

RESEARCH

Open Access



Single-cell landscape of intratumoral heterogeneity and tumor microenvironment remodeling in pre-nodal metastases of breast cancer

Kaidong Liu^{1†}, Huiming Han^{1†}, Kai Xiong^{1†}, Songmei Zhai¹, Xiuqi Yang¹, Xinmiao Yu², Bo Chen¹, Mingyue Liu¹, Qi Dong¹, Hongxue Meng^{3*} and Yunyan Gu^{1*} 

Abstract

Background The metastasis of cancer cells is influenced by both their intrinsic characteristics and the tumor microenvironment (TME). However, the molecular mechanisms underlying pre-nodal metastases of breast cancer remain unclear.

Methods We integrated a total of 216,963 cells from 54 samples across 6 single-cell datasets to profile the cellular landscape differences between primary tumors and pre-nodal metastases.

Results We revealed three distinct metastatic epithelial cell subtypes (Epi1, Epi2 and Epi3), which exhibited different metastatic mechanisms. Specifically, the marker gene *KCNK15* of the Epi1 subtype exhibited increased gene expression along the cell differentiation trajectory and was specifically regulated by the transcription factor *ASCL1*. In the Epi3 subtype, we highlighted *NR2F1* as a regulator targeting the marker gene *MUCL1*. Additionally, we found that the Epi2 and Epi3 subtypes shared some regulons, such as *ZEB1* and *NR2C1*. Similarly, we identified specific subtypes of stromal and immune cells in the TME, and discovered that vascular cancer-associated fibroblasts might promote capillary formation through CXCL9⁺ macrophages in pre-nodal metastases. All three subtypes of metastatic epithelial cells were associated with poor prognosis.

Conclusions In summary, this study dissects the intratumoral heterogeneity and remodeling of the TME in pre-nodal metastases of breast cancer, providing novel insights into the mechanisms underlying breast cancer metastasis.

Keywords Breast cancer, Pre-nodal metastasis, Single-cell analysis, Tumor microenvironment remodeling

[†]Kaidong Liu, Huiming Han and Kai Xiong are contributed equally to this work.

*Correspondence:

Hongxue Meng
menghongxue@hrbmu.edu.cn

Yunyan Gu
guyunyan@ems.hrbmu.edu.cn

Full list of author information is available at the end of the article



Background

Globally, breast cancer remains one of the most common malignancies among women [1]. The incidence of breast cancer has been steadily increasing, partly due to increased awareness and improved screening programs [2]. It is estimated that approximately 20–30% of breast cancer cases have evidence of metastasis at the time of diagnosis, and a significant proportion of the remaining cases may eventually develop metastatic disease [3]. Breast cancer metastasis plays an important role in disease progression and is a leading cause of mortality among breast cancer patients [4]. The metastatic process is complex, involving genetic alterations, biological processes, and interactions with cells in the tumor microenvironment. Thus, it is essential to understand the heterogeneity of breast cancer metastasis and gain novel insights into preventing its occurrence and progression.

In recent years, single-cell RNA sequencing (scRNA-seq) technology has revolutionized cancer metastasis research by providing unparalleled insights into the complex cellular heterogeneity [5, 6]. Several scRNA-seq studies related to breast cancer metastasis have been largely focused on cells in the tumor microenvironment (TME), revealing that the immune and stromal landscapes are highly heterogeneous [7, 8]. It is crucial to recognize that cancer cells themselves, as well as the interactions between cancer cells and the TME, play critical roles in breast cancer metastasis. Xu et al. identified nine subclusters of cancer cell using scRNA-seq and highlighted a specific subcluster of breast cancer stem cells that exhibited metastatic potential from primary tumor to lymph node [9]. Nevertheless, the underlying metastatic mechanisms in different cancer cell subclusters, and their interplay with the TME, remain incompletely understood.

According to recent studies, patients with metastatic breast cancer displayed heterogeneity in distant metastases, with bone metastasis being the most prevalent, followed by lung and liver metastases, while brain metastasis accounted for the least frequency [3, 10]. In view of the high heterogeneity of cancer cells in origin, phenotype and function, it remains challenging to elucidate the tissue tropism, or the preferential metastasis to specific organs, of different cancer cell subtypes in breast cancer metastasis. Zou et al. employed scRNA-seq to depict intratumoral heterogeneity and immunosuppressive microenvironment in breast cancer liver and brain metastases. Unfortunately, a comparative analysis of the differences between liver and brain metastases in breast cancer has not been conducted [11], hindering the understanding of the distinct metastatic preferences and mechanisms. In addition, few studies have looked into the tissue tropism of different cancer cell subtypes in the

metastatic process. Therefore, systematic and in-depth research on investigating the tissue tropism and metastatic mechanisms of different cancer cell subtypes, as well as their interactions with the TME, is crucial for gaining profound insights into the mechanisms underlying breast cancer metastasis.

In this study, we integrated public scRNA-seq data to profile the cellular landscape of breast cancer metastasis, comprehensively characterizing the metastatic epithelial cell subtypes and the TME remodeling between primary tumors and pre-nodal metastases. Our analyses revealed that metastatic epithelial cell subtypes exhibit diverse metastatic approaches, with a particular focus on the Epi1 and Epi3 subtypes. In summary, our comprehensive investigation on breast cancer metastasis highlights the possible heterogeneity and plasticity of cancer cells and the TME, providing novel insights into the mechanisms underlying breast cancer metastasis.

Methods

Data sources

Bulk data of breast cancer was downloaded from The Cancer Genome Atlas (TCGA, <https://portal.gdc.cancer.gov/>) using the R package "GDCRNATools" (v 2.27.2). The molecular data, such as somatic mutations and gene expression, as well as clinical characteristics including tumor stage and survival time, were obtained accordingly. Single-cell data of breast cancer was downloaded from the Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) and the National Genomics Data Center (NGDC, <https://www.cnbc.ac.cn/>). To ensure consistency and minimize technical variability, only treatment-naïve samples analyzed using the 10× Genomics platform were included for subsequent analysis. The detailed data information used in this study are listed in Supplementary Table 1.

Single-cell data processing

Based on the available clinical stage or paired lymph node metastasis information, the single-cell samples were classified into two groups: the primary tumor group and the pre-nodal metastasis group. The R package "Seurat" (v 4.3.0) was employed to process the single-cell data. All functions were executed using their default parameters, unless otherwise instructed. Cells with fewer than 200 detected genes or more than 20% mitochondrial content were excluded, as these could indicate low-quality cells. Potential doublets were removed by filtering out cells that had more than three median absolute deviations (MADs) of expressed genes. To mitigate batch effects, we utilized the Harmony algorithm (v 0.1.1) to integrate the gene-cell matrices derived from different samples [12]. After batch effect

correction, a shared nearest-neighbor graph was generated using the FindNeighbors function with the parameter: `dims=1:30`, and clusters were identified using the FindClusters function. The resulting clusters were visualized using the t-Distributed Stochastic Neighbor Embedding (t-SNE), a non-linear dimensional reduction method. Subsequently, cluster-specific genes were calculated using the FindAllMarkers function with the parameter: `only.pos=T`. We annotated the clusters into specific cell types by comparing the cluster-specific genes with canonical cell markers. Sub-clustering of major cell populations, including epithelial cells, T cells, macrophages, B cells, fibroblasts, and endothelial cells, was performed using the same approach.

Identification of metastasis-associated cell subtypes

We utilized the "Scissor" (v 2.0.0) algorithm to link the metastatic status of TCGA bulk data with the single-cell data of breast cancer [13]. The metastatic status of TCGA bulk data was determined based on the clinical stage of the patients, where patients without lymph node metastasis (N0) and distant metastasis (M0) were classified as non-metastatic status, while patients with either lymph node or distant metastasis were classified as metastatic status. The Scissor function was run on epithelial cells with the following parameters: `alpha=0.01`, `family="binomial"`. Scissor+ cells were associated with metastatic status, while Scissor-cells were associated with non-metastatic status.

Copy number variation and clonality analysis

Initial copy number variations (CNVs) for each cell were estimated using the R package "inferCNV" (v 1.16.0, <https://github.com/broadinstitute/inferCNV>). The CNVs of epithelial cells were calculated, with fibroblasts and endothelial cells serving as the reference. After creating the inferCNV object, we performed the inferCNV analysis with the following parameters: `cutoff=0.1`, `denoise=T`. The CNV score was calculated as the quadratic sum of the CNV regions. To infer the clonal CNV changes, we employed the Hidden Markov Model framework and utilized the "subcluster" analysis mode to assign copy number states to genomic regions. By consulting the genomic cytoband information, p- or q-arm level changes were converted to equivalent CNVs based on their locations. The "UPhyloplot2" (v 2.3) algorithm was used to generate intra-tumor phylogenetic trees [14], where the length of each branch is proportional to the percentage of cells. The characteristic CNV changes for each subclone were manually inferred based on the output files of inferCNV.

Functional enrichment analysis

We identified differentially expressed genes (DEGs) for each cluster using the FindMarkers function in Seurat. Only the differentially up-regulated genes were obtained for further analysis. Gene Ontology (GO) enrichment analysis was conducted on these differentially up-regulated genes using the R package "clusterProfiler" (v 4.8.2).

