Journal of
Translational Medicine

RESEARCH Open Access

Single-cell transcriptome analysis reveals immune microenvironment changes and insights into the transition from DCIS to IDC with associated prognostic genes

Yidi Sun[1], Zhuoyu Pan[2], Ziyi Wang[1], Haofei Wang[1], Leyi Wei[3,4], Feifei Cui[1*], Quan Zou[5,6*] and Zilong Zhang[1*]

Abstract

Background Ductal carcinoma in situ (DCIS) of the breast is an early stage of breast cancer, and preventing its progression to invasive ductal carcinoma (IDC) is crucial for the early detection and treatment of breast cancer. Although single-cell transcriptome analysis technology has been widely used in breast cancer research, the biological mechanisms underlying the transition from DCIS to IDC remain poorly understood.

Results We identified eight cell types through cell annotation, finding significant differences in T cell proportions between DCIS and IDC. Using this as a basis, we performed pseudotime analysis on T cell subpopulations, revealing that differentially expressed genes primarily regulate immune cell migration and modulation. By intersecting WGCNA results of T cells highly correlated with the subtypes and the differentially expressed genes, we identified six key genes: FGFBP2, GNLY, KLRD1, TYROBP, PRF1, and NKG7. Excluding PRF1, the other five genes were significantly associated with overall survival in breast cancer, highlighting their potential as prognostic biomarkers.

Conclusions We identified immune cells that may play a role in the progression from DCIS to IDC and uncovered five key genes that can serve as prognostic markers for breast cancer. These findings provide insights into the mechanisms underlying the transition from DCIS to IDC, offering valuable perspectives for future research. Additionally, our results contribute to a better understanding of the biological processes involved in breast cancer progression.

Keywords Single-cell transcriptomics, T cells, DCIS, IDC

*Correspondence:
Feifei Cui
feifeicui@hainanu.edu.cn
Quan Zou
zouquan@nclab.net
Zilong Zhang
zhangzilong@hainanu.edu.cn
[1] School of Computer Science and Technology, Hainan University, Haikou 570228, China
[2] International Business School, Hainan University, Haikou 570228, China
[3] Centre for Artificial Intelligence driven Drug Discovery, Faculty of Applied Science, Macao Polytechnic University, Macao SAR, China
[4] School of Informatics, Xiamen University, Xiamen, China
[5] Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu 610054, China
[6] Yangtze Delta Region Institute (Quzhou), University of Electronic Science and Technology of China, Quzhou 324000, China

Introduction

Breast cancer is one of the most common malignant tumors affecting women's health [1, 2]. Ductal carcinoma in situ of the breast (DCIS) is a type of breast cancer that occurs due to the proliferation of malignant cells within the breast ducts. In DCIS, the cancer cells are confined to the ducts, whereas in invasive ductal carcinoma (IDC), the cancer cells have penetrated the ductal basement membrane and invaded the surrounding breast tissue [3–6]. DCIS is considered a precursor lesion to IDC [7, 8], some researchers have proposed that the progression from DCIS to IDC may serve as a model and cascade for the development of invasive ductal carcinoma in humans [9–12]. Therefore, it is crucial to study the differences in gene expression and changes in the immune microenvironment between DCIS and IDC. Studies have shown that the loss of FAT1 and β-catenin is associated with the progression from DCIS to IDC and poor prognosis. The expression of FAT1, either alone or in combination with β-catenin, may serve as a biomarker for predicting the prognosis of breast cancer patients [13]. Compared to DCIS, IDC exhibits higher levels of HER2 expression, making it more prone to develop into metastatic disease and impacting the disease free survival (DFS) of patients [14–16].

Research on single-cell transcriptomics in breast cancer has rapidly expanded, unveiling the complex cellular dynamics within tumors [17]. In 2017, Chung et al. employed single-cell transcriptomics to dissect the intricate interplay between heterogeneous tumor cells, stromal cells, and immune cells. Their analysis at single-cell resolution allowed for precise clustering of cancerous and non-cancerous cells, highlighting how interactions between tumor and immune cells drive the significant intratumoral heterogeneity observed in breast cancer [18]. Building on this, Nguyen et al. in 2018 analyzed single-cell transcriptomes of human breast cancer epithelial cells, identifying three distinct epithelial subtypes and laying the groundwork for understanding the systemic alterations that occur during breast cancer progression [19]. Bartoschek et al. further leveraged single-cell data to classify tumor-associated fibroblasts (CAFs) into three unique subpopulations, each with distinct gene expression profiles and functional roles, opening new possibilities for targeted therapies against these fibroblast subtypes [20]. Savas et al. explored the role of tumor-infiltrating lymphocytes in triple-negative breast cancer by analyzing single-cell transcriptomes of T cells, linking immune cell presence to patient prognosis and providing fresh perspectives on effective treatment strategies [21]. In 2021, Xu et al. mapped the single-cell transcriptomic landscapes of both primary and lymph node metastatic breast cancer, offering novel insights into metastatic mechanisms and inspiring new approaches to inhibit the spread of breast cancer [22]. Davis et al. highlighted the critical role of oxidative phosphorylation in preventing metastasis by creating predictive models based on single-cell transcriptomics [23–26]. Despite these significant advancements, there remains a notable gap in analyzing single-cell transcriptomic data specifically for DCIS and IDC. Understanding the differences between these stages, as well as the underlying mechanisms and biological significance of their progression, has received limited attention, underscoring the need for deeper investigation in this area.

In this study, we focused on exploring the differences between DCIS and IDC using single-cell transcriptome data from the Gene Expression Omnibus (GEO) database. We investigated changes in various immune pathways and identified several key genes. Our findings suggest that these genes may play crucial roles in the progression of breast cancer and could serve as potential prognostic factors.

Results

The landscape of single-cell transcriptomics in breast cancer

In this study, we analyzed 13 datasets from the GSE195861 dataset, which includes 7 samples of DCIS and 6 samples of IDC. The overall workflow of this work is illustrated in Fig. 1. Each sample underwent rigorous quality control, including the removal of mitochondrial genes and outliers, cell cycle correction, batch effect removal, and data normalization. After these processes, a total of 19,474 cells were retained for further analysis. Dimensionality reduction was performed using the PCA method, retaining the top 50 principal components [27]. Using the KNN method to construct an adjacency matrix, followed by clustering with the Louvain method, the 19,474 cells were divided into 29 clusters. These clusters were visualized using t-SNE, and Fig. 2A shows the single-cell transcriptomic landscape for the two different subgroups, while Fig. 2B depicts the single-cell transcriptomic atlas for all 13 samples. Subsequently, we annotated the clustering results for further analysis using the HumanPrimaryCellAtlasData dataset from singleR as the reference (Fig. 2C, D). We identified eight distinct cell types: 4231 macrophages, 5239 T cells, 7003 epithelial cells, 1597 B cells, 745 monocytes, 292 fibroblasts, 286 natural killer cells, and 81 neutrophils (Fig. 2E). Finally, we observed a considerable difference in the proportion of T cells between the two subgroups, as shown in the cell proportion plots (Fig. 2F, G).

