Mathematical Biosciences and Engineering

MBE, 19(6): 5793–5812.
DOI: 10.3934/mbe.2022271
Received: 18 January 2022
Revised: 22 March 2022
Accepted: 28 March 2022
Published: 06 April 2022

http://www.aimspress.com/journal/MBE

Research article

Analysis of single-cell RNA-sequencing data identifies a hypoxic tumor subpopulation associated with poor prognosis in triple-negative breast cancer

Yi Shi, Xiaoqian Huang, Zhaolan Du and Jianjun Tan*

Department of Biomedical Engineering, Faculty of Environment and Life, Beijing University of Technology, Beijing International Science and Technology Cooperation Base for Intelligent Physiological Measurement and Clinical Transformation, Beijing 100124, China

* **Correspondence:** E-mail: tanjianjun@bjut.edu.cn; Tel: +8601067392001.

Abstract: Triple-negative breast cancer (TNBC) is an aggressive subtype of mammary carcinoma characterized by low expression levels of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2). Along with the rapid development of the single-cell RNA-sequencing (scRNA-seq) technology, the heterogeneity within the tumor microenvironment (TME) could be studied at a higher resolution level, facilitating an exploration of the mechanisms leading to poor prognosis during tumor progression. In previous studies, hypoxia was considered as an intrinsic characteristic of TME in solid tumors, which would activate downstream signaling pathways associated with angiogenesis and metastasis. Moreover, hypoxia-related genes (HRGs) based risk score models demonstrated nice performance in predicting the prognosis of TNBC patients. However, it is essential to further investigate the heterogeneity within hypoxic TME, such as intercellular communications. In the present study, utilizing single-sample Gene Set Enrichment Analysis (ssGSEA) and cell-cell communication analysis on the scRNA-seq data retrieved from Gene Expression Omnibus (GEO) database with accession number GSM4476488, we identified four tumor subpopulations with diverse functions, particularly a hypoxia-related one. Furthermore, results of cell-cell communication analysis revealed the dominant role of the hypoxic tumor subpopulation in angiogenesis- and metastasis-related signaling pathways as a signal sender. Consequently, regard the TNBC cohorts acquired from The Cancer Genome Atlas (TCGA) and GEO as train set and test set respectively, we constructed a risk score model with reliable capacity for the prediction of overall survival (OS), where *ARTN* and *L1CAM* were identified as risk factors promoting angiogenesis and metastasis of tumors.

The expression of *ARTN* and *L1CAM* were further analyzed through tumor immune estimation resource (TIMER) platform. In conclusion, these two marker genes of the hypoxic tumor subpopulation played vital roles in tumor development, indicating poor prognosis in TNBC patients.

Keywords: single-cell RNA-sequencing; triple-negative breast cancer; cell-cell communication; hypoxia; prognosis

1. Introduction

Breast cancer (BRCA) is the most frequently diagnosed cancer and the leading cause of cancer mortality among females [1]. Triple-negative breast cancer (TNBC) is a subtype of BRCA with the absence of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2), accounting for approximately 15–20% of all breast carcinomas [2]. Compared to other subtypes, TNBC has a more aggressive clinical process and worse prognosis due to the extremely heterogeneous tumor microenvironment (TME) [3]. TME is a complicated ecological system comprising vasculature, extracellular matrix (ECM), cytokines, growth factors, and lots of various cell populations [4], which has a strong association with tumorigenesis, tumor cells proliferation, invasion, and metastasis of BRCA [5]. Investigating the mechanisms by which TME influences tumor progression in TNBC could facilitate improving the prognosis of patients.

Single-cell RNA-sequencing (scRNA-seq) is an innovative technology that enables the investigation of the transcriptome of individual cells with the purpose of identification of cell subpopulations with analogical transcriptional patterns [6]. In contrast with traditional bulk RNA-sequencing (RNA-seq), scRNA-seq allows revealing the significant heterogeneity within TME [7,8]. Previous study has discovered three distinct subpopulations of breast epithelial cells and reconstructed a continuous differentiation trajectory [9], which provided comprehensive insights into the human mammary epithelium. Similarly, cancer-associated fibroblasts (CAFs) were revealed to exist in three spatially and functionally diverse subpopulations in the mouse breast cancer model [10]. Furthermore, the prognostic capacity of each CAFs subpopulation was validated in clinical cohorts, confirming the potential prognostic implications of cell subpopulations within TME.

Because of the imbalance between the rapid proliferation of tumor cells and the inadequate oxygen supply of vascular tissue, hypoxic TME is an inherent characteristic of solid tumors [11]. In the condition of hypoxia, tumor cells could secrete a variety of vascular growth factors that promote abnormal angiogenesis, consequently increasing the oxygen supply to meet tumor progression. Meanwhile, the invasive and metastatic capacities of tumor cells might also be greatly enhanced for the purpose of seeking a more appropriate growing environment [12,13]. In recent studies, prognostic prediction models have been established and validated based on hypoxia-related genes (HRGs) [14] and hypoxia-immune-related genes [15], expounding their views that HRGs within hypoxic TME were significant factors contributing to the poor prognosis of TNBC patients. However, the heterogeneity within hypoxic TME was ignored in these studies, which may play a vital role during tumor progression. To overcome this deficiency, we explored the heterogeneity within hypoxic TME utilizing scRNA-seq data and corresponding analytical methods, such as Gene Set Enrichment Analysis (GSEA) and cell-cell communication analysis in this study. Furthermore, we identified a hypoxia-related subpopulation and investigated its involved signaling pathways and prognostic implications for TNBC patients.

2. Materials and methods

2.1. Data acquisition

The counts matrix of TNBC scRNA-seq was collected from Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) with accession number GSM4476488 [16]. The BRCA RNA-seq expression profiles of The Cancer Genome Atlas (TCGA) were downloaded through the UCSC Xena platform (https://xena.ucsc.edu/) [17], consisting of tumor samples of 1050 patients with corresponding clinical features. Subtyping by R package genefu [18] with the PAM50 model obtained 188 TNBC patients. Furthermore, we retrieved the data of 107 TNBC patients from GEO with accession number GSE58812 [19].

2.2. Analysis of scRNA-seq data

Seurat package [20] was used to analyze the raw counts matrix of scRNA-seq data containing 33,694 genes and 532 cells. Imported into R by CreateSeuratObject with genes detected in at least one cell, cells with the proportion of mitochondrial genes higher than 15% and doublet cells detected by DoubletFinder package were further filtered. After data normalization through NormalizeData, the top 2000 highly variable genes were identified using FindVariableFeatures for downstream analysis. Subsequently, we performed ScaleData to scale the data and remove the effects of certain variables such as the proportion of mitochondrial genes. Furthermore, we conducted principal component analysis (PCA) to reduce the dimensionality. The first 15 significant principal components were applied by *t*-Distributed Stochastic Neighbor Embedding (*t*-SNE) for nonlinear dimensionality reduction and FindNeighbors for cells clustering. Finally, the resolution of FindClusters was set to 0.8 thus identifying seven clusters. Cell types were annotated by typical marker genes such as *EPCAM*, *VWF*, *PTPRC* and *PDGFRA* along with the SingleR package [21]. The genomic copy number profiles of cells were inferred to distinguish tumor cells from non-malignant (i.e., normal) cell types through the CopyKAT package with default parameters [16]. For the epithelial cells extracted, the resolution was adjusted to 1.3 after re-clustering. FindMarkers was utilized to identify marker genes with the $\log_2$FC (fold change) > 0.2 and p-value < 0.05. Several plotting functions built-in the Seurat package were applied for visualization such as DimPlot, VlnPlot, and FeatureScatter.

2.3. Single-sample Gene Set Enrichment Analysis

Single-sample Gene Set Enrichment Analysis (ssGSEA) is an extension of GSEA that calculates separate enrichment scores for each paired sample and gene set. Each ssGSEA enrichment score represents the extent to which genes in a particular gene set are coordinately up- or down-regulated within single sample. Acquiring the collection of hallmark gene sets from the Molecular Signatures Database (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp) [22] as an input, which includes 50 gene sets representing specific well-defined biological states or processes, we performed ssGSEA and visualized the enrichment scores using R package escape.

2.4. Cell-cell communication analysis

Analysis of cell-cell communication was conducted based on R package CellChat [23] built-in human ligand-receptor interactions database, consisting of 1939 interactions among 427 ligands and 357 receptors. Through the integration of the expression levels with prior known ligand-receptor interactions, we compute the communication probability and infer intercellular communication networks. Afterwards, we summarized the communication probabilities of all interactions associated with each signaling pathway, and identified outputting communication patterns primarily regulating different signaling pathways.

2.5. Tumor immune estimation resource (TIMER) analysis tool

To determine the correlation between gene expression and immune infiltration in TNBC, we applied TIMER 2.0 (http://timer.cistrome.org/) web platform, comprising of more 10,000 TCGA samples across 32 cancer types. We investigated the infiltration levels of six types of immune cells, including CD8+ T cells, CD4+ T cells, B cells, neutrophils, macrophages, and dendritic cells. Furthermore, we utilized the Gene_DE and Gene_Corr modules of TIMER to explore the expression of *ARTN* and *L1CAM* in TNBC, along with their correlation.