Trajectory analysis

We utilized the Monocle (v2.28.0) algorithm to construct single-cell pseudotime trajectory of epithelial cell subtypes [15]. The DEGs identified by the FindAllMarkers function in Seurat were used to order the cells in the trajectory. Subsequently, we determined the root of the trajectory based on the degree of cell differentiation and de-differentiation. The CytoTRACE (v 0.3.3) algorithm was utilized to predict and visualize the cell differentiation scores along the trajectory [16]. Cell de-differentiation scores were computed using the AUCell algorithm in Seurat based on a set of epithelial–mesenchymal transition (EMT) gene sets obtained from the Molecular Signatures Database (MSigDB, <https://www.gsea-msigdb.org/gsea/msigdb>). The branch displaying high cell differentiation scores and low cell de-differentiation scores was selected as the root. The branch expression analysis modeling (BEAM) function in Monocle was then performed on the branch point to identify genes that exhibited branch-dependent expression patterns, where cell fate 1 corresponded to the state 2, and cell fate 2 corresponded to the state 1.

SCENIC analysis

We utilized the pySCENIC (v0.12.1) algorithm to identify specific gene regulatory networks between epithelial cell subtypes [17]. First, the grnboost2 function was used to predict potential regulatory interactions between transcription factors (TFs) and their target genes. Then, the inferred regulatory networks were pruned using cis-regulatory motifs, resulting in the identification of regulons. Finally, the AUCell function was used to score the activity of each regulon in each cell. The regulon specificity score (RSS) between epithelial cell subtypes was calculated using the calcRSS function in R. Furthermore, Pearson correlation analysis was performed to identify functional modules within the TF regulons and to calculate the correlations between epithelial cell subtypes. The binding site information of the TFs in each module was obtained from the JASPAR database (<https://jaspar.elixir.no/>). CYTOSCAPE software (v3.8.2, <https://cytoscape.org/>) was used to visualize the gene regulatory networks.

Distant metastasis analysis

We collected two breast cancer bone metastasis samples from the GSE190772 dataset and one breast cancer brain metastasis sample from the GSE143423 dataset. The single-cell data processing for these samples followed the same procedure as described above. Based on the DEGs of metastatic epithelial cell subtypes, we calculated Epi scores for each distant metastatic sample using the AddModuleScore function in Seurat. The Wilcoxon rank-sum test was applied to determine whether the Epi scores showed statistically significant differences between different metastatic epithelial cell subtypes.

Cell interaction analysis

We utilized the CellChat (v 1.6.1) algorithm to investigate the potential interactions between different cell types [18]. The combined Seurat object, comprising both epithelial cells and other cell types in the TME, was used as input to the algorithm. After creating the CellChat object, we set up the reference database using the secretory signaling pathways. The specific receptor-ligand interactions and communication probabilities between different cell types were inferred using the computeCommunProb and computeCommunProbPathway functions, respectively.

Survival analysis

In addition to the TCGA data, we collected three bulk datasets of breast cancer from the GEO, namely GSE3143, GSE7390, GSE20685. We downloaded the gene expression profiles and survival data of patients from these datasets for further analysis. Univariate Cox regression analysis was employed to assess the prognostic impact of the DEGs in the metastatic epithelial cell subtypes. Based on the prognosis-related DEGs, we calculated the enrichment score for each bulk sample using single sample gene set enrichment analysis (ssGSEA). Subsequently, the samples were divided into high-score or low-score groups along the median enrichment score. The survival differences between the high-score and low-score groups were tested using the log-rank test. Survival curves were fitted using the Kaplan–Meier method in the R package "survival" and visualized using the ggsurvplot function in the R package "survminer".

Correlation to mutation and clinical characteristics

To investigate the correlation between metastatic cell subtypes and mutations, we utilized the mutation profiles from both the TCGA and METABRIC datasets. The mutation profiles were downloaded from the cBioPortal database (<https://www.cbioportal.org/>). We applied a filtering step to retain only non-silent gene mutations, and further selected gene mutations that were detected in at least 5% of the samples for subsequent analysis. As

mentioned in the survival analysis, we also classified the samples into high-score or low-score groups along the median enrichment score. In addition, we collected the clinical information of the patients in the METABRIC data. Fisher's exact test was used to evaluate if there were significant differences in the frequency of gene mutations, ER status, PR status, HER2 status and histologic grade between the high-score and low-score groups.

Association of metastatic epithelial cell subtypes with immunotherapy

To explore the impact of immune-checkpoint blockade (ICB) on metastatic epithelial cell subtypes, we collected single-cell data from patients receiving anti-PD1 (n=29) and patients receiving anti-PD1 following neoadjuvant chemotherapy (n=11) [19]. This single-cell dataset includes paired pre-treatment and on-treatment information. Based on the DEGs of metastatic epithelial cell subtypes, we calculated Epi score for each epithelial cell of the ICB data using the AddModuleScore function in Seurat. Subsequently, we standardized Epi scores to a range of 0–100 and employed ternary plots to visualize the difference between different metastatic epithelial cell subtypes.

Statistical analysis

Apart from the statistical analyses described above, all other statistical analyses were performed using the R language (v 4.2.2, <https://www.r-project.org/>). The significance of the *P* value is indicated as follows: **P* < 0.05; ***P* < 0.01; ****P* < 0.001; ns: No significance.

Result

The heterogeneous cellular composition between primary tumors and pre-nodal metastases

To investigate the cellular diversity of breast cancer metastasis and non-metastasis, we integrated data from 17 primary tumor samples and 37 pre-nodal metastasis samples across six single-cell datasets (Fig. 1a). After quality control and batch effect removal, we acquired a total of 68,619 cells from primary tumors and 148,344 cells from pre-nodal metastases (Fig. 1b). Next, using canonical marker genes, we cataloged 216,963 cells into seven distinct cell types: epithelial cells, endothelial cells, fibroblasts, macrophages, T cells, B cells and plasma cells (Fig. 1c). The primary tumors and pre-nodal metastases exhibited comparable cellular proportions of epithelial cells and macrophages (Fig. 1d). However, we observed an accumulation of T cells and a depletion of fibroblasts in pre-nodal metastases compared to primary tumors (Fig. 1d). Furthermore, pre-nodal metastases showed higher inter-sample heterogeneity in cellular composition compared to primary tumors (Supplementary Fig. 1).

