



Article

# Delving into the Heterogeneity of Different Breast Cancer Subtypes and the Prognostic Models Utilizing scRNA-Seq and Bulk RNA-Seq

Jieyun Xu, Shijie Qin, Yunmeng Yi, Hanyu Gao, Xiaoqi Liu, Fei Ma and Miao Guan \*

Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Wenyuan Road 1, Nanjing 210023, China

\* Correspondence: 08326@njnu.edu.cn

**Abstract:** Background: Breast cancer (BC) is the most common malignancy in women with high heterogeneity. The heterogeneity of cancer cells from different BC subtypes has not been thoroughly characterized and there is still no valid biomarker for predicting the prognosis of BC patients in clinical practice. Methods: Cancer cells were identified by calculating single cell copy number variation using the inferCNV algorithm. SCENIC was utilized to infer gene regulatory networks. CellPhoneDB software was used to analyze the intercellular communications in different cell types. Survival analysis, univariate Cox, least absolute shrinkage and selection operator (LASSO) regression and multivariate Cox analysis were used to construct subtype specific prognostic models. Results: Triple-negative breast cancer (TNBC) has a higher proportion of cancer cells than subtypes of HER2+ BC and luminal BC, and the specifically upregulated genes of the TNBC subtype are associated with antioxidant and chemical stress resistance. Key transcription factors (TFs) of tumor cells for three subtypes varied, and most of the TF-target genes are specifically upregulated in corresponding BC subtypes. The intercellular communications mediated by different receptor–ligand pairs lead to an inflammatory response with different degrees in the three BC subtypes. We establish a prognostic model containing 10 genes (risk genes: *ATP6AP1*, *RNF139*, *BASP1*, *ESR1* and *TSKU*; protective genes: *RPL31*, *PAK1*, *STARD10*, *TFPI2* and *SIAH2*) for luminal BC, seven genes (risk genes: *ACTR6* and *C2orf76*; protective genes: *DIO2*, *DCXR*, *NDUFA8*, *SULT1A2* and *AQP3*) for HER2+ BC, and seven genes (risk genes: *HPGD*, *CDC42* and *PGK1*; protective genes: *SMYD3*, *LMO4*, *FABP7* and *PRKRA*) for TNBC. Three prognostic models can distinguish high-risk patients from low-risk patients and accurately predict patient prognosis. Conclusions: Comparative analysis of the three BC subtypes based on cancer cell heterogeneity in this study will be of great clinical significance for the diagnosis, prognosis and targeted therapy for BC patients.



**Citation:** Xu, J.; Qin, S.; Yi, Y.; Gao, H.; Liu, X.; Ma, F.; Guan, M. Delving into the Heterogeneity of Different Breast Cancer Subtypes and the Prognostic Models Utilizing scRNA-Seq and Bulk RNA-Seq. *Int. J. Mol. Sci.* **2022**, *23*, 9936. <https://doi.org/10.3390/ijms23179936>

Academic Editor: Fabio Puglisi

Received: 20 July 2022

Accepted: 31 August 2022

Published: 1 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** breast cancer; cancer cells; intercellular communication; scRNA-seq; bulk RNA-seq; prognosis

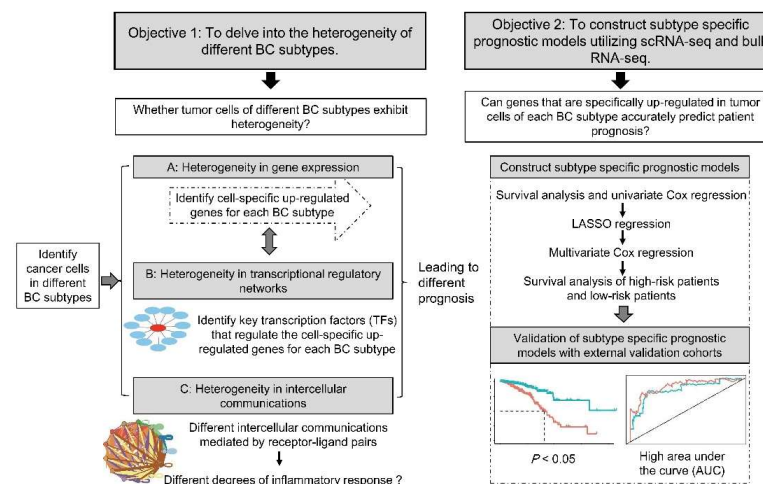
## 1. Introduction

Breast cancer (BC) is the most common malignancy in women, with 2.3 million new cases worldwide in 2020, accounting for one quarter of all cancer cases and one sixth of cancer deaths [1]. BC is a highly heterogeneous tumor and heterogeneity usually exists between subtypes [2]. Clinical classification of BC based on the expression of the estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2) has resulted in three broad subtypes, which are luminal BC, HER2-positive (HER2+) BC and triple-negative breast cancer (TNBC). Among all BC subtypes, luminal BC accounts for approximately 70% [3], TNBC accounts for 15–20% [4] and HER2+ BC accounts for 15–20% [5]. Each BC subtype has unique molecular characteristics, prognosis, clinical behavior, and treatment modalities [6–8]. For example, luminal BC patients have a better prognosis and are mainly treated with endocrine therapy and chemotherapy [9]. HER2+

BC has a poor prognosis and rapid progression, and is mainly treated with chemotherapy and anti-HER2 therapy [10]. TNBC is an aggressive subtype with terrible prognosis, strong drug resistance and high mortality. Current therapeutic options are still relatively limited for TNBC [11]. However, the cellular heterogeneity as drivers of cancer progression across different BC subtypes has not been well characterized. Meanwhile, there is still no valid biomarker for predicting the prognosis of BC patients for three subtypes in clinical practice.

Identifying new biomarkers in BC is crucial for accurate prognostic prediction and determining candidate targets of treatment. Delineation of subtypes should be considered when searching for new clinically and prognostically relevant biomarkers due to high heterogeneity across different BC subtypes. Of note, specifically expressed genes in cancer cells of different BC subtypes lead to different clinical phenotypes [12–14], which may explain the heterogeneity among different BC subtypes. However, most of the previous studies on BC prognosis-related biomarker mining are based on bulk RNA-seq [15], which only provide the average level of gene expression in different cell populations and do not take into account the widespread transcriptional heterogeneity in different cell populations [16,17].

Single-cell RNA sequencing (scRNA-seq) is a promising technology that allows transcriptome analysis of individual cells, identifying different cell types and precisely characterizing the transcription of each cell [18–20]. scRNA-seq provides great insight into the diversity of cell states and the heterogeneity of cell populations, making it a useful tool for dissecting the properties of multiple cell types in and around BC tumors. Researchers have used scRNA-seq to analyze the tumor heterogeneity of different subtypes of BC and identified cell clusters associated with poor prognosis or treatment response [21,22]. Single-cell profiles of different immune cells in the BC tumor microenvironment reveal specific immune cell subpopulations that could be potential immunotherapeutic targets [23,24]. In addition, intercellular communication is also one of the central issues in scRNA-seq of BC, as intercellular communication between tumor microenvironments and cells drives cancer progression and influences response to existing therapies [25–27]. Taken together, the objective of this study was (1) to delve into the heterogeneity of different BC subtypes; and (2) to construct subtype specific prognostic models utilizing scRNA-seq and bulk RNA-seq (Figure 1). In this work, we identified 13,517 cancer cells out of 73,866 cells based on copy number variations (CNVs) in scRNA-seq data from three BC subtype samples. Moreover, we identified specifically upregulated genes and key transcription factors (TFs) in cancer cells of three BC subtypes. Intercellular communications mediated by receptor–ligand pairs between cancer cells and other cells were analyzed. Then, we applied survival analysis, univariate Cox analysis, least absolute shrinkage and selection operator (LASSO) regression analysis and multivariate Cox analysis to construct a prognostic model of each BC subtype using cancer cell specific upregulated genes.



**Figure 1.** Workflow of delving into the heterogeneity of different breast cancer (BC) subtypes and constructing subtype specific prognostic models. In objective 1, a comparative analysis of three

BC subtypes was performed based on the heterogeneity of cancer cells in gene expression, transcriptional regulatory networks and cellular communication. In objective 2, survival analysis, univariate Cox, least absolute shrinkage and selection operator (LASSO) regression and multivariate Cox analysis were used to construct subtype specific prognostic models using cancer cell specific upregulated genes. The accuracy of the prognostic models was then validated using external validation cohorts.

## 2. Materials and Methods

### 2.1. Data Collection and Processing

The scRNA-seq data for this study was downloaded from the Gene Expression Omnibus (GEO) database (GSE176078), and 15 samples including 5 samples from each subtype (luminal BC, HER2+ BC and TNBC) with similar age and high cell numbers were selected (Table S1). Three broad BC subtypes are distinguished by the expression of ER, PR and HER2: luminal BC (ER+, PR+/-), HER2+ BC (HER2+, ER+/-, PR+/-) and TNBC (ER-, PR-). The breast cancer gene expression profiles and corresponding clinical information for the training cohort were obtained from The Cancer Genome Atlas (TCGA) database (<https://cancergenome.nih.gov/>) on 29 December 2021, containing 678 luminal BC patients, 159 HER2+ BC patients, and 116 TNBC patients after excluding patients with incomplete clinic pathological data. The validation cohort for the subtype of HER2+ BC was derived from the METABRIC dataset (<http://www.cbioportal.org/>) on 6 March 2022, including 220 HER2+ BC patients. The validation cohort for subtypes of luminal BC and TNBC were downloaded from the GEO database (GSE25066) containing 143 and 122 patients, respectively. All BC patients in the validation cohort have complete survival information.

### 2.2. Single-Cell Data Integration and Analysis

The Seurat package (version 4) was used to integrate scRNA-seq. Cells with the number of expressing genes below 200 or above 8000 were removed. In addition, cells with mitochondrial content greater than 20% were filtered out, and a total of 73,866 cells with single-cell transcriptome data were obtained for further analysis. The Harmony algorithm was used to eliminate batch effects. The top 2000 feature variables with the highest variance were selected for subsequent descending and clustering. Uniform manifold approximation and projection (UMAP) dimensionality reduction was used to project cells in two dimensions. The AddModuleScore function in the Seurat package was used to evaluate the gene set scores of the cells [28,29]. Gene sets in the inflammatory response hallmark were mainly obtained from the msigdb R package [30]. CellPhoneDB software was used to analyze intercellular communications and discover significant receptor–ligand pairs; only highly interacting receptor–ligand pairs were preserved with  $p$  value < 0.01 and means > 0.6 [31].

