# scientific reports

OPEN

# Single-cell profiling transcriptomic reveals cellular heterogeneity and cellular crosstalk in breast cancer lymphatic node, bone, and brain metastases

Longyu Zhu[1], Miaomiao Liu[2], Yuguang Shang[1], Jingge Cheng[3], Hongye Zhao[4], Jun Zhang[1]✉ & Dongxing Shen[1]✉

Breast cancer is the most common malignant tumor in the world, and its metastasis is the main cause of death in breast cancer patients. However, the differences between primary breast cancer tissue and lymphatic node, bone, and brain metastases at the single-cell level are not fully understood. We analyzed the microenvironment heterogeneity in samples of primary breast cancer ($n = 4$), breast cancer lymphatic node metastasis ($n = 4$), breast cancer brain metastasis ($n = 3$), and breast cancer bone metastasis ($n = 2$) using single-cell sequencing data from the GEO database. The malignant epithelial cells were characterized by InferCNV algorithm. The cell-cell communication was analyzed using CellChat package. The biological function of cell subpopulations was analyzed using gene set variation analysis. The expression of STMN1 was analyzed using immunohistochemical staining. The proportion of pCAFs in breast cancer was explored using multispectral immunohistochemical staining. We identified seven cell clusters in primary and metastatic breast cancer (Lymphatic node, brain, and bone metastases) by analyzing single-cell transcriptomic profiles. T-NK and B cells dominated breast cancer with lymphatic node metastasis, whereas fibroblasts were prevalent in brain metastases and primary breast cancer. We identified five T cells (T memory, CD8 +T cells, regulatory T cells, natural killer cells, CD4 +T cells), three B cells (naïve B cells, memory B cells, plasma B cells), and five cancer-associated fibroblasts (CAFs) subpopulations (Smooth muscle cells (SMC), pericyte, antigen-presenting CAFs (apCAFs), proliferative CAFs (pCAFs), and matrix CAFs (mCAFs)). Notably. pCAFs dominated breast cancer with lymphatic node, bone, and brain metastasis. Furthermore, we identified four malignant epithelial cell subpopulations: G0, G1, G2, and G3. The G2 cell population exhibited strong invasion ability, it can differentiate into G3 with strong proliferative ability and proliferation-related G1 cell population after metastasis. Cell-cell communication demonstrated an interaction between pCAFs and metastasis-associated malignant epithelial cells. Finally, we discovered that in advanced breast cancer, the proportion of pCAF increased and was associated with a poor prognosis of breast cancer. This study elucidated the potential cellular origins and drivers of breast cancer metastases to lymphatic nodes, brain, and bone, utilizing single-cell transcriptomic profiles. Furthermore, it demonstrated that increased pCAFs were associated with advanced breast cancer and a poor prognosis.

**Keywords** Breast cancer, Distinct metastasis, Malignant epithelial cells, pCAFs, Prognosis

Breast cancer is a heterogeneous disease involving genetic and environmental factors[1]. The global breast cancer statistics report shows that in 2022, there were an estimated 2.3 million new cases and 665,684 deaths worldwide in 2022, making it the second most prevalent malignant tumor worldwide[2]. New therapies for breast cancer

[1]Department of Radiotherapy Oncology, The Fourth Hospital of Hebei Medical University, No. 169, Tianshan Street, Hebei, Shijiazhuang 050035, Hebei Province, China. [2]The Fifth Department of Oncology, Hebei General Hospital, Shijiazhuang 050035, Hebei, China. [3]Department of Thoracic Surgery, The Fourth Hospital of Hebei Medical University, Hebei, Shijiazhuang 050035, China. [4]Department of Dermatology, The Fourth Hospital of Hebei Medical University, Hebei, Shijiazhuang 050035, China. ✉email: zhangjun1965@163.com; eastarshen@163.com

have been released in recent years, including targeted therapy and immunotherapy, opening up new horizons for the treatment of breast cancer patients[3]. However, breast cancer is frequently diagnosed only after metastasis has occurred. Generally, the prognosis of breast cancer is poor because it starts as a local disease and can spread to lymphatic nodes[4] or distant organs such as bones[5] and brain[6], which presents significant challenges for the treatment of breast cancer. Thus, understanding the mechanisms of breast cancer metastasis may help develop targeted therapies against metastasis.

The TME is composed of immune cells and non-immune cells such as CAFs, adipocytes, and endothelial cells. These different TME components play a role in cancer initiation, progression, metastasis, and treatment resistance in breast cancer[7,8]. In breast cancer, different metastatic sites have different levels of immune cell enrichment. Compared with primary breast tumors (PT), bone metastases (OM) have greater neutrophil infiltration, whereas lymphatic node metastases (LM) have fewer macrophages[9]. In breast cancer, the TME shows infiltration of PD-L1-amplified CD8 + T cells[10]. CAFs within the TME can remodel the extracellular matrix (ECM), and play an important role in the interaction between tumor cells and surrounding cells[11]. CAF is associated with the clinicopathological characteristics of tumors and plays an important role in the pathogenesis of tumors[12]. For example, in breast cancer, tumor cells co-cultured with CAFs exhibit increased expression of IL-6 and IL-8, resulting in increased invasiveness and angiogenic capacity[13]. CAFs exhibit significant heterogeneity in breast cancer and display different expression levels of various molecules and biological behaviors in different metastatic parts of breast cancer[14]. In metastatic breast cancer, bone metastases are characterized by significant upregulation of stromal PDPN, FSP1, and PDGFRα, lung metastases show a notable increase in interstitial PDGFRβ expression, and liver metastases exhibit reduced interstitial levels of FSP1 and PDGFRα[15]. Furthermore, different subtypes of CAFs in axillary lymphatic nodes drive breast cancer metastasis via complementary mechanisms[16]. Therefore, exploring the molecular pathways and intercellular crosstalk in breast cancer with distinct metastasis may help to gain insights into the mechanisms of breast cancer development and metastasis, and provide new targets and strategies for breast cancer treatment.

In recent years, the heterogeneity of in situ breast cancer and the mechanisms underlying the development of metastasis have been increasingly reported. For instance, Hou et al. have documented the cellular heterogeneity present in ductal subtype breast cancer, highlighting the differential gene expressions and biological functions between Type 1 and Type 2 ductal epithelial cells[17]. Additionally, Sanjaya et al. have identified several mechanisms that contribute to lymph node and liver metastasis in breast cancer, including loss of differentiation, epithelial-to-mesenchymal transition, and autophagy[18]. However, these studies did not specifically investigate the cell-cell interactions within the TME of breast cancer. To fill this gap, we analyzed the cellular composition of the TME of patients with breast cancer LM, brain metastasis (BM), and OM using single-cell RNA sequencing (scRNA-seq) data. We focused on the main cell types, including T cells, B cells, and CAFs, and their related biological functions. In addition, we analyzed the interaction between proliferative CAFs (pCAFs) and other cells and explored the role of pCAFs in the prognosis of breast cancer patients. These findings will provide important evidence to support the diagnosis, treatment, and molecular mechanisms of metastasis in breast cancer.

## Materials and methods

### Data sources
The scRNA-seq datasets GSE225600, GSE186344, and GSE190772 of breast cancer were downloaded from the GEO (https://www.ncbi.nlm.nih.gov/) database. GSE225600 included 4 PT and 4 breast cancer LM samples. GSE186344 and GSE190772 datasets contained 3 breast cancer BM samples and 2 breast cancer OM samples, respectively. Bulk RNA-seq data and clinical information of breast cancer were downloaded from The Cancer Genome Atlas (TCGA, https://tcga-data.nci.nih.gov/tcga/) database.

### ScRNA-seq data quality control and cell annotation
Transcripts were mapped to the human reference genome (GRCh38) using Cell Ranger V6.1.2. ScRNA-seq data was processed using Seurat v4.1.1[19], and cells with higher than 10% mitochondrial content, higher than 5% hemoglobin content, and less than 200 and more than 20,000 expressed genes were removed. Data normalization, cell clustering, and dimensional reduction were performed using the Seurat package[20]. The "FindVariableFeatures" function was utilized to select 2,000 highly variable genes from the corrected expression matrix. Subsequently, the principal component analysis was conducted using the "RunPCA" function, retaining the top 20 principal components for further analysis. After correcting batch effects by "RunHarmony" function in the harmony package, cells were clustered using the "FindClusters" function (resolution 0.7). The nonlinear dimensionality reduction was performed using the "RunUMAP" function[21]. Cell clusters were annotated using common cell mark genes based on the cellmark2.0 database[22].

### Single-cell differential gene analysis and functional enrichment analysis
The differentially expressed genes (DEGs) between different groups were identified using "FindMarkers" in the Seurat package, based on the |avg_log2FC|>0.3 and adjusted p value<0.05. The DEGs were then subjected to Gene ontology (GO, including Biological Process (BP), Molecular Function (MF), and Cellular Component (CC) analysis) and Kyoto Encyclopedia of Genes and Genome (KEGG) enrichment analysis. The significantly enriched pathways were screened using $p < 0.05$, and the top 30 GOs and 20 KEGG pathways were presented.

### Gene set variation analysis (GSVA)
The gene set variation analysis (GSVA) package in R was used to analyze the most enriched hallmark pathways for each cell in primary and metastatic tumor samples drawing data from the Molecular Signature Database (https://www.gsea-msigdb.org/gsea/msigdb/ index.jsp). The differences in pathway enrichment scores between

different cell clusters were calculated using the "limma" package. $P < 0.05$ indicated that the pathway was differentially enriched by the two groups.

### Transcription factor analysis
To calculate regulon specificity scores (RSS) for fibroblast subpopulations, SCENIC analysis was performed using pySCENIC Python software package. The co-expression modules of transcription factors (TFs) and their target genes were identified using GRNBoost2. RcisTarget was used to analyze the genes in each co-expression module to identify transcription factor binding motifs. The AUCell package was used to assess the activity of regulons in each cell. The R package "ggplot" was used to draw a scatter plot showing the RSS of each fibroblast subpopulation, marking the top 5 regulators.