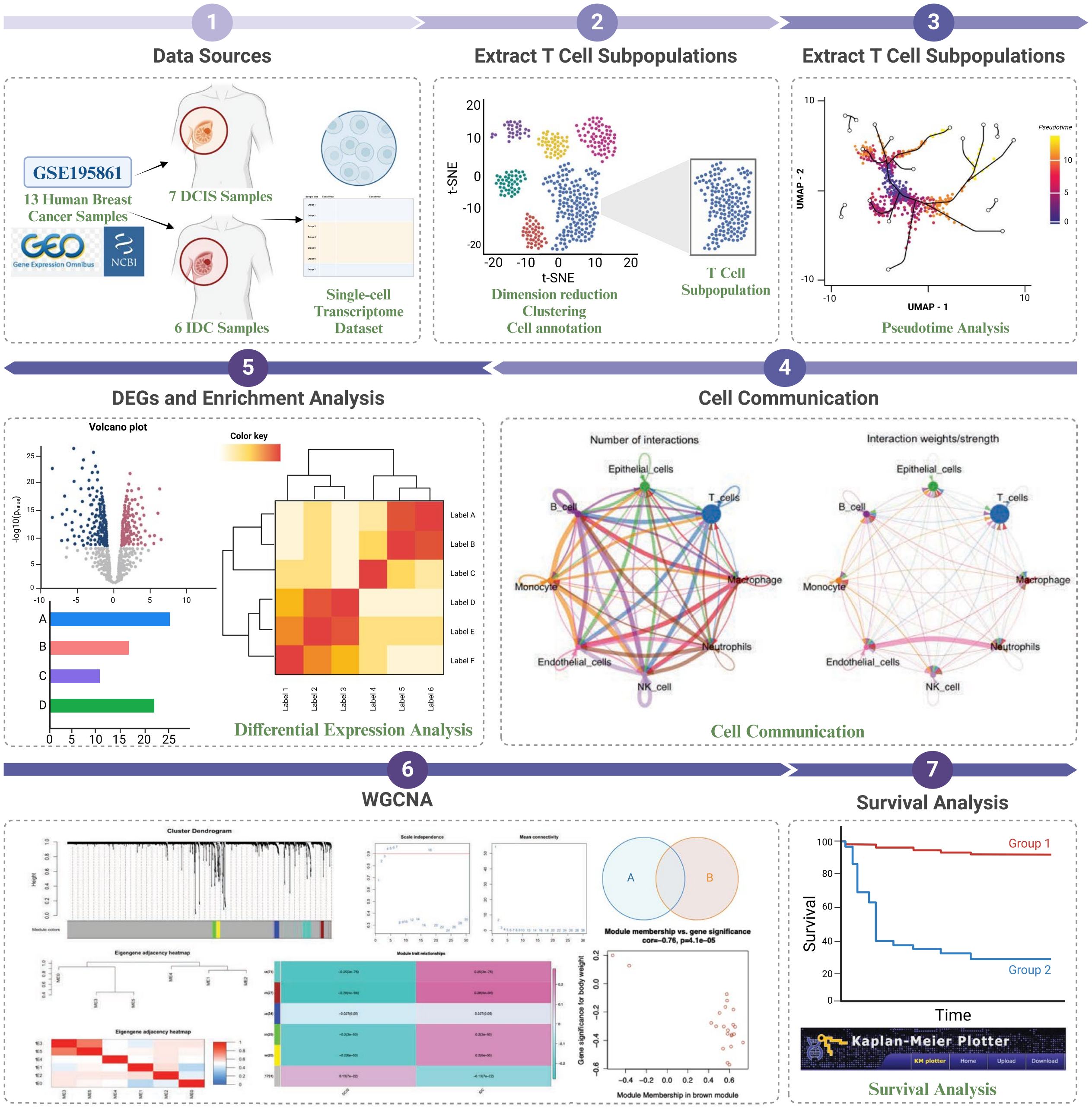

Fig. 1 The workflow of the present study

Cell–cell communication

To understand the relationships among different cell type subpopulations, we performed cell–cell communication analysis on the annotated data. Here, focusing on the IDC subgroup, we identified extensive communication among the eight cell types (Fig. 3A, B). Among all the interactions, we found that the interactions between fibroblasts and neutrophils were the strongest (Fig. 3C). However, this interaction pathway was not enriched in the DCIS subgroup. Previous studies have shown that the interaction between fibroblasts and neutrophils is often linked to inflammatory and immune responses in the human body [28–31], which aligns with the idea that changes in cellular interactions may occur as DCIS progresses to IDC. We also analyzed all possible ligand-receptor pairs mediating interactions between T cells and other cell types (Fig. 3D), identifying the key pairs involved in interactions between T cells and fibroblasts, natural killer

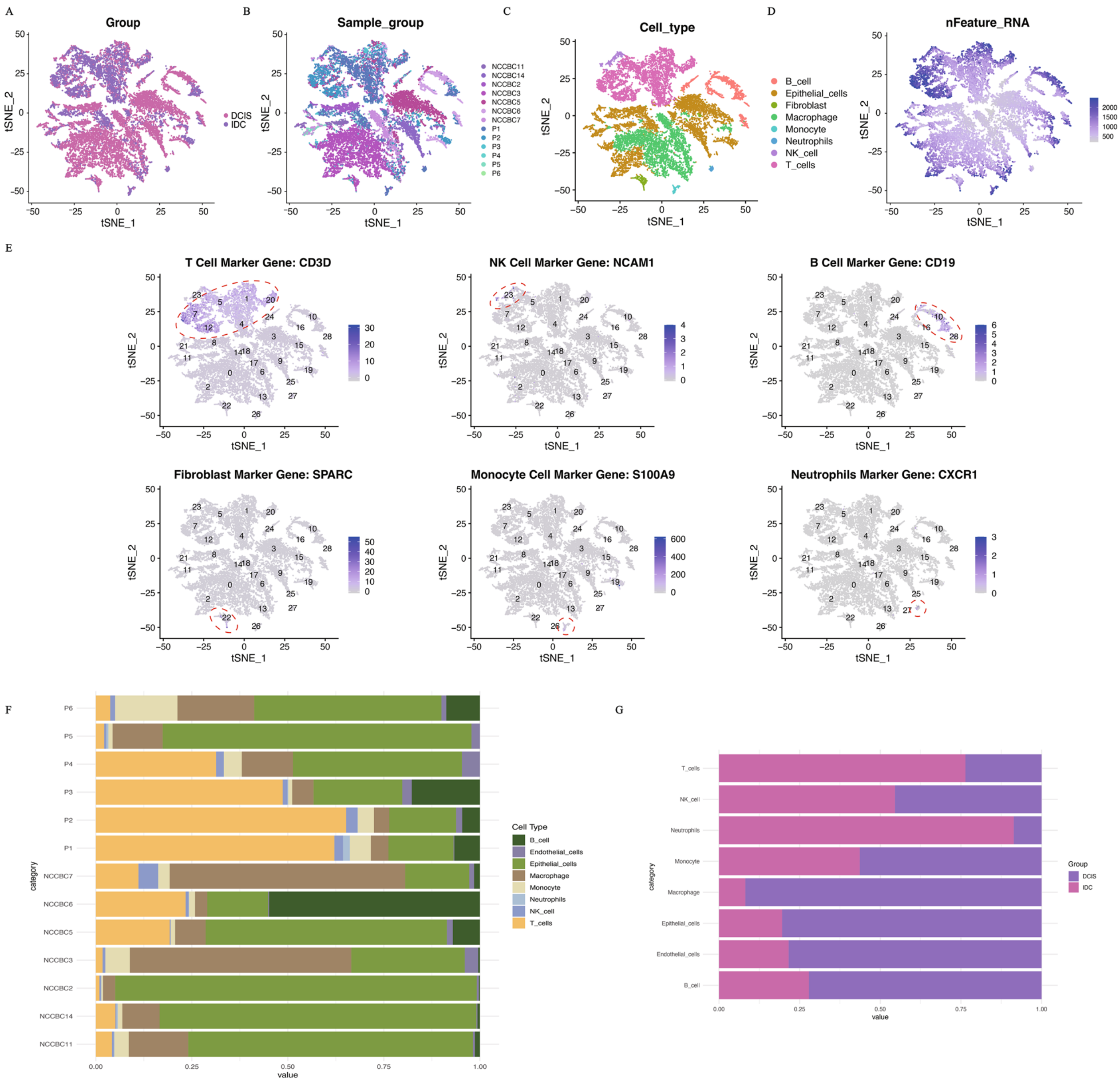

Fig. 2 Overview of the 19,474 single cells from DCIS and IDC samples. **A** The t-SNE visualization results display the distribution of single-cell transcriptomic data grouped into DCIS and IDC. **B** The t-SNE visualization results display the distribution of single-cell transcriptomic data grouped by different samples (specifically, 7 DCIS samples starting with NCCBC and 6 IDC samples starting with P). **C** The t-SNE visualization results display the distribution of single-cell transcriptomic data grouped by cell type. **D** The t-SNE visualization results display the distribution of the number of transcripts (UMIs) detected in the single-cell transcriptomic data. **E** Expression of marker genes for the cell types. **F** Proportional representation of different cell types by 13 samples. **G** Proportional representation of different cell types by two breast cancer types (DCIS or IDC)

cells, and neutrophils. Among these, the top four ligand-receptor pairs mediating interactions between T cells and fibroblasts or natural killer cells THBS1-SDC1, COL2A1-SDC1, COL4A2-SDC1, and TNC-SDC1 primarily play roles in cell adhesion and migration. Additionally, the two ligand-receptor pairs that primarily mediate interactions between T cells and neutrophils THBS4-CD36 and THBS1-CD36 are mainly involved in inflammatory responses and cell adhesion. Figure 3E, F illustrate the top four signaling pathways involved in interactions between T cells and other cell types. It is evident that the eight different cell types play distinct roles within these pathways. As shown in the clustering tree in Fig. 3G, these cell types work collaboratively to exert their effects.

Pseudo temporal analysis of T cell

Temporal analysis can simulate the developmental trajectory of cells. In this study, we isolated single-cell transcriptomic data from T cells and standardized the data using the monocle3 package for dimensionality reduction. Subsequently, we visualized the reduced-dimensional data using the t-SNE method. Figure 4A illustrates the expression of six key genes in T cells, indicating their presence across T cell populations. To further delineate trajectories, we examined the distribution of different subgroups within T cell populations and the distribution of cells across various cell cycles while correcting for batch effects (Fig. 4C, D). Subsequently, T cells were clustered into nine clusters, and different clusters were labeled accordingly (Fig. 4E). Figure 4B depicts the expression of six key genes in the quality-controlled data, demonstrating their uniform expression across T cell populations. Figure 4F presents a bubble plot of marker genes in the nine different clusters, while Fig. 4G displays the marker genes for each cluster. In detail, SCGB2A2, a marker gene for cluster 1, is used as a biomarker for the diagnosis and prognosis of breast cancer, with elevated expression levels in metastatic breast cancer [32, 33]; CD8A, a marker gene for cluster 2, CD8A is primarily expressed on cytotoxic T cells and plays a role in the immune process by participating in the regulation of T cell antigen recognition [34–36]; GZMK, marking cluster 3, participates in the activation of immune cells and cytotoxic effects [37–39]; GZMB, marking cluster 4, regulates the cytotoxicity of immune cells against target cells [40, 41]; EEF1B2, marking cluster 5, plays a role in protein synthesis [42]; FOXP3, marking cluster 6, regulates the quantity and function of regulatory T cells, thereby impacting tumor immunity [43–46]; CXCL13, a marker gene for cluster 7, is a chemokine that can induce the migration of immune cells [47, 48]; TXNIP, a marker gene for cluster 8, acts as an endogenous inhibitor of thioredoxin, often involved in regulating cellular redox balance and influencing the status of tumor cells [49–51]; Finally, TOX2, marking cluster 9, is involved in cell proliferation, differentiation, and migration processes [52, 53].

