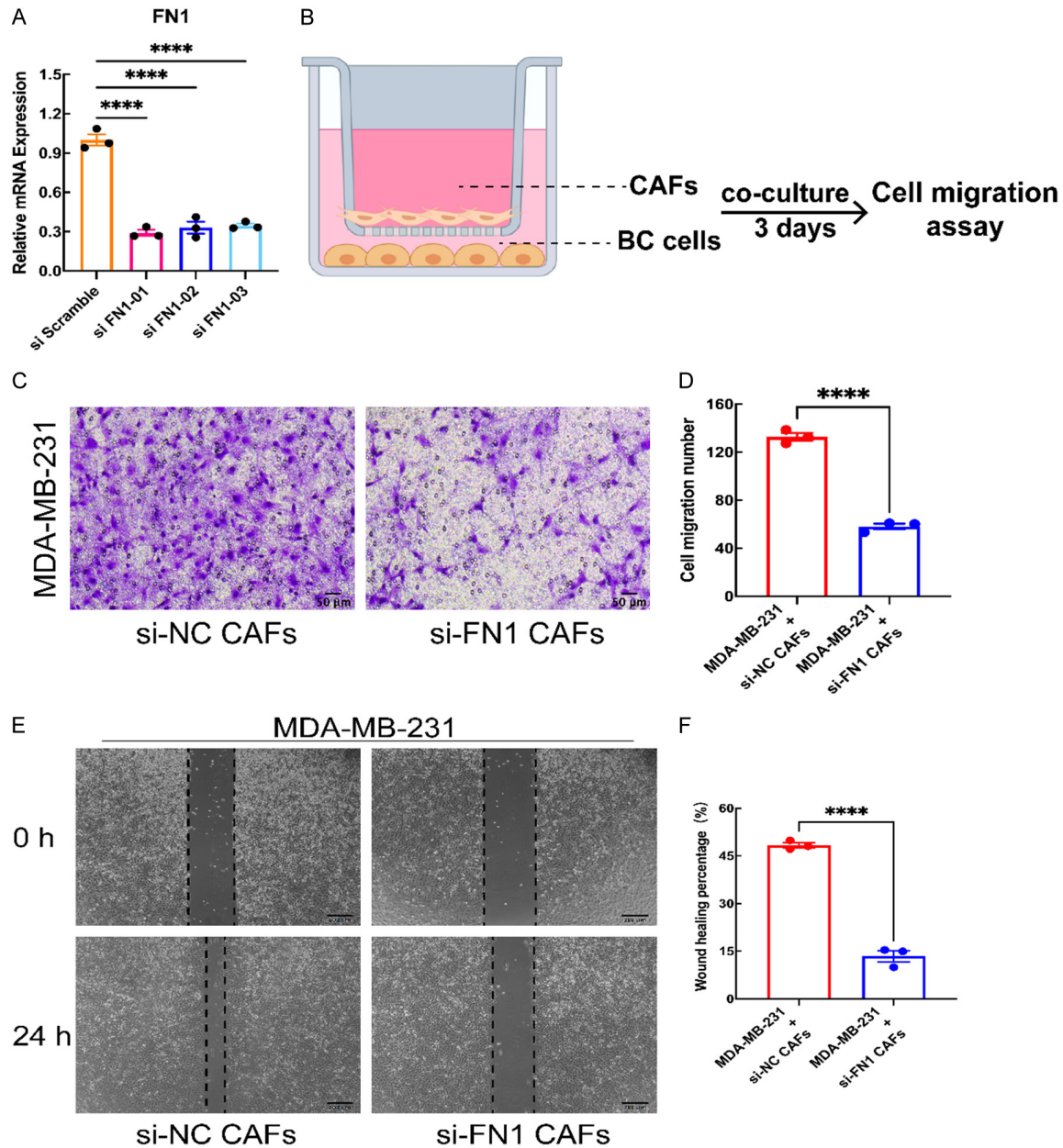
Figure 6. The effect of FN1 downregulation in CAFs on breast cancer cell migration. (A) The knockdown of FN1 in CAFs was verified by qRT-PCR. (B) A co-culture model of breast cancer cell (BC cells) and CAFs. Transwell (C, D) and wound healing assays (E, F) showed that knockdown FN1 in CAFs inhibited the migration of MDA-MB-231 cells. Bar = 50 μm (C), 200 μm (E).

15 individuals diagnosed with breast cancer, revealing a significant upregulation of MXRA5 in breast cancer tissue (**Figure 7G**), the patient's clinical information was shown in Table S5. We also examined MXRA5 expression in 47 human breast cancer tissue samples using immunohistochemistry, which showed that MXRA5 was upregulated in both the stromal and intratumoral regions (**Figure 8A**, **8B**). A more compre-

hensive analysis of high and low MXRA5 expression groups in TCGA-BRCA revealed a positive correlation between MXRA5 and molecules associated with EMT, ECM, TGFβ and other pathways (**Figure 8C-H**), suggesting a potential pro-metastatic function for MXRA5 in human breast cancer. Furthermore, we analyzed the relationship between MXRA5 expression and clinicopathological characteristics of breast
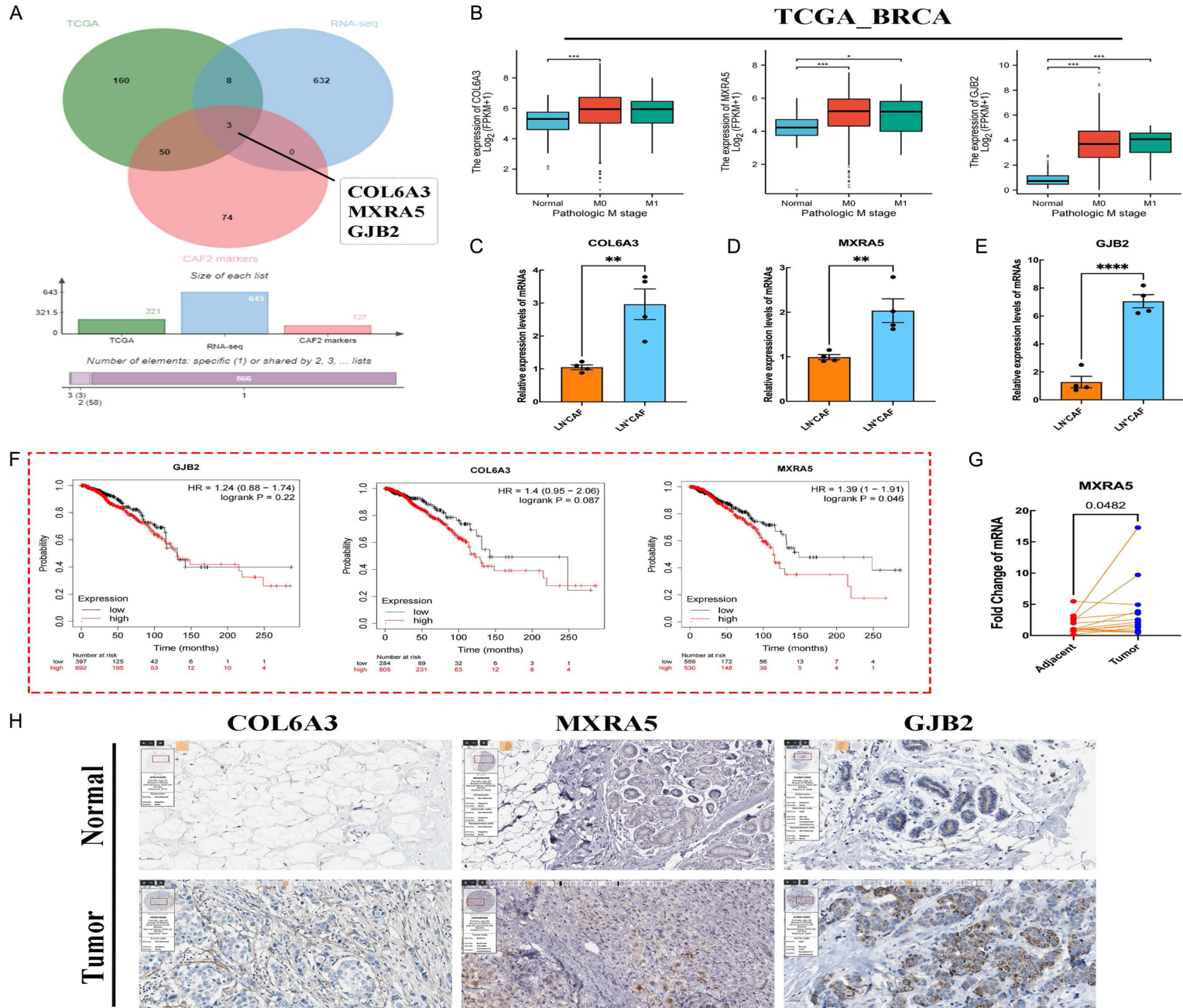
Single-cell analysis of cancer-associated fibroblasts subtypes in breast cancer