### Identification of malignant cells
Based on single-cell gene expression and chromosome sequencing data, the inferCNV v1.6.0 package [https://github.com/broadinstitute/inferCNV] in R language was used to distinguish malignant epithelial cells (ECs) from non-malignant ECs. The settings for the inferCNV analysis are as follows: cutoff = 0.1, cluster by groups = TRUE, and hidden markov model (HMM) = TRUE. To minimize false positives in copy number variation (CNV) inference, we employed the default Bayesian latent mixture model to assess the posterior probability of variants in each cell, utilizing a threshold of 0.5. Then, the CNV scores of all genes were hierarchically clustered on ECs and reference cells (T-NK cells) by k-means algorithm. Finally, the subclusters with relatively higher CNV scores were considered malignant cells. To illustrate tumor clonality and evolution, the "sub-cluster" model was further applied to classify malignant cells into 8 clusters based on different CNV patterns generated by HMM. Each CNV was annotated as a gain or loss of p-arm or q-arm based on chromosome cell banding information. Subclones containing identical arm-level CNVs were merged to construct an evolutionary tree. Uphyloplot2 was used to visualize evolutionary phylogenetic dendrograms to represent subclonal CNV structures. There are at most 8 branches at the end of the tree, and less than 5% of the cells will not be output to the graph.

### Trajectory analysis
Trajectory analysis was performed using the Monocle 2 package to reveal epithelial cell differentiation trajectories. An integrated expression matrix with batch effects removed was used as input data, and unit trajectories and evolutionary orders were inferred using default parameters. Highly variable genes associated with cell trajectories were identified using the graph_test function.

### Cell-cell interactions
The CellChat package was used to predict and visualize biologically relevant cell-to-cell communications. Specifically, the createCellChat function was employed to generate a CellChat object. Subsequently, annotate the objects with labels and identify the overexpressed genes. The communication probabilities were inferred using the computeCommunProb function, and the intercellular communication for each specific cell signaling pathway was predicted using the computeCommunProbPathway function.

### Survival analysis
The patients were categorized into high and low-expression groups using the R package "survminer", with the optimal index serving as the cutoff value. Kaplan-Meier (KM) survival analysis was employed to assess the overall survival (OS) of patients in both the high and low-expression groups, and comparisons were made using a two-sided log-rank test.

### Multispectral immunohistochemical (mIHC) staining
Breast cancer tissues were obtained from The Fourth Hospital of Hebei Medical University. Detailed patient information is shown in Table S1. ITGAV (Proteintech, 27096-1-AP) and POSTN (Proteintech, 66491-1-Ig) were used as primary antibodies.

Immunofluorescence staining was performed using AlphaTSA Multiplex IHC Kit (AXT37100031, Alphaxbio) according to the manufacturer's instructions. Briefly, the tissue chip was dewaxed and hydrated using xylene and alcohol washing, and then antigen retrieval and sealing were performed. Sections were then blocked and incubated with primary and secondary antibodies, followed by fluorescent staining. Finally, cell nuclei were counterstained with DAPI and enclosed in Mounting Medium. ZEN (v3.1) software was used for film reading.

### Immunohistochemical (IHC) staining
Formalin-fixed, paraffin-embedded tissue blocks were serially cut into 4-µm-thick sections, which were dewaxed in xylene and rehydrated through ethanol solutions. For antigen retrieval, slides were microwaved with citrate buffer (pH 6.0) for 10 min. The slides were then incubated with a primary antibody (Proteintech, 11157-1-AP) at 4 °C overnight. The slides were then incubated with secondary antibodies (Abcam, ab288151) for 1 h at room temperature, and then developed using the Ultravision DAB Plus Substrate Detection System (Thermo Fischer Scientific, TA-125-QHDX) at room temperature, then hematoxylin stained, dehydrated, and coverslipped.

### Statistical analysis
The chi-square test was applied to categorical variables for comparison between two groups. The Mann-Whitney U test was used to compare categorical variables and non-normally distributed variables between two groups, while the Kruskal-Wallis test was used to compare multiple groups. All statistical analyses were performed using R software v 4.0.5, and a two-sided P value < 0.05 was considered statistically significant.

## Results

### Single-cell transcriptional landscape of primary and metastatic breast cancer

We obtained a total of 58,504 cells from breast cancer samples, among which 18,159, 16,800, 21,723, and 1822 cells were collected from PT, LM, BM, and OM, respectively. After dimensionality reduction and unsupervised cell clustering, these cells were clustered into 29 clusters (Fig. 1A). The clusters were annotated with marker genes and identified 7 cell clusters, including T-natural killer (NK) cells, B plasma cells, myeloid cells, fibroblast, epithelial, endothelial, and astrocyte (Fig. 1B and D, Table S2). LM samples had a relatively high proportion of T-NK and B cells. Fibroblasts accounted for more in BM and PT samples, but less in LM and OM samples.
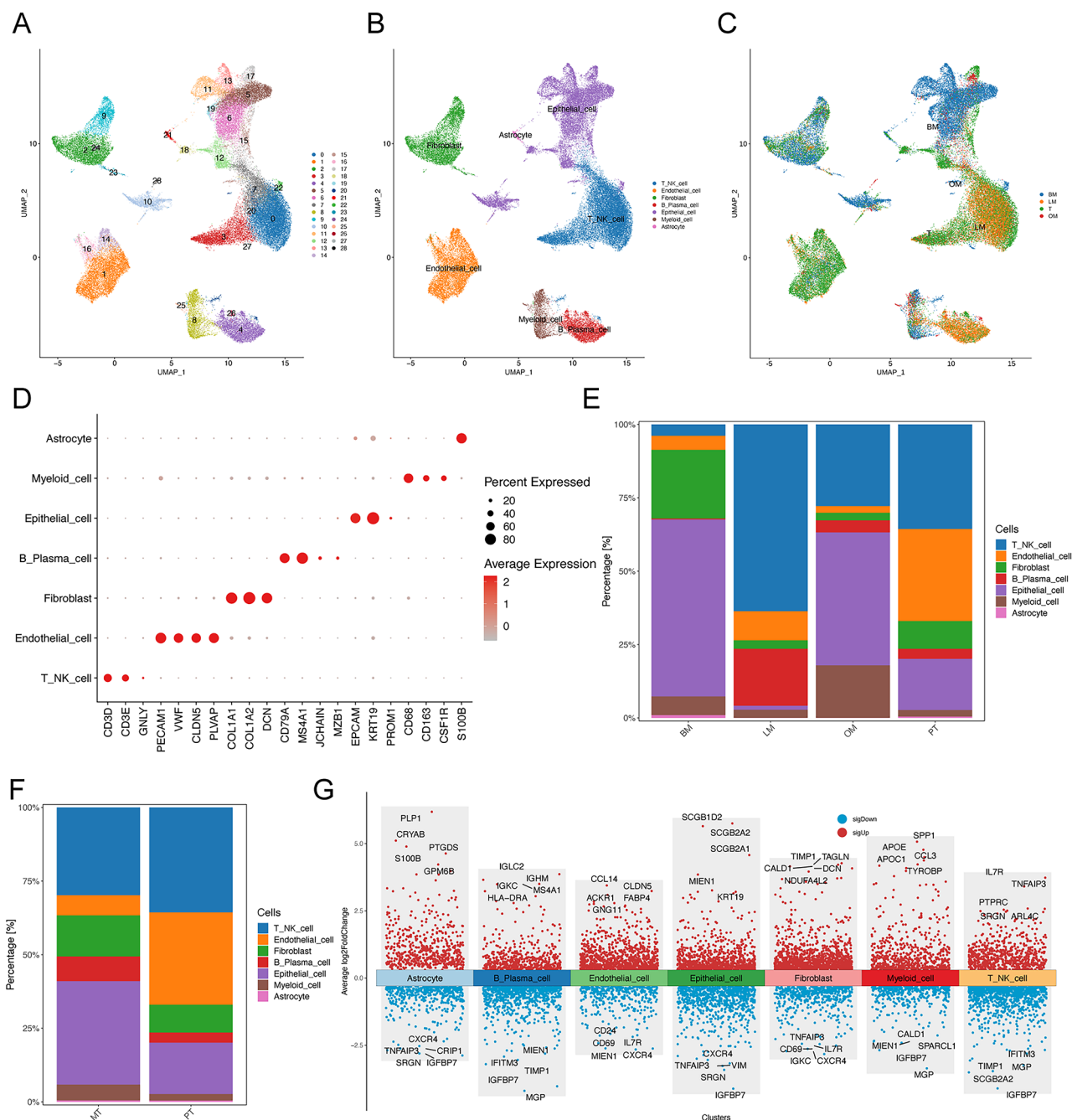


**Fig. 1.** Single-cell transcriptional landscape of primary and metastatic breast cancer. (**A**), A UMAP diagram showing the cell clusters. (**B**), Cell annotation results. (**C**), UMAP showing the origin of cells. (**D**), The expression of marker genes in different cell clusters. (**E**), The proportion of seven cell clusters in primary, lymphatic node, brain, and bone metastatic samples. (**F**), The fraction of seven cell types in primary and metastatic samples. (**G**), Differentially expressed genes between primary and metastatic samples in seven cell clusters. PT: Primary tumor; BM: brain metastases; LM: lymphatic node metastases; OM: bone metastases; MT: metastases tumor.

Epithelial cells accounted for more in BM, OM, and PT samples, but less in LM samples (Fig. 1E). Next, we compared the difference in cell proportions between PT and all MT samples and found that epithelial cells were the major cell differences in TME between PT and MT samples (Fig. 1F). In addition, we also analyzed the differentially expressed genes between PT and MT samples in different cell types and found that the upregulated genes (such as SCGB1D2, TNFAIP3, TIMP1, DCN, CALD1) in MT samples were closely related to inflammation and matrix remodeling (Fig. 1G). The upregulated genes in PT samples were closely related to metastasis (IGFBP7, CXCR4, VIM), regulation of the immune response and TME (IGFBP7, TIMP1, MGP, CD69), and inflammation (CXCR4, TNFAIP3) of breast cancer (Fig. 1G).