We further subdivided the overall dataset into four partitions and identified the trajectory of T cell development (Fig. 4H, I). By defining the earliest time-point interval, we identified the root cells and obtained the developmental trajectory of T cells (Fig. 4J). Visualizing the differential genes along the trajectory at different time points, we generated violin plots for the top three significant genes (Fig. 4K). Among them, TNFRSF18 primarily plays a role in immune regulation, mainly involved in modulating T cell activity. TNFRSF4 encodes a cell membrane receptor known to modulate cell proliferation and immune responses. TXNIP is primarily involved in inflammation response and metabolic regulation in the body.

Differential expression analysis of T cell subgroups

Previous analysis has already shown a significant difference in T-cell expression between the DCIS and IDC subgroups. To explore the underlying mechanisms, we conducted differential expression analysis on single-cell transcriptomic data from the T-cell populations in the DCIS and IDC groups, resulting in a total of 234 differentially expressed genes (Supplementary Table 1). GO enrichment analysis and GSVA on these differentially expressed genes revealed that they were primarily enriched in four pathways: “Leukocyte mediated immunity,” “lymphocyte mediated immunity,” “positive regulation of leukocyte activation,” and “positive regulation of cell activation.” This suggests that during the progression of breast cancer, these differentially expressed genes may play a role in activating and regulating immune-related pathways, particularly those involved in leukocyte and lymphocyte activation (Fig. 5B).

The GSVA results indicate that, compared to the IDC subgroup, the DCIS subgroup shows significant upregulation in the “HALLMARK_INFLAMMATORY_RESPONSE” gene set, which includes genes associated with inflammatory responses and can be used to analyze inflammation-related biological processes and signaling pathways. Conversely, the DCIS subgroup shows significant downregulation in the “HALLMARK_ESTROGEN_RESPONSE_EARLY” gene set, which primarily contains genes related to the early cellular response to estrogen stimulation (Fig. 5A).

(See figure on next page.)

Fig. 3 Cellular communication network between IDC microenvironment. **A** Cellular interaction network. The size of the dots represents the number of cells, the thickness of the line represents the number of cell interactions. **B** Another presentation of cellular interaction networks. The thickness of the line represents the intensity of intercellular interaction. **C** The interaction between each type of cell and other cells. **D** Bubble diagram of all potential ligand-receptor interactions between T cells and other cell types. The x-axis shows the interacting cell types, the y-axis shows ligand-receptor pairs. Dot color indicates interaction likelihood, and dot size represents corresponding P-values. **E** Heatmap of the role of each cell type in the cell interaction network. The role is: sender, receiver, mediator, and influencer. **F** Heatmap showing the contribution of incoming and outgoing signals for each cell type. The bar at the top indicates the proportion of each cell type's ligands. **G** Heatmap of global communication mode. Color represents contribution

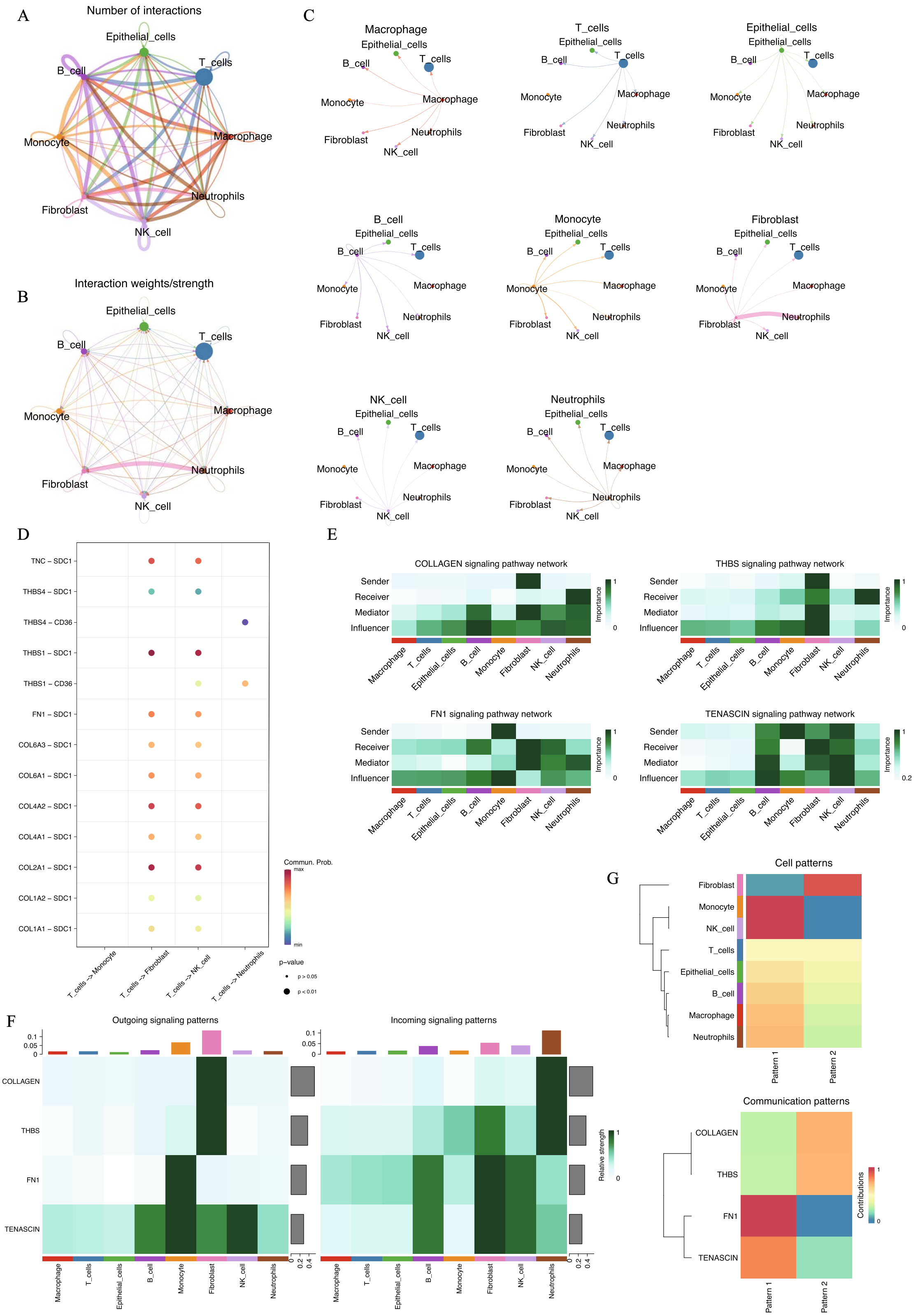

Fig. 3 (See legend on previous page.)

WGCNA

WGCNA (Weighted Gene Co-expression Network Analysis) is used to identify gene sets with the strongest correlation to a phenotype of interest. We selected a clustering number of 6 as optimal (Fig. 5C, D), then analyzed the correlation between these 6 gene modules and the two subtypes. Figure 5E shows the correlation between different modules. The results indicated that the gene set in the "brown" module had the strongest correlation with the two subtypes (Fig. 5F, Supplementary Table 2). We extracted the genes from this module for further analysis. Moreover, we validated the correlation between the genes and the subtypes, and the scatter plot results showed a strong association (Fig. 5G). Finally, we intersected the key genes from the WGCNA modules with the differentially expressed genes, resulting in 6 key genes: FGFBP2, GNLY, KLRD1, TYROBP, PRF1, and NKG7 for further analysis (Fig. 5H).

Survival analysis of key genes

As previously mentioned, the six key genes identified in our study might play a crucial role in the progression from DCIS to IDC. To evaluate the impact of these genes on breast cancer prognosis, we conducted Kaplan–Meier survival analysis. Figure 6 indicates that the expression levels of five out of the six genes are significantly associated with overall survival in patients. Specifically, low expression of FGFBP2 and TYROBP is linked to better prognosis, while high expression of GNLY, KLRD1, and NKG7 is associated with improved outcomes. These results suggest that FGFBP2, TYROBP, GNLY, KLRD1, and NKG7 could potentially serve as prognostic biomarkers in breast cancer.

Discussion

Currently, despite extensive research into the genes involved in the immune microenvironment of breast cancer [54–56], there have been no reports focusing on the differences and changes between the two subtypes, DCIS and IDC.