Figure 7. MXRA5 as a novel target in CAF2 subtype mediating tumor metastasis. A. Venn diagram showing the intersection of differential genes for high and low CAF2 typing groups in TCGA, differential genes for high and low CAF2 typing groups in 6 CAF progenitors, and CAF2 subpopulation marker molecules from single-cell sequencing. B. Expression of COL6A3, GJB2 and MXRA5 genes in TCGA-BRCA in normal, metastatic and non-metastatic groups. C-E. Expression of COL6A3, GJB2 and MXRA5 genes in metastatic and non-metastatic groups in primary cells of CAF (n = 4, **P < 0.01, ****P < 0.0001). F. Kaplan-Meier survival curves for COL6A3, GJB2 and MXRA5. G. mRNA expression of MXRA5 in adjacent and breast cancer tissues (n = 15, P = 0.0482). H. Protein expression of COL6A3, GJB2 and MXRA5 in normal and breast cancer from the HPA database, bar = 50 μm.
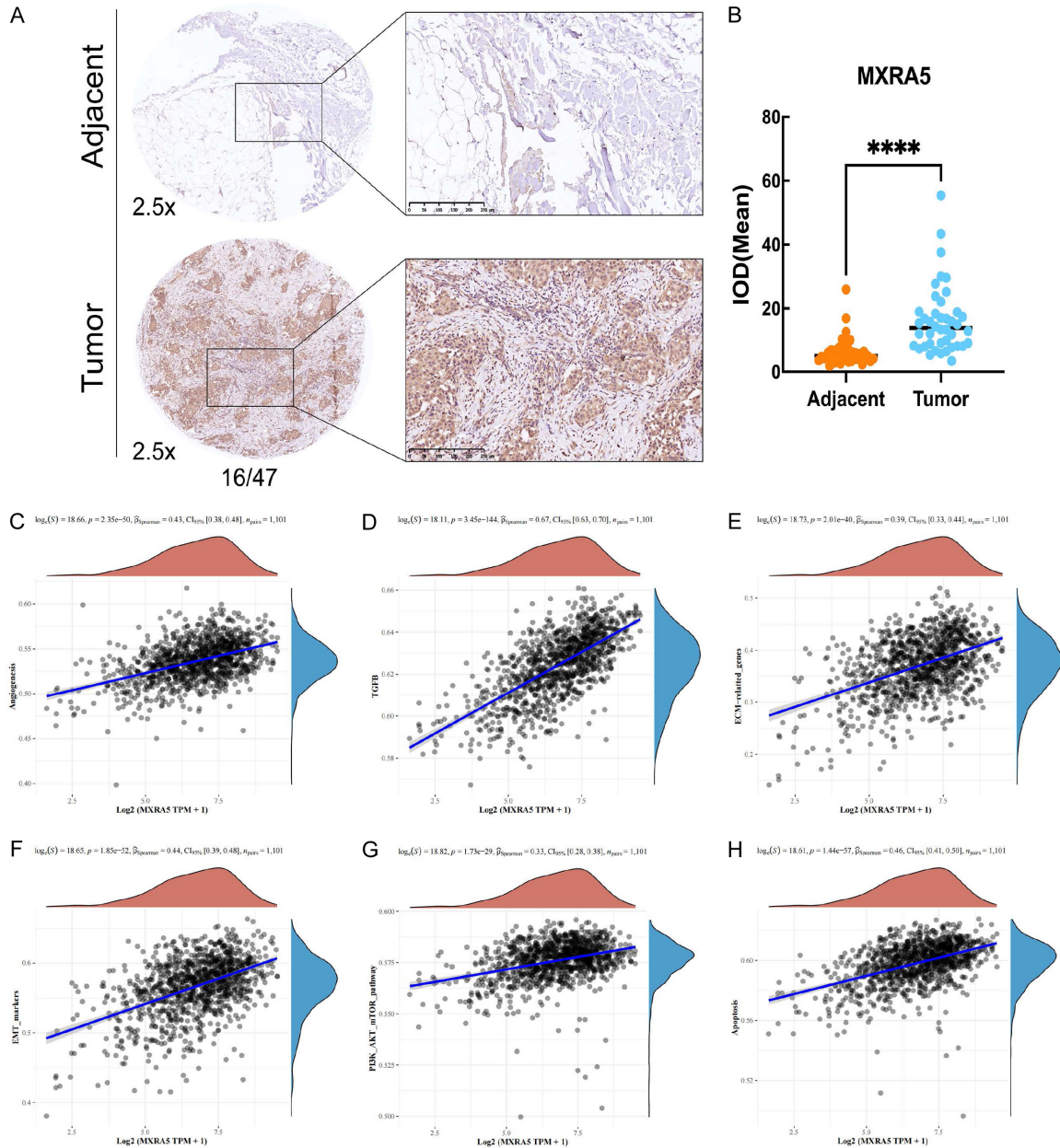
Figure 8. The expression and pro-metastatic pathway of MXRA5 in breast cancer. A, B. The expression of MXRA5 in 47 human breast cancer tissue microarray (n = 42, P < 0.0001), bar = 100 μm. C-H. EMT, TGFB pathway and others were positively correlated with MXRA5 expression.

cancer, revealing significant associations with tumor size (T, P = 0.008), lymph node involve- ment (N, P = 0.005), and HER2 status (P = 0.05), as shown in Table S3. We also investi-

gated the relative risk of MXRA5 in relation to breast cancer prognosis. One-way Cox-regression analysis indicated that MXRA5 could be an independent risk factor, with elevated expression correlating with increased risk of mortality (P = 0.075) and significant correlation with lymph node and distant metastasis (P < 0.001), as indicated in Table S4. These results suggest MXRA5 as a novel molecular target in the CAF2 subtype for regulating breast cancer metastasis.

Discussion

Recently, the heterogeneity and plasticity of cancer-associated fibroblasts have garnered increasing interest among researchers. CAFs play crucial roles in tumor progression and represent the predominant cell phenotype in the breast tumor microenvironment, which encompasses a range of cell subtypes with phenotypic, functional, and spatial heterogeneity [20]. This diversity presents significant challenges in targeting various CAF subsets, arising not only from the lack of specific markers to individually distinguish these CAF subpopulations but also from the obscure molecular mechanisms linking CAF subsets and cancer cells [17, 21]. In our work, we focused on the role of the CAF2 subtype in breast cancer progression, performing systematic classification and prognostic assessment of CAF subsets based on single-cell RNA sequencing (scRNA-seq) data from breast cancer. Furthermore, we identified a metastasis-related CAF subtype and discovered MXRA5 as a significant molecular target, thereby enhancing our understanding of the treatment and pathogenesis of metastatic breast cancer.