### Identification of major T and B cell types

We conducted a comprehensive analysis of T and B cells utilizing single-cell transcriptomes. A total of 18,504 T-NK cells were regrouped and categorized into T-memory, CD8 + T cells, Treg, NK cells, and CD4 + T cells based on marker gene expression (Fig. 2A and B, Table S2). Among these cell types, CD8 + T cells represented the largest proportion of the BM samples, while Treg cells accounted for the largest proportion of the PT samples (Fig. 2C). GSVA results showed that T-NK cells in MT samples significantly activated DNA repair, oxidative phosphorylation, and WNT-beta-catenin signaling pathways; while T-NK cells in PT samples significantly activated interferon-gamma response, complement, and IL2-STAT5 signaling pathways (Fig. 2D). The CD8 + T cell subset in MT samples significantly activated oxidative phosphorylation, DNA repair, and MYC-targets-v1 signaling pathways, while the CD8 + T cell subset in PT samples significantly activated TNF-alpha signaling via NF-κB, hypoxia, and interferon-gamma response pathways (Figure S1B).

A total of 4029 B cells were re-grouped and divided into naïve B, memory B, and plasma B cells according to marker gene expression (Fig. 2E and F, Table S2). As shown in Fig. 2G, the predominant cell type in LM, OM, and PT samples was naïve B cells. The GSVA results showed that B cells in the MT sample significantly activated allograft rejection, MYC-targets-v1, KRAS signaling-up signaling, and B cells in the PT sample significantly activated metabolism-related pathways, such as xenobiotic metabolism, glycolysis, apoptosis (Fig. 2H). Naïve B cells in MT samples significantly activated allograft rejection, oxidative phosphorylation, DNA repair, and interferon-ALPHA response pathways, while naïve B cells in PT samples significantly activated apoptosis, TNFA signaling via NF-κB, and inflammatory response pathways (Figure S1G).

### Identification of major fibroblast types

The GSVA results indicated that fibroblasts in MT samples significantly activated oxidative phosphorylation, DNA repair, coagulation, and glycolysis pathways. In contrast, fibroblasts in PT samples significantly activated G2M-checkpoint, TNFA signaling via NF-κB, and KRAS signaling DN pathways (Fig. 3A). We re-clustered 7,317 fibroblasts and identified 5 subgroups based on their marker genes: smooth muscle cells (SMC, MYH11), pericyte (FRZB), antigen-presenting cancer-associated fibroblasts (apCAFs, HLA-DRB1), pCAFs (TOP2A), and matrix CAFs (mCAFs, MMP11) (Fig. 3B and C). GO and KEGG enrichment analysis showed that these subgroups were closely associated with responses to mechanical stimuli, reactive oxygen genes, epithelial cell proliferation, immune system function, and cell migration (Figure S2). As illustrated in Fig. 3D and E, the proportion of pericytes and mCAFs was significantly higher in BM and OM samples, respectively, when compared to PT samples. Differential gene expression analysis of each fibroblast subpopulation revealed that in the mCAFs subpopulation (Fig. 3F), APOD and CTHRC1 were significantly upregulated in MT samples compared to PT samples. Similarly, in the pCAFs subpopulation, STMN1 and CENPF were significantly upregulated in the MT samples (Fig. 3F, MT vs. PT). Furthermore, we identified ETV1 and E2F2 as the key regulatory transcription factors for the mCAFs and pCAFs subpopulations, respectively (Fig. 3G). These findings suggested that mCAFs and pCAFs subpopulations might play a role in the metastasis of breast cancer.

### Identification of malignant ECs

To identify the clonal structure and cellular origin of malignant cells, we analyzed CNV and clonality of ECs from PT, BM, OM, and LM samples using the inferCNV algorithm. A total of 521, 1,2519, 795, and 30 malignant ECs were identified in the PT (Fig. 4A), BM (Fig. 4B), OM (Fig. 4C), and LM (Fig. 4D) samples, respectively. Loss of chromosomal regions 13q and 7q was observed in malignant ECs from branch J of the PT, BM, OM, and LM samples (Fig. 4A and C). Copy number losses of 115 genes located on chromosomes 7 and 13 were shared among subclonal cell populations with 13q and 7q gains in PT and BM samples (Fig. 4E). Copy number losses of 37 genes located on these chromosomes were identified in subclonal cell populations with 13q and 7q gains in the PT and OM samples (Fig. 4F). Subclonal cell populations with 13q and 7q gains in LM samples are shown in Figure S3A. Venn plots indicated that 23 gene copy number variations were shared between the BM and OM samples (Fig. 4G). Of these 23 genes, the high expression of SBDS, POR, YWHAG, and MDH2 was closely correlated with the prognosis of breast cancer (Figure S3B, $p < 0.05$). The proportion of malignant ECs was significantly higher in OM and BM samples compared to PT and LM samples (Fig. 4H and I).

### Identification of major malignant ECs

Subsequently, we re-clustered 13,865 malignant ECs in the PT and MT samples and identified 4 subclusters: G0, G1, G2, and G3 (Fig. 5A). The G2 subcluster accounts for more of the PT, LM, and OM samples, while the G0 subgroup accounts for more of the BM samples (Fig. 5B and C). GSVA result showed that G1 subcluster was significantly activated MYC-targets-V1, MYC-targets-V2, and E2F-targets pathways (Fig. 5D). G2 subcluster was significantly activated epithelial-mesenchymal transition and hedgehog signaling pathways (Fig. 5D). G3 subcluster was significantly activated G2M-checkpoint, oxidative phosphorylation, and reactive oxygen species pathways (Fig. 5D). We analyzed the cancer stem cell (CSC) marker genes[23–25], the characteristic genes associated with tumor proliferation ability[26,27], and the characteristic genes related to tumor migration ability
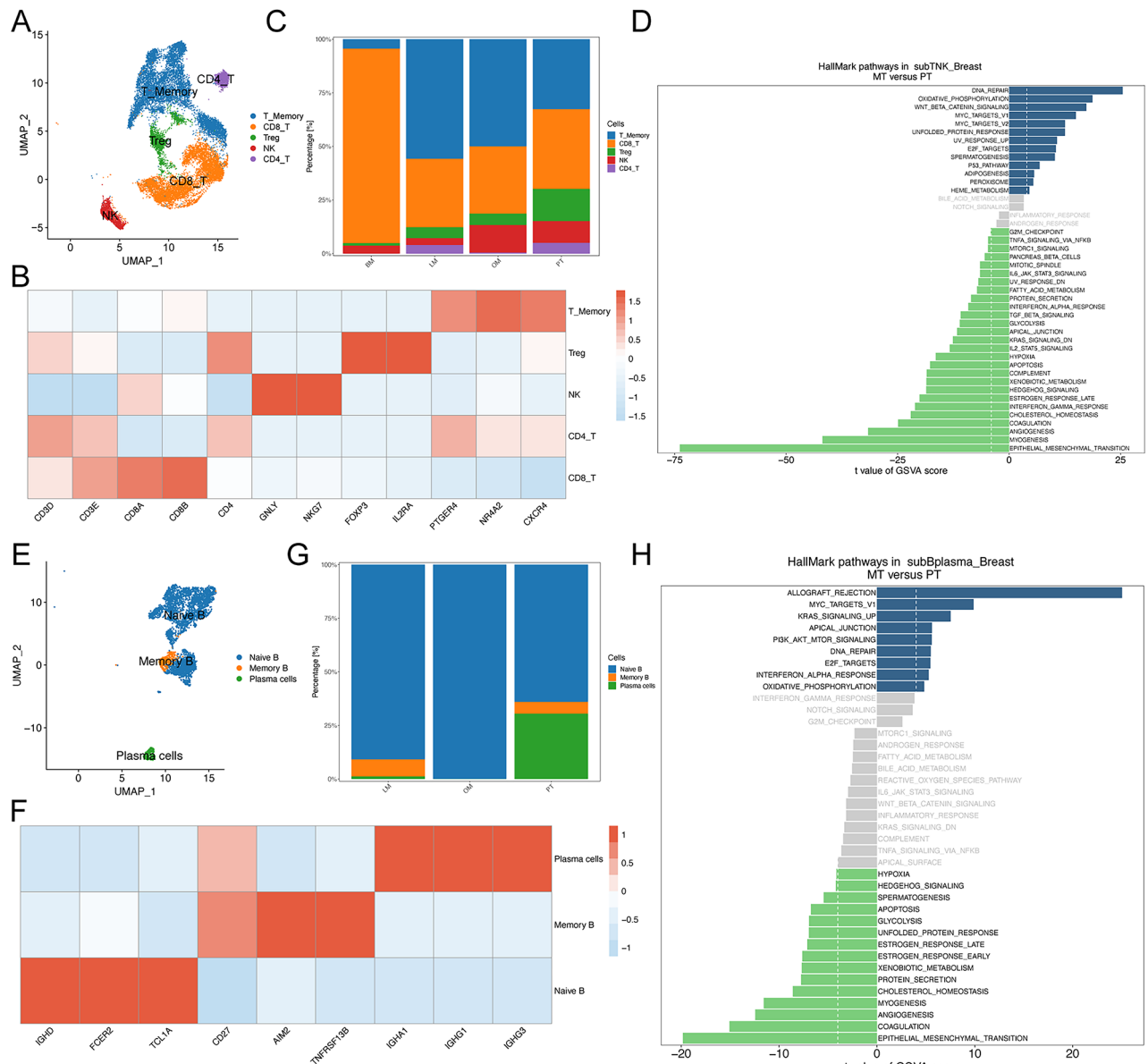
**Fig. 2.** T and B cell repopulation. (**A**), T cell repopulation results. (**B**), The expression of marker genes in different T cell types. (**C**), The proportion of T cell types in primary, lymphatic node, brain, and bone metastatic samples. (**D**), Gene set variation analysis (GSVA) reveals differences in hallmark pathways of T-NK cells in primary tumor (PT) and metastasis (MT) samples. (**E**), (**B**) cell repopulation results. (**F**), The expression of marker genes in different B cell types. (**G**), The proportion of B cell types in primary, lymphatic node, brain, and bone metastatic samples. (**H**), Gene set variation analysis (GSVA) reveals differences in hallmark pathways of B cells in primary tumor (PT) and metastasis (MT) samples. PT: Primary tumor; BM: brain metastases; LM: lymphatic node metastases; OM: bone metastases; MT: metastases tumor.