2.6. Construction of the risk score model for TNBC prognosis

Utilizing the TNBC patients extracted from TCGA-BRCA cohorts as the train set, univariate Cox proportional hazards regression analysis was performed to identify prognostic genes (p-value < 0.05) associated with overall survival (OS). Whereafter, multivariate Cox proportional hazards regression analysis was implemented to construct a prognostic model for the calculation of risk score as Eq (1), where the β_i and Exp_i represented the multivariate Cox regression coefficient and expression level of gene i respectively. The number of prognostic genes was denoted by n. To ascertain how well the risk score model predicts the survival of patients at the time point of one, three and five years, receiver operating characteristic (ROC) curves were drawn by survivalROC package to obtain the area under curve (AUC) values. Afterward, TNBC patients of the GSE58812 dataset were divided into high- and low-risk groups according to the median risk score as the cutoff value. The performance of the model was further validated through Kaplan-Meier curves with the log-rank test.

$$\text{Risk score} = \sum_i^n \beta_i \times Exp_i \tag{1}$$

3. Results and discussion

3.1. Investigation of the heterogeneity within hypoxic TME

A total of 483 cells and 22,237 genes were remained after quality control, consisting of 196 epithelial cells, 250 fibroblasts, 13 endothelial cells, and 24 monocytes. On the basis of the genomic copy number profiles distinguishing cellular malignant status, we divided the epithelial cells into the tumor and normal ones. Cell numbers of the former (157) were nearly four times that of the latter (39), suggesting that a large proportion of the epithelial cells were malignant. The final annotation results of cell types were shown in Figure 1A, while the expression levels of typical marker genes were

illustrated in Figure A1. To further explore the heterogeneity within epithelial cells, we extracted them and identified two normal subpopulations along with four tumor subpopulations (Figure 1B). For each subpopulation, marker genes were arranged in descending order according to log_2FC, of which the top five genes were illustrated in Figure 1C.

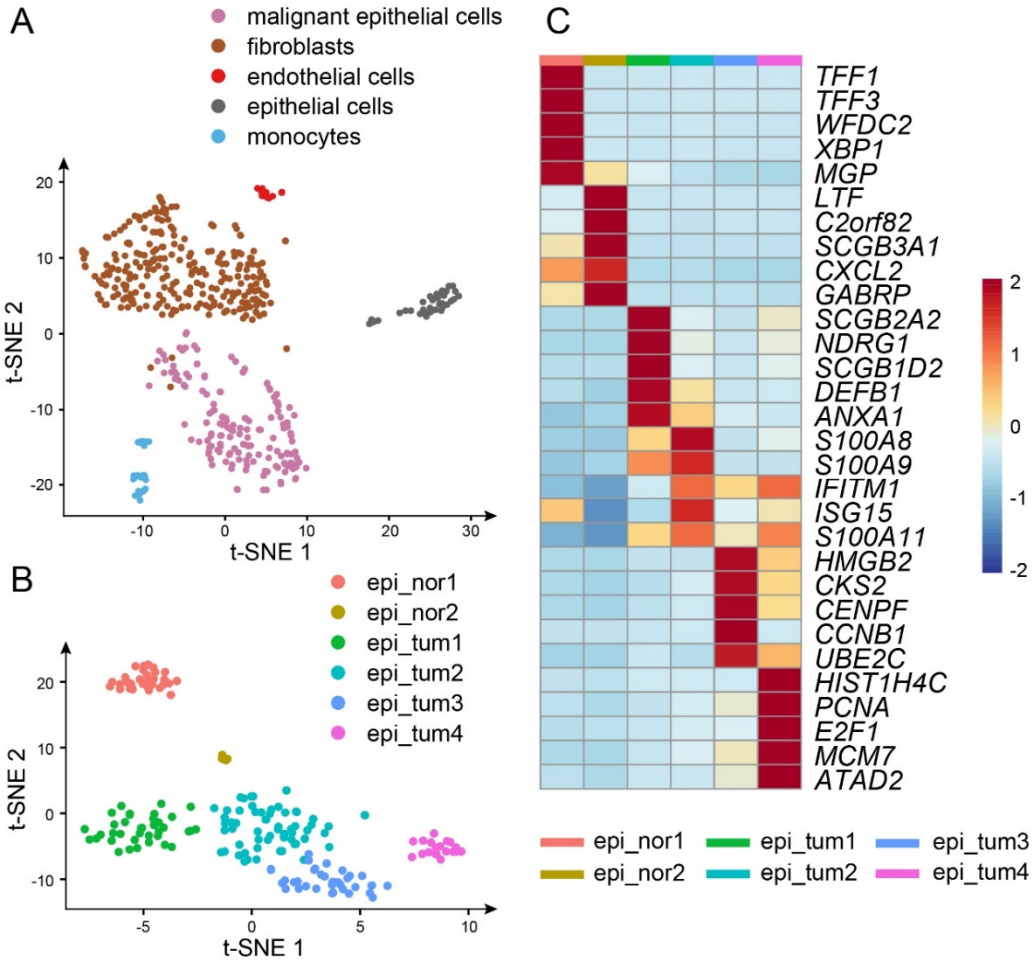

Figure 1. Dimensional reduction plot by t-SNE for (A) all cell types after quality control; (B) subpopulations of epithelial cells. (C) Heat map of the top five genes ordered by log_2FC among all subpopulations of epithelial cells.

Notably, the co-expression of homologous genes was detected in some subpopulations of epithelial cells. *TFF1* and *TFF3*, both belonging to the same trefoil factor family, were overexpressed in the normal subpopulation 1 of epithelial cells (epi_nor1). It has been demonstrated that *TFF1* was strongly expressed under estrogen transcriptional control in a type of human breast cancer cells containing estrogen receptor, but at low levels in normal mammary tissue [24,25]. In line with previous studies, the expression level of estrogen receptor 1 (*ESR1*) was up-regulated in epi_nor1 compared to other subpopulations (Figure A1). Consistently, ssGSEA results further confirmed that epi_nor1 was intensely involved in the response to estrogen (Figure 2A). Furthermore, the Pearson correlation coefficient among *TFF1*, *TFF3*, and *ESR1* in epithelial cells reached 0.97, 0.85, and 0.86 respectively (Figure 2B), which indicates that the expression of *TFF3* might be under estrogen regulation similar to that of *TFF1*. Moreover, S100 calcium binding protein A8 (*S100A8*) and S100 calcium binding

protein A9 (*S100A9*) were up-regulated in epi_tum2, which can form a heterodimer with biological activity [26]. The expression of *S100A8* and *S100A9* was regulated by tamoxifen, an estrogen receptor inhibitor, in breast cancer tissue [27]. Meanwhile, mammaglobin A (*SCGB2A2*) and lipophilin B (*SCGB1D2*), members of the secretoglobin superfamily, were identified as the marker genes of epi_tum1. These two genes were reported to be co-expressed in breast cancer [28] and several malignant tumors of the female genital tract [29], where their proteins form a covalent compound [30,31]. It has been also proved that *SCGB1D2* was expressed in estrogen receptor-positive tumors with more frequency [28]. These results suggest that the co-expression of above mentioned gene pairs could be regulated by *ESR1*-positive epi_nor1.

As the ssGSEA results characterized diverse biological states for subpopulations of epithelial cells shown in Figure 3, there were no hallmark gene sets enriched in epi_nor2, indicating its comparatively weak biological behaviors contrary to other subpopulations. Moreover, gene sets associated with cell division (G2/M checkpoint, DNA repair, and mitotic spindle assembly) as well as target genes of *MYC* and E2F transcription factors family were significantly enriched in epi_tum3 and epi_tum4. It has been proved that the transcription factors coded by *MYC* [32–34] and E2F family [35] target genes related to DNA synthesis and cell cycle, thus regulating cell division. Although the enrichment scores of epi_tum2 on these gene sets are slightly less than epi_tum3 and epi_tum4, there appear to be similarities among them in some gene sets such as DNA repair and target genes of *MYC*. These results suggest that these three subpopulations might be in the process of mitosis and proliferation, of which epi_tum3 and epi_tum4 share similar malignant states. It was further confirmed that genes encoding proteins involved in oxidative phosphorylation were enriched significantly in epi_tum2, epi_tum3, and epi_tum4 due to excess energy consumption by cell proliferation. In contrast to them, up-regulation of genes associated with hypoxia in epi_tum1 indicates it was a vital subpopulation within hypoxic TME. To clarify how the hypoxic tumor subpopulation (epi_tum1) influences the behaviors of other subpopulations through intercellular interactions, we further performed cell-cell communication analysis.