Several studies have reported the crucial functions of different CAF subpopulations. Costa et al. [22] identified four CAF subsets (CAF-S1-CAF-S4) in breast cancer subtypes and confirmed that CAF-S1 promotes immunosuppression in triple-negative breast cancers [23]. Interestingly, these four CAF clusters were also found in metastatic lymph nodes of breast cancer [9]. CAF-S1 (FAP^{High} α-SMA^{High}) modulated cancer cell migration and drove the EMT, while the highly contractile CAF-S4 (CD29^{High} α-SMA^{High}) led to cancer cell invasion through ECM remodeling. Markers such as alpha-smooth muscle actin (α-SMA), fibroblast activa-

tion protein (FAP), platelet-derived growth factor receptors α and β (PDGFRα and PDGFRβ), and fibroblast-specific protein-1 (FSP1) are commonly used for CAF identification [21]. We performed immunofluorescence staining on primary CAFs using FAP and α-SMA, revealing that the CAFs were irregularly spindle-shaped. However, due to the limited size of clinical specimens, we were unable to identify each of the eight CAF subtypes in this study. In other tumor types, Elyada et al. identified three distinct CAF subpopulations in pancreatic ductal adenocarcinoma (PDAC): myofibroblastic CAFs (myCAFs), inflammatory CAFs (iCAFs) and antigen-presenting CAFs (apCAFs) [12]. They also found that the iCAF and myCAF populations could interconvert [24], and apCAFs could differentiate into myCAFs, confirming the plasticity and dynamics of CAFs. Different CAF subsets in various tumors exhibit distinct tumor-promoting or tumor-restraining effects. Consequently, further studies are crucial to understand the specific targetable markers and pathways of fibroblast subsets to in the development of more effective CAF-targeted therapies.

In this study, we identified eight CAF subtypes from four breast cancer single-cell datasets using scRNA-seq analysis. A recent study delineated nine CAF phenotypes from tumors of 14 breast cancer patients and identified these CAF subtypes in four additional cancer types [25]. Their results indicated high expression of MMP11, POSTN, and COL1A2 in mCAFs, findings largely consistent with our identification of markers in the CAF2 subgroup. They also reported the association of mCAFs with matrix remodeling and migration by upregulating TGF-β signaling and EMT pathways, a conclusion corroborated by our analysis. Furthermore, our study highlighted the significance of fibronectin 1 (FN1) and midkine (MK) signaling as key pathways in fibroblast interactions with other cells. FN1, a multifunctional extracellular matrix protein, interacts with various components on the cell surface and in the extracellular matrix [26]. As Peng Z et al. reported, fibronectin from fibroblasts promotes tumor growth and angiogenesis through interactions with liver cancer cells [27]. Additionally, a multi-omics analysis indicated that FN1 derived from tumor-associated macrophages and fibroblasts facilitates metastasis in hepatocellular carci-

noma [28]. MK, a growth factor overexpressed in several human carcinomas [29], plays a crucial role in cell growth, metastasis, migration, and angiogenesis [30]. In our research, immunohistochemical analysis of FN1 expression in tissue microarrays showed significant upregulation in the stromal region of breast cancer compared to normal breast tissues. Knockdown of FN1 in CAFs notably inhibited breast cancer cell migration. Friedman et al. identified ECM-CAFs based on FN1 expression [31]. Our RNA-seq analysis of CAFs isolated from breast cancer patients revealed that genes highly expressed in the CAF2 subgroup were predominantly enriched in cases with axillary lymph node metastasis. Overall, these findings reinforce our viewpoint that the CAF2 subset (characterized by high FN1 expression) is closely associated with tumor metastasis. However, as our study is descriptive, further research is required to validate these conclusions.

The definition and classification of CAF subpopulations have become well documented; therefore, we did not perform extensive naming and functional analyses on all CAF subtypes. Instead, we concentrated on the function of the CAF2 subtype and its impact on breast cancer prognosis. The tumor microenvironment is a complex, heterogeneous system comprising various cellular elements, including tumor cells, immune cells, and fibroblasts. The intercellular communication among these cell types dictates tumor progression and heterogeneity [32]. For example, interactions between CAFs and gallbladder cancer cells have been shown to promote proliferation, EMT, and cancer stem-like features via the TSP4/integrin α2/HSF1 axis [33]. IL-6 signaling also mediates crosstalk between cancer cells and CAFs in the tumor microenvironment [34]. Our results revealed that CAF2 primarily communicates with tumor and myeloid cells through COL6A3-CD44 signaling. COL6A3, an integral component of Collagen VI in connective tissues, is uniquely expressed in stromal cells within the tumor microenvironment [35]. As COL6A3 is a secreted protein, targeting it would not specifically affect stromal cells. Notably, interactions between CAFs and immune cells are another crucial factor in promoting tumor progression [36]. Recent reviews have summarized possible immunoinhibitory mechanisms induced by CAF-immune cell crosstalk, involving various immune cells like TAMs, TANs, NK cells, and Tregs, among others [37]. Timperi et al. reported that a STAB1+TREM2^{high} lipid-associated macrophage subpopulation creates an immunosuppressive microenvironment in breast cancer, driven by the CAF-mediated CXCL12-CXCR4 axis [38]. In our study, the high CAF2 score group was characterized by increased macrophage infiltration, particularly a higher expression of M2 macrophages and lower expression of M1 macrophages. This suggests that CAF2 may be involved in macrophage differentiation and polarization, potentially contributing to breast cancer progression. However, further experiments are needed to explore the mechanism of CAF2-TAM communication in the tumor microenvironment.

Finally, our analysis suggests that MXRA5 might be a critical molecular target in the CAF2 subset associated with breast cancer metastasis. RT-qPCR analysis confirmed the elevated expression of MXRA5 in 15 paired breast cancer tissues. Similarly, tissue microarray analysis revealed upregulated MXRA5 in breast cancer tissues, significantly correlating with poor prognosis. MXRA5, a member of the MXRA protein family, is known as a secreted glycoprotein that inhibits fibrosis and inflammation [39]. Its role in cancer remains unclear, with only a handful of studies identifying MXRA5 as a significant factor in tumor progression, including in lung, gastric, and pancreatic cancers [40-42]. One study reported that MXRA5 promotes migration and EMT of pancreatic cancer cells by activating Akt-mTOR signaling [42]. Additionally, MXRA5 overexpression was found to enhance trophoblast cell invasion via the MAPK pathway [43]. These findings suggest that MXRA5 likely plays a crucial role in regulating cancer metastatic potential, supporting our hypothesis. However, its role in breast cancer had not been explored prior to our study. Our TCGA-BRCA analysis indicated that high MXRA5 expression significantly correlates with advanced breast cancer stages and lymph node metastasis. We report for the first time that MXRA5 expression is upregulated in breast cancer and CAFs associated with lymph node metastasis. Thus, we propose that MXRA5 could be a vital target for regulating the interaction between the CAF2 subtype and breast cancer cells, though further research is needed to substantiate this hypothesis.

This study has several limitations which should be acknowledged. Firstly, the identification of fibroblast clusters and the assessment of cancer prognosis in CAF subsets were based on retrospective data from public databases. This aspect would benefit from more prospective datasets (across different breast cancer subtypes) and advanced computational methods, such as artificial intelligence (AI). Secondly, although we strived for robust clustering analysis of CAFs in breast cancer single-cell datasets, our study remains primarily descriptive. Specific markers for these CAF subsets in human breast cancer should be identified using more precise experimental methods, like multiple-immunofluorescence (mIFC) staining and fluorescence-activated cell sorting (FACS). In addition, exploring the spatial distribution of CAF subtypes using spatial transcriptomics would also be a valuable research direction. Lastly, our study's computational and multiomic nature necessitates further experimental validation. While we have confirmed the differential expression of MXRA5 and the CAF2 subtype in breast cancer, the mechanisms by which the CAF2 subtype regulates MXRA5, affecting breast cancer metastasis, still require more extensive in vitro and in vivo exploration.

Conclusion

In summary, our study reveals the heterogeneity of cancer-associated fibroblasts in breast cancer and identifies a metastasis-associated CAF subtype - FN1+CAF2. Furthermore, we report, for the first time, the high expression of MXRA5 in breast cancer tissues, which may serve as a potential marker mediating the effects of CAF2 on breast cancer metastasis through TGFβ and EMT pathways. Our findings expand the understanding of CAF heterogeneity, offering new insights to assist in driving novel therapeutic developments for breast cancer metastasis.

Acknowledgements

Informed consent was obtained from all patients.

Disclosure of conflict of interest

None.

Address correspondence to: Bo Zhang, Department of Thyroid and Breast Surgery, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, Hubei, China. Tel: +86-15308620447; Fax: +86-27-85726139; E-mail: union_bzhang@hust.edu.cn; Jianguo Zhao, Department of Thyroid and Breast Surgery, Wuhan No. 1 Hospital, Wuhan 430022, Hubei, China. E-mail: 3040114010@qq.com

References

[1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A and Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021; 71: 209-249.