in each subpopulation[28,29]. The results indicated that the G2 subpopulation exhibited high expression levels of stemness and migration-related genes, while the G3 subpopulation demonstrated elevated expression of proliferation-related genes (Fig. 5E and G), suggesting that G2 cells have an EMT-induced CSC phenotype. Moreover, the pseudotime trajectory axis showed dynamic properties and heterogeneity of malignant epithelial cells (Fig. 5H). Specifically, G3 cells with an EMT-induced CSC phenotype were initially observed, followed by G2 cells exhibiting proliferative characteristics. Dormancy-like tumor G0 and proliferation-associated G1 cells were located in separate trajectory branches, indicating distinct differentiation states. We also found that genes that were gradually highly expressed from pre-branch to cell fate2 significantly activated the MYC-targets-V1 and oxidative phosphorylation pathways, and G1 cells were located in cell fate2 at the end of the differentiation trajectory (Fig. 5H and J). The invasion score of G2 cells was elevated during the early stages of pseudo-time and gradually declined over time, while the proliferation score of G3 cells peaked in the middle stages of pseudo-time. In summary, G2 stem cells exhibiting strong invasive capabilities may represent malignant cells with
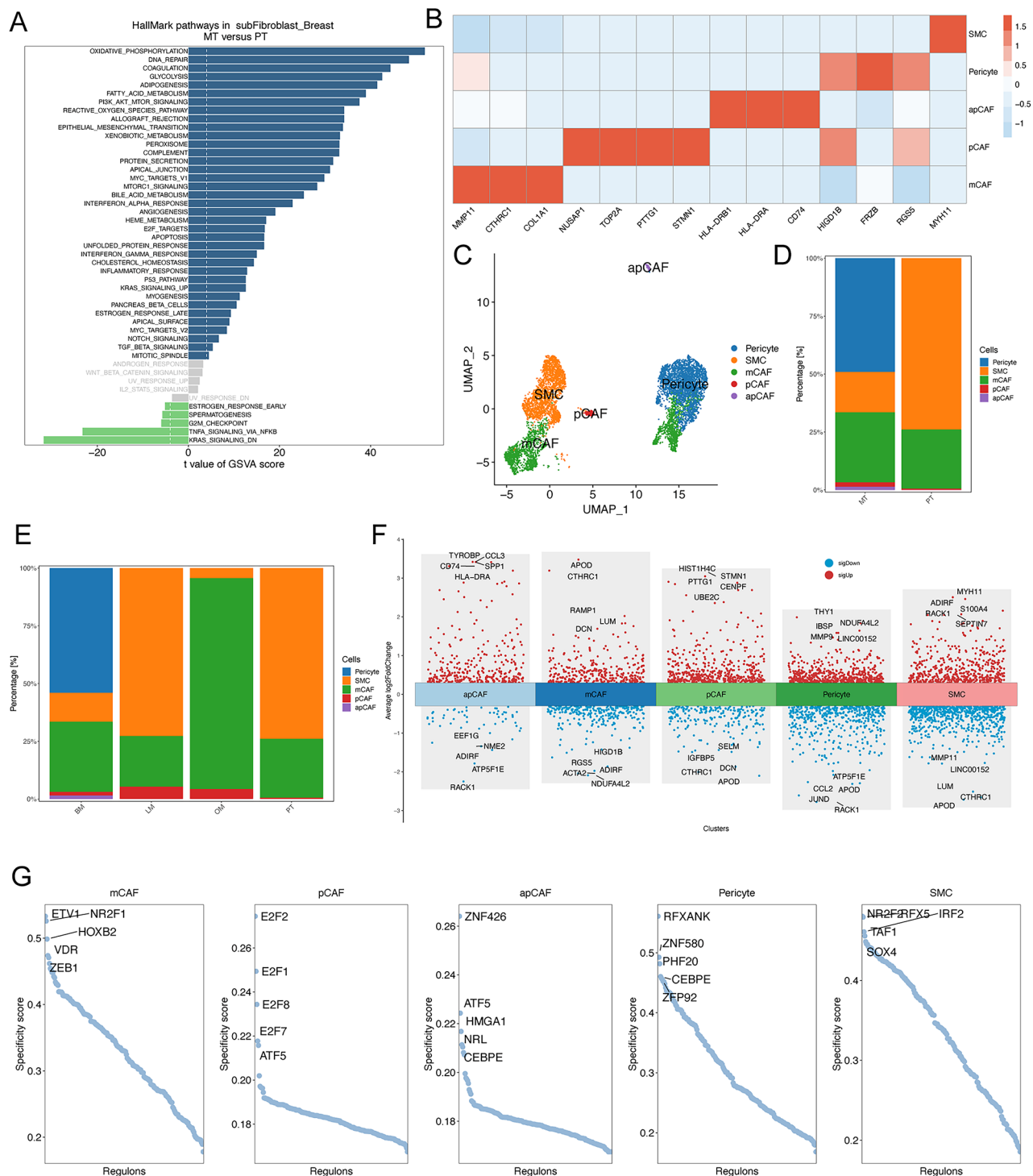
**Fig. 3**. Fibroblast repopulation. (**A**), Gene set variation analysis (GSVA) reveals differences in hallmark pathways of fibroblasts in primary tumor (PT) and metastasis (MT) samples. (**B**), The expression of marker genes in different fibroblast types. (**C**), Fibroblast repopulation results. (**D**), Percentage of fibroblast subpopulations in primary tumor (PT) and metastasis (MT) samples. (**E**), The proportion of fibroblast types in primary, lymphatic node, brain, and bone metastatic samples. (**F**), Differentially expressed genes between primary tumor (PT) and metastasis (MT) samples in different fibroblast types. (**G**), Scatter plot showing the regulon specificity score (RSS) in each fibroblast subtype, with the top 5 transcription factors highlighted. PT: Primary tumor; BM: brain metastases; LM: lymphatic node metastases; OM: bone metastases; MT: metastases tumor.
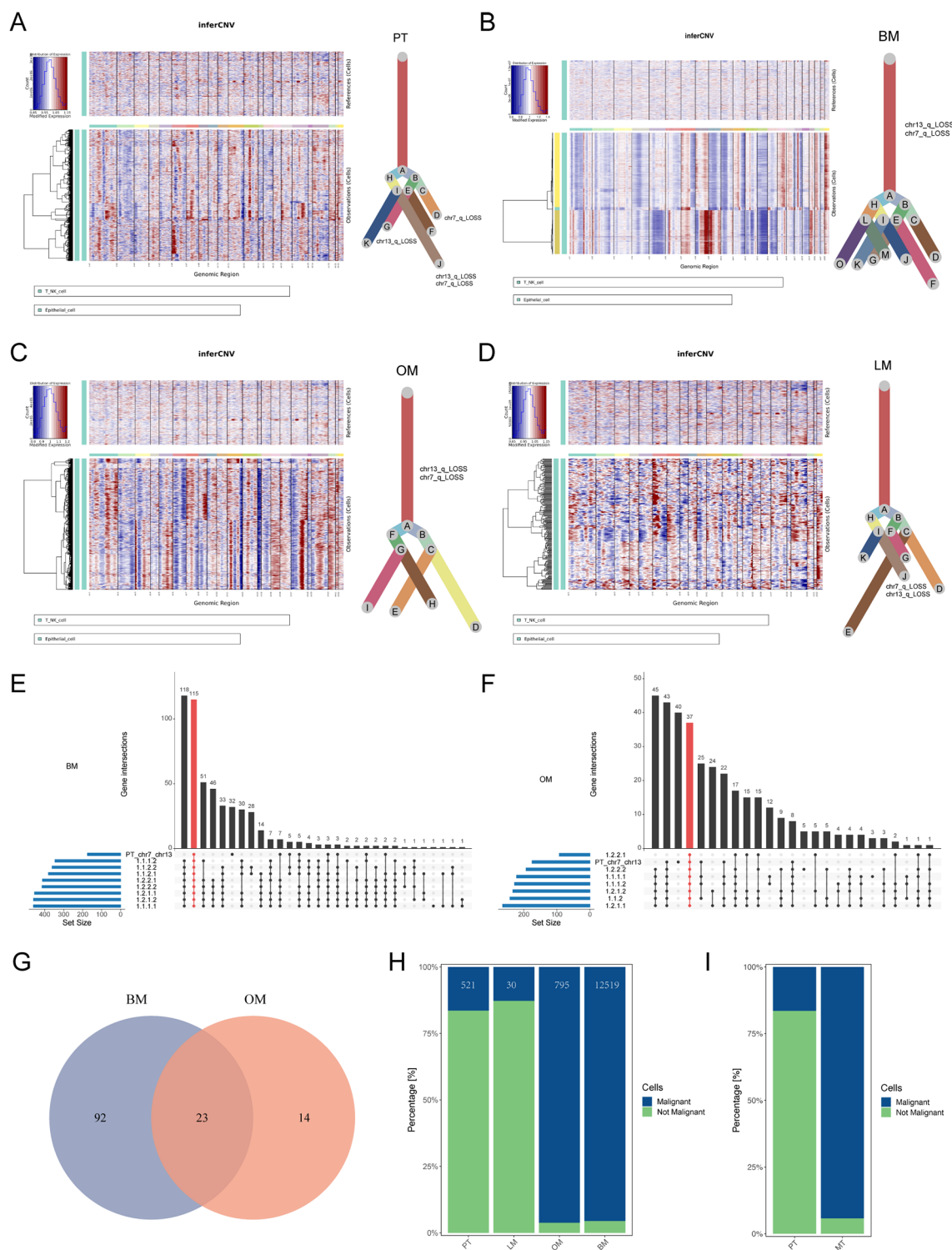
**Fig. 4**. Identification of malignant epithelial cells. (**A**–**D**), the copy number variation (CNV) and phylogenetic tree of epithelial cells in primary, lymphatic node, brain, and bone metastatic samples. UpSet plots reveal the number of genes shared by subclones with 13q and 7q loss in PT and BM (**E**), and PT and OM (**F**). Red bars and dots represent genes shared by all subclones. **G**, Venn diagram displaying shared signature genes between primary, brain, and bone metastatic malignant cells. **H**, The fraction of malignant epithelial cells in primary and different metastatic samples. **I**, The proportion of malignant epithelial cells in primary tumor (PT) and metastasis (MT) samples. PT: Primary tumor; BM: brain metastases; LM: lymphatic node metastases; OM: bone metastases; MT: metastases tumor.
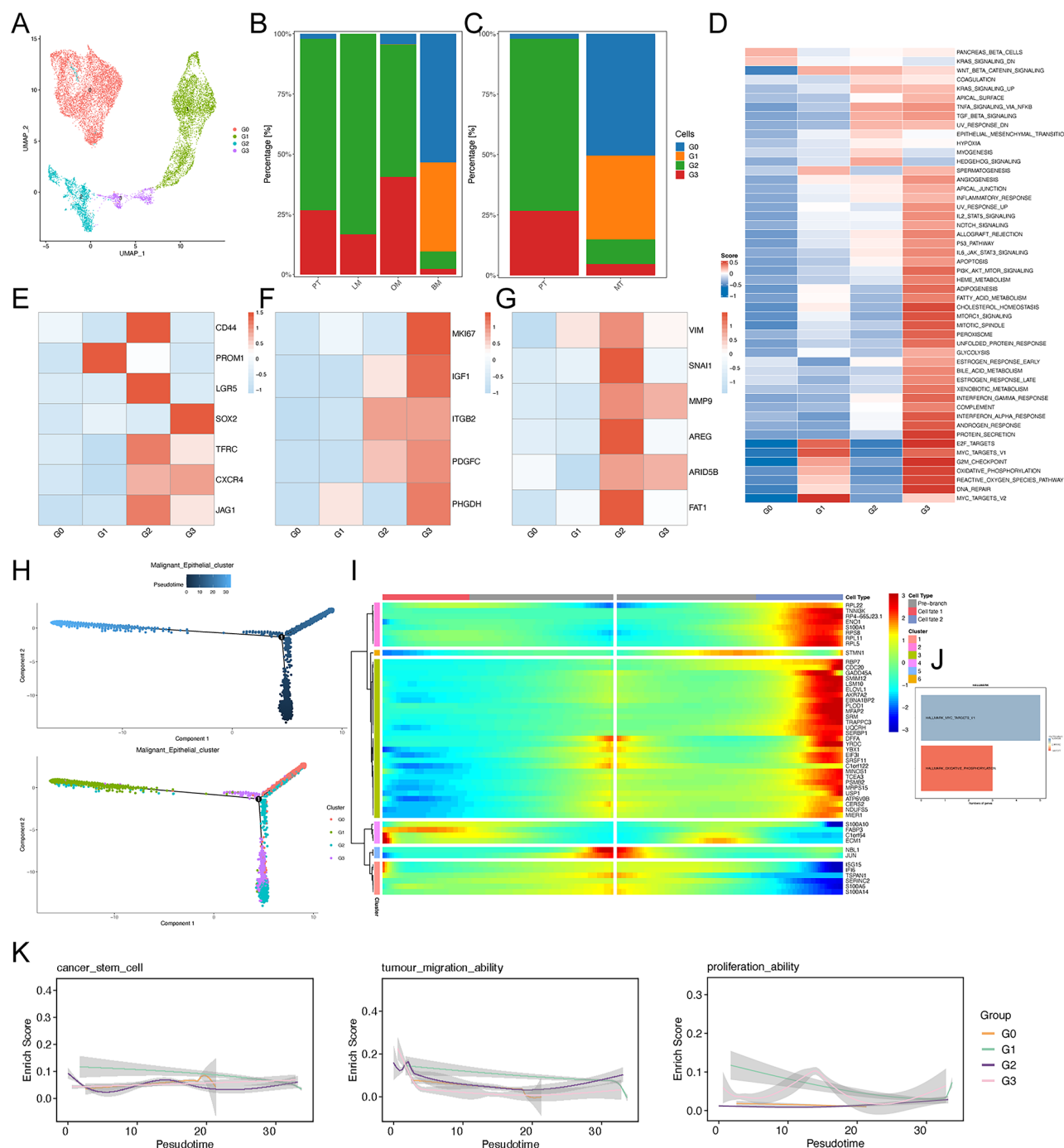
**Fig. 5**. Malignant epithelial cell repopulation. (**A**), The results of malignant epithelial cell repopulation. (**B**), The proportion of malignant epithelial cell types in primary, lymphatic node, brain, and bone metastatic samples. (**C**), The proportion of malignant epithelial cell types in primary tumor (PT) and metastasis (MT) samples. (**D**), Gene set variation analysis (GSVA) reveals differences in hallmark pathways across malignant cell subpopulations. (**E**), The heat map shows the expression of cancer stem cell (CSCs) markers (**E**), characteristic genes for tumor proliferation (**F**), and characteristic genes for tumor migration (**G**) in various subpopulations of malignant cells. (**H**), Temporal analysis results of each subpopulation of malignant epithelial cells. (**I**), Temporal analysis of top50 gene expression heat map. (**J**), Genes in the pseudo-time-related cluster 3 significantly activate the MYC_TARGETS_V1 and OXIDATIVE_PHOSPHORYLATION pathways. (**K**), Changes in cancer cell stemness, proliferation, and invasion pathways along the pseudo-timeline. T: Primary tumor; BM: brain metastases; LM: lymphatic node metastases; OM: bone metastases; MT: metastases tumor.