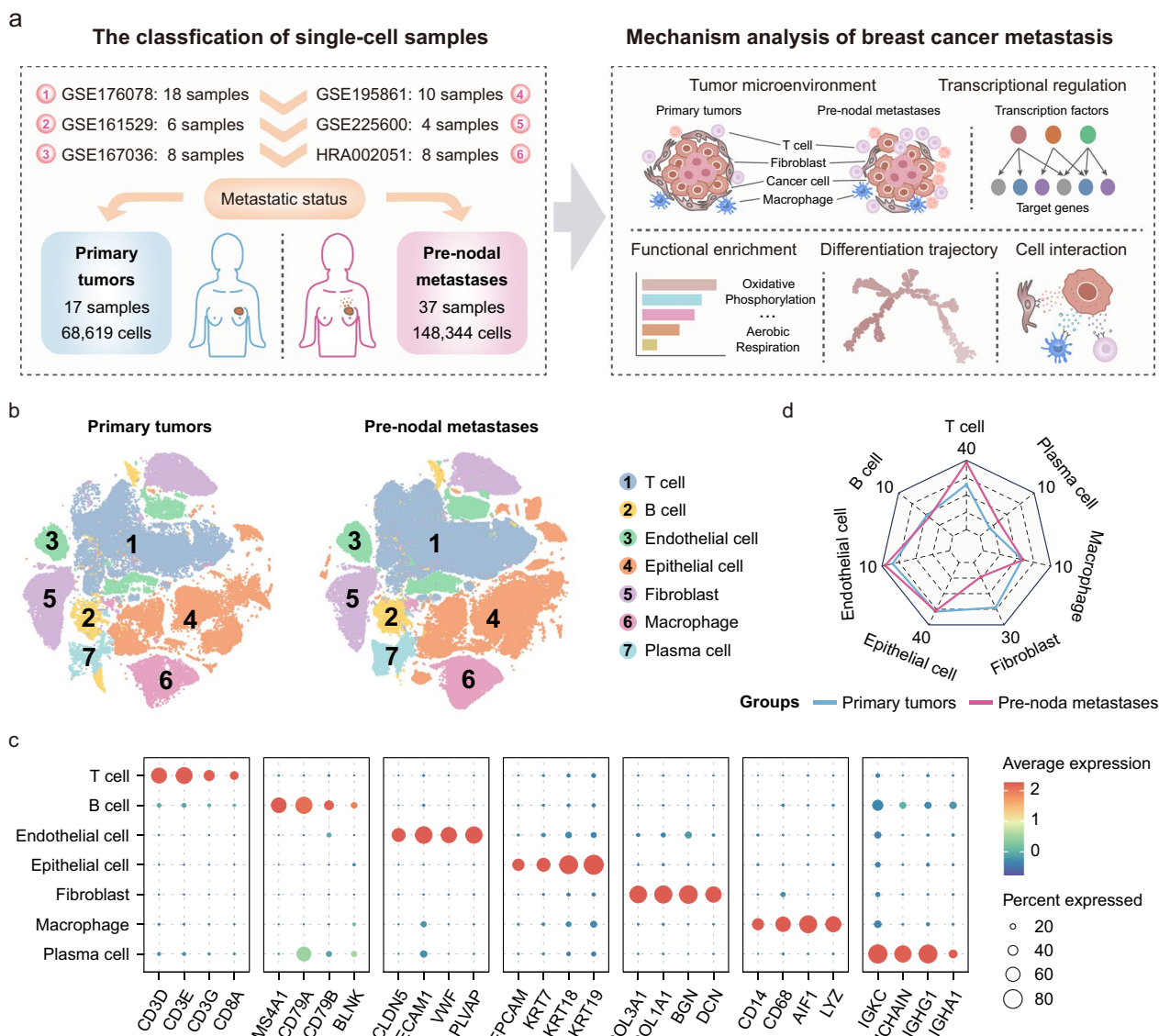


Fig. 1 Single-cell transcriptomic analysis of primary tumor and pre-nodal metastasis in breast cancer. **a** Workflow depicting the data collection, processing, and analysis in this study. **b** t-SNE plot displaying the distribution of 216,963 cells, categorized by primary tumor and pre-nodal metastasis. **c** Bubble plot showing the average expression of canonical marker genes for seven cell types. The size of each bubble represents the fraction of cell types expressing the gene. The color intensity of bubbles represents the scaled expression level of genes. **d** Radar plot showing the relative cell proportions of seven cell types in each group. The numbers surrounding the panel indicate the maximum range of cell proportions, starting from zero

Integration of phenotype-associated bulk data reveals metastatic epithelial cell subtypes

We next focused on epithelial cells, given their crucial role as the origin of breast cancer cells. We performed re-clustering of the 67,784 epithelial cells, comprising 21,641 cells from primary tumors and 46,143 cells from pre-nodal metastases (Fig. 2a). After integrating the metastatic phenotype of bulk data by Scissor, 4961 cells were identified as Scissor+ cells, and 3484 cells were

classified as Scissor- cells (Fig. 2b). Consistent with the clinical information of patients, we found that 4717 of the 4961 Scissor+ cells (95.1%) originated from patients with lymph node metastases. In contrast, Scissor- cells were evenly distributed in the primary tumors (1723, 49.5%) and pre-nodal metastases (1761, 50.5%) (Fig. 2b). Based on the results from Scissor, we classified epithelial cells into 8 subtypes (Fig. 2c). Epi1-3 subtypes accounted for the majority of the metastatic subtypes

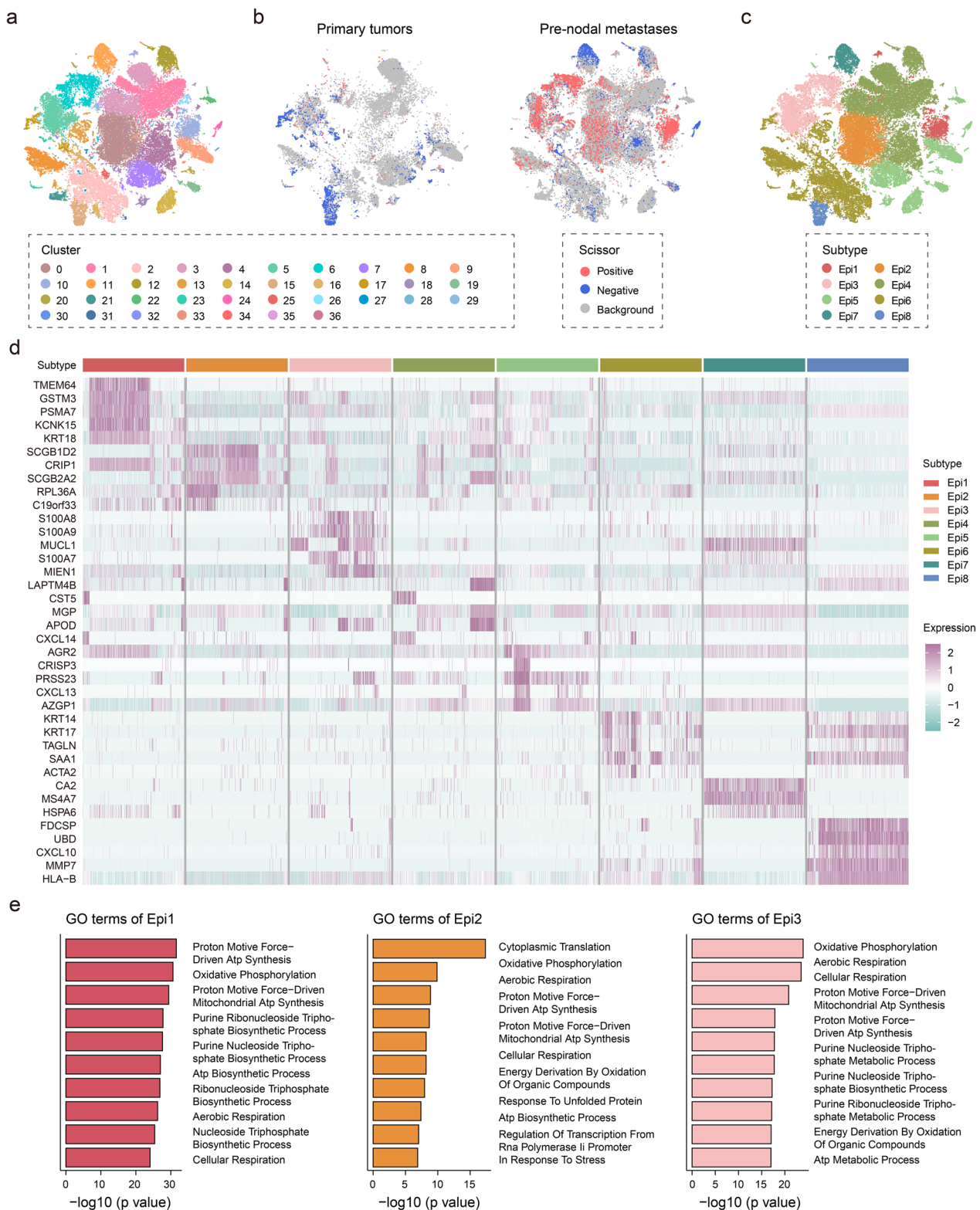


Fig. 2 Reclustering analysis of epithelial cells. **a** t-SNE plot displaying the distribution of 67,784 epithelial cells, colored by clusters. **b** t-SNE plots displaying the distribution of Scissor-selected epithelial cells, categorized by primary tumor and pre-nodal metastasis. The red and blue dots represent cells associated with the metastatic and non-metastatic phenotypes, respectively. **c** t-SNE plot displaying the distribution of 67,784 epithelial cells, colored by subtypes. **d** Heatmap showing the expression levels of representative DEGs across epithelial cell subtypes. **e** Barplots showing the GO enrichment of DEGs in Epi1-3 subtypes

and were characterized by the expression of metastatic signatures such as *PSMA7*, *CRIP1*, *S100A8*, and *S100A9* (Fig. 2d). Previous studies on breast, gastric and cervical cancers have demonstrated that high expression of *PSMA7*, *CRIP1*, *S100A8*, and *S100A9* enhances the migration and invasion capabilities of cancer cells, thereby promoting tumor metastasis [20–24]. Gene ontology (GO) analysis indicated that the up-regulated genes specific to Epi1-3 subtypes were significantly enriched in oxidative phosphorylation, proton motive force–driven ATP synthesis, and aerobic respiration (Fig. 2e). In contrast, the up-regulated genes in Epi4-8 subtypes were enriched in gland development and wound healing (Supplementary Fig. 2).

Metastatic epithelial cell subtypes exhibit heterogeneous genomic variations

The presence of transcriptomic heterogeneity among epithelial cell subtypes prompted us to investigate their genomic variations. Copy number variation (CNV) analysis revealed that metastatic Epi1-3 subtypes exhibited high levels of copy number amplifications (Fig. 3a). Among non-metastatic epithelial cell subtypes, the Epi7 subtype showed high copy number amplifications, while the Epi8 subtype displayed high copy number deletions (Fig. 3a). Notably, we found that both Epi1-3 and Epi7-8 subtypes demonstrated higher CNV scores compared to Epi4-6 subtypes (Fig. 3b). Furthermore, the clonality analysis results revealed distinct clonal evolution patterns between primary tumors and pre-nodal metastases (Fig. 3c). Based on the evolutionary trees, we found that the majority of metastatic epithelial cells exhibited copy number amplifications in chromosome 13. Additionally, we detected canonical CNVs in several genomic regions of pre-nodal metastases, including chromosomes 8, 12, and 20 (Fig. 3c), which is consistent with a previous study investigating lymph node metastases in breast cancer [25].

