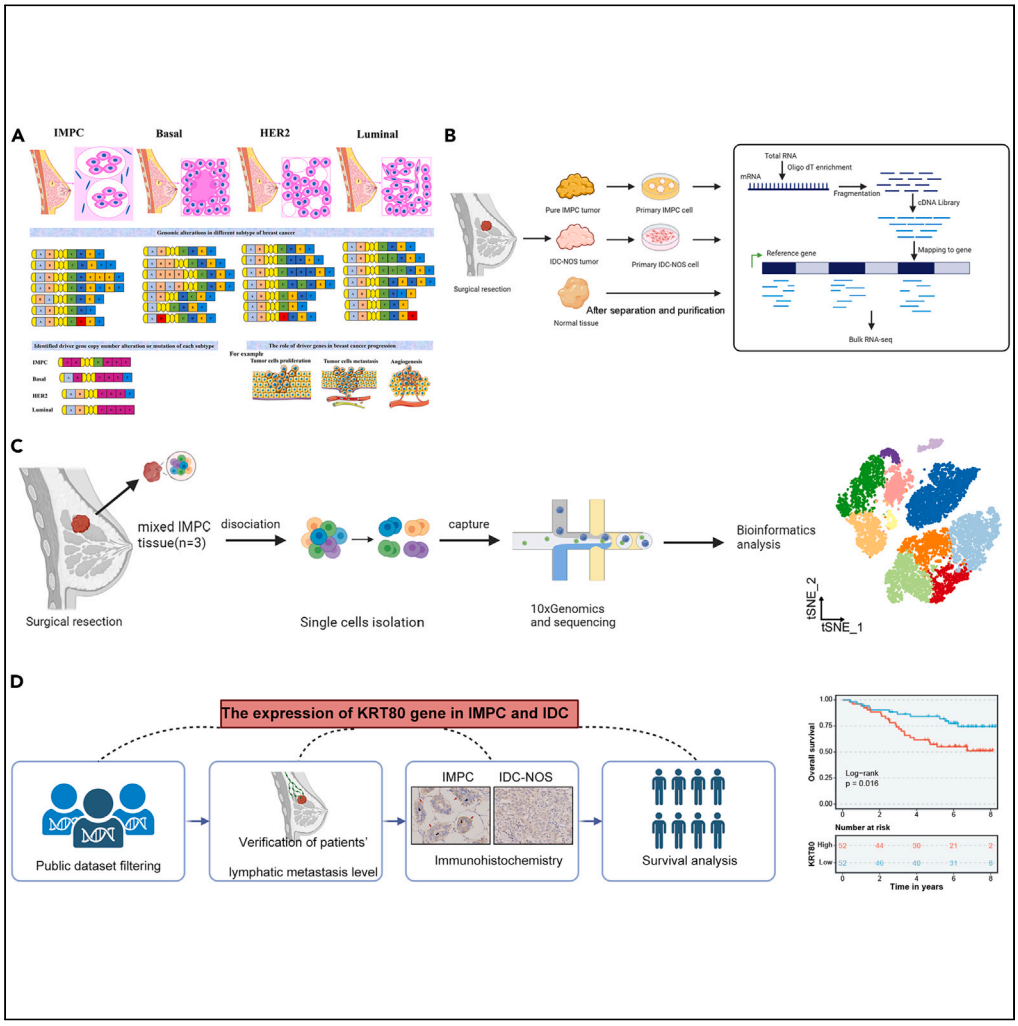
iScience

CellPress
OPEN ACCESS

Article

Deciphering the composition and key driver genes of breast invasive micropapillary carcinoma by multi-omics analysis



Yongjie Xie, Ziyun Liu, Jie Zhang, ..., Yiping Zou, Yaoyao Zhou, Xiaobin Shang

yiping-zou@foxmail.com (Y.Z.)
1457893602@qq.com (Y.Z.)
shangxiaobin@tmu.edu.cn (X.S.)

Highlights

Our study revealed the heterogeneity and the molecular characteristics of mixed IMPC

Positive correlation between KRT80 and clinicopathological characteristics in IMPC

Delineated lineage distinctions across diverse IMPC tumor cell clusters

Our findings helped to develop precision medical treatment for IMPC based on its features

Xie et al., iScience *27*, 111178
November 15, 2024
https://doi.org/10.1016/j.isci.2024.111178

iScience

CellPress
OPEN ACCESS

Article

Deciphering the composition and key driver genes of breast invasive micropapillary carcinoma by multi-omics analysis

Yongjie Xie,[1,6] Ziyun Liu,[1,6] Jie Zhang,[2,6] Guangming Li,[3] Bo Ni,[1] Chunlei Shi,[4] Yiping Zou,[1,*] Yaoyao Zhou,[2,*] and Xiaobin Shang[5,7,*]

SUMMARY

In this study, we delved into the intrinsic cellular components and transcriptomic signatures characterizing breast-invasive micropapillary carcinoma (IMPC). Employing bulk RNA sequencing, we conducted differential gene expression and functional profiles across breast cancer tissues. Single-cell transcriptome sequencing was performed on mixed IMPC samples. Moreover, a multicenter retrospective cohort of IMPC patients validated the critical role of KRT80. Our findings illuminated heightened activity in redox reactions and metabolism-related functions within IMPC compared to other tissue types. The single-cell atlas of IMPC demonstrated substantial heterogeneity predominantly driven by two distinct cell subsets: epithelioid and interstitial cells. Pseudotime analysis unveiled unique cell trajectories, and we found positive correlation between KRT80 expression and clinicopathological characteristics in IMPC. High KRT80 expression was associated with shorter overall survival for IMPC patients. This investigation unmasked extensive heterogeneity within breast IMPC tumors, delineating lineage distinctions across diverse cell clusters. It unveils potential prospective therapeutic targets with clinical relevance.

INTRODUCTION

Breast cancer is a complex and heterogeneous disease. Accurate classification of breast cancer can promote the clinical treatment effect. Breast cancer classification has been aided by the histological stratification of breast tumors, predominantly based on the expression of the ERBB2 receptor (HER2), PR, and ER receptor (ER). The luminal A shows ER+, PR+, HER2-, and low ki67. Luminal B has two types, one showing ER+, PR-/low, HER2-, and high ki67, and the other showing ER+, HER2+, PR, and ki67 expressed. HER2+ subtype: ER-, PR-, and HER2+. Triple-negative subtype: ER-, PR-, and HER2-. PAM50 is an important expression profiling-based multi-gene detection method [Reduction Analysis of Microarray 500], based on predictive analysis of 50 gene chips (15 hormone receptor-related genes, 3 HER2-related genes, 21 proliferation-related genes, and 11 basal-related genes). Through the detection of the PAM50 expression profile in breast cancer, different molecular expression matrices can be formed, and tumors can be divided into five subtypes: luminal A, luminal B, HER2 enriched, basal-like, and normal-like. The characteristics of the basal-like subtype in PAM50 are similar to triple-negative breast cancer with ER-, PR-, and HER2-. Invasive micropapillary carcinoma (IMPC) has been documented in multiple human organs,[1–6] showcasing a distinctive growth pattern, high mortality rates, and exceedingly aggressive biological characteristics. Notably, IMPC of the breast was initially reported and subsequently incorporated into the World Health Organization (WHO) classification in 2003.[7,8] Histologically, IMPC cell clusters consist of small, hollow, or morula-like formations of cancer cells enveloped by prominent stromal spaces. Our research has demonstrated that IMPC tumor cells exhibit a distinct reverse polarity, commonly referred to as an "inside-out" growth pattern,[9] whereby the apical pole of the cells faces the stroma, not the luminal surface. IMPC tumor cells show positive EMA and MUC1 in immunohistochemistry. This distinct growth pattern of the tumor can be identified as neoplastic foci within the realms of lymphovascular invasion (LVI), lymph node metastasis (LNM), and *in vitro* cultures.[10] Under electron microscopy, the surface of IMPC tumor cell clusters facing the stroma exhibit abundant microvilli. The corresponding cytoplasm comprises numerous medium and coarse motor fibers, mitochondria, and Golgi apparatus. Clinical characteristics of IMPC are

[1]Department of Pancreatic Cancer, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Center for Cancer, Tianjin 300060, China
[2]Tianjin Medical University, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin, China
[3]Department of Gastrointestinal Surgery, The Second Affiliated Hospital of Guangzhou Medical University, Guangzhou, China
[4]Department of colorectal Surgery, The Wuhu Hospital of Traditional Chinese Medicine, Anhui, China
[5]Department of Minimally Invasive Esophageal Surgery, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Center for Cancer, Tianjin, China
[6]These authors contributed equally
[7]Lead contact
*Correspondence: yiping-zou@foxmail.com (Y.Z.), 1457893602@qq.com (Y.Z.), shangxiaobin@tmu.edu.cn (X.S.)
https://doi.org/10.1016/j.isci.2024.111178

associated with elevated rates of recurrence, lymphovascular invasion (LVI), and lymph node metastasis (LNM). Building upon these observations, Fu et al. postulated and subsequently validated the theory that IMPC propagates and spreads through distinct papillary cell clusters.[9,11–20]

Bulk RNA sequencing (RNA-seq) can identify oncogenic drivers and has made a breakthrough in the treatment of various malignant tumors such as melanoma.[21,22] Single-cell RNA-seq, a recently developed powerful tool, can discover gene characteristics, explore tumor heterogeneity and subpopulations, and detect the reconstruction of evolutionary lineages.[23–25]

Hence, we integrated RNA sequencing and single-cell transcriptome sequencing to unveil the evident intratumoral heterogeneity of IMPC. Through a comparative analysis of the differentiation trajectories within the two IMPC cell subgroups, pivotal genes at the differentiation node such as KRT80 were identified. This led to a comprehensive description of the molecular attributes characterizing the various IMPC cell subgroups. These results indicate that IMPC tumor cells infiltrate and disseminate through distinctive papillary cell clusters by modulating diverse subpopulations of IMPC cells.

RESULTS

Genomic variation and copy-number profile analysis reveals distinct molecular mechanisms in different breast cancer subtypes

Based on the characterization of the expression of different molecules and pathological features, we can classify breast cancer into four categories: IMPC, basal, HER2, and luminal (Figure 1A). To explore the different molecular mechanisms involved in the progression of different breast cancer subtypes, we focused on genomic variation, comparing driver copy-number variation (CNV) across subtypes (Figure 1B). After removing the sex chromosomes, we present a panoramic view of autosomal copy-number variation in breast cancer. More consistently, all four subtypes showed large segments of copy-number amplification on chromosome 1, chromosome 8, and chromosome 20. However, the difference was that IMPC showed more copy-number amplification and fewer copy-number deletions; the basal subtype described large segments of copy-number deletions on most chromosomes; the HER2 subtype with most amplification on chromosome 17; and the luminal subtype, although it accounted for a large proportion of the breast cancer samples, many samples did not show copy-number variants (Figure 1B). According to the GISTIC2 algorithm software, we identified significant copy-number variant sites. Apparently, IMPC exhibited significant amplification and deletion of copy number at more loci than other subtypes. (Figure 1C). By calculating the percentage of different genes experiencing different copy-number events in each subtype, we identified genes with high amplification or deletion events at IMPC. There were 62 genes that underwent amplification in all IMPC samples, of which 51 genes were amplified in at least three-quarters of IMPC samples. We can see that these amplificated oncogenes were clearly associated with multiple cancer signaling pathways, and 62 amplificated genes were primarily associated with maintaining active pathways of cellular function. (Figures 1D and S1A–S1C). Compared to amplification, the percentage of predominantly occurring deletions in individual genes was not significant between IMPC and other subtypes (Figure 1E). From the aforementioned analysis, it was clear that IMPC was more prone to copy-number variation than other subtypes, especially the copy-number amplification.

