

Integrated analysis of single-cell RNA-seq and bulk RNA-seq unravels the heterogeneity of cancer-associated fibroblasts in TNBC

Xiaoqing Wu¹, Wenping Lu¹, Weixuan Zhang¹, Dongni Zhang¹, Heting Mei¹, Mengfan Zhang¹, Yongjia Cui¹, Zhili Zhuo¹

¹Department of Oncology, Guang'anmen Hospital, China Academy of Chinese Medical Sciences, Beijing 100053, People's Republic of China

Correspondence to: Wenping Lu; email: lu_wenping@sina.com, <https://orcid.org/0000-0001-5589-1048>

Keywords: triple-negative breast cancer, heterogeneity, cancer-associated fibroblasts, bioinformatic

Received: July 25, 2023

Accepted: October 3, 2023

Published: November 13, 2023

Copyright: © 2023 Wu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Background: The treatment of triple-negative breast cancer (TNBC) is one of the main focuses and key difficulties because of its heterogeneity, and the source of this heterogeneity is unclear.

Methods: Single-cell RNA (scRNA) and transcriptomics data of TNBC and normal breast samples were retrieved from Gene Expression Omnibus (GEO) database and TCGA-BRCA database. These cells were clustered using the t-SNE and UMAP method, and the marker genes for each cluster were found. We annotated the clusters using the published literature, CellMarker database and “SingleR” R package.

Results: A total of 1535 cells and 21785 genes from 6 TNBC patients and 2068 cells and 15868 genes from 3 normal breast tissues were used for downstream analyses. The scRNA data were divided into 14 clusters labeled into 8 cell types, including epithelial cells, immunocytes, CAFs/fibroblasts and etc. In the TNBC samples, CAFs were divided into three clusters and labelled as prCAFs, myCAFs and emCAFs, and the marker genes were DCN, FAP and RGS5, respectively. The prCAF subgroup is functionally characterized by promoting proliferation and multi drug resistance; myCAF subgroup is involved in constituting the extracellular matrix and collagen production, matrix composition and collagen production, and the emCAF functionally characterized by energy metabolism.

Conclusions: TNBC has inter- and intra-tumor heterogeneity, and CAF is one of the sources of this heterogeneity. CD74, SASH3, CD2, TAGAP and CCR7 served as significant marker genes with prognostic and therapeutic value.

INTRODUCTION

Female breast cancer has surpassed lung cancer as the most prevalent cancer, with an estimated 2.3 million new cases (11.7%), and triple-negative breast cancer (TNBC) accounts for 11%-20% of all breast cancers [1]. It is characterized by the lack of estrogen receptor, progesterone receptor and human epidermal growth factor receptor 2 expression, as assessed by immunohistochemistry [2]. Its biological aggressiveness and propensity to metastasize are higher than those of any other pathological type of breast cancer [3].

Patients with TNBC have a poor prognosis as well [4]. Therapeutic approaches to TNBC are limited by the lack of therapeutic targets. Conventional chemotherapy, consisting mainly of anthracyclines and paclitaxel, has been the mainstay of treatment [5]. However, the heterogeneity of TNBC contributes to reducing the effectiveness of chemotherapy [6]. Although the majority of TNBC patients receive conventional chemotherapy, some patients will still experience recurrent metastases. This suggests that eliminating most cancer cells in tumor tissue does not inhibit progression [7]. A study showed that multiclonal seeding from individual

clones in TNBC tissue leads to multisite metastasis [8]. Oncogene mutations, amplifications, loss-of-function mutations in tumor suppressors, and large-scale chromosomal alterations are all recognized mechanisms that have previously been shown to drive cancer evolution and generate subpopulations of cancer cells for metastatic spread and growth [9, 10]. However, recent advances suggest that epigenetics and transcriptional reprogramming are now considered key factors driving tumor heterogeneity and evolution [11]. Single-cell RNA sequencing (scRNA-seq) uses optimized next-generation sequencing technology to define the global gene expression profile of individual cells, which helps isolate heterogeneity previously hidden in cell populations [12]. In a patient-derived breast cancer xenograft model, single-cell RNA sequencing revealed that both primary tumors and micro-metastases show transcriptional heterogeneity that is highly predictive of poor patient survival [13]. These results may be evidence of tumor progression guided by tumor heterogeneity. Therefore, identifying the diversity of TNBC cells and subsets of associated cells at the single-cell level will contribute to the precise treatment of TNBC.

The tumor microenvironment, which includes the surrounding cells and molecules that interact with cancer cells, plays a critical role in the development of tumors, including the occurrence, progression, and immune suppression of the tumors [14, 15], cancer-associated fibroblasts are an important part of that. CAFs as the prominent stromal cell type in solid tumors, can enhance tumor phenotypes, especially cancer cell proliferation and invasion, neoangiogenesis, inflammation and extracellular matrix (ECM) remodeling [16, 17]. Abundant stromal myofibroblasts are associated with aggressive adenocarcinoma in human breast tumors and predict disease recurrence in humans [18]. In addition, the infiltration of CAFs contributes to angiogenesis, drug resistance and reduces anti-tumor immunity [3, 19, 20]. Most studies have focused on the relationship between CAFs and the tumor microenvironment and the immune cells within it, while its heterogeneity in human cancers is far from complete. In human TNBC samples with single-cell sequencing, CAFs clustered into two states: the first with features of myofibroblasts, and the second with high expression of growth factors and immunomodulatory molecules [21]. A study classified CAFs in human breast cancer into four subpopulations with different properties and activation levels. Two subpopulations of myofibroblasts with immunosuppressive effects (CAF-S1 and CAF-S4) accumulate differentially in TNBC [22]. Three CAF subpopulations were identified in 4T1 tumors transplanted from BALB/c mice, including myofibroblastic CAFs, enriched in α smooth muscle actin and several

other contractile proteins; inflammatory CAFs with elevated expression of inflammatory cytokines; and CAF subpopulations expressing major histocompatibility complex class II proteins [23]. It has been previously shown that CAFs express high levels of alpha-smooth muscle actin (α -SMA/Acta2), fibroblast activation protein (Fap), fibroblast-specific protein 1 (S100a4), CD74, etc. [24, 25]. However, these markers are not exhaustive or can characterize certain subgroups. Therefore, the use of single-cell analysis is necessary to identify intra- and inter-tumor heterogeneity in TNBC tumors. Based on the above, we proposed the hypothesis that CAFs in TNBC are heterogeneous and that this heterogeneity may contribute to differences in patient prognosis.

In the present study, to explore the heterogeneity of CAFs in TNBC and the prognostic value of CAFs, we identified the biomarkers and functions of three subpopulations of CAFs and molecular fractionation of TNBC through scRNA-seq and bulk RNA-seq, described the tumor heterogeneity in TNBC and developed a prognostic model based on CAF-related genes.

MATERIALS AND METHODS

Data extraction

scRNA-seq was retrieved from GSE75688, GSE118389 and GSE226391 in the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). A total of 21785 genes and 1534 cells were obtained in 6 TNBC patients and 21921 genes and 4840 cells in 3 normal breast tissues. Transcriptomics data from GSE19615 and GSE21653 in the GEO database and 107 TNBC patients were screened by clinical information. Transcriptomics data of 142 TNBC patients were screened by corresponding clinical information from TCGA-BRCA in the TCGA database (<https://portal.gdc.cancer.gov/>). RNA-seq data were normalized, and batch effects were removed by the R packages “limma” and “sva”.

scRNA-seq quality control and analyses

The extracted data need to be quality controlled. Data quality control and analysis were applied using the “seurat” R package. First, the percentage of mitochondrial genes was tested, and a percentage less than 5% was considered satisfactory for cell quality control [26]. Second, the sequencing depth and counts of each cell were detected, and low-quality single-cell data were removed. Third, the first 1500 highly variable genes were extracted for subsequent analysis. After that, PCA was performed for linear

dimensionality reduction. These cells were clustered using the louvain clustering method, and the marker genes for each cluster were found. In the analysis for the TNBC samples, the PC value is defined as 14 according to the ElbowPlot function, as shown in Supplementary Figure 1. The resolution is defined as 0.5 according to the “clustree” package, as shown in Supplementary Figure 2. In the analysis for the normal tissue, the PC value is defined as 9, the resolution is defined as 0.5, as shown in Supplementary Figures 3, 4. K-nearest neighbor value defaults to 20 according to the number of cells. We annotated the clusters using the published literature, the CellMarker database (<http://xteam.xbio.top/CellMarker/index.jsp>) and referring to the annotation results of the “SingleR” R package. The marker genes for each cluster were considered to be the most significantly different genes in PCs. Tumor cells and normal cells were identified by the R package “copykat”. The distribution of cell types for each sample was counted and visualized. Finally, cell trajectory analysis was performed using the “monocle” R package to derive the trajectory for each type of cell in terms of state, clustering and pseudotime distribution to find the differential genes on each branch. The results of each of the above steps were visualized.

Biological functional and signaling pathway analyses

To identify the function of each CAF subgroup, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were performed using the genes and expressions of each subgroup. GO and KEGG enrichment analyses using cluster marker genes were performed with R with the aid of the packages clusterProfiler, enrichplot, and ggplot2. Terms with FDR values of < 0.05 were considered significantly enriched. GO analyses included GO biological process, GO molecular function and GO cellular component. The top 10 terms of all of the above analyses were visualized by barplot.

Identification of molecular subtypes

We intended to cluster the sample with the inner feature in tumor samples and extracted biological correlation coefficients, and the R package “ConsensusClusterPlus” was used to take the above action. We selected an optimum k value of 2 to 9 in consideration of stability and clustering performance. A suitable k value was chosen based on consensus CDF and delta area. Kaplan–Meier analysis was conducted between clusters via the “survival” R package. We also examined the differences in clinical characteristics in clusters and visualized them.

Definition and comparison of the tumor immune infiltration microenvironment (TIME)

We used ESTIMATE to assess the tumor microenvironment by the R package “estimate” and obtained ESTIMATE scores, immune scores, stromal scores and tumor purity. By using the CIBERSORT algorithm, we calculated the absolute abundance of 22 types of immune cells. By contrasting the scores between clusters, it was possible to identify the infiltrating cells. The “ggpubr” package was used to visualize the results above. Differences in common immune checkpoints between clusters were analyzed by the R package “limma”. Kaplan–Meier analysis was conducted between clusters via the “survival” R package.

Construction and validation of CAF-related gene prognostic features

CAF signature genes were found in the pseudotime analysis, which were used to build prognostic prediction model. To screen out CAF-related genes that may be related to prognosis, we performed Cox regression with the R packages “survival” and “survminer” to obtain candidate genes. The Lasso algorithm was executed to construct a prognostic model by the “glmnet” R package. We determined the risk score using the formula:

$$Risk\ score = \sum_{i=1}^n (coef_i * Exp_i)$$

The mathematical meaning of the representation of this formula is to calculate the expression of each candidate gene multiplied by the regression coefficient of the multivariate Cox and then to sum all the values. We divided all cases into a high-risk group and a low-risk group based on the median risk score. The training set was composed of GEO data, and the testing set was composed of TCGA data. Receiver operating characteristic (ROC) curve analysis was used to confirm that the prognostic model was stable. To establish the risk score as a significant prognostic factor, univariate and multivariate Cox regression was conducted.

Establishment and verification of the nomogram

The R program “rms” was used to finish the establishment, and “regplot” was used to complete the visualization. The model was built taking into account nomogram risk scores and clinical features. We offer a measurable tool to forecast overall survival at 1, 3, and 5 years. Calibration curves were created to evaluate the effectiveness of the nomogram.

Extraction of hub genes

Hub genes, which are high-degree genes, have high connectivity in the protein-protein interaction network. The PPI network was constructed by the STRING database (<https://cn.string-db.org/>) based on CAF-related genes, followed by reconstruction with Cytoscape version 3.6.1. Nodes with confidence of interactive relationship larger than 0.95 were used for building the network. The number of adjacent nodes of each gene was counted by R.

Statistical analysis

All the data were processed and analyzed by using R software (version: 4.1.2). In comparisons of two classes of data, we used the Mann–Whitney U test (for nonnormally distributed data) or Student's t test (for normally distributed data). Spearman's correlation test (for nonnormally distributed data) or Pearson's correlation test (for normally distributed data) was applied to evaluate the correlation between two groups of data. We used Fisher's exact or chi-square tests to assess the association between two categorical variables. We utilized the FDR, evaluated by the Benjamini–Hochberg method, to adjust for multiple tests.

Availability of data and materials

All data generated or analysed during this study are included in this published article.

RESULTS

Normalization and dimensionality reduction of scRNA-seq data

The flow of analyses in the presented study is shown in Figure 1. In the GSE226391, three samples, GSM7074398, GSM7074399 and GSM7074340, were of too low quality and were therefore excluded. Ultimately, a total of 1535 cells and 21785 genes from 6 TNBC patients and 2068 cells and 15868 genes from 3 normal breast tissues were used for downstream analyses. Visualization of sequencing depth, gene quantity, and mitochondrial gene content showed that scRNA data were available, as shown in Figure 2A, 2B. The percent of mitochondrial genes was independent of the counts of genes, and the sequencing depth was positively correlated with the counts of genes with a coefficient of 0.63 and 0.93 in TNBC and normal breast individually (Figure 2C, 2D). The relationship between normalized variance and average expression of all genes for each cell was visualized as shown in Figure 2E, 2F, and the 1500 genes with

the largest normalized variables were selected for subsequent analysis.