[2] Waks AG and Winer EP. Breast cancer treatment: a review. JAMA 2019; 321: 288-300.

[3] Loibl S, Poortmans P, Morrow M, Denkert C and Curigliano G. Breast cancer. Lancet 2021; 397: 1750-1769.

[4] Rivenbark AG, O'Connor SM and Coleman WB. Molecular and cellular heterogeneity in breast cancer: challenges for personalized medicine. Am J Pathol 2013; 183: 1113-1124.

[5] Quail DF and Joyce JA. Microenvironmental regulation of tumor progression and metastasis. Nat Med 2013; 19: 1423-1437.

[6] Bartoschek M, Oskolkov N, Bocci M, Lövrot J, Larsson C, Sommarin M, Madsen CD, Lindgren D, Pekar G, Karlsson G, Ringnér M, Bergh J, Björklund Å and Pietras K. Spatially and functionally distinct subclasses of breast cancer-associated fibroblasts revealed by single cell RNA sequencing. Nat Commun 2018; 9: 5150.

[7] Ishii G, Ochiai A and Neri S. Phenotypic and functional heterogeneity of cancer-associated fibroblast within the tumor microenvironment. Adv Drug Deliv Rev 2016; 99: 186-196.

[8] Luo H, Xia X, Huang LB, An H, Cao M, Kim GD, Chen HN, Zhang WH, Shu Y, Kong X, Ren Z, Li PH, Liu Y, Tang H, Sun R, Li C, Bai B, Jia W, Liu Y, Zhang W, Yang L, Peng Y, Dai L, Hu H, Jiang Y, Hu Y, Zhu J, Jiang H, Li Z, Caulin C, Park J and Xu H. Pan-cancer single-cell analysis reveals the heterogeneity and plasticity of cancer-associated fibroblasts in the tumor microenvironment. Nat Commun 2022; 13: 6619.

[9] Pelon F, Bourachot B, Kieffer Y, Magagna I, Mermet-Meillon F, Bonnet I, Costa A, Givel AM, Attieh Y, Barbazan J, Bonneau C, Fuhrmann L, Descroix S, Vignjevic D, Silberzan P, Parrini MC, Vincent-Salomon A and Mechta-Grigoriou F. Cancer-associated fibroblast heterogeneity in

axillary lymph nodes drives metastases in breast cancer through complementary mechanisms. Nat Commun 2020; 11: 404.

[10] Houthuijzen JM, de Bruijn R, van der Burg E, Drenth AP, Wientjens E, Filipovic T, Bullock E, Brambillasca CS, Pulver EM, Nieuwland M, de Rink I, van Diepen F, Klarenbeek S, Kerkhoven R, Brunton VG, Scheele CLGJ, Boelens MC and Jonkers J. CD26-negative and CD26-positive tissue-resident fibroblasts contribute to functionally distinct CAF subpopulations in breast cancer. Nat Commun 2023; 14: 183.

[11] Hu H, Piotrowska Z, Hare PJ, Chen H, Mulvey HE, Mayfield A, Noeen S, Kattermann K, Greenberg M, Williams A, Riley AK, Wilson JJ, Mao YQ, Huang RP, Banwait MK, Ho J, Crowther GS, Hariri LP, Heist RS, Kodack DP, Pinello L, Shaw AT, Mino-Kenudson M, Hata AN, Sequist LV, Benes CH, Niederst MJ and Engelman JA. Three subtypes of lung cancer fibroblasts define distinct therapeutic paradigms. Cancer Cell 2021; 39: 1531-1547, e1510.

[12] Elyada E, Bolisetty M, Laise P, Flynn WF, Courtois ET, Burkhart RA, Teinor JA, Belleau P, Biffi G, Lucito MS, Sivajothi S, Armstrong TD, Engle DD, Yu KH, Hao Y, Wolfgang CL, Park Y, Preall J, Jaffee EM, Califano A, Robson P and Tuveson DA. Cross-species single-cell analysis of pancreatic ductal adenocarcinoma reveals antigen-presenting cancer-associated fibroblasts. Cancer Discov 2019; 9: 1102-1123.

[13] ChallaSivaKanaka S, Vickman RE, Kakarla M, Hayward SW and Franco OE. Fibroblast heterogeneity in prostate carcinogenesis. Cancer Lett 2022; 525: 76-83.

[14] Puram SV, Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S, Rodman C, Luo CL, Mroz EA, Emerick KS, Deschler DG, Varvares MA, Mylvaganam R, Rozenblatt-Rosen O, Rocco JW, Faquin WC, Lin DT, Regev A and Bernstein BE. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. Cell 2017; 171: 1611-1624, e1624.

[15] Tirosh I, Izar B, Prakadan SM, Wadsworth MH 2nd, Treacy D, Trombetta JJ, Rotem A, Rodman C, Lian C, Murphy G, Fallahi-Sichani M, Dutton-Regester K, Lin JR, Cohen O, Shah P, Lu D, Genshaft AS, Hughes TK, Ziegler CG, Kazer SW, Gaillard A, Kolb KE, Villani AC, Johannessen CM, Andreev AY, Van Allen EM, Bertagnolli M, Sorger PK, Sullivan RJ, Flaherty KT, Frederick DT, Jané-Valbuena J, Yoon CH, Rozenblatt-Rosen O, Shalek AK, Regev A and Garraway LA. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science 2016; 352: 189-196.

[16] Su S, Chen J, Yao H, Liu J, Yu S, Lao L, Wang M, Luo M, Xing Y, Chen F, Huang D, Zhao J, Yang L,

Liao D, Su F, Li M, Liu Q and Song E. CD10(+) GPR77(+) cancer-associated fibroblasts promote cancer formation and chemoresistance by sustaining cancer stemness. Cell 2018; 172: 841-856, e816.

[17] Chen X and Song E. Turning foes to friends: targeting cancer-associated fibroblasts. Nat Rev Drug Discov 2019; 18: 99-115.

[18] Galbo PM Jr, Zang X and Zheng D. Molecular features of cancer-associated fibroblast subtypes and their implication on cancer pathogenesis, prognosis, and immunotherapy resistance. Clin Cancer Res 2021; 27: 2636-2647.

[19] Han HJ, Russo J, Kohwi Y and Kohwi-Shigematsu T. SATB1 reprogrammes gene expression to promote breast tumour growth and metastasis. Nature 2008; 452: 187-193.

[20] Rimal R, Desai P, Daware R, Hosseinnejad A, Prakash J, Lammers T and Singh S. Cancer-associated fibroblasts: origin, function, imaging, and therapeutic targeting. Adv Drug Deliv Rev 2022; 189: 114504.

[21] Sahai E, Astsaturov I, Cukierman E, DeNardo DG, Egeblad M, Evans RM, Fearon D, Greten FR, Hingorani SR, Hunter T, Hynes RO, Jain RK, Janowitz T, Jorgensen C, Kimmelman AC, Kolonin MG, Maki RG, Powers RS, Puré E, Ramirez DC, Scherz-Shouval R, Sherman MH, Stewart S, Tlsty TD, Tuveson DA, Watt FM, Weaver V, Weeraratna AT and Werb Z. A framework for advancing our understanding of cancer-associated fibroblasts. Nat Rev Cancer 2020; 20: 174-186.

[22] Costa A, Kieffer Y, Scholer-Dahirel A, Pelon F, Bourachot B, Cardon M, Sirven P, Magagna I, Fuhrmann L, Bernard C, Bonneau C, Kondratova M, Kuperstein I, Zinovyev A, Givel AM, Parrini MC, Soumelis V, Vincent-Salomon A and Mechta-Grigoriou F. Fibroblast heterogeneity and immunosuppressive environment in human breast cancer. Cancer Cell 2018; 33: 463-479, e410.

[23] Kieffer Y, Hocine HR, Gentric G, Pelon F, Bernard C, Bourachot B, Lameiras S, Albergante L, Bonneau C, Guyard A, Tarte K, Zinovyev A, Baulande S, Zalcman G, Vincent-Salomon A and Mechta-Grigoriou F. Single-cell analysis reveals fibroblast clusters linked to immunotherapy resistance in cancer. Cancer Discov 2020; 10: 1330-1351.