Bulk RNA-seq transcriptome analysis and characteristics of primary IMPC cells and IDC-NOS in primary culture

Copy-number amplification often leads to an increase in gene mRNA levels. Therefore, we conducted transcriptome-level comparisons between IMPC and IDC to explore the downstream mechanisms. To enrich tumor cells, we cultured primary cells from 4 patients with pure IMPC and 5 patients with IDC-NOS and collected their corresponding normal tissue for bulk RNA-seq (Figure 2A; Table S1). IMPC primary cells were clustered together to form several cell groups, and the cell clusters grew almost without adherence (Figures S2A–S2C). IDC-NOS primary cells grew in a single layer adherently, and the cell morphology was similar to that of MDA-MB-231 or MCF7 cells (Figures S2D–S2F). Then, we planted primary cultured IDC-NOS and IMPC cells in collagen or Matrigel to culture in three dimensions for immunofluorescence analysis. All tumor samples were subjected to immunohistochemistry to assess the levels of MUC1/EMA and SLEX. Primary cells in three dimensions were stained with immunofluorescent antibodies. We found significant differences between IMPC and IDC-NOS at both the tissue and cellular levels. The IMPC samples had typical histological morphology and a unique cell surface protein expression pattern with positive MUC1/EMA (Figures S2 and S3).

After isolating and purifying primary IMPC tumor clusters, we performed bulk RNA-seq to identify transcriptomic differences between IMPC and IDC-NOS cells (Figures 2A and 2B). We analyzed bulk RNA data using principal-component analysis (PCA). We found obvious separation in IMPC cells, IDC-NOS cells, and corresponding normal cells for each population in the PCA plot (Figure 2C). We identified a total of 1,882 differentially expressed genes: 977 genes were upregulated in IMPC primary cells, and 905 genes were upregulated in IDC-NOS primary cells (Figure 2D). The Gene Ontology (GO) term enrichment analyses showed that genes upregulated in IMPC primary cells were mainly enriched for metabolism-related functions, such as hormone regulation, organic hydroxy metabolism, steroid metabolism, alpha-linolenic acid metabolism (Figure 2E), and a stronger external motor protein assembly (ciliary formation). Genes upregulated in IDC-NOS primary cells were mainly enriched for cell-cell interaction-related functions, such as extracellular matrix organization, adherens junctions, and cell junction organization (Figure 2F). The Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses showed that genes upregulated in IMPC primary cells were also mainly enriched for metabolism-related pathways, such as insulin secretion, fatty acid metabolism, arginine and proline metabolism, and histidine metabolism. Genes upregulated in IDC-NOS primary cells were mainly enriched for typical cancer-related pathways and cell-cell interaction-related pathways, such as focal adhesion, ECM-receptor interaction and cytokine-cytokine receptor interaction (Figures 2G and 2H). The gene set enrichment analysis (GSEA) of the two subtypes with the KEGG database showed that IMPC

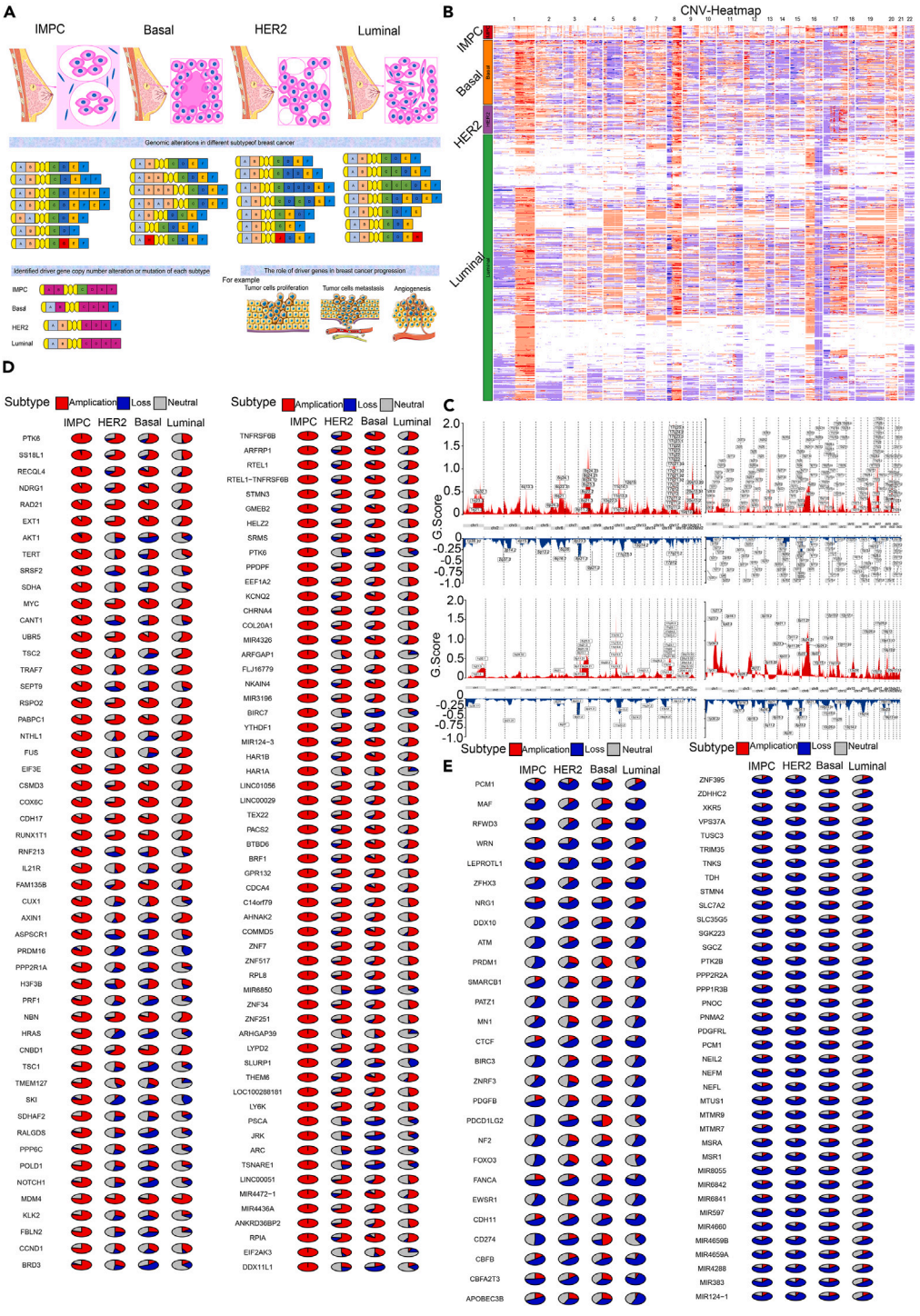


Figure 1. The CNV landscape of different breast cancer subtypes

(A) The flowchart depicting pathological characteristics and CNV analysis process of breast cancer.

(B) The heatmap shows amplification and deletion events of IMPC and other subtypes.

(C) Density plot denoting CNV amplitude in specific chromosomal loci. The top left panel belongs to HER2 subtype, the top right is IMPC subtype, the bottom left is luminal subtype and the bottom right is basal subtype.

(D and E) The pie chart illustrates CNV events percentage of the most amplificated oncogenes and other genes in (D) and of the most deleted oncogenes and other genes in (E).

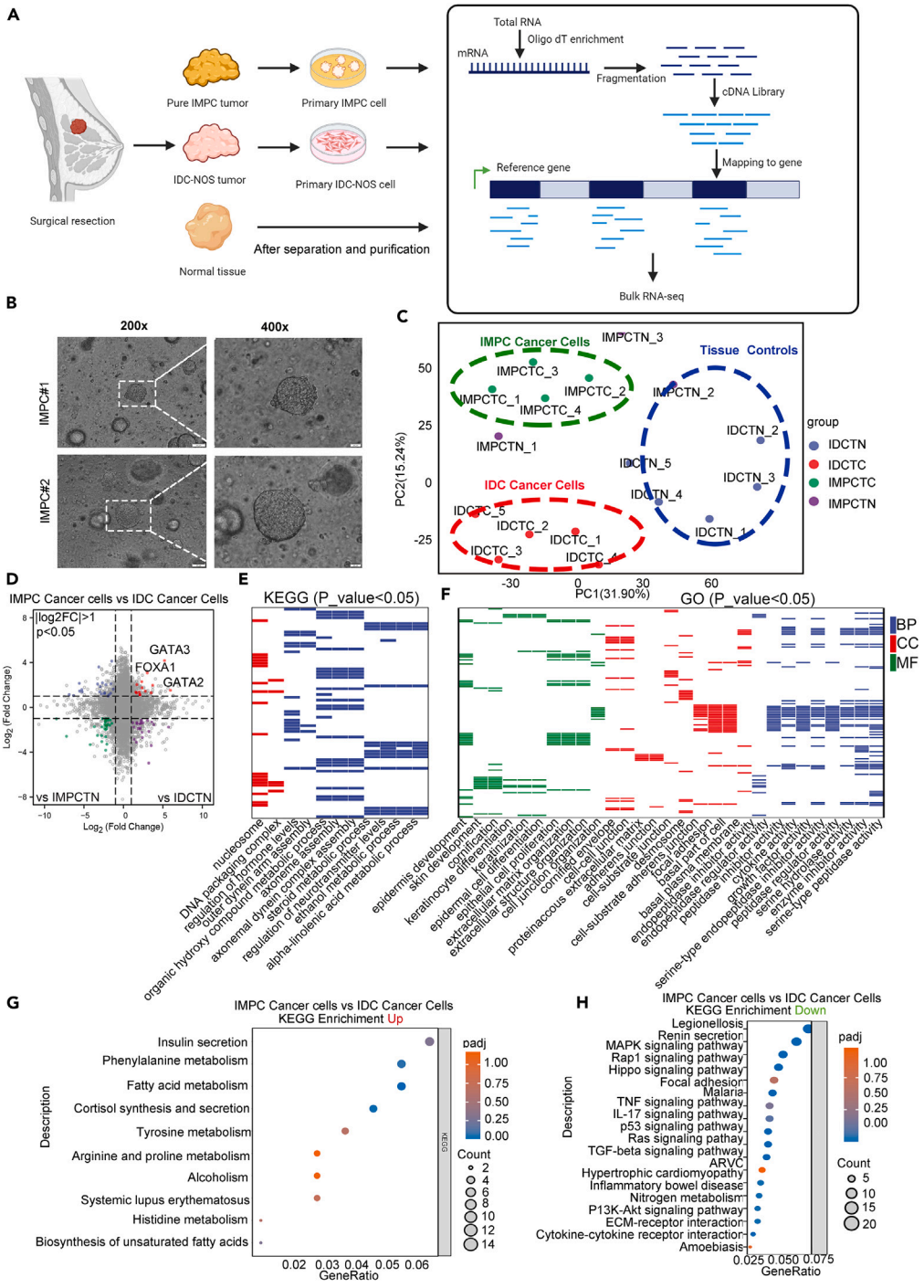


Figure 2. Bulk transcriptome differences between IMPC and IDC subtypes

(A) The experimental flowchart describes the process of breast cancer tissue sampling and analysis.

(B) Optical microscope observing IMPC tumor cells morphological changes.

(C) PCA plot of tumor and normal samples from IMPC and IDC.

(D) Volcano plot tells different expression genes in IMPC and IDC.

(E) GO pathways heatmap of upregulated genes of IMPC.

(F) GO pathways heatmap of upregulated genes of IDC.

(G and H) Bubble plot shows KEGG pathways enriched by IMPC upregulated genes in G and IMPC downregulated genes in (H).

primary cells were mainly enriched for metabolism-related pathways (especially stearic acid, niacin, and nicotinamide metabolism) and pathways related to vitamin transporters, vitamin digestion and absorption, and lysosomes. This suggests that IMPC tumor primary cell clusters lack a central vascular bundle but have a stronger demand for nutrients from the extracellular matrix, leading to a more vigorous metabolism. The IDC-NOS tumor primary cells were enriched for typical cancer-related pathways, such as the cell cycle and cell proliferation (Figure S4).

Meanwhile, we collected 73 IMPC samples and 51 invasive carcinoma no special type (ICNST) samples from GSE66418 in the GEO database for comparison. Breast cancer datasets in TCGA were integrated for subtype-specific comparisons. Consistent with our analysis of the self-assessment data, IMPC tumor samples were associated with fat metabolism and bile acid metabolism, while the other subtypes were enriched in cell proliferation-related and ECM-receptor interaction pathways (Figures S5A–S5E).