3.2. Effects of the hypoxic tumor subpopulation on tumor progression

The cell-cell communication network was illustrated in Figure 4A, where the width of the edges represents the strength of the communication. As shown more intuitively in Figure 4B, the outputting/inputting communication strength of all cell subpopulations demonstrated that fibroblasts and endothelial cells served as the dominant signal sender/receiver, though monocytes have the highest expression levels of ligands/receptors (Figure A1). Furthermore, we identified statistically significant signaling pathways with $p < 0.05$, of which 39 signaling pathways involved by epi_tum1 as a signal sender were illustrated in Figure 5. There were six outputting communication patterns coordinating with each other to participate in these signaling pathways. Particularly, epi_nor2 was filtered out since it has relatively weak communications with other subpopulations at the level of signaling pathways. Distinct from the other three tumor subpopulations, epi_tum1 has a unique outputting communication pattern corresponding to eight signaling pathways, where ligands and receptors involved were listed in Table 1. In MPZ and OCLN signaling pathways, the communication way was cell-cell contact and the ligands and receptors were identical, which primarily mediated cell adhesion and tight junction between cells respectively [36,37].

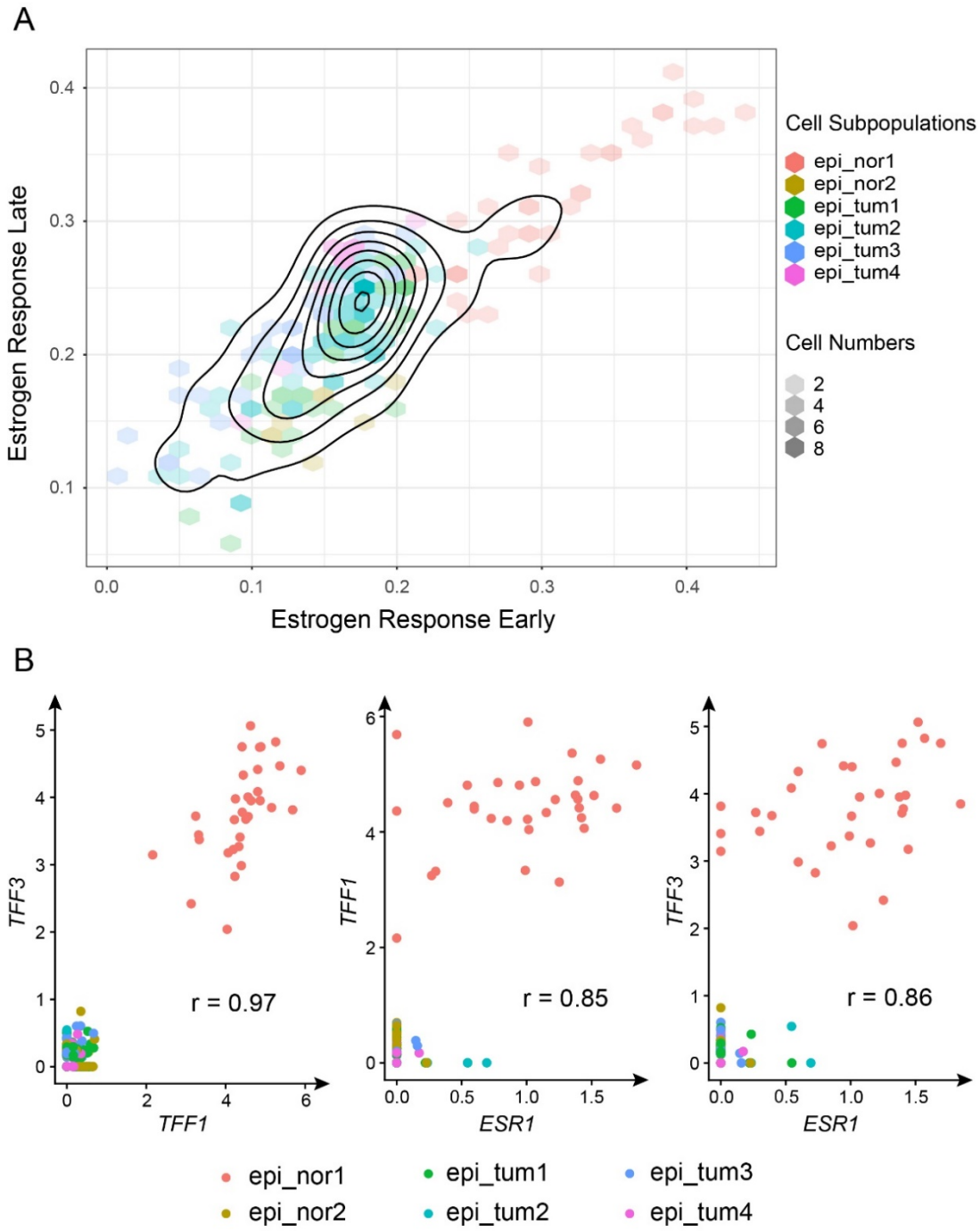

Figure 2. Scatter plot showing (A) the ssGSEA enrichment score of early and late response to estrogen in each subpopulation of epithelial cells (Each hexagon represents a certain number of cells, indicated by color shades); (B) the co-expression among *TFF1*, *TFF3* and *ESR1* in epi_nor1, with the Pearson correlation coefficient reached 0.97, 0.85 and 0.86 respectively.

Figure 3. Heat map showing the ssGSEA enrichment scores of certain gene sets in all subpopulations of epithelial cells, where each column represents a cell.

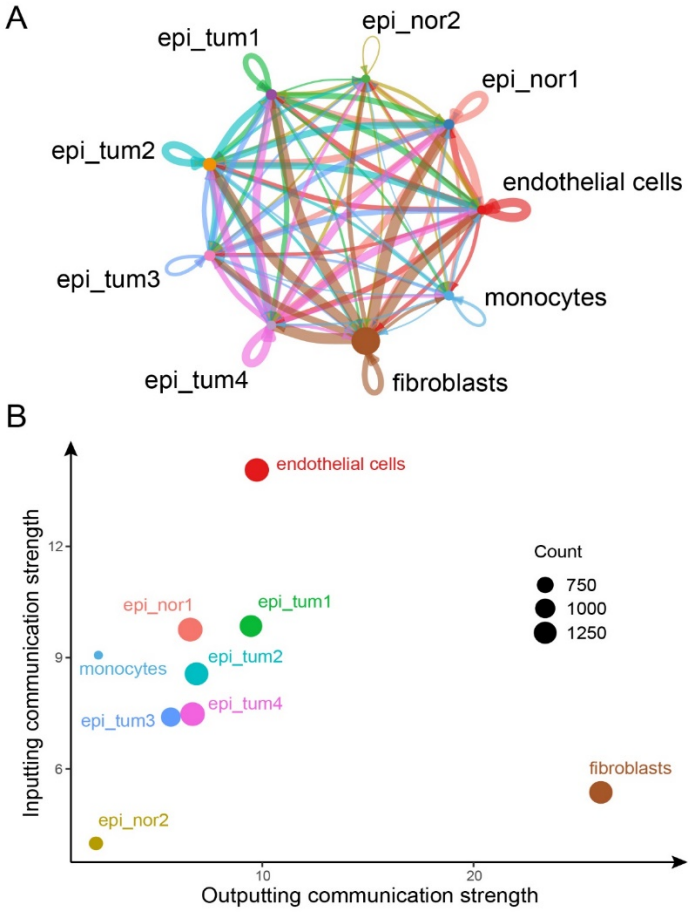

Figure 4. (A) Network diagram of cell-cell communication with the width of edges representing the strength of the communication, where the arrow points from the ligand cells to the receptor cells. (B) Scatter plot showing the outputting and inputting communication strength of all cell subpopulations.

Table 1. Signaling pathways corresponding to the epi_tum1 outputting communication pattern.