[24] Öhlund D, Handly-Santana A, Biffi G, Elyada E, Almeida AS, Ponz-Sarvise M, Corbo V, Oni TE, Hearn SA, Lee EJ, Chio II, Hwang CI, Tiriac H, Baker LA, Engle DD, Feig C, Kultti A, Egeblad M, Fearon DT, Crawford JM, Clevers H, Park Y and Tuveson DA. Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. J Exp Med 2017; 214: 579-596.

[25] Cords L, Tietscher S, Anzeneder T, Langwieder C, Rees M, de Souza N and Bodenmiller B. Cancer-associated fibroblast classification in single-cell and spatial proteomics data. Nat Commun 2023; 14: 4294.

[26] Lin TC, Yang CH, Cheng LH, Chang WT, Lin YR and Cheng HC. Fibronectin in cancer: friend or foe. Cells 2019; 9: 27.

[27] Peng Z, Hao M, Tong H, Yang H, Huang B, Zhang Z and Luo KQ. The interactions between integrin α(5)β(1) of liver cancer cells and fibronectin of fibroblasts promote tumor growth and angiogenesis. Int J Biol Sci 2022; 18: 5019-5037.

[28] Zhang L, Zhang C, Xing Z, Lou C, Fang J, Wang Z, Li M, He H and Bai H. Fibronectin 1 derived from tumor-associated macrophages and fibroblasts promotes metastasis through the JUN pathway in hepatocellular carcinoma. Int Immunopharmacol 2022; 113: 109420.

[29] Tsutsui J, Kadomatsu K, Matsubara S, Nakagawara A, Hamanoue M, Takao S, Shimazu H, Ohi Y and Muramatsu T. A new family of heparin-binding growth/differentiation factors: increased midkine expression in Wilms' tumor and other human carcinomas. Cancer Res 1993; 53: 1281-1285.

[30] Filippou PS, Karagiannis GS and Constantinidou A. Midkine (MDK) growth factor: a key player in cancer progression and a promising therapeutic target. Oncogene 2020; 39: 2040-2054.

[31] Friedman G, Levi-Galibov O, David E, Bornstein C, Giladi A, Dadiani M, Mayo A, Halperin C, Pevsner-Fischer M, Lavon H, Mayer S, Nevo R, Stein Y, Balint-Lahat N, Barshack I, Ali HR, Caldas C, Nili-Gal-Yam E, Alon U, Amit I and Scherz-Shouval R. Cancer-associated fibroblast compositions change with breast cancer progression linking the ratio of S100A4(+) and PDPN(+) CAFs to clinical outcome. Nat Cancer 2020; 1: 692-708.

[32] Hinshaw DC and Shevde LA. The tumor microenvironment innately modulates cancer progression. Cancer Res 2019; 79: 4557-4566.

[33] Shi Y, Sun L, Zhang R, Hu Y, Wu Y, Dong X, Dong D, Chen C, Geng Z, Li E and Fan Y. Thrombospondin 4/integrin α2/HSF1 axis promotes proliferation and cancer stem-like traits of gallbladder cancer by enhancing reciprocal crosstalk between cancer-associated fibroblasts and tumor cells. J Exp Clin Cancer Res 2021; 40: 14.

[34] Karakasheva TA, Lin EW, Tang Q, Qiao E, Waldron TJ, Soni M, Klein-Szanto AJ, Sahu V, Basu D, Ohashi S, Baba K, Giaccone ZT, Walker SR, Frank DA, Wileyto EP, Long Q, Dunagin MC, Raj A, Diehl JA, Wong KK, Bass AJ and Rustgi AK. IL-6 mediates cross-talk between tumor cells and activated fibroblasts in the tumor microenvironment. Cancer Res 2018; 78: 4957-4970.

[35] Kim GB, Fritsche J, Bunk S, Mahr A, Unverdorben F, Tosh K, Kong H, Maldini CR, Lau C, Srivatsa S, Jiang S, Glover J, Dopkin D, Zhang CX, Schuster H, Kowalewski DJ, Goldfinger V, Ott M, Fuhrmann D, Baues M, Boesmueller H, Schraeder C, Schimmack G, Song C, Hoffgaard F, Roemer M, Tsou CC, Hofmann M, Treiber T, Hutt M, Alten L, Jaworski M, Alpert A, Missel S, Reinhardt C, Singh H, Schoor O, Walter S, Wagner C, Maurer D, Weinschenk T and Riley JL. Quantitative immunopeptidomics reveals a tumor stroma-specific target for T cell therapy. Sci Transl Med 2022; 14: eabo6135.

[36] De Jaeghere EA, Denys HG and De Wever O. Fibroblasts fuel immune escape in the tumor microenvironment. Trends Cancer 2019; 5: 704-723.

[37] Mao X, Xu J, Wang W, Liang C, Hua J, Liu J, Zhang B, Meng Q, Yu X and Shi S. Crosstalk between cancer-associated fibroblasts and immune cells in the tumor microenvironment: new findings and future perspectives. Mol Cancer 2021; 20: 131.

[38] Timperi E, Gueguen P, Molgora M, Magagna I, Kieffer Y, Lopez-Lastra S, Sirven P, Baudrin LG, Baulande S, Nicolas A, Champenois G, Meseure D, Vincent-Salomon A, Tardivon A, Laas E, Soumelis V, Colonna M, Mechta-Grigoriou F, Amigorena S and Romano E. Lipid-associated macrophages are induced by cancer-associated fibroblasts and mediate immune suppression in breast cancer. Cancer Res 2022; 82: 3291-3306.

[39] Poveda J, Sanz AB, Fernandez-Fernandez B, Carrasco S, Ruiz-Ortega M, Cannata-Ortiz P, Ortiz A and Sanchez-Niño MD. MXRA5 is a TGF-β1-regulated human protein with anti-inflammatory and anti-fibrotic properties. J Cell Mol Med 2017; 21: 154-164.

[40] Xiong D, Li G, Li K, Xu Q, Pan Z, Ding F, Vedell P, Liu P, Cui P, Hua X, Jiang H, Yin Y, Zhu Z, Li X, Zhang B, Ma D, Wang Y and You M. Exome sequencing identifies MXRA5 as a novel cancer gene frequently mutated in non-small cell lung carcinoma from Chinese patients. Carcinogenesis 2012; 33: 1797-1805.

[41] Liu L, Pang H, He Q, Pan B, Sun X, Shan J, Wu L, Wu K, Yao X and Guo Y. A novel strategy to identify candidate diagnostic and prognostic biomarkers for gastric cancer. Cancer Cell Int 2021; 21: 335.

[42] Peng SQ, Zhu XR, Zhao MZ, Zhang YF, Wang AR, Chen MB and Ye ZY. Identification of matrix-remodeling associated 5 as a possible molecular oncotarget of pancreatic cancer. Cell Death Dis 2023; 14: 157.

[43] Ding L, Li S, Zhang Y, Gai J and Kou J. MXRA5 is decreased in preeclampsia and affects trophoblast cell invasion through the MAPK pathway. Mol Cell Endocrinol 2018; 461: 248-255.