Metastatic epithelial cell subtypes show orchestrated cell differentiation

To understand the dynamics of epithelial cells, we constructed a cell trajectory to infer the differentiation relationship between primary tumors and pre-nodal metastases. Among the eight epithelial cell subtypes, we observed three different states (Fig. 3d). Epi1-3 subtypes were primarily distributed in state 1, while the Epi7 subtype was predominant in state 3 (Fig. 3e). Subsequently, the trajectory was labeled based on the epithelial–mesenchymal transition (EMT) score and CytoTRACE score, with state 3 being considered a potential root due to its low EMT score and high CytoTRACE score (Fig. 3f, g). Notably, metastatic Epi1-3 subtypes dominated the end of the trajectory

with high CytoTRACE score, indicating an orchestrated differentiation of epithelial cells during metastasis (Fig. 3g). Using branched expression analysis modeling (BEAM) analysis, we identified 37 branch-dependent genes that could regulate the cell differentiation process from pre-branch (state 3) to cell fate 2 (state 1) and cell fate 1 (state 2) (Fig. 3h). Importantly, we observed elevated expression of *KCNK15* and *MUCL1* along the inferred pseudotime from the pre-branch to cell fate 2 (Fig. 3i, j).

Identification of customized transcriptional regulatory programs in metastatic epithelial cell subtypes

As mentioned above, metastatic epithelial cell subtypes displayed extensive transcriptomic heterogeneity, which prompted us to examine the regulon activity of these subtypes. By calculating the regulon specificity score (RSS), we observed that *ASCL1* exhibited the highest RSS in the Epi1 subtype, while *RARB* and *NR2F1* were specifically enriched in the Epi2 and Epi3 subtypes, respectively (Fig. 4a). In addition, we found that the Epi2 and Epi3 subtypes exhibited some shared regulons, such as *ZEB1* and *NR2C1* (Fig. 4a). Based on the SCENIC results, we further constructed a transcriptional regulatory network to explore potential metastatic mechanisms of the Epi1-3 subtypes (Fig. 4b). Our results indicated that *ASCL1* was a key TF targeting the majority of DEGs specific to the Epi1 subtype (Fig. 2d). Moreover, *KCNK15*, a target gene of *ASCL1*, exhibited elevated expression along the inferred pseudotime determined by trajectory analysis (Fig. 3i). In contrast, we did not observe any DEGs regulated by TFs in the transcriptional regulatory network specific to the Epi2 subtype. In the Epi3 subtype, *NR2F1* was highlighted as a regulator of gene encoding mucin protein *MUCL1*. Notably, the target gene *MUCL1* was not only a DEG specific to the Epi3 subtype but also played an important role on the pseudotime trajectory (Fig. 3j).

Next, we sought to identify specific modules for each epithelial cell subtype, which were composed of a set of regulons. Through correlation analysis, a total 357 regulons were clustered into 8 major modules (Fig. 4c). These modules were enriched not only with subtype-specific TFs but also with conserved binding motifs, which contributed to co-ordinated activation of the modules. For example, we identified a co-regulatory module (*ASCL1*, *LHX1*, *NHLH1*, *KLF1*, etc.) specific to the Epi1 subtype, where some TFs such as *KLF1* [26] and *LHX1* [27] were associated with EMT. We also observed the activation of module 5-associated *NR2F1* in the Epi3 subtype. Furthermore, some shared modules, such as module 3 and module 8, might represent cell states that were associated with the metastasis process rather than specific cell subtypes.

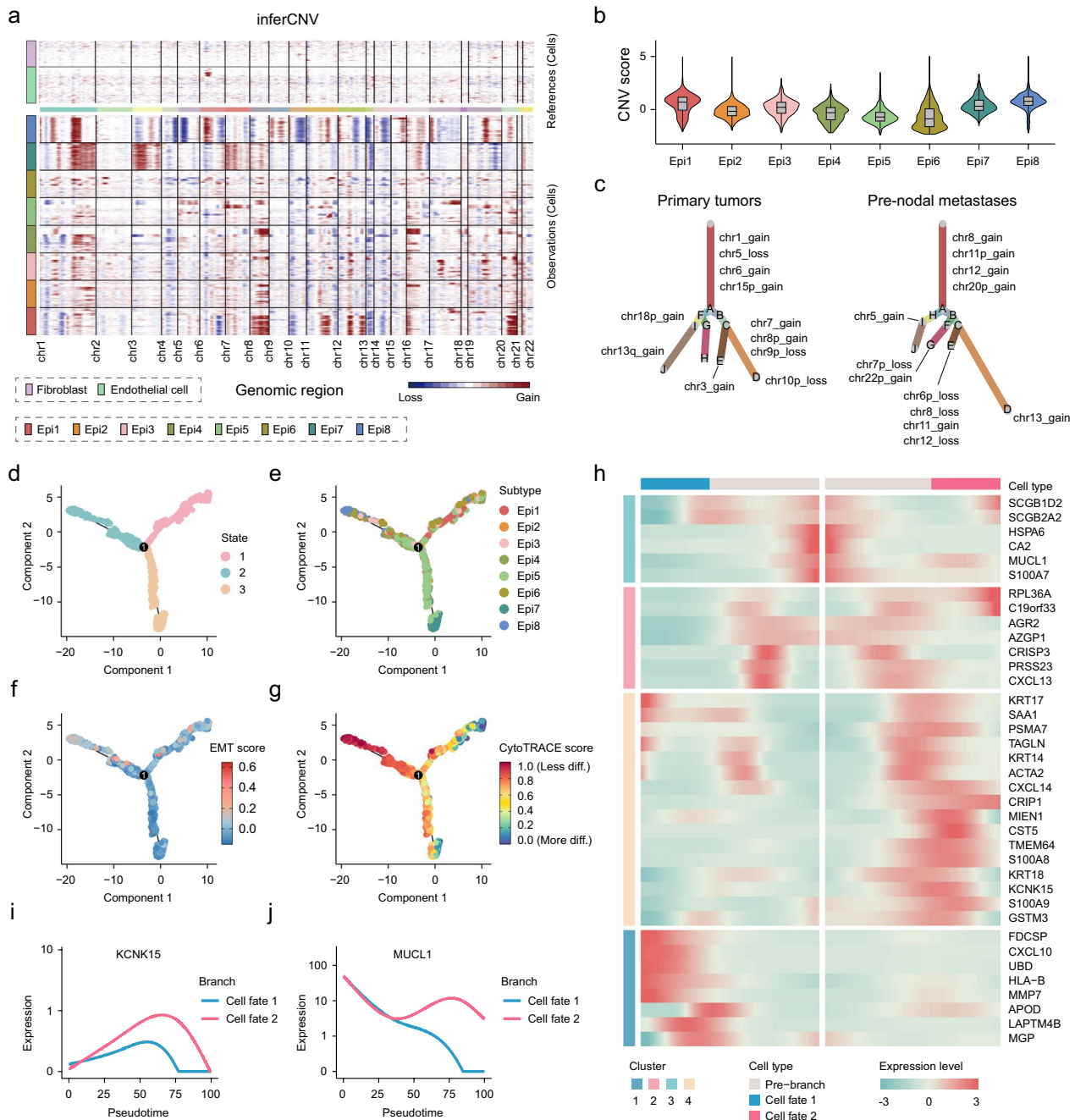


Fig. 3 Single-cell copy number variation analysis and trajectory analysis of epithelial cell subtypes. **a** The hierarchical heatmap showing the CNV profiles of epithelial cell subtypes. **b** Violin plots showing the CNV scores of epithelial cell subtypes. **c** Clonality trees showing the developmental course of epithelial cells, categorized by primary tumor and pre-nodal metastasis. The length of each branch is scaled according to percentage of cells in the subclone. **d–g** Monocle plots showing the differentiation trajectories of epithelial cells, colored by states (**d**), epithelial cell subtypes (**e**), EMT scores (**f**), and CtoTRACE scores (**g**). **h** Heatmap showing the expression patterns of 37 branch-dependent genes across 3 states, which were identified by branched expression analysis modeling (BEAM) using the top 5 DEGs of each epithelial cell subtype. **i, j** Two-dimensional plots showing the expression patterns of two representative genes in cells of cell fate 1 (blue) and cell fate 2 (pink), respectively, along the pseudotime