### 2.3. Identification of Cancer Cells by Calculating Cell Copy Number Variations

InferCNV [32] was used to infer large-scale CNVs in single-cell gene expression data. Control cells which contain 500 B cells, 500 T cells and 500 endothelial cells were randomly selected, serving as cells with a normal copy number, and the random seed was set to 123 to ensure reproducibility [33]. Normal epithelial cells and cancer cells were separated by checking the iterative clustering effect and calculating the copy number variation score [34,35].

### 2.4. Identification and Functional Analysis of Differentially Expressed Genes

The FindMarkers function in the Seurat package was used to find the differentially expressed genes with filter criteria ( $|\logFC| > 0.25$  and  $FDR < 0.01$ ). In each BC subtype, overlapped genes differentially upregulated in cancer cells compared with the normal epithelial cells group, and differentially upregulated compared with cancer cells of other subtypes group, were defined as cancer cell-specific upregulated genes. The Cluster-

Profiler [36] package was used for Gene Ontology (GO) biological process (BP) analysis. Significantly enriched GO BP terms were identified by a  $q$  value smaller than 0.05.

### 2.5. Inference of Gene Regulatory Networks (GRNs) in Tumor Cells

SCENIC [37] workflow was applied to normalize expression matrices of 5000 randomly selected epithelial cells to infer transcriptional regulatory networks in tumor cells of the three BC subtypes. Firstly, GENE Network Inference with Ensemble of trees (GENIE3) [38] was utilized to infer the co-expression modules between TFs and candidate target genes. Secondly, RcisTarget was used to identify TFs and the genes that they directly regulate. Finally, each formed regulon was scored in each cell using AUCell [39]. Key TFs for tumor cells of each BC subtype were identified with  $p < 0.01$ . The transcriptional regulatory networks of key TFs were visualized via Cytoscape software.

### 2.6. Construction of Prognostic Risk Model

Patients in the TCGA database with a follow-up time of more than 30 days were included to establish a prognostic risk model. To screen for genes associated with prognosis in BC subtypes, batch screening was performed by survival analysis and univariate Cox regression, and  $p$  values less than 0.05 were included. To screen for the most significant genes affecting overall survival (OS) in the three BC subtypes, LASSO regression analysis was used to remove genes with collinear correlations to reduce the number of prognostic genes. Finally, the remaining genes were subjected to multivariate Cox regression to screen independent prognostic genes and construct the prognostic model. The mathematical formula for calculating risk score is: Risk score =  $h_0(t) \times \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)$ . Herein,  $h_0(t)$  is a constant,  $x_n$  represents the prognostic genes,  $\exp$  represents the expression level of these genes,  $\beta_n$  represents the multiple regression coefficients of the prognostic genes. Survival and survminer packages were used to construct survival curves to evaluate survival differences between high-risk and low-risk groups. The discriminant ability of the model was evaluated by the receiver operating characteristic (ROC) curve [40].

### 2.7. Data Statistics and Visualization

All graph constructions in this study were performed using R package software (version 4). The Wilcox test was used to determine statistical differences in different groups. Survival curves were measured using the Kaplan–Meier method, and the statistical significance of differences was determined using the log-rank test. ROC curves were used to assess the predictive power of the prediction models.  $p < 0.05$  was defined as statistically significant.

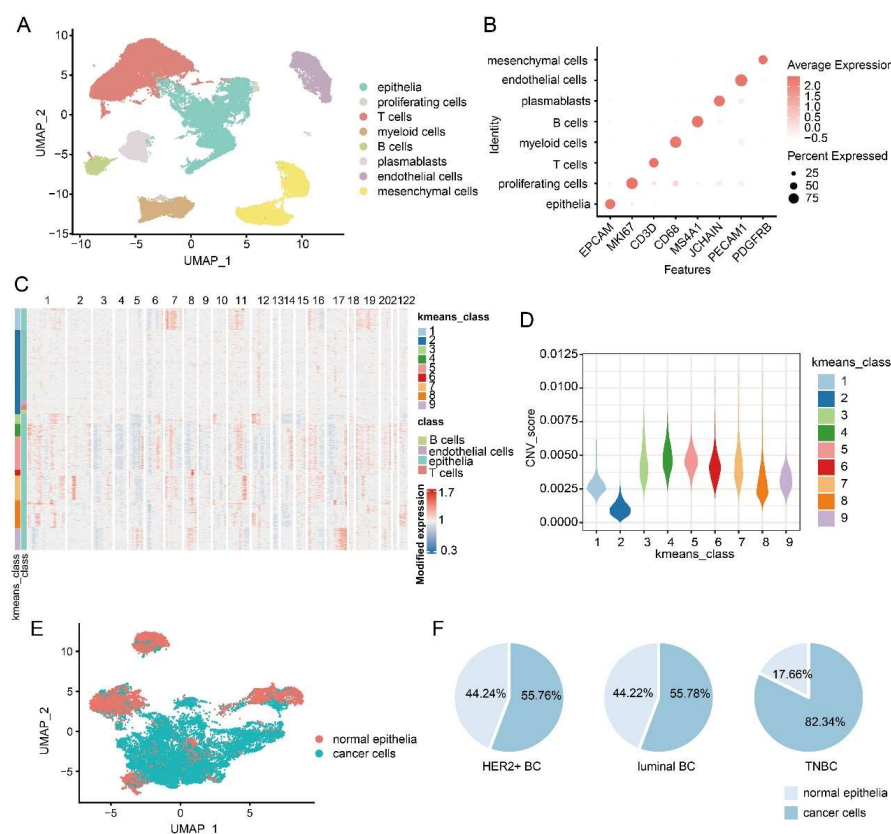
## 3. Results

### 3.1. TNBC Has a Higher Proportion of Tumor Cells

After integrating single-cell data from different BC subtypes, 73,866 cells from 15 samples were obtained, and eight major cell types including epithelial cells (EPCAM+), proliferating cells (MKI67+), T cells (CD3D+), myeloid cells (CD68+), B cells (MS4A1+), plasma cells (JCHAIN+), endothelial cells (PECAM1+) and mesenchymal cells (PDGFRB+) were identified by cell markers (Figure 2A,B). All markers and cell types have good specificity, indicating that the identification of cell types is accurate and efficient (Figure 2B). Cancer cells of BC originate from epithelial cells, so we defined cancer cells in all epithelial cells by calculating CNVs inferred from single-cell gene expression data with reference cells of immune cells (B cells and T cells) and endothelial cells. Unsupervised clustering with nine clusters was applied to distinguish cells with high CNVs and low CNVs (Figure 2C). As expected, CNVs levels in most epithelial clusters were obviously higher than reference cells. The second cluster containing reference cells (immune cells and endothelial cells) have the lowest coefficient of variation among the nine clusters (Figure 2D). Therefore, the epithelial cells in the second cluster were defined as normal epithelial cells, and the epithelial cells in other clusters were defined as cancer cells. A total of 13,517 cancer cells



out of 19,534 epithelial cells were identified, indicating that cancer cells accounted for a large proportion (69.2%) of BC epithelial cells (Figure 2E). Among the three subtypes of BC, TNBC had the highest proportion of cancer cells in epithelial cells, up to 82.34%, while luminal BC (55.78%) and HER2+ BC (55.76%) had a similar relative lower proportion than TNBC (Figure 2F).

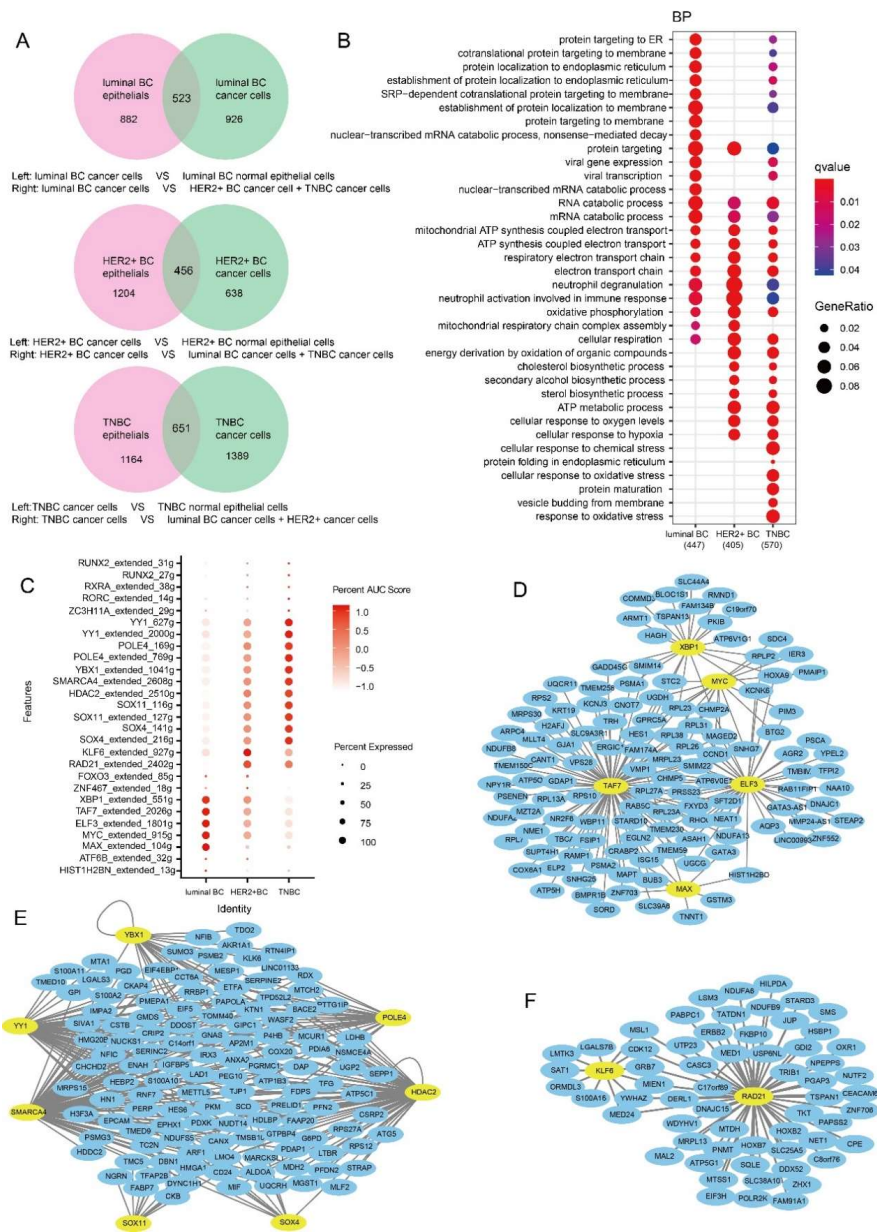


**Figure 2.** Identification of tumor cells from epithelia. (A) Uniform manifold approximation and projection (UMAP) cluster plot of different BC cell types. (B) Expression levels of cellular markers corresponding to different BC cell types. (C) Unsupervised clustering of inferred large-scale copy number variations (CNVs) to identify cancer cells from epithelia cells. Epithelial cells and reference cells (B cells, T cells and endothelial cells) are in the y-axis and chromosomal regions in the x-axis. (D) Violin plot showing the differences of CNVs scores among 9 clusters. (E) UMAP cluster plot showing the distribution of normal epithelial cells and cancer cells. (F) The proportion of cancer cells and normal epithelial cells of the three subtypes of BC.