Single-cell analysis of cancer-associated fibroblasts subtypes in breast cancer

Table S1. Characteristics of patients whose tumors were used for isolation of CAFs and RNA-seq

Patients ID	Age	RNA-seq	Tumor size (cm)	ER	PR	HER2	Lymph node metastasis	Histology grade
BC01	48	Yes	4.3*1.8*1.3	+	+	-	-	III
BC02	41	Yes	2.1*2.0*1	-	-	-	-	II
BC03	67	Yes	2.2*1.6*1.2	+	+	-	-	III
BC04	74	No	2.3*2*1.4	+	-	-	-	III
BC05	57	Yes	2.1*1.8*1.2	+	+	-	+	II
BC06	44	Yes	2.4*1.5*1	+	+	+	+	III
BC07	48	Yes	2.5*2*2	-	-	+	+	III
BC08	56	No	2*1.5*1	+	+	-	+	II

Table S2. Characteristics of patients whose tumors were used for construction of the tissue microarray

Patients number	Gender	Age	Neoadjuvant chemotherapy	ER (%)	PR (%)	HER2	Ki-67 (Li)	Lymph node metastasis	TNM Stage
P01	female	56	-	+	-	+	30%	+	T2N1M0
P02	female	58	-	-	+	+	20%	-	T1N0M0
P03	female	51	-	+	-	-	15%	-	T2N0M0
P04	female	55	-	+	+	+	35%	-	T1N0M0
P05	female	58	-	-	-	-	70%	+	T1N1M0
P06	female	40	+	+	-	-	90%	+	T2N1M0
P07	female	45	-	+	+	-	30%	-	T1N0M0
P08	female	57	-	+	-	+	20%	-	T1N0M0
P09	female	72	+	+	+	-	5%	+	T2N2M0
P10	female	59	-	+	+	-	10%	-	T2N0M0
P11	female	68	-	+	+	-	15%	+	T1N2M0
P12	female	48	-	+	+	-	30%	-	T2N0M0
P13	female	30	-	-	-	+	15%	-	T1N0M0
P14	female	65	-	+	+	+	30%	-	T2N0M0
P15	female	46	-	+	+	-	<5%	-	T1N0M0
P16	female	57	-	+	+	+	80%	+	T2N1M0
P17	female	68	-	+	+	-	40%	+	T1N1M0
P18	female	65	-	NA	NA	NA	-	-	Tis
P19	female	38	+	+	+	-	20%	-	T2N0M0
P20	female	50	-	-	-	-	60%	-	T2N0M0
P21	female	66	-	-	-	+	20%	+	T2N1M0
P22	female	49	-	+	+	-	5%	-	T1N0M0
P23	female	47	-	+	-	+	40%	+	T2N2M0
P24	female	56	-	+	+	+	25%	-	T2N0M0
P25	female	69	-	+	+	-	30%	-	T1N0M0
P26	female	43	-	+	+	-	40%	+	T1N1M0
P27	female	49	-	NA	NA	NA	-	-	Tis
P28	female	28	+	+	+	-	20%	-	T1N0M0
P29	female	57	+	+	+	-	30%	-	T1N0M0
P30	female	49	-	-	-	+	20%	+	T2N1M0
P31	female	65	-	+	+	-	5%	-	T2N0M0
P32	female	62	-	+	+	-	20%	-	T1N0M0
P33	female	32	+	-	-	+	40%	+	T3N1M0
P34	female	51	-	+	+	-	60%	-	T1N0M0
P35	female	50	-	+	+	-	15%	+	T1N2M0

1

P36	female	51	-	+	+	-	10%	-	T2N0M0
P37	female	64	-	+	+	-	20%	+	T1N1M0
P38	female	78	-	+	+	-	10%	+	T2N2M0
P39	female	45	-	+	+	-	10%	+	T1N2M0
P40	female	43	-	+	+	+	>10%	+	T2N2M0
P41	female	62	-	+	+	-	10%	-	T2N0M0
P42	female	55	-	+	+	-	25%	-	T1N0M0
P43	female	47	-	+	+	-	>15%	+	T2N1M0
P44	female	40	-	+	+	-	<10%	+	T2N2M0
P45	female	83	-	+	-	-	30%	-	T2N0M0
P46	female	45	-	+	-	+	50%	+	T1N1M0
P47	female	48	-	+	-	-	20%	+	T2N2M0

Blue areas are patients with neoadjuvant chemotherapy.

Table S3. Correlation between MXRA5 expression and clinicopathologic characteristics of breast cancer patients in TCGA

Characteristics	Low expression of MXRA5	High expression of MXRA5	P value
n	543	544	
Pathologic T stage, n (%)			0.008
T1	118 (10.9%)	160 (14.8%)	
T2	323 (29.8%)	308 (28.4%)	
T3&T4	99 (9.1%)	76 (7%)	
Pathologic N stage, n (%)			0.005
N0	272 (25.5%)	244 (22.8%)	
N1	183 (17.1%)	176 (16.5%)	
N2	40 (3.7%)	76 (7.1%)	
N3	38 (3.6%)	39 (3.7%)	
Pathologic M stage, n (%)			0.887
M0	438 (47.4%)	467 (50.5%)	
M1	10 (1.1%)	10 (1.1%)	
Pathologic stage, n (%)			0.287
Stage I	85 (8%)	97 (9.1%)	
Stage II	327 (30.8%)	292 (27.5%)	
Stage III	114 (10.7%)	130 (12.2%)	
Stage IV	9 (0.8%)	9 (0.8%)	
Age, n (%)			0.063
≤ 60	286 (26.3%)	317 (29.2%)	
> 60	257 (23.6%)	227 (20.9%)	
HER2 status, n (%)			0.050
Negative	263 (36.1%)	297 (40.7%)	
Indeterminate	8 (1.1%)	4 (0.5%)	
Positive	60 (8.2%)	97 (13.3%)	

Table S4. Univariate analysis and multivariate analysis of the correlation of MXRA5 expression with OS among breast cancer

Characteristics	Total (N)	Univariate analysis		Multivariate analysis	
		Hazard ratio (95% CI)	*P* value	Hazard ratio (95% CI)	*P* value
Pathologic T stage	1,083				
T1	277	Reference		Reference	
T2	631	1.334 (0.889-2.003)	0.164	0.942 (0.519-1.710)	0.844
T3&T4	175	1.931 (1.208-3.088)	0.006	2.111 (1.023-4.355)	0.043
Pathologic N stage	1,067				
N0	516	Reference		Reference	
N1	358	1.947 (1.322-2.865)	< 0.001	1.613 (0.927-2.809)	0.091
N2	116	2.522 (1.484-4.287)	< 0.001	1.837 (0.849-3.978)	0.123
N3	77	4.191 (2.318-7.580)	< 0.001	4.115 (1.694-9.995)	0.002
Pathologic M stage	925				
M0	905	Reference		Reference	
M1	20	4.266 (2.474-7.354)	< 0.001	3.389 (1.338-8.588)	0.010
Age	1,086				
≤ 60	603	Reference		Reference	
> 60	483	2.024 (1.468-2.790)	< 0.001	2.838 (1.735-4.643)	< 0.001
HER2 status	717				
Negative	560	Reference		Reference	
Positive	157	1.593 (0.973-2.609)	0.064	1.254 (0.716-2.197)	0.428
MXRA5	1,086				
Low	542	Reference		Reference	
High	544	1.340 (0.971-1.848)	0.075	1.083 (0.659-1.780)	0.753