Single-cell transcriptomic profile and cell subtypes of mixed IMPC

Pure IMPC is relatively rare in clinical practice, and IMPC often appears in the mixed form. Our previous studies have proven that regardless of the proportion of IMPC in mixed IMPC, it indicates more lymph node metastasis and a worse prognosis. Therefore, single-cell RNA-seq was performed on 3 patients with mixed IMPC (Figure 3A; Table S2), whose scRNA-seq data matrices were merged and subjected to batch effect removal (CCA and IntegrateData packages). After data preprocessing, we generated single-cell RNA-seq profiles for a total of 14,057 cells. To explore the cellular diversity in mixed IMPC, we applied the uniform manifold approximation and projection (UMAP) algorithm for variably expressed genes across all cells and identified 11 main clusters (Figure 3B). Then, we identified cluster-specific marker genes by performing differential gene expression analysis to define each cluster. Well-known cell type markers, such as *KRT 18*, *KRT 8*, *KRT 7*, *MUC 1*, and *EPCAM*[26] (Figure 3C), were identified as markers of epithelial cells (clusters 0, 1, 2, 3, 4, 5, 6, 7, 9, and 10), while *CD74*, *HLA-DRA*, *CD68*, and *CCL5* were identified as marked for the immune cell cluster (cluster 8).[27] At the same time, we calculated the large-scale chromosomal CNV in a single cell based on the averaged expression profile across intervals of the chromosome[28] (Figure S6). The CNV landscape distinguished malignant epithelial cells (cluster 0–7) from all cells in mixed IMPC, which underwent copy-number amplification or loss of large segments. Based on the differentially expressed genes of each cluster and specific marker genes of known cell types, we identified the following cell clusters: The cells in the cluster with high expression of the basal cell marker genes S100A6, S100A16, and SH3BGRL3 were identified as basal cells (cluster 0, 20.33%). The cells in the cluster with high expression of the FOS, BTG2, CYR61, CTGF, and extracellular matrix COL1A1 genes were identified as villous stromal cells (cluster 1, 19.94%). The cells in cluster 2, which expressed the ciliated cell marker genes ATF3, ARL14, and CTGF, were identified as ciliated cells (cluster 2, 14.90%) (Figures 3B–3D). We established a gene expression database for IMPC and IDC-NOS based on bulk RNA-seq and further performed differential gene correlation analysis with bulk RNA-seq and single-cell RNA-seq data (Figure 3E). We found that the gene expression of clusters 3 and 6 was highly consistent with the gene expression of IMPC cells according to bulk RNA-seq, and the gene expression of cluster 4 was also highly consistent with the gene expression of IDC cells according to bulk RNA-seq. Therefore, we identified IMPC-type cells (clusters 3 and 6, 16.77%) and IDC-NOS type cells (cluster 4, 13.44%) from the epithelial cell cluster (Figure 3F). We further compared cluster 3 and cluster 6 and found that cluster 3 overexpressed duct epithelial markers, such as *KRT8*, *KRT18*, *KRT19*, and *MUC1*, and the stem cell marker *CD24*.[29] Cluster 6 overexpressed the stromal cell marker genes *IGFBP5*, *SEPP1*, *MGP*, and *TFF3*. *SEPP1* and *MGP* are involved in focal adhesion and calcium ion binding, and *TFF3* is related to angiogenesis and lymphangiogenesis. Therefore, we identified cluster 3 as epithelial-like IMPC and cluster 6 as interstitial-like IMPC. This result reveals the heterogeneity within IMPC tumors. Cluster 5 highly expressed stromal genes, such as *MYL9*, *IGFBP2*, and *TAGLN*, and the correlation analysis with bulk RNA-seq showed that cluster 5 had a strong correlation with IDC-NOS, so we identified them as stromal-like IDC cells (cluster 5, 13.33%). Cluster 7 highly expressed mesenchymal genes, such as *WFDC2*, *FDCSP*, *CD24*, and *ALDH1A3*, which were strongly correlated with IDC-NOS genes by bulk RNA-seq, so they were defined as mesenchymal-like IDC cells (cluster 7, 0.76%) (Figures 3D–3F). Gene set variation analysis (GSVA) showed that the differentially expressed genes in cluster 3 were mainly enriched in peroxisome activity, fatty acid metabolism, glycolysis, xenobiotic metabolism, oxidative phosphorylation, reactive oxygen species pathways, etc. Cluster 6 was mainly enriched in fatty acid metabolism, late and early estrogen response, xenobiotic metabolism, estrogen response peroxisome, and oxidative phosphorylation (Figure 3G). GSVA of the differentially expressed genes of cluster 4 showed enrichment of the androgen response, transforming growth factor beta (TGF-beta) pathway, P53 pathway, and tumor necrosis factor alpha (TNF-alpha) pathway (Figure 3G).

Cell trajectory analysis reveals the evolutionary relationship between IDC-NOS and IMPC

The cellular mechanisms of tumor initiation in IMPC are unclear, so we performed pseudotime trajectory analysis on IMPC cells (clusters 3 and 6) and IDC-NOS cells using Monocle 3. The pseudotime trajectories of the two clusters of cells were modeled according to their gene expression profiles. We found that cluster 4 cells could not only transform into cluster 5 cells (stromal-like IDC cells) but also convert to cluster 3 and cluster 6 cells (Figures 4A–4E). To explore the potential markers and therapeutic targets of IMPC, we performed differential gene expression analysis on the differentiation nodes. IMPC cells (clusters 3 and 6) were enriched in the IMPC-specific gene *MUC1* and its related gene *MUCL1*, which are involved in IMPC progression, and the *NDUFB4*, *DHCR24*, *NQO1*, and *SEPP1* genes, which are related to oxidoreductase. The expression level of these genes was decreased in the IDC-NOS cells (cluster 4). In contrast, the expression levels of the key transcription factors GATA3, NUPR1, and SPDEF exhibited a marked minimize during the transition in transcriptional states from cluster 4 to cluster 3 and cluster 6 (Figure 4F). We also performed differential gene expression analysis on clusters 3 and 6 The differentially expressed genes at the transition nodes from cluster 3 to cluster 6 were *CLDN4*, *KRT80*, *IGFBP5*, *ENO1*, *LDHA*, *PDLIM7*, and *PIP*. We observed that the expression levels of *CLDN4*, which is related to the epithelium, decreased significantly in the transcription state of cluster 6, while the expression level of KRT80 increased significantly (Figure 4F). The highly expressed genes of cluster 6 were related to insulin and insulin-like growth factor

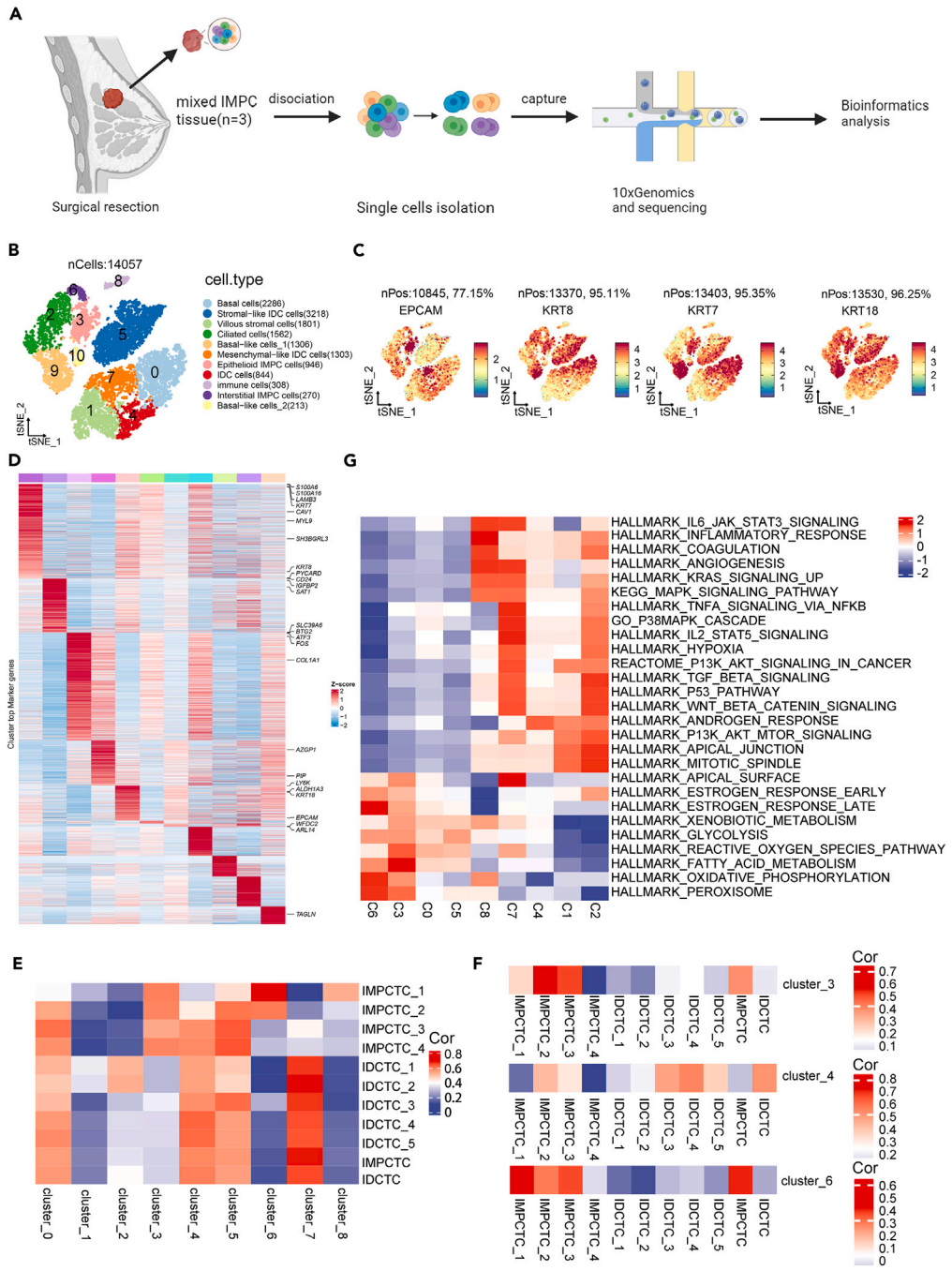


Figure 3. Single-cell transcriptome analysis of mixed IMPC

(A) Single-cell sampling analysis flowchart.

(B) TSNE plot dividing different tumor parenchymal cells, immune cells, and stromal cells.

(C) Marker genes expression map of single cell subpopulation.

(D) Heatmap depicting overexpressed genes in specific cell subpopulations.

(E) Correlation heatmap of bulk different expression genes and epithelial subpopulation genes.

(F) Three represented clusters from (E).

(G) GSVA enrichment heatmap of epithelial cells cluster. The pathways were collected from the Msigdb database.