### 3.2. Functions of Specifically Upregulated Genes and Variation of GRNs in Three BC Subtypes

The number of genes specifically upregulated in cancer cells was 523, 456 and 651 for subtypes of luminal BC, HER2+ BC and TNBC, respectively (Figure 3A). GO enrichment analysis demonstrated that the specific upregulated genes in the three BC subtypes were related to functions such as protein targeting, RNA catabolic processes, cellular respiration, mitochondrial ATP synthesis coupled electron transport, and the respiratory electron transport chain, all of which are essential for the survival of cancer cells (Figure 3B). In addition, these specific upregulated genes in three BC subtypes were also associated with neutrophil degranulation and neutrophil activation (Figure 3B). Neutrophils are the first responders to inflammation and infection, suggesting that there is an inflammatory response in BC. Moreover, pathways of protein targeting to membrane and the nuclear-transcribed mRNA catabolic process were enriched by specific upregulated genes of luminal BC cancer cells compared to HER2+ BC and TNBC. The specific upregulated genes in tumor cells of HER2+ BC and TNBC were associated with energy derivation by the oxidation of

organic compounds, ATP metabolic process, cholesterol biosynthetic process, secondary alcohol biosynthetic process, and sterol biosynthetic process (Figure 3B). These pathways modulated by specific upregulated genes of cancer cells demonstrated that cancer cells need more energy to synthesize more substances. Most importantly, we found that cancer cells of HER2+ BC and TNBC, especially TNBC, specifically expressed genes that responded to chemical and oxidative stress compared with luminal BC and HER2+ BC, which were related to the high drug resistance of TNBC (Figure 3B) [41–43].



**Figure 3.** Functions and gene regulatory networks of subtype-specific upregulated genes. (A) Intersection of differentially upregulated genes in cancer cells compared with normal epithelial cells (left) and differentially upregulated compared with other subtypes (right). (B) GO biological process function enrichment of specific upregulation genes in cancer cells of three BC subtypes. (C) Dotplot showing TFs specifically enriched in cancer cells of different BC subtypes. (D–F) Key TFs and their highly upregulated target genes in (D) luminal BC, (E) TNBC and (F) HER2+ BC tumor cells.

We used the SCENIC algorithm to analyze single cell transcriptional regulatory networks in tumor cells to identify the enriched key TFs of the three BC subtypes (Figure 3C).

We found that XBP1, TAF7, ELF3, MYC, MAX, etc., were more enriched in tumor cells of luminal BC than the other two subtypes. The highly active TFs in TNBC were YY1, YBX1, SOX11, SOX4, POLE4, SMARCA4 and HDAC2, etc., while in HER2+ BC they were FOXO3, ZNF467, and especially RAD21 and KLF6 (Figure 3C). Since TFs enriched in BC subtypes may regulate gene-specific expression, we further identified overlapping genes between the target genes of TFs and cancer cell-specific upregulated genes of different BC subtypes. There were 317, 151 and 512 upregulated target genes for subtypes of luminal BC, HER2+ BC and TNBC, respectively. Highly upregulated target genes with  $\text{avg\_log}_2\text{FC} > 0.5$  were visualized in Figure 3D,F.

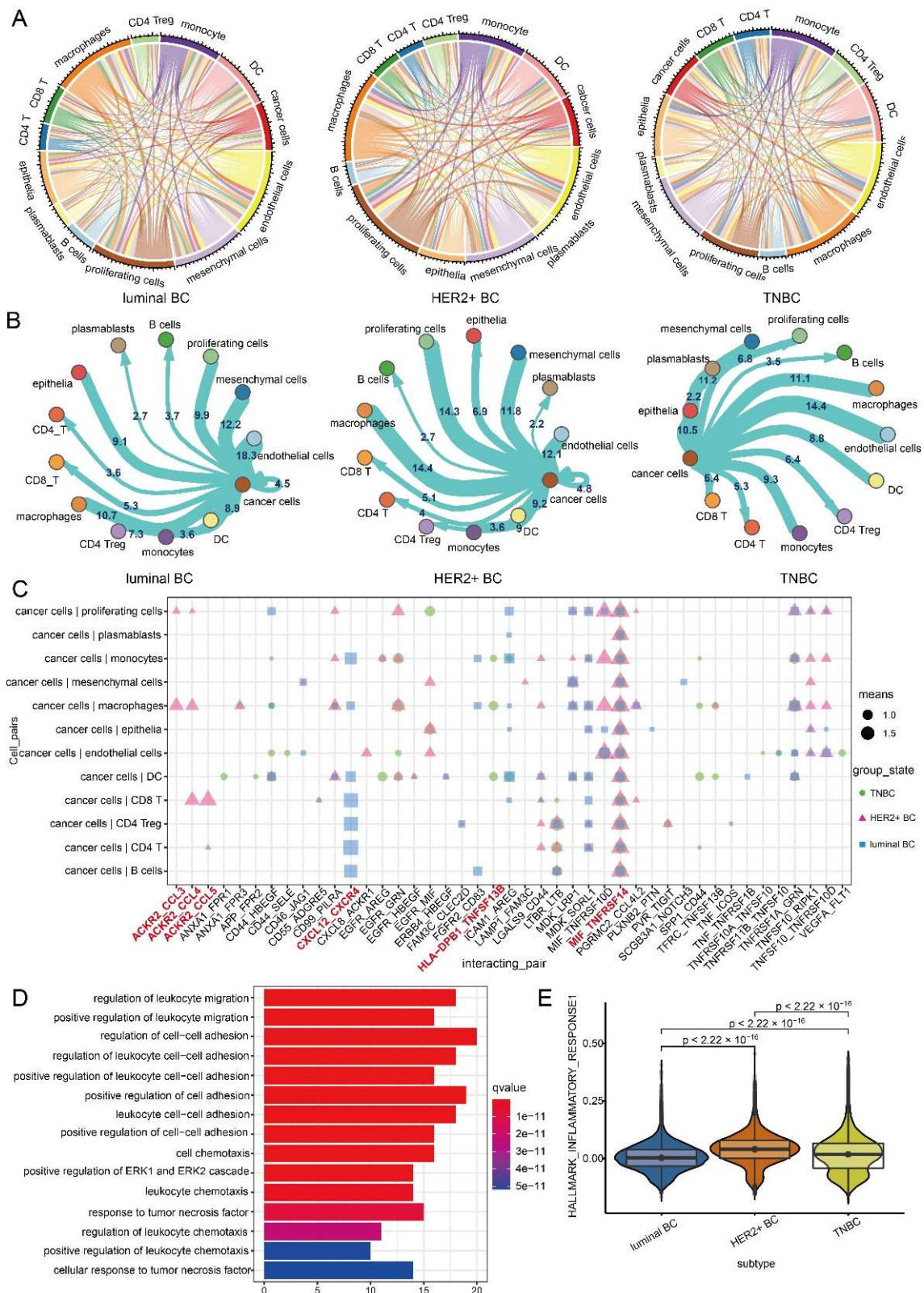
### 3.3. Intercellular Communication of Cancer Cells Leads to BC Inflammatory Microenvironment

We used the CellphoneDB algorithm to analyze the intercellular communication of three BC subtypes (Figures 4A and S1). The top cell types with the highest frequency of communication with cancer cells were mesenchymal cells and endothelial cells in luminal BC, macrophages and proliferating cells in HER2+ BC, and plasmablasts and endothelial cells in TNBC, respectively (Figure 4B). We also analyzed the ligand–receptor pairs with high interactions between cancer cells and other cells in three BC subtypes (Figure 4C). We found that these receptor and ligand genes were significantly enriched in leukocyte migration, cell chemotaxis, and tumor necrosis factor response (Figure 4D), and also the inflammatory character of the BC tumor microenvironment was consistent with our previous results that there is an inflammatory response in BC. Among the receptor–ligand pairs with high interactions, cancer cells of the three BC subtypes communicate with all other cell types through the MIF-TNFRSF14. Moreover, cancer cells of the luminal BC communicate with immune cells through CXCL12\_CXCR4. Cancer cells of HER2+ BC communicate with macrophages and CD8 T cells through receptor–ligand pairs of ACKR2-CCL3, ACKR2-CCL4 and ACKR2-CCL5 (Figure 4C). Cancer cells of TNBC communicate with macrophages and DC through the receptor–ligand pair of HLA–DPB1\_TNFSF13B (Figure 4C). Interestingly, most of these ligands and receptors are involved in the induction of inflammatory responses, which suggest that receptors and ligands that communicate between cancer cells and other cells play a key role in inducing inflammatory responses. We also speculate that the different ligand–receptor pairs that communicate between tumor cells and other cells to trigger an inflammatory response will also lead to different degrees of inflammatory response in the three BC subtypes. As expected, among the three BC subtypes, the inflammatory response scores of TNBC and HER2+ BC were significantly higher than those of luminal BC (Figure 4D).