Here, we preprocessed and annotated single-cell transcriptome data from DCIS and IDC, identifying eight cell types. We found that T cells and neutrophils showed the most significant proportional differences between the two subtypes. Related immunofluorescence experiments indicated that various cell types, including T cells and neutrophils, exhibit significant tissue- and breast cancer subtype-specific differences. For instance, activated GZMB + CD8 + T cells are less prevalent in IDC than in DCIS, and T cell receptor clonality is significantly higher in DCIS compared to IDC. T cells expressing the immune checkpoint protein TIGIT are more frequently observed in DCIS, while high PD-L1 expression and amplification of CD274 (encoding PD-L1) are detected only in triple-negative IDC [15]. However, T cells are more numerous in IDC [57], which aligns with our findings of a higher proportion of T cells in IDC. We infer that this may be due to increased T cell presence in IDC, but their activity might be suppressed.

The migration of fibroblasts is related to the movement of cancer cells from the primary site to specific metastatic locations. Many inflammatory chemokines can regulate the migration of tumor-associated neutrophils (TANs) and various other immune cells [58, 59]. In this study, our cell communication analysis revealed a significant interaction between fibroblasts and neutrophils in IDC, which was not observed in DCIS. This finding further indicates that the progression of DCIS to IDC is closely associated with the interactions between fibroblasts and neutrophils.

Notably, in the T cell subpopulations, the differentially expressed genes between DCIS and IDC are primarily enriched in pathways related to the activation of immune cells, including leukocytes and lymphocytes. GSEA results indicate that these genes are associated with biological processes and signaling pathways related to inflammatory responses, as well as early estrogen responses. This further supports the aforementioned conclusions.

Through the intersection of WGCNA and differential expression analysis results, we identified six key genes that may play important roles in the progression from DCIS to IDC. Survival analysis revealed that the expression differences of five of these genes—FGFBP2, TYROBP, GNLY, KLRD1, and NKG7—are significantly associated with breast cancer prognosis. Previous studies

(See figure on next page.)

Fig. 4 Pseudo temporal analysis of T cell clustering. **A** Expression distribution of 6 key genes in untreated T cell clusters using t-SNE clustering. **B** Expression distribution of 6 key genes in t-SNE clustering in T cell clusters after quality control. **C** t-SNE examination of the distribution of two different subgroups in T cell single-cell data, excluding batch effects. **D** Cluster based on different cell cycles after removing batch effects of subtypes. **E** The T cell data after quality control is clustered into 9 different subgroups. **F** Bubble plots of marker gene expression in 9 different subgroups. **G** Expression of marker genes for 9 different subgroups. **H** Grouping according to developmental sequence. **I** Development trajectory recognition. **J** Constructing temporal developmental trajectories after defining root cells. **K** Violin plots of the first three differentially expressed genes at different time points

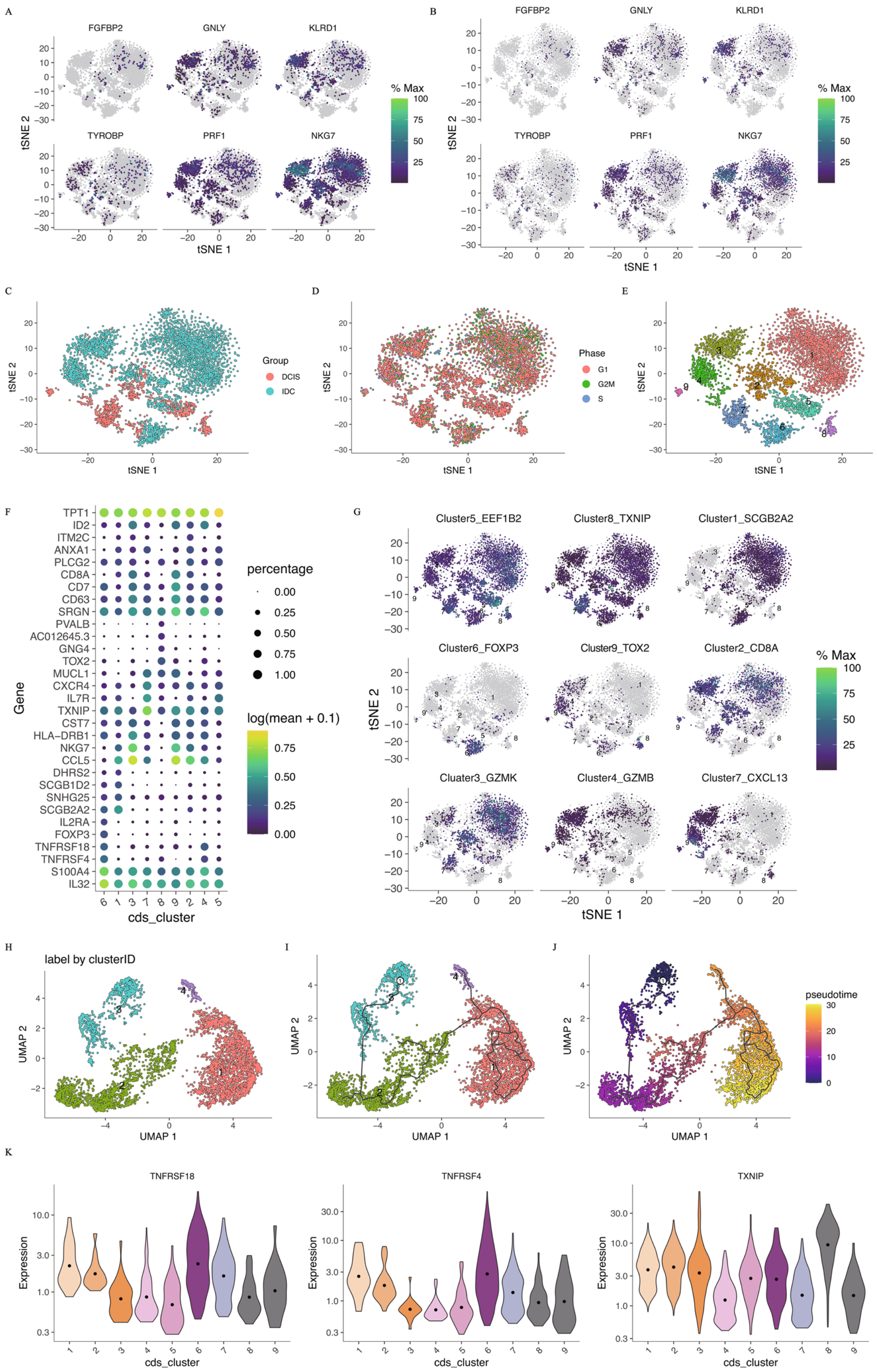

Fig. 4 (See legend on previous page.)