enhanced metastatic potential. Following metastasis, these cells can differentiate into G3 cells, which possess robust proliferative abilities, and proliferation-related G1 cells. G0 cells, which are associated with dormant tumors, are terminally differentiated cells (Fig. 5K).

### Cell-cell interactions

Next, we analyzed the interaction between various subpopulations of fibroblasts and malignant cells in PT samples, and found that pCAFs had the largest number of interactions with other cells (Fig. 6A). The above results indicated that G2 cells were metastasis-related malignant epithelial cells. Thus, we analyzed significant cell-cell interaction pathways using CellChat to investigate the cell-to-cell interaction pathway between G2 cells and pCAFs, and discovered that KITL-KIT was found to be present only in the cellular communication of pCAFs to G2 cells, suggesting that KITL-KIT was identified as a significant interaction pair between pCAFs and G2 cells (Fig. 6B), and the corresponding pathway being the KIT signaling pathway. The only cells involved
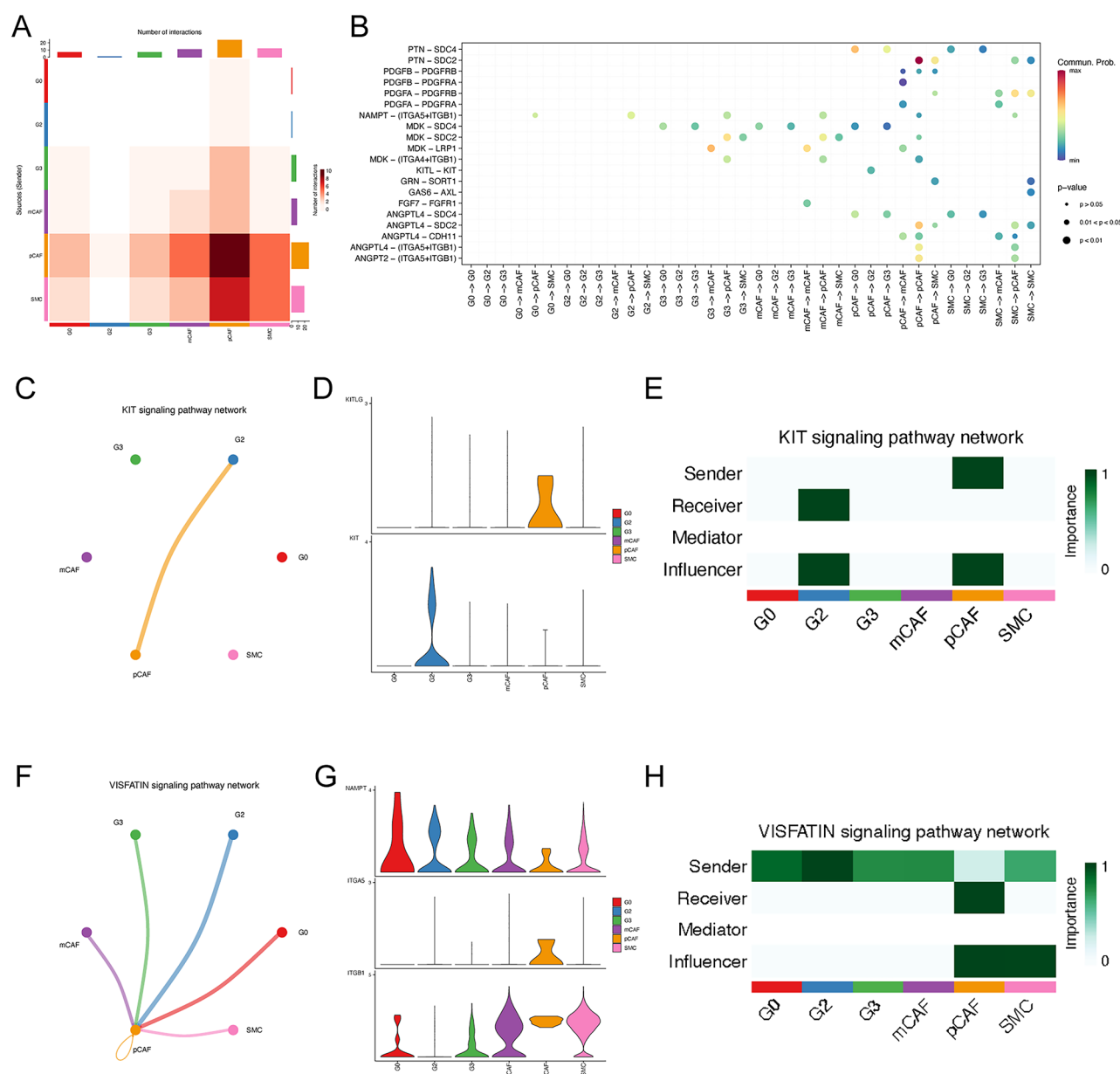