Table S5. Clinical information for breast cancer patients were used for RT-qPCR

| Patients ID	Age	Neoadjuvant chemotherapy	Tumor size (cm)	ER (%)	PR (%)	HER2	Ki-67 (Li)	Lymph node metastasis	Histology grade
01	38	No	3.5*2.5*2.5	-	-	3+	30%	+	III
02	67	No	3.5*2.2*1.5	-	-	-	30%	-	III
03	35	No	3.5*3*2	+	+	2+, FISH-	10%	-	II
04	61	No	3*2*1.5	-	-	2+, FISH-	30%	-	III
05	71	No	2*2*1.5	+	+	1+	30%	-	II
06	67	No	3*2.8*2.5	+	+	2+, FISH-	30%	+	II
07	51	No	3*1.5*1.5	+	+	3+	70%	+	III
08	63	No	3.5*2.7*2	+	-	-	30%	+	III
09	31	No	1.5*1.5*1	+	+	2+, FISH+	30%	+	II
10	71	No	2.5*1.5*1	+	-	2+, FISH-	15%	+	II
11	38	No	2.2*2*1.5	+	+	-	20%	-	II
12	52	No	3.3*3*2.5	+	+	2+, FISH+	20%	-	III
13	71	No	2.5*1.7*1	+	+	1+	10%	+	II
14	40	No	2.5*1.5	+	+	1+	<10%	+	II
15	51	No	2.5*2*1.5	+	+	-	10%	-	II
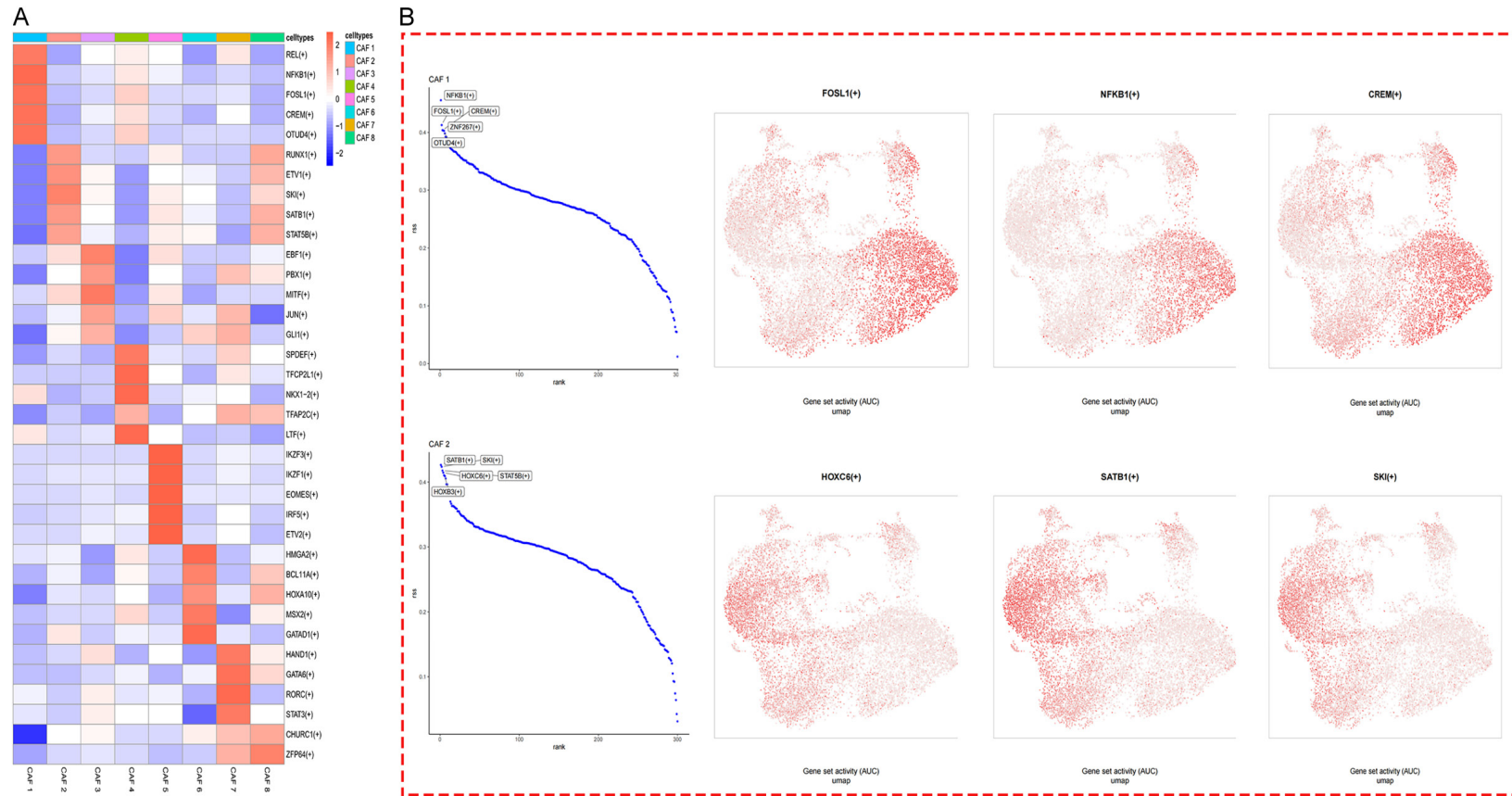
Figure S1. Transcription factors associated with caf isoforms in breast cancer. A. Heatmap of gene expression of transcription factors associated with CAF1-8 cell subtypes. B. Transcription factors such as NFKB1 and SATB1 were significantly expressed in CAF1-2.

Single-cell analysis of cancer-associated fibroblasts subtypes in breast cancer
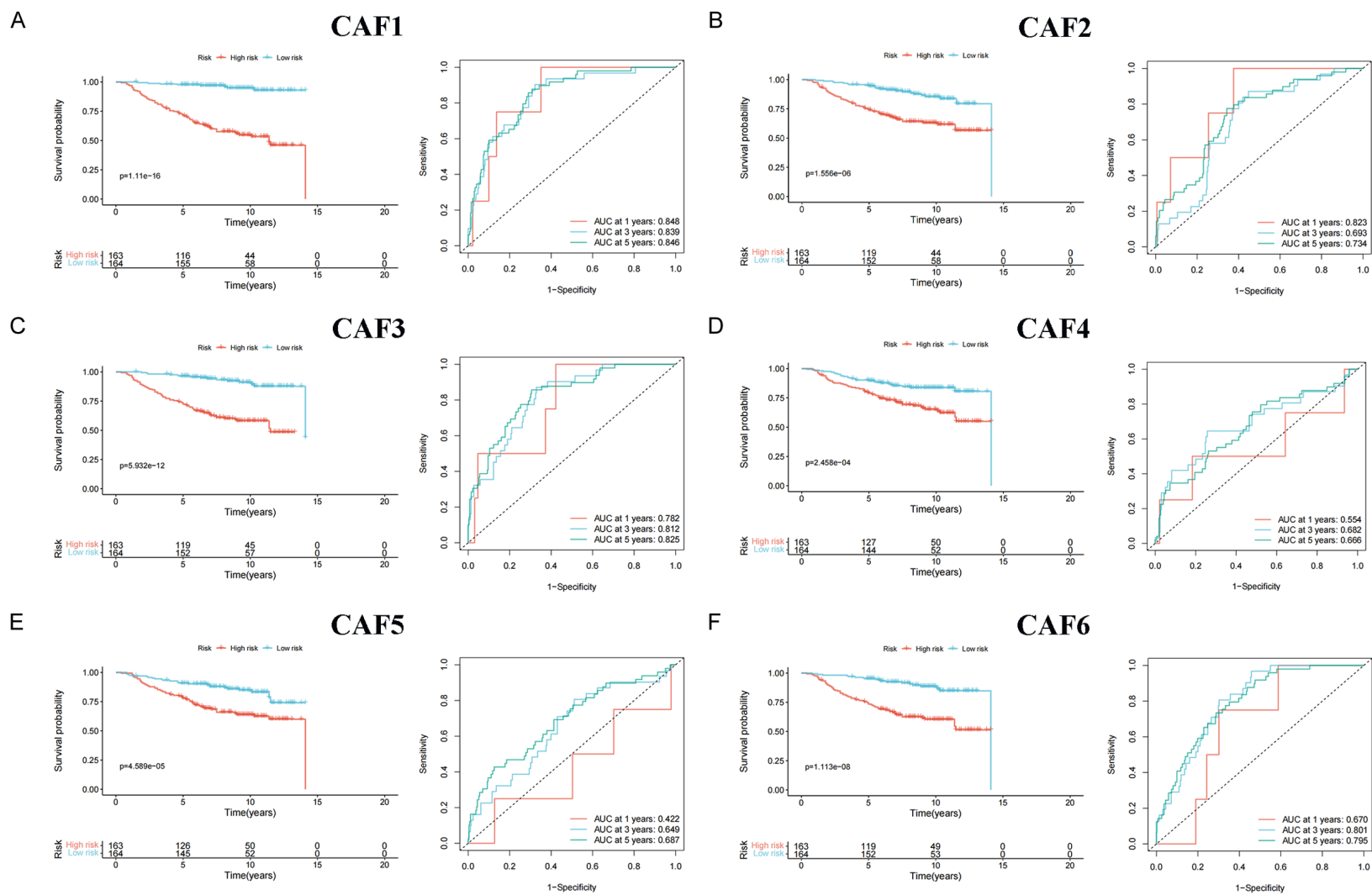
A