### 3.4. Prognostic Model Construction Using Cancer Cell-Specific Upregulated Genes

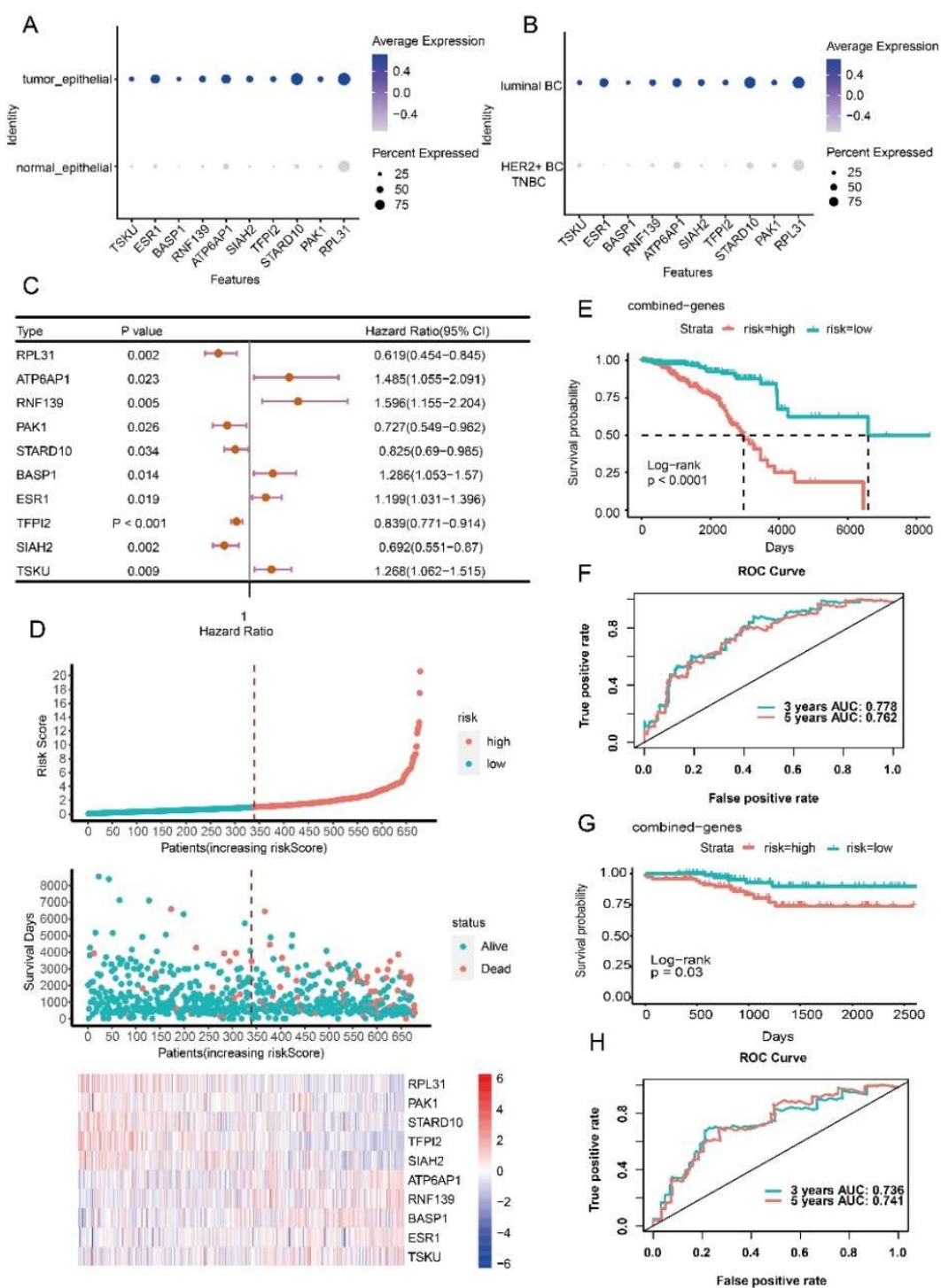
Three prognostic risk models for three BC subtypes (luminal BC, HER2+ BC and TNBC) were obtained using survival analysis, univariate Cox regression, LASSO regression analysis, and multivariate Cox regression. The prognostic risk model of luminal BC included ten genes (*RPL31*, *PAK1*, *STARD10*, *TFPI2*, *SIAH2*, *ATP6AP1*, *RNF139*, *BASP1*, *ESR1* and *TSKU*) (Figure S2). Seven genes (*DIO2*, *ACTR6*, *DCXR*, *NDUFA8*, *SULT1A2*, *C2orf76* and *AQP3*) were contained in the prognostic risk model of HER2+ BC (Figure S3). Seven combinations (*SMYD3*, *HPGD*, *LMO4*, *FABP7*, *CDC42*, *PGK1* and *PRKRA*) constructed the prognostic risk model of TNBC (Figure S4). These prognostic factors were specific differentially upregulated in cancer cells compared with normal epithelial cells and differentially upregulated compared with the cancer cells of the other subtypes (Figures 5A,B, 6A,B and 7A,B). Among genes in the prognostic risk model for luminal BC, *RPL31*, *PAK1*, *STARD10*, *TFPI2* and *SIAH2* were protective factors, while *ATP6AP1*, *RNF139*, *BASP1*, *ESR1* and *TSKU* were risk factors. *DIO2*, *DCXR*, *NDUFA8* and *AQP3* genes were protective factors of HER2+ BC, while *ACTR6*, *SULT1A2* and *C2orf76* genes were risk factors. *SMYD3*, *LMO4*, *FABP7* and *PRKRA* were protective factors of TNBC, while *HPGD*, *CDC42* and *PGK1* were risk factors (Figures 5C, 6C and 7C).



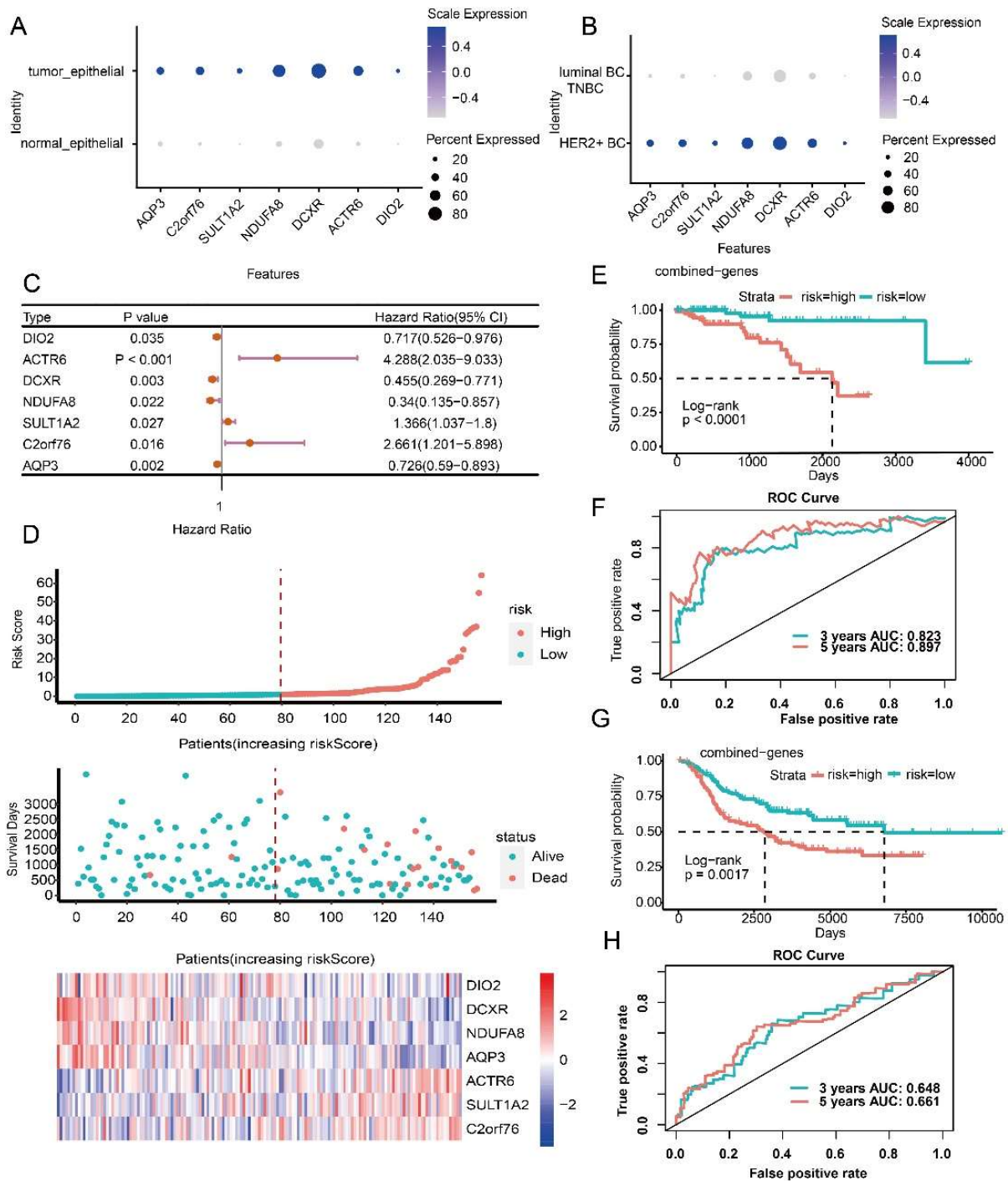


**Figure 4.** Cell communication of cancer cells leads to BC inflammatory microenvironment. (A) Overall intercellular communication profiles of the three BC subtypes. (B) Intercellular communication frequency maps of cancer cells of three BC subtypes. (C) Receptor–ligand pairs for communication between cancer cells and other cells of the three BC subtypes. (D) GO biological process enrichment analysis of receptor–ligand genes for communication between BC cancer cells and other cells. (E) Scores of cellular inflammatory responses in the three BC subtypes.