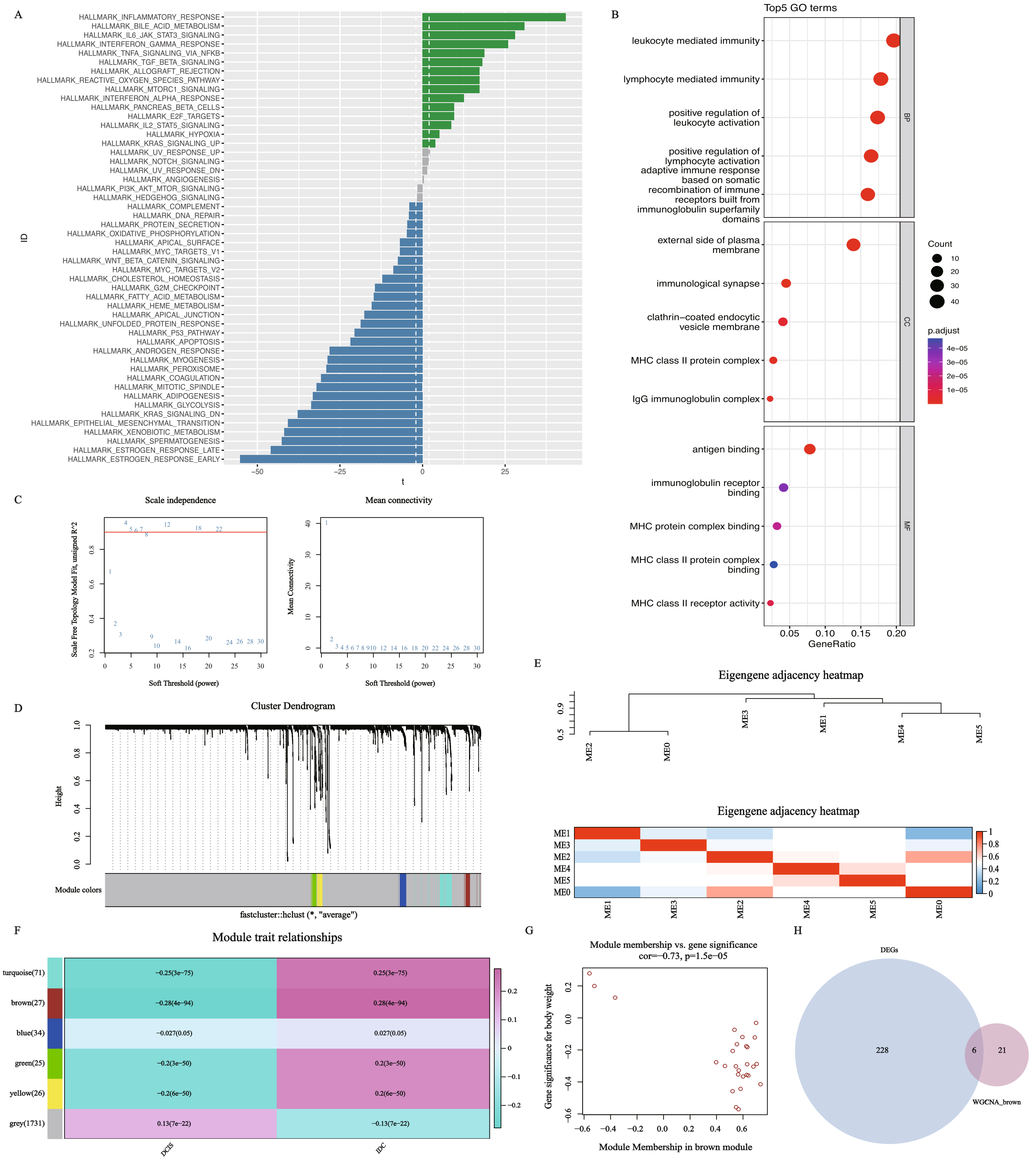

Fig. 5 Differential expression analysis and presentation of WGCNA results. **A** Differences in pathway activities scored per cell by GSVA between DCIS and IDC of T cells (DCIS = 1237 and IDC = 4002 cells). Shown are t values from a linear model, corrected for patient of IDC. **B** Bubble plot depicting GO enrichment results of DEGs in T cells. **C** WGCNA searching for the best soft threshold. **D** Construct a co expression network based on the optimal soft threshold to draw a gene clustering tree, with each color representing a module. **E** Heatmap of correlation between different modules. **F** Heatmap of correlation and significance between different modules and features (DCIS and IDC). **G** Scatter plot of gene phenotype correlation in brown module. **H** Venn diagram shows the intersection of 234 differentially expressed genes and 27 brown modules genes

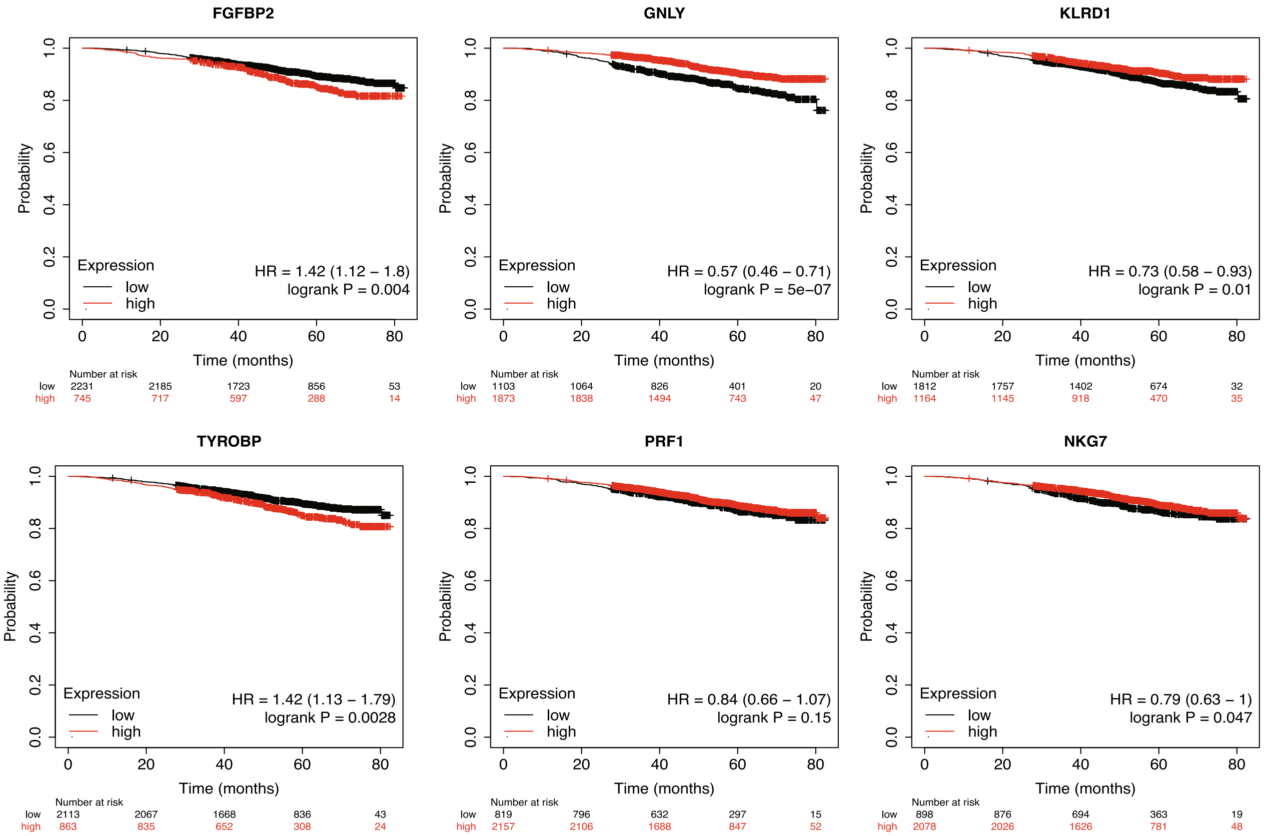

Fig. 6 K-M survival analysis of 6 key genes. Divide high and low expression groups based on the median expression of each gene, red represents the high expression group, black represents the low expression group

have shown that FGFBP2 is expressed at higher levels in IDC compared to DCIS [60], the role of the TYROBP gene in the progression from DCIS to IDC is rarely reported. However, there are reports indicating that high expression of TYROBP is associated with bone metastasis and poor prognosis in breast cancer [61, 62]. This aligns with our study's findings that low expression of FGFBP2 and TYROBP is associated with a better prognosis in breast cancer. In a 2023 report, the authors constructed a model using the genes IDO1, GNLY, IRF1, CTLA4, and CXCR6 to predict the prognosis of triple-negative breast cancer [63]. Another study reported that GNLY expression is significantly associated with the prognosis of primary breast cancer [64], Cai et al. developed the cr-TILCD8TSig tool, which includes KLRD1 among seven genes, to provide independent prognostic assessment for breast cancer [65]. Additionally, NKG7 has been identified as an intrinsic therapeutic target in T cells for enhancing antitumor cytotoxicity and cancer immunotherapy [66]. In our findings, high expression levels of GNLY, KLRD1, and NKG7 are significantly associated with a favorable prognosis in breast cancer. These five genes could serve as potential prognostic biomarkers for breast cancer. Notably, among the six key genes we identified, PRF1 did not exhibit significant prognostic value. However, related literature has demonstrated its crucial role in mediating cytotoxicity and its association with inflammatory responses [67].

At the same time, our study has some limitations. First, we only used the DCIS and IDC data from the GSE195861 dataset for single cell transcriptome analysis. While we conducted a comparative analysis between the two breast cancer subtypes, further validation with additional single cell data from both DCIS and IDC will be necessary. Secondly, our study relied solely on transcriptome data; future research could benefit from integrating multi-omics data for a more comprehensive exploration.

Conclusions

Through our research on the two breast cancer subtypes, DCIS and IDC, we identified six key genes—FGFBP2, GNLY, KLRD1, TYROBP, PRF1, and NKG7—that are significantly differentially expressed between the two. We demonstrated that five of these genes are significantly associated with breast cancer survival, suggesting they may serve as potential prognostic markers. Our study partially elucidates the mechanisms underlying the progression from DCIS to IDC, providing valuable insights for researchers in the field.

Materials and methods

Data source

The single-cell transcriptomic data originates from the GSE195861 dataset in the GEO database, corresponding to the published article "Single-Cell Transcriptome Profiling Reveals Intratumoral Heterogeneity and Molecular Features of Ductal Carcinoma In Situ" [68]. In this study, 13 breast cancer samples were selected from the GSE195861 dataset as the raw data, which including 7 samples of DCIS and 6 samples of IDC. Batch reading of these 13 samples was conducted using the R package Seurat (version: 4.3.0) to construct Seurat objects and append grouping information for each sample.

Quality control of single-cell transcriptomic data

The raw data was preprocessed using the Seurat package, excluding samples with fewer than 200 and more than 2500 transcripts per single cell and samples with mitochondrial gene percentages exceeding ten percent. Using the "CellCycleScoring" function from the Seurat package to perform cell cycle scoring and mitigate the influence of the cell cycle on single-cell transcriptomic data; utilizing the "NormalizeData" function from the Seurat package to perform logarithmic transformation and normalization of the data using the LogNormalize method; Using the "FindVariableFeatures" function to identify highly variable genes, applying variance stabilizing transformation (vst) method to stabilize variance and retain the top 1000 genes with the highest variability, and finally using the "ScaleData" function to remove batch effects between samples.