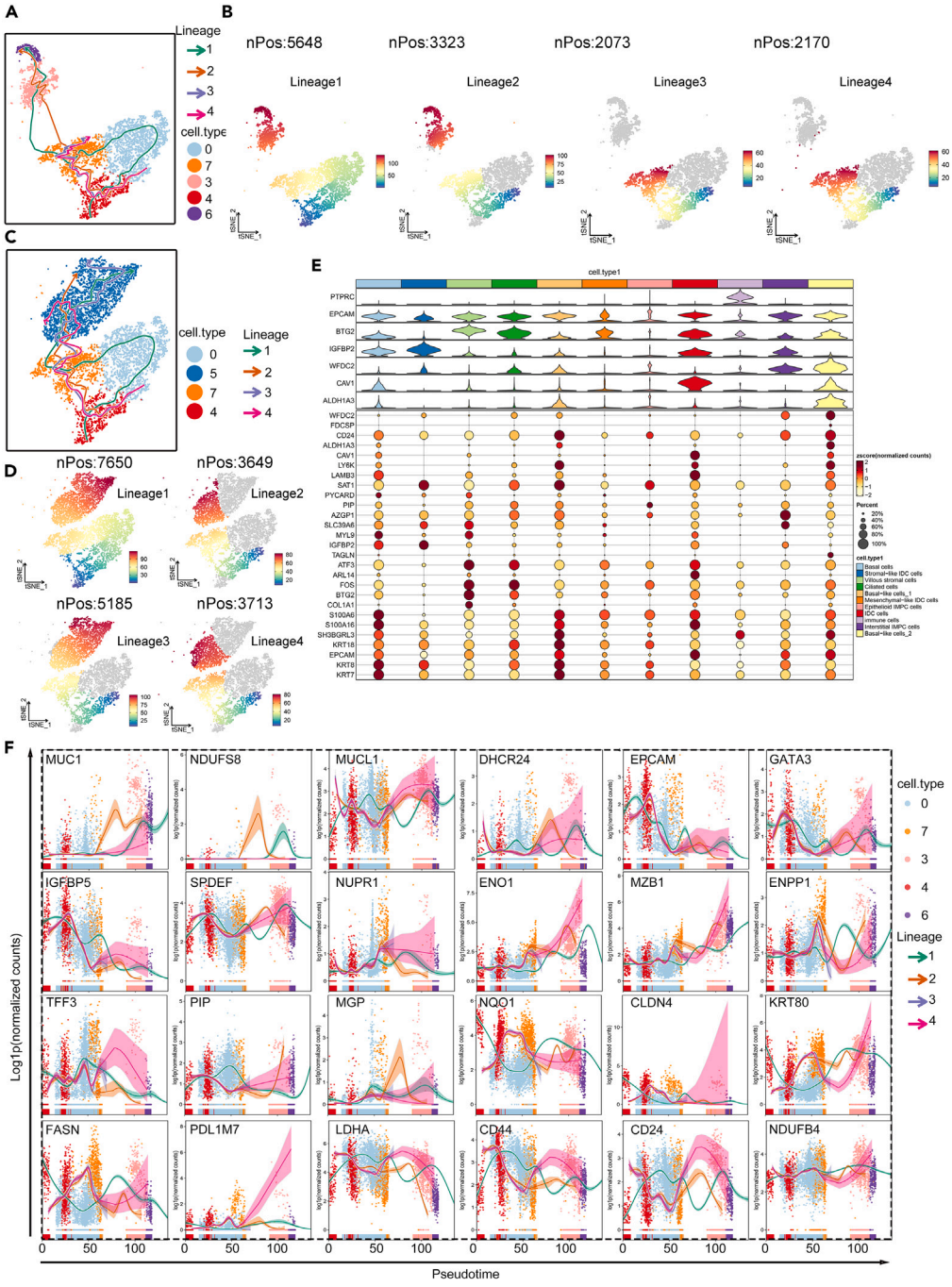


Figure 4. Pseudo-time trajectory of mixed IMPC epithelial cell clusters

(A) Four lineages predicted by the Slingshot method of distinct IMPC clusters and IDC clusters.

(B) TSNE plot of pseudo-time value in different clusters, the redder the color, the more it represents a late stage of evolution.

(C) Trajectories of IDC-dominated epithelial cell subpopulations.

(D) Pseudo-time value of epithelial IDC clusters in corresponding trajectories.

(E) Complex heatmap describes the characteristic high-expression genes of each cell population. The size of the dot means the proportion of cells expressing this gene in all cells of this cluster.

(F) Combination of line and point charts displays genes that undergo changes in expression levels with the evolution trajectory of different cell types.

signaling pathways (*IGFBP5*, *GRB14*, and *ENPP1*) as well as matrix assembly, cilia, and microvilli formation (*PIP* and *MZB1*). Cluster 3 highly expressed the glycolysis-related gene *ENO1* and the cytoskeletal mitosis-related gene *PDLIM7* (Figure 4F).

Correlations between the expression of KRT80 and IMPC clinicopathological characteristics and prognosis

KEGG analysis and GSEA of bulk RNA-seq data for IMPC primary cells showed enrichment in lipid metabolism-related pathways (Figures 2G and S5). Among the internal subgroups of IMPC cells, the GSVA of interstitial-like IMPC cells (cluster 6) in the late stage of pseudotemporal differentiation mainly showed enrichment in fatty acid metabolism. *KRT80*, localizing near desmosomal plaques in earlier stages of differentiation but then dispersing throughout the cytoplasm in terminally differentiating cells, was significantly upregulated in cluster 6. Therefore, we believe that *KRT80* may be an important factor in the development of IMPC. Next, we selected 102 formalin-fixed and paraffin-embedded (FFPE) tumor tissues from IMPC patients to perform immunohistochemical staining using IDC ($n = 79$) as a control. Immunohistochemical experiments confirmed that the differential expression of KRT80 protein between IMPC and IDC was uniform, and 43% of tumor cell clusters in IMPC showed the phenomenon that the expression of KRT80 in the outer circle of the cell group was higher than that in the inner cells (Figure 5A). We analyzed the potential relationships between the expression of KRT80 in IMPC and clinicopathological characteristics, the expression of KRT80 in IMPC was significantly positively correlated with lymph node metastasis and lymphatic vessel invasion in our cohort and other datasets (Figures 5B–5I). In addition, when we analyzed the relationship between the expression of KRT80 and the prognosis in IMPC patients, we found that the overall survival and disease-free survival times of patients with high KRT80 expression were significantly shorter than those of patients with low KRT80 expression (Figures 5J and 5K). This indicates that KRT80 could be an independent risk factor affecting the prognosis of IMPC patients.

Exploring the potential mechanism and phenotypic changes of KRT80 in IMPC and IDC cells

Subsequently, to further explore the specific mechanisms and phenotypic variations of KRT80 in IMPC and IDC cells, we employed primary cell lines sorted from IMPC and IDC and verified by transfecting with KRT80 overexpression and knock down plasmids. We found that after overexpression of KRT80, stemness markers (SOX2, OCT4, Snail1, and vimentin) and key proteins in the glycolytic pathway (HIF1α, MYC, and LDHA) exhibited significant positive correlations (Figures 6A–6E). Simultaneously, we also reached the corresponding conclusion through the cell line with the knockdown of KRT80. Based on this, we infer that KRT80 can affect the tumor differentiation and motility of IMPC through glycolysis level and stemness ability, and promote its lymphatic and lymph node metastasis (Figures S7A–S7E). Subsequently, we also explored the crucial mechanism by which KRT80 affects glycolysis and stemness. Through mechanism analysis, we found that the JAK/STAT1 pathway was also activated, and subsequent administration of a STAT1 inhibitor revealed that downstream glycolysis phenotypes and stemness-related molecules were inhibited. Thus, we speculated through the research results that KRT80 can modify its phenotype via the JAK/STAT1 pathway and impact its cell differentiation and glycolysis flux (Figure 6F).

The therapeutic target of IMPC patients by analyzing the differentially expressed genes in the cell trajectory

Analysis of the METABRIC and GEO databases revealed that the differentially expressed genes CLDN4, NDUFB4, and DHCR24 were correlated with diminished overall and disease-free survival within the differentiation nodes between IMPC and IDC-NOS type. The differentially expressed genes *KRT80*, *ENO1*, *LDHA*, and *PDLIM7* in the differentiation node between epithelial-like IMPC and interstitial-like IMPC cells were significantly associated with poorer overall and disease-free survival rates ($p < 0.05$), the expression levels of the key transcription factors GATA3 and NUPR1 play a role in the transition from cluster 4 to cluster 3 and cluster 6 transcriptional status, potentially inhibiting lymph node metastasis and extending survival (Figure 7; Figure S8). Summarizing the aforementioned, *KRT80*, *CLDN4*, *ENO1*, *PDLIM7*, and *PIP* were the most important genes associated with tumor metastasis and poorer survival in patients with breast cancer and could be identified as potential therapeutic targets to block the transition from IDC-NOS cells to epithelial-like IMPC cells and interstitial-like IMPC cells.

Sensitivity of the KRT80 and other potential therapeutic targets to chemotherapeutic agents in IMPC

We predicted the correlation between therapeutic target genes and chemotherapeutic drugs in 71 IMPC samples (GSE66418) by combining the drug sensitivity data of about 1,000 cell lines from GDSC. Apparently, KRT80 was insensitive to all conventional chemotherapeutic agents, including 5-fluorouracil, tamoxifen, and vinblastine (Figure S9A). CLDN4 was also insensitive to 5-fluorouracil but was sensitive to osimertinib (Figure S9B). ENO1 was resistant to cyclophosphamide while responding to selumetinib (Figure S9C). Conventional platinum, vincristine, and cyclophosphamide drugs were ineffective against PDLIM7, but pictilisib was effective in it (Figure S9D). Like ENO1, PIP could also respond to selumetinib (Figure S9E). Therefore, in order to predict the drugs that can effectively treat IMPC patients with high KRT80 expression, we included CMap small molecule drug prediction analysis and inferred the top 10 drugs that may be effective for KRT80 (score <0). However, the top 10 drugs with positive scores were counterproductive and might promote KRT80 expression. Hence, we suggested that DNA inhibitor and cyclooxygenase inhibitors were preferred in the treatment of patients with high expression of KRT80 (Figure S9F).

DISCUSSION

IMPC is known for its high rates of metastasis and recurrence regardless of the site of origin and has a unique growth and metastasis pattern of cell cluster invasion and metastasis. The understanding of the internal structure of the IMPC tumor cell cluster is still at the stage of microscopic observation, and its intrinsic molecular characteristics and mechanism are unclear. In order to explore the heterogeneity of IMPC tumor

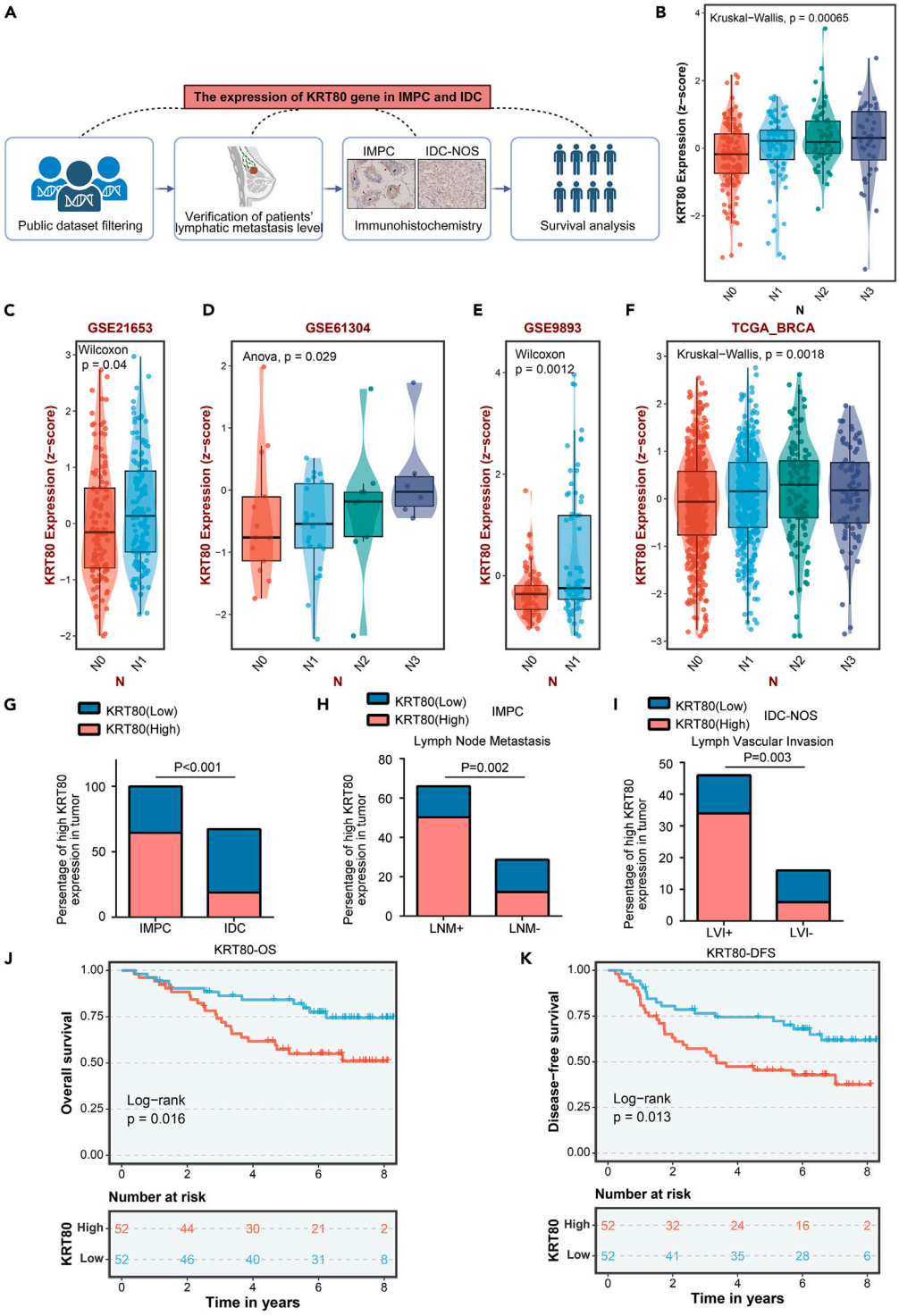


Figure 5. KRT80 expression was related to tumor metastasis and patients' survival

(A) Flowchart of collecting tumor tissues, experimental verifying, and bioinformatical analyzing.