Signaling pathway	Ligand	Log_2FC	p-value	Receptor	Communication way
ANGPTL	*ANGPTL4*	1.017	0.002	*ITGA5_ITGB1*	Secreted signaling
ANGPTL	*ANGPTL4*	1.017	0.002	*CDH5*	Secreted signaling
ANGPTL	*ANGPTL4*	1.017	0.002	*CDH11*	Secreted signaling
ANGPTL	*ANGPTL4*	1.017	0.002	*SDC1*	Secreted signaling
ANGPTL	*ANGPTL4*	1.017	0.002	*SDC2*	Secreted signaling
ANGPTL	*ANGPTL4*	1.017	0.002	*SDC3*	Secreted signaling
ANGPTL	*ANGPTL4*	1.017	0.002	*SDC4*	Secreted signaling
CALCR	*ADM*	1.358	< 0.001	*CALCRL*	Secreted signaling
EDN	*EDN1*	0.435	< 0.001	*EDNRB*	Secreted signaling
GDNF	*ARTN*	0.354	< 0.001	*GFRA1_RET*	Secreted signaling
L1CAM	*L1CAM*	0.240	< 0.001	*ITGA4_ITGB7*	Cell-cell contact
L1CAM	*L1CAM*	0.240	< 0.001	*L1CAM*	Cell-cell contact
MPZ	*MPZL1*	0.355	< 0.001	*MPZL1*	Cell-cell contact
OCLN	*OCLN*	0.371	< 0.001	*OCLN*	Cell-cell contact
VEGF	*PGF*	-0.781	0.042	*FLT1*	Secreted signaling
VEGF	*VEGFA*	1.963	< 0.001	*FLT1*	Secreted signaling
VEGF	*VEGFA*	1.963	< 0.001	*KDR*	Secreted signaling
VEGF	*VEGFA*	1.963	< 0.001	*FLT1_KDR*	Secreted signaling
VEGF	*VEGFB*	0.103	0.027	*FLT1*	Secreted signaling

Angiogenesis-related signaling pathways were revealed to be participated predominantly by epi_tum1 such as CALCR, VEGF, EDN, and ANGPTL (Figure 6A–D). Calcitonin receptor-like receptor (*CALCRL*) was one receptor of adrenomedullin (*ADM*), a 52 amino acid peptide with 24% homology to calcitonin gene-related peptide (*CGRP*) initially discovered in 1993 [38]. *ADM* could promote proliferation and migration of endothelial cells thus facilitating vasodilation and angiogenesis [38,39], which was functionally similar to vascular endothelial growth factor A (*VEGFA*) [40,41]. Except for being synthesized and secreted by endothelial cells [42], *ADM* has also been reported as an autocrine growth factor within TME [43]. As shown in Figure 6A,B, endothelial cells are the only signal receivers of CALCR and VEGF signaling pathways. Compared to other cell subpopulations, epi_tum1 is the dominant signal sender of these two signaling pathways promoting the proliferation and migration of endothelial cells, proving that the autocrine growth factor *ADM* was secreted primarily by the hypoxic tumor subpopulation within TME. Moreover, endothelin 1 (*EDN1*) and angiopoietin like 4 (*ANGPTL4*) are ligands secreted by epi_tum1 involved in the EDN and ANGPTL signaling pathways, whose roles in the regulation of angiogenesis have also been demonstrated in previous studies [44–47]. As illustrated in Figure 6C–D, fibroblasts are also participating in both signaling pathways. In summary, the hypoxic tumor subpopulation within TME could accelerate abnormal angiogenesis and tumor progression through the above signaling pathways due to its up-regulation of *ADM*, *VEGFA*, *EDN1*, and *ANGPTL4*, which were shown in Figure A1.

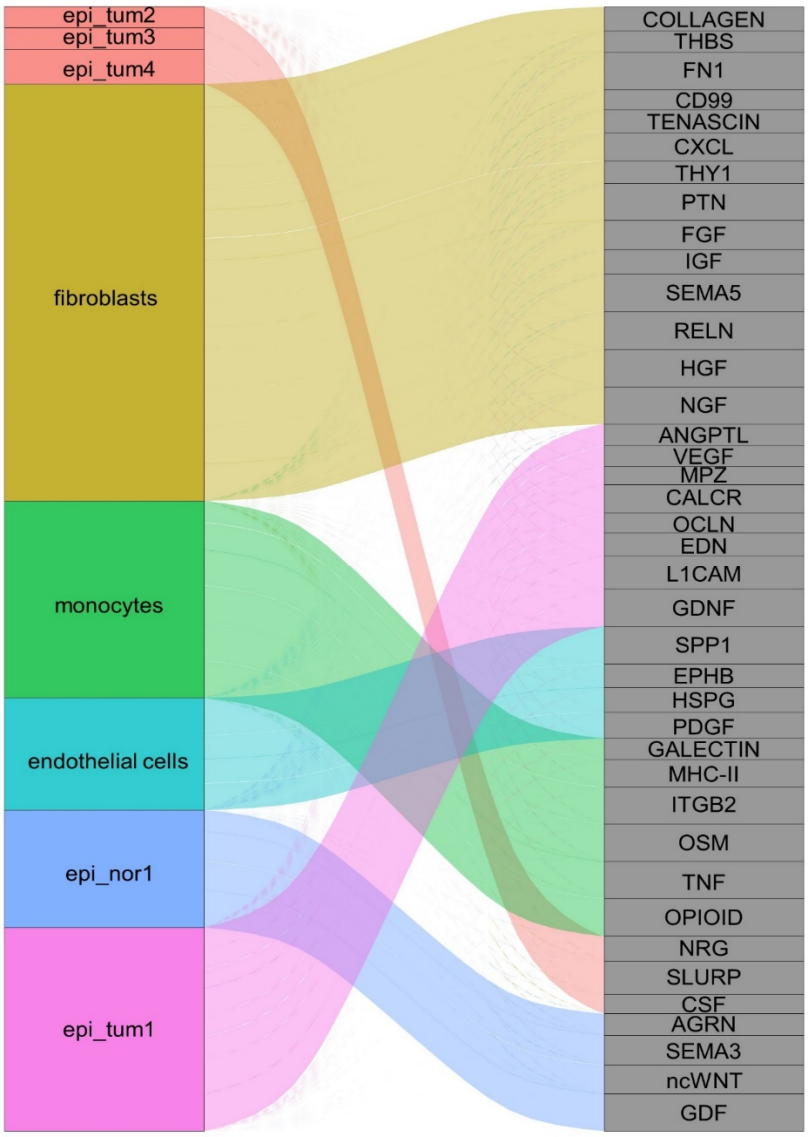

Figure 5. Alluvial diagram presenting the correspondence between six outputting communication patterns with epi_tum1 involved 39 signaling pathways. Each color of line represents a distinct outputting communication pattern, and whether the dashed or solid line represents the communication strength of the signaling pathway involved.

Besides, epi_tum1 serves as the primary signal sender of GDNF and L1CAM signaling pathways (Figure 6E–F). Artemin (*ARTN*) was a member of the glial cell derived neurotrophic factor (*GDNF*) family ligands [48], the expression of which was regulated by estrogen [49]. It has been demonstrated that *ARTN* was highly expressed in breast cancer [50] and could enhance the metastasis and invasion of estrogen receptor-positive cells [51,52]. Consistent with previous studies, *ARTN* secreted by epi_tum1 mainly targets the *ESR1*-positive subpopulation (epi_nor1) within TME as shown in Figure 6E. L1 cell adhesion molecule (*L1CAM*) and *ANGPTL4* have been demonstrated to accelerate the vascular metastasis of tumor cells from the breast to the lungs [53], whose expression was mediated by hypoxia inducible factor 1 subunit alpha (*HIF1A*), especially under hypoxia [53]. The expression levels of *ARTN* and *L1CAM* were illustrated in Figure A1. These results suggest that the hypoxic tumor subpopulation contributes to the metastasis and invasion of tumors.

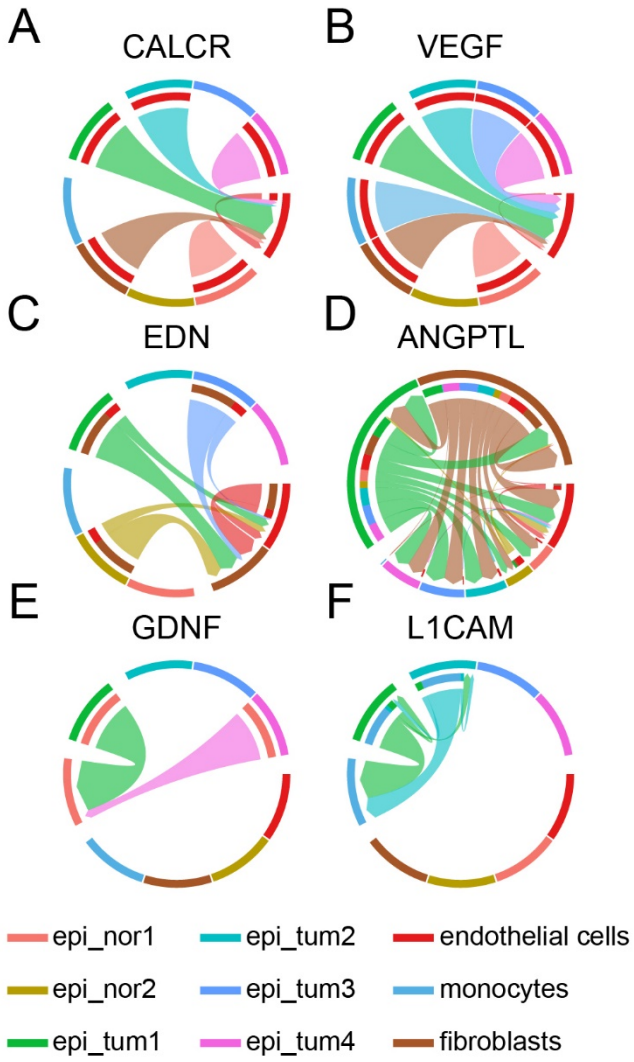

Figure 6. Chord diagram visualizing the cell-cell communication in (A) CALCR; (B) VEGF; (C) EDN; (D) ANGPTL; (E) GDNF and (F) L1CAM signaling pathways. The color of the arrow represents the ligand cells and the direction points to the receptor cells.