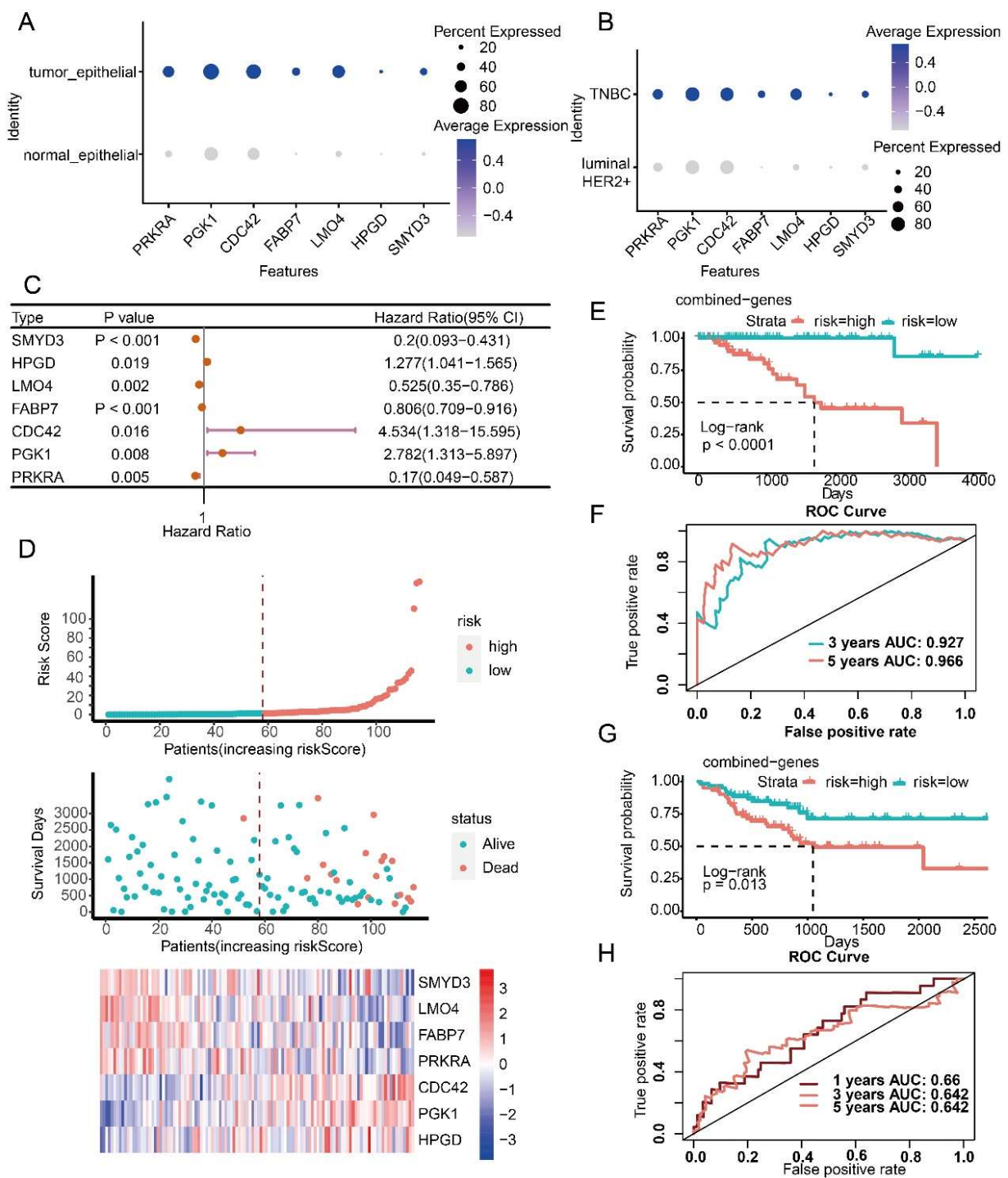




**Figure 5.** Construction and evaluation of the prognostic risk model for luminal BC. (A) Expression levels of prognostic factors between normal and cancer cells. (B) Expression levels of prognostic factors in cancer cells of luminal BC and other subtypes. (C) HR and *p* values of prognostic factors by univariate Cox regression. (D) Risk curves and scatter plots of sample survival probability for each sample reordered by risk score, heatmap of expression of prognostic factors in low-risk and high-risk groups. (E) Differences in overall survival between high-risk and low-risk groups in the training cohort of luminal BC. (F) Receiver operating characteristic curve (ROC) analysis of the sensitivity and specificity of the prognostic model in the training cohort of luminal BC. (G) Differences in overall survival between high-risk and low-risk groups in the external validation cohort of luminal BC. (H) ROC analysis of the sensitivity and specificity of the prognostic model in the external validation cohort of luminal BC.



**Figure 6.** Construction and evaluation of the prognostic risk model for HER2+ BC. (A) Expression levels of prognostic factors between normal and cancer cells. (B) Expression levels of prognostic factors in cancer cells of HER2+ BC and other subtypes. (C) HR and *p* values of prognostic factors by univariate Cox regression. (D) Risk curves and scatter plots of sample survival probability for each sample reordered by risk score, heatmap of expression of prognostic factors in low-risk and high-risk groups. (E) Differences in overall survival between high-risk and low-risk groups in the training cohort of HER2+ BC. (F) Receiver operating characteristic curve (ROC) analysis of the sensitivity and specificity of the prognostic model in the training cohort of HER2+ BC. (G) Differences in overall survival between high-risk and low-risk groups in the external validation cohort of HER2+ BC. (H) ROC analysis of the sensitivity and specificity of the prognostic model in the external validation cohort of HER2+ BC.



**Figure 7.** Construction and evaluation of the prognostic risk model for TNBC. (A) Expression levels of prognostic factors between normal and cancer cells. (B) Expression levels of prognostic factors in cancer cells of TNBC and other subtypes. (C) HR and *p* values of prognostic factors by univariate Cox regression. (D) Risk curves and scatter plots of sample survival probability for each sample reordered by risk score, heatmap of expression of prognostic factors in low-risk and high-risk groups. (E) Differences in overall survival between high-risk and low-risk groups in the training cohort of TNBC. (F) Receiver operating characteristic curve (ROC) analysis of the sensitivity and specificity of the prognostic model in the training cohort of TNBC. (G) Differences in overall survival between high-risk and low-risk groups in the external validation cohort of TNBC. (H) ROC analysis of the sensitivity and specificity of the prognostic model in the external validation cohort of TNBC.

Based on the established prognostic risk model of each BC subtype, patients were divided into high risk and low risk groups according to the median risk score. The results showed that in the three BC subtypes, the mortality rate of the high-risk group was higher than that of the low-risk group, and the expression of risk factors increased with the increase of risk value, while the protective factors decreased with the increase of risk value (Figures 5D, 6D and 7D). As expected, the overall survival of the high-risk group was worse than that of the low-risk group ( $p < 0.0001$ ) (Figures 5E, 6E and 7E). To evaluate the performance of the prognostic risk model, the ROC curve was drawn to confirm the prognostic accuracy of the risk score. The area under the curve (AUC) of the prognostic risk model were 0.778 and 0.762 for luminal BC, 0.823 and 0.897 for HER2+ BC, 0.927 and 0.966 for TNBC at 3-years and 5-years, respectively (Figures 5F, 6F and 7F). These results illustrate the outstanding predictive performance of the three prognostic models of specific upregulated gene composition. We also obtained results in the external validation cohort that were consistent with the training cohort; the overall survival of the high-risk group was significantly worse than that of the low-risk group with  $p$  smaller than 0.05 (Figures 5G, 6G and 7G). The AUC of the prognostic risk model in the validation cohort were 0.736 and 0.741 for luminal BC, 0.648 and 0.661 for HER2+ BC at 3 and 5 years, and 0.66 and 0.642 for TNBC at 1 and 3 years, respectively (Figures 5H, 6H and 7H). Above all, these results suggest that the three prognostic models are accurate and reliable.

#### 4. Discussion

BC is the most common malignant tumor with high heterogeneity in women around the world and the main cause of cancer-related death in women [44,45]. The high heterogeneity of BC presents a challenge for classification and treatment. Exploring the heterogeneity of different BC subtypes is considered a key step towards the goal of subtype-specific targeted therapy in BC clinical practice [46]. In this study, cancer cells were identified from scRNA-seq data of BC samples based on CNVs. We illustrate the heterogeneity of cancer cells in gene expression, transcriptional regulatory networks, and the intercellular communications of different BC subtypes, which lead to different clinical performance and prognosis of corresponding BC subtypes (Figure 1).

We found that TNBC had a higher proportion of cancer cells than other subtypes. Obviously, the high proportion of cancer cells is associated with difficulty in treating and susceptibility to recurrence of TNBC. The number of specifically upregulated genes in cancer cells of luminal BC, HER2+ BC and TNBC were 523, 456 and 651, respectively. Functional analysis of these genes demonstrated that cancer cells of HER2+ BC and TNBC, especially TNBC, expressed genes in antioxidant and anti-chemical stress pathways compared with that of luminal BC. Cancer cells grow in a hypoxic environment, and reactive oxygen and nitrogen species (RONS) are elevated in a variety of cancer cells under hypoxia [41]. The cytotoxicity of anticancer drugs such as tamoxifen, paclitaxel and  $As_2O_3$  is associated with the accumulation of  $O_2$ ,  $H_2O_2$  and NO [42,43]. The chemical resistance of cancer cells to these anticancer drugs is proportional to the activity of antioxidant genes. The unique expression of antioxidant and anti-chemical stress genes in TNBC explains the high drug resistance of TNBC compared with other BC subtypes, which may provide new ideas for the treatment of TNBC.

TFs have long been recognized as important aspects in maintaining cellular identity and function, and the increasing or decreasing expression of TF can significantly affect cellular function [47]. The results of the transcriptional regulatory networks in tumor cells of three BC subtypes indicate that the key TFs in tumor cells of different BC subtypes varied. TFs, such as XBP1, TAF7, ELF3, MYC, MAX, etc., were more enriched in tumor cells of luminal BC than the other two subtypes. The highly activated TFs in TNBC were YY1, YBX1, SOX11, SOX4, POLE4, SMARCA4 and HDAC2, etc., while in HER2+ BC they were FOXO3, ZNF467, and especially RAD21 and KLF6. Most of the target genes regulated by these key TFs are the specific enriched and upregulated genes of corresponding subtypes, revealing that these TFs might be of great potential value in breast cancer treatment and



drug development. Our TF enrichment results in tumor cells are consistent with many previous studies which demonstrated that TF of YY1 could drive many aggressive cancer phenotypes [48]. The TF of YBX1 is upregulated in TNBC and plays a role in TNBC invasion by regulating glycolysis and EMT-related gene expression [49]. The TF of SOX11 is a critical regulator of basal-like breast cancer growth, invasion, and basal-like gene expression [50]. The TF of MYC is a cancer driver that regulates many biological activities that contribute to tumorigenesis [51]. SOX4 increases breast cancer cell viability, migration, and invasion in vitro, and enhances tumor growth and metastasis in vivo [50]. SMARCA4 is closely related to tumor immune evasion [52]. HDAC3 is strongly expressed in a subgroup with more aggressive breast cancer [53]. Nuclear expression of XBP1s correlates with shorter breast cancer survival [54]. RAD21 expression is associated with early recurrence and treatment resistance in sporadic breast cancer [55].