**Fig. 6.** Cell-cell interaction in primary samples. (**A**), Heat map showing the number of interactions between CAFs and cancer cell subpopulations. (**B**), The interaction between CAFs and cancer cell subpopulations is regulated by ligand pairs. (**C**), Chord diagram demonstrate cell-cell interactions involving the KIT signaling pathway. (**D**), The expression of KITL and KIT in different cell types. (**E**), The situation of each cell signaling network in the KIT signaling pathway. (**F**), Chord diagram demonstrate cell-cell interactions involving VISFATIN signaling pathway. (**G**), The expression of NAMPT, ITGA5, and ITGB1 in different cell types. (**H**), Cell signaling networks in the VISFATIN signaling pathway.
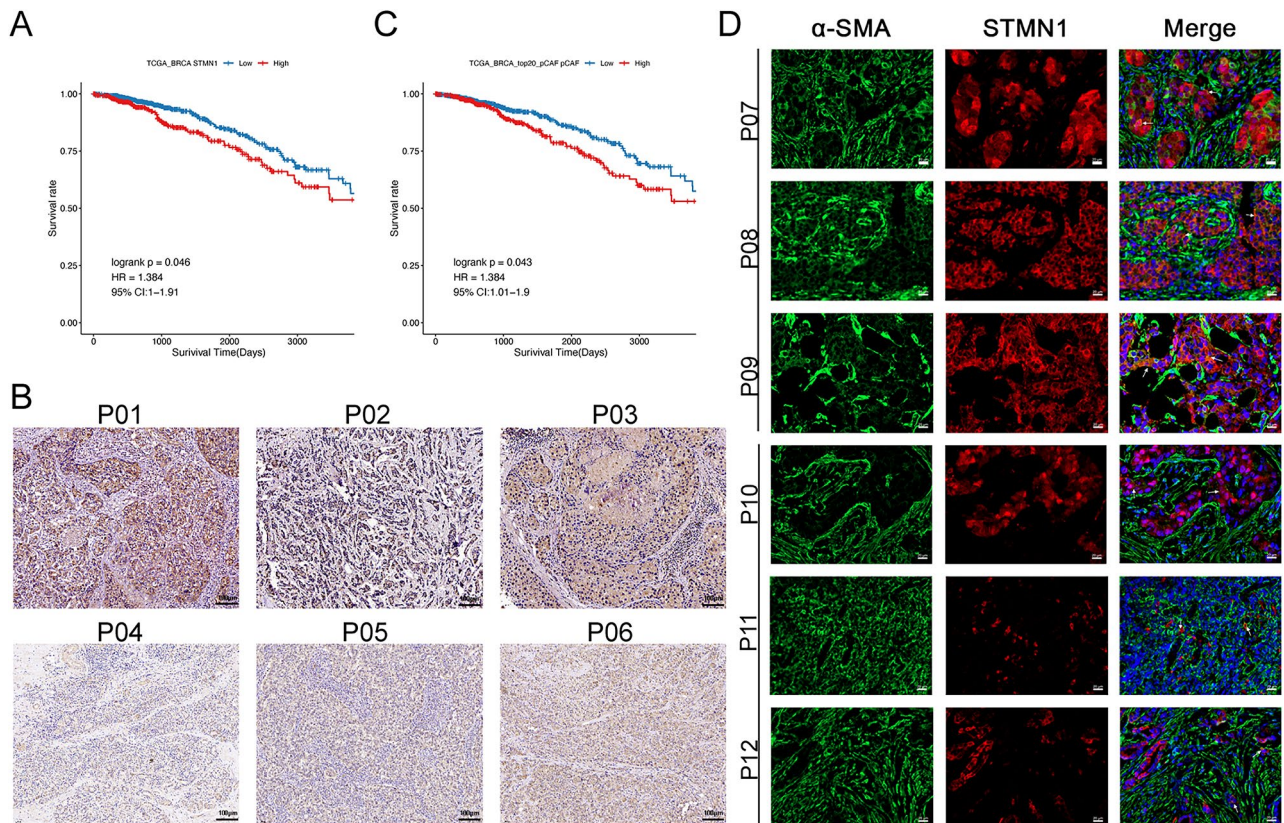
**Fig. 7.** Increased pCAFs were associated with advanced breast cancer and poor prognosis. (**A**), The marker gene STMN1 of pCAFs is associated with poor prognosis of breast cancer. (**B**), Immunohistochemical analysis of STMN1 expression in early and advanced breast cancer samples. Scale bar represents 100 μm. (**C**), The role of pCAFs in the prognosis of breast cancer patients. (**D**), Multispectral immunohistochemical analysis of the proportion of pCAFs cells (a-SMA + STMN1+) in early and advanced breast cancer samples. Scale bar represents 20 μm.

in this pathway are pCAFs and G2 cells (Fig. 6C). Notably, KITL was only expressed in pCAFs cells, while its receptor KIT was only expressed in G2 cells (Fig. 6D). These results suggested that pCAFs increased the invasion and metastasis capabilities of G2 tumor cells by binding to its receptor KIT through KITL (Fig. 6E). Moreover, we analyzed the expression of KITL (KITLG) and KIT in various cell types from PT (Figure S4A) and MT (Figure S4B) samples. Our findings indicated that KITLG was highly expressed exclusively in pCAFs of PT, while other cell types exhibited minimal to no expression. Conversely, KIT was predominantly expressed in the G2 subpopulation of malignant cells in PT; notably, the G2 subpopulation of malignant cells in MT samples did not express KIT, and the remaining cells showed minimal to no expression. Furthermore, we examined cell-cell interactions between fibroblast subpopulations and malignant cell subpopulations in MT samples and found no evidence of a KITL-KIT ligand-receptor pair. Additionally, we assessed the expression of KITL (KITLG) and KIT in PT and MT samples from TCGA breast cancer datasets, revealing that KITL (KITLG) expression was lower in PT samples. In contrast, KIT expression was higher in PT samples (Figure S4C, S4D); however, the statistically insignificant difference may be attributed to the limited number of metastatic samples analyzed (only six cases). These results suggested that the expression of KIT in the G2 subpopulation of malignant cells diminished at the metastasis site, leading to a reduction in the functional regulation mediated by the KITL-KIT ligand-receptor pair.

The regulatory receptor-ligand pair for pCAFs in G2 cells was NAMPT-(ITGA5 + ITGB1), corresponding to the pathway VISFATIN (Fig. 6B). Cells that regulate pCAFs through the VISFATIN pathway include G0, G3, mCAFs, and SMC, among which G2 cells are the most likely (Fig. 6F). Moreover, NAMPT was expressed in a variety of cells, while ITGA5 was only expressed in pCAFs (Fig. 6G). These findings suggested that G0, G2, G3, mCAF, and SMC bind to pCAFs receptors (ITGA5 + ITGB1) through NAMPT to regulate the function of pCAFs (Fig. 6H).

### Increased pCAFs were associated with advanced breast cancer and poor prognosis

To demonstrate the clinical relevance of pCAFs in breast cancer, we analyzed the impact of the marker gene STMN1 in pCAFs on the prognosis of breast cancer patients. Our findings indicated that high expression levels of STMN1 were associated with a poor prognosis in breast cancer (Fig. 7A). Subsequently, we performed IHC staining on samples from both early and advanced breast cancer and observed that STMN1 was highly

expressed in advanced breast cancer samples (Fig. 7B). To further investigate the influence of pCAFs on breast cancer prognosis, we analyzed the pCAFs score in TCGA breast cancer samples, revealing that samples with elevated pCAFs scores exhibited significantly worse survival outcomes (Fig. 7C). Additionally, mIHC results demonstrated an increased proportion of pCAFs cells (a-SMA + STMN1+) in late-stage samples (Fig. 7D). These results indicated that increased pCAFs were associated with advanced breast cancer and poor prognosis.

## Discussion

Distant metastasis to lymphatic nodes, brain, and bone is the leading cause of breast cancer death and is associated with poor prognosis. In the current study, we have provided a comprehensive cellular landscape of breast cancer metastases to these different sites. Our findings have revealed unique characteristics of tumor cells and uncovered the regulatory role of pCAFs by metastatic malignant epithelial cells in breast cancer. Furthermore, we also found that increased pCAFs were associated with advanced breast cancer and a poor prognosis. These findings will provide important evidence to support the diagnosis, treatment, and molecular mechanisms of metastasis in breast cancer.

Within the TME, T cells, B cells, NK cells, and other types of lymphocytes play an important role in the progression of cancer[30]. In TNBC, compared to the immunosuppressive microenvironment of HER2 + breast cancer, triple-negative breast cancer had a higher proportion of Tregs and CD8 + T cell depletion, accompanied by more plasma cells[31]. We found that breast cancer patients with BM had significantly more CD8 + T cells and significantly fewer B cells than breast cancer patients with LM and OM. This finding provides a new perspective for us to understand the immune microenvironment under different transfer modes. A high proportion of CD8 + T cells often signifies that the TME is in an "inflammatory state," which may contribute to the inhibition of tumor growth and spread. Research demonstrated that the proportion of CD8 + T cells was significantly elevated in kataegic breast BM, which might reflect an enhanced immune system response in these tumors[32]. In addition, higher proportion of CD8 + T cells was observed in smaller breast cancer BM, suggesting that an immune response might be initiated during the early stages of breast cancer BM; however, as the tumor progresses, the mechanisms of immune evasion might increasingly prevail[33]. Zou et al. reported that the infiltration of NK and CD8 + T cells was higher in breast cancer with live metastasis compared to breast cancer with MB[34]. After tumor liver metastasis occurs, the liver microenvironment changes, resulting in the "retention" of systemic CD8 + T cells in the liver and their apoptosis, thus evading the body's anti-tumor immunity and immunotherapy[35]. Together, these findings suggested that the increase in CD8 + T cells observed in breast cancer BM might suggest a sustained immune response from the patient that inhibited tumor progression; however, this elevation could also be associated with T cell exhaustion, which inhibited the immune response and consequently facilitated tumor immune escape.