Identification of clusters and cell types

To better specify the types of these cells, we used the t-SNE and UMAP method to classify them into 14 clusters in TNBC, as shown in Figure 3A. The 14 clusters were labeled into eight cell types based on known marker expression, including epithelial cells, T cells, tumor-associated fibroblasts, macrophages, tumor stem cells, smooth muscle cells, tissue stem cells, and endothelial cells. The gene markers for each cell cluster are shown in the scatter plot in Figure 3B. There was heterogeneity in the distribution of cell types per tumor sample, as shown in Figure 3C. Whereas in normal breast, these cells were clustered by 9 clusters in t-SNE and UMAP, as shown in Figure 4. The 9 clusters were labeled into 6 types, including endothelial cells, epithelial cells, fibroblasts, keratinocytes, monocytes and tissue stem cells. Interestingly, we found that the cells labeled CAFs were divided into three subpopulations, whereas in normal breast tissue there was only one cluster of fibroblasts. We suggested that this may be due to heterogeneity within CAFs as well.

Heterogeneity of CAFs and functional analyses

In TNBC, clusters 7, 11 and 13 were labeled CAFs, and the marker genes for these clusters were DCN, FAP and RGS5, respectively, which are widely reported markers of CAFs. We performed a separate t-SNE analysis of CAFs and found that they were clearly divided into three subgroups, as shown in Figure 5A. The marker genes for each subgroup can be seen in the heatmap in Figure 5B.

To identify the function of each subgroup, we performed GO and KEGG analyses to identify their functional differences, as shown in Figure 5D–5I. The cluster 0 subgroup is functionally characterized by promoting proliferation and multi drug resistance, activating PI3K signalling and ABC transporters, the cluster 1 subgroup is functionally characterized by constituting the extracellular matrix and collagen production, and the cluster 2 subgroup is functionally characterized by energy metabolism, in which Notch signaling pathway and Apelin pathway are activated. In the present research, we labeled cluster 0 as prCAFs, cluster 1 as myCAFs, and cluster 2 as emCAFs, as shown in Figure 5C.

Molecular subgroup of TNBC based on CAF characteristic genes

The distribution of clusters in the trajectory diagram is shown in Figure 6A. The differentiation trajectory

diagram showed that there are five main states of these cells (Figure 6B), and it can be seen that epithelial cells are distributed in all five states, and CAFs are mainly distributed in the first state, with the largest number of

cell types in this state, as shown in Figure 6C. The pseudotime analysis diagram showed that CAFs are mainly distributed at the starting point from the pseudotime line, as shown in Figure 6D.

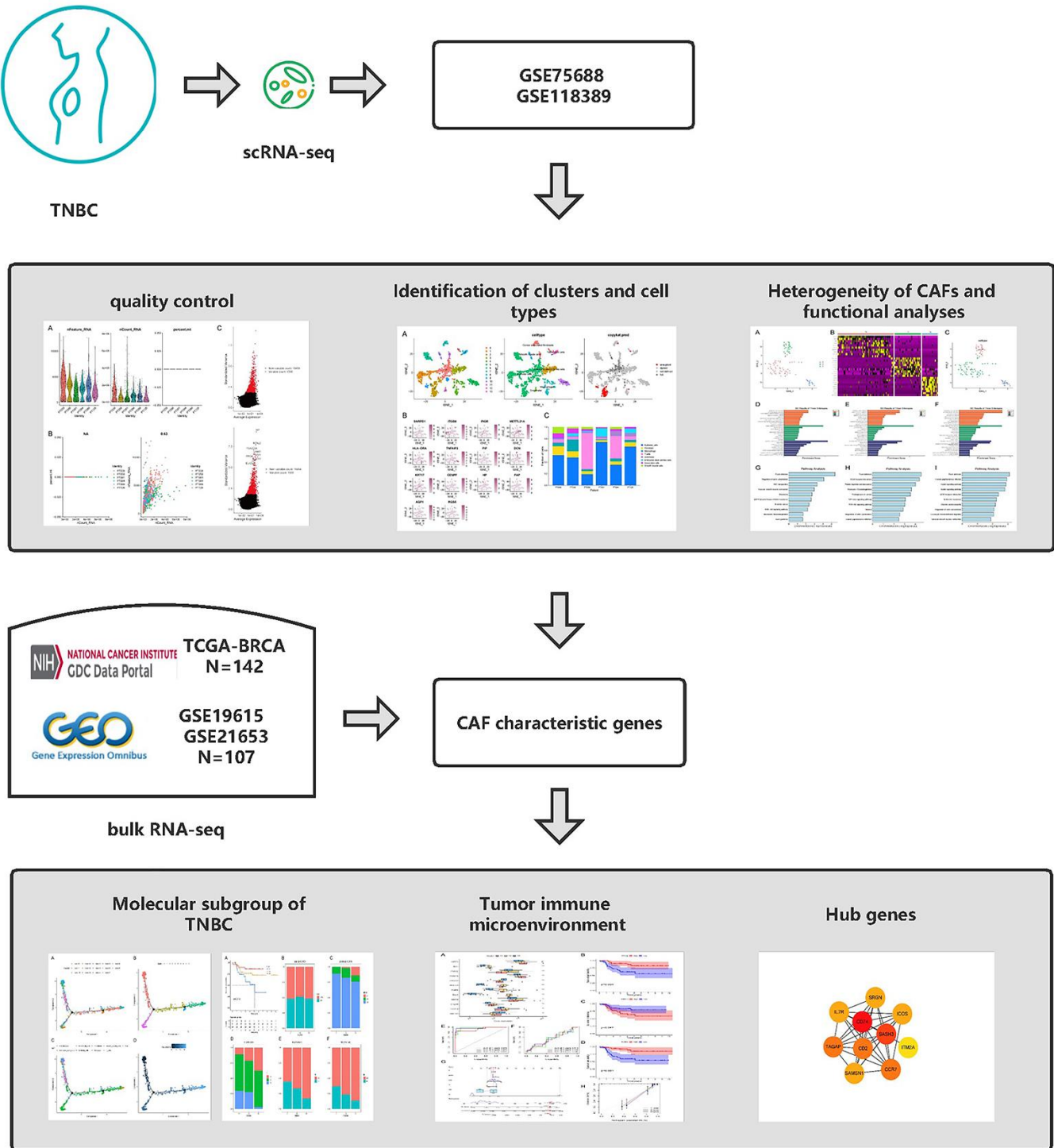


Figure 1. Graphical abstract of the analysis process.

There were 450 characteristic genes in the first state, and these genes were further analyzed. The 107 patients from GSM19615 and GSM21653 were divided into 3 clusters based on the expression of these genes. The clinical characteristics of patients were shown as Table 1. The Kaplan–Meier analysis (Figure 7A) showed that the survival probability of patients in cluster C3 was the longest and that of patients in cluster C1 was the shortest ($P=0.019$). Age, grade, lymph node metastasis (N) and distant metastasis (M) did not differ among the three clusters ($P>0.05$), while the size of the tumor (T) differed from each cluster ($P=0.029$), as shown in Figure 7B–7F.

Tumor immune microenvironment (TIME) differs between clusters

To investigate the correlation between cluster and TIME, ESTIMATE and CIBERSORT analyses were performed. The results showed that there were differences in the ESTIMATE score, immune score, stromal score and tumor purity between C1 and C2 and between C2 and C3 ($P<0.001$), as shown in Figure 8A–8D. This suggested that there are differences in the

TIME and stromal microenvironment among clusters. Among the three clusters, C2 had the least tumor cell infiltration and the most immune cells and stromal components. There were differences between clusters in multiple immune cells, including T cells, as shown in Figure 8E. Calculation of CAF scores in the tumor microenvironment using the Xcell, MCPcounter and EPIC algorithms is shown in Figure 7F. Kaplan–Meier analysis showed that CD4 memory activated T cells, CD4 memory resting T cells and gamma delta T cells were positively related to overall survival, while regulatory T cells and M0 macrophages were negatively related to overall survival (Figure 8G–8K).

A checkpoint distribution among the three clusters was performed to predict the response to immunotherapy. There were differences among clusters in the distribution of genes encoding the immune checkpoints WEE1, SHH, PVRIG, PRMT5, PIK3CA, PDCD1, EGFR, DLL4, DDR2, CTLA4, CD274 and BRCA (Figure 9A). In Kaplan–Meier analyses, high expression of CTLA4 and PVRIG and low expression of PRMT5 were related to longer overall survival (Figure 9B–9D).

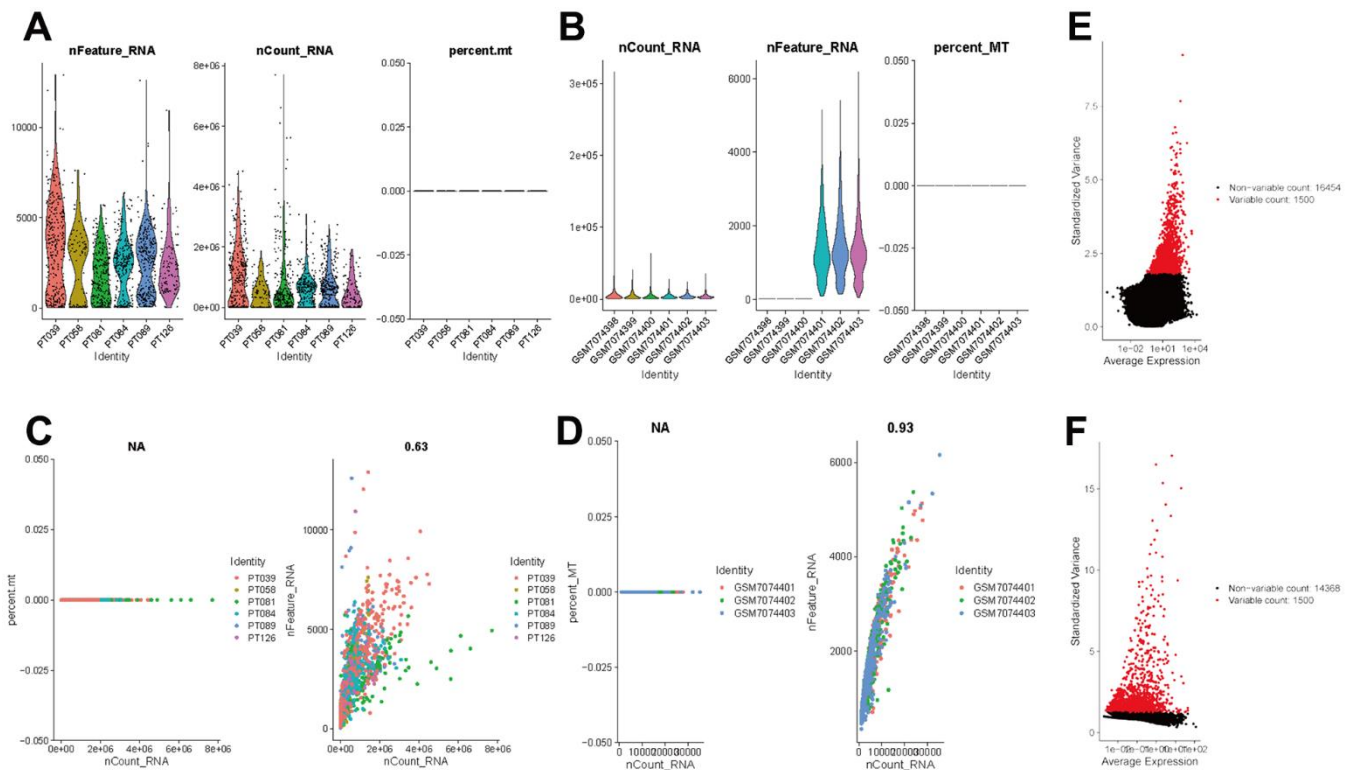


Figure 2. Normalization and quality control of scRNA-seq. (A) Single-cell sequencing depth, counts and fraction of reads mapped to mitochondrial genes in TNBC samples. (B) Single-cell sequencing depth, counts and fraction of reads mapped to mitochondrial genes in breast tissues. (C) Correlation of gene count and percent of mitochondrial genes and features in TNBC. (D) Correlation of gene count and percent of mitochondrial genes and features in breast tissue. (E) The first 1500 genes were screened as variant genes in TNBC samples. (F) The first 1500 genes were screened as variant genes in normal breast.