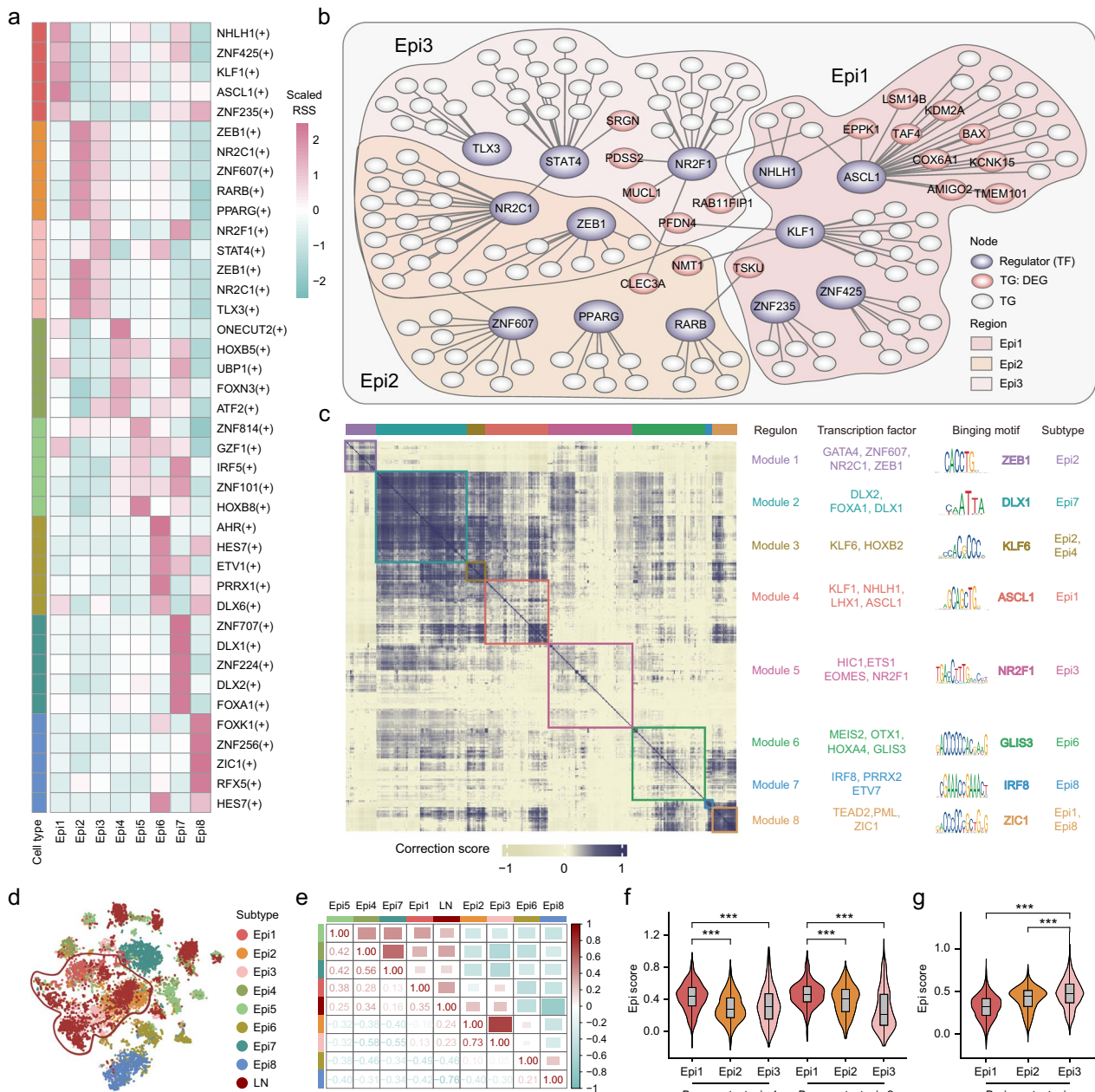


Fig. 4 Transcriptional regulatory analysis of metastatic epithelial cell subtypes. **a** Heatmap showing the top 5 enrichment of transcription factors (TFs) in each epithelial cell subtype. **b** Network showing the TFs and their predicted target genes. The red nodes represent the differentially expressed genes (DEGs) in Epi1-3. The colors of the regions correspond to the transcriptional regulatory network specific to the Epi1 (red), Epi2 (yellow), and Epi3 (pink) subtypes. **c** Heatmap showing TF modules identified by Pearson correlation based on TF activities in epithelial cell subtypes. **d** t-SNE plots displaying the distribution of epithelial cell subtypes and lymph node epithelial cells. LN, lymph node. **e** Heatmap showing the correlation of TFs between epithelial cell subtypes and lymph node epithelial cells. Pearson correlation analysis was applied to calculate the correlation coefficients. **f, g** Violin plots showing the Epi scores of Epi1-3 subtypes in two breast cancer bone metastasis samples (**f**) and one breast cancer brain metastasis sample (**g**)

Metastatic epithelial cell subtypes are correlated with different distant tissue metastases

To investigate the metastatic potential of different epithelial cell subtypes, we analyzed 25 paired metastatic lymph

node samples derived from the 37 pre-nodal metastases. By integrating with the above eight epithelial cell subtypes, we observed that epithelial cells from metastatic lymph node samples were closely clustered with the

Epi1-3 subtypes (Fig. 4d). The correlation analysis of the SCENIC results also revealed that epithelial cells from metastatic lymph node samples harbored the strongest correlations with the Epi1 subtype (Fig. 4e). The above observations prompted us to investigate whether the Epi1-3 subtypes had different effects on distant metastases. Based on the DEGs of the Epi1-3 subtypes, we found that the Epi1 subtype exhibited significantly higher scores than the Epi2-3 subtypes in two breast cancer bone metastasis samples ($P < 0.001$, Wilcoxon rank-sum test, Fig. 4f). In contrast, the Epi3 subtype showed significantly higher scores than the Epi1-2 subtypes in one breast cancer brain metastasis sample ($P < 0.001$, Wilcoxon rank-sum test, Fig. 4g).

Remodeling of the tumor microenvironment in pre-nodal metastases

Cancer metastasis is closely linked to the support of the TME. Single-cell analysis unveiled a depletion of fibroblasts and an accumulation of T cells in pre-nodal metastases (Fig. 1d). Unsupervised clustering of T cells revealed 17 clusters, including eight naïve T cell clusters, four natural killer (NK) cell clusters, two CD4⁺T cell clusters, and three CD8⁺T cell clusters (Fig. 5a, b). All T cell clusters were shared among primary tumors and pre-nodal metastases, but they exhibited heterogeneous cell proportions (Fig. 5b). Specifically, NK cells and CD8⁺T cells showed significant enrichment in pre-nodal metastases (NK cells: $P = 0.011$, CD8⁺T cells: $P = 0.013$, Wilcoxon rank-sum test, Fig. 5c). In contrast, naïve T cells were enriched in primary tumors ($P = 0.025$, Wilcoxon rank-sum test, Fig. 5c). Although there was no significant change in cell proportions, CD4⁺T cells displayed remarkable down-regulation of effector markers and up-regulation of exhaustion markers in pre-nodal metastases (Fig. 5d). In addition, NK cells and CD8⁺T cells showed notable down-regulation of cytotoxic markers in pre-nodal metastases (Fig. 5d).

Next, subclustering of fibroblasts revealed 17 clusters (Fig. 5e). Following the fibroblast classification proposed by Cords et al. [28], these clusters were classified into three cancer-associated fibroblast (CAF) subtypes: matrix CAFs (mCAFs), inflammatory CAFs (iCAFs), and vascular CAFs (vCAFs) (Fig. 5f). Notably, pre-nodal metastases exhibited a relatively higher proportion of vCAFs compared to primary tumors (Fig. 5g), implying an important role of vCAFs in breast cancer metastasis.

Although we did not observe obvious differences in the cell proportions of macrophages, endothelial cells, and B cells between primary tumors and pre-nodal metastases (Fig. 1d), subclustering of these cell types unveiled heterogeneity in their subtypes (Fig. 5g). We identified six subtypes of macrophages, with FOLR2⁺ macrophages enriched in pre-nodal metastases and SPP1⁺ macrophages enriched in primary tumors (Supplementary Fig. 3a, b). Endothelial cells consisted of four subtypes, with the most predominant subtype being venous endothelial cells, which exhibited an increase in pre-nodal metastases (Supplementary Fig. 3c, d). For B cells, we identified three major subtypes, with memory B cells and plasma cells exhibiting a decrease in pre-nodal metastases compared to primary tumors (Supplementary Fig. 3e, f).

vCAF promotes capillary formation through CXCL9⁺ macrophages in pre-nodal metastases

To decipher the crosstalk between cellular components in the TME, we constructed cellular interaction networks using potential receptor-ligand pairs in primary tumors and pre-nodal metastases, respectively (Fig. 6a, b). We observed that the crosstalk among different cellular components was much lower in pre-nodal metastases than in primary tumors. In particular, the crosstalk between fibroblasts and endothelial cells, regardless of subtypes, was predicted to be dominant in primary tumors, whereas it was notably reduced in pre-nodal metastases

(See figure on next page.)