GSVA results indicated that T-NK cell subsets significantly activated different signaling pathways in PT and MT samples, and these pathways were closely related to the biological behavior of tumors and the TME. Oxidative phosphorylation is the primary pathway through which cells generate energy (ATP) via the mitochondrial respiratory chain. It has been reported that in breast cancer with LM samples, the oxidative phosphorylation pathway was significantly activated, indicating that metastatic breast cancer cells may rely more on aerobic respiration to meet their energy needs to support their growth and invasion capabilities[36]. Aberrant activation of the Wnt signaling pathway is a critical driver of breast cancer recurrence and progression. Studies indicated that activation of the Wnt/β-catenin signaling pathway can promote the proliferation and migration of breast cancer cells, thereby accelerating tumor growth and metastasis[37]. IFN-γ is a type II interferon mainly produced by activated T cells and NK cells. In HER2 + breast cancer, tumor-infiltrating natural killer cells (TI-NK) produce CCL5 and IFN-ɣ when activated, thereby triggering an effective anti-tumor immune response[38]. IFNγ can promote anti-tumor immune responses through multiple mechanisms, including directly inhibiting tumor cell proliferation, promoting tumor cell apoptosis, and enhancing the anti-tumor activity of T cells[39]. In the present study, T-NK cells in MT samples significantly activated oxidative phosphorylation and WNT-beta-catenin signaling pathways; while T-NK cells in PT samples significantly activated interferon-gamma response signaling pathways. These findings suggest that studying oxidative phosphorylation and activation of the WNT-β-catenin signaling pathway in PT samples may provide insights into developing therapeutic strategies for breast cancer metastasis.

Previous studies have shown that CAF cells play an essential role in the process of tumor invasion and metastasis. The GSVA results indicated that fibroblasts in MT samples significantly activated oxidative phosphorylation and glycolysis pathways. In contrast, fibroblasts in PT samples significantly activated G2M-checkpoint pathways. It has been reported that during the evolvement of the early dissemination of breast cancer, the early-disseminated cancer cell cluster undergoes a transition between glycolysis and oxidative phosphorylation and tends to distribute along the tumor border[36], suggesting that oxidative phosphorylation may play an important role in lymph node metastasis. The G2/M DNA damage checkpoint prevents cells with genomic DNA damage from entering mitosis (M phase). Inhibition of the G2/M checkpoint can disrupt the DNA damage response in cancer cells, allowing them to divide despite the presence of DNA damage, which exacerbates genomic instability and can lead to cell apoptosis[40–42]. Therefore, investigating the activation of various pathways, including oxidative phosphorylation and the G2/M DNA checkpoint, in MT and PT samples may yield valuable insights for the treatment and development of therapeutic strategies against breast cancer metastasis.

Furthermore, mCAFs were elevated in OM, and pCAFs were increased in breast cancer with LM, BM, and OM compared to PT. It has been demonstrated that mCAFs could enhance the growth and metastasis of breast cancer[43]. In comparison to PT samples, mCAFs exhibited higher expression levels of APOD and CTHRC1, whereas pCAFs demonstrated increased expression of STMN1 and CENPF in MT samples. Elevated levels of APOD and CTHRC1 in breast cancer have been associated with poor prognosis[44,45], suggesting that mCAFs

may be correlated with unfavorable outcomes in breast cancer. STMN1 is a protein linked to cancer metastasis, and its expression in breast cancer correlates with tumor metastasis and chemotherapy resistance[46]. Additionally, overexpression of CENPF was associated with poor prognosis in breast cancer with BM and might promote tumor BM by activating the PI3K-AKT-mTORC1 signaling pathway[47]. In the present study, we found that increased pCAFs were associated with advanced breast cancer and poor prognosis. These findings suggested that higher levels of pCAFs might correlated with the metastasis of breast cancer.

Communication between CAFs and cancer cells promotes tumor metastasis[48]. Cell-cell communication demonstrated an interaction between pCAFs and metastasis-associated malignant ECs. Notably, we found that KITL was expressed in pCAFs cells, while its receptor KIT was expressed in metastasis-associated malignant ECs. KIT is a receptor tyrosine kinase that activates signaling cascades involved in biological processes such as cell proliferation, differentiation, migration, and survival by binding to stem cell binding factors (SCF, also known as KITL)[49]. The c-Kit/KitL signaling pathway is also implicated in cell migration/mobilization[50]. Accordingly, pCAFs might increase the invasion and metastasis capabilities of malignant ECs by binding to its receptor KIT through KITL in metastatic breast cancer. In addition, the regulatory receptor-ligand pair for pCAFs in malignant ECs was NAMPT-(ITGA5 + ITGB1), corresponding to the VISFATIN pathway. It has been reported that VISFATIN/NAMPT has catabolic and pro-inflammatory properties that induce an inflammatory phenotype of fibroblasts in rheumatoid arthritis[51] and promote chondrocyte apoptosis and extracellular matrix degradation in osteoarthritis[52,53]. Moreover, the main receptors of the VISFATIN pathway, ITGA5 and ITGB1, exhibited a positive association with chondrogenic marker genes and a negative correlation with osteogenic marker genes[54]. In this study, ITGA5 was expressed in pCAFs in breast cancer and pCAFs were increased in breast cancer with OM, indicating that malignant ECs might regulate the function of pCAFs through NAMPT binding to pCAFs receptors (ITGA5 + ITGB1), thereby promoting BM of breast cancer.

The study presented provides a comprehensive analysis of the single-cell transcriptional landscape in primary and metastatic breast cancer, offering valuable insights into the cellular heterogeneity and interactions within the tumor microenvironment. However, it has its limitations and areas for improvement. Frist, single-cell sequencing technologies, while powerful, can introduce technical noise and variability. Future studies could benefit from incorporating multiple replicates or using complementary technologies to validate the results. Secondly, the study identifies key cell-cell interactions, but the complexity of the tumor microenvironment suggests that additional interactions and regulatory mechanisms may be at play. Further exploration of the immune contexture, including other cell types and their interactions, could provide a more complete picture. Finally, while the study provides insights into potential therapeutic targets, such as pCAFs, the direct translation of these findings into clinical applications requires further investigation, including preclinical testing and clinical trials.

## Conclusion

In summary, this study elucidated the potential cellular origins and drivers of breast cancer metastases to lymph nodes, the brain, and bone, utilizing single-cell transcriptomic profiles. Five subpopulations of CAFs were identified, with mCAFs and pCAFs potentially playing significant roles in the metastasis of breast cancer. Additionally, a subpopulation of malignant ECs with strong invasive capabilities was identified, which can differentiate into malignant ECs exhibiting robust proliferative abilities. Further investigations revealed a strong interaction between pCAFs and metastasis-related malignant ECs in breast cancer. Finally, we found that the proportion of pCAF increased in advanced breast cancer and was associated with a poor prognosis of breast cancer. These findings have implications for understanding the metastatic process, identifying novel therapeutic targets, and stratifying patients with breast cancer for more personalized treatment strategies.

## Data availability

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material . All data are available from the TCGA database (https://xenabrowser.net/datapages/) and GEO (https://ncbi.nlm.nih.gov/geo/) database within the article. GEO database under accession number GSE225600, GSE186344, and GSE190772.

## References

1. Barzaman, K. et al. Breast cancer: Biology, biomarkers, and treatments. *Int. Immunopharmacol.* **84**, 106535. https://doi.org/10.1016/j.intimp.2020.106535 (2020).
2. Bray, F. et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **74**, 229–263. https://doi.org/10.3322/caac.21834 (2024).
3. Ye, F. et al. Advancements in clinical aspects of targeted therapy and immunotherapy in breast cancer. *Mol. Cancer*. **22**, 105. https://doi.org/10.1186/s12943-023-01805-y (2023).
4. Marino, M. A., Avendano, D., Zapata, P., Riedl, C. C. & Pinker, K. Lymph Node Imaging in Patients with Primary Breast Cancer: Concurrent Diagnostic Tools. *The oncologist* 25, e231-e242, (2020). https://doi.org/10.1634/theoncologist.2019-0427
5. Tahara, R. K., Brewer, T. M., Theriault, R. L. & Ueno, N. T. Bone metastasis of breast Cancer. *Adv. Exp. Med. Biol.* **1152**, 105–129. https://doi.org/10.1007/978-3-030-20301-6_7 (2019).
6. Sun, H., Xu, J., Dai, S., Ma, Y. & Sun, T. Breast cancer brain metastasis: current evidence and future directions. *Cancer Med.* **12**, 1007–1024. https://doi.org/10.1002/cam4.5021 (2023).
7. Place, A. E., Huh, J., Polyak, K. & S. & The microenvironment in breast cancer progression: biology and implications for treatment. *Breast Cancer Res.* 13 https://doi.org/10.1186/bcr2912 (2011).
8. Dias, A. S., Almeida, C. R., Helguero, L. A. & Duarte, I. F. Metabolic crosstalk in the breast cancer microenvironment. *Eur. J. cancer (Oxford England: 1990)*. **121**, 154–171. https://doi.org/10.1016/j.ejca.2019.09.002 (2019).

9. Lee, H., Na, K. J. & Choi, H. Differences in Tumor Immune Microenvironment in Metastatic sites of breast Cancer. *Front. Oncol.* **11**, 649004. https://doi.org/10.3389/fonc.2021.649004 (2021).

10. Gruosso, T. et al. Spatially distinct tumor immune microenvironments stratify triple-negative breast cancers. *J. Clin. Investig.* **129**, 1785–1800. https://doi.org/10.1172/jci96313 (2019).

11. Gaggioli, C. et al. Fibroblast-led collective invasion of carcinoma cells with differing roles for RhoGTPases in leading and following cells. *Nat. Cell Biol.* **9**, 1392–1400. https://doi.org/10.1038/ncb1658 (2007).

12. Su, S. et al. CD10(+)GPR77(+) Cancer-Associated fibroblasts promote Cancer formation and chemoresistance by sustaining Cancer Stemness. *Cell* **172**, 841–856e816. https://doi.org/10.1016/j.cell.2018.01.009 (2018).

13. Eiro, N. et al. Cancer-associated fibroblasts affect breast cancer cell gene expression, invasion and angiogenesis. *Cell. Oncol. (Dordrecht Netherlands).* **41**, 369–378. https://doi.org/10.1007/s13402-018-0371-y (2018).

14. Hu, D. et al. Cancer-associated fibroblasts in breast cancer: challenges and opportunities. *Cancer Commun. (London England).* **42**, 401–434. https://doi.org/10.1002/cac2.12291 (2022).

15. Kim, H. M., Jung, W. H. & Koo, J. S. Expression of cancer-associated fibroblast related proteins in metastatic breast cancer: an immunohistochemical analysis. *J. Translational Med.* **13** https://doi.org/10.1186/s12967-015-0587-9 (2015).