Development of a prognostic predictive model and a nomogram

The 33 genes associated with prognosis were screened using univariate regression analysis and then compressed to 16 genes using LASSO regression, shown as Table 2. The AUC values of the ROC curves in the training and test sets validate that the model is stable. The 1-year AUC of the training set was

0.976, the 3-year AUC was 0.972, and the 5-year AUC was 0.927 (Figure 9E). The 1-year AUC of the testing set was 0.514, the 2-year AUC was 0.548, and the 3-year AUC was 0.540 (Figure 9F). Univariate and multivariate Cox regressions were implemented for clinical characteristics and risk scores, respectively, and the results showed that the risk score was related to overall survival (univariate: HR= 1.007, 95% CI (1.004–1.009), P<0.001; multivariate: HR= 1.005,

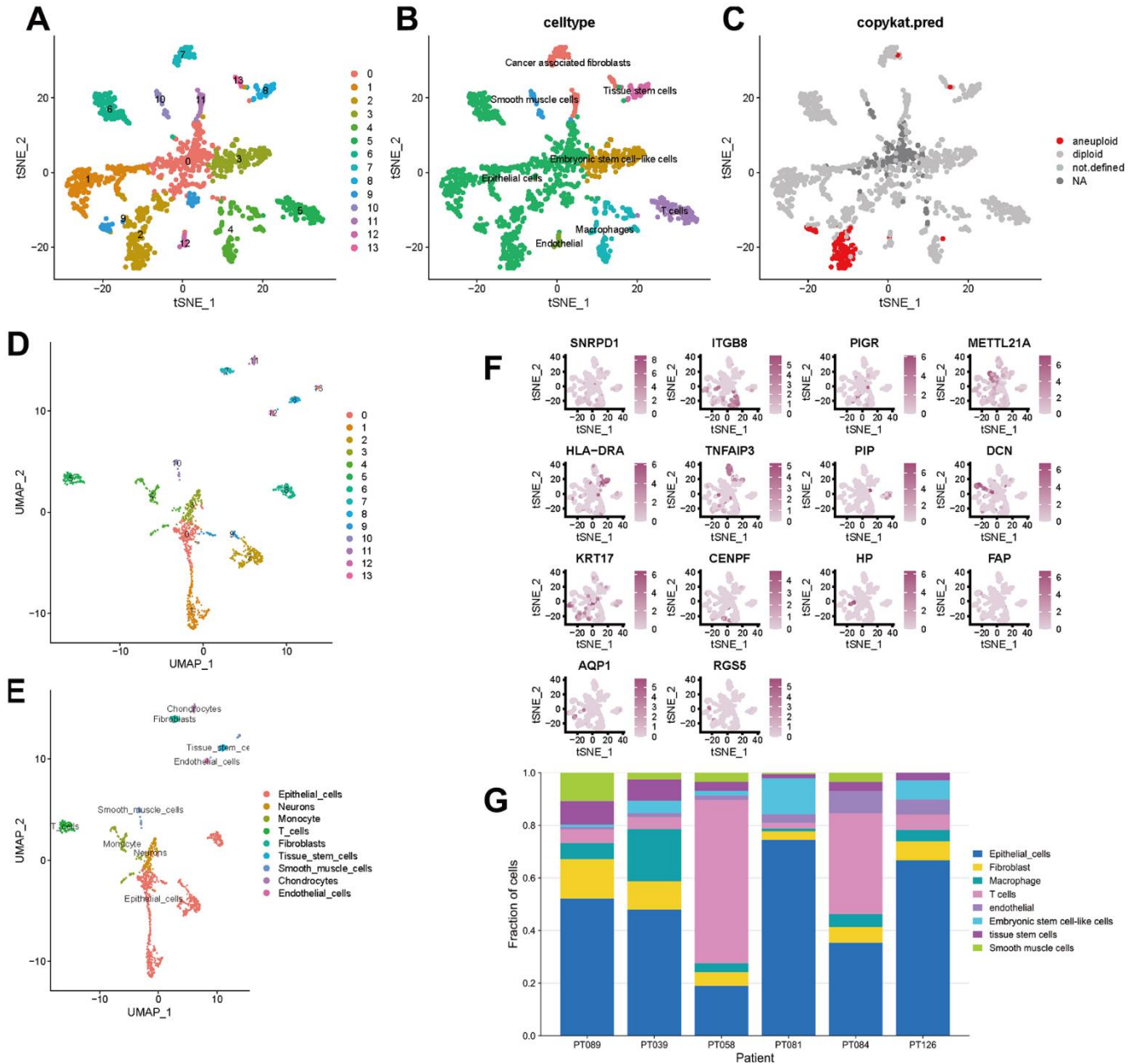


Figure 3. t-SNE and UMAP clustering of TNBC samples. (A–C) Clusters, cell type annotations and aneuploid cells at the single-cell level in TNBC samples by t-SNE method. (D, E) Clusters and cell type annotations at the single-cell level in TNBC samples by UMAP method. (F) Scatter plot of marker gene expression in each cluster. (G) Inter-tumor heterogeneity of triple-negative breast cancer at the single-cell level.

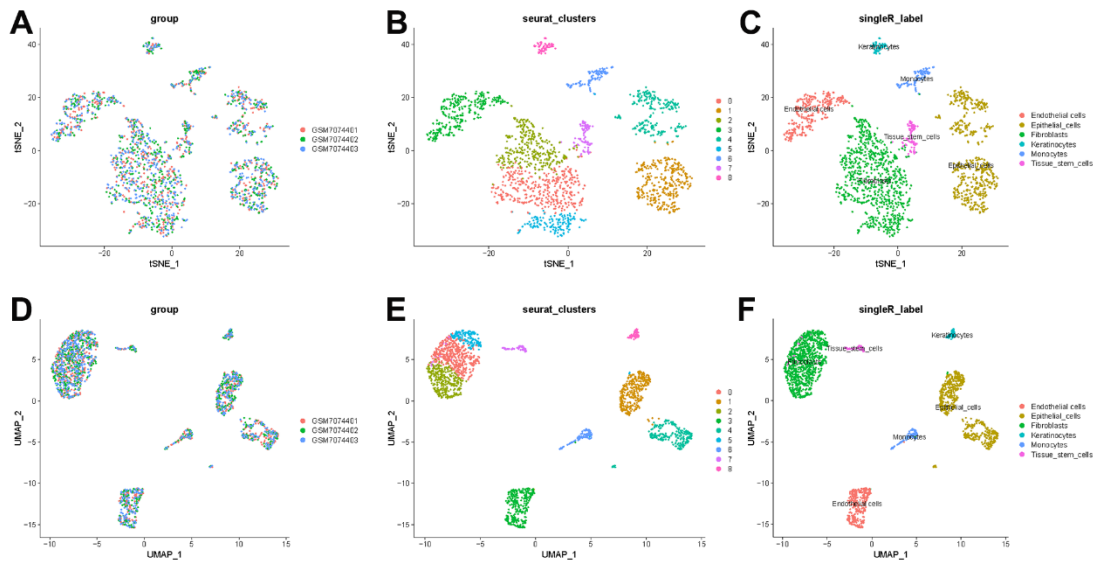


Figure 4. t-SNE and UMAP clustering of normal breast samples. (A) Inter-tumor heterogeneity of triple-negative breast cancer by t-SNE method. (B, C) Clusters and cell type annotations at the single-cell level in TNBC samples by t-SNE method. (D) Inter-tumor heterogeneity of triple-negative breast cancer by UMAP method. (E, F) Clusters and cell type annotations at the single-cell level in TNBC samples by UMAP method.

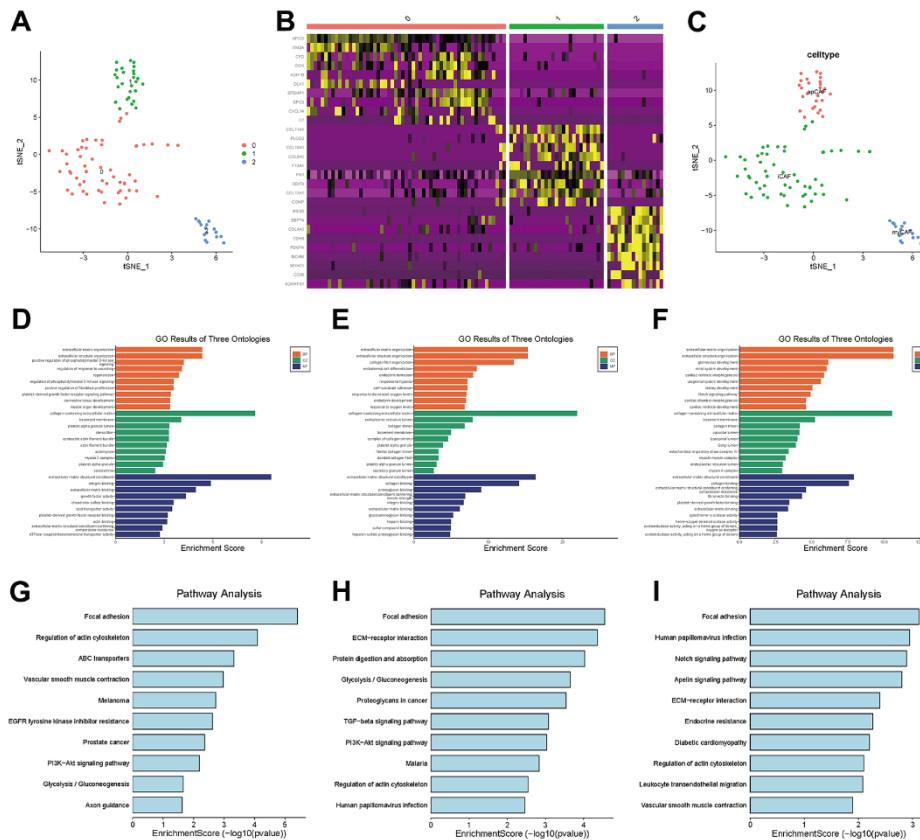


Figure 5. CAF subgroup analysis. (A) Scatter plot of the distribution of clusters in CAFs. (B) Heatmap of the distribution of gene markers in clusters. (C) Type annotation of CAF subgroups. (D) Results of GO analyses of prCAF subgroups. (E) Results of GO analyses of myCAF subgroups. (F) Results of GO analyses of emCAFs. (G) Results of KEGG pathway analyses of prCAF subgroups. (H) Results of KEGG pathway analyses of myCAF subgroups. (I) Results of KEGG pathway analyses of prCAF subgroups.

Table 1. TNBC patient information (bulk RNA-seq).

Characteristics	TCGA cohort N = 143	GEO cohort N = 107
Age		
<=53	63(44.06%)	54(50.47%)
>53	80(55.94%)	53(49.53%)
Gender		
Female	143(100.00%)	107(100%)
Grade		
G1	0	5(4.67%)
G2	0	14(13.08)
G3	0	88(82.24%)
Unknown	143(100%)	0
Stage		
I	23(16.08%)	0
II	96(67.13%)	0
III	23(16.08%)	0
IV	1(0.01%)	0
Unknown	0	107(100%)
Survival status		
Alive	124(86.71%)	80(14.62%)
Dead	19(13.29%)	27(85.38%)
The median follow-up time (year)	2.12	4.02

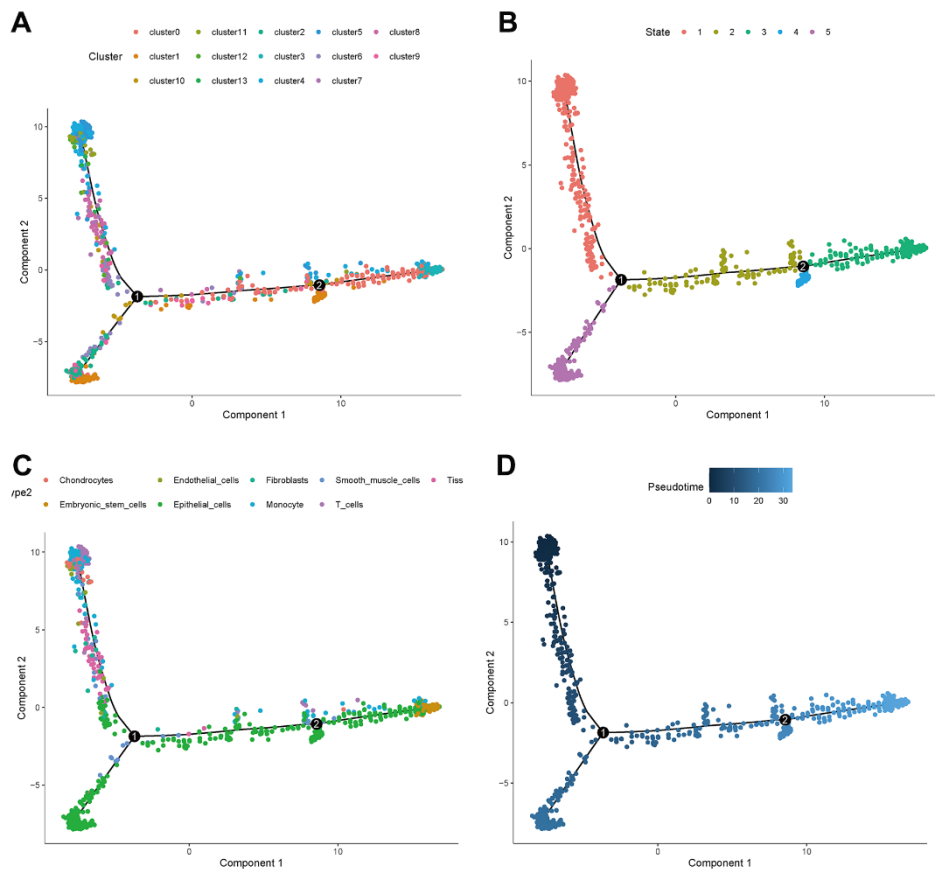


Figure 6. Trajectory analyses of a single cell. (A) Trajectory diagram of cell clusters. **(B)** Trajectory diagram of cell status. **(C)** Trajectory diagram of cell types. **(D)** Trajectory diagram of cell differentiation time.

95% CI (1.002–1.007), $P < 0.001$). To visualize the model, a nomogram was applied (Figure 9G). Calibration curves showed that the nomogram predicted 1-year survival more accurately and may underestimate 3- and 5-year survival (Figure 9H). The hub genes were CD74, SASH3, CD2, TAGAP and CCR7 by degree of nodes (Figure 10).