3.3. Prognostic implications of the hypoxic tumor subpopulation

Totally 10 ligands secreted by epi_tum1 involved in above eight signaling pathways were listed in Table 1, of which *PGF* and *VEGFB* were filtered out due to not meeting the criteria for marker genes of epi_tum1 ($log_2FC > 0.2$ and p-value < 0.05). For the rest eight genes, we investigated their potential correlations with immune cells infiltration in TNBC using TIMER. As shown in Figure 7, the infiltration levels of dendritic cells had significant positive correlations with the expression levels of *L1CAM* (R = 0.196, p = 9.58e-03), *EDN1* (R = 0.272, p = 2.89e-04), *VEGFA* (R = 0.162, p = 3.24e-02), and *ADM* (R = 0.202, p = 7.65e-03). The infiltration levels of macrophages had significant positive correlations with the expression levels of *ANGPTL4* (R = 0.15, p = 4.79e-02), *OCLN* (R = 0.182, p = 1.60e-02), and *MPZL1* (R = 0.249, p = 9.42e-04). The infiltration levels of CD4+ T cells had significant positive correlations with the expression levels of *L1CAM* (R = 0.296, p = 7.50e-05) and *EDN1* (R = 0.253, p = 7.57e-04). The infiltration levels of neutrophils had significant positive

correlations with the expression levels of *MPZL1* (R = 0.177, p = 1.98e-02). On the other hand, the infiltration levels of B cells had significant negative correlations with the expression levels of *ARTN* (R = -0.175, p = 2.11e-02), *ANGPTL4* (R = -0.273, p = 2.70e-04), *VEGFA* (R = -0.236, p = 1.69e-03), and *ADM* (R = -0.284, p = 1.5e-04). These results suggested that these genes modulate infiltration of immune cells into tumor tissues in TNBC.

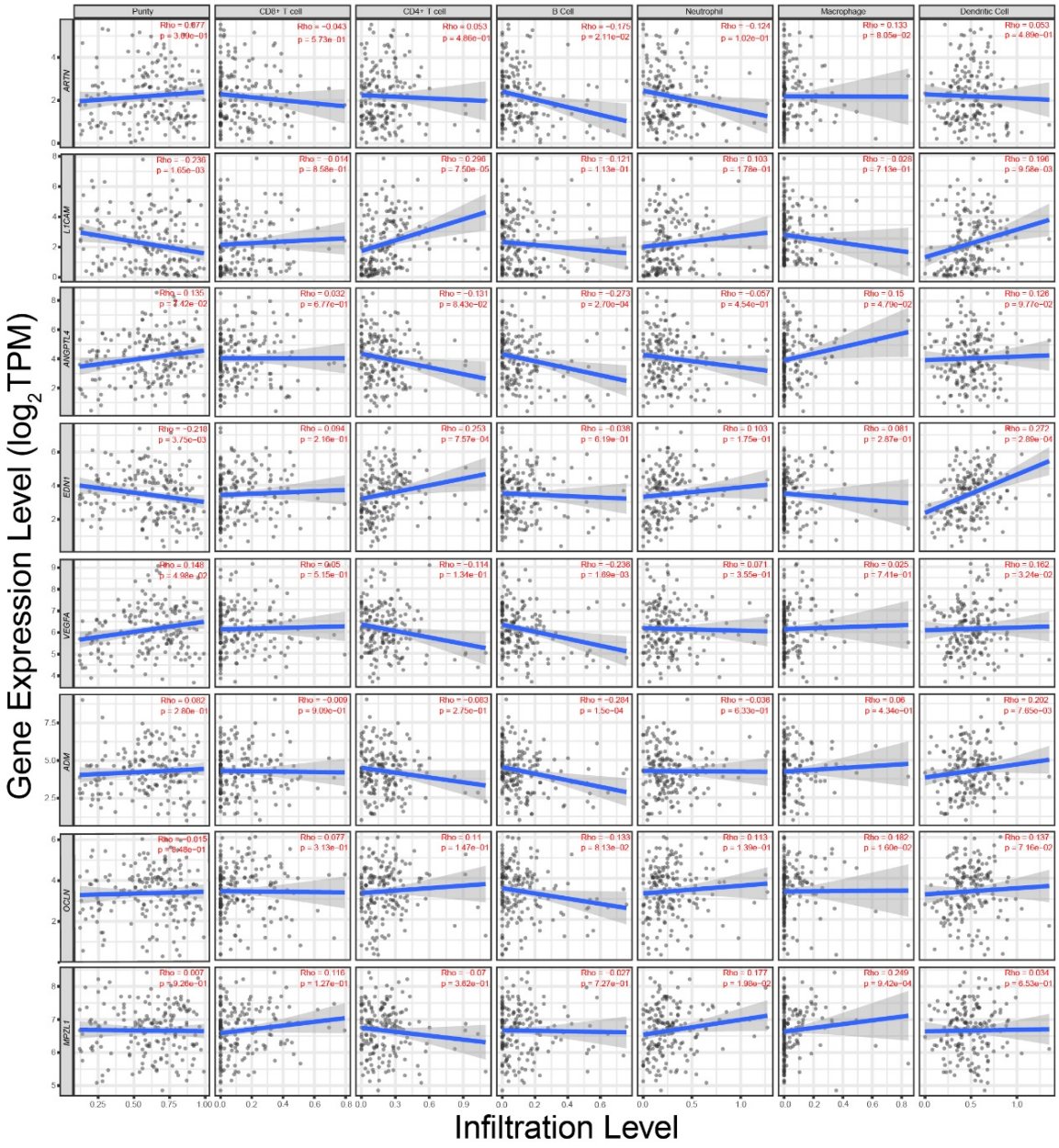

Figure 7. Correlation of expression levels of eight marker genes of epi_tum1 with immune infiltration levels in TNBC.

Then, we performed univariate Cox regression analysis to identify whether these eight genes were associated with OS of TNBC patients. As shown in Figure 8A, *ARTN* and *L1CAM* were considered as risk factors associated with prognosis due to hazard ratio (HR) > 1 and p-value < 0.05, then used for the construction of the prognostic model. The risk score was calculated as follows: Risk score = 0.3893

× Exp_{ARTN} + 0.1824 × Exp_{L1CAM}. The time-dependent ROC curves were shown in Figure 8B, where the AUC values reached 0.802, 0.728, and 0.637 at the time point of one, three, and five years respectively, indicating the reliability of the model in terms of predicting OS of TNBC patients. Moreover, as the Kaplan-Meier curves are shown in Figure 8C, the high-risk group was associated with worse OS, suggesting high expression levels of *ARTN* and *L1CAM* imply poor prognosis.

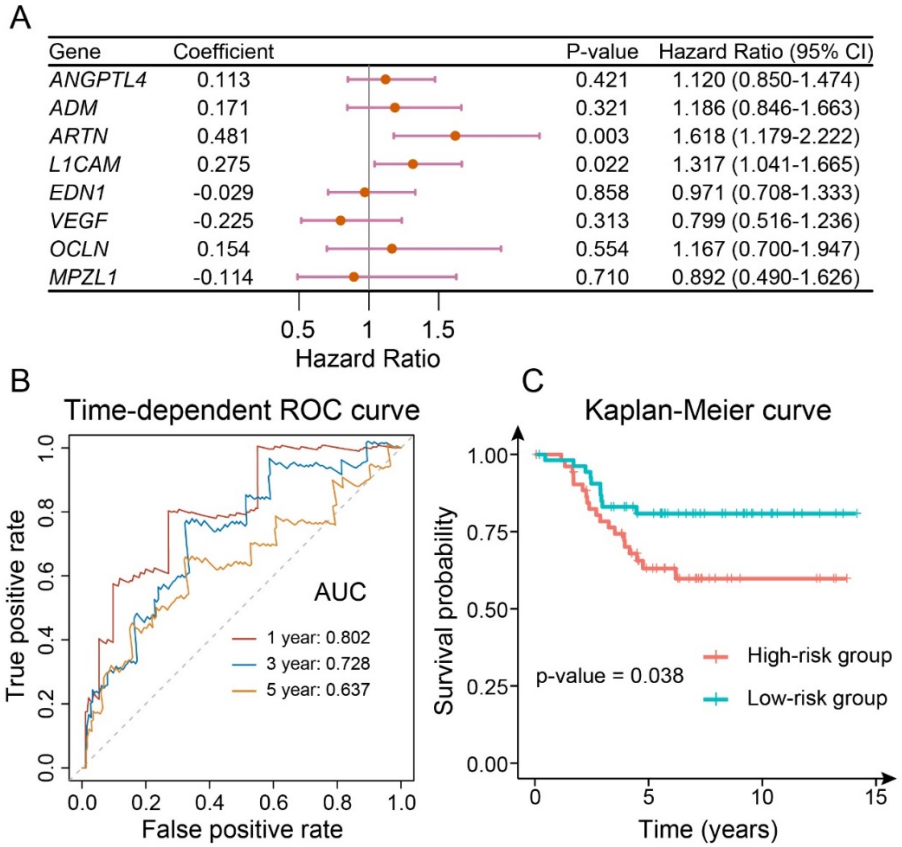

Figure 8. (A) Forest plot showing the results of univariate Cox proportional hazards regression analysis. (B) Time-dependent ROC curve for predicting the OS of TCGA patients. (C) Kaplan-Meier curve showing the OS of GEO patients stratified by median risk score.