Dimensionality reduction, clustering, and cell annotation of single-cell data

After data preprocessing, it is necessary to reduce the dimensionality of high-dimensional single-cell transcriptomic data for further analysis [69, 70]. Principal component analysis (PCA) is currently the most commonly used method for dimensionality reduction. Using the "RunPCA" function from the Seurat package to perform dimensionality reduction on the data and retain the top 50 principal components. Then, using the "FindNeighbors" function to identify K-nearest neighbor (KNN) relationships between cells in the single-cell transcriptomic data, and using the Louvain method with the "FindClusters" function to implement clustering of cells. The SingleR package provides an objective identification of cell types and integrates well with the Seurat package [71]. The celldex package allows for direct retrieval and access to seven built-in datasets from the SingleR package, download the HumanPrimaryCellAtlasData dataset for cell annotation and export the cell annotation results [72]. Following the aforementioned processing steps, a Seurat object containing 2000 genes and 19474 cells was obtained for subsequent analysis.

Cell communication

Cell–cell communication analysis was conducted using the CellChat package, with the patchwork package used for plot arrangement [73, 74]. The ligand-receptor interaction database was set to "CellChatDB.human." The "computeCommunProb" function was used to calculate the probability of communication between cells, while the "filterCommunication" function was used to filter out communication data observed in fewer than 10 cells. The "subsetCommunication" function was applied to infer the CellChat network, and the "computeCommunProbPathway" function was used to infer the probability of cell–cell communication at the signaling pathway level.

Pseudotime analysis

To track the developmental trajectory of T cell populations, we conducted pseudotime analysis on T cells using the Monocle3 package. We constructed a CellDataSet (CDS) object with the "new_cell_data_set" function, standardized the CDS object using a log transformation, and reduced dimensionality with the "PCA" method. We removed batch effects with the "align_cds" function, then used t-Distributed Stochastic Neighbor Embedding (t-SNE) for dimensionality reduction and the "leiden" method for clustering. Afterward, we used the "top_markers" function to identify the most significant marker genes. Finally, we defined a function to choose the earliest time interval and visualized the results using the Uniform Manifold Approximation and Projection (UMAP) method of the "plot_cells" function.

Differential expression analysis of T cells across different groups

Use the "subset" function to extract the T-cell cluster and define the DCIS and IDC groups. Then, use the "FindMarkers" function to set the two groups as two separate idents. For differential expression analysis, choose MAST, a common method for single-cell data analysis. Additionally, filter out genes detected in fewer than 50% of the cells and those with a fold change less than 2 between groups. The final result is 234 differentially expressed genes.

Enrichment analysis

Gene Set Variation Analysis (GSVA) enrichment analysis can reveal the enrichment levels of gene sets in different biological pathways [75–78]. In this study, we used the GSVA method to analyze the biological significance of differentially expressed gene sets. We downloaded the "h.all.v7.0.symbols.gmt" dataset from the Molecular signatures database (MSigDB) (https://www.gsea-msigdb.org/gsea/msigdb/human/collections.jsp) as a standard dataset for enrichment analysis. The "gsva" function was used to perform gene set variation analysis, and we visualized the results using the "pheatmap" and "ggplot" functions. Gene Ontology (GO) enrichment analysis helps in understanding the enrichment of gene sets in specific biological processes, molecular functions, and cellular components [79]. Enrichment analysis can be conducted using the "enrichGO" function from the clusterProfiler package.

WGCNA

Weighted gene co-expression network analysis (WGCNA) enables the construction of gene co-expression networks. In this study, standardized data were analyzed using the built-in R function "hclust" with the "average" method specified as the clustering method to identify and remove outlier samples. The "pickSoftThreshold" function was employed to compute the optimal power value and visualize it. Subsequently, the "blockwiseModules" function was used to construct the network and identify highly correlated gene expression modules. Relevant network parameters were extracted, and a correlation plot between modules was drawn based on gene expression levels. The "moduleEigengenes" function was utilized to calculate the eigengenes of gene modules, and a relationship matrix between modules and samples was extracted. Correlation analysis between modules and phenotypes was performed using the "cor" function from the WGCNA package, and p-values were calculated. The correlation between genes and modules was also computed. Finally, the "verboseScatterplot" function was used to generate scatterplots illustrating the correlation between genes and phenotypes.

Survival analysis

Kaplan–Meier Plotter (https://kmplot.com/analysis/) is an online analysis tool frequently used in bioinformatics analyses. In this study, Kaplan–Meier Plotter was used to create survival curves, with samples grouped by the median expression level of each gene.

Statistical analysis

The data processing in this study was conducted using R software (version: 4.2.2) (https://www.r-project.org/). Student's t-test was employed for significance analysis. Pearson correlation was used for correlation analysis. Kaplan–Meier analysis was utilized to assess survival differences between high and low expression groups of genes, with statistical significance defined as *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and $p < 0.05$ considered statistically significant. Visualize the t-Distributed Stochastic Neighbor Embedding (t-SNE) dimensionality reduction results using the "DimPlot" and "FeaturePlot" functions from the Seurat package, and create a cell proportion plot using the ggplot2 package.

Abbreviations

DCIS	Ductal carcinoma in situ
IDC	Invasive ductal carcinoma
DFS	Disease-free survival
vst	Variance stabilizing transformation
PCA	Principal component analysis
KNN	K-nearest neighbors
CDS	CellDataSet
MSigDB	Molecular signatures database
WGCNA	Weighted gene co-expression network analysis
TANs	Tumor-associated neutrophils
GEO	Gene expression omnibus
GO	Gene ontology
GSVA	Gene Set variation analysis
UMAP	Uniform manifold approximation and projection
t-SNE	T-distributed stochastic neighbor embedding

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12967-024-05706-6.

Additional file 1: Figure R1. Kaplan-Meier Survival Curve of the Cumulative Expression of Six Key Genes. High cumulative expression of the six key genes is significantly associated with better prognosis compared to low expression

Additional file 2: Figure R2. Immune Infiltration Status of the GSE66301 Dataset.Boxplot of Infiltration Levels of 11 Immune Cell Types in DCIS. Boxplot of Infiltration Levels of 12 Immune Cell Types in IDC

Additional file 3: Figure R3. Clustering and cell annotation results of single-cell transcriptomics.The clustering analysis divided the single-cell transcriptomics data into 29 distinct clusters.The single-cell transcriptomics data were annotated into eight different cell types: B cells, Epithelial cells, Fibroblast, Macrophage, Monocyte, Neutrophils, NK cells, and T cells

Additional file 4: Figure R4. The expression distribution of four T cell-specific marker genes, CD3D, CD3E, CD3G, and CD2, in the single-cell transcriptomics data

Additional file 5: Table 1. Differential expression genes of T cell subpopulations DCIS and IDC

Additional file 6: Table 2. The genes with the strongest correlation with DCIS and IDC subtypes in WGCNA results

Acknowledgements

Not applicable.

Author contributions

Y.S. analyzed the data and explained the results obtained, was a major contributor in writing the manuscript. Z.P. organized the research data. Z.W. corrected the content of the manuscript. H.W. verified the research method. L.W. collected research resources. F.C. provided research funding and revised the manuscript. Q.Z. verified the research results. Z.Z. provided research funding, conceived the idea, reviewed and revised the manuscript. All authors read and approved the final manuscript.

Funding
The work was supported by the National Natural Science Foundation of China (No. 62102064, No. 62261018, 62262015), Science and Technology special fund of Hainan Province (ZDYF2024GXJS01), and Hainan Provincial Natural Science Foundation of China (324MS009).

Availability of data and materials
The datasets analysed during the current study are available in the GEO repository, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE195861/.

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.

Received: 27 June 2024 Accepted: 25 September 2024
Published online: 03 October 2024