Genes specifically upregulated in cancer cells of luminal BC, HER2+ BC and TNBC were all enriched in neutrophil degranulation and neutrophil activation pathways. Neutrophil granule protein released after neutrophil activation is associated with tumor progression, and this differential granule mobilization may lead to the migration of neutrophils and associated cancer cells [56]. Meanwhile, neutrophils are the first responders of inflammation and infection [57], and the activation of neutrophils suggests the existence of inflammatory reactions in BC tumor tissues, which was consistent with the results from intercellular communications analysis.

Receptor–ligand pairs that mediate cellular communication between cancer cells and other cells are significantly enriched in the pathways of leukocyte migration, tumor necrosis factor response, and cell chemotaxis. Cancer cells in all three BC subtypes communicate with all other cell types through the MIF-TNFRSF14. TNFRSF14 is a membrane-bound receptor that activates the NF- $\kappa$ B pathway, leading to a pro-inflammatory response [58,59]. Cancer cells of luminal BC communicate with immune cells through CXCL12\_CXCR4. The CXCLs\_CXCR family is known to induce neutrophil recruitment in the inflammatory response [60]. Cancer cells of HER2+ BC communicate with macrophages and CD8\_T cells through ACKR2-CCL3, ACKR2-CCL4 and ACKR2-CCL5. CCL3, CCL4 and CCL5 are pro-inflammatory chemokines, which are mainly used as attractants of leukocytes (monocytes and neutrophils) and are considered to mediate chronic and acute inflammation [61]. Cancer cells of TNBC communicate with macrophages and DC through the receptor–ligand pair of HLA–DPB1\_TNFSF13B. TNFSFs are key molecules in local and systemic inflammatory networks [62]. These results suggest that receptor–ligand pairs for communication between cancer cells and other cells play a key role in inducing inflammatory responses in BC tissues. Inflammation is related to the occurrence and malignant progression of most cancers, and the malignant progression of cancer cells can be promoted through the recruitment and activation of inflammatory cells. Both external and internal inflammation can lead to immunosuppression, thus providing a preferred environment for the development of cancer cells [63–65]. The different ligand–receptor pairs communicate between tumor cells and other cells, triggering inflammatory response with different degrees in the three BC subtypes. Among the three BC subtypes, the inflammatory response scores of TNBC and HER2+ BC were significantly higher than those of luminal BC. Obviously, the high inflammatory response of TNBC and HER2 + BC is inseparable from the poor prognosis of TNBC and HER2+ BC. Our study shows that the heterogeneity exhibited by cancer cells of the three BC subtypes is correlated with the clinical prognosis of each subtype, suggesting the contribution of heterogeneity from cancer cells in shaping subtype heterogeneity.

We combined bulk RNA-seq and scRNA-seq to construct subtype specific prognostic models using cancer cell-specific upregulation genes. Three prognostic models can distinguish high-risk patients from low-risk patients and effectively predict the survival rate of patients, although there is no direct association between prognostic factors for each BC subtype (Figure S5). There were some significant enriched GO terms of these prognostic genes for each subtype, but most terms only contain very few genes. (Figure S5). Moreover, prognostic factors of three BC subtypes were able to distinguish the three BC

subtypes to some extent, especially luminal BC and TNBC. We suspect the poor differentiation for TNBC is due to the high heterogeneity of TNBC (Figure S6). Based on our subtype-specific prognostic models, *HPGD*, *CDC42*, and *PGK1* are the risk factors of TNBC, *ACTR6*, *SULT1A2* and *C2orf76* are the risk factors of HER2+ BC, while *ATP6AP1*, *RNF139*, *BASP1*, *ESR1* and *TSKU* are the risk factors of luminal BC. *HPGD*, which encodes alcohol dehydrogenase, was reported as a marker of poor prognosis of breast cancer [66]. The overexpression of *CDC42* is associated with poor prognosis in BC because *CDC42* enhances the migration of cancer cells [67]. *PGK1* overexpression is associated with the mutations of common tumor suppressor genes *TP53* and *CDH1* [68]. *ACTR6* can be used as a marker of poor prognosis in lung cancer [69]. Overexpression of *SULT1A2* in BC tissues is significantly correlated with BC staging [70]. High expression of *ATP6AP1* in breast cancer is associated with poor prognosis [71]. *RNF139* is a driver gene closely associated with breast cancer [72]. *BASP1* was reported to be highly expressed in cancer and promotes the proliferation of cancer [73]. *ESR1* mutation has become a key mechanism of endocrine therapy resistance in breast cancer [74]. High expression of *TSKU* is associated with a poor prognosis of NSCLC patients [75]. Above all, these risk factors may serve as targets for BC treatment in addition to prognosis of the survival of BC patients. Our study continues the basic, but broadly applicable, clinical subtype classification of the data source (luminal BC, HER2+ BC and TNBC) [76]. Of note, luminal BC can be subdivided into luminal A, luminal B and luminal-HER2 groups [77], and the accurate classification of them needs to refer to more indicators such as KI67, CCNB1, MKI67 and MYBL2 [78]. It is still difficult for us to simply distinguish ER+ into luminal A and luminal B based on the existing information. Taking luminal BC as a subtype contributes to the search for representative biomarkers which are broadly applicable to luminal BC, but it has limitation in exploring the heterogeneity of breast cancer. We hope that more clinical data will be available to explore the heterogeneity of luminal BC in the future.

## 5. Conclusions

In this study, we selected and reanalyzed scRNA-seq data of cancer cells from three BC subtypes. The proportion of cancer cells in TNBC was higher than that of HER2+ BC and luminal BC. There exists heterogeneity both in GRNs and intercellular communication in cancer cells of different BC subtypes. Cancer cells of TNBC subtype uniquely upregulated genes were related to antioxidant and anti-chemical stress, which are associated with high drug resistance and poor prognosis in TNBC. The intercellular communication of cancer cells leads to the inflammatory response of breast tissue, and the different receptor–ligand pairs mediated intercellular communication that induce an inflammatory response with different degrees in three BC subtypes, leading to different prognosis. Finally, we constructed subtype-specific prognostic models using cancer cell-specific upregulated genes, and the prognostic models had high predictive performance. In the context of the emphasis on personalized treatment of BC, these prognostic factors, in addition to predicting patient survival, may also serve as some drug targets for clinical treatment, and they provide some ideas for drug design in BC.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms23179936/s1>.

**Author Contributions:** Conceptualization, J.X. and S.Q.; visualization, J.X., Y.Y. and S.Q.; writing draft, J.X.; editing draft, M.G., F.M., J.X. and S.Q.; data curation, H.G. and X.L.; supervision, S.Q., M.G. and F.M.; funding acquisition, M.G. and F.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by grants from the National Natural Science Foundation of China under Grant No.31970477 (F.M.), and a Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

**Informed Consent Statement:** The datasets used in this study are all from public databases, including GEO, TCGA and METABRIC databases. These public databases allow researchers to download and analyze public data for scientific purposes, so this study does not require ethical approval.

**Data Availability Statement:** All the data supporting this article are available in the Section 2 and Supplementary Files of the manuscript, or can be obtained by contacting the author.

**Acknowledgments:** We are also grateful to our peers for their suggestions and review of the manuscript. All the authors have agreed to the manuscript.