DISCUSSION

Breast cancer is a highly heterogeneous disease and this heterogeneity leads to the development of drug resistance and progression [17], in order to reveal the source of this heterogeneity as well as to explore therapeutic targets, our study was implemented. In the presented research, the intra- and inter-tumor heterogeneity was identified, and three subpopulations

of CAFs and five branches in the differentiation trajectory were recognized by the scRNA-seq data. The three subgroups are prCAFs, which promote proliferation and multi drug resistance; myCAFs, which are involved in constituting the extracellular matrix and collagen production; and emCAFs, which are functionally characterized by energy metabolism. The naming of these subpopulations was proposed only for the present study, and we looked for established subpopulation names that matched the results of the present study as closely as possible, and for the other subpopulations, naming them according to their functions. The differentiation trajectory showed five statuses of cells. We found that CAFs were enriched in branch 1 and appeared in the early stages of differentiation. We divided the patients into three clusters based on the differentially expressed

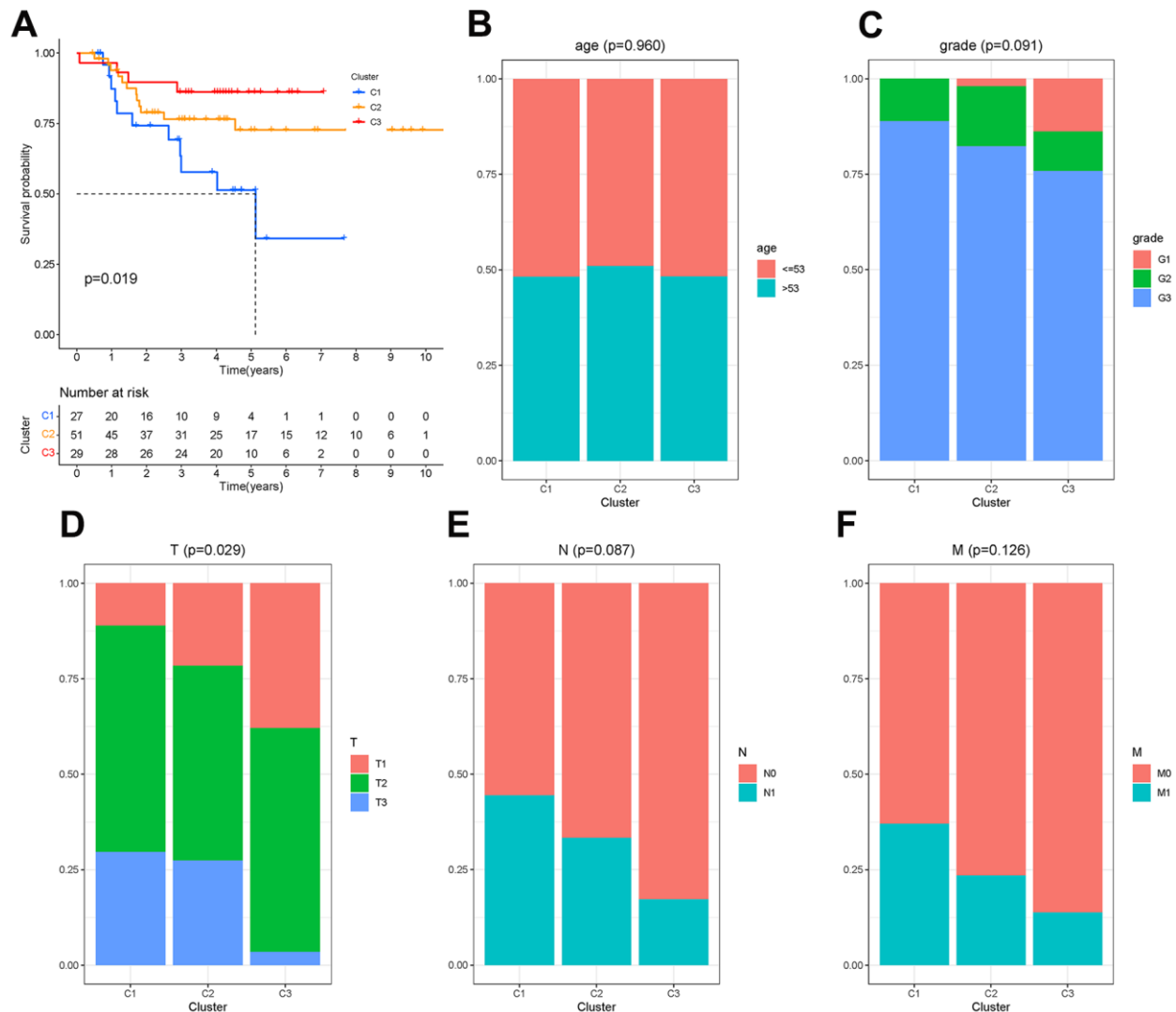


Figure 7. Molecular clusters of TNBC. (A) Kaplan–Meier analysis among molecular clusters. (B–F) Distribution of clinical characteristics in clusters.

genes in the branch containing CAFs. In Kaplan–Meier analysis, C3 patients had a higher survival probability, C2 had the second highest survival probability and C1 had the lowest survival probability. The clinical features and TIME were characterized by clusters. Tumor size and tumor microenvironment varied among clusters. There were differences in T cells and M0 macrophages among clusters. High infiltration of CD4+ memory activated T cells, CD4+ memory resting T cells and gamma delta T cells was related to higher survival probability, while low infiltration of M0 macrophages and regulatory T cells was related to higher survival probability. Immune check-point-related genes varied among clusters. In addition, a prognostic model and nomogram were developed and verified.

In the scRNA-seq data, we found that the sample consisted mainly of epithelial cells, and cancer cells were almost concentrated in the epithelial cells. This result is consistent with the fact that breast cancer originates from the epithelial cells of the breast. Other cells within the sample were cancer-associated fibroblasts, tissue stem cells, smooth muscle cells, embryonic stem cell-like cells, endothelial cells, macrophages and T cells. Cancer-associated fibroblasts were clearly divided into three clusters, and their marker genes were DCN, FAP and RGS5. DCN is considered a marker for one of these subgroups in CAFs of colorectal cancer and high-grade serous ovarian cancer samples [27, 28]. FAP is thought to be a CAF-specific expression protein [29]. RGS5 serves as the gene marker of myCAF [30, 31]. Single-cell

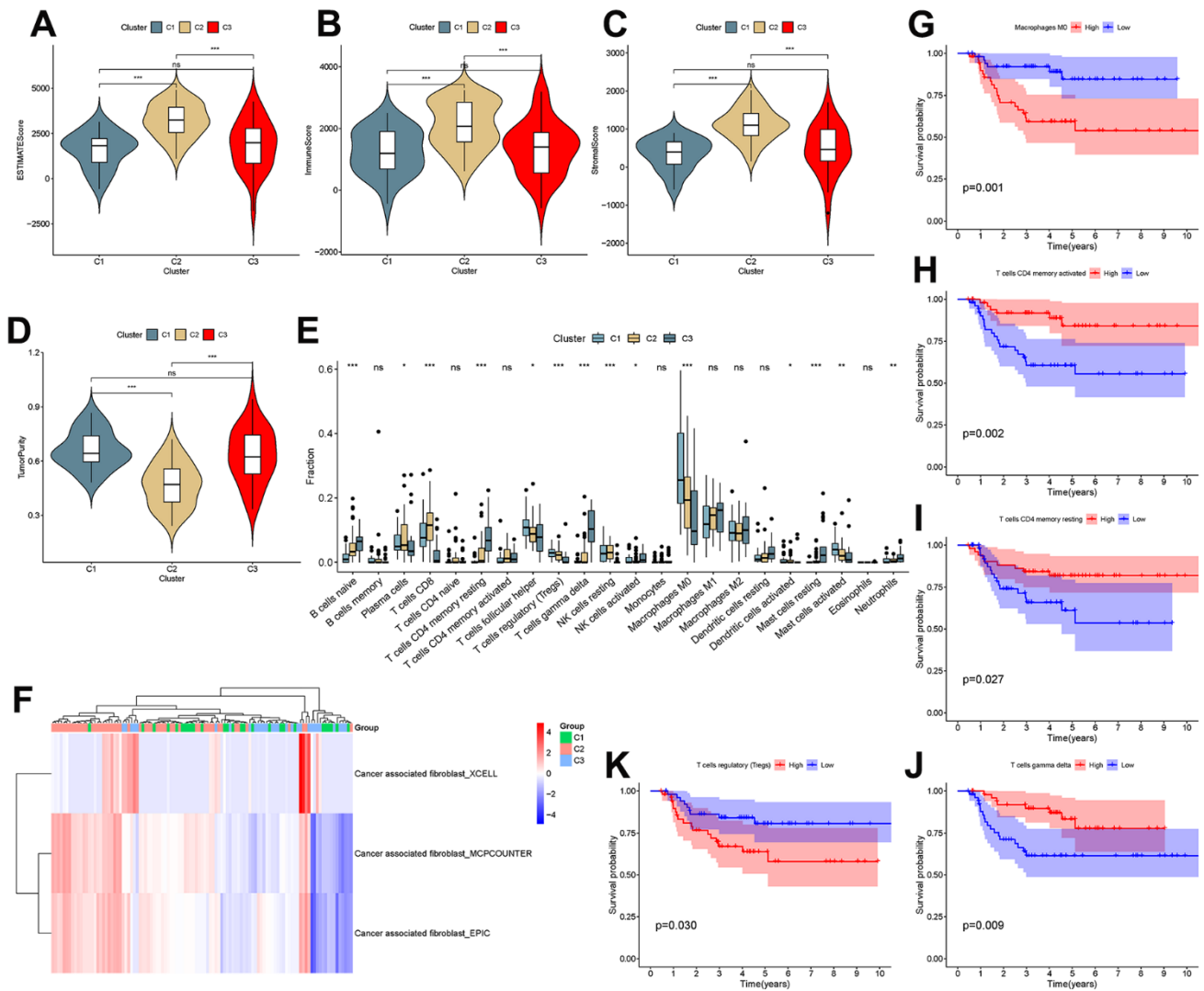


Figure 8. Analysis of the tumor microenvironment among molecular subtypes. (A–D) ESTIMATE score, immune score, stromal score and tumor purity among clusters. (E) Comparison of infiltration of 22 immune cells among clusters. (F–J) Immune cells associated with survival in triple-negative breast cancer patients.

Table 2. CAF-related genes to construct a prognostic model.

Gene name	Coefficient	HR	HR.95 L	HR.95H	pvalue
SRGN	1.55	4.72	1.47	15.12	0.01
ITM2A	1.17	3.21	1.02	10.09	0.05
CD74	-0.59	0.55	0.27	1.15	0.11
SAMSN1	-2.58	0.08	0.01	0.52	0.01
CD2	0.91	2.49	0.69	8.91	0.16
IL7R	0.63	1.89	0.86	4.18	0.12
SASH3	1.75	5.76	1.31	25.44	0.02
TAGAP	1.94	6.96	1.42	34.10	0.02
CCR7	-2.22	0.11	0.02	0.53	0.01
ICOS	-3.42	0.03	0.01	0.18	<0.001
ALDOA	2.14	8.49	2.82	25.56	<0.001
MRFAP1	2.33	10.23	2.74	38.15	<0.001
LCLAT1	2.19	8.96	2.19	36.73	<0.001
PCNA	1.41	4.11	1.52	11.10	0.01
MAN2B1	-0.94	0.39	0.14	1.10	0.08
ZNF445	1.83	6.23	1.67	23.29	0.01

RNA sequencing in patients with intrahepatic cholangiocarcinoma divides CAFs into inflammatory and growth factor-rich and myCAF subpopulations, showing distinct ligand–receptor interactions. myCAFs express hyaluronate synthase 2, whereas type I collagen is not expressed and promotes intrahepatic cholangiocarcinoma [32]. The results were consistent with the results of our study. Six CAF subpopulations have been identified in 4T1 tumors, including 1) myofibroblast CAFs, enriched in α -SMA54re and several other contractile proteins; 2) CAFs with elevated expression of “inflammatory” inflammatory cytokines; and 3) CAF subpopulations expressing major histocompatibility complex class II proteins, which are typically expressed in antigen-presenting cells [23]. The expression of activated CAF-specific genes significantly correlated with patient survival [33]. Four major subgroups of fibroblasts were identified in gastric cancer samples: myofibroblasts, pericytes, extracellular matrix CAFs and immunomodulatory CAFs [34]. The function of each subgroup is different. High iCAF levels may indicate a high degree of malignancy, while high myCAF levels may indicate a poor response to treatment [30]. ICAFs appear to promote tumors better than myCAFs by producing chemokines and cytokines [35]. myCAFs may deposit large amounts of ECM to impede drug delivery [36]. In breast cancer, scRNA-seq in mouse models detected four distinct CAF phenotypes [37], termed vascular CAFs, matrix CAFs, cycling CAFs and developmental CAFs [37]. A scRNA-seq study using a triple-negative breast cancer mouse model further identified 2 CAF phenotypes, pCAF and sCAF, showed that CAFs functions change

with tumor progression [38]. Our results for CAFs subgroups differ from the previous studies in perspective of concern, which focused more on the relationship between CAFs and immune cells, whereas the present study identified the pathways and mechanisms by which CAFs acts. Lena Cords also provided 8 CAF classification: vascular CAFs, matrix CAFs, immune CAFs, tumor-like CAFs, interferon-response CAFs, antigen-presenting CAFs, dividing CAFs, and reticular-like CAFs [39]. For triple-negative breast cancer, 52% of patients have low peripheral lymphocyte infiltration and poor PD-1/PD-L1 treatment [40], which is supported by the single-cell clustering results of the presented research, so targeting the relevant pathways and mechanisms may create a breakthrough.