References

1. Obeagu EI, Babar Q, Vincent C, Udenze CL, Eze R, Okafor CJ, Ifionu BI, Amaeze AA, Amaeze FN. Therapeutic targets in breast cancer signaling: a review. J Pharm Res Int. 2021;33(56A):82–99.
2. Aizaz M, Khan M, Khan F, Munir A, Ahmad S, Obeagu E. Burden of breast cancer: developing countries perspective. Int J Innov Appl Res. 2023;11(1):31–7.
3. Burstein HJ, Polyak K, Wong JS, Lester SC, Kaelin CM. Ductal carcinoma in situ of the breast. N Engl J Med. 2004;350(14):1430–41.
4. Sgroi DC. Preinvasive breast cancer. Ann Rev Pathol: Mechan Dis. 2010;5:193–221.
5. Gradishar WJ, Anderson BO, Balassanian R, Blair SL, Burstein HJ, Cyr A, Elias AD, Farrar WB, Forero A, Giordano SH, et al. Invasive breast cancer version 1.2016, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw. 2016;14(3):324–54.
6. Sun L, Qiu Y, Ching W-K, Zhao P, Zou Q. PCB: a pseudotemporal causality-based Bayesian approach to identify EMT-associated regulatory relationships of AS events and RBPs during breast cancer progression. PLoS Comput Biol. 2023;19(3):e1010939.
7. Kole AJ, Park HS, Johnson SB, Kelly JR, Moran MS, Patel AA. Overall survival is improved when DCIS accompanies invasive breast cancer. Sci Rep. 2019;9(1):9934.
8. Hanley K, Wang J, Bourne P, Yang Q, Gao AC, Lyman G, Tang P. Lack of expression of androgen receptor may play a critical role in transformation from in situ to invasive basal subtype of high-grade ductal carcinoma of the breast. Hum Pathol. 2008;39(3):386–92.
9. London SJ, Connolly JL, Schnitt SJ, Colditz GA. A prospective study of benign breast disease and the risk of breast cancer. JAMA. 1992;267(7):941–4.
10. Miki Y, Suzuki T, Sasano H. Intracrinology of sex steroids in ductal carcinoma in situ (DCIS) of human breast: comparison to invasive ductal carcinoma (IDC) and non-neoplastic breast. J Steroid Biochem Mol Biol. 2009;114(1):68–71.
11. Li L, Algabri YA, Liu Z-P. Identifying diagnostic biomarkers of breast cancer based on gene expression data and ensemble feature selection. Curr Bioinform. 2023;18(3):232–46.
12. Su R, Liu X, Wei L. MinE-RFE: determine the optimal subset from RFE by minimizing the subset-accuracy-defined energy. Brief Bioinform. 2020;21(2):687–98.
13. Wang L, Lyu S, Wang S, Shen H, Niu F, Liu X, Liu J, Niu Y. Loss of FAT1 during the progression from DCIS to IDC and predict poor clinical outcome in breast cancer. Exp Mol Pathol. 2016;100(1):177–83.
14. Goh CW, Wu J, Ding S, Lin C, Chen X, Huang O, Chen W, Li Y, Shen K, Zhu L. Invasive ductal carcinoma with coexisting ductal carcinoma in situ (IDC/DCIS) versus pure invasive ductal carcinoma (IDC): a comparison of clinicopathological characteristics, molecular subtypes, and clinical outcomes. J Cancer Res Clin Oncol. 2019;145(7):1877–86.
15. Gil Del Alcazar CR, Huh SJ, Ekram MB, Trinh A, Liu LL, Beca F, Zi X, Kwak M, Bergholtz H, Su Y, et al. Immune escape in breast cancer during in situ to invasive carcinoma transition. Cancer Discov. 2017;7(10):1098–115.
16. Dai C, Jiang Y, Yin C, Su R, Zeng X, Zou Q, Nakai K, Wei L. scIMC: a platform for benchmarking comparison and visualization analysis of scRNA-seq data imputation methods. Nucleic Acids Res. 2022;50(9):4877–99.
17. Duan H, Zhang Y, Qiu H, Fu X, Liu C, Zang X, Xu A, Wu Z, Li X, Zhang Q, et al. Machine learning-based prediction model for distant metastasis of breast cancer. Comput Biol Med. 2024;169:107943.
18. Chung W, Eum HH, Lee H-O, Lee K-M, Lee H-B, Kim K-T, Ryu HS, Kim S, Lee JE, Park YH, et al. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. Nat Commun. 2017;8(1):15081.
19. Nguyen QH, Pervolarakis N, Blake K, Ma D, Davis RT, James N, Phung AT, Willey E, Kumar R, Jabart E. Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity. Nat Commun. 2018;9(1):2028.
20. Bartoschek M, Oskolkov N, Bocci M, Lövrot J, Larsson C, Sommarin M, Madsen CD, Lindgren D, Pekar G, Karlsson G. Spatially and functionally distinct subclasses of breast cancer-associated fibroblasts revealed by single cell RNA sequencing. Nat Commun. 2018;9(1):5150.
21. Savas P, Virassamy B, Ye C, Salim A, Mintoff CP, Caramia F, Salgado R, Byrne DJ, Teo ZL, Dushyanthen S. Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis. Nat Med. 2018;24(7):986–93.
22. Xu K, Wang R, Xie H, Hu L, Wang C, Xu J, Zhu C, Liu Y, Gao F, Li X. Single-cell RNA sequencing reveals cell heterogeneity and transcriptome profile of breast cancer lymph node metastasis. Oncogenesis. 2021;10(10):66.
23. Davis RT, Blake K, Ma D, Gabra MBI, Hernandez GA, Phung AT, Yang Y, Maurer D, Lefebvre AE, Alshetaiwi H. Transcriptional diversity and bioenergetic shift in human breast cancer metastasis revealed by single-cell RNA sequencing. Nat Cell Biol. 2020;22(3):310–20.
24. Ren L, Li J, Wang C, Lou Z, Gao S, Zhao L, Wang S, Chaulagain A, Zhang M, Li X. Single cell RNA sequencing for breast cancer: present and future. Cell Death Discovery. 2021;7(1):104.
25. Wang J, Chen Y, Zou Q. Inferring gene regulatory network from single-cell transcriptomes with graph autoencoder model. PLoS Genet. 2023;19(9):e1010942.
26. Su R, Wu H, Xu B, Liu X, Wei L. Developing a multi-dose computational model for drug-induced hepatotoxicity prediction based on toxicogenomics data. IEEE/ACM Trans Comput Biol Bioinf. 2019;16(4):1231–9.
27. Sun Y, Kong L, Huang J, Deng H, Bian X, Li X, Cui F, Dou L, Cao C, Zou Q, et al. A comprehensive survey of dimensionality reduction and clustering methods for single-cell and spatial transcriptomics data. Brief Funct Genomics. 2024. https://doi.org/10.1093/bfgp/elae023.
28. Liu Y, Zhang X, Gu W, Su H, Wang X, Wang X, Zhang J, Xu M, Sheng W. Unlocking the crucial role of cancer-associated fibroblasts in tumor metastasis: mechanisms and therapeutic prospects. J Adv Res. 2024. https://doi.org/10.1016/j.jare.2024.05.031.
29. Dragoni G, Ke BJ, Picariello L, Ceni E, Mello T, Verstockt B, Abdurahiman S, Biscu F, Innocenti T, De Hertogh G, et al. P099 neutrophil extracellular traps enhance profibrotic activity of intestinal fibroblasts in Crohn's disease through TLR2/NF-kB pathway. J Crohn's Colitis. 2024;18(Supplement_1):i379.
30. Cavagnero KJ, Li F, Dokoshi T, Nakatsuji T, O'Neill AM, Aguilera C, Liu E, Shia M, Osuoji O, Hata T, et al. CXCL12+ dermal fibroblasts promote neutrophil recruitment and host defense by recognition of IL-17. J Exp Med. 2024;221(4):e20231425.
31. Lin D, Zhai X, Qi X, Zhou Q, Liu Y, Lin Y, Liu J. Senescent cancer-associated fibroblasts facilitate tumor associated neutrophil recruitment suppressing tumor immunity. J Transl Med. 2024;22(1):231.
32. Talaat IM, Hachim MY, Hachim IY, Ibrahim RAE-R, Ahmed MAER, Tayel HY. Bone marrow mammaglobin-1 (SCGB2A2) immunohistochemistry expression as a breast cancer specific marker for early detection of bone marrow micrometastases. Sci Rep. 2020;10(1):13061.