Considered as metastasis-related marker genes of the hypoxic tumor subpopulation in this study, the roles of *ARTN* and *L1CAM* in angiogenesis have also been studied. *ARTN* has been reported to stimulate *de novo* tumor angiogenesis in ER-positive BRCA partially mediated by up-regulation of *VEGFA* [54]. Furthermore, the sixth Ig-like domain of *L1CAM* (L1Ig6) has been demonstrated as a pro-angiogenic factor in vivo [55–57]. A recent study has developed and verified a risk score model based on TNBC-specific differential HRGs (dHRGs) for prognostic prediction in TNBC patients [14]. Through differential expression analysis and survival analysis, they identified 48 dHRGs associated with prognosis. The eventual 3-gene dHRGs signature consisted of *ALDOA*, *PFKL*, and *PGK1*, which performed well in predicting prognosis and distinguishing OS between high- and low-risk groups. All three genes (*ALDOA*, *PFKL*, and *PGK1*) considered as risk factors in TNBC patients by them were members of the hypoxia gene set containing 200 HRGs. In contrast to the previous study [14], neither *ARTN* nor *L1CAM* of the prognostic model constructed in this study belongs to the hypoxia

gene set. However, both of them were transcriptionally activated by *HIF1A* according to previous studies [53,58]. Specifically, *ARTN* was identified as a responsive factor to hypoxia that promotes tumor progression in hepatocellular carcinoma [58]. Therefore, the relationship between *ARTN* and hypoxia in TNBC deserves to be further investigated. Figure 9A showed the expression levels of *ARTN* and *L1CAM* in tumor tissues of different BRCA subtypes and normal tissues, which were investigated using TIMER. It indicated that both *ARTN* and *L1CAM* were up-regulated in TNBC (i.e., BRCA-Basal) compared to normal tissues. We further analyzed their expression correlation in four subtypes of BRCA. As illustrated in Figure 9B, *ARTN* and *L1CAM* had a significant positive correlation in TNBC, suggesting that their crucial roles in tumor progression. In summary, the hypoxic tumor subpopulation within TME tends to overexpress angiogenesis- and metastasis-related genes in response to hypoxia, thus presenting a self-perpetuating state. In addition, the marker genes of the hypoxic tumor subpopulation *ARTN* and *L1CAM* were indicators of poor prognosis for TNBC patients.

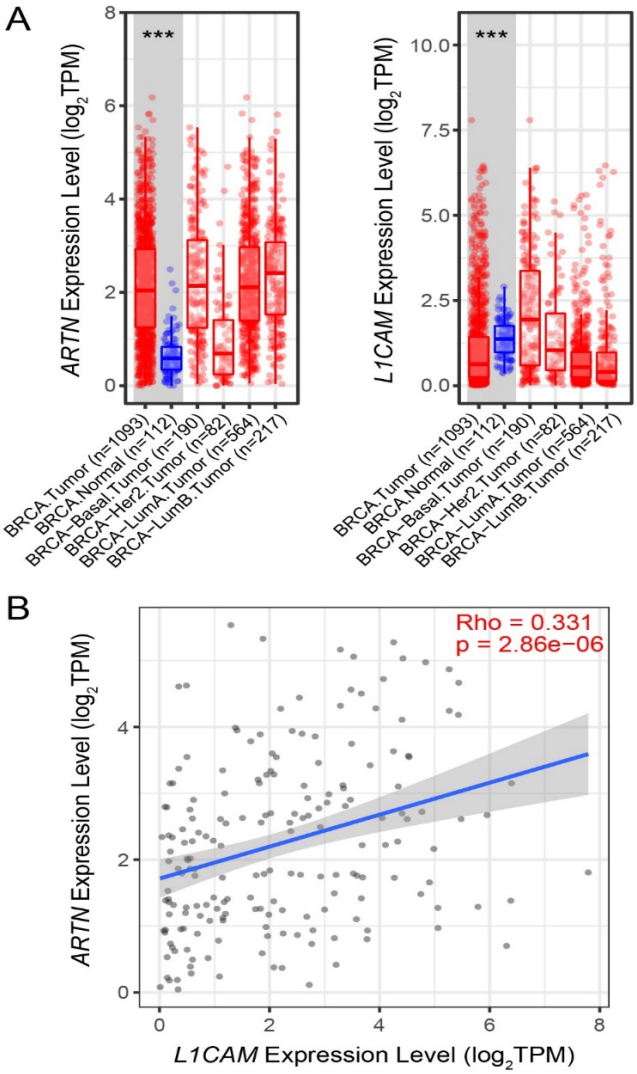

Figure 9. (A) The expression levels of *ARTN* and *L1CAM* in tumor tissues of different BRCA subtypes and normal tissues; (B) The correlation of expression levels between *ARTN* and *L1CAM* in TNBC.

4. Conclusions

In conclusion, we revealed the considerable heterogeneity within TME based on scRNA-seq data of TNBC, focusing on the roles of different cell subpopulations, especially epi_nor1 and epi_tum1 with high expression levels of *ESR1* and HRGs respectively. Subsequently, the intercellular communication network was inferred to further explore how distinct communication patterns were coordinately involved in diverse signaling pathways. Particularly, the hypoxic tumor subpopulation was identified as the predominant signal sender of angiogenesis- and metastasis-related signaling pathways such as CALCR, VEGF, GDNF, and L1CAM. Finally, *ARTN* and *L1CAM* were applied for the construction of the risk score model, both of which were risk factors during tumor progression, indicating the hypoxic tumor subpopulation with self-perpetuating capacity could cause a dismal prognosis.

Acknowledgments

This work was supported by the Beijing Natural Science Foundation (No. 2202002) and the Chinese Natural Science Foundation project (No. 21173014).

Conflict of interest

The authors declare that they have no conflicts of interest.

References

1. F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, A. Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA Cancer J. Clin*., **68** (2018), 394–424. https://doi.org/10.3322/caac.21492
2. R. Dent, M. Trudeau, K. I. Pritchard, W. M. Hanna, H. K. Kahn, C. A. Sawka, et al., Triple-negative breast cancer: clinical features and patterns of recurrence, *Clin. Cancer Res.*, **13** (2007), 4429–4434. https://doi.org/10.1158/1078-0432.CCR-06-3045
3. A. Marra, G. Viale, G. Curigliano, Recent advances in triple negative breast cancer: the immunotherapy era, *BMC Medicine*, **17** (2019), 90. https://doi.org/10.1186/s12916-019-1326-5
4. G. Zarrilli, G. Businello, M. V. Dieci, S. Paccagnella, V. Carraro, R. Cappellesso, et al., The tumor microenvironment of primitive and metastatic breast cancer: implications for novel therapeutic strategies, *Int. J. Mol. Sci.*, **21** (2020), 8102. https://doi.org/10.3390/ijms21218102
5. S. D. Soysal, A. Tzankov, S. E. Muenst, Role of the tumor microenvironment in breast cancer, *Pathobiology*, **82** (2015), 142–152. https://doi.org/10.1159/000430499
6. P. V. Loo, T. Voet, Single cell analysis of cancer genomes, *Curr. Opin. Genet. Dev.*, **24** (2014), 82–91. https://doi.org/10.1016/j.gde.2013.12.004
7. A. A. Pollen, T. J. Nowakowski, J. Shuga, X. Wang, A. A. Leyrat, J. H. Lui, et al., Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex, *Nat. Biotechnol.*, **32** (2014), 1053–1058. https://doi.org/10.1038/nbt.2967