The CAFs marker FAP mentioned in the presented study has been the subject of several preclinical studies and is considered a promising target. Treatment of xenograft models of lung, pancreatic and head and neck cancers with a novel antibody-metanicinol conjugate, FAP5-DM1, resulted in long-term inhibition of tumor growth and complete regression with no signs of intolerance [41]. Depletion of FAP-positive stromal cells with the FAP-targeted immunotoxin α FAP-PE38 reduced the recruitment of tumor-infiltrating immune cells in the tumor microenvironment and inhibited tumor growth [42]. There are also novel FAP-targeted immunotherapies, such as DNA vaccination [43], chimeric antigen receptor T cells [44, 45] or lysing adenovirus [46, 47]. The FAP-specific antibody sibrotuzumab was

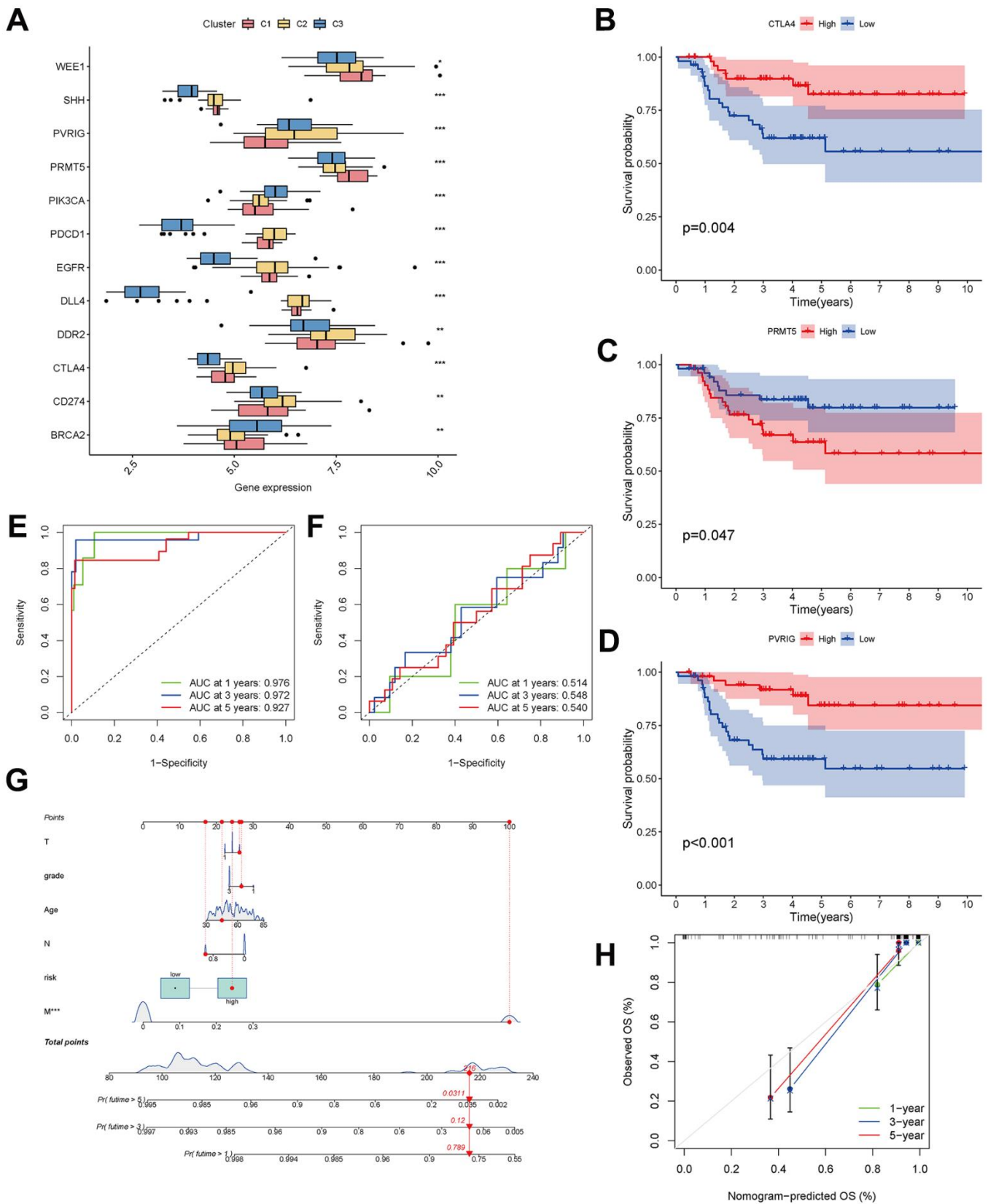


Figure 9. Analysis of CAF-related genes between molecular subtypes. (A) The expression of immune checkpoint-related genes among clusters. **(B–D)** The expression of checkpoint-related genes associated with survival in triple-negative breast cancer patients. **(E, F)** ROC curves for the training and testing sets. **(G)** Nomogram predicting patient survival prognosis. **(H)** Calibration curve of the nomogram.

considered clinically safe and effective in phase I trials in advanced cancer [48, 49]. However, no beneficial effects were shown in phase II trials in metastatic colorectal cancer [50].

Interestingly, we found a cluster of embryonic stem cell-like cells in these triple-negative breast cancer samples. Poorly differentiated basal-like breast cancer with similarities in gene expression to embryonic and induced pluripotent stem cells [51]. The extensive proliferation, migration and invasion required for mammaryogenesis do not occur in the resting adult breast, but they do resemble the processes that mediate breast cancer progression. Forced coexpression of luminal keratins 8 and 18 with vimentin in human breast cancer cells *in vitro* increases motility, invasiveness, and proliferation [52]. Similarly, basal-like breast cancers frequently exhibit an undifferentiated phenotype and coexpress keratin and vimentin proteins of the myoepithelium and tubular epithelium [53]. The epithelial-to-mesenchymal transition commonly observed during invasive tumorigenesis may represent a return to an embryonic-like state and may also promote a stem cell-like state in breast cells [54, 55]. Based on the above studies, we speculated that this fraction of cells may be embryonic stem cell-like mammary epithelial cells generated by the EMT process due to tumor tissue generation.

We identified 5 hub genes, including CD74, SASH3, CD2, TAGAP and CCR7. CD74 is a type II transmembrane protein that is primarily expressed on

antigen-presenting cells and functions intracellularly as a major histocompatibility complex class II chaperone [56, 57]. Its expression is highly correlated with tumor progression, especially in chronic lymphocytic leukemia [58]. SASH3 has also been reported by other bioinformatic analyses as a tumor microenvironment-related gene with prognostic value for breast cancer [59]. SASH3 is important for T-cell proliferation, activation and cell survival, and deficiency or mutation of SASH3 may lead to a decrease in CD4 T-cell lymphocytes, a decrease in T-cell proliferation, a decrease in cell cycle progression and an increase in T-cell apoptosis [60]. Our findings suggested that T-cell infiltration differed among clusters. CD2 promotes adhesion and signaling and counteracts human T-cell depletion. Low CD2 expression is shown in patients with endometrial or ovarian cancer. CD2 downregulation may attenuate antitumor T-cell responses and have implications for checkpoint immunotherapy [61]. In weighted gene co-expression network analysis, TAGAP was considered a key marker associated with prognosis in bladder cancer [62], endometrial cancer [63], and prostate cancer [64]. In tumors, CCL chemokine expression is present in cancer-associated fibroblasts [65]. The CCR7 chemokine axis plays an important role in controlling the migration of tumor cells to the lymphatic system and metastasis and therefore contributes to cancer expansion [66]. In esophageal squamous carcinoma, activation of CCR7 increases VEGF-A expression in cancer cells by increasing angiogenesis through activation of NF- κ B [67]. It also causes EMT and migration of cancer cells [68, 69].

In this study, three CAF subgroups and marker genes were identified, their functions were annotated, and a prognostic model based on CAF-related genes was constructed. However, some limitations in our study should be acknowledged. Clinical information corresponding to single-cell data was not available for this study, resulting in failure to analyze the relationship between CAF percentage and survival in individual samples. The functions of the CAF subgroup proposed in this study need to be verified by biological tests. In the future, we will perform flow cytometry to classify CAFs and investigate their effects on the proliferation, migration and apoptosis of TNBC cells.

CONCLUSIONS

In summary, we identified three subgroups: prCAFs (marked by DCN), myCAFs (marked by FAP) and emCAFs (marked by RGS5). Their functions are promoting angiogenesis, constructing matrix composition and collagen production, having myofibroblast

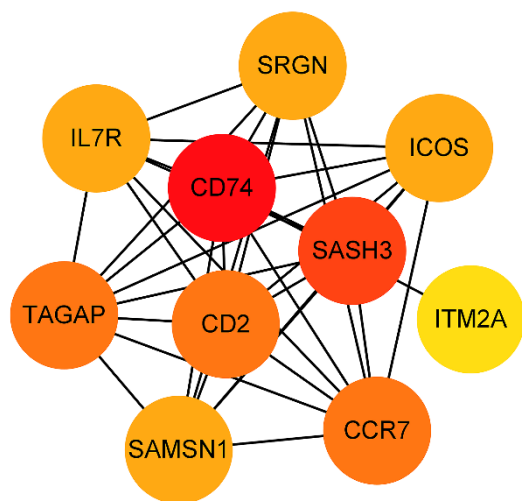


Figure 10. Hub genes obtained from the prognostic predictive model. Note: The dots represent proteins, the lines represent the interaction between proteins, and the color of the dots represents the degree (red: high; yellow: low).

characteristics and inducing endocrine resistance. CAFs are a source of inter- and intra-tumor heterogeneity in TNBC tumors. A prognostic model based on CAF-related genes was constructed and verified by a testing set. Five hub genes, CD74, SASH3, CD2, TAGAP and CCR7, served as significant marker genes with prognostic value.

AUTHOR CONTRIBUTIONS

Study conception and design: Wenping Lu, Xiaoqing Wu; data collection: Dongni Zhang, Yongjia Cui, Zhili Zhuo; analysis and interpretation of results: Xiaoqing Wu, Cuihong Jiang; draft manuscript preparation: Xiaoqing Wu and Cuihong Jiang. All authors reviewed the results and approved the final version of the manuscript.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest to report regarding the present study.

FUNDING

The authors received no specific funding for this study.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021; 71:209–49. <https://doi.org/10.3322/caac.21660> PMID:33538338
2. So JY, Ohm J, Lipkowitz S, Yang L. Triple negative breast cancer (TNBC): Non-genetic tumor heterogeneity and immune microenvironment: Emerging treatment options. *Pharmacol Ther.* 2022; 237:108253. <https://doi.org/10.1016/j.pharmthera.2022.108253> PMID:35872332
3. Karami Fath M, Azargoonjahromi A, Kiani A, Jalalifar F, Osati P, Akbari Oryani M, Shakeri F, Nasirzadeh F, Khalesi B, Nabi-Afjadi M, Zalpoor H, Mard-Soltani M, Payandeh Z. The role of epigenetic modifications in drug resistance and treatment of breast cancer. *Cell Mol Biol Lett.* 2022; 27:52. <https://doi.org/10.1186/s11658-022-00344-6> PMID:35764927
4. de Jong VMT, Wang Y, Ter Hoeve ND, Opdam M, Stathonikos N, Józwiak K, Hauptmann M, Cornelissen S, Vreuls W, Rosenberg EH, Koop EA, Varga Z, van Deurzen CH, et al. Prognostic Value of Stromal Tumor-Infiltrating Lymphocytes in Young, Node-Negative, Triple-Negative Breast Cancer Patients Who Did Not Receive (neo)Adjuvant Systemic Therapy. *J Clin Oncol.* 2022; 40:2361–74. <https://doi.org/10.1200/JCO.21.01536> PMID:35353548
5. Chen K, Lu P, Beeraka NM, Sukocheva OA, Madhunapantula SV, Liu J, Sinelnikov MY, Nikolenko VN, Bulygin KV, Mikhaleva LM, Reshetov IV, Gu Y, Zhang J, et al. Mitochondrial mutations and mitoepigenetics: Focus on regulation of oxidative stress-induced responses in breast cancers. *Semin Cancer Biol.* 2022; 83:556–69. <https://doi.org/10.1016/j.semcancer.2020.09.012> PMID:33035656
6. McAndrews KM, Chen Y, Darpolor JK, Zheng X, Yang S, Carstens JL, Li B, Wang H, Miyake T, Correa de Sampaio P, Kirtley ML, Natale M, Wu CC, et al. Identification of Functional Heterogeneity of Carcinoma-Associated Fibroblasts with Distinct IL6-Mediated Therapy Resistance in Pancreatic Cancer. *Cancer Discov.* 2022; 12:1580–97. <https://doi.org/10.1158/2159-8290.CD-20-1484> PMID:35348629
7. Liedtke C, Mazouni C, Hess KR, André F, Tordai A, Mejia JA, Symmans WF, Gonzalez-Angulo AM, Hennessy B, Green M, Cristofanilli M, Hortobagyi GN, Pusztai L. Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. *J Clin Oncol.* 2008; 26:1275–81. <https://doi.org/10.1200/JCO.2007.14.4147> PMID:18250347
8. Hoadley KA, Siegel MB, Kanchi KL, Miller CA, Ding L, Zhao W, He X, Parker JS, Wendl MC, Fulton RS, Demeter RT, Wilson RK, Carey LA, et al. Tumor Evolution in Two Patients with Basal-like Breast Cancer: A Retrospective Genomics Study of Multiple Metastases. *PLoS Med.* 2016; 13:e1002174. <https://doi.org/10.1371/journal.pmed.1002174> PMID:27923045
9. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Børresen-Dale AL, Boyault S, Burkhardt B, Butler AP, et al, and Australian Pancreatic Cancer Genome Initiative, and ICGC Breast Cancer Consortium, and ICGC MMLL-Seq Consortium, and ICGC PedBrain. Signatures of mutational processes in human cancer. *Nature.* 2013; 500:415–21. <https://doi.org/10.1038/nature12477> PMID:23945592
10. Angus L, Smid M, Wilting SM, van Riet J, Van Hoeck A, Nguyen L, Nik-Zainal S, Steenbruggen TG, Tjan-Heijnen VC, Labots M, van Riel JM, Bloemendal HJ, Steeghs N,