33. Zafrakas M, Petschke B, Donner A, Fritzsche F, Kristiansen G, Knüchel R, Dahl E. Expression analysis of mammaglobin A (SCGB2A2) and lipophilin B (SCGB1D2) in more than 300 human tumors and matching normal tissues reveals their co-expression in gynecologic malignancies. BMC Cancer. 2006;6:1–13.
34. Chu J, Tang S, Li T, Fan H. The role of CD8A in the immune microenvironment of breast cancer. Front Biosci-Landmark. 2024;29(2):73.
35. Niu D, Chen Y, Mi H, Mo Z, Pang G. The epiphany derived from T-cell–inflamed profiles: pan-cancer characterization of CD8A as a biomarker spanning clinical relevance, cancer prognosis, immunosuppressive environment, and treatment responses. Front Genet. 2022;13:974416.
36. Zheng Z, Guo Y, Huang X, Liu J, Wang R, Qiu X, Liu S. CD8A as a prognostic and immunotherapy predictive biomarker can be evaluated by MRI radiomics features in bladder cancer. Cancers. 2022;14(19):4866.
37. Duquette D, Harmon C, Zaborowski A, Michelet X, O'Farrelly C, Winter D, Koay H-F, Lynch L. Human Granzyme K is a feature of innate T cells in blood, tissues, and tumors, responding to cytokines rather than TCR stimulation. J Immunol. 2023;211(4):633–47.
38. Mogilenko DA, Shpynov O, Andhey PS, Arthur L, Swain A, Esaulova E, Brioschi S, Shchukina I, Kerndl M, Bambouskova M, et al. Comprehensive profiling of an aging immune system reveals clonal GZMK(+) CD8(+) T cells as conserved hallmark of inflammaging. Immunity. 2021;54(1):99–115.
39. Bade B, Boettcher HE, Lohrmann J, Hink-Schauer C, Bratke K, Jenne DE, Virchow JC Jr, Luttmann W. Differential expression of the granzymes A, K and M and perforin in human peripheral blood lymphocytes. Int Immunol. 2005;17(11):1419–28.
40. Kim T-D, Lee SU, Yun S, Sun H-N, Lee SH, Kim JW, Kim HM, Park S-K, Lee CW, Yoon SR, et al. Human microRNA-27a* targets Prf1 and GzmB expression to regulate NK-cell cytotoxicity. Blood. 2011;118(20):5476–86.
41. Revell PA, Grossman WJ, Thomas DA, Cao X, Behl R, Ratner JA, Lu ZH, Ley TJ. Granzyme B and the downstream granzymes C and/or F are important for cytotoxic lymphocyte functions1. J Immunol. 2005;174(4):2124–31.
42. Khalyfa A, Bourbeau D, Chen E, Petroulakis E, Pan J, Xu S, Wang E. Characterization of elongation factor-1A (eEF1A-1) and eEF1A-2/S1 protein expression in normal and <em>wasted</em> Mice *. J Biol Chem. 2001;276(25):22915–22.
43. Zheng Y, Rudensky AY. Foxp3 in control of the regulatory T cell lineage. Nat Immunol. 2007;8(5):457–62.
44. Rudensky AY. Regulatory T cells and Foxp3. Immunol Rev. 2011;241(1):260–8.
45. Gavin MA, Rasmussen JP, Fontenot JD, Vasta V, Manganiello VC, Beavo JA, Rudensky AY. Foxp3-dependent programme of regulatory T-cell differentiation. Nature. 2007;445(7129):771–5.
46. Hori S, Nomura T, Sakaguchi S. Control of regulatory T cell development by the transcription factor Foxp3. Science. 2003;299(5609):1057–61.
47. Kazanietz MG, Durando M, Cooke M. CXCL13 and its receptor CXCR5 in cancer: inflammation, immune response, and beyond. Front Endocrinol. 2019. https://doi.org/10.3389/fendo.2019.00471.
48. Klimatcheva E, Pandina T, Reilly C, Torno S, Bussler H, Scrivens M, Jonason A, Mallow C, Doherty M, Paris M, et al. CXCL13 antibody for the treatment of autoimmune disorders. BMC Immunol. 2015;16(1):6.
49. Mohammad Alhawiti N, Al Mahri S, Azhar Aziz M, Shafi Malik S, Mohammad S. TXNIP in metabolic regulation: physiological role and therapeutic outlook. Curr Drug Targets. 2017;18(9):1095–103.
50. Parikh H, Carlsson E, Chutkow WA, Johansson LE, Storgaard H, Poulsen P, Saxena R, Ladd C, Schulze PC, Mazzini MJ, et al. TXNIP regulates peripheral glucose metabolism in humans. PLoS Med. 2007;4(5):e158.
51. Pan M, Zhang F, Qu K, Liu C, Zhang J. TXNIP: a double-edged sword in disease and therapeutic outlook. Oxid Med Cell Longev. 2022;2022:7805115.
52. Xu W, Zhao X, Wang X, Feng H, Gou M, Jin W, Wang X, Liu X, Dong C. The transcription factor Tox2 drives T follicular helper cell development via regulating chromatin accessibility. Immunity. 2019;51(5):826–39.
53. Jiang B, Chen W, Qin H, Diao W, Li B, Cao W, Zhang Z, Qi W, Gao J, Chen M. TOX3 inhibits cancer cell migration and invasion via transcriptional regulation of SNAI1 and SNAI2 in clear cell renal cell carcinoma. Cancer Lett. 2019;449:76–86.
54. Tower H, Ruppert M, Britt K. The immune microenvironment of breast cancer progression. Cancers. 2019. https://doi.org/10.3390/cancers11091375.
55. Xu Q, Chen S, Hu Y, Huang W. Landscape of immune microenvironment under immune cell infiltration pattern in breast cancer. Front Immunol. 2021. https://doi.org/10.3389/fimmu.2021.711433.
56. Tekpli X, Lien T, Røssevold AH, Nebdal D, Borgen E, Ohnstad HO, Kyte JA, Vallon-Christersson J, Fongaard M, Due EU, et al. An independent poor-prognosis subtype of breast cancer defined by a distinct tumor immune microenvironment. Nat Commun. 2019;10(1):5499.
57. Speiser DE, Verdeil G. More T cells versus better T cells in patients with breast cancer. Cancer Discov. 2017;7(10):1062–4.
58. Ben-Baruch A. The tumor-promoting flow of cells into, within and out of the tumor site: regulation by the inflammatory axis of TNFα and chemokines. Cancer Microenviron. 2012;5(2):151–64.
59. Sadeghalvad M, Mohammadi-Motlagh H-R, Rezaei N. Immune microenvironment in different molecular subtypes of ductal breast carcinoma. Breast Cancer Res Treat. 2021;185(2):261–79.
60. Tang S, Wang Q, Sun K, Song Y, Liu R, Tan X, Li H, Lv Y, Yang F, Zhao J, et al. Metabolic heterogeneity and potential immunotherapeutic responses revealed by single-cell transcriptomics of breast cancer. Apoptosis. 2024. https://doi.org/10.1007/s10495-024-01952-7.
61. Moretta L, Moretta A. Unravelling natural killer cell function: triggering and inhibitory human NK receptors. EMBO J. 2004;23(2):255–9.
62. Lu J, Peng Y, Huang R, Feng Z, Fan Y, Wang H, Zeng Z, Ji Y, Wang Y, Wang Z. Elevated TYROBP expression predicts poor prognosis and high tumor immune infiltration in patients with low-grade glioma. BMC Cancer. 2021;21(1):723.
63. Li T, Chen S, Zhang Y, Zhao Q, Ma K, Jiang X, Xiang R, Zhai F, Ling G. RETRACTED ARTICLE: ensemble learning-based gene signature and risk model for predicting prognosis of triple-negative breast cancer. Funct Integr Genomics. 2023;23(2):81.
64. Milovanović J, Todorović-Raković N, Vujasinović T, Greenman J, Mandušić V, Radulovic M. Can granulysin provide prognostic value in primary breast cancer? Pathol – Res Pract. 2022;237:154039.
65. Cai D, Cai D, Zou Y, Chen X, Jian Z, Shi M, Lin Y, Chen J. Construction and validation of chemoresistance-associated tumor-infiltrating exhausted-like CD8+ T cell signature in breast cancer: cr-TILCD8TSig. Front Immunol. 2023;14:1120886.
66. Wen T, Barham W, Li Y, Zhang H, Gicobi JK, Hirdler JB, Liu X, Ham H, Peterson Martinez KE, Lucien F, et al. NKG7 Is a T-cell–intrinsic therapeutic target for improving antitumor cytotoxicity and cancer immunotherapy. Cancer Immunol Res. 2022;10(2):162–81.
67. Guan X, Guo H, Guo Y, Han Q, Li Z, Zhang C. Perforin 1 in cancer: mechanisms, therapy, and outlook. Biomolecules. 2024. https://doi.org/10.3390/biom14080910.
68. Tokura M, Nakayama J, Prieto-Vila M, Shiino S, Yoshida M, Yamamoto T, Watanabe N, Takayama S, Suzuki Y, Okamoto K, et al. Single-cell transcriptome profiling reveals intratumoral heterogeneity and molecular features of ductal carcinoma in situ. Can Res. 2022;82(18):3236–48.
69. Zhang Z, Cui F, Wang C, Zhao L, Zou Q. Goals and approaches for each processing step for single-cell RNA sequencing data. Brief Bioinform. 2021. https://doi.org/10.1093/bib/bbaa314.
70. Zhang Z, Cui F, Lin C, Zhao L, Wang C, Zou Q. Critical downstream analysis steps for single-cell RNA sequencing data. Brief Bioinform. 2021. https://doi.org/10.1093/bib/bbab105.
71. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, Chak S, Naikawadi RP, Wolters PJ, Abate AR, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol. 2019;20(2):163–72.
72. Mabbott NA, Baillie JK, Brown H, Freeman TC, Hume DA. An expression atlas of human primary cells: inference of gene function from coexpression networks. BMC Genomics. 2013;14(1):632.
73. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan C-H, Myung P, Plikus MV, Nie Q. Inference and analysis of cell-cell communication using cell chat. Nat Commun. 2021;12(1):1088.
74. Zhang Z, Cui F, Cao C, Wang Q, Zou Q. Single-cell RNA analysis reveals the potential risk of organ-specific cell types vulnerable to SARS-CoV-2 infections. Comput Biol Med. 2021;140:105092.
75. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-Seq data. BMC Bioinform. 2013;14(1):7.
76. Zhang ZY, Sun ZJ, Gao D, Hao YD, Lin H, Liu F. Excavation of gene markers associated with pancreatic ductal adenocarcinoma based on

interrelationships of gene expression. IET Syst Biol. 2024. https://doi.org/10.1049/syb2.12090.

77. Ren L, Huang D, Liu H, Ning L, Cai P, Yu X, Zhang Y, Luo N, Lin H, Su J, et al. Applications of single-cell omics and spatial transcriptomics technologies in gastric cancer (Review). Oncol Lett. 2024;27(4):152.
78. Li H, Pang Y, Liu B. BioSeq-BLM: a platform for analyzing DNA, RNA, and protein sequences based on biological language models. Nucleic Acids Res. 2021;49(22):e129.
79. Zhang Z, Cui F, Zhou M, Wu S, Zou Q, Gao B. Single-cell RNA sequencing analysis identifies key genes in brain metastasis from lung adenocarcinoma. Curr Gene Ther. 2021;21(4):338–48.

Publisher's Note