**Conflicts of Interest:** All authors declare that there is no conflict of interest.

## References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; 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*, 209–249. [[CrossRef](#)] [[PubMed](#)]
2. Januškevičienė, I.; Petrikaitė, V. Heterogeneity of breast cancer: The importance of interaction between different tumor cell populations. *Life Sci.* **2019**, *239*, 117009. [[CrossRef](#)] [[PubMed](#)]
3. Franchet, C.; Duprez-Paumier, R.; Lacroix-Triki, M. Molecular taxonomy of luminal breast cancer in 2015. *Bull. Cancer* **2015**, *102*, S34–S46. [[CrossRef](#)]
4. Garrido-Castro, A.C.; Lin, N.U.; Polyak, K. Insights into Molecular Classifications of Triple-Negative Breast Cancer: Improving Patient Selection for Treatment. *Cancer Discov.* **2019**, *9*, 176–198. [[CrossRef](#)]
5. Kreutzfeldt, J.; Rozeboom, B.; Dey, N.; De, P. The trastuzumab era: Current and upcoming targeted HER2+ breast cancer therapies. *Am. J. Cancer Res.* **2020**, *10*, 1045–1067.
6. Perou, C.M.; Sørlie, T.; Eisen, M.B.; van de Rijn, M.; Jeffrey, S.S.; Rees, C.A.; Pollack, J.R.; Ross, D.T.; Johnsen, H.; Akslen, L.A.; et al. Molecular portraits of human breast tumours. *Nature* **2000**, *406*, 747–752. [[CrossRef](#)]
7. Sørlie, T.; Perou, C.M.; Tibshirani, R.; Aas, T.; Geisler, S.; Johnsen, H.; Hastie, T.; Eisen, M.B.; van de Rijn, M.; Jeffrey, S.S.; et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10869–10874. [[CrossRef](#)]
8. Sorlie, T.; Tibshirani, R.; Parker, J.; Hastie, T.; Marron, J.S.; Nobel, A.; Deng, S.; Johnsen, H.; Pesich, R.; Geisler, S.; et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 8418–8423. [[CrossRef](#)]
9. Alfarsi, L.; Johnston, S.; Liu, D.X.; Rakha, E.; Green, A.R. Current issues with luminal subtype classification in terms of prediction of benefit from endocrine therapy in early breast cancer. *Histopathology* **2018**, *73*, 545–558. [[CrossRef](#)]
10. Veitch, Z.; Khan, O.F.; Tilley, D.; Ribnikar, D.; Kostaras, X.; King, K.; Tang, P.; Lupichuk, S. Real-World Outcomes of Adjuvant Chemotherapy for Node-Negative and Node-Positive HER2-Positive Breast Cancer. *J. Natl. Compr. Canc. Netw.* **2019**, *17*, 47–56. [[CrossRef](#)]
11. Kumar, P.; Aggarwal, R. An overview of triple-negative breast cancer. *Arch. Gynecol. Obstet.* **2016**, *293*, 247–269. [[CrossRef](#)]
12. Ren, C.; Han, C.; Fu, D.; Wang, D.; Chen, H.; Chen, Y.; Shen, M. Circulating tumor cells in breast cancer beyond the genotype of primary tumor for tailored therapy. *Int. J. Cancer* **2016**, *138*, 1586–1600. [[CrossRef](#)]
13. Lane, A.N.; Tan, J.; Wang, Y.; Yan, J.; Higashi, R.M.; Fan, T.W. Probing the metabolic phenotype of breast cancer cells by multiple tracer stable isotope resolved metabolomics. *Metab. Eng.* **2017**, *43*, 125–136. [[CrossRef](#)]
14. Liu, X.; Peng, Y.; Wang, J. Integrative analysis of DNA methylation and gene expression profiles identified potential breast cancer-specific diagnostic markers. *Biosci. Rep.* **2020**, *40*, BSR20201053. [[CrossRef](#)]
15. Wang, Y.; Mashock, M.; Tong, Z.; Mu, X.; Chen, H.; Zhou, X.; Zhang, H.; Zhao, G.; Liu, B.; Li, X. Changing Technologies of RNA Sequencing and Their Applications in Clinical Oncology. *Front. Oncol.* **2020**, *10*, 447. [[CrossRef](#)]
16. Zheng, S.; Zou, Y.; Liang, J.Y.; Xiao, W.; Yang, A.; Meng, T.; Lu, S.; Luo, Z.; Xie, X. Identification and validation of a combined hypoxia and immune index for triple-negative breast cancer. *Mol. Oncol.* **2020**, *14*, 2814–2833. [[CrossRef](#)]
17. Tang, J.; Luo, Y.; Wu, G. A glycolysis-related gene expression signature in predicting recurrence of breast cancer. *Aging (Albany NY)* **2020**, *12*, 24983–24994. [[CrossRef](#)]
18. Tang, F.; Barbacioru, C.; Wang, Y.; Nordman, E.; Lee, C.; Xu, N.; Wang, X.; Bodeau, J.; Tuch, B.B.; Siddiqui, A.; et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **2009**, *6*, 377–382. [[CrossRef](#)]
19. Olsen, T.K.; Baryawno, N. Introduction to Single-Cell RNA Sequencing. *Curr. Protoc. Mol. Biol.* **2018**, *122*, e57. [[CrossRef](#)]
20. The Biology of Genomes. Single-cell sequencing tackles basic and biomedical questions. *Science* **2012**, *336*, 976–977. [[CrossRef](#)]
21. Karaayvaz, M.; Cristea, S.; Gillespie, S.M.; Patel, A.P.; Mylvaganam, R.; Luo, C.C.; Specht, M.C.; Bernstein, B.E.; Michor, F.; Ellisen, L.W. Unravelling subclonal heterogeneity and aggressive disease states in TNBC through single-cell RNA-seq. *Nat. Commun.* **2018**, *9*, 3588. [[CrossRef](#)] [[PubMed](#)]
22. Chung, W.; Eum, H.H.; Lee, H.O.; Lee, K.M.; Lee, H.B.; Kim, K.T.; Ryu, H.S.; Kim, S.; Lee, J.E.; Park, Y.H.; et al. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nat. Commun.* **2017**, *8*, 15081. [[CrossRef](#)] [[PubMed](#)]

23. AlJanahi, A.A.; Danielsen, M.; Dunbar, C.E. An Introduction to the Analysis of Single-Cell RNA-Sequencing Data. *Mol. Ther. Methods Clin. Dev.* **2018**, *10*, 189–196. [[CrossRef](#)] [[PubMed](#)]
24. Savas, P.; Virassamy, B.; Ye, C.; Salim, A.; Mintoff, C.P.; Caramia, F.; Salgado, R.; Byrne, D.J.; Teo, Z.L.; Dushyanthen, S.; et al. Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis. *Nat. Med.* **2018**, *24*, 986–993. [[CrossRef](#)] [[PubMed](#)]
25. Zhou, J.X.; Taramelli, R.; Pedrini, E.; Knijnenburg, T.; Huang, S. Extracting Intercellular Signaling Network of Cancer Tissues using Ligand-Receptor Expression Patterns from Whole-tumor and Single-cell Transcriptomes. *Sci. Rep.* **2017**, *7*, 8815. [[CrossRef](#)]
26. Ding, S.; Chen, X.; Shen, K. Single-cell RNA sequencing in breast cancer: Understanding tumor heterogeneity and paving roads to individualized therapy. *Cancer Commun.* **2020**, *40*, 329–344. [[CrossRef](#)]
27. AlMusawi, S.; Ahmed, M.; Nateri, A.S. Understanding cell-cell communication and signaling in the colorectal cancer microenvironment. *Clin. Transl. Med.* **2021**, *11*, e308. [[CrossRef](#)]
28. Butler, A.; Hoffman, P.; Smibert, P.; Papalexi, E.; Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **2018**, *36*, 411–420. [[CrossRef](#)]
29. Stuart, T.; Butler, A.; Hoffman, P.; Hafemeister, C.; Papalexi, E.; Mauck, W.M., 3rd; Hao, Y.; Stoeckius, M.; Smibert, P.; Satija, R. Comprehensive Integration of Single-Cell Data. *Cell* **2019**, *177*, 1888–1902.e1821. [[CrossRef](#)]
30. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. [[CrossRef](#)]
31. Efremova, M.; Vento-Tormo, M.; Teichmann, S.A.; Vento-Tormo, R. CellPhoneDB: Inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* **2020**, *15*, 1484–1506. [[CrossRef](#)] [[PubMed](#)]
32. Patel, A.P.; Tirosh, I.; Trombetta, J.J.; Shalek, A.K.; Gillespie, S.M.; Wakimoto, H.; Cahill, D.P.; Nahed, B.V.; Curry, W.T.; Martuza, R.L.; et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **2014**, *344*, 1396–1401. [[CrossRef](#)] [[PubMed](#)]
33. Wu, F.; Fan, J.; He, Y.; Xiong, A.; Yu, J.; Li, Y.; Zhang, Y.; Zhao, W.; Zhou, F.; Li, W.; et al. Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer. *Nat. Commun.* **2021**, *12*, 2540. [[CrossRef](#)] [[PubMed](#)]
34. Maynard, A.; McCoach, C.E.; Rotow, J.K.; Harris, L.; Haderk, F.; Kerr, D.L.; Yu, E.A.; Schenk, E.L.; Tan, W.; Zee, A.; et al. Therapy-Induced Evolution of Human Lung Cancer Revealed by Single-Cell RNA Sequencing. *Cell* **2020**, *182*, 1232–1251.e1222. [[CrossRef](#)]
35. Li, F.; Wen, H.; Bukhari, I.; Liu, B.; Guo, C.; Ren, F.; Tang, Y.; Mi, Y.; Zheng, P. Relationship Between CNVs and Immune Cells Infiltration in Gastric Tumor Microenvironment. *Front. Genet.* **2022**, *13*, 869967. [[CrossRef](#)]
36. Yu, G.; Wang, L.G.; Han, Y.; He, Q.Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *Omics J. Integr. Biol.* **2012**, *16*, 284–287. [[CrossRef](#)]
37. Aibar, S.; González-Blas, C.B.; Moerman, T.; Huynh-Thu, V.A.; Imrichova, H.; Hulselmans, G.; Rambow, F.; Marine, J.C.; Geurts, P.; Aerts, J.; et al. SCENIC: Single-cell regulatory network inference and clustering. *Nat. Methods* **2017**, *14*, 1083–1086. [[CrossRef](#)]
38. Huynh-Thu, V.A.; Irrthum, A.; Wehenkel, L.; Geurts, P. Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE* **2010**, *5*, e12776. [[CrossRef](#)]
39. Zhao, X.; Yang, J.; Yang, C. The Neuronal Transcription Factor Creb3l1 Potential Upregulates Ntrk2 in the Hypertensive Microenvironment to Promote Vascular Smooth Muscle Cell-Neuron Interaction and Prevent Neurons from Ferroptosis: A Bioinformatic Research of scRNA-seq Data. *Dis. Markers* **2022**, *2022*, 8339759. [[CrossRef](#)]
40. Moons, K.G.; Altman, D.G.; Reitsma, J.B.; Ioannidis, J.P.; Macaskill, P.; Steyerberg, E.W.; Vickers, A.J.; Ransohoff, D.F.; Collins, G.S. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and elaboration. *Ann. Intern. Med.* **2015**, *162*, W1–W73. [[CrossRef](#)]
41. Liu, J.; Wang, Z. Increased Oxidative Stress as a Selective Anticancer Therapy. *Oxid. Med. Cell Longev.* **2015**, *2015*, 294303. [[CrossRef](#)]
42. Moldogazieva, N.T.; Lutsenko, S.V.; Terentiev, A.A. Reactive Oxygen and Nitrogen Species-Induced Protein Modifications: Implication in Carcinogenesis and Anticancer Therapy. *Cancer Res.* **2018**, *78*, 6040–6047. [[CrossRef](#)]
43. Moldogazieva, N.T.; Mokhosoev, I.M.; Feldman, N.B.; Lutsenko, S.V. ROS and RNS signalling: Adaptive redox switches through oxidative/nitrosative protein modifications. *Free Radic. Res.* **2018**, *52*, 507–543. [[CrossRef](#)]
44. DeSantis, C.; Ma, J.; Bryan, L.; Jemal, A. Breast cancer statistics, 2013. *CA Cancer J. Clin.* **2014**, *64*, 52–62. [[CrossRef](#)]
45. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2020. *CA Cancer J. Clin.* **2020**, *70*, 7–30. [[CrossRef](#)]
46. Ma, W.; Zhao, F.; Yu, X.; Guan, S.; Suo, H.; Tao, Z.; Qiu, Y.; Wu, Y.; Cao, Y.; Jin, F. Immune-related lncRNAs as predictors of survival in breast cancer: A prognostic signature. *J. Transl. Med.* **2020**, *18*, 442. [[CrossRef](#)]
47. Lin, W.W.; Xu, L.T.; Chen, Y.S.; Go, K.; Sun, C.; Zhu, Y.J. Single-Cell Transcriptomics-Based Study of Transcriptional Regulatory Features in the Mouse Brain Vasculature. *Biomed. Res. Int.* **2021**, *2021*, 7643209. [[CrossRef](#)]
48. Agarwal, N.; Theodorescu, D. The Role of Transcription Factor YY1 in the Biology of Cancer. *Crit. Rev. Oncog.* **2017**, *22*, 13–21. [[CrossRef](#)]
49. Lai, Y.W.; Hsu, W.J.; Lee, W.Y.; Chen, C.H.; Tsai, Y.H.; Dai, J.Z.; Yang, C.C.; Lin, C.W. Prognostic Value of a Glycolytic Signature and Its Regulation by Y-Box-Binding Protein 1 in Triple-Negative Breast Cancer. *Cells* **2021**, *10*, 1890. [[CrossRef](#)]