8. B. Treutlein, D. G. Brownfield, A. R. Wu, N. F. Neff, G. L. Mantalas, F. H. Espinoza, et al., Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq, *Nature*, **509** (2014), 371–375. https://doi.org/10.1038/nature13173
9. Q. H. Nguyen, N. Pervolarakis, K. Blake, D. Ma, R. T. Davis, N. James, et al., Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity, *Nat. Commun.*, **9** (2018), 2028. https://doi.org/10.1038/s41467-018-04334-1
10. M. Bartoschek, N. Oskolkov, M. Bocci, J. Lövrot, C. Larsson, M. Sommarin, et al., Spatially and functionally distinct subclasses of breast cancer-associated fibroblasts revealed by single cell RNA sequencing, *Nat. Commun.*, **9** (2018), 5150. https://doi.org/10.1038/s41467-018-07582-3
11. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: the next generation, *Cell*, **144** (2011), 646–674. https://doi.org/10.1016/j.cell.2011.02.013
12. B. Muz, P. de la Puente, F. Azab, A. K. Azab, The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy, *Hypoxia (Auckl)*, **3** (2015), 83–92. https://doi.org/10.2147/HP.S93413
13. X. Jing, F. Yang, C. Shao, K. Wei, M. Xie, H. Shen, et al., Role of hypoxia in cancer therapy by regulating the tumor microenvironment, *Mol. Cancer*, **18** (2019), 157. https://doi.org/10.1186/s12943-019-1089-9
14. X. Sun, H. Luo, C. Han, Y. Zhang, C. Yan, Identification of a hypoxia-related molecular classification and hypoxic tumor microenvironment signature for predicting the prognosis of patients with triple-negative breast cancer, *Front. Oncol.*, **11** (2021), 700062. https://doi.org/10.3389/fonc.2021.700062
15. X. Yang, X. Weng, Y. Yang, M. Zhang, Y. Xiu, W. Peng, et al., A combined hypoxia and immune gene signature for predicting survival and risk stratification in triple-negative breast cancer, *Aging (Albany NY)*, **13** (2021), 19486–19509. https://doi.org/10.18632/aging.203360
16. R. Gao, S. Bai, Y. C. Henderson, Y. Lin, A. Schalck, Y. Yan, et al., Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes, *Nat. Biotechnol.*, **39** (2021), 599–608. https://doi.org/10.1038/s41587-020-00795-2
17. M. J. Goldman, B. Craft, M. Hastie, K. Repečka, F. McDade, A. Kamath, et al., Visualizing and interpreting cancer genomics data via the Xena platform, *Nat. Biotechnol.*, **38** (2020), 675–678. https://doi.org/10.1038/s41587-020-0546-8
18. D. M. Gendoo, N. Ratanasirigulchai, M. S. Schroder, L. Paré, J. S. Parker, A. Prat, et al., Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer, *Bioinformatics*, **32** (2016), 1097–1099. https://doi.org/10.1093/bioinformatics/btv693
19. P. Jezequel, D. Loussouarn, C. Guerin-Charbonnel, L. Campion, A. Vanier, W. Gouraud, et al., Gene-expression molecular subtyping of triple-negative breast cancer tumours: importance of immune response, *Breast Cancer Res.*, **17** (2015), 43. https://doi.org/10.1186/s13058-015-0550-y
20. Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck III, S. Zheng, A. Butler, et al., Integrated analysis of multimodal single-cell data, *Cell*, **184** (2021), 3573–3587. https://doi.org/10.1016/j.cell.2021.04.048
21. D. Aran, A. P. Looney, L. Liu, E. Wu, V. Fong, A. Hsu, et al., Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage, *Nat. Immunol.*, **20** (2019), 163–172. https://doi.org/10.1038/s41590-018-0276-y

22. A. Liberzon, C. Birger, H. Thorvaldsdottir, M. Ghandi, J. P. Mesirov, P. Tamayo, The Molecular Signatures Database (MSigDB) hallmark gene set collection, *Cell Syst.*, **1** (2015), 417–425. https://doi.org/10.1016/j.cels.2015.12.004
23. S. Jin, C. F. Guerrero-Juarez, L. Zhang, I. Chang, R. Ramos, C. H. Kuan, et al., Inference and analysis of cell-cell communication using CellChat, *Nat. Commun.*, **12** (2021), 1088. https://doi.org/10.1038/s41467-021-21246-9
24. J. F. Prud'homme, F. Fridlansky, M. Le Cunff, M. Atger, C. Mercier-Bodart, M. F. Pichon, et al., Cloning of a gene expressed in human breast cancer and regulated by estrogen in MCF-7 cells, *DNA*, **4** (1985), 11–21. https://doi.org/10.1089/dna.1985.4.11
25. M. C. Rio, J. P. Bellocq, J. Y. Daniel, C. Tomasetto, R. Lathe, M. P. Chenard, et al., Breast cancer-associated pS2 protein: synthesis and secretion by normal stomach mucosa, *Science*, **241** (1988), 705–708. https://doi.org/10.1126/science.3041593
26. T. Vogl, A. Stratis, V. Wixler, T. Völler, S. Thurainayagam, S. K. Jorch, et al., Autoinhibitory regulation of S100A8/S100A9 alarmin activity locally restricts sterile inflammation, *J. Clin. Invest*., **128** (2018), 1852–1866. https://doi.org/10.1172/JCI89867
27. Q. Fang, S. Yao, G. Luo, X. Zhang, Identification of differentially expressed genes in human breast cancer cells induced by 4-hydroxyltamoxifen and elucidation of their pathophysiological relevance and mechanisms, *Oncotarget*, **9** (2018), 2475–2501. https://doi.org/10.18632/oncotarget.23504
28. N. O'Brien, T. M. Maguire, N. O'Donovan, N. Lynch, A. D. Hill, E. McDermott, et al., Mammaglobin a: a promising marker for breast cancer, *Clin. Chem.*, **48** (2002), 1362–1364. https://doi.org/10.1093/clinchem/48.8.1362
29. M. Zafrakas, B. Petschke, A. Donner, F. Fritzsche, G. Kristiansen, R. Knüchel, et al., Expression analysis of mammaglobin A (SCGB2A2) and lipophilin B (SCGB1D2) in more than 300 human tumors and matching normal tissues reveals their co-expression in gynecologic malignancies, *BMC Cancer*, **6** (2006), 88. https://doi.org/10.1186/1471-2407-6-88
30. D. Carter, J. F. Douglass, C. D. Cornellison, M. W. Retter, J. C. Johnson, A. A. Bennington, et al., Purification and characterization of the mammaglobin/lipophilin B complex, a promising diagnostic marker for breast cancer, *Biochemistry*, **41** (2002), 6714–6722. https://doi.org/10.1021/bi0159884
31. T. L. Colpitts, P. Billing-Medel, P. Friedman, E. N. Granados, M. Hayden, S. Hodges, et al., Mammaglobin is found in breast tissue as a complex with BU101, *Biochemistry*, **40** (2001), 11048–11059. https://doi.org/10.1021/bi010284f
32. S. Robson, S. Pelengaris, M. Khan, c-Myc and downstream targets in the pathogenesis and treatment of cancer, *Recent Pat. Anticancer Drug Discov.*, **1** (2006), 305–326. https://doi.org/10.2174/157489206778776934
33. P. C. Fernandez, S. R. Frank, L. Wang, M. Schroeder, S. Liu, J. Greene, et al., Genomic targets of the human c-Myc protein, *Genes Dev.*, **17** (2003), 1115–1129. https://doi.org/10.1101/gad.1067003
34. J. H. Patel, A. P. Loboda, M. K. Showe, L. C. Showe, S. B. McMahon, Analysis of genomic targets reveals complex functions of MYC, *Nat. Rev. Cancer*, **4** (2004), 562–568. https://doi.org/10.1038/nrc1393
35. C. Attwooll, E. L. Denchi, K. Helin, The E2F family: specific functions and overlapping interests, *EMBO J.*, **23** (2004), 4709–4716. https://doi.org/10.1038/sj.emboj.7600481