16. Pelon, F. et al. Cancer-associated fibroblast heterogeneity in axillary lymph nodes drives metastases in breast cancer through complementary mechanisms. *Nat. Commun.* **11**, 404. https://doi.org/10.1038/s41467-019-14134-w (2020).

17. Hou, J. et al. Unveiling heterogeneity and prognostic markers in ductal breast cancer through single-cell RNA-seq. *Cancer Cell. Int.* **24**, 266. https://doi.org/10.1186/s12935-024-03325-1 (2024).

18. Sanjaya, A., Ratnawati, H., Adhika, O. A. & Rahmatilah, F. R. The heterogeneity of breast cancer metastasis: a bioinformatics analysis utilizing single-cell RNA sequencing data. *Breast Cancer Res. Treat.* **208**, 379–390. https://doi.org/10.1007/s10549-024-07428-1 (2024).

19. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587e3529. https://doi.org/10.1016/j.cell.2021.04.048 (2021).

20. Gribov, A. et al. SEURAT: visual analytics for the integrated analysis of microarray data. *BMC Med. Genomics.* **3** https://doi.org/10.1186/1755-8794-3-21 (2010).

21. Sun, H. et al. Multi-omics analysis-based macrophage differentiation-associated papillary thyroid cancer patient classifier. *Transl Oncol.* **43**, 101889. https://doi.org/10.1016/j.tranon.2024.101889 (2024).

22. Zhang, X. et al. CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Res.* **47**, D721–d728. https://doi.org/10.1093/nar/gky900 (2019).

23. Takaishi, S. et al. Identification of gastric cancer stem cells using the cell surface marker CD44. *Stem Cells (Dayton Ohio).* **27**, 1006–1020. https://doi.org/10.1002/stem.30 (2009).

24. Walcher, L. et al. Cancer Stem cells-origins and biomarkers: perspectives for targeted personalized therapies. *Front. Immunol.* **11**, 1280. https://doi.org/10.3389/fimmu.2020.01280 (2020).

25. Bocci, F. et al. Toward understanding cancer stem cell heterogeneity in the tumor microenvironment. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 148–157. https://doi.org/10.1073/pnas.1815345116 (2019).

26. Menon, S. S., Guruvayoorappan, C., Sakthivel, K. M. & Rasmi, R. R. Ki-67 protein as a tumour proliferation marker. *Clin. Chim. Acta.* **491**, 39–45. https://doi.org/10.1016/j.cca.2019.01.011 (2019).

27. de la Rojo, M., Chapman, E. & Zhang, D. D. NRF2 and the hallmarks of Cancer. *Cancer cell.* **34**, 21–43. https://doi.org/10.1016/j.ccell.2018.03.022 (2018).

28. Aiello, N. M. & Kang, Y. Context-dependent EMT programs in cancer metastasis. *J. Exp. Med.* **216**, 1016–1026. https://doi.org/10.1084/jem.20181827 (2019).

29. Katoh, M. Function and cancer genomics of FAT family genes (review). *Int. J. Oncol.* **41**, 1913–1918. https://doi.org/10.3892/ijo.2012.1669 (2012).

30. Hinshaw, D. C. & Shevde, L. A. The Tumor Microenvironment innately modulates Cancer Progression. *Cancer Res.* **79**, 4557–4566. https://doi.org/10.1158/0008-5472.Can-18-3962 (2019).

31. Ding, S. et al. Single-cell atlas reveals a distinct immune profile fostered by T cell-B cell crosstalk in triple negative breast cancer. *Cancer Commun. (London England).* **43**, 661–684. https://doi.org/10.1002/cac2.12429 (2023).

32. Álvarez-Prado, Á. Immunogenomic analysis of human brain metastases reveals diverse immune landscapes across genetically distinct tumors. *Cell. Rep. Med.* **4**, 100900. https://doi.org/10.1016/j.xcrm.2022.100900 (2023).

33. Duchnowska, R. et al. Immune response in breast cancer brain metastases and their microenvironment: the role of the PD-1/PD-L axis. *Breast Cancer Res.* **18** https://doi.org/10.1186/s13058-016-0702-8 (2016).

34. Zou, Y. et al. The single-cell Landscape of Intratumoral Heterogeneity and the immunosuppressive microenvironment in liver and brain metastases of breast Cancer. *Adv. Sci. (Weinheim Baden-Wurttemberg Germany).* **10**, e2203699. https://doi.org/10.1002/advs.202203699 (2023).

35. Yu, J. et al. Liver metastasis restrains immunotherapy efficacy via macrophage-mediated T cell elimination. *Nat. Med.* **27**, 152–164. https://doi.org/10.1038/s41591-020-1131-x (2021).

36. Liu, Y. M. et al. Combined single-cell and spatial transcriptomics reveal the metabolic evolvement of breast Cancer during early dissemination. *Adv. Sci. (Weinh).* **10**, e2205395. https://doi.org/10.1002/advs.202205395 (2023).

37. Xu, X., Zhang, M., Xu, F. & Jiang, S. Wnt signaling in breast cancer: biological mechanisms, challenges and opportunities. *Mol. Cancer.* **19**, 165. https://doi.org/10.1186/s12943-020-01276-5 (2020).

38. Santana-Hernández, S. et al. NK cell-triggered CCL5/IFNγ-CXCL9/10 axis underlies the clinical efficacy of neoadjuvant anti-HER2 antibodies in breast cancer. *J. Exp. Clin. Cancer Res.* **43**, 10. https://doi.org/10.1186/s13046-023-02918-4 (2024).

39. Jorgovanovic, D., Song, M., Wang, L. & Zhang, Y. Roles of IFN-γ in tumor progression and regression: a review. *Biomark. Res.* **8**, 49. https://doi.org/10.1186/s40364-020-00228-x (2020).

40. Jung, Y., Kraikivski, P., Shafiekhani, S., Terhune, S. S. & Dash, R. K. Crosstalk between Plk1, p53, cell cycle, and G2/M DNA damage checkpoint regulation in cancer: computational modeling and analysis. *NPJ Syst. Biol. Appl.* **7**, 46. https://doi.org/10.1038/s41540-021-00203-8 (2021).

41. Barnaba, N. & LaRocque, J. R. Targeting cell cycle regulation via the G2-M checkpoint for synthetic lethality in melanoma. *Cell. Cycle.* **20**, 1041–1051. https://doi.org/10.1080/15384101.2021.1922806 (2021).

42. Hsieh, H. J. et al. Systems biology approach reveals a link between mTORC1 and G2/M DNA damage checkpoint recovery. *Nat. Commun.* **9**, 3982. https://doi.org/10.1038/s41467-018-05639-x (2018).

43. Gui, Y. et al. Metastatic breast Carcinoma-Associated fibroblasts have enhanced Protumorigenic properties related to increased IGF2 expression. *Clin. cancer Research: Official J. Am. Association Cancer Res.* **25**, 7229–7242. https://doi.org/10.1158/1078-0432.Ccr-19-1268 (2019).

44. Jankovic-Karasoulos, T. et al. Elevated levels of tumour apolipoprotein D independently predict poor outcome in breast cancer patients. *Histopathology* **76**, 976–987. https://doi.org/10.1111/his.14081 (2020).

45. Wang, Z. et al. CTHRC1 is a potential prognostic biomarker and correlated with macrophage infiltration in breast Cancer. *Int. J. Gen. Med.* **15**, 5701–5713. https://doi.org/10.2147/ijgm.S366272 (2022).

46. Kuang, X. Y. et al. The phosphorylation-specific association of STMN1 with GRP78 promotes breast cancer metastasis. *Cancer Lett.* **377**, 87–96. https://doi.org/10.1016/j.canlet.2016.04.035 (2016).

47. Sun, J. et al. Overexpression of CENPF correlates with poor prognosis and tumor bone metastasis in breast cancer. *Cancer Cell Int.* **19** https://doi.org/10.1186/s12935-019-0986-8 (2019).

48. Ren, Y. et al. Paracrine and epigenetic control of CAF-induced metastasis: the role of HOTAIR stimulated by TGF-ß1 secretion. *Mol. Cancer*. **17** https://doi.org/10.1186/s12943-018-0758-4 (2018).
49. Martinez-Anton, A., Gras, D., Bourdin, A., Dubreuil, P. & Chanez, P. KIT as a therapeutic target for non-oncological diseases. *Pharmacol. Ther*. **197**, 11–37. https://doi.org/10.1016/j.pharmthera.2018.12.008 (2019).
50. Heissig, B. et al. Recruitment of stem and progenitor cells from the bone marrow niche requires MMP-9 mediated release of kit-ligand. *Cell* **109**, 625–637. https://doi.org/10.1016/s0092-8674(02)00754-7 (2002).
51. Brentano, F. et al. Pre-B cell colony-enhancing factor/visfatin, a new marker of inflammation in rheumatoid arthritis with proinflammatory and matrix-degrading activities. *Arthritis Rheum*. **56**, 2829–2839. https://doi.org/10.1002/art.22833 (2007).
52. Cheleschi, S. et al. MicroRNA-34a and MicroRNA-181a Mediate Visfatin-Induced apoptosis and oxidative stress via NF-κB pathway in human osteoarthritic chondrocytes. *Cells* **8** https://doi.org/10.3390/cells8080874 (2019).
53. Liu, C., Cheng, P., Liang, J., Zhao, X. & Du, W. Circular RNA circ_0128846 promotes the progression of osteoarthritis by regulating miR-127-5p/NAMPT axis. *J. Orthop. Surg, Res*. **16**, 307. https://doi.org/10.1186/s13018-021-02428-z (2021).
54. Liao, Z. et al. Single-cell transcriptome analysis reveals aberrant stromal cells and heterogeneous endothelial cells in alcohol-induced osteonecrosis of the femoral head. *Commun. Biology*. **5**, 324. https://doi.org/10.1038/s42003-022-03271-6 (2022).

## Acknowledgements

## Author contributions

L.Z.: wrote the manuscript; M.L. and Y.S.: performed bioinformatics analysis; J.C. and H.Z.: investigation and figure preparation; L.Z. :engaged in the clinical sample collection and immunofluorescence experiments; D.S. and J.Z.: conceptualization, methodology and supervision. All authors contributed to the article and approved the submitted version.

## Declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-85531-z.

**Correspondence** and requests for materials should be addressed to J.Z. or D.S.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.