Fig. 5 Remodeling of the pre-nodal metastasis microenvironment. **a** t-SNE plots displaying the distribution of T cell subtypes between primary tumor and pre-nodal metastasis. **b** Heatmap showing the average expression of canonical marker genes for T cell subtypes. Barplot above the heatmap indicates the relative proportions of T cell clusters in primary tumor and pre-nodal metastasis. **c** Scatterplots showing the relative proportions of four T cell subtypes between primary tumor and pre-nodal metastasis. The Wilcoxon rank-sum test was applied to determine statistical significance. Each point represents a sample. **d** Bubble plot showing the expression pattern of signature genes for four T cell subtypes between primary tumor and pre-nodal metastasis. The size of each bubble represents the fraction of T cell subtypes expressing the gene. The color intensity of bubbles represents the scaled expression level of genes. **e** t-SNE plots displaying the distribution of fibroblast subtypes between primary tumor and pre-nodal metastasis. **f** Heatmap showing the average expression of canonical marker genes for fibroblast subtypes. Barplot above the heatmap indicates the relative proportions of fibroblast clusters in primary tumor and pre-nodal metastasis. **g** Line plots showing the differences in cell proportions of specific cell subtypes between primary tumor and pre-nodal metastasis. Each point represents a specific cell subtypes. The top circular plots depict the relative cell proportions

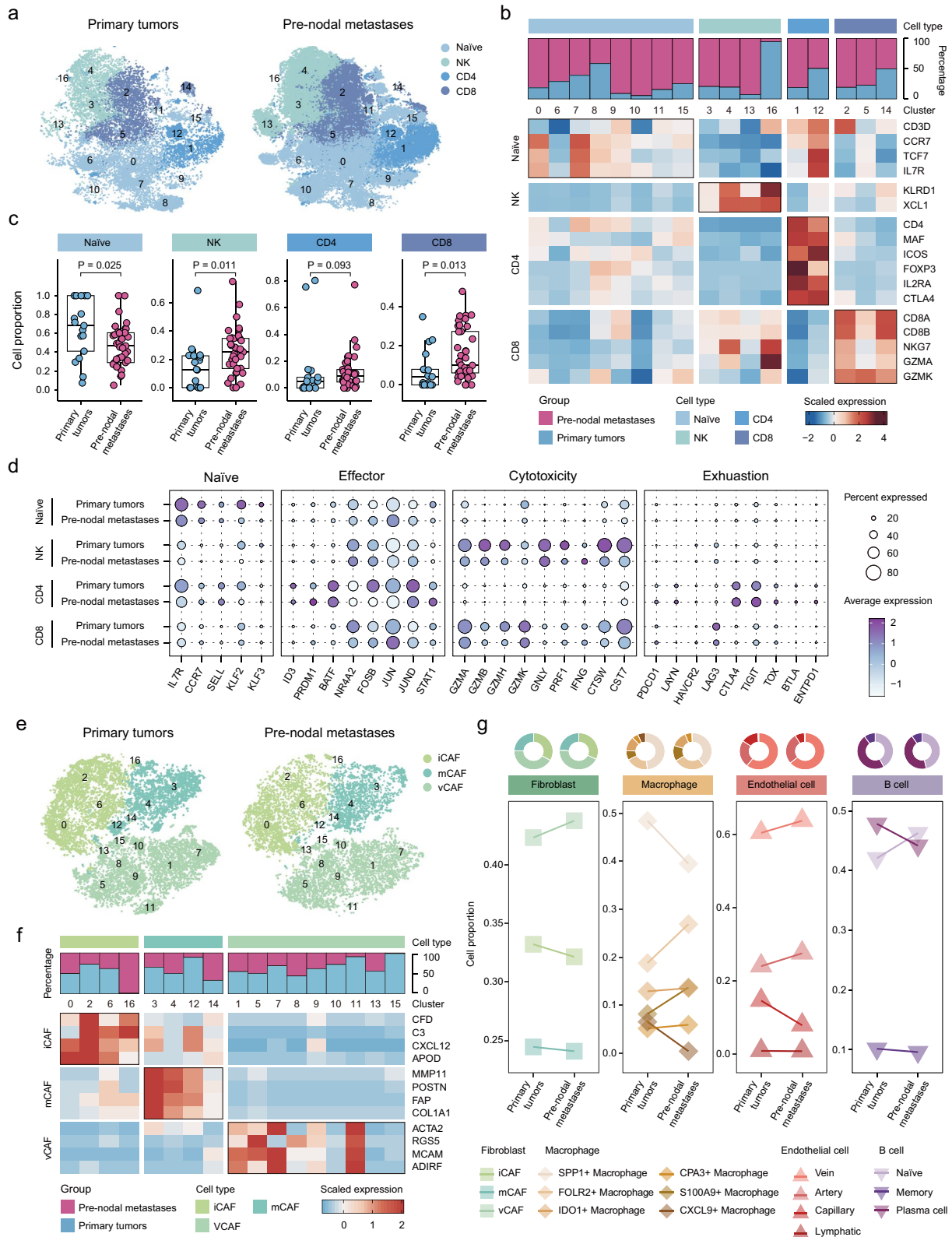


Fig. 5 (See legend on previous page.)

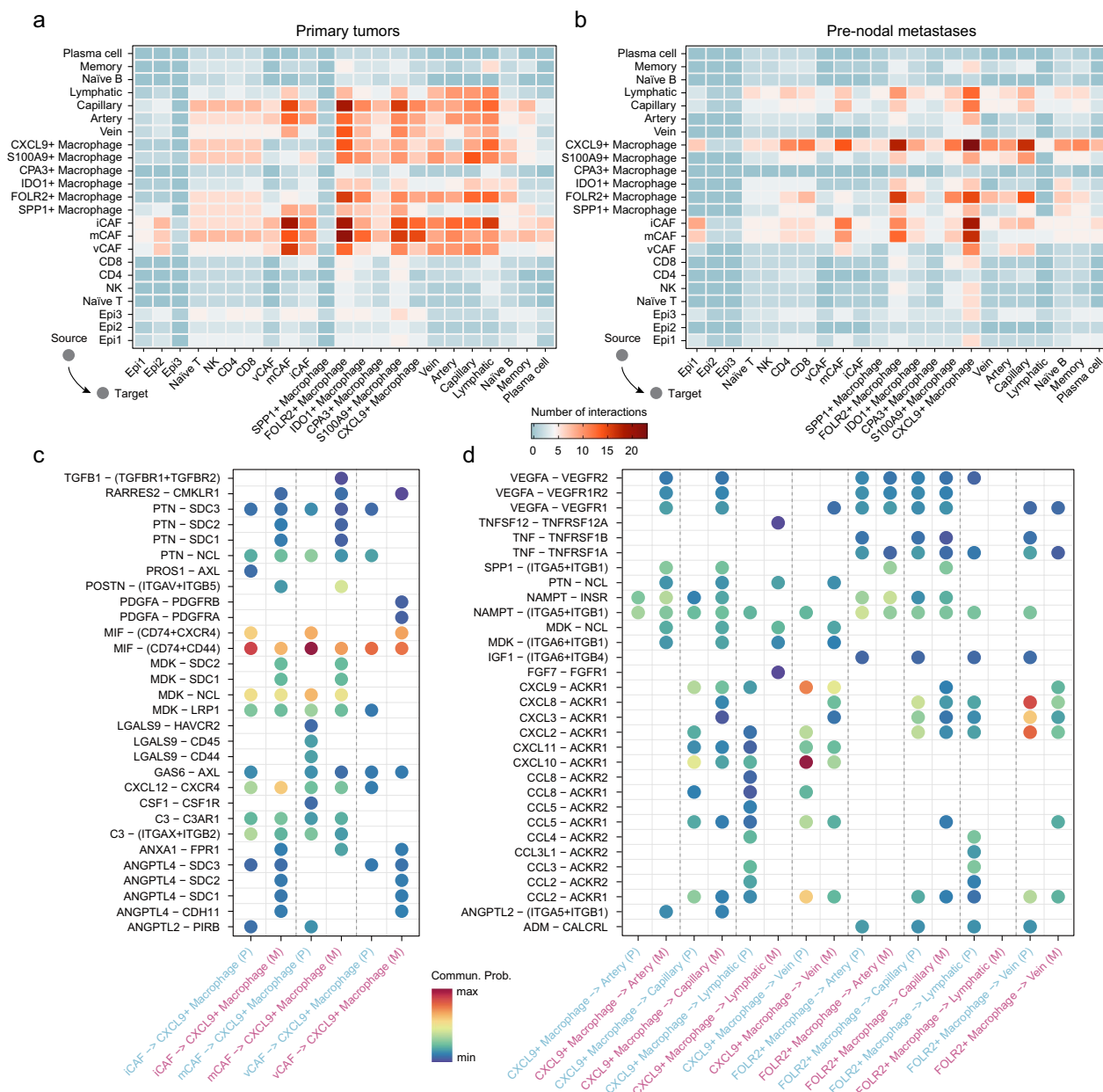


Fig. 6 Cellular interactions in the breast cancer microenvironment. **a, b** Heatmaps showing the overall interaction strength between specific cell subtypes in primary tumor (**a**) and pre-nodal metastasis (**b**). **c** Bubble plots showing the differences of specific ligand-receptor interactions from fibroblast subtypes to CXCL9⁺ macrophages between primary tumor and pre-nodal metastasis. **d** Bubble plots showing the differences of specific ligand-receptor interactions from CXCL9⁺ macrophages and FOLR2⁺ macrophages to endothelial cell subtypes between primary tumor and pre-nodal metastasis

(Fig. 6a, b, Supplementary Fig. 4). In addition, the cross-talk between immune cells and stromal cells, such as FOLR2⁺ macrophages with fibroblasts and endothelial cells, was also reduced in pre-nodal metastases compared with primary tumors (Fig. 6a, b).

Despite the overall crosstalk being decreased in pre-nodal metastases, we found that CXCL9⁺ macrophages

exhibited increased interactions with fibroblasts and endothelial cells (Fig. 6a, b). We found that CXCL9⁺ macrophages in pre-nodal metastases were more regulated by fibroblasts, regardless of their subtypes, through *RARRES2-CMKLR1* and *ANXA1-FPR1* than in primary tumors (Fig. 6c). Furthermore, CXCL9⁺ macrophages in pre-nodal metastases were also regulated by iCAFs and

mCAFs through multiple receptor-ligand pairs including *PTN-SDC1*, *MDK-SDC1*, and *ANXA1-FPR1* (Fig. 6c). Notably, *CXCL9*⁺ macrophages in pre-nodal metastases exhibited specific interactions with vCAFs through *PDGFA-PDGFRB* and *PDGFA-PDGFR2* (Fig. 6c). Subsequently, we found that *CXCL9*⁺ macrophages in pre-nodal metastases further regulated endothelial cells, especially capillary endothelial cells, through multiple receptor-ligand pairs, such as *VEGFA-VEGFR1*, *VEGFA-VEGFR2*, and *CXCL8-ACKR1* (Fig. 6d).