(B–F) Boxplot displaying KRT80 expression of different lymph node stages in our own tumor tissues (B), GSE12653 (C), GSE61304 (D), GSE9893 (E), and TCGA (F). Data are represented as median +/− 1.5 interquartile range. N0 represents no lymph node metastasis.

(G–I) Stacked bar chart of high and low KRT80 samples relative to total IMPC or IDC samples in (G), total positive lymph node metastasis or negative lymph node metastasis samples in (H), total positive lymph vascular invasion or negative lymph vascular invasion samples in (I).

(J and K) KM curves plotting the relationship of KRT80 expression with patients' overall survival in (J) and disease-free survival in (K).

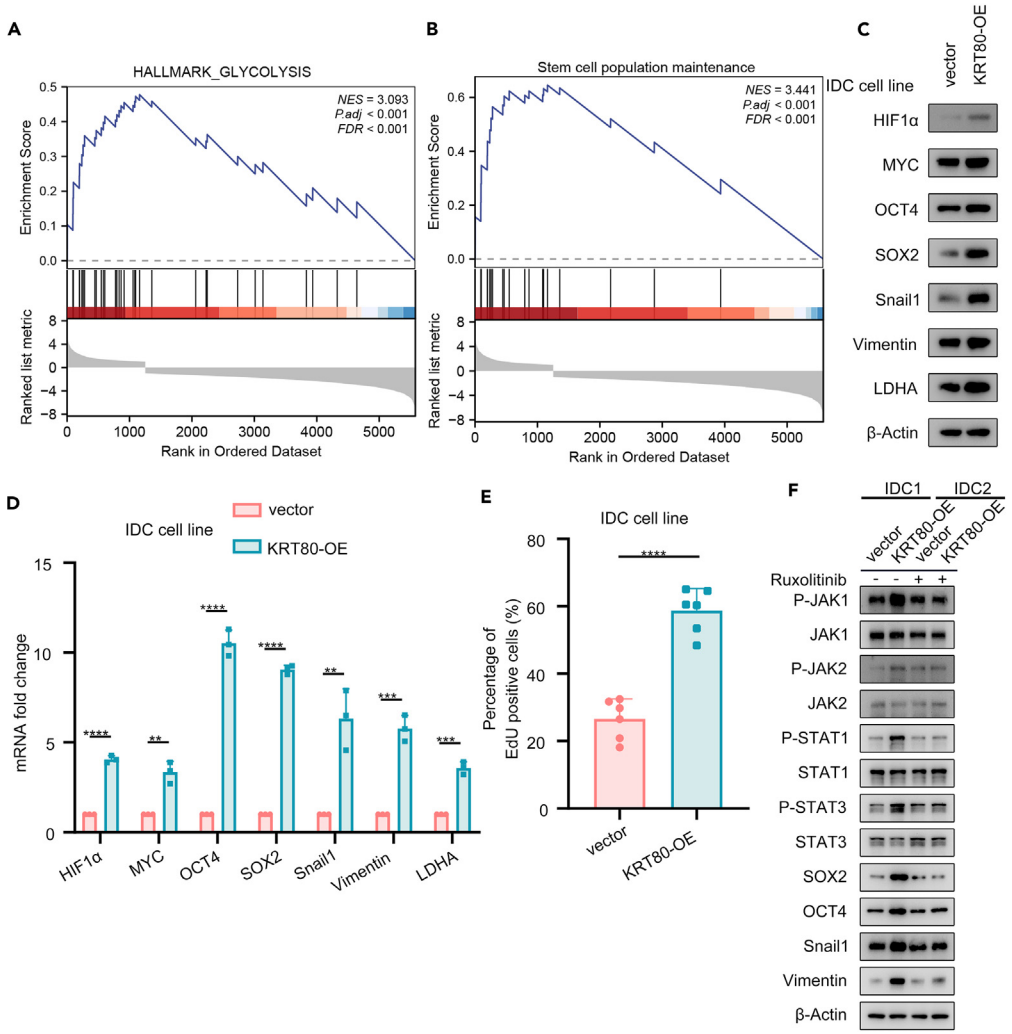


Figure 6. Exploring the potential mechanism and phenotypic changes of KRT80 in IDC cells

(A and B) The gene set enrichment analysis (GSEA) between KRT80-high and KRT80-low group.

(C) The protein expression level of each group by western blot.

(D) The mRNA expression level of each group by qPCR. Data are represented as mean ± SEM.

(E) The percentage of Edu positive cells of each group by flow cytometry. Data are represented as mean ± SEM.

(F) The activation of JAK-STAT signaling pathway and the related protein expression level of each group by western blot. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$ and n.s., non-significant.

cell clusters and the evolutionary trajectory of IMPC cell subsets, we sequenced IMPC at the single-cell transcriptome level to investigate the transcriptome characteristics of IMPC. Many studies have shown that the formation of tumor cell clusters can significantly enhance metastatic potential[30,31] and induce a variety of molecular properties, including increased stem cell-like properties,[32] evasion of natural killer cell targeting,[33] radioresistance,[34] and resistance to metabolic stress.[35] Our three-dimensional culture of IMPC primary cells proved that IMPC involves tumor cell clusters with reversed polarity. The characteristics of IMPC cluster growth and invasion support the highly invasive and metastatic biological behavior of IMPC.

Insulin can stimulate the proliferation of certain human breast cancer cell lines *in vitro* by utilizing the phosphatidylinositol 3-kinase and mitogen-activated protein kinase/Akt signaling pathways. It is also an antiapoptotic agent and enhances tumor cell migration and invasion ability.[36] Abnormal fatty acid metabolism often occurs in tumor cells, and tumor cells prefer to synthesize fatty acids *de novo* rather than exogenously (as in normal cells).[37] From the results of GO and KEGG enrichment analysis, we can see that IMPC is a metabolically abnormal tumor. The unique growth pattern and biological behavior of IMPC may be related to the metabolic abnormalities that occur in its tumor cells.

The identification of cell subpopulations in the single-cell transcription map of the mixed IMPC is an exciting finding. Fu et al. observed abundant microvilli on the cell surface of the side of the IMPC cancer nest in contact with the stroma,[9] and the cytoplasm was rich in filaments

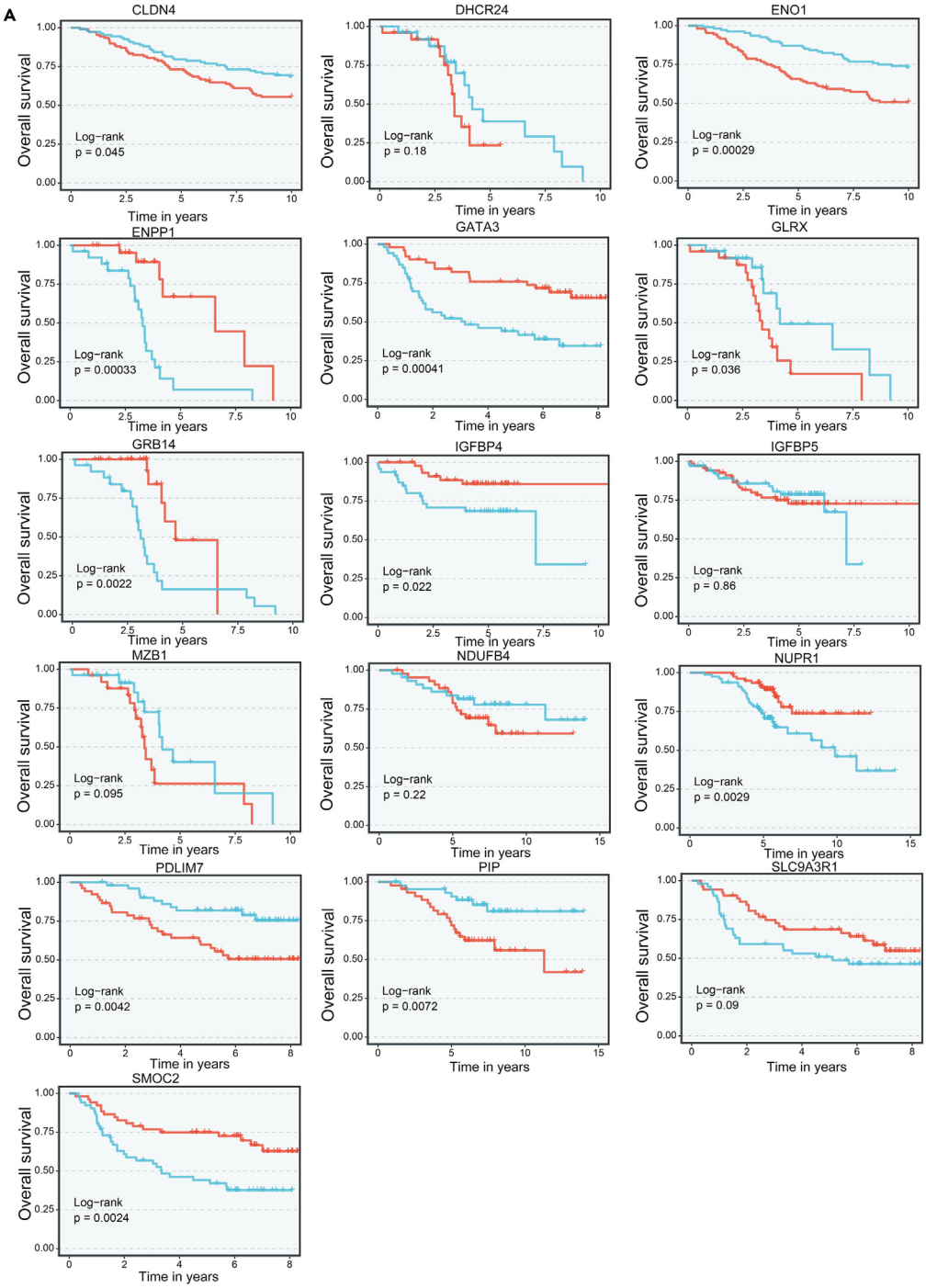


Figure 7. Cell trajectory-related genes took a role in IMPC patients' survival

(A) Survival curve showing overall survival difference in high and low expression groups. The red curve means a high expression group and the shorter line means a poor prognosis.

under an electron microscope. In our study, the IMPC atlas is largely consistent with previous morphological observations of IMPC. Villous stromal cells (cluster 1) are rich in mitochondrial dynein, which can promote the transfer of cell mass to the matrix. The highly expressed villus cell marker genes *CTR61* and *CYR61* are related to adhesion, the RhoA signaling pathway, angiogenesis, tumor vascular endothelial migration, and the formation of extracellular matrix.[38,39] High expression of the marker gene *ATF3* in the ciliated cell population (cluster 2) can enhance epithelial-mesenchymal transition and promote cancer cell proliferation and migration.[40,41] These two groups of cells can

functionally enhance the overall mobility of IMPC tumor cell clusters and promote the migration and invasion of IMPC to the tumor stroma and vascular endothelium.

It is worth mentioning that the infiltrating immune cells (cluster 8) expressed the tumor immune microenvironment marker *IFI30*. There were $CD68^+$ macrophages, which indicated that IMPC patients had innate immune responses.[42] *IFI27* was highly expressed in the immune cell cluster and participates in innate immunity, interferon γ signaling, and type I interferon-induced apoptosis, which can enable rapid and stable release of cytochrome *c* in mitochondria. *IFI27* also functions in *TNFSF10*-induced cell apoptosis.[43–46] These gene expression changes are important for identifying early biomarkers and predicting the immunotherapy response in IMPC, and the characterization of tumor-infiltrating immune cells may reveal the strategy of IMPC immune escape.