36. T. Yu, L. Liang, X. Zhao, Y. Yin, Structural and biochemical studies of the extracellular domain of Myelin protein zero-like protein 1, *Biochem. Biophys. Res. Commun.*, **506** (2018), 883–890. https://doi.org/10.1016/j.bbrc.2018.10.161
37. K. M. McCarthy, I. B. Skare, M. C. Stankewich, M. Furuse, S. Tsukita, R. A. Rogers, et al., Occludin is a functional component of the tight junction, *J. Cell Sci.*, **109** (1996), 2287–2298. https://doi.org/10.1242/jcs.109.9.2287
38. J. Sakata, T. Shimokubo, K. Kitamura, S. Nakamura, K. Kangawa, H. Matsuo, et al., Molecular cloning and biological activities of rat adrenomedullin, a hypotensive peptide, *Biochem. Biophys. Res. Commun.*, **195** (1993), 921–927. https://doi.org/10.1006/bbrc.1993.2132
39. K. Miyashita, H. Itoh, N. Sawada, Y. Fukunaga, M. Sone, K. Yamahara, et al., Adrenomedullin promotes proliferation and migration of cultured endothelial cells, *Hypertens. Res.*, **26** Suppl (2003), S93–98. https://doi.org/10.1291/hypres.26.S93
40. N. Ferrara, H. P. Gerber, J. LeCouter, The biology of VEGF and its receptors, *Nat. Med.*, **9** (2003), 669–676. https://doi.org/10.1038/nm0603-669
41. M. I. Lin, W. C. Sessa, Vascular endothelial growth factor signaling to endothelial nitric oxide synthase: more than a FLeeTing moment, *Circ. Res.*, **99** (2006), 666–668. https://doi.org/10.1161/01.RES.0000245430.24075.a4
42. S. Sugo, N. Minamino, K. Kangawa, K. Miyamoto, K. Kitamura, J. Sakata, et al., Endothelial cells actively synthesize and secrete adrenomedullin, *Biochem. Biophys. Res. Commun.*, **201** (1994), 1160–1166. https://doi.org/10.1006/bbrc.1994.1827
43. M. J. Miller, A. Martinez, E. J. Unsworth, C. J. Thiele, T. W. Moody, T. Elsasser, et al., Adrenomedullin expression in human tumor cell lines. Its potential role as an autocrine growth factor, *J. Biol. Chem.*, **271** (1996), 23345–23351. https://doi.org/10.1074/jbc.271.38.23345
44. K. Dawas, M. Loizidou, A. Shankar, H. Ali, I. Taylor, Angiogenesis in cancer: the role of endothelin-1, *Ann. R. Coll. Surg. Engl.*, **81** (1999), 306–310.
45. N. Zhu, L. Gu, J. Jia, X. Wang, L. Wang, M. Yang, W. Yuan, Endothelin-1 triggers human peritoneal mesothelial cells' proliferation via ERK1/2-Ets-1 signaling pathway and contributes to endothelial cell angiogenesis, *J. Cell. Biochem.*, **120** (2019), 3539–3546. https://doi.org/10.1002/jcb.27631
46. Y. Katanasaka, Y. Kodera, Y. Kitamura, T. Morimoto, T. Tamura, F. Koizumi, Epidermal growth factor receptor variant type III markedly accelerates angiogenesis and tumor growth via inducing c-myc mediated angiopoietin-like 4 expression in malignant glioma, *Mol. Cancer*, **12** (2013), 31. https://doi.org/10.1186/1476-4598-12-31
47. R. Kolb, P. Kluz, Z. W. Tan, N. Borcherding, N. Bormann, A. Vishwakarma, et al., Obesity-associated inflammation promotes angiogenesis and breast cancer via angiopoietin-like 4, *Oncogene*, **38** (2019), 2351–2363. https://doi.org/10.1038/s41388-018-0592-6
48. M. S. Airaksinen, M. Saarma, The GDNF family: signalling, biological functions and therapeutic value, *Nat. Rev. Neurosci.*, **3** (2002), 383–394. https://doi.org/10.1038/nrn812
49. J. Kang, P. X. Qian, V. Pandey, J. K. Perry, L. D. Miller, E. T. Liu, et al., Artemin is estrogen regulated and mediates antiestrogen resistance in mammary carcinoma, *Oncogene*, **29** (2010), 3228–3240. https://doi.org/10.1038/onc.2010.71
50. J. Kang, J. K. Perry, V. Pandey, G. C. Fielder, B. Mei, P. X. Qian, et al., Artemin is oncogenic for human mammary carcinoma cells, *Oncogene*, **28** (2009), 2034–2045. https://doi.org/10.1038/onc.2009.66

51. A. Banerjee, Z. S. Wu, P. Qian, J. Kang, V. Pandey, D. X. Liu, et al., ARTEMIN synergizes with TWIST1 to promote metastasis and poor survival outcome in patients with ER negative mammary carcinoma, *Breast Cancer Res.*, **13** (2011), R112. https://doi.org/10.1186/bcr3054
52. A. Banerjee, P. Qian, Z. S. Wu, X. Ren, M. Steiner, N. M. Bougen, et al., Artemin stimulates radio- and chemo-resistance by promoting TWIST1-BCL-2-dependent cancer stem cell-like behavior in mammary carcinoma cells, *J. Biol. Chem.*, **287** (2012), 42502–42515. https://doi.org/10.1074/jbc.M112.365163
53. H. Zhang, C. C. Wong, H. Wei, D. M. Gilkes, P. Korangath, P. Chaturvedi, et al., HIF-1-dependent expression of angiopoietin-like 4 and L1CAM mediates vascular metastasis of hypoxic breast cancer cells to the lungs, *Oncogene*, **31** (2012), 1757–1770. https://doi.org/10.1038/onc.2011.365
54. A. Banerjee, Z. S. Wu, P. X. Qian, J. Kang, D. X. Liu, T. Zhu, et al., ARTEMIN promotes de novo angiogenesis in ER negative mammary carcinoma through activation of TWIST1-VEGF-A signaling, *PLoS One*, **7** (2012), e50098. https://doi.org/10.1371/journal.pone.0050098
55. A. Friedli, E. Fischer, Novak-Hofer I, S. Cohrs, K. Ballmer-Hofer, P. A. Schubiger, et al., The soluble form of the cancer-associated L1 cell adhesion molecule is a pro-angiogenic factor, *Int. J. Biochem. Cell Biol.*, **41** (2009), 1572–1580. https://doi.org/10.1016/j.biocel.2009.01.006
56. H. Hall, J. A. Hubbell, Matrix-bound sixth Ig-like domain of cell adhesion molecule L1 acts as an angiogenic factor by ligating alphavbeta3-integrin and activating VEGF-R2, *Microvasc. Res.*, **68** (2004), 169–178. https://doi.org/10.1016/j.mvr.2004.07.001
57. H. Hall, V. Djonov, M. Ehrbar, M. Hoechli, J. A. Hubbell, Heterophilic interactions between cell adhesion molecule L1 and alphavbeta3-integrin induce HUVEC process extension in vitro and angiogenesis in vivo, *Angiogenesis*, **7** (2004), 213–223. https://doi.org/10.1007/s10456-004-1328-5
58. M. Zhang, W. Zhang, Z. Wu, S. Liu, L. Sun, Y. Zhong, et al., Artemin is hypoxia responsive and promotes oncogenicity and increased tumor initiating capacity in hepatocellular carcinoma, *Oncotarget*, **7** (2016), 3267–3282. https://doi.org/10.18632/oncotarget.6572

Appendix

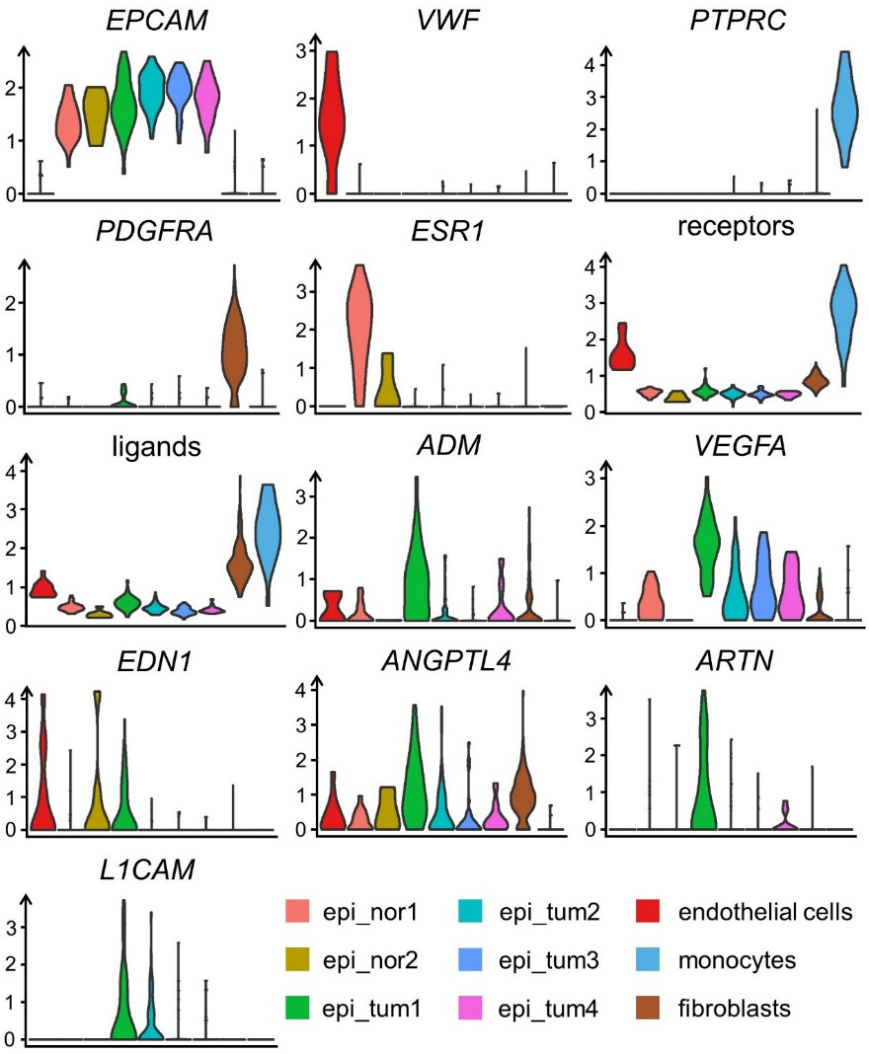

Figure A1. Violin plot presenting the relative expression levels of certain genes among all cell subpopulations.