scientific reports

Check for updates

OPEN

Construction and validation of antibody dependent cell phagocytosis related risk model in breast cancer

Xiwei Jiang, Jinqing Zhang & Xiaoming Liu✉

Antibody-dependent cellular phagocytosis (ADCP) has potential in breast cancer (BRCA) treatment, but the prognostic significance of ADCP-associated regulators (PRs) in BRCA remains unclear, necessitating this study. Firstly, candidate genes were identified by crossing differentially expressed genes (DEGs) with PRs, and prognostic genes were selected from these. A risk model based on SIAH2 and PIGR was constructed and validated in the GSE42568 cohort. We also assessed the relationship between expression of prognostic genes and clinical characteristics. From these, independent prognostic factors were selected and a nomogram was constructed to predict survival in BRCA patients. Immune microenvironment analysis revealed that PIGR had the strongest positive correlation with naive B cells, while SIAH2 showed the strongest negative correlation with activated memory CD4 T cells. Moreover, drug sensitivity analysis revealed that 109 drugs differed between high- and low-risk groups. Meanwhile, single-cell analysis identified epithelial cells (EPCs) as key cells, with PIGR highly expressed in the middle stage and SIAH2 in the early stage of cell differentiation. Finally, reverse transcription quantitative polymerase chain reaction (RT-qPCR) confirmed the up-regulation of SIAH2 and down-regulation of PIGR in BRCA. These findings suggest that SIAH2 and PIGR are potential prognostic genes and novel therapeutic targets for BRCA management.

Keywords Breast cancer, Antibody-dependent cellular phagocytosis, SIAH2, PIGR, Immune microenvironment, Risk model

Breast cancer (BRCA) is a prevalent malignancy in women, representing a significant health concern. As of the 2020 global cancer statistics, BRCA has overtaken lung cancer to become the most frequently diagnosed cancer type[1]. BRCA is categorized into Luminal A type, Luminal B type, HER2 positive type, and triple negative type based on variations in targeted receptor expressions[2]. The occurrence of breast cancer is usually attributed to mutations in BRCA susceptibility gene, TP53 tumor suppressor genes, abnormal mismatch repair, and the activation of classic cancer promoting pathways, such as PI3K/AKT/mTOR, Wnt/β-catenin, and the progesterone-induced RANKL/RANK system. At present, the treatment methods for BRCA mainly encompass surgical resection, endocrine therapy, chemotherapy, molecular targeted therapy and radiation therapy, Trastuzumab, known to bind to HER2 and impede the PI3K/AKT pathway activation, has demonstrated broad clinical utility in managing BRCA[3]. However, the mortality rate associated with BRCA remains very high. In recent years, many new biomarkers have emerged for BRCA, for example, serum LRG1[4], BUBR1[5], etc. Given the considerable heterogeneity across different BRCA subtypes, integrating other significant factors is imperative for enhancing the analysis and prognostication of BRCA outcomes.

Antibody dependent cellular phagocytosis (ADCP) refers to the activation of monocytes, macrophages, neutrophils, and dendritic cells for engulfing diseased cells by binding to corresponding antibodies[6]. Cellular phagocytosis is an important component of the immune system's immune regulation. By activating the cellular phagocytosis function in cancer tissue has shown promise for effectively eliminated while preserving normal tissue. This plays an important role in cancer treatment. However, existing research predominantly concentrates on ADCP and its application in cancer treatment, with limited attention given to the prognostic value of ADCP-related regulators (PRs) in cancer. In this study, we combined with single cell analysis to deeply explore the potential role of PRs in breast cancer, providing a new theoretical basis for their application in cancer prognosis evaluation.

Department of Thyroid and Breast Surgery, Shandong Provincial Third Hospital, Shandong University, Jinan 250031, China. ✉email: liuxiaodao_go@163.com

Based on the vital role of ADCP in the occurrence and evolution of BRCA, we will further understand the role of ADCP related genes in regulating immune cell and immune microenvironment progression, thereby providing certain reference value for the pathogenesis and prognosis of BRCA. This study is based on the BRCA related datasets from The Cancer Genome Atlas (TCGA) database and Gene Expression Omnibus (GEO) database. Through bioinformatics methods, the prognostic value of PRs in BRCA was revealed and analyzed, and their potential molecular regulatory mechanisms in BRCA were explored, providing theoretical support for the development of new treatment methods and drugs in clinical experience.

Results

The 133 candidate genes were mainly related to BRCA immune-related functions

Through conducting differential expression analysis between the BRCA and normal groups, we identified a total of 5870 DEGs. Among these DEGs, 3336 genes were found to be up-regulated in BRCA while 2534 genes were down-regulated (Supplementary Fig. S1a,b). A sum of 133 candidate genes were acquired by intersecting 5870 DEGs and 730 PRs (Supplementary Fig. S1c). Subsequently, enrichment analysis was conducted for these candidate genes to facilitate the understanding of their potential biological functions. These candidate genes were significantly enriched to 507 GO entries, containing 435 biological processes (BPs), 27 cellular components (CCs), and 45 molecular functions (MFs). Specifically, the signaling pathways significantly enriched for candidate genes included lymphocyte differentiation, B cell differentiation, B cell proliferation, and so on (Supplementary Fig. S1d). At the same time, the 133 candidate genes were also significantly enriched in 9 KEGG pathways, such as FoxO signaling pathway, AMPK signaling pathway and B cell receptor signaling pathway, etc. (Supplementary Fig. S1e). These results suggested that candidate genes might influence BRCA progression by affecting the function of B cells, lymphocytes, etc.

Prognostic gene expression correlated with survival in BRCA patients

SIAH2 and PIGR were identified as prognostic genes (Fig. 1a,b). The ROC curve demonstrated that the AUC values of SIAH2 and PIGR exceeded 0.8, indicating their excellent diagnostic potential for BRCA (Fig. 1c). Moreover, K–M survival curve revealed a notable disparity in the survival rates of BRCA patients based on the expression levels of SIAH2 and PIGR. Notably, individuals with low expression of PIGR exhibited comparatively shorter survival periods (Fig. 1d,e).

Risk model displayed strong proficiency in assessing the risk of BRCA patients

A risk model was constructed based on SIAH2 and PIGR. The risk score was calculated as: riskScore = (− 0.308) × SIAH2 expression level + (− 0.203) × PIGR expression level + (− 0.295). Figure 2a,b displayed the risk curves of the TCGA-BRCA cohort and GSE42568 cohort, respectively. K–M survival curve data showed that, in these two cohorts, survival time was significantly shorter in high risk patients than in low risk patients (Fig. 2c,d).