Metastatic epithelial cell subtypes are associated with unfavorable clinical features

Next, we explored the clinical implications of the Epi1-3 subtypes. In the TCGA dataset, the group with high enrichment scores was significantly associated with poor overall survival (OS) ($P=1.24e-05$, log-rank test, Fig. 7a). Moreover, in three independent breast cancer

datasets, the Kaplan–Meier survival curves also demonstrated impaired OS in the high-score groups (GSE20685: $P=4.51e-05$; GSE7390: $P=0.024$; GSE3143: $P=0.179$; log-rank test, Fig. 7b, c, d). Among the frequently mutated genes (>5%) in breast cancer, we observed a significantly higher frequency of *TP53* mutations in the high-score groups in both the TCGA and METABRIC datasets (TCGA: $P=3.49e-03$; METABRIC: $P=4.42e-14$; Fisher’s exact test, Fig. 7e, f). By contrast, the high-score group exhibited a significantly lower frequency of *CDH1* mutations compared with the low-score group (TCGA: $P=1.12e-08$; METABRIC: $P=6.08e-06$; Fisher’s exact test, Fig. 7e, f). In the clinical information of the METABRIC dataset, the high-score group also exhibited unfavorable clinical features, including a significantly higher frequency of PR negative status and a higher histologic grade (PR: $P=3.91e-03$; Grade: $P=1.94e-15$; Fisher’s exact test, Fig. 7g).

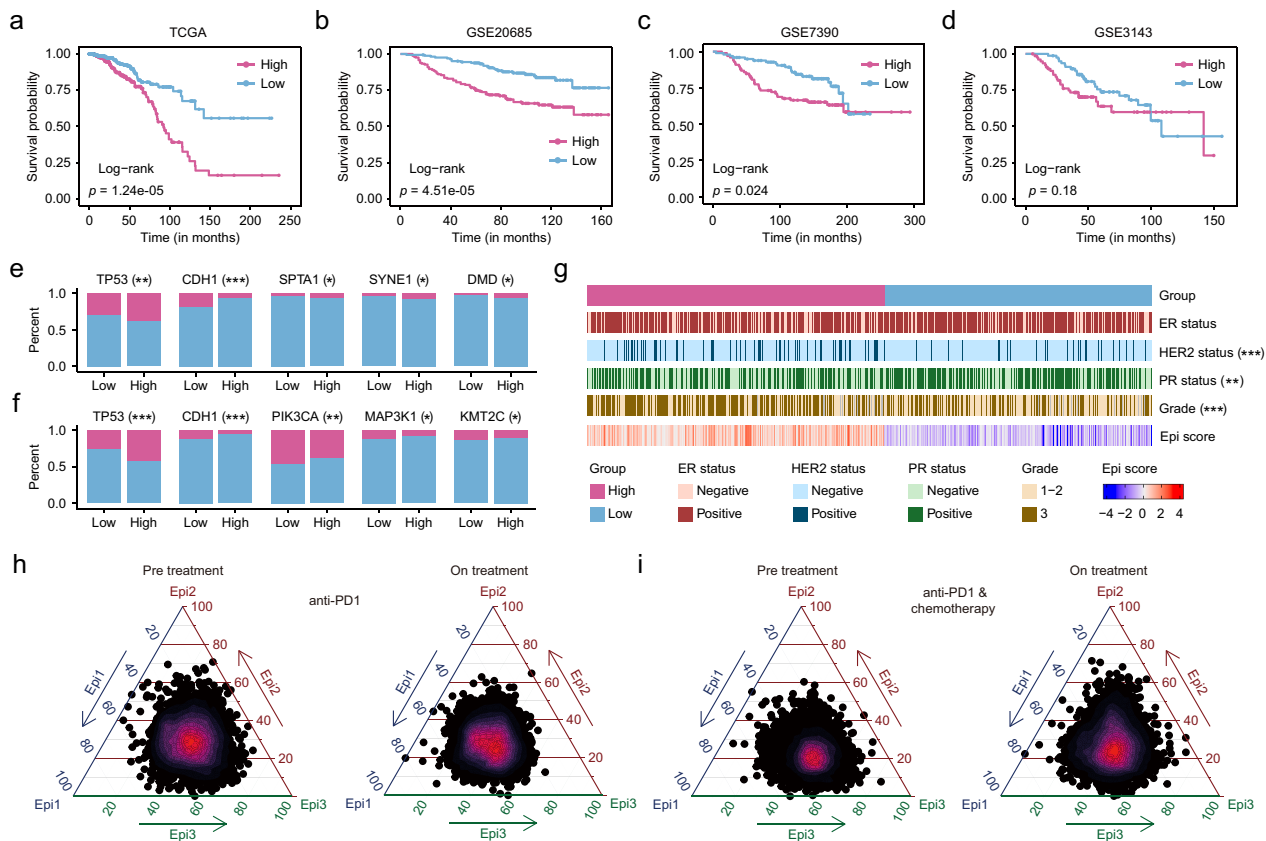


Fig. 7 The clinical implications of metastatic epithelial cell subtypes. **a–d** Kaplan–Meier survival curves showing the survival differences between the high-score and low-score groups classified by metastatic epithelial cell subtypes in TCGA (**a**), GSE20685 (**b**), GSE7390 (**c**), and GSE3143 (**d**). **e, f** Barplots showing the differences of gene mutations between the high-score and low-score groups classified by metastatic epithelial cell subtypes in TCGA (**e**) and METABRIC (**f**). **g** Barplots showing the differences of clinical characteristics between the high-score and low-score groups classified by metastatic epithelial cell subtypes. **h, i** Ternary plots showing the relative Epi scores of Epi1–3 subtypes before and after treatment for patients who only received anti-PD1 treatment (**h**) and patients who received neoadjuvant chemotherapy followed by anti-PD1 treatment (**i**)

Epi3 subtype shows a response to anti-PD1 immunotherapy

In the single-cell immunotherapy data, we extracted epithelial cells for further analysis. For patients who only received anti-PD1 treatment, we found that the Epi2 subtype had the lowest Epi scores, while the Epi1 and Epi3 subtypes displayed relatively higher Epi scores (Fig. 7h). In addition, the Epi1 and Epi2 subtypes were barely affected by immunotherapy, while the Epi3 subtype exhibited a decrease in Epi scores during immunotherapy (Fig. 7h). Among patients who received neoadjuvant chemotherapy followed by anti-PD1 treatment, we also observed that the Epi2 subtype displayed the lowest Epi scores, and the Epi3 subtype showed a decrease in Epi scores (Fig. 7i).

Discussion

In this study, we compared the cellular landscape between primary tumors and pre-nodal metastases in breast cancer. We identified three distinct metastatic epithelial cell subtypes that exhibited distinct metastatic mechanisms, including diverse genomic alterations, coordinated cellular differentiation processes, customized transcriptional regulatory networks, and specific propensities for distant metastasis. Additionally, we discovered particular subtypes of stromal and immune cells within the TME, along with their intricate interactions, which were instrumental in facilitating the pre-nodal spread of breast cancer.

Considering that not all cancer cells possess the ability to metastasize, we employed the Scissor algorithm to integrate single-cell data with phenotype-associated bulk data. We observed a scarcity of Scissor+ epithelial cells in primary tumors, indicating the accuracy of identifying metastatic epithelial cells. Furthermore, metastatic epithelial cell subtypes exhibited specific gene expression patterns, such as *KRT18* in the Epi1 subtype, *CRIP1* in the Epi2 subtype, and *MUCL1*, *S100A8*, and *S100A9* in the Epi3 subtype. The up-regulation of *KRT18* and *MUCL1* has been shown to promote the invasion of breast cancer cells [21, 29]. *CRIP1* has been found to possess the ability to promote metastasis in both gastric and cervical cancer [22, 24]. *S100A8* and *S100A9* also contribute to tumor development, growth and metastasis by disrupting tumor metabolism and the TME [20]. These findings suggest that there are distinct metastatic mechanisms among metastatic epithelial cell subtypes.

Through cell trajectory and SCENIC analysis, we elucidated the potential metastatic mechanisms for the Epi1 and Epi3 subtypes. In the Epi1 subtype, the marker gene *KCNK15* exhibited increased gene expression along the trajectory and was specifically regulated by *ASCL1*. For the Epi3 subtype, the marker gene *MUCL1* also showed increased expression along the trajectory and was specifically regulated by *NR2F1*. Although we did not identify

a specific metastatic mechanism in the Epi2 subtype, we found that the Epi2 and Epi3 subtypes exhibited shared regulons, such as *ZEB1* and *NR2C1*. Notably, the EMT-related regulon *ZEB1* is known to contribute to the metastatic spread of breast cancer cells, and its interaction with ER α may alter the tissue tropism of metastatic cancer cells towards bone [30]. Intriguingly, the marker genes of the Epi2 and Epi3 subtypes demonstrated significant down-regulation compared to the Epi1 subtype.