Our pseudo-time trajectory analysis shows that IDC-NOS represents cells in the initial stage of differentiation; these cells can differentiate into epithelial-like IMPC cells and interstitial-like IMPC cells. We found that the differentially expressed gene *TM4SF1* of the IDC-NOS cell group (cluster 4) was related to the regulation of apoptosis[47,48]; *TM4SF1* facilitates epithelial-mesenchymal transition,[49] self-renewal,[50] tumor angiogenesis,[51] and invasion pseudopodia formation[52] and regulates related signaling pathways to promote the migration and invasion of cancer cells. Differently, the differentially expressed genes of epithelial-like IMPC cells have the functions of promoting mitosis, regulating tight junctions and hormonal activity, participating in redox stress, and regulating the activity of transcription factors. Among the highly expressed genes in epithelial-like IMPC cells, *CD24* is related to epithelial differentiation. LI et al. also confirmed that the proportion of $CD24^+$ tumor cells in IMPC was higher than that in IDC-NOS.[11] The highly expressed transcription factors *GATA3*, *SPDEF*, and *NDUFB4* in epithelial-like IMPC can protect IMPC cells from stress-induced cell death by binding to the promoter and activating its transcription. These genes may be the reason the IMPC tumor cell cluster can proliferate and resist cell apoptosis in the nest.

KRT80, as an intermediate filament protein responsible for the structural integrity of epithelial cells, is differentially expressed in the two subgroups of IMPC and is highly expressed in the relatively small interstitial-like IMPC. The expression of KRT80 was significantly different in a single IMPC cell cluster but not in the whole tumor tissue. KRT80 was positively correlated with lymph node metastasis and lymphatic vascular invasion in IMPC patients, which further showed that KRT80 affects the malignant biological behavior of IMPC by regulating the structural integrity of epithelial cells. In IMPC tumor cell clusters, the expression of the intercellular adhesion molecule E-Ca increased in epithelial-like IMPC cells and decreased in interstitial-like IMPC cells. Furthermore, we found that *FASN*, *CLDN4*, and *EPCAM*, which are connected to the epithelium, were significantly reduced in cluster 6, which may be the reason the IMPC tumor cell mass detaches from the matrix. Overall, epithelial-like IMPC cells express genes that promote mitosis, regulate tight junctions, maintain a stem cell-like phenotype to resist apoptosis, promote epithelial cell growth, and regulate the activity of transcription factors; these properties allow IMPC cell clusters to form, grow, and proliferate. KRT80 and most of the pseudo-time trajectories' genes that promote the evolution of IMPC tumor formation were associated with poor prognosis in breast cancer patients and could be potential therapeutic targets. Conventional chemotherapeutic agents such as paclitaxel, platinum, and cyclophosphamide are generally used for breast cancer treatment, but the analysis of pan-cancer cell line drugs predicted that evolutionary-related genes, including KRT80, were resistant to conventional chemotherapeutic agents. Therefore, in the treatment of IMPC patients, we should select precisely targeted drugs and develop an effective treatment strategy. However, our study has some defects, these associations between potential therapeutic targets and drugs are predicted by bioinformatics algorithms. We do not have enough research on drug mechanisms and we need more *in vivo/in vitro* experiments of drugs targeting the potential genes and multi-center clinical drug trials.

In summary, using a combination of single-cell transcriptome sequencing and bulk RNA-seq, we generated a single-cell transcription atlas of mixed IMPC; this revealed the heterogeneity of mixed IMPC and the molecular characteristics of each subgroup and allowed us to trace the lineage level of these cells. We also identified the key genes at the subgroup differentiation node within IMPC, such as KRT80. These findings will help to develop precision medical treatment for IMPC based on its features and immune environment. These findings also support the hypothesis that the reverse polarity of tumor cell clusters facilitates the invasion and metastasis of IMPC.

Limitations of the study

The small sample size of our study may affect its statistical strength and the authenticity of the results. We do not have enough research on drug mechanisms, and we need more multi-center clinical drug trials for compensating the limitations of experiments.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Xiaobin Shang (shangxiaobin@tmu.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- RNA-seq count data and single-cell RNA-seq have been deposited in the National Omics Data Encyclopedia (NODE) database and are publicly available as of the date of publication. Accession numbers are listed in the key resources table.
- This paper does not report the original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank all the participants included in this study. This work was supported by grants from Beijing-Tianjin-Hebei Basic Research Project of Tianjin Science and Technology Bureau (Grant Numbers: 22JCZXJC00040); Public Health Science and Technology Major Project of the Tianjin Science and Technology Bureau (grant numbers: 21ZXGWSY00020); Tianjin Key Medical Discipline (Specialty) Construction Project (grant numbers: TJYXZDXK-010A).

AUTHOR CONTRIBUTIONS

Writing – original draft, Y.X. and Z.L.; investigation, J.Z. and B.N.; methodology, C.S. and G.L.; project administration, Y.Z.; visualization, Y.Z.; funding acquisition, X.S. All authors have read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Analysis of copy number alterations
  - Immunohistochemistry
  - Primary tumor cell culture
  - 3D culture and immunofluorescence staining of primary tumor cells
  - Bulk-RNA-seq
  - Differential expression analysis
  - Enrichment analysis of differentially expressed genes
  - Single-cell RNA sequencing analysis
  - Processing and analysis of single-cell RNA sequencing data
  - Gene set variation analysis (GSVA) and gene set enrichment analysis (GSEA) of IMPC
  - Pseudotiming analysis of IMPC
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.111178.

Received: April 7, 2024
Revised: August 22, 2024
Accepted: October 14, 2024
Published: October 16, 2024

REFERENCES

1. Amin, M.B., Ro, J.Y., el-Sharkawy, T., Lee, K.M., Troncoso, P., Silva, E.G., Ordóñez, N.G., and Ayala, A.G. (1994). Micropapillary variant of transitional cell carcinoma of the urinary bladder. Histologic pattern resembling ovarian papillary serous carcinoma. Am. J. Surg. Pathol. *18*, 1224–1232.
2. Kitagawa, H., Nakamura, M., Tani, T., Tajima, H., Nakagawara, H., Ohnishi, I., Takamura, H., Kayahara, M., Ohta, T., Zen, Y., et al. (2007). A pure invasive micropapillary carcinoma of the pancreatic head: long disease-free survival after pancreatoduodenectomy and adjuvant chemotherapy with gemcitabine. Pancreas *35*, 190–192.
3. Luna-More, S., Gonzalez, B., Acedo, C., Rodrigo, I., and Luna, C. (1994). Invasive micropapillary carcinoma of the breast. A new special type of invasive mammary carcinoma. Pathol. Res. Pract. *190*, 668–674.
4. Miyaoka, Y., Fujiwara, A., Kotani, S., Tsukano, K., Ogawa, S., Yamanouchi, S., Kusunoki, R., Fujishiro, H., Kohge, N., Yamamoto, T., and Amano, Y. (2016). Primary micropapillary carcinoma of the colon with submucosal invasion: A case report. Endosc. Int. Open *4*, E744–E747.
5. Monroig-Bosque, P.D.C., Morales-Rosado, J.A., Roden, A.C., Churg, A., Barrios, R., Cagle, P., Ge, Y., Allen, T.C., Smith, M.L., Larsen, B.T., et al. (2019). Micropapillary adenocarcinoma of lung: Morphological criteria and diagnostic reproducibility among pulmonary pathologists. Ann. Diagn. Pathol. *41*, 43–50.
6. Yang, Y., Kaimakliotis, H.Z., Williamson, S.R., Koch, M.O., Huang, K., Barboza, M.P., Zhang, S., Wang, M., Idrees, M.T., Grignon, D.J., et al. (2020). Micropapillary urothelial carcinoma of urinary bladder displays immunophenotypic features of luminal and p53-like subtypes and is not a variant of adenocarcinoma. Urol. Oncol. *38*, 449–458.
7. Siriaunkgul, S., and Tavassoli, F.A. (1993). Invasive micropapillary carcinoma of the breast. Mod. Pathol. *6*, 660–662.
8. FA, T., and P, D. (2003). WHO classtfication of tumours,pathology &genetics,tumours ofthe breast and femals genital organs. Lyonia *10*.
9. Fu, L., Ikuo, M., Fu, X.Y., Liu, T.H., and Shinichi, T. (2004). [Relationship between biologic behavior and morphologic features of invasive micropapillary carcinoma of the breast]. Zhonghua Bing li xue za zhi= Chinese Journal of Pathology *33*, 21–25.
10. Yang, Y.L., Liu, B.B., Zhang, X., and Fu, L. (2016). Invasive Micropapillary Carcinoma of the Breast: An Update. Arch. Pathol. Lab Med. *140*, 799–805.
11. Li, W., Liu, F., Lei, T., Xu, X., Liu, B., Cui, L., Wei, J., Guo, X., Lang, R., Fan, Y., et al. (2010). The clinicopathological significance of CD44+/CD24-/low and CD24+ tumor cells in invasive micropapillary carcinoma of the breast. Pathol. Res. Pract. *206*, 828–834.
12. Li, W., Yang, D., Wang, S., Guo, X., Lang, R., Fan, Y., Gu, F., Zhang, X., Niu, Y., Yan, X., and Fu, L. (2011). Increased expression of CD146 and microvessel density (MVD) in invasive micropapillary carcinoma of the breast: Comparative study with invasive ductal carcinoma-not otherwise specified. Pathol. Res. Pract. *207*, 739–746.
13. Liu, F., Lang, R., Wei, J., Fan, Y., Cui, L., Gu, F., Guo, X., Pringle, G.A., Zhang, X., and Fu, L. (2009). Increased expression of SDF-1/CXCR4 is associated with lymph node metastasis of invasive micropapillary carcinoma of the breast. Histopathology *54*, 741–750.
14. Liu, F., Yang, M., Li, Z., Guo, X., Lin, Y., Lang, R., Shen, B., Pringle, G., Zhang, X., and Fu, L. (2015). Invasive micropapillary mucinous carcinoma of the breast is associated with