CAF1

B

CAF2

C

CAF3

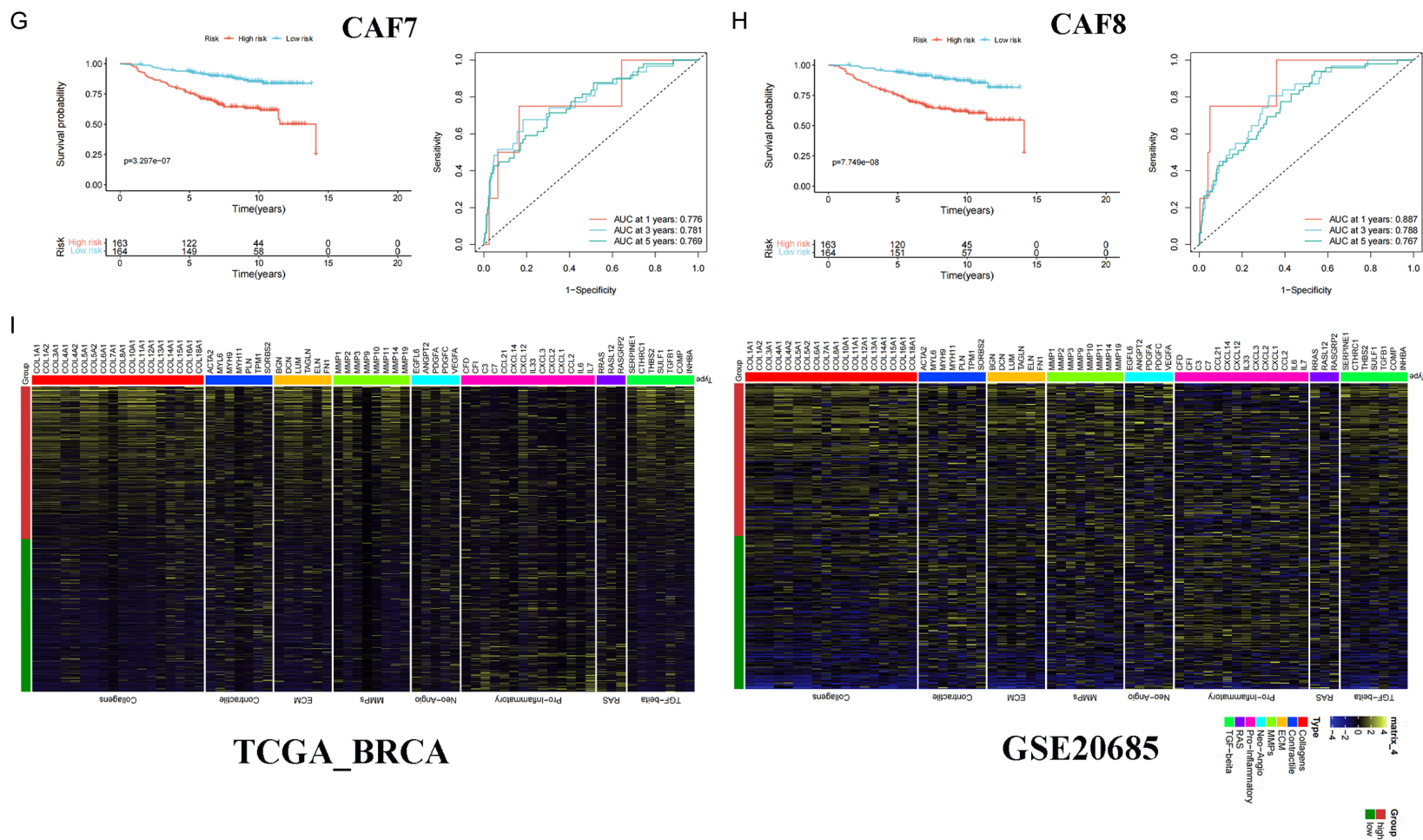D

CAF4

E

CAF5

F

CAF6

Figure S2. Expression of cancer-promoting genes in breast cancer with high CAF2 scores. A-H. Survival curves for the GES20685 dataset showed that among breast cancers with different CAF subtypes, the group subtype with a higher score had a worse prognosis. I. Heatmap of gene expression associated with collagen, ECM, MMPs, TGF-β, Neo-Angio, Contractile, RAS, and Pro-Inflammatory certain functions in the TCGA-BRCA and GSE20685 datasets in the high-CAF2 scoring subgroups.

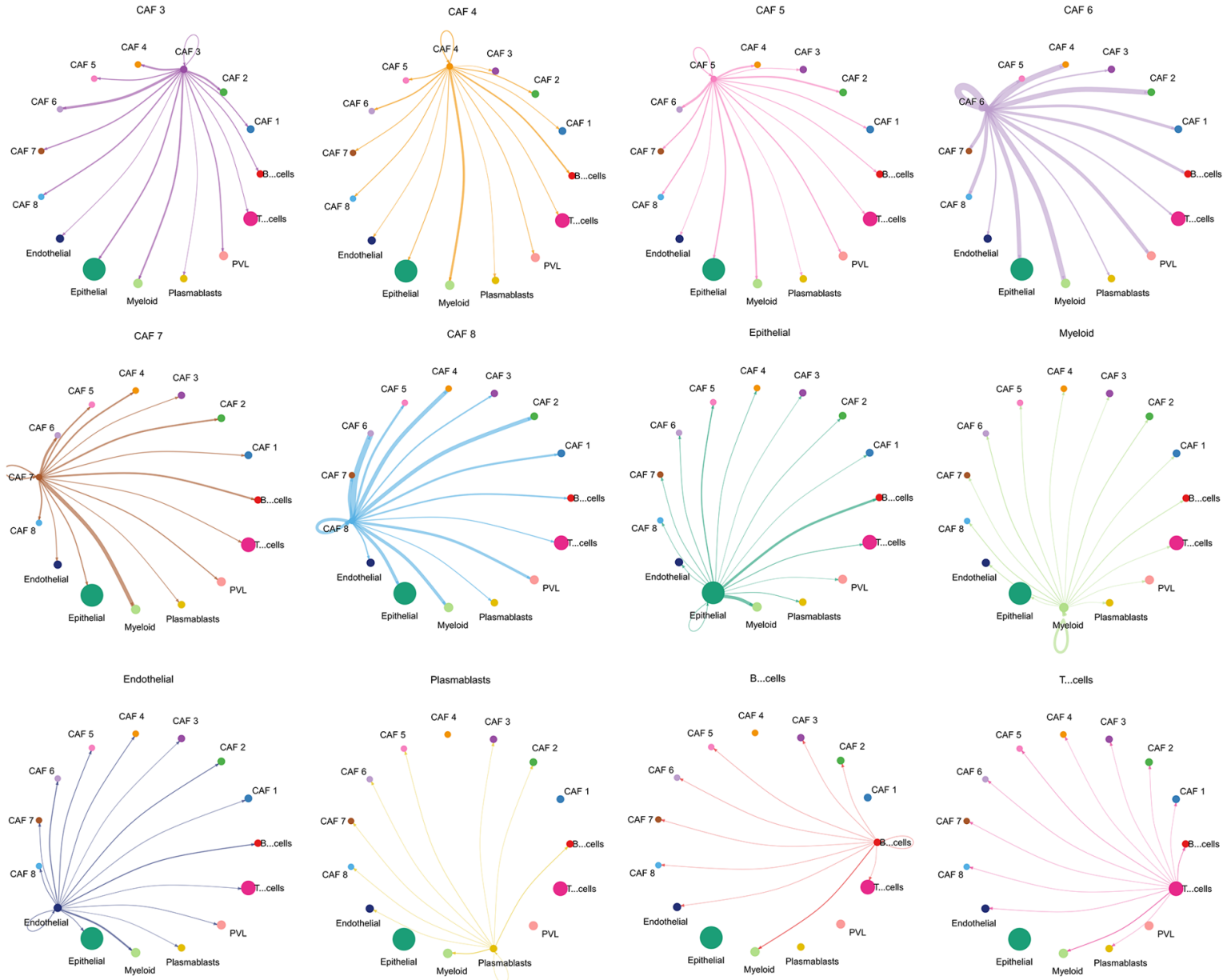# Single-cell analysis of cancer-associated fibroblasts subtypes in breast cancer

Figure S3. The strength of interactions between individual subclass cell populations and other subclass cell populations.