We have also found that the TME remodeling in pre-nodal metastases played a crucial role in the metastatic process. The increased abundance of T cells in pre-nodal metastases was primarily manifested in exhausted T cells. vCAFs exhibited an increase in pre-nodal metastases, accompanied by a relative decrease in iCAFs and mCAFs. Moreover, we observed a decrease of overall cell interactions in pre-nodal metastases, particularly in the interactions between fibroblasts and endothelial cells. However, we revealed an increased indirect interaction between fibroblasts and endothelial cells, where vCAFs primarily regulated capillary endothelial cells by modulating CXCL9⁺ Macrophages.

Our study has several limitations. First, we only determined the potential mechanisms of metastatic epithelial cell subtypes through single-cell analysis. Further experimental validation may provide a comprehensive understanding of breast cancer metastasis. Second, the analysis of tissue tropism in distant metastasis was performed on only three samples, necessitating further research in larger cohorts that include matched primary and metastatic samples. And third, the cellular interaction analysis primarily relied on transcriptomic predictions. Further confirmation of these predicted interactions is necessary through high-dimensional multiplex in situ or spatial transcriptome analysis.

In summary, we characterized three metastatic epithelial cell subtypes, elucidated their distinct metastatic characteristics, and observed the TME remodeling in pre-nodal metastases of breast cancer. Although metastatic epithelial cell subtypes were generally associated with poor prognosis, we found that the proportion of the Epi3 subtype decreased after anti-PD1 immunotherapy. Collectively, our study provided novel insights for future studies of breast cancer metastasis and may facilitate the development of combination therapies to prevent metastasis.

Conclusions

Overall, our systematic comparison of the cellular landscapes of primary tumors and pre-nodal metastases has not only shed light on the intrinsic features of metastatic epithelial cell subtypes but also elucidated the transformative effects on the TME during the early stages of breast cancer metastasis.

Abbreviations

BEAM	Branched expression analysis modeling
CAF	Cancer-associated fibroblast
CNV	Copy number variation
DEG	Differentially expressed gene
EMT	Epithelial–mesenchymal transition
GEO	Gene expression omnibus
GO	Gene ontology
iCAF	Inflammatory CAF
ICB	Immune-checkpoint blockade
MAD	Median absolute deviation
mCAF	Matrix CAF
NGDC	National genomics data center
NK	Natural killer
OS	Overall survival
RSS	Regulon specificity score
scRNA-seq	Single-cell RNA sequencing
ssGSEA	Single sample gene set enrichment analysis
TCGA	The cancer genome atlas
TF	Transcription factor
TME	Tumor microenvironment
t-SNE	T-distributed stochastic neighbor embedding
vCAF	Vascular CAF

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-024-05625-6>.

Additional file 1.

Acknowledgements

Not applicable.

Author contributions

KDL, HMH and YYG conceived the research; KDL and HMH carried out the data collection; KDL and HMH performed the data processing and analysis; KDL wrote the manuscript; HMH, HXM and YYG revised the manuscript; SMZ, XQY, KX, XMY, BC, and MYL provided valuable suggestions; All author read and approved the final manuscript.

Funding

This work was supported by grants from the National Natural Science Foundation of China (32270710; 82072985), the Outstanding Youth Foundation of Heilongjiang Province of China (YQ2021H005), the Scientific Research Project of Provincial Scientific Research Institutes of Heilongjiang Province (CZKYF2024-1-A010), and the Heilongjiang Province Innovation Base Award Project (JD2023SJ03).

Availability of data and materials

All data analyzed in this study can be downloaded from public databases. The detailed data information was provided in Supplementary Table 1.

Declarations**Ethics approval and consent to participate**

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests in this section.

Author details

¹Department of Systems Biology, College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China. ²Department of Human Anatomy, Harbin Medical University, Harbin, China. ³Department of Pathology, Harbin Medical University Cancer Hospital, Harbin, China.

Received: 4 July 2024 Accepted: 15 August 2024

Published online: 29 August 2024

References

- Kamangar F, Dores GM, Anderson WF. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. *J Clin Oncol.* 2023;41:5209–24.
- Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin.* 2022;72:7–33.
- Liang Y, Zhang H, Song X, Yang Q. Metastatic heterogeneity of breast cancer: molecular mechanism and potential therapeutic targets. *Semin Cancer Biol.* 2020;60:14–27.
- Welch DR, Hurst DR. Defining the hallmarks of metastasis. *Cancer Res.* 2019;79:3011–27.
- Liu Y, et al. Single-cell and spatial transcriptomics reveal metastasis mechanism and microenvironment remodeling of lymph node in osteosarcoma. *BMC Med.* 2024;22:200.
- Yang S, et al. Single-cell and spatial transcriptome profiling identifies the transcription factor BHLHE40 as a driver of EMT in metastatic colorectal cancer. *Cancer Res.* 2024;84:2202–17.
- Liu T, et al. Single cell profiling of primary and paired metastatic lymph node tumors in breast cancer patients. *Nat Commun.* 2022;13:6823.
- Mao X, et al. Single-cell and spatial transcriptome analyses revealed cell heterogeneity and immune environment alternations in metastatic axillary lymph nodes in breast cancer. *Cancer Immunol Immunother.* 2023;72:679–95.
- Xu K, et al. Single-cell RNA sequencing reveals cell heterogeneity and transcriptome profile of breast cancer lymph node metastasis. *Oncogenesis.* 2021;10:66.
- Ibragimova MK, Tsyganov MM, Kravtsova EA, Tsydenova IA, Litviakov NV. Organ-specificity of breast cancer metastasis. *Int J Mol Sci.* 2023;24:15625.
- 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 (Weinh).* 2023;10: e2203699.
- Korsunsky I, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods.* 2019;16:1289–96.
- Sun D, et al. Identifying phenotype-associated subpopulations by integrating bulk and single-cell sequencing data. *Nat Biotechnol.* 2022;40:527–38.
- Kurtenbach S, Cruz AM, Rodriguez DA, Durante MA, Harbour JW. Uphyloplot2: visualizing phylogenetic trees from single-cell RNA-seq data. *BMC Genomics.* 2021;22:419.
- Qiu X, et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods.* 2017;14:979–82.
- Gulati GS, et al. Single-cell transcriptional diversity is a hallmark of developmental potential. *Science.* 2020;367:405–11.
- Van de Sande B, et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat Protoc.* 2020;15:2247–76.
- Jin S, et al. Inference and analysis of cell–cell communication using Cell Chat. *Nat Commun.* 2021;12:1088.
- Bassez A, et al. A single-cell map of intratumoral changes during anti-PD1 treatment of patients with breast cancer. *Nat Med.* 2021;27:820–32.
- Chen Y, Ouyang Y, Li Z, Wang X, Ma J. S100A8 and S100A9 in Cancer. *Biochim Biophys Acta Rev Cancer.* 2023;1878: 188891.
- Li QH, et al. Small breast epithelial mucin promotes the invasion and metastasis of breast cancer cells via promoting epithelial-to-mesenchymal transition. *Oncol Rep.* 2020;44:509–18.
- Wu Z, et al. CRIP1 Reshapes the gastric cancer microenvironment to facilitate development of lymphatic metastasis. *Adv Sci (Weinh).* 2023;10: e2303246.
- Xia S, et al. Proteasome subunit alpha type 7 promotes proliferation and metastasis of gastric cancer through MAPK signaling pathway. *Dig Dis Sci.* 2022;67:880–91.
- Zhang LZ, Huang LY, Huang AL, Liu JX, Yang F. CRIP1 promotes cell migration, invasion and epithelial–mesenchymal transition of cervical cancer by activating the Wnt/beta-catenin signaling pathway. *Life Sci.* 2018;207:420–7.

25. Bao L, et al. Coexisting genomic aberrations associated with lymph node metastasis in breast cancer. *J Clin Invest*. 2018;128:2310–24.
26. Chen Y, Hong C, Qu J, Chen J, Qin Z. Knockdown of lncRNA PCAT6 suppresses the growth of non-small cell lung cancer cells by inhibiting macrophages M2 polarization via miR-326/KLF1 axis. *Bioengineered*. 2022;13:12834–46.
27. Tian Y, Wen F, Wang S, Lv N. LHX1 as a potential biomarker regulates EMT induction and cellular behaviors in uterine corpus endometrial carcinoma. *Clinics (Sao Paulo)*. 2022;77:100103.
28. Cords L, et al. Cancer-associated fibroblast classification in single-cell and spatial proteomics data. *Nat Commun*. 2023;14:4294.
29. Kang DS, et al. Ectopic expression of a truncated isoform of hair keratin 81 in breast cancer alters biophysical characteristics to promote metastatic propensity. *Adv Sci (Weinh)*. 2023;11: e2300509.
30. Mohammadi Ghahhari N, et al. Cooperative interaction between ERalpha and the EMT-inducer ZEB1 reprograms breast cancer cells for bone metastasis. *Nat Commun*. 2022;13:2104.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.