- et al. The genomic landscape of metastatic breast cancer highlights changes in mutation and signature frequencies. *Nat Genet.* 2019; 51:1450–8.
<https://doi.org/10.1038/s41588-019-0507-7>
PMID:[31570896](https://pubmed.ncbi.nlm.nih.gov/31570896/)
11. Tavernari D, Battistello E, Dheilly E, Petruzzella AS, Mina M, Sordet-Dessimoz J, Peters S, Krueger T, Gfeller D, Riggi N, Oricchio E, Letovanec I, Ciriello G. Nongenetic Evolution Drives Lung Adenocarcinoma Spatial Heterogeneity and Progression. *Cancer Discov.* 2021; 11:1490–507.
<https://doi.org/10.1158/2159-8290.CD-20-1274>
PMID:[33563664](https://pubmed.ncbi.nlm.nih.gov/33563664/)
 12. Potter SS. Single-cell RNA sequencing for the study of development, physiology and disease. *Nat Rev Nephrol.* 2018; 14:479–92.
<https://doi.org/10.1038/s41581-018-0021-7>
PMID:[29789704](https://pubmed.ncbi.nlm.nih.gov/29789704/)
 13. Davis RT, Blake K, Ma D, Gabra MBI, Hernandez GA, Phung AT, Yang Y, Maurer D, Lefebvre AE, Alshetaiwi H, Xiao Z, Liu J, Locasale JW, et al. Transcriptional diversity and bioenergetic shift in human breast cancer metastasis revealed by single-cell RNA sequencing. *Nat Cell Biol.* 2020; 22:310–20.
<https://doi.org/10.1038/s41556-020-0477-0>
PMID:[32144411](https://pubmed.ncbi.nlm.nih.gov/32144411/)
 14. Jahandideh A, Yarizadeh M, Noei-Khesht Masjedi M, Fatehnejad M, Jahandideh R, Soheili R, Eslami Y, Zokaei M, Ahmadvand A, Ghalamkarpour N, Kumar Pandey R, Nabi Afjadi M, Payandeh Z. Macrophage's role in solid tumors: two edges of a sword. *Cancer Cell Int.* 2023; 23:150.
<https://doi.org/10.1186/s12935-023-02999-3>
PMID:[37525217](https://pubmed.ncbi.nlm.nih.gov/37525217/)
 15. Razi S, Haghparast A, Chodari Khameneh S, Ebrahimi Sadrabadi A, Aziziyan F, Bakhtiyari M, Nabi-Afjadi M, Tarhriz V, Jalili A, Zalpoor H. The role of tumor microenvironment on cancer stem cell fate in solid tumors. *Cell Commun Signal.* 2023; 21:143.
<https://doi.org/10.1186/s12964-023-01129-w>
PMID:[37328876](https://pubmed.ncbi.nlm.nih.gov/37328876/)
 16. Gentric G, Mieulet V, Mechta-Grigoriou F. Heterogeneity in Cancer Metabolism: New Concepts in an Old Field. *Antioxid Redox Signal.* 2017; 26:462–85.
<https://doi.org/10.1089/ars.2016.6750>
PMID:[27228792](https://pubmed.ncbi.nlm.nih.gov/27228792/)
 17. Shi W, Chen Z, Liu H, Miao C, Feng R, Wang G, Chen G, Chen Z, Fan P, Pang W, Li C. COL11A1 as a novel biomarker for breast cancer with machine learning and immunohistochemistry validation. *Front Immunol.* 2022; 13:937125.
<https://doi.org/10.3389/fimmu.2022.937125>
PMID:[36389832](https://pubmed.ncbi.nlm.nih.gov/36389832/)
 18. Toullec A, Gerald D, Despouy G, Bourachot B, Cardon M, Lefort S, Richardson M, Rigail G, Parrini MC, Lucchesi C, Bellanger D, Stern MH, Dubois T, et al. Oxidative stress promotes myofibroblast differentiation and tumour spreading. *EMBO Mol Med.* 2010; 2:211–30.
<https://doi.org/10.1002/emmm.201000073>
PMID:[20535745](https://pubmed.ncbi.nlm.nih.gov/20535745/)
 19. Ruhland MK, Loza AJ, Capietto AH, Luo X, Knolhoff BL, Flanagan KC, Belt BA, Alspach E, Leahy K, Luo J, Schaffer A, Edwards JR, Longmore G, et al. Stromal senescence establishes an immunosuppressive microenvironment that drives tumorigenesis. *Nat Commun.* 2016; 7:11762.
<https://doi.org/10.1038/ncomms11762>
PMID:[27272654](https://pubmed.ncbi.nlm.nih.gov/27272654/)
 20. Yang X, Lin Y, Shi Y, Li B, Liu W, Yin W, Dang Y, Chu Y, Fan J, He R. FAP Promotes Immunosuppression by Cancer-Associated Fibroblasts in the Tumor Microenvironment via STAT3-CCL2 Signaling. *Cancer Res.* 2016; 76:4124–35.
<https://doi.org/10.1158/0008-5472.CAN-15-2973>
PMID:[27216177](https://pubmed.ncbi.nlm.nih.gov/27216177/)
 21. Wu SZ, Roden DL, Wang C, Holliday H, Harvey K, Cazet AS, Murphy KJ, Pereira B, Al-Eryani G, Bartonicek N, Hou R, Torpy JR, Junankar S, et al. Stromal cell diversity associated with immune evasion in human triple-negative breast cancer. *EMBO J.* 2020; 39:e104063.
<https://doi.org/10.15252/emboj.2019104063>
PMID:[32790115](https://pubmed.ncbi.nlm.nih.gov/32790115/)
 22. Costa A, Kieffer Y, Scholer-Dahirel A, Pelon F, Bourachot B, Cardon M, Sirven P, Magagna I, Fuhrmann L, Bernard C, Bonneau C, Kondratova M, Kuperstein I, et al. Fibroblast Heterogeneity and Immunosuppressive Environment in Human Breast Cancer. *Cancer Cell.* 2018; 33:463–79.e10.
<https://doi.org/10.1016/j.ccell.2018.01.011>
PMID:[29455927](https://pubmed.ncbi.nlm.nih.gov/29455927/)
 23. Sebastian A, Hum NR, Martin KA, Gilmore SF, Peran I, Byers SW, Wheeler EK, Coleman MA, Loots GG. Single-Cell Transcriptomic Analysis of Tumor-Derived Fibroblasts and Normal Tissue-Resident Fibroblasts Reveals Fibroblast Heterogeneity in Breast Cancer. *Cancers (Basel).* 2020; 12:1307.
<https://doi.org/10.3390/cancers12051307>
PMID:[32455670](https://pubmed.ncbi.nlm.nih.gov/32455670/)
 24. Musa M. Single-cell analysis on stromal fibroblasts in the microenvironment of solid tumours. *Adv Med Sci.* 2020; 65:163–9.
<https://doi.org/10.1016/j.advms.2019.12.001>
PMID:[31972467](https://pubmed.ncbi.nlm.nih.gov/31972467/)
 25. Gorchs L, Hellevik T, Bruun JA, Camilio KA, Al-Saad S, Stuge TB, Martinez-Zubiaurre I. Cancer-associated

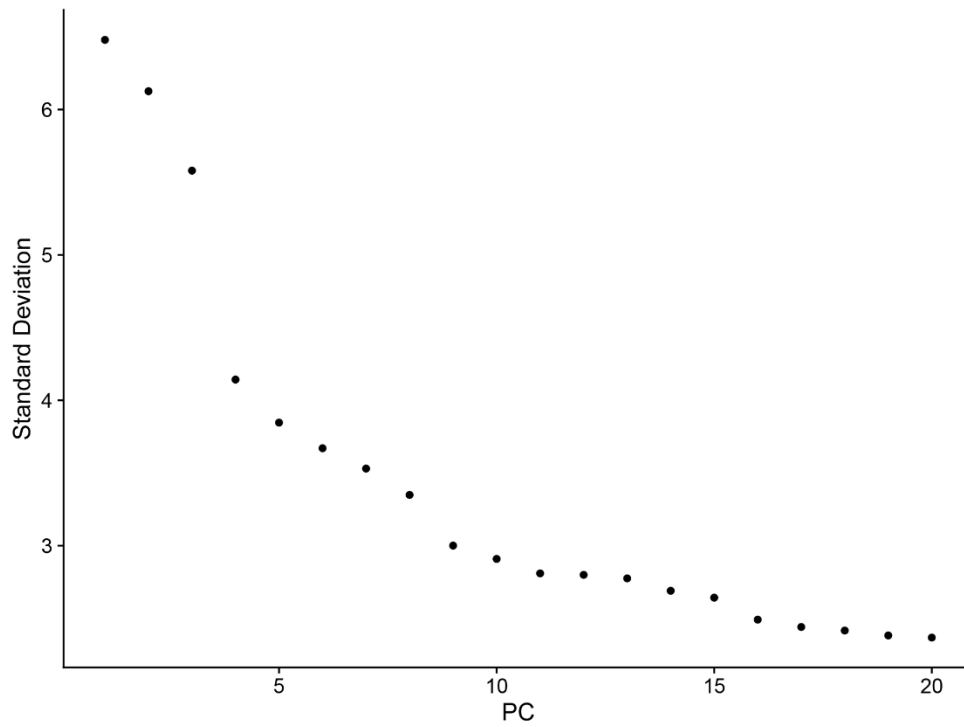
- fibroblasts from lung tumors maintain their immunosuppressive abilities after high-dose irradiation. *Front Oncol.* 2015; 5:87.
<https://doi.org/10.3389/fonc.2015.00087>
 PMID:26029659
26. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, Hoffman P, Stoeckius M, Papalexi E, et al. Integrated analysis of multimodal single-cell data. *Cell.* 2021; 184:3573–87.e29.
<https://doi.org/10.1016/j.cell.2021.04.048>
 PMID:34062119
 27. Li H, Courtois ET, Sengupta D, Tan Y, Chen KH, Goh JLL, Kong SL, Chua C, Hon LK, Tan WS, Wong M, Choi PJ, Wee LJ, et al. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat Genet.* 2017; 49:708–18.
<https://doi.org/10.1038/ng.3818> PMID:28319088
 28. Izar B, Tirosh I, Stover EH, Wakiro I, Cuoco MS, Alter I, Rodman C, Leeson R, Su MJ, Shah P, Iwanicki M, Walker SR, Kanodia A, et al. A single-cell landscape of high-grade serous ovarian cancer. *Nat Med.* 2020; 26:1271–9.
<https://doi.org/10.1038/s41591-020-0926-0>
 PMID:32572264
 29. Chen X, Song E. Turning foes to friends: targeting cancer-associated fibroblasts. *Nat Rev Drug Discov.* 2019; 18:99–115.
<https://doi.org/10.1038/s41573-018-0004-1>
 PMID:30470818
 30. Han C, Liu T, Yin R. Biomarkers for cancer-associated fibroblasts. *Biomark Res.* 2020; 8:64.
<https://doi.org/10.1186/s40364-020-00245-w>
 PMID:33292666
 31. Chen Z, Zhou L, Liu L, Hou Y, Xiong M, Yang Y, Hu J, Chen K. Single-cell RNA sequencing highlights the role of inflammatory cancer-associated fibroblasts in bladder urothelial carcinoma. *Nat Commun.* 2020; 11:5077.
<https://doi.org/10.1038/s41467-020-18916-5>
 PMID:33033240
 32. Affo S, Nair A, Brundu F, Ravichandra A, Bhattacharjee S, Matsuda M, Chin L, Filliol A, Wen W, Song X, Decker A, Worley J, Caviglia JM, et al. Promotion of cholangiocarcinoma growth by diverse cancer-associated fibroblast subpopulations. *Cancer Cell.* 2021; 39:883.
<https://doi.org/10.1016/j.ccell.2021.05.010>
 PMID:34129825
 33. Lin W, Noel P, Borazanci EH, Lee J, Amini A, Han IW, Heo JS, Jameson GS, Fraser C, Steinbach M, Woo Y, Fong Y, Cridebring D, et al. Single-cell transcriptome analysis of tumor and stromal compartments of pancreatic ductal adenocarcinoma primary tumors and metastatic lesions. *Genome Med.* 2020; 12:80.
<https://doi.org/10.1186/s13073-020-00776-9>
 PMID:32988401
 34. Li X, Sun Z, Peng G, Xiao Y, Guo J, Wu B, Li X, Zhou W, Li J, Li Z, Bai C, Zhao L, Han Q, et al. Single-cell RNA sequencing reveals a pro-invasive cancer-associated fibroblast subgroup associated with poor clinical outcomes in patients with gastric cancer. *Theranostics.* 2022; 12:620–38.
<https://doi.org/10.7150/thno.60540> PMID:34976204
 35. Öhlund D, Handly-Santana A, Biffi G, Elyada E, Almeida AS, Ponz-Sarvise M, Corbo V, Oni TE, Hearn SA, Lee EJ, Chio II, Hwang CI, Tiriach H, et al. Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer. *J Exp Med.* 2017; 214:579–96.
<https://doi.org/10.1084/jem.20162024>
 PMID:28232471
 36. Biffi G, Oni TE, Spielman B, Hao Y, Elyada E, Park Y, Preall J, Tuveson DA. IL1-Induced JAK/STAT Signaling Is Antagonized by TGF β to Shape CAF Heterogeneity in Pancreatic Ductal Adenocarcinoma. *Cancer Discov.* 2019; 9:282–301.
<https://doi.org/10.1158/2159-8290.CD-18-0710>
 PMID:30366930
 37. Bartoschek M, Oskolkov N, Bocci M, Lötvot J, Larsson C, Sommarin M, Madsen CD, Lindgren D, Pekar G, Karlsson G, Ringnér M, Bergh J, Björklund Å, Pietras K. Spatially and functionally distinct subclasses of breast cancer-associated fibroblasts revealed by single cell RNA sequencing. *Nat Commun.* 2018; 9:5150.
<https://doi.org/10.1038/s41467-018-07582-3>
 PMID:30514914
 38. Friedman G, Levi-Galibov O, David E, Bornstein C, Giladi A, Dadiani M, Mayo A, Halperin C, Pevsner-Fischer M, Lavon H, Mayer S, Nevo R, Stein Y, et al. Cancer-associated fibroblast compositions change with breast cancer progression linking the ratio of S100A4+ and PDPN+ CAFs to clinical outcome. *Nat Cancer.* 2020; 1:692–708.
<https://doi.org/10.1038/s43018-020-0082-y>
 PMID:35122040
 39. Cords L, Tietscher S, Anzeneder T, Langwieder C, Rees M, de Souza N, Bodenmiller B. Cancer-associated fibroblast classification in single-cell and spatial proteomics data. *Nat Commun.* 2023; 14:4294.
<https://doi.org/10.1038/s41467-023-39762-1>
 PMID:37463917
 40. Cortes J, Rugo HS, Cescon DW, Im SA, Yusof MM, Gallardo C, Lipatov O, Barrios CH, Perez-Garcia J, Iwata H, Masuda N, Torregroza Otero M, Gokmen E, et al, and KEYNOTE-355 Investigators. Pembrolizumab plus