poor prognosis. Breast Cancer Res. Treat. *151*, 443–451.
15. Wei, J., Cui, L., Liu, F., Fan, Y., Lang, R., Gu, F., Guo, X., Tang, P., and Fu, L. (2010). E-selectin and Sialyl Lewis X expression is associated with lymph node metastasis of invasive micropapillary carcinoma of the breast. Int. J. Surg. Pathol. *18*, 193–200.
16. Li, W., Han, Y., Wang, C., Guo, X., Shen, B., Liu, F., Jiang, C., Li, Y., Yang, Y., Lang, R., et al. (2018). Precise pathologic diagnosis and individualized treatment improve the outcomes of invasive micropapillary carcinoma of the breast: a 12-year prospective clinical study. Mod. Pathol. *31*, 956–964.
17. Guo, X., Chen, L., Lang, R., Fan, Y., Zhang, X., and Fu, L. (2006). Invasive Micropapillary Carcinoma of the Breast. Am. J. Clin. Pathol. *126*, 740–746.
18. Li, S., Yang, C., Zhai, L., Zhang, W., Yu, J., Gu, F., Lang, R., Fan, Y., Gong, M., Zhang, X., and Fu, L. (2012). Deep sequencing reveals small RNA characterization of invasive micropapillary carcinomas of the breast. Breast Cancer Res. Treat. *136*, 77–87.
19. Liu, B., Zheng, X., Meng, F., Han, Y., Song, Y., Liu, F., Li, S., Zhang, L., Gu, F., Zhang, X., and Fu, L. (2018). Overexpression of β1 integrin contributes to polarity reversal and a poor prognosis of breast invasive micropapillary carcinoma. Oncotarget *9*, 4338–4353.
20. Meng, F., Liu, B., Xie, G., Song, Y., Zheng, X., Qian, X., Li, S., Jia, H., Zhang, X., Zhang, L., et al. (2017). Amplification and overexpression of PSCA at 8q24 in invasive micropapillary carcinoma of breast. Breast Cancer Res. Treat. *166*, 383–392.
21. Chapman, P.B., Hauschild, A., Robert, C., Haanen, J.B., Ascierto, P., Larkin, J., Dummer, R., Garbe, C., Testori, A., Maio, M., et al. (2011). Improved survival with vemurafenib in melanoma with BRAF V600E mutation. N. Engl. J. Med. *364*, 2507–2516.
22. Cursons, J., Souza-Fonseca-Guimaraes, F., Foroutan, M., Anderson, A., Hollande, F., Hediyeh-Zadeh, S., Behren, A., Huntington, N.D., and Davis, M.J. (2019). A Gene Signature Predicting Natural Killer Cell Infiltration and Improved Survival in Melanoma Patients. Cancer Immunol. Res. *7*, 1162–1174.
23. Navin, N.E. (2015). The first five years of single-cell cancer genomics and beyond. Genome Res. *25*, 1499–1507.
24. Tanay, A., and Regev, A. (2017). Scaling single-cell genomics from phenomenology to mechanism. Nature *541*, 331–338.
25. Peng, J., Sun, B.F., Chen, C.Y., Zhou, J.Y., Chen, Y.S., Chen, H., Liu, L., Huang, D., Jiang, J., Cui, G.S., et al. (2019). Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. Cell Res. *29*, 725–738.
26. Gužvić, M., Braun, B., Ganzer, R., Burger, M., Nerlich, M., Winkler, S., Werner-Klein, M., Czyz, Z.T., Polzer, B., and Klein, C.A. (2014). Combined genome and transcriptome analysis of single disseminated cancer cells from bone marrow of prostate cancer patients reveals unexpected transcriptomes. Cancer Res. *74*, 7383–7394.
27. Zhang, F., Wei, K., Slowikowski, K., Fonseka, C.Y., Rao, D.A., Kelly, S., Goodman, S.M., Tabechian, D., Hughes, L.B., Salomon-Escoto, K., et al. (2019). Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry. Nat. Immunol. *20*, 928–942.
28. Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science *344*, 1396–1401.
29. Tarhriz, V., Bandehpour, M., Dastmalchi, S., Ouladsahebmadarek, E., Zarredar, H., and Eyvazi, S. (2019). Overview of CD24 as a new molecular marker in ovarian cancer. J. Cell. Physiol. *234*, 2134–2142.
30. Aceto, N., Bardia, A., Miyamoto, D.T., Donaldson, M.C., Wittner, B.S., Spencer, J.A., Yu, M., Pely, A., Engstrom, A., Zhu, H., et al. (2014). Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. Cell *158*, 1110–1122.
31. Cheung, K.J., and Ewald, A.J. (2016). A collective route to metastasis: Seeding by tumor cell clusters. Science *352*, 167–169.
32. Gkountela, S., Castro-Giner, F., Szczerba, B.M., Vetter, M., Landin, J., Scherrer, R., Krol, I., Scheidmann, M.C., Beisel, C., Stirnimann, C.U., et al. (2019). Circulating Tumor Cell Clustering Shapes DNA Methylation to Enable Metastasis Seeding. Cell *176*, 98–112.
33. Lo, H.C., Xu, Z., Kim, I.S., Pingel, B., Aguirre, S., Kodali, S., Liu, J., Zhang, W., Muscarella, A.M., Hein, S.M., et al. (2020). Resistance to natural killer cell immunosurveillance confers a selective advantage to polyclonal metastasis. Nat. Can. (Ott.) *1*, 709–722.
34. Haeger, A., Alexander, S., Vullings, M., Kaiser, F.M.P., Veelken, C., Flucke, U., Koehl, G.E., Hirschberg, M., Flentje, M., Hoffman, R.M., et al. (2020). Collective cancer invasion forms an integrin-dependent radioresistant niche. J. Exp. Med. *217*, e20181184.
35. Labuschagne, C.F., Cheung, E.C., Blagih, J., Domart, M.C., and Vousden, K.H. (2019). Cell Clustering Promotes a Metabolic Switch that Supports Metastatic Colonization. Cell Metabol. *30*, 720–734.
36. Rose, D.P., and Vona-Davis, L. (2012). The cellular and molecular mechanisms by which insulin influences breast cancer risk and progression. Endocr. Relat. Cancer *19*, R225–R241.
37. Röhrig, F., and Schulze, A. (2016). The multifaceted roles of fatty acid synthesis in cancer. Nat. Rev. Cancer *16*, 732–749.
38. Chen, C.C., Mo, F.E., and Lau, L.F. (2001). The angiogenic factor Cyr61 activates a genetic program for wound healing in human skin fibroblasts. J. Biol. Chem. *276*, 47329–47337.
39. Han, J.S., Macarak, E., Rosenbloom, J., Chung, K.C., and Chaqour, B. (2003). Regulation of Cyr61/CCN1 gene expression through RhoA GTPase and p38MAPK signaling pathways. Eur. J. Biochem. *270*, 3408–3421.
40. Yin, X., Wolford, C.C., Chang, Y.S., McConoughey, S.J., Ramsey, S.A., Aderem, A., and Hai, T. (2010). ATF3, an adaptive-response gene, enhances TGF{beta} signaling and cancer-initiating cell features in breast cancer cells. J. Cell Sci. *123*, 3558–3565.
41. Gokulnath, M., Partridge, N.C., and Selvamurugan, N. (2015). Runx2, a target gene for activating transcription factor-3 in human breast cancer cells. Tumour Biol. *36*, 1923–1931.
42. Satoh, J.i., Obayashi, S., Misawa, T., Tabunoki, H., Yamamura, T., Arima, K., and Konno, H. (2008). Neuromyelitis optica/Devic's disease: gene expression profiling of brain lesions. Neuropathology *28*, 561–576.
43. Gytz, H., Hansen, M.F., Skovbjerg, S., Kristensen, A.C.M., Hørlyck, S., Jensen, M.B., Fredborg, M., Markert, L.D., McMillan, N.A., Christensen, E.I., and Martensen, P.M. (2017). Apoptotic properties of the type 1 interferon induced family of human mitochondrial membrane ISG12 proteins. Biol. Cell. *109*, 94–112.
44. Liu, N., Zuo, C., Wang, X., Chen, T., Yang, D., Wang, J., and Zhu, H. (2014). miR-942 decreases TRAIL-induced apoptosis through ISG12a downregulation and is regulated by AKT. Oncotarget *5*, 4959–4971.
45. Papac-Milicevic, N., Breuss, J.M., Zaujec, J., Ryban, L., Plyushch, T., Wagner, G.A., Fenzl, S., Dremsek, P., Cabaravdic, M., Steiner, M., et al. (2012). The interferon stimulated gene 12 inactivates vasculoprotective functions of NR4A nuclear receptors. Circ. Res. *110*, e50–e63.
46. Rosebeck, S., and Leaman, D.W. (2008). Mitochondrial localization and pro-apoptotic effects of the interferon-inducible protein ISG12a. Apoptosis *13*, 562–572.
47. Stingl, J., Eirew, P., Ricketson, I., Shackleton, M., Vaillant, F., Choi, D., Li, H.I., and Eaves, C.J. (2006). Purification and unique properties of mammary epithelial stem cells. Nature *439*, 993–997.
48. Sun, Y., Xu, Y., Xu, J., Lu, D., and Wang, J. (2015). Role of TM4SF1 in regulating breast cancer cell migration and apoptosis through PI3K/AKT/mTOR pathway. Int. J. Clin. Exp. Pathol. *8*, 9081–9088.
49. Tang, Q., Chen, J., Di, Z., Yuan, W., Zhou, Z., Liu, Z., Han, S., Liu, Y., Ying, G., Shu, X., and Di, M. (2020). TM4SF1 promotes EMT and cancer stemness via the Wnt/β-catenin/SOX2 pathway in colorectal cancer. J. Exp. Clin. Cancer Res. *39*, 232.
50. Gao, H., Chakraborty, G., Zhang, Z., Akalay, I., Gadiya, M., Gao, Y., Sinha, S., Hu, J., Jiang, C., Akram, M., et al. (2016). Multi-organ Site Metastatic Reactivation Mediated by Non-canonical Discoidin Domain Receptor 1 Signaling. Cell *166*, 47–62.
51. Shih, S.C., Zukauskas, A., Li, D., Liu, G., Ang, L.H., Nagy, J.A., Brown, L.F., and Dvorak, H.F. (2009). The L6 protein TM4SF1 is critical for endothelial cell function and tumor angiogenesis. Cancer Res. *69*, 3272–3277.
52. Yang, J.C., Zhang, Y., He, S.J., Li, M.M., Cai, X.L., Wang, H., Xu, L.M., and Cao, J. (2017). TM4SF1 Promotes Metastasis of Pancreatic Cancer via Regulating the Expression of DDR1. Sci. Rep. *7*, 45895.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies name		
Anti-HIF1α Antibody (for WB)	Abcam	ab51608; RRID: AB_880418
Anti-c-Myc Antibody (for WB)	Abcam	ab32072; RRID: AB_731658
Anti-OCT4 Antibody (for WB)	Abcam	ab200834; RRID: AB_2924374
Anti-SOX2 Antibody (for WB)	Abcam	ab92494; RRID: AB_10585428
Anti-Snail Antibody (for WB)	Abcam	ab216347; RRID: AB_2910593
Anti-Vimentin Antibody (for WB)	Abcam	ab20346; RRID: AB_445527
Anti-LDHA Antibody (for WB)	Abcam	ab300637; RRID: AB_300637
Anti-Jak1 Antibody (for WB)	CST	Cat#3344; RRID: AB_2265054
Anti-Phospho-Jak1 (Tyr1034/1035) Antibody (for WB)	CST	Cat#3331; RRID: AB_2265057
Anti-Jak2 Antibody (for WB)	CST	Cat#3230; RRID: AB_2128522
Anti-Phospho-Jak2 (Tyr1007/1008) Antibody (for WB)	CST	Cat#3771; RRID: AB_330403
Anti-Stat1 Antibody (for WB)	CST	Cat#14994; RRID: AB_2737027
Anti-Phospho-Stat1 (Tyr701) Antibody (for WB)	CST	Cat#9167; RRID: AB_561284
Anti-Stat3 Antibody (for WB)	CST	Cat#12640; RRID: AB_2629499
Anti-Phospho-Stat3 (Tyr705) Antibody (for WB)	CST	Cat#9145; RRID: AB_2491009
Anti-β-Actin Antibody (for WB)	Beijing Ray Antibody Biotech	RM2001L; RRID: AB_2756462
Deposited data		
Single-cell sequencing data and bulk transcriptome data	NODE	
IMPC copy number alterations	GEO	GSE37035
IMPC bulk transcriptome data	GEO	GSE66418
Software and algorithms		
R software version 4.2.0	R Core Team	https://www.r-project.org/
PennCNV software	Perl/C-based	https://penncnv.openbioinformatics.org/en/latest/
GISTIC2 software	MATLAB	https://broadinstitute.github.io/gistic2/
SPSS version 19.0	C++	https://www.ibm.com/cn-zh/products/spss-statistics
Cell Ranger 1.3.1	10× genomic	https://www.10xgenomics.com/cn/support/software/cell-ranger/

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

For bulk transcriptome, 18 fresh tissues from 9 female patients with IMPC and IDC-NOS, as well as normal breast tissues and FFPE tumor tissues of IMPC and IDC-NOS, were collected at the Department of Pathology in Wuhu Hospital of Traditional Chinese Medicine, The Second Affiliated Hospital of Guangzhou Medical University, and The Affiliated Tumor Hospital of Tianjin Medical University. Clinicopathological information from these patients is detailed in Table S1. For single-cell RNA-seq, the tissue samples from the 3 female cases of mixed IMPC tumors (comprising IMPC with an IDC-NOS component) were obtained from the Second Affiliated Hospital of Guangzhou Medical University. Clinicopathological information from three patients is detailed in Table S2. Because breast cancer occurs almost exclusively in women, only female was used in the analysis of these patients. No information about ancestry, race, ethnicity, or socioeconomic status was collected for any of the participants. All patients provided informed consent before tissue collection. The study received approval from the Ethics Committee of the Affiliated Cancer Hospital of Tianjin Medical University.