50. Shepherd, J.H.; Uray, I.P.; Mazumdar, A.; Tsimelzon, A.; Savage, M.; Hilsenbeck, S.G.; Brown, P.H. The SOX11 transcription factor is a critical regulator of basal-like breast cancer growth, invasion, and basal-like gene expression. *Oncotarget* **2016**, *7*, 13106–13121. [[CrossRef](#)]
51. Lourenco, C.; Resetca, D.; Redel, C.; Lin, P.; MacDonald, A.S.; Ciaccio, R.; Kenney, T.M.G.; Wei, Y.; Andrews, D.W.; Sunnerhagen, M.; et al. MYC protein interactors in gene transcription and cancer. *Nat. Rev. Cancer* **2021**, *21*, 579–591. [[CrossRef](#)] [[PubMed](#)]
52. Peng, L.; Li, J.; Wu, J.; Xu, B.; Wang, Z.; Giamas, G.; Stebbing, J.; Yu, Z. A Pan-Cancer Analysis of SMARCA4 Alterations in Human Cancers. *Front. Immunol.* **2021**, *12*, 762598. [[CrossRef](#)] [[PubMed](#)]
53. Müller, B.M.; Jana, L.; Kasajima, A.; Lehmann, A.; Prinzler, J.; Budczies, J.; Winzer, K.J.; Dietel, M.; Weichert, W.; Denkert, C. Differential expression of histone deacetylases HDAC1, 2 and 3 in human breast cancer—Overexpression of HDAC2 and HDAC3 is associated with clinicopathological indicators of disease progression. *BMC Cancer* **2013**, *13*, 215. [[CrossRef](#)] [[PubMed](#)]
54. Wang, M.; Ruan, S.; Ming, J.; Dong, F. Nuclear expression of XBP1s is correlated with breast cancer survival: A retrospective analysis based on tissue microarray. *Onco Targets Ther.* **2017**, *10*, 5927–5934. [[CrossRef](#)]
55. Yan, M.; Xu, H.; Waddell, N.; Shield-Artin, K.; Haviv, I.; McKay, M.J.; Fox, S.B. Enhanced RAD21 cohesin expression confers poor prognosis in BRCA2 and BRCA1, but not BRCA1 familial breast cancers. *Breast Cancer Res.* **2012**, *14*, R69. [[CrossRef](#)]
56. Xiong, S.; Dong, L.; Cheng, L. Neutrophils in cancer carcinogenesis and metastasis. *J. Hematol. Oncol.* **2021**, *14*, 173. [[CrossRef](#)]
57. Lodge, K.M.; Cowburn, A.S.; Li, W.; Condliffe, A.M. The Impact of Hypoxia on Neutrophil Degranulation and Consequences for the Host. *Int. J. Mol. Sci.* **2020**, *21*, 1183. [[CrossRef](#)]
58. Choi, E.K.; Kim, W.K.; Sul, O.J.; Park, Y.K.; Kim, E.S.; Suh, J.H.; Yu, R.; Choi, H.S. TNFRSF14 deficiency protects against ovariectomy-induced adipose tissue inflammation. *J. Endocrinol.* **2014**, *220*, 25–33. [[CrossRef](#)]
59. Choi, E.K.; Park, H.J.; Sul, O.J.; Rajasekaran, M.; Yu, R.; Choi, H.S. Carbon monoxide reverses adipose tissue inflammation and insulin resistance upon loss of ovarian function. *Am. J. Physiol. Endocrinol. Metab.* **2015**, *308*, E621–E630. [[CrossRef](#)]
60. Khalil, B.A.; Elemam, N.M.; Maghazachi, A.A. Chemokines and chemokine receptors during COVID-19 infection. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 976–988. [[CrossRef](#)]
61. Lazennec, G.; Richmond, A. Chemokines and chemokine receptors: New insights into cancer-related inflammation. *Trends Mol. Med.* **2010**, *16*, 133–144. [[CrossRef](#)]
62. Fan, H.; Lu, B.; Cao, C.; Li, H.; Yang, D.; Huang, L.; Ding, T.; Wu, M.; Lu, G. Plasma TNFSF13B and TNFSF14 Function as Inflammatory Indicators of Severe Adenovirus Pneumonia in Pediatric Patients. *Front. Immunol.* **2020**, *11*, 614781. [[CrossRef](#)]
63. Singh, N.; Baby, D.; Rajguru, J.P.; Patil, P.B.; Thakkannavar, S.S.; Pujari, V.B. Inflammation and cancer. *Ann. Afr. Med.* **2019**, *18*, 121–126. [[CrossRef](#)]
64. Todoric, J.; Antonucci, L.; Karin, M. Targeting Inflammation in Cancer Prevention and Therapy. *Cancer Prev. Res.* **2016**, *9*, 895–905. [[CrossRef](#)]
65. Zappavigna, S.; Cossu, A.M.; Grimaldi, A.; Bocchetti, M.; Ferraro, G.A.; Nicoletti, G.F.; Filosa, R.; Caraglia, M. Anti-Inflammatory Drugs as Anticancer Agents. *Int. J. Mol. Sci.* **2020**, *21*, 2605. [[CrossRef](#)]
66. Greenland, K.J.; Jantke, I.; Jenatschke, S.; Bracken, K.E.; Vinson, C.; Gellersen, B. The human NAD<sup>+</sup>-dependent 15-hydroxyprostaglandin dehydrogenase gene promoter is controlled by Ets and activating protein-1 transcription factors and progesterone. *Endocrinology* **2000**, *141*, 581–597. [[CrossRef](#)]
67. Bray, K.; Gillette, M.; Young, J.; Loughran, E.; Hwang, M.; Sears, J.C.; Vargo-Gogola, T. Cdc42 overexpression induces hyperbranching in the developing mammary gland by enhancing cell migration. *Breast Cancer Res.* **2013**, *15*, R91. [[CrossRef](#)]
68. He, Y.; Luo, Y.; Zhang, D.; Wang, X.; Zhang, P.; Li, H.; Ejaz, S.; Liang, S. PGK1-mediated cancer progression and drug resistance. *Am. J. Cancer Res.* **2019**, *9*, 2280–2302.
69. Zheng, X.; Weigert, A.; Reu, S.; Guenther, S.; Mansouri, S.; Bassaly, B.; Gattenlöhner, S.; Grimminger, F.; Pullamsetti, S.; Seeger, W.; et al. Spatial Density and Distribution of Tumor-Associated Macrophages Predict Survival in Non-Small Cell Lung Carcinoma. *Cancer Res.* **2020**, *80*, 4414–4425. [[CrossRef](#)]
70. Chao, Y.; Ou, Q.; Shang, J. Expression and prognostic value of SULT1A2 in bladder cancer. *Exp. Ther. Med.* **2021**, *22*, 779. [[CrossRef](#)]
71. Wang, J.; Liu, Y.; Zhang, S. Prognostic and immunological value of ATP6AP1 in breast cancer: Implications for SARS-CoV-2. *Aging* **2021**, *13*, 16904–16921. [[CrossRef](#)]
72. Xu, Y.; Luo, H.; Hu, Q.; Zhu, H. Identification of Potential Driver Genes Based on Multi-Genomic Data in Cervical Cancer. *Front. Genet.* **2021**, *12*, 598304. [[CrossRef](#)]
73. Jaikumarr Ram, A.; Girija As, S.; Jayaseelan, V.P.; Arumugam, P. Overexpression of BASP1 Indicates a Poor Prognosis in Head and Neck Squamous Cell Carcinoma. *Asian Pac. J. Cancer Prev.* **2020**, *21*, 3435–3439. [[CrossRef](#)]
74. Carausu, M.; Bidard, F.C.; Callens, C.; Melaabi, S.; Jeannot, E.; Pierga, J.Y.; Cabel, L. ESR1 mutations: A new biomarker in breast cancer. *Expert Rev. Mol. Diagn.* **2019**, *19*, 599–611. [[CrossRef](#)]
75. Huang, H.; Zhang, D.; Fu, J.; Zhao, L.; Li, D.; Sun, H.; Liu, X.; Xu, J.; Tian, T.; Zhang, L.; et al. Tsukushi is a novel prognostic biomarker and correlates with tumor-infiltrating B cells in non-small cell lung cancer. *Aging (Albany NY)* **2021**, *13*, 4428–4451. [[CrossRef](#)]
76. Wu, S.Z.; Al-Eryani, G.; Roden, D.L.; Junankar, S.; Harvey, K.; Andersson, A.; Thennavan, A.; Wang, C.; Torpy, J.R.; Bartonicek, N.; et al. A single-cell and spatially resolved atlas of human breast cancers. *Nat. Genet.* **2021**, *53*, 1334–1347. [[CrossRef](#)]

77. Harbeck, N.; Penault-Llorca, F.; Cortes, J.; Gnant, M.; Houssami, N.; Poortmans, P.; Ruddy, K.; Tsang, J.; Cardoso, F. Breast cancer. *Nat. Rev. Dis. Prim.* **2019**, *5*, 66. [[CrossRef](#)]
78. Hu, Z.; Fan, C.; Oh, D.S.; Marron, J.S.; He, X.; Qaqish, B.F.; Livasy, C.; Carey, L.A.; Reynolds, E.; Dressler, L.; et al. The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genom.* **2006**, *7*, 96. [[CrossRef](#)] [[PubMed](#)]