- Chemotherapy in Advanced Triple-Negative Breast Cancer. *N Engl J Med.* 2022; 387:217–26.
<https://doi.org/10.1056/NEJMoa2202809>
PMID:[35857659](https://pubmed.ncbi.nlm.nih.gov/35857659/)
41. Ostermann E, Garin-Chesa P, Heider KH, Kalat M, Lamche H, Puri C, Kerjaschki D, Rettig WJ, Adolf GR. Effective immunoconjugate therapy in cancer models targeting a serine protease of tumor fibroblasts. *Clin Cancer Res.* 2008; 14:4584–92.
<https://doi.org/10.1158/1078-0432.CCR-07-5211>
PMID:[18628473](https://pubmed.ncbi.nlm.nih.gov/18628473/)
 42. Fang J, Xiao L, Joo KI, Liu Y, Zhang C, Liu S, Conti PS, Li Z, Wang P. A potent immunotoxin targeting fibroblast activation protein for treatment of breast cancer in mice. *Int J Cancer.* 2016; 138:1013–23.
<https://doi.org/10.1002/ijc.29831> PMID:[26334777](https://pubmed.ncbi.nlm.nih.gov/26334777/)
 43. Loeffler M, Krüger JA, Niethammer AG, Reisfeld RA. Targeting tumor-associated fibroblasts improves cancer chemotherapy by increasing intratumoral drug uptake. *J Clin Invest.* 2006; 116:1955–62.
<https://doi.org/10.1172/JCI26532> PMID:[16794736](https://pubmed.ncbi.nlm.nih.gov/16794736/)
 44. Wang LC, Lo A, Scholler J, Sun J, Majumdar RS, Kapoor V, Antzis M, Cotner CE, Johnson LA, Durham AC, Solomides CC, June CH, Puré E, Albelda SM. Targeting fibroblast activation protein in tumor stroma with chimeric antigen receptor T cells can inhibit tumor growth and augment host immunity without severe toxicity. *Cancer Immunol Res.* 2014; 2:154–66.
<https://doi.org/10.1158/2326-6066.CIR-13-0027>
PMID:[24778279](https://pubmed.ncbi.nlm.nih.gov/24778279/)
 45. Kakarla S, Chow KK, Mata M, Shaffer DR, Song XT, Wu MF, Liu H, Wang LL, Rowley DR, Pfizenmaier K, Gottschalk S. Antitumor effects of chimeric receptor engineered human T cells directed to tumor stroma. *Mol Ther.* 2013; 21:1611–20.
<https://doi.org/10.1038/mt.2013.110>
PMID:[23732988](https://pubmed.ncbi.nlm.nih.gov/23732988/)
 46. Freedman JD, Duffy MR, Lei-Rossmann J, Muntzer A, Scott EM, Hagel J, Campo L, Bryant RJ, Verrill C, Lambert A, Miller P, Champion BR, Seymour LW, Fisher KD. An Oncolytic Virus Expressing a T-cell Engager Simultaneously Targets Cancer and Immunosuppressive Stromal Cells. *Cancer Res.* 2018; 78:6852–65.
<https://doi.org/10.1158/0008-5472.CAN-18-1750>
PMID:[30449733](https://pubmed.ncbi.nlm.nih.gov/30449733/)
 47. de Sostoa J, Fajardo CA, Moreno R, Ramos MD, Farrera-Sal M, Alemany R. Targeting the tumor stroma with an oncolytic adenovirus secreting a fibroblast activation protein-targeted bispecific T-cell engager. *J Immunother Cancer.* 2019; 7:19.
<https://doi.org/10.1186/s40425-019-0505-4>
PMID:[30683154](https://pubmed.ncbi.nlm.nih.gov/30683154/)
 48. Welt S, Divgi CR, Scott AM, Garin-Chesa P, Finn RD, Graham M, Carswell EA, Cohen A, Larson SM, Old LJ. Antibody targeting in metastatic colon cancer: a phase I study of monoclonal antibody F19 against a cell-surface protein of reactive tumor stromal fibroblasts. *J Clin Oncol.* 1994; 12:1193–203.
<https://doi.org/10.1200/JCO.1994.12.6.1193>
PMID:[8201382](https://pubmed.ncbi.nlm.nih.gov/8201382/)
 49. Scott AM, Wiseman G, Welt S, Adjei A, Lee FT, Hopkins W, Divgi CR, Hanson LH, Mitchell P, Gansen DN, Larson SM, Ingle JN, Hoffman EW, et al. A Phase I dose-escalation study of sibtrotuzumab in patients with advanced or metastatic fibroblast activation protein-positive cancer. *Clin Cancer Res.* 2003; 9:1639–47.
PMID:[12738716](https://pubmed.ncbi.nlm.nih.gov/12738716/)
 50. Hofheinz RD, al-Batran SE, Hartmann F, Hartung G, Jäger D, Renner C, Tanswell P, Kunz U, Amelsberg A, Kuthan H, Stehle G. Stromal antigen targeting by a humanised monoclonal antibody: an early phase II trial of sibtrotuzumab in patients with metastatic colorectal cancer. *Onkologie.* 2003; 26:44–8.
<https://doi.org/10.1159/000069863> PMID:[12624517](https://pubmed.ncbi.nlm.nih.gov/12624517/)
 51. Mizuno H, Spike BT, Wahl GM, Levine AJ. Inactivation of p53 in breast cancers correlates with stem cell transcriptional signatures. *Proc Natl Acad Sci USA.* 2010; 107:22745–50.
<https://doi.org/10.1073/pnas.1017001108>
PMID:[21149740](https://pubmed.ncbi.nlm.nih.gov/21149740/)
 52. Hendrix MJ, Seftor EA, Seftor RE, Trevor KT. Experimental co-expression of vimentin and keratin intermediate filaments in human breast cancer cells results in phenotypic interconversion and increased invasive behavior. *Am J Pathol.* 1997; 150:483–95.
PMID:[9033265](https://pubmed.ncbi.nlm.nih.gov/9033265/)
 53. Livasy CA, Reading FC, Moore DT, Boggess JF, Lininger RA. EGFR expression and HER2/neu overexpression/amplification in endometrial carcinosarcoma. *Gynecol Oncol.* 2006; 100:101–6.
<https://doi.org/10.1016/j.ygyno.2005.07.124>
PMID:[16157366](https://pubmed.ncbi.nlm.nih.gov/16157366/)
 54. Spike BT, Engle DD, Lin JC, Cheung SK, La J, Wahl GM. A mammary stem cell population identified and characterized in late embryogenesis reveals similarities to human breast cancer. *Cell Stem Cell.* 2012; 10:183–97.
<https://doi.org/10.1016/j.stem.2011.12.018>
PMID:[22305568](https://pubmed.ncbi.nlm.nih.gov/22305568/)
 55. Qi J, Yan J, Idrees M, Almutairi SM, Rasheed RA, Hussein UA, Abdel-Maksoud MA, Wang R, Huang J, Huang C, Wang N, Huang D, Hui Y, Li C. Five EMT-Related Gene Signatures Predict Acute Myeloid Leukemia Patient Outcome. *Dis Markers.* 2022; 2022:7826393.

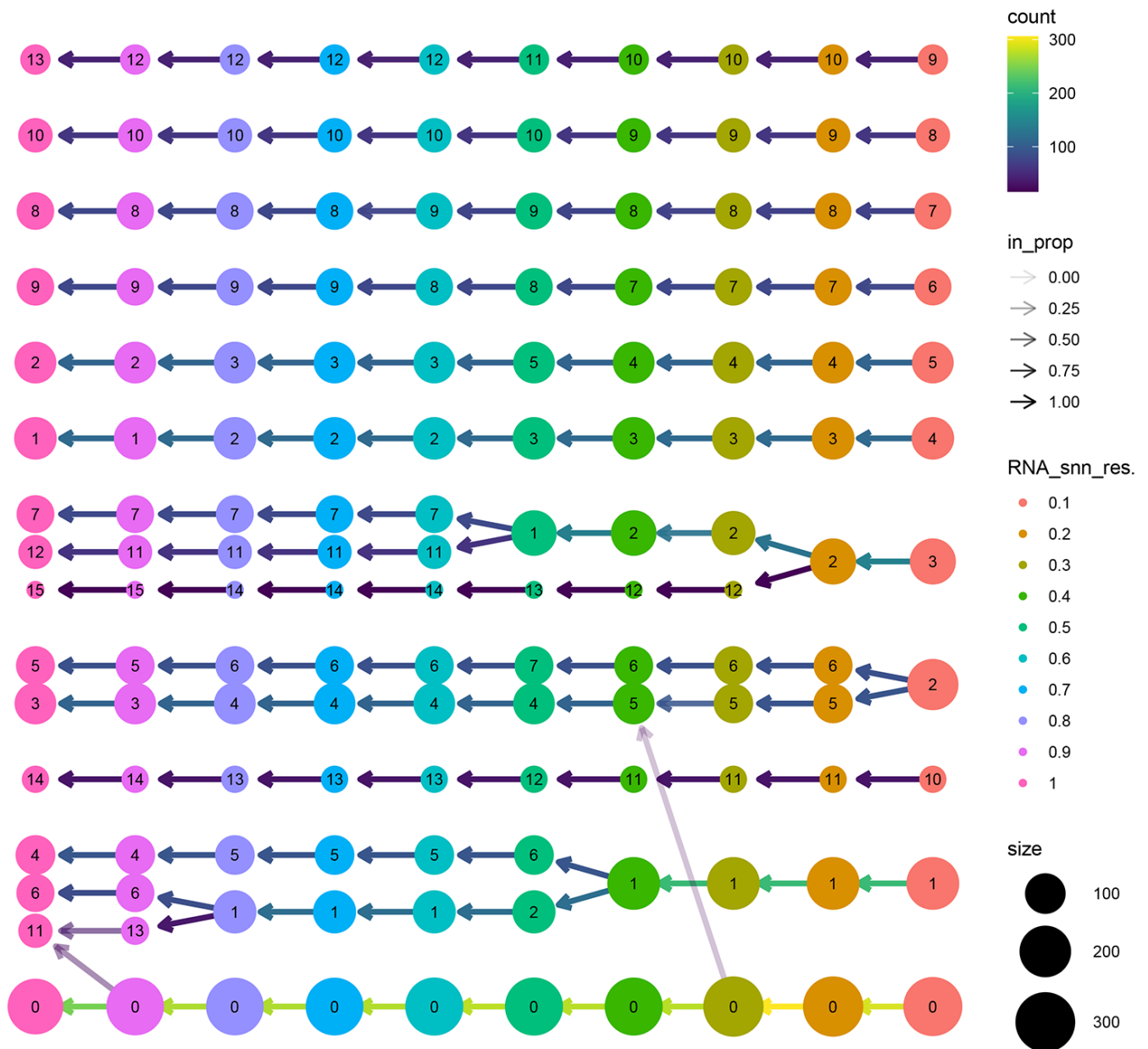
- <https://doi.org/10.1155/2022/7826393>
PMID:[36246561](https://pubmed.ncbi.nlm.nih.gov/36246561/)
56. Stumptner-Cuvelette P, Benaroch P. Multiple roles of the invariant chain in MHC class II function. *Biochim Biophys Acta*. 2002; 1542:1–13.
[https://doi.org/10.1016/s0167-4889\(01\)00166-5](https://doi.org/10.1016/s0167-4889(01)00166-5)
PMID:[11853874](https://pubmed.ncbi.nlm.nih.gov/11853874/)
57. Gu Y, Wu L, Hameed Y, Nabi-Afjadi M. Overcoming the challenge: cell-penetrating peptides and membrane permeability. *Biomaterials and Biosensors*. 2023; 2:7.
<https://doi.org/10.58567/bab02010002>
58. Griffin DO, Holodick NE, Rothstein TL. Human B1 cells in umbilical cord and adult peripheral blood express the novel phenotype CD20+ CD27+ CD43+ CD70-. *J Exp Med*. 2011; 208:67–80.
<https://doi.org/10.1084/jem.20101499>
PMID:[21220451](https://pubmed.ncbi.nlm.nih.gov/21220451/)
59. Shi Y, Chen S, Xing H, Jiang G, Wu N, Liu Q, Sakamoto N, Kuno T, Sugiura R, Xiao Q, Jin F, Fang Y, Yao F. Comprehensive Analysis of Prognostic Microenvironment-Related Genes in Invasive Breast Cancer. *Front Oncol*. 2022; 11:576911.
<https://doi.org/10.3389/fonc.2021.576911>
PMID:[35047378](https://pubmed.ncbi.nlm.nih.gov/35047378/)
60. Delmonte OM, Bergerson JRE, Kawai T, Kuehn HS, McDermott DH, Cortese I, Zimmermann MT, Dobbs AK, Bosticardo M, Fink D, Majumdar S, Palterer B, Pala F, et al. SASH3 variants cause a novel form of X-linked combined immunodeficiency with immune dysregulation. *Blood*. 2021; 138:1019–33.
<https://doi.org/10.1182/blood.2020008629>
PMID:[33876203](https://pubmed.ncbi.nlm.nih.gov/33876203/)
61. Demetriou P, Abu-Shah E, Valvo S, McCuaig S, Mayya V, Kvalvaag A, Starkey T, Korobchevskaya K, Lee LY, Friedrich M, Mann E, Kutuzov MA, Morotti M, et al, and Oxford IBD Cohort Investigators. A dynamic CD2-rich compartment at the outer edge of the immunological synapse boosts and integrates signals. *Nat Immunol*. 2020; 21:1232–43.
<https://doi.org/10.1038/s41590-020-0770-x>
PMID:[32929275](https://pubmed.ncbi.nlm.nih.gov/32929275/)
62. Figueroa JD, Han SS, Garcia-Closas M, Baris D, Jacobs EJ, Kogevinas M, Schwenn M, Malats N, Johnson A, Purdue MP, Caporaso N, Landi MT, Prokunina-Olsson L, et al. Genome-wide interaction study of smoking and bladder cancer risk. *Carcinogenesis*. 2014; 35:1737–44.
<https://doi.org/10.1093/carcin/bgu064>
PMID:[24662972](https://pubmed.ncbi.nlm.nih.gov/24662972/)
63. Guo C, Tang Y, Zhang Y, Li G. Mining TCGA Data for Key Biomarkers Related to Immune Microenvironment in Endometrial cancer by Immune Score and Weighted Correlation Network Analysis. *Front Mol Biosci*. 2021; 8:645388.
<https://doi.org/10.3389/fmolb.2021.645388>
PMID:[33869285](https://pubmed.ncbi.nlm.nih.gov/33869285/)
64. Zhao X, Lei Y, Li G, Cheng Y, Yang H, Xie L, Long H, Jiang R. Integrative analysis of cancer driver genes in prostate adenocarcinoma. *Mol Med Rep*. 2019; 19:2707–15.
<https://doi.org/10.3892/mmr.2019.9902>
PMID:[30720096](https://pubmed.ncbi.nlm.nih.gov/30720096/)
65. Yang T, Chen M, Yang X, Zhang X, Zhang Z, Sun Y, Xu B, Hua J, He Z, Song Z. Down-regulation of KLF5 in cancer-associated fibroblasts inhibit gastric cancer cells progression by CCL5/CCR5 axis. *Cancer Biol Ther*. 2017; 18:806–15.
<https://doi.org/10.1080/15384047.2017.1373219>
PMID:[28934010](https://pubmed.ncbi.nlm.nih.gov/28934010/)
66. Salem A, Alotaibi M, Mroueh R, Basheer HA, Afarinkia K. CCR7 as a therapeutic target in Cancer. *Biochim Biophys Acta Rev Cancer*. 2021; 1875:188499.
<https://doi.org/10.1016/j.bbcan.2020.188499>
PMID:[33385485](https://pubmed.ncbi.nlm.nih.gov/33385485/)
67. Cai QY, Liang GY, Zheng YF, Tan QY, Wang RW, Li K. CCR7 enhances the angiogenic capacity of esophageal squamous carcinoma cells *in vitro* via activation of the NF-κB/VEGF signaling pathway. *Am J Transl Res*. 2017; 9:3282–92.
PMID:[28804546](https://pubmed.ncbi.nlm.nih.gov/28804546/)
68. Cheng S, Han L, Guo J, Yang Q, Zhou J, Yang X. The essential roles of CCR7 in epithelial-to-mesenchymal transition induced by hypoxia in epithelial ovarian carcinomas. *Tumour Biol*. 2014; 35:12293–8.
<https://doi.org/10.1007/s13277-014-2540-6>
PMID:[25168373](https://pubmed.ncbi.nlm.nih.gov/25168373/)
69. Fu D, Hu Z, Xu X, Dai X, Liu Z. Key signal transduction pathways and crosstalk in cancer: Biological and therapeutic opportunities. *Transl Oncol*. 2022; 26:101510.
<https://doi.org/10.1016/j.tranon.2022.101510>
PMID:[36122506](https://pubmed.ncbi.nlm.nih.gov/36122506/)