METHOD DETAILS

Analysis of copy number alterations

To characterize CNVs, every sample was profiled on Affymetrix SNP 6.0 arrays downloaded from the GEO database GSE37035. To identify CNVs, we used the "PennCNV" software package. This algorithm employs a hidden Markov model to segment the total signal intensity for both alleles (log R ratio, or LRR) and allelic intensity ratio between the two alleles (B allele frequency, or BAF) for each probe across the genome. After obtaining LRR and BAF, we used the "DNAcopy" of the R package to produce segment files. The GISTIC2 software was used to identify significant CNV regions and peaks, we set the threshold at $q < 0.25$ and confidence at 0.95.

Immunohistochemistry

Four-micrometer serial tissue sections were meticulously extracted from archived formalin-fixed paraffin-embedded tissue blocks, meticulously dewaxed, and subsequently rehydrated via xylene and graded alcohol washes. Antigen retrieval procedures were meticulously carried out in EDTA buffer for precisely 2 min and 30 s. Subsequently, a 3% hydrogen peroxide treatment lasting 10 min was executed to impede endogenous peroxidase activity, followed by a 10-min incubation period with normal goat serum to eliminate nonspecific background staining. Thereafter, primary antibodies targeting EMA (ZSGB-bio, ZM-0095, monoclonal, China) and FASN (Abcam, ab128870, monoclonal, UK) were diligently incubated at 4°C overnight. Antigens were then sequentially detected using secondary biotin-labeled antibodies and peroxidase-conjugated streptavidin. The chromogen utilized was 3,3-diaminobenzidine, with the sections subsequently counterstained with hematoxylin.

Primary tumor cell culture

Samples from 4 IMPC and 5 IDC-NOS cases were obtained for the purpose of primary cell culture. The samples underwent two washes in normal saline and were then finely dissected into small fragments (<1 mm). Subsequently, 1.5 mL of the cell dispersing enzyme EZ solution was introduced, and digestion occurred at 37°C for 2 h within an oscillator. Once digestion was complete (noted by a notable release of cells from the tissue under microscopic examination), the cells were filtered through a 308-μm nylon net. The cell precipitates were then suspended in 3 mL of modified medium and transferred to a 6-cm dish. Primary tumor cells from both IMPC and IDC-NOS were cultured in DMEM/F12 enriched with 5% horse serum, 10 μg/mL insulin, 20 ng maximum EGF, 0.5 μg/mL hydrocortisone, 100 ng/mL cholera toxin, 10 μM RhoA kinase inhibitor (YMEC 27632), 1% L-glutamine, 1% pyruvate, 0.05% bovine pituitary extract, and 1% penicillin-streptomycin. The cell cultures were maintained in an incubator at 37°C with 5% CO2.

3D culture and immunofluorescence staining of primary tumor cells

The primary cell suspensions of IMPC and IDC-NOS were combined with collagens A, B, and C (from Osaka, Japan) at an 8:1:1 ratio on ice. Subsequently, 30 μL per drop was seeded onto a glass coverslip coated with fibronectin, positioned in a 6-well plate (Corning, USA). Following a 30-min incubation period in a 37°C, 5% CO2 environment, the collagen drops solidified, after which 2 mL of medium was added for sustained culture. The floating IMPC and IDC-NOS cells within the collagen droplets were then fixed using 4% paraformaldehyde and permeabilized with 0.2% Triton X-100. Subsequently, antibodies targeting sLex (BD, 551344, CSLEX1, monoclonal, diluted at 1:150, USA) or EMA (ZSGB-Bio, ZM-0095, monoclonal, China) were incubated overnight at 4°C. The secondary antibody was applied at room temperature for 1 h. DAPI (Solarbio, Beijing, China) was utilized for nuclear staining, and imaging was performed using confocal microscopy.

Bulk-RNA-seq

The primary cells cultured from the 4 IMPC cases, 5 IDC-NOS cases, and adjacent normal breast tissues underwent bulk RNA-seq analysis. In each sample, a total of 3 μg of RNA was utilized as input material for bulk RNA-seq. The NEBNext UltraTM RNA Library Prep Kit for Illumina (NEB, USA) was employed to construct the sequencing library, with an index code added to the sample's attribute sequence. Briefly, mRNA was isolated from total RNA using poly T oligomers linked to magnetic beads. Within the NEBNext first-strand synthesis reaction buffer (5X), divalent cations were heat-denatured. Random hexamer primers and M-MuLV reverse transcriptase (RNaseH-) were used to synthesize the first cDNA strand. Subsequently, DNA polymerase I and ribonuclease H were employed to generate the second cDNA strand, ensuring the remaining overhangs were converted into blunt ends through exonuclease/polymerase activity. Following adenylation at the DNA fragment's 3′ end, hybridization occurred with the NEBNext adapter possessing a hairpin loop structure. Purification of the library fragment, ranging from 250–300 bp, was carried out using the AMPureXP system (Beckman Coulter, Beverly, USA). Subsequently, the cDNAs selected based on size and junction sequence were enzymatically treated with 3 μL USER enzyme (NEB, USA) at 37°C for 15 min followed by a 5-min incubation at 95°C before PCR amplification. Phusion high-fidelity DNA polymerase, universal PCR primers, and index (X) primers were utilized for the PCR step. Lastly, the PCR products were purified using the AMPureXP system, and the library quality was assessed using an Agilent BioAnalyzer 2100 instrument. The TruSeqPE Cluster Kit v3-CBOT-HS (Illumina) was used for clustering the samples indexed on the CBOT clustering generation system. Following cluster formation, the libraries were sequenced on the Illumina HiSeq platform, generating paired-end reads of 125 bp/150 bp. For the primary suspension cell mass preparation of the IMPC tumor, upon the suspension cell mass reaching a diameter of 2–3 mm, the IMPC cell mass was dissociated and cultured for 3–5 days. Subsequently, the IMPC cell mass was harvested for bulk RNA-seq analysis.

Differential expression analysis

Differential expression analysis of IMPC tumor cells and IDC-NOS tumor cells (with two biological replicates per condition) from bulk RNA-seq data was conducted using the DESeq2 R package (version 1.20.0). DESeq2 offers statistical tools for identifying differential expression in digital gene expression data by employing a model based on the negative binomial distribution. The resultant *p*-values were adjusted utilizing the Benjamini and Hochberg methods to control the false discovery rate. A threshold of padj <0.05 and |log2(fold change)| > 1 was established to define significant differential expression levels.

Enrichment analysis of differentially expressed genes

Gene Ontology (GO) enrichment analysis of differentially expressed genes was conducted using the clusterProfiler R package (version 3.8.1), with gene length bias correction applied. GO terms exhibiting a corrected *p*-value below 0.05 were deemed significantly enriched by differentially expressed genes. The Kyoto Encyclopedia of Genes and Genomes (KEGG) database serves as a valuable resource for comprehending the high-level functions and biological mechanisms of systems ranging from cells to organisms and ecosystems, leveraging molecular-level data, particularly large-scale molecular datasets generated through genome sequencing and other high-throughput experimental technologies (http://www.genome.jp/kegg/). The clusterProfiler R package (version 3.8.1) was utilized to assess the statistical enrichment of differentially expressed genes in KEGG pathways. The Reactome database amalgamates diverse reactions and biological pathways of model species, with Reactome pathways exhibiting a corrected *p*-value under 0.05 being considered significantly enriched by differentially expressed genes. Furthermore, the Disease Ontology (DO) database delineates human gene functionality and disease associations, with DO pathways featuring a corrected *p*-value less than 0.05 is regarded as significantly enriched by differentially expressed genes. Lastly, the DisGeNET database integrates genes linked to human diseases, and DisGeNET pathways surpassing a corrected *p*-value of 0.05 are perceived as significantly enriched by differentially expressed genes. The clusterProfiler R package (version 3.8.1) was employed to evaluate the statistical enrichment of differentially expressed genes across Reactome, DO, and DisGeNET pathways.

Single-cell RNA sequencing analysis

The tissue samples from the 3 cases of mixed IMPC tumors (comprising IMPC with an IDC-NOS component) were enzymatically dissociated into single cells using Accutase. These isolated single cells underwent two washes with 3 mL of Hank's solution (lacking calcium and magnesium ions) and were subsequently resuspended in 3 mL of Hank's solution (without calcium and magnesium ions). The single cells were stained with trypan blue, and their viability was determined to be in the range of 79%–91%. A single-cell suspension was loaded into a chromium microfluidic chip via the Chromium Single Cell 3′ v2 chemical approach, with barcoding accomplished using a 10X Chromium Controller (10X Genomics). The RNA, together with cell barcodes, underwent reverse transcription, and the sequencing library was prepared utilizing reagents from the Chromium Single Cell 3′ v2 Kit (10X Genomics). Subsequently, the library was sequenced using an Illumina instrument (NovaSeq) (refer to Figure 3A). A total of 9876 captured cells were sequenced in parallel, each exhibiting an average read depth of 30 M per cell to generate a single-cell cDNA library.

Processing and analysis of single-cell RNA sequencing data

The construction of a 10X single-cell sequencing 3′ end counting library was accomplished using Cell Ranger 1.3.1 (10X Genomics). Cell Ranger counts were employed to align each library to the indexed GRCh38 genome. The resultant data matrix was imported into R, and Seurat (Version 3.1.3) was utilized for unsupervised clustering. Cell filtration criteria required that each gene be expressed in a minimum of three cells, with at least 200 genes expressed in each cell. Cells expressing either too few or too many genes (<200 genes, >7500 genes) and those displaying high mitochondrial gene coverage (>5%) were excluded. Following quality control measures, a total of 8230 single cells and 22,215 genes were retained. Normalization of each sample was performed using "NormalizeData." To mitigate batch effects, "CCA" and "IntegrateData" functions were applied to merge samples and eliminate batch variations. Subsequently, "ScaleData" was employed for data normalization, the "PCEbowPlot" function was utilized to estimate the optimal number of principal components, and graph-based clustering was executed to group cells. This method primarily entailed generating a nearest-neighbor matrix followed by employing a community search algorithm to identify the most suitable clusters. The UMAP dimensionality reduction technique was employed for data visualization, while the "FindAllMarkers" function was used to identify genes specifically expressed within each cluster and annotate cell types.

Gene set variation analysis (GSVA) and gene set enrichment analysis (GSEA) of IMPC

GSVA (version 1.32.0) was used to evaluate the activation of classical pathways and metabolism-related pathways between different subsets of cells, and the list of pathway genes was obtained from classical literature and the Molecular Signatures Database. Genes were clustered by multiplying the average log-fold change (avg_logFC) of differentially expressed genes by the logarithm base 10 of the *p*-value. The "FindAllMarkers" function in Seurat was utilized to identify genes specifically expressed in each subgroup within single-cell RNA-seq data. Furthermore, the specific expression genes of IMPC cells and IDC-NOS cells were compared based on bulk RNA-seq outcomes. Gene set enrichment analyses were conducted through GO and KEGG analyses. To elucidate pathways showing significant enrichment during the transition from the primitive stage to the primary stage, GSEA, and KEGG gene cluster enrichment analyses were performed. Gene sets with *p*-values below 0.05 were deemed enriched.

Pseudotiming analysis of IMPC

Monocle3 was employed to construct single-cell pseudotime trajectories illustrating cell differentiation. The differentially expressed genes between clusters 3, 4, and 6, as well as clusters 3 and 6, were identified using the "DifferentialGeneTest" function, establishing the initial point of the pseudospatial trajectory via the "OrderCells" function. Subsequently, dimensionality reduction of the data was carried out using the "DDRTree" function, followed by data visualization utilizing the "Plot_Cell_Trajectory" tool. Differentially expressed genes across pseudotime were determined through the "DifferentialGeneTest" function.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were conducted using R (http://www.r-project.org) and SPSS version 19.0 (SPSS Inc., Chicago, IL, USA). To assess the prognostic significance of specific gene overexpression in the IMPC group, METABRIC expression data and clinical information sourced from the cBioPortal website were obtained. Patients were stratified into high and low expression groups based on gene expression levels. Clinicopathological characteristics were compared utilizing the Mann-Whitney U-test, while correlations were examined using the Spearman rank test. Overall Survival (OS) and Disease-Free Survival (DFS) durations were calculated from the date of surgery to death and locoregional recurrence and/or distant metastasis, respectively. Follow-up durations ranged between 2 and 120 months. OS and DFS curves were generated using the Kaplan-Meier method, and differences between the curves were assessed via the log rank test. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$ and n.s., non-significant.