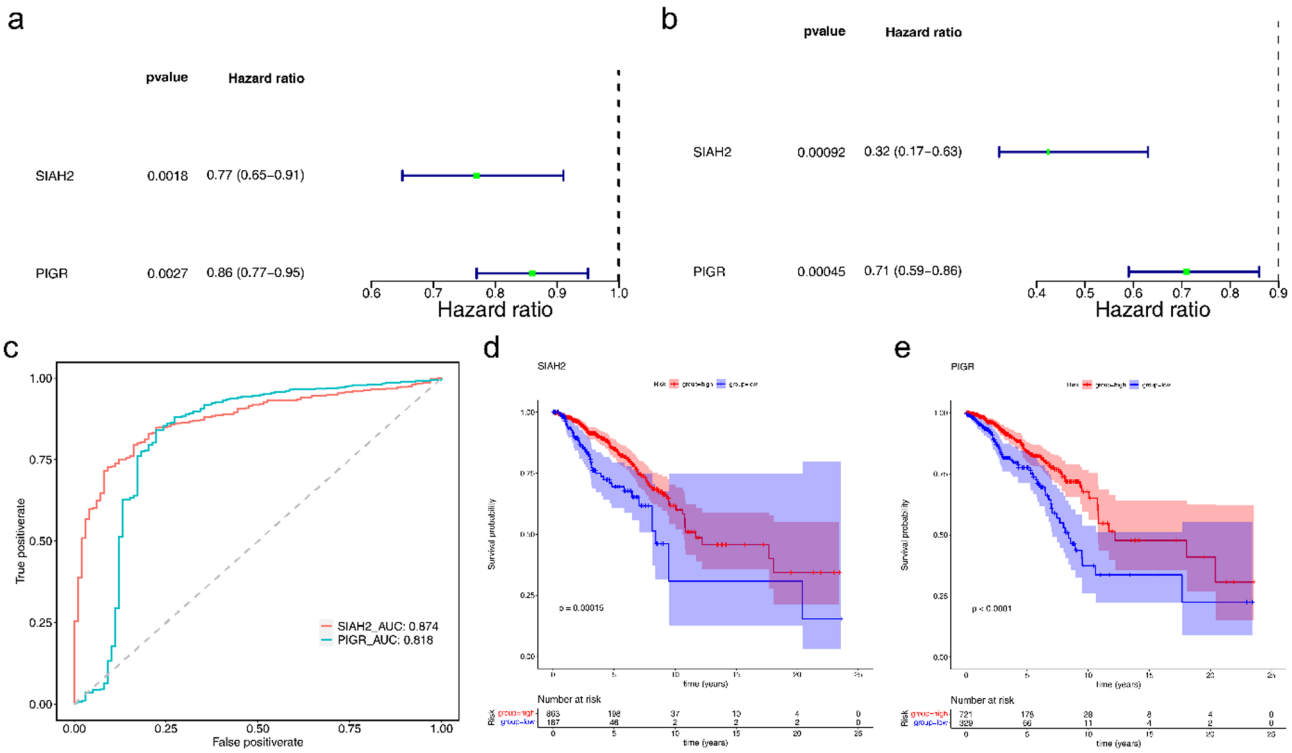

Fig. 1. Prognostic genes and survival analysis of BRCA patients. (**a**,**b**) Identification of BRCA prognostic genes. (**c**) ROC curves of BRCA. (**d**,**e**) K–M curve analysis of BRCA.

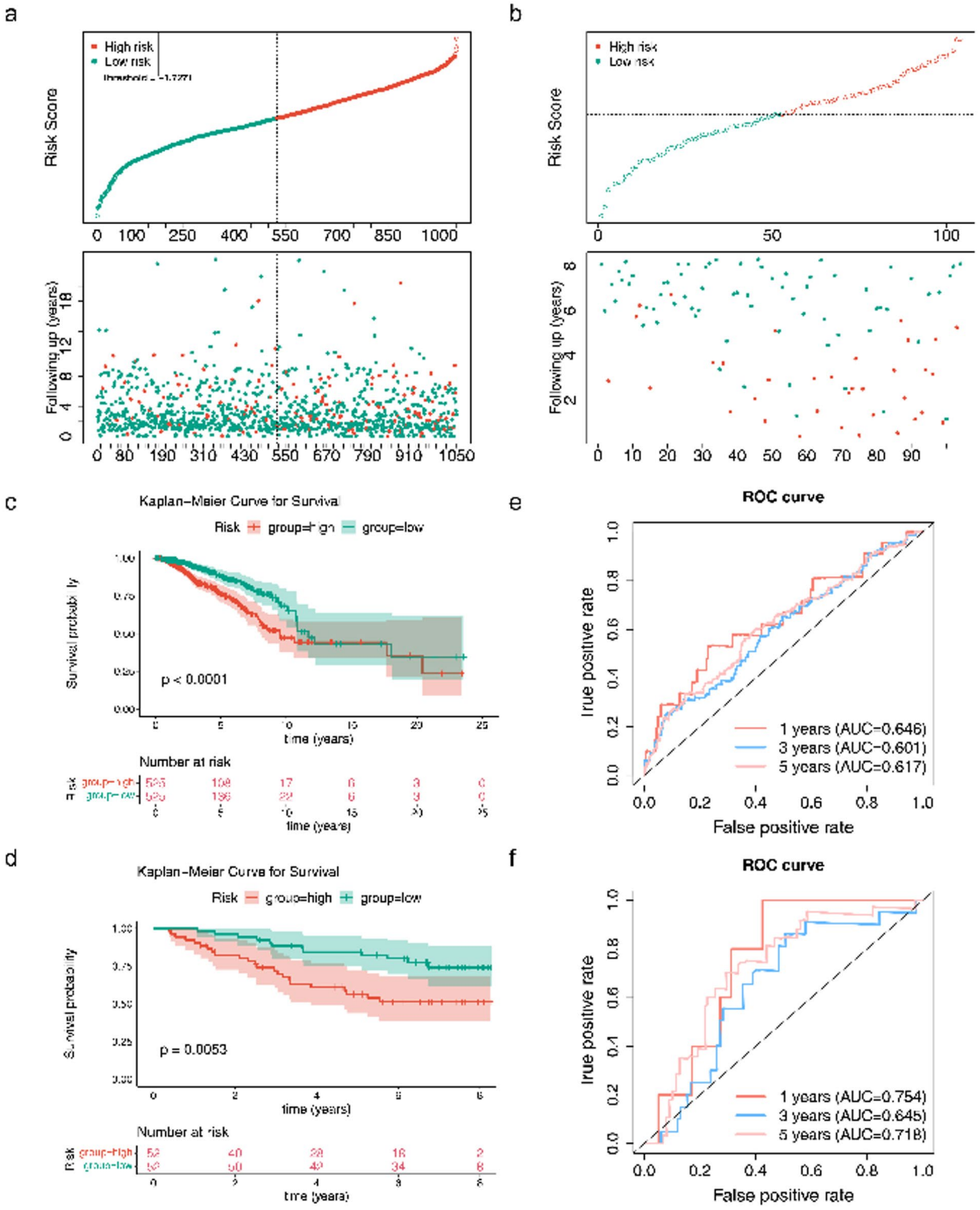

Fig. 2. Nomogram development and validation. (**a**,**b**) A risk modell was constructed based on SIAH2 and PIGR. (**c**,**d**) K–M survival curves in high- and low-risk patients. (**e**) ROC curves of high- and low-risk patients at 1, 3, and 5 years in the TCGA-BRCA dataset. (**f**) ROC curves for high and low risk patients in the GSE42568 group at 1, 3 and 5 years.

Additionally, the AUC values for 1-, 3-, and 5-year in the TCGA-BRCA cohort were 0.646, 0.601, and 0.617, respectively (Fig. 2e), while the AUC values for the same time period in the GSE42568 cohort were 0.754, 0.645, and 0.718, respectively (Fig. 2f), indicating a certain degree of predictive capacity in the risk model. These findings suggested that risk models constructed from prognostic genes had a certain predictive ability for BRCA and has been verified in another cohort.

The prognostic genes were related to clinical features

We further explored the relationship between prognostic genes and clinical features in TCGA-BRCA cohort. We observed that PIGR was significantly different between age, N-stage, stage, and T-stage (Fig. 3a), whereas SIAH2 was significantly different only in age (Fig. 3b).

A nomogram could effectively predict patients with BRCA

Screening in the TCGA BRCA cohort found that age, N stage and M stage were independent prognostic factors (Fig. 4a,b). A nomogram was developed to estimate the survival rates of BRCA patients at 1-, 3-, and 5-year intervals (Fig. 4c). The calibration curve demonstrated a high level of accuracy for the nomogram model with a slope close to 1 (Fig. 4d).

Association between immune microenvironment and risk score in patients with BRCA

In TCGA-BRCA cohort, we observed variations in the levels of 22 immune cell infiltrations across different risk groups (Supplementary Fig. S2a). Specifically, memory B cells, resting dendritic cells, naive B cells, resting mast cells, resting memory CD4 T cells, monocytes, plasma cells, and CD8 T cells exhibited significantly higher infiltration levels in the low-risk group. Resting NK cells, M0 macrophages, and activated memory CD4 T cells were the opposite (Supplementary Fig. S2b). Moreover, analysis of correlations indicated a significant positive association between PIGR and naive B cells (correlation coefficient (cor) = 0.32). Conversely, SIAH2 displayed a notable negative link with activated memory CD4 T cells (cor = − 0.21) (Supplementary Fig. S2c). In addition, we observed three immune checkpoints (CTLA4, IDO1, and PVR) were significantly higher in the high risk group (Supplementary Fig. S2d). Similarly, the correlation analysis between prognostic genes and immune checkpoints revealed that PIGR exhibited the strongest positive association with CD226 (cor = 0.27), whereas SIAH2 demonstrated the strongest negative correlation with PVR (cor = − 0.28) (Supplementary Fig. S2e). Moreover, Immunophenotype Score (IPS) analysis showed significant differences in all four components of PD1 and CTLA4 in both high and low risk groups, suggesting that the expression status of CTLA4 and PD1 may be significantly associated with the level of risk to which patients were exposed ($p < 0.05$) (Supplementary Fig. S2f).

109 Drugs differed between high and low-risk groups, with missense mutation being the most common gene mutation

To understand the drugs associated with prognostic genes, predictive analysis was performed, which showed that SIAH2 predicted 42 drugs and PIGR predicted 37 drugs. The prognostic genes-drug network was shown in Supplementary Fig. S3a. In addition, the rank sum test found 109 drugs differed between high and low risk groups, and the top 20 drugs with the smallest p-values were selected for presentation ($p < 0.05$) (Supplementary Fig. S3b). Next, mutation analysis was performed to understand the mutation profiles of the genes. The results showed that the highest mutation frequency in the low-risk group was TP53 and in the high-risk group was PIK3CA. The type of mutation with the highest mutation frequency in these genes was missense_mutation (Supplementary Fig. S3c,d). Moreover, Tumor Mutational Burden (TMB) scores differed between high and low risk groups, with higher TMB scores in the high risk group (Supplementary Fig. S3e).

A total of 7 cell types were annotated

After identifying prognostic genes through transcriptome analysis, we conducted further investigation into the expression distribution of these prognostic genes at the single-cell level. The data from the single-cell dataset GSE176078 were quality controlled (Supplementary Fig. S4), followed by normalisation of the data to screen for top 2000 highly variable genes (Fig. 5a), and principal component analysis (PCA) of the first 30 principal components of the highly variable genes ($p < 0.05$) yielded 21 cell clusters (Fig. 5b,c). Finally, 7 cell types were identified (Fig. 5d), which were visualized by Uniform Manifold Approximation and Projection (UMAP) and showed the expression of marker genes (Fig. 5e). Cellular functional enrichment analyses revealed that the different types of cell clusters were associated with a variety of biological pathways, including Reuptake of GABA, Aromatic amines can be N-hydroxylated or N-dealkylated by CYP1A2, Ficolins bind to repetitive carbohydrate structures on the target cell surface and so on (Fig. 5f).

The strongest interaction was between endothelial cells and monocyte

To further screen key cells, the distribution of prognostic genes expression in different cell types was demonstrated by bubble plot (Fig. 6a). The results showed that SIAH2 and PIGR were highly expressed in Epithelial cells (EPCs), thus defining EPCs as key cells. Moreover, the results of intercellular communication network analysis showed that the strongest interaction was between Endothelial cells and Monocyte (Fig. 6b,c). Subsequently, bubble diagram demonstrated receptor-ligand interactions between different cells (Fig. 6d).

PIGR was highly expressed during the middle stage, while SIAH2 was highly expressed in the early stage

In order to understand the trajectories of EPCs, cell trajectory analysis was performed, which determined the starting point of the proposed temporal trajectory for EPCs development. This analysis indicated that as the cells progressed away from this starting point, they underwent maturation in their development (Fig. 7a). Based on

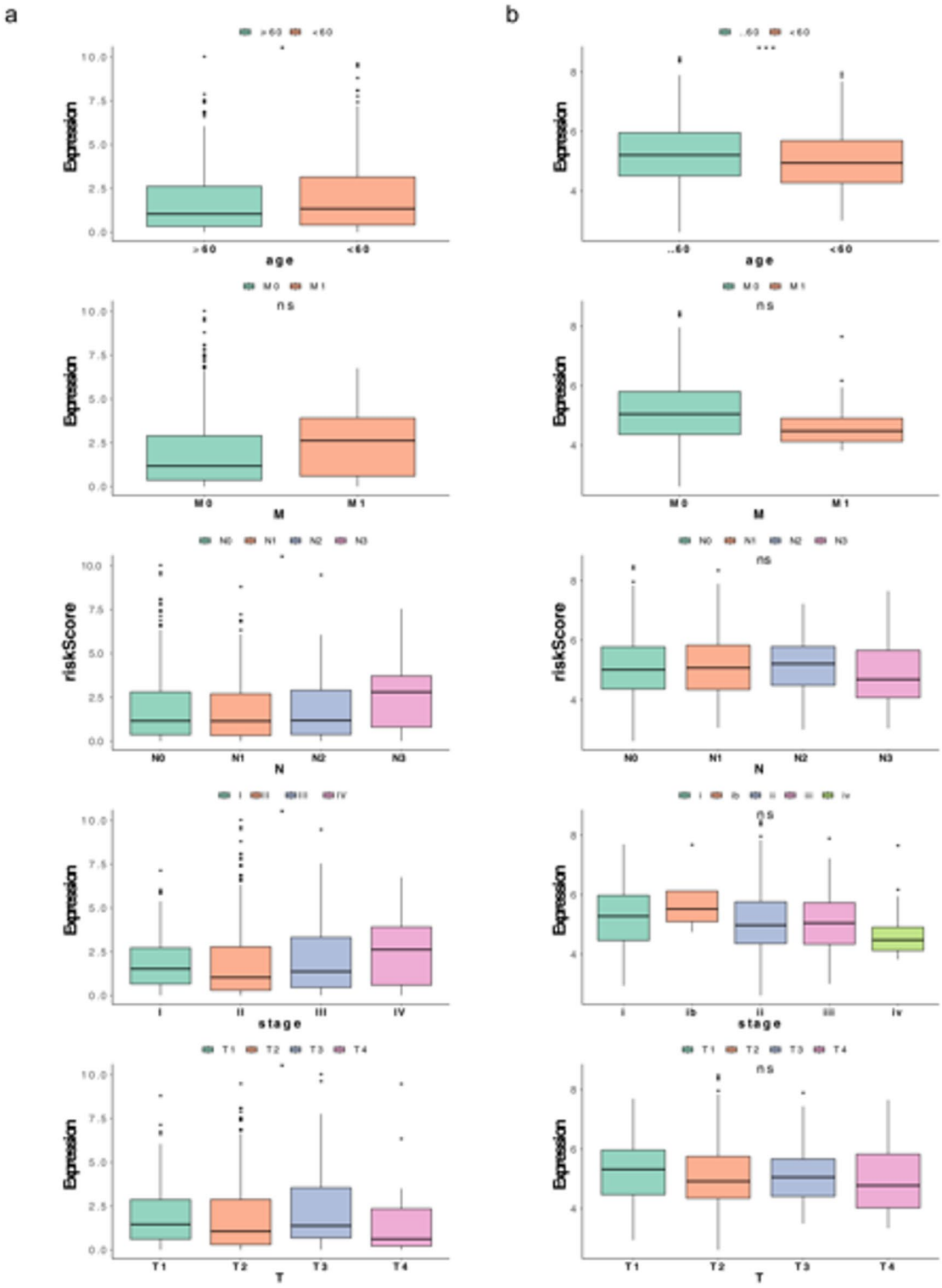

Fig. 3. The prognostic genes were related to clinical features. (**a**,**b**) The relationship between prognostic genes and clinical features.

their developmental stage, EPCs were divided into 7 periods, each represented by different colors. According to gene expression patterns, the samples were classified into cell populations at various differentiation states (Fig. 7b). Furthermore, during cell differentiation, the PIGR gene was highly expressed at the mid-differentiation stage, while the SIAH2 gene was highly expressed at the early differentiation stage (Fig. 7c).

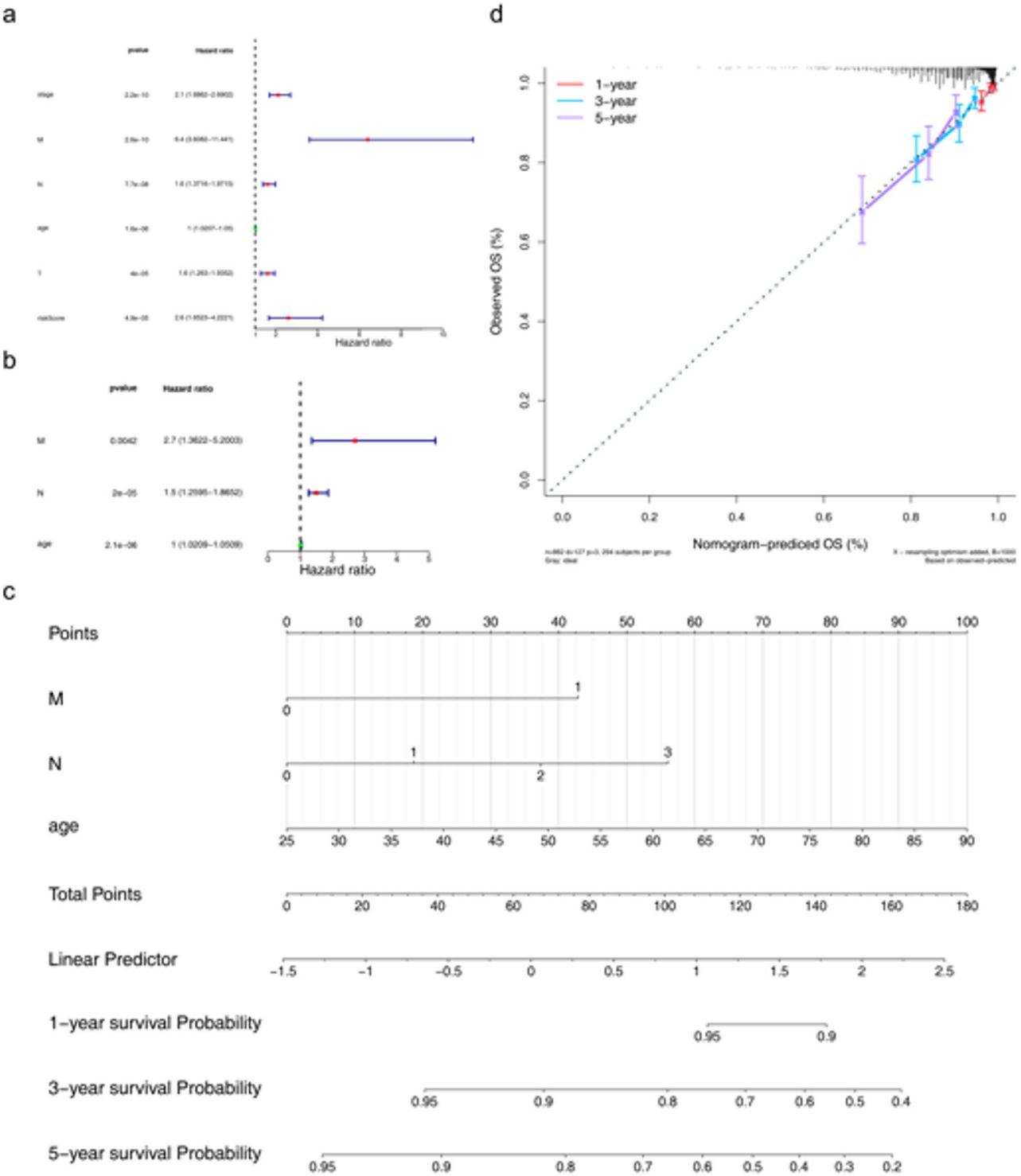

Fig. 4. Nomogram model construction. (**a**,**b**) Univariate and multivariate Cox regression analyses of risk scores and clinical characteristics in the training set. (**c**) Column line graph of independent prognostic factors for BRCA. (**d**) The BRCA column chart correction curve.

Differentially expressed genes in epithelial tumor cells were mainly enriched in Immunoglobulin complex and cell cycle pathways

Subsequently, we analyzed the differentially expressed genes in the high and low risk groups of epithelial cells. Firstly, 20,862 epithelial cells and 6900 endothelial cells were extracted from the single-cell dataset GSE176078, and the Copy Number Variation (CNV) Score of each cell was calculated based on the copy number quantitative analysis formula. As shown in Fig. 8a, cluster 2 represented endothelial cells, and the rest of the clusters were

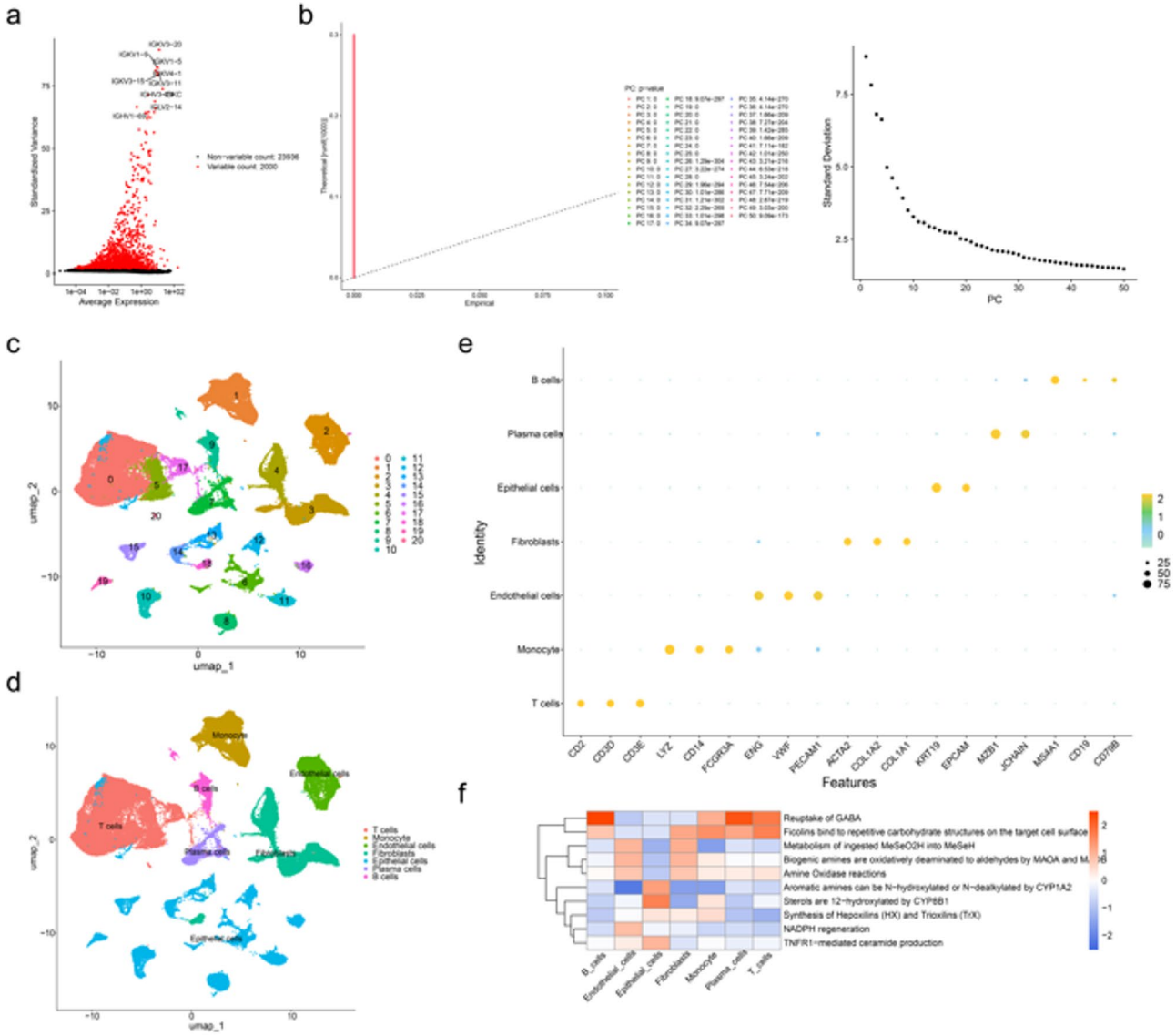

Fig. 5. A total of 7 cell types were annotated. (**a**) Screening for highly variable genes. (**b**) Jackstraw plot and Elbow Plot for principal component analysis. (**c**) UMAP clustering map for 21 clusters. (**d**) UMAP clustering map for 7 cell types. (**e**) Expression map of marker genes in clusters of cells. The size of the circle indicated the number of cells, the larger the circle, the higher the number of cells in the cluster; the yellower the colour, the higher the expression level of the marker genes. (**f**) The results of functional enrichment for cell clusters. The colours represented the degree of enrichment, with blue being a negative correlation and red a positive correlation.

epithelial cells. Next, the significance of the difference in CNV Score between epithelial and endothelial cells was calculated, and the results showed that there was a significant difference in CNV Score between them (Supplementary Fig. S5). Then, the epithelial cells were categorized into high and low risk groups based on the median risk score, and Fig. 8b demonstrated the clustering results in UMAP. Moreover, 445 differentially expressed genes were screened by differential expression analysis. And 228 significant GO pathways were enriched based on adj $p < 0.05$, covering 182 BPs, 28 CCs, and 18 MFs, of which the Top5 pathways were shown in Fig. 8c. In addition, 8 significant KEGG pathways were also derived from the enrichment analysis (adj $p < 0.05$) (Fig. 8d).

Expression evaluation of prognostic genes was implemented

The expression of prognostic genes was detected via qRT-PCR. The findings demonstrated a significant up-regulation of SIAH2 in BRCA compared to the control group, while PIGR exhibited a substantial down-regulation (Fig. 9).

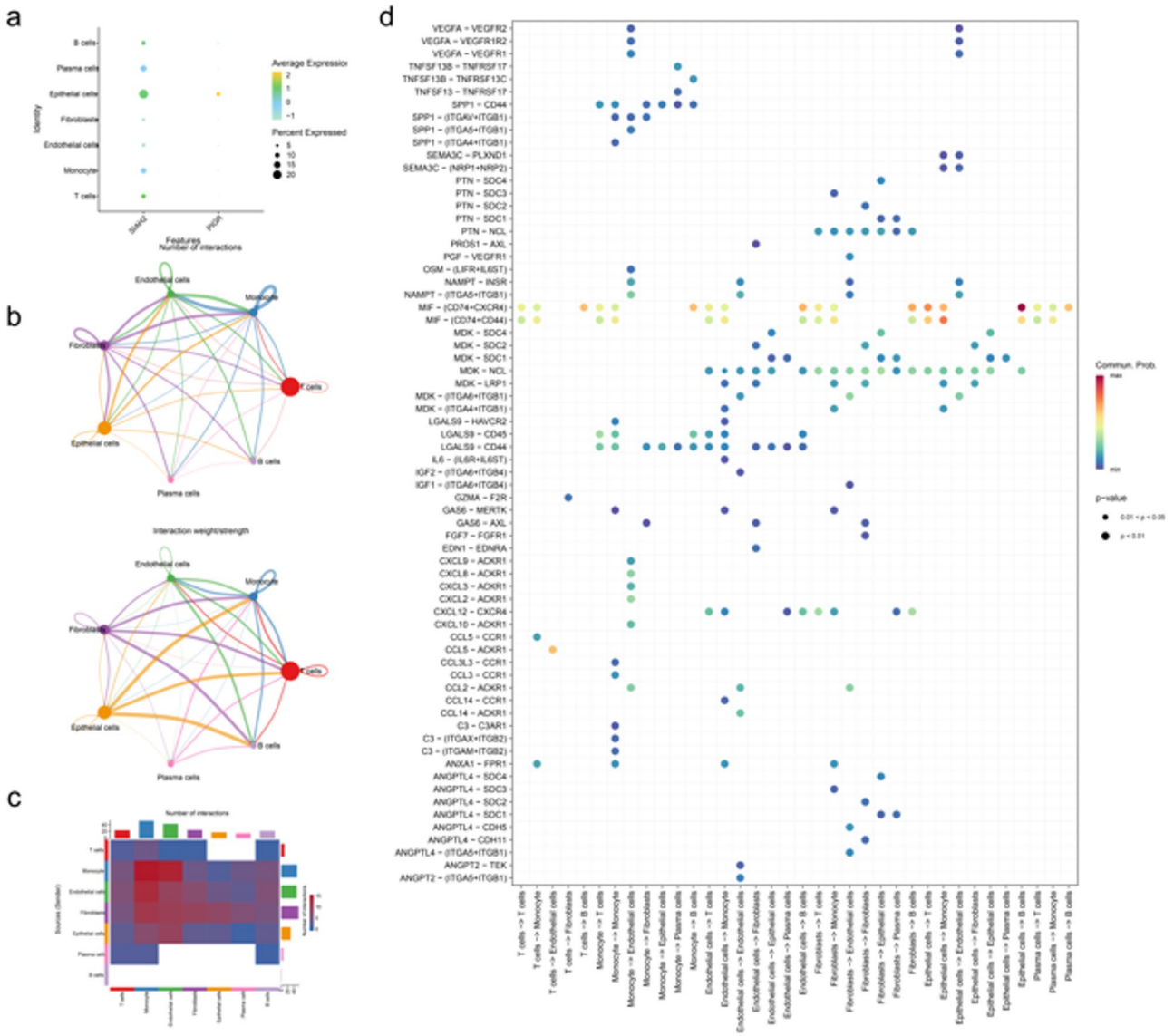

Fig. 6. The strongest interaction was between Endothelial cells and Monocyte. (**a**) Expression of prognostic genes in different cells. The gradual change in colour from green to yellow indicated a gradual increase in gene expression. (**b**) Cellular communication between all cell types. The thickness of the line represented the strength of the cellular interaction. (**c**) Heatmap of cellular communication between all cell types. The colour gradient from blue to red indicated an increase in the number of interactions. (**d**) Diagram of receptor-ligand communication between different cell types. The colours were graduated from blue to red, with redder colours indicating stronger interactions between the receptor and the ligand.

Discussion

BRCA is the most prevalent malignant tumour in women worldwide, and is also the second major factor in cancer related deaths. Immunotherapy has become increasingly important in many treatments for BRCA. ADCP is an immune elimination mechanism that is currently considered an important mechanism of monoclonal antibody therapy for cancer[7]. PRs have been confirmed to play a major role in the treatment of ADCP, but no research has reported the value of PRs in the prognosis of BRCA.

By taking the intersection of 5870 DEGs and 730 PRs, 133 candidate genes were ultimately obtained. Enrichment analysis of 133 candidate genes revealed that the regulation of monocyte differentiation, lymphocyte differentiation, white blood cell differentiation, regulation of the immune response by signaling pathways of cell surface receptors, cell cycle processes, B cell receptor signaling pathways, and primary immunodeficiency were associated with these candidate genes. This indicated a significant relationship between these candidate genes and the immune microenvironment.

Immune microenvironment is a hot topic in current tumor treatment. Among these, the PD-1/PD-L1 monoclonal antibody immune checkpoints inhibitor therapy for T cells has become the first-line treatment plan for BRCA. However, the significance and role of B cells in the anti-tumor immune process are rarely reported.

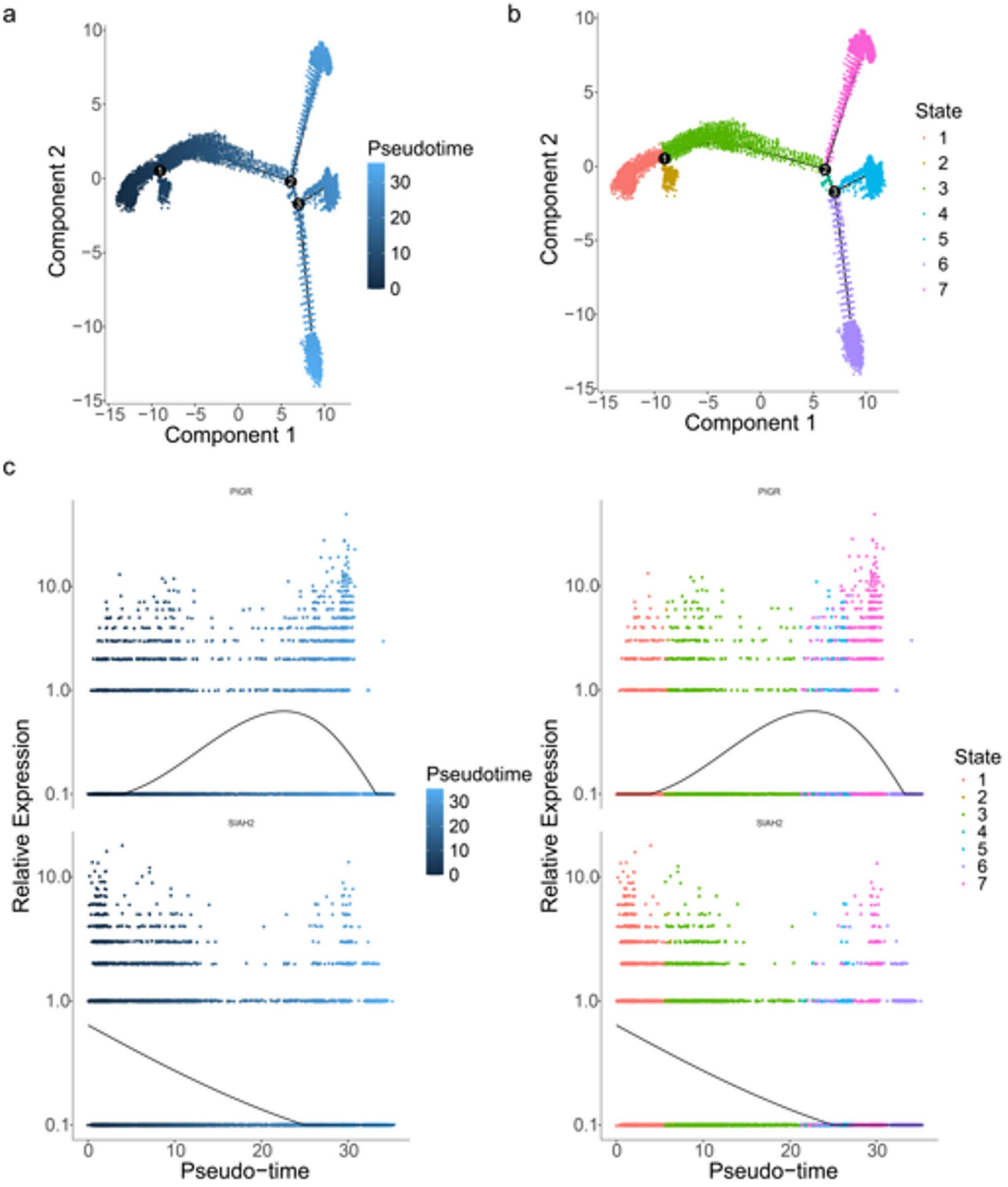

Fig. 7. PIGR was highly expressed during the middle stage, while SIAH2 was highly expressed in the early stage. (**a**) The cell trajectory plot displayed the changes in cell progression during the pseudotime process. (**b**) Divided into 7 periods according to the state of cell development, corresponding to different colours. (**c**) Dynamic expression profiles of prognostic genes. The vertical coordinate was the relative expression and the horizontal coordinate was the expression of prognostic genes under different time stages.

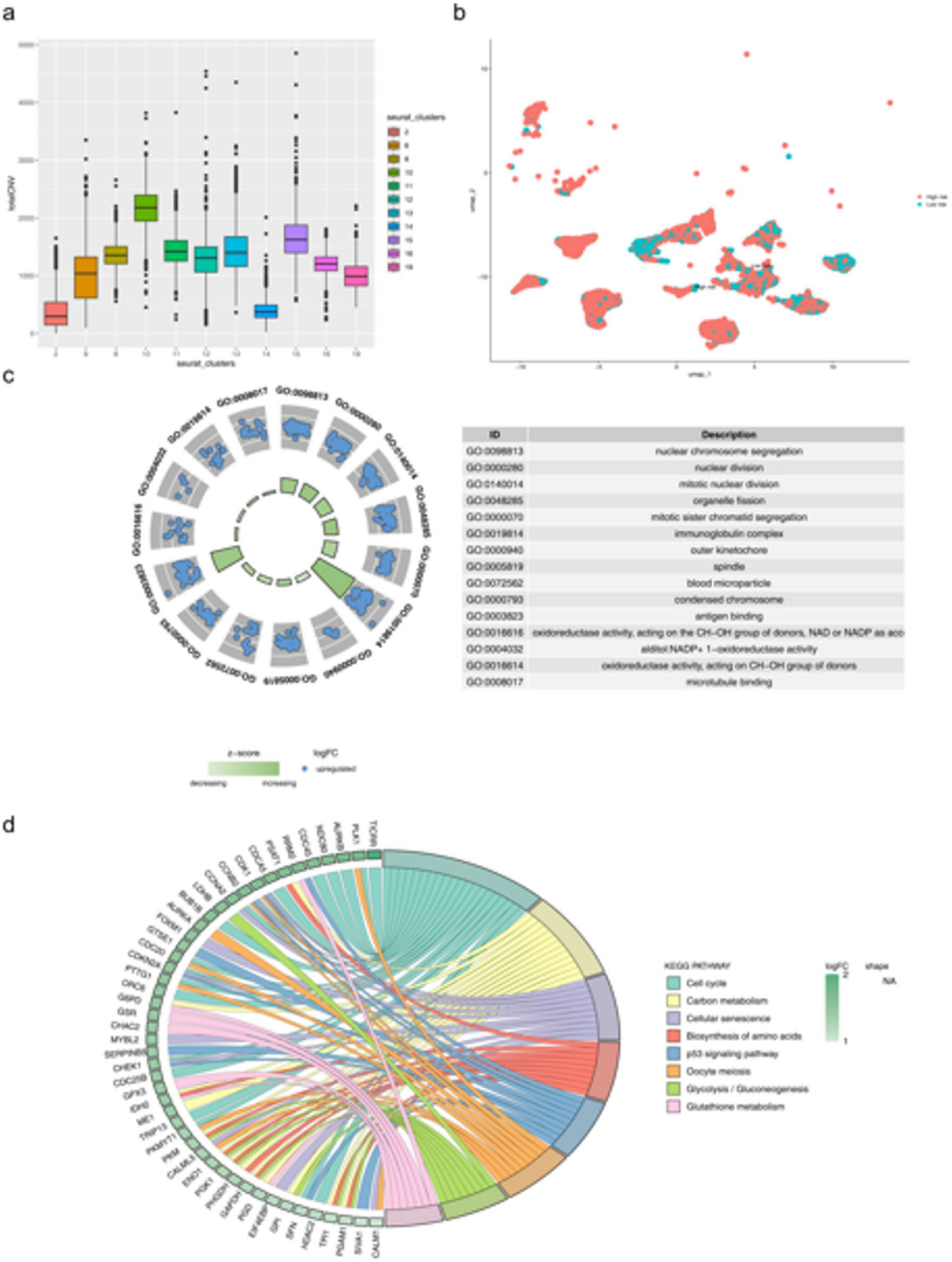

Fig. 8. Differentially expressed genes in epithelial tumor cells were mainly enriched in immunoglobulin complex and cell cycle pathways. (**a**) CNV Score distribution of epithelial and endothelial cells. (**b**) UMAP clustering results of epithelial cells categorized into high and low risk groups based on the median risk score. (**c**) Top 5 significant GO pathways enriched based on adj $p < 0.05$. (**d**) Top 8 significant KEGG pathways enriched based on adj $p < 0.05$.

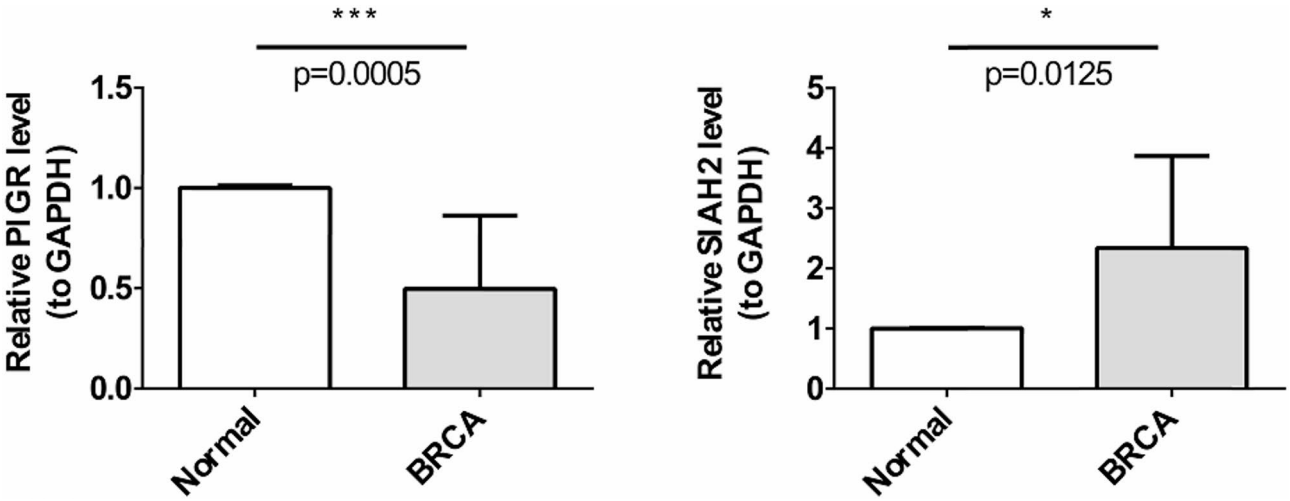

Fig. 9. The expression of prognostic genes was detected via qRT-PCR.

When faced with tumor antigen stimulation, B cells can secrete tumor specific antibodies and control tumor progression through mechanisms such as complement dependent cytotoxicity (CDC), antibody dependent cell-mediated cytotoxicity (ADCC), and ADCP[8,9]. The role of B cells in BRCA and its prognosis are still controversial. Some studies have found that a high level of invasion of B cells indicates a better prognosis of patients[10,11], while other studies have also shown that a high level of B cells invasion can lead to a worse prognosis of patients[12]. Therefore, the role and molecular mechanism of B cells in BRCA needs to be further investigated.

Univariate and multivariate Cox regression analyses of the 133 candidate genes in the TCGA dataset suggested that SIAH2 and PIGR play a critical role in the prognosis of BRCA patients. SIAH2 (Siah E3 Ubiquitin Protein Ligase 2) was a protein coding gene, and the translated SIAH2 protein functions as a ubiquitin ligase, adapting to hypoxic environments, regulating DNA damage, cell apoptosis, angiogenesis, and cell proliferation[13]. Therefore, SIAH2 plays an essential biological role in tumor occurrence and evolution, but the specific molecular mechanism is not fully understood. Studies have shown that overexpression of SIAH2 plays a role in promoting tumor progression in tissues such as prostate cancer[14], lung cancer[15], and liver cancer[16]. Similarly, SIAH2 also has a role in the promotion of breast cancer initiation, development and drug resistance[17–22]. This is the same as the conclusion of this study that high SIAH2 expression is associated with increased invasion of breast cancer and decreased overall survival rate of patients, and there are significant differences in the age of clinical characteristics. However, some studies have shown that SIAH2 plays a protective role in Luminal breast cancer[23,24]. This may be due to the inherent heterogeneity of breast cancer. In addition, studies have found that SIAH2 regulates the Hippo pathway in Triple negative breast cancer (TNBC) cells, affecting cell proliferation, stemness, chemotherapy sensitivity, and drug resistance. It is an important molecule in the mechanism of TNBC chemotherapy resistance[25]. In another study, SIAH2 also promoted the survival, apoptosis and invasion of breast cancer cells by regulating ERK and PI3K signaling pathways, thus playing a key role in the malignant transformation of breast cancer[26]. It was also found that SIAH2 promotes the sensitivity of breast cancer cells to tamoxifen by regulating the expression of ER - α in estrogen receptor (ER) positive breast cancer cells. Its low expression may be a mechanism leading to tamoxifen resistance[27]. Therefore, we need to further study the mechanism of SIAH2 in different breast cancer subtypes.

The Polymeric Immunoglobulin Receptor (PIGR) can be translated to produce the PIGR protein, which is a specific receptor for dimeric IgA (d IgA) and pentameric IgM (p IgM). It can mediate transmembrane transport and act as a bridge between innate and adaptive immunity[28]. The expression level of PIGR in different cancer tissues is different. Its high expression in liver cancer[29] and colon cancer[30]promotes tumor growth and metastasis, but it is low expression in lung cancer[31], ovarian cancer[32] and endometrial cancer[33]. At present, the role of PIGR in breast cancer is not completely clear, and the regulation and expression of PIGR by immune cells in different subtypes of breast cancer are also different[34,35]. The study found that PIGR is one of the down regulated genes in breast cancer tissue, and its down-regulation may play an inhibitory role in the occurrence of breast cancer, and is related to immune response, cell adhesion and migration[36]. In addition, PIGR is considered to be the key PANoptosis related gene, and its expression decline may be related to the progress of breast cancer and changes in the immune microenvironment, thus helping to predict the prognosis of breast cancer patients[37]. At the same time, PIGR is considered to be an important part of the competing endogenous RNA (ceRNA) network in breast cancer, and its expression changes may become a potential marker for prognosis evaluation of breast cancer[38]. Therefore, Our hypothesis was that PIGRs had different roles in the development of different subtypes of breast cancer, and its specific role and regulatory mechanism need further study. At the same time, through analysis in the TCGA dataset, the expression of PIGR gene was low in breast cancer, and there was significant difference among age, N stage, stage and T stage. the gene PIGR has significant differences between age, N stage, stage, and T stage.

This study constructed a risk model based on SIAH2 and PIGR, and the results showed that the model has a certain predictive ability for BRCA, which was validated in another cohort. Previous studies have shown that the risk model combining self-monitoring learning, data standardization and large-scale data sets can improve the diagnostic accuracy of pancreatic cancer and breast cancer, thus providing support for clinical decision-making, optimizing treatment plans, and improving patient prognosis[39]. Jacqueline C Stocking et al.'s research also suggests that risk models can help doctors make precise decisions, improve postoperative care, enhance patient outcomes, and optimize hospital resource allocation[40]. In addition, the risk model combining machine learning and clinical data can predict the risk of distant metastasis of breast cancer, improve the early diagnosis rate, improve the prognosis, and support the rational allocation of medical resources[41]. Our research results also show that by evaluating gene expression and clinical characteristics, risk models can help clinicians to conduct accurate stratification and prognosis assessment, provide basis for personalized treatment, and thus improve the survival rate and quality of life of breast cancer patients. However, this model still needs further validation in future clinical practice.

Through analysis of the intercellular communication network, we found that the interaction between endothelial cells and monocytes is the strongest. At the same time, the risk model based on SIAH2 and PIGR in this study shows potential value in predicting the prognosis of breast cancer. Therefore, we propose a hypothesis that PRs may affect the development and prognosis of tumors through cellular communication networks. Studies have found that SIAH2, as a protective gene, may play an inhibitory role in the progression of breast cancer by regulating cell communication and immune response, thus affecting the risk prediction in the prognosis model. This makes SIAH2 have important clinical significance in the prognosis evaluation of different breast cancer subtypes, especially Luminal BC subtype[42]. These findings further support our hypothesis. However, there are currently no research reports indicating that PIGR directly affects tumor development and prognosis through cellular communication networks. Therefore, future research can further explore the role of PIGR in cell communication and provide new insights and references for the treatment of breast cancer.

By further conducting immune microenvironment analysis, we discovered that there were significant differences in 11 different types of immune cells between the two groups at risk. Among them, PIGR showed a significant positive correlation with naive B cells, while SIAH2 displayed a notable negative connection with activated memory CD4 T cells. At present, there is little research on the relationship between PIGR and human B cells. It is only understood that overexpression of PIGR may lead to abnormal B cell receptor signaling, but the corresponding regulatory mechanism has not been elucidated[43], which requires further exploration. Similarly, previous report[44] has shown that high expression of SIAH2 promotes tumor growth by enhancing the proliferation and recruitment of regulatory T cells within tumors. This indicates that SIAH2 can affect the activation of immune cells, thereby altering the tumor immune microenvironment. However, the intrinsic relationship and mechanism between SIAH2 and activated memory CD4 T cells are still not fully understood.

In addition, immune checkpoints molecules such as CTLA-4, IDO1, and PVR exhibited relatively higher expressed levels in the high risk group. Currently, the targeted immunotherapy against CTLA-4 and PD-L1/PD-1 immune checkpoints has been recognized[45], igniting interest in exploring alternative immunotherapeutic targets. In BRCA, SIAH2 has a significant negative correlation with the PVR immune checkpoint. PVR, predominantly expressed on epithelial and myeloid cells is frequently upregulated in tumor cells. Patients with elevated PVR expression demonstrate reduced responsiveness to PD-1 monotherapy or combination therapy of PD-1 and CTLA-4[46]. Therefore, PVR immune checkpoints are poised to emerge as pivotal targets for immunotherapy in BRCA.

Conclusions

This study preliminarily explored the role of PRs in the prognosis of BRCA through multiple datasets from multiple databases. Subsequently, clinical specimens were collected for qRT-PCR and IHC validation. There are still some shortcomings in this study. Firstly, the small sample size of BRCA organizations may lead to insufficient statistical power and representativeness, thereby affecting the broad applicability of research results. Secondly, in the process of bioinformatics analysis, algorithm and parameter selection may introduce potential biases. Finally, this study mainly focuses on the prognostic significance and recurrence risk of PRs in BRCA. Therefore, in the future, the sample size should be expanded for validation, and multi center studies should be conducted to improve sample diversity and representativeness of results. In addition, although multiple methods were used for mutual verification in this study, potential biases in bioinformatics analysis can still be further reduced in the future by improving algorithms, optimizing parameters, and strengthening data quality control. In addition, more experimental studies are needed in the future to explore the specific molecular mechanisms underlying BRCA occurrence.

Materials and methods

Data acquisition

TCGA (https://portal.gdc.cancer.gov/) database offered RNA-seq data and survival information for the TCGA-BRCA cohort. This cohort contained 1171 samples, including 99 normal tissue samples and 1072 BRCA tissue samples, which was used as a training set. The cohort GSE42568, which contained survival information, was obtained from GEO (https://www.ncbi.nlm.nih.gov/) database using the sequencing platform GPL570. This validation set comprised 17 normal tissue samples and 104 BRCA tissue samples. The single-cell dataset GSE176078 (sequencing platform: GPL18573), also from the GEO database, contained a total of 26 tumour tissue samples and was sequenced using high-throughput sequencing for analysing the expression distribution of prognostic genes in the single-cell data. Additionally, a sum of 730 ADCP-related regulators (PRs) were gained from the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas9 system[47,48].

Identification of differentially expressed genes (DEGs) in TCGA-BRCA cohort

The DESeq2 package (v 1.36.0)[49] was used for differential expression analysis to identify DEGs between normal and BRCA samples in the TCGA-BRCA cohort. The conditions for screening were adj.$p < 0.05$ & |log2FoldChange(FC)| > 1. The ggplot2 package (v 3.3.6)[50] was employed to plot volcano map, while the pheatmap package (v 1.0.12)[51] was utilized to create heat map, both of which were employed to illustrate the expression of DEGs. Importantly, only up- and down-regulated top40 genes were selected as shown by heat map.

Recognition of candidate genes and functional enrichment analysis

Differentially expressed PRs were obtained by taking intersections of DEGs with PRs via VennDiagram package (v 1.7.3)[52], which were recorded as candidate genes. To deeply explore the biological functions involved in the candidate genes, enrichment analysis was performed in TCGA-BRCA cohort based on the clusterProfiler package (v 4