SUPPLEMENTARY MATERIALS

Supplementary Figures

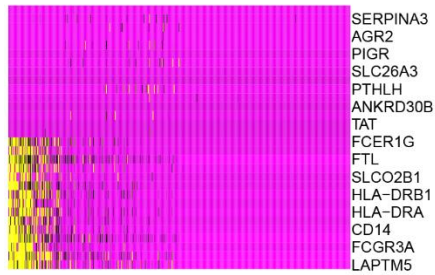


Supplementary Figure 1. Elbow plot of PCA in normal breast tissue data.

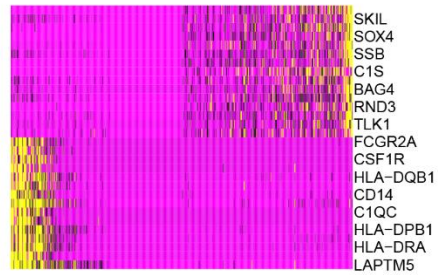


Supplementary Figure 2. Plot of identification of resolution in normal breast tissue data.

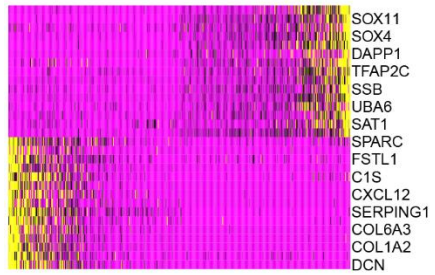
PC_1



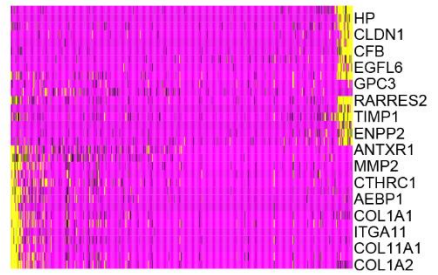
PC_2



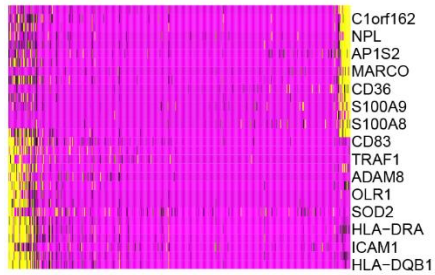
PC_3



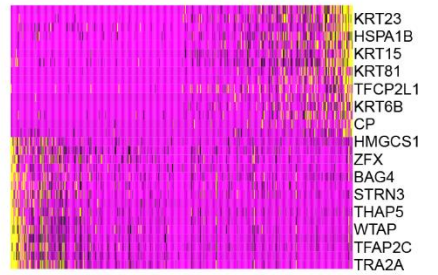
PC_4



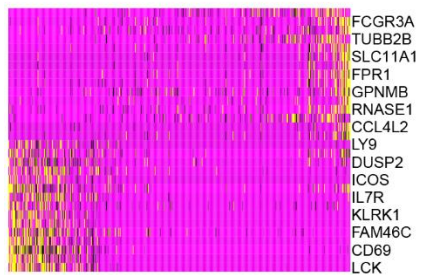
PC_5



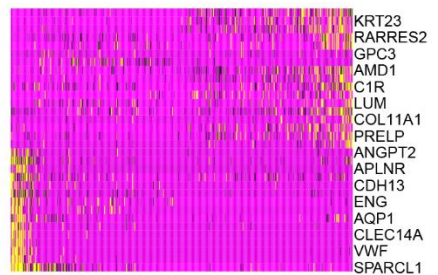
PC_6



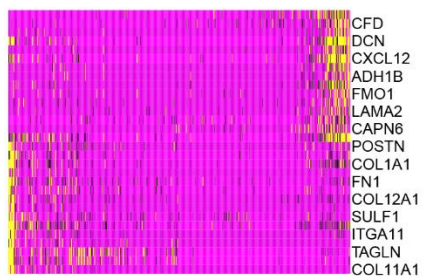
PC_7



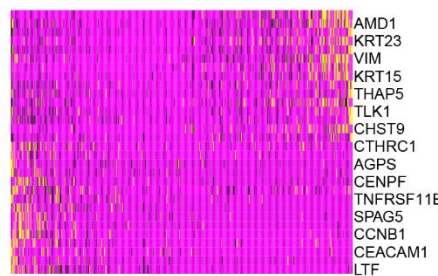
PC_8

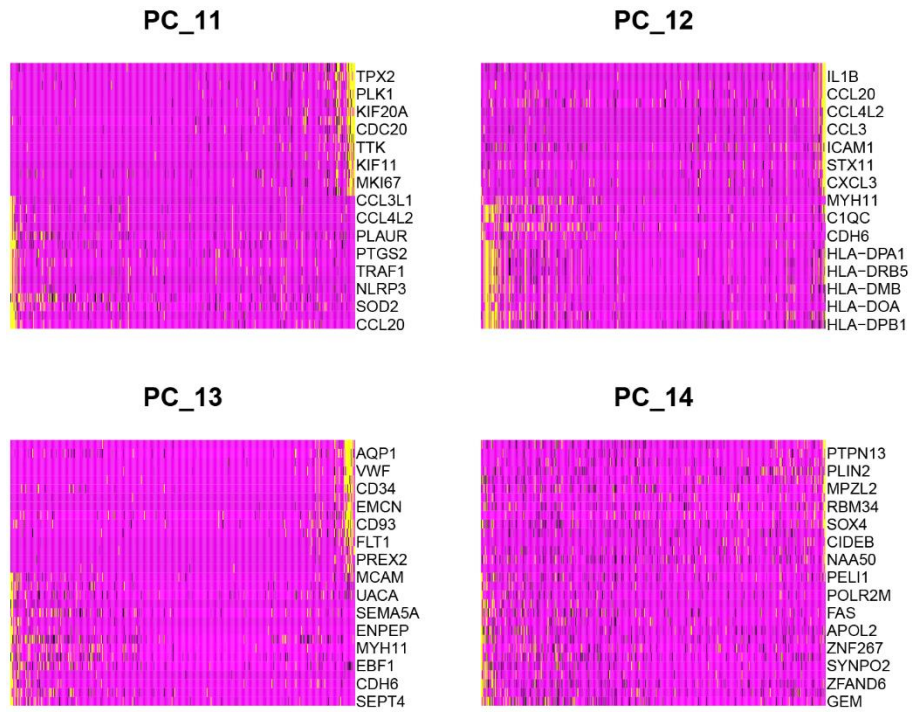


PC_9

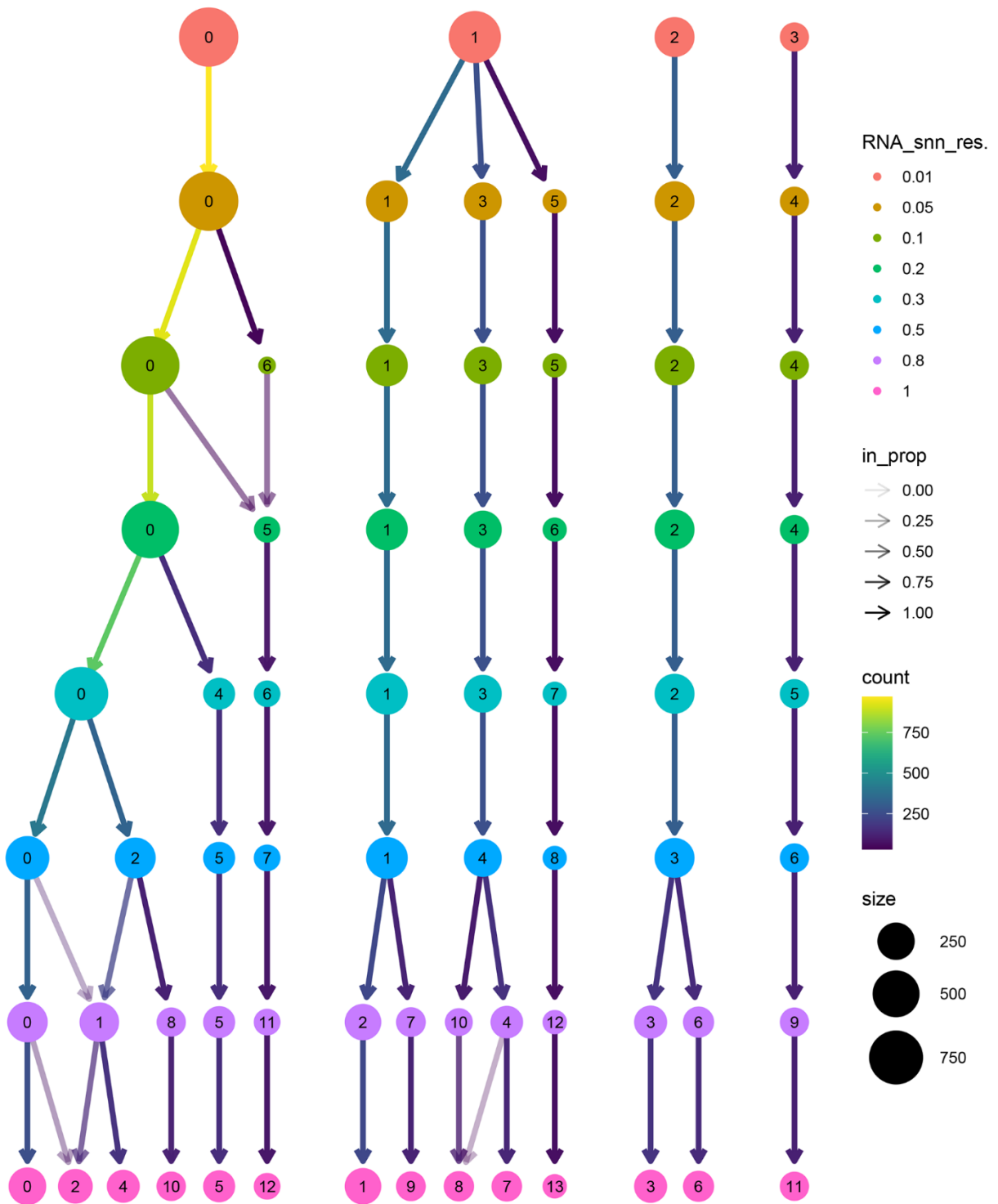


PC_10





Supplementary Figure 3. Heatmap plot of PCA in TNBC data.



Supplementary Figure 4. Plot of identification of resolution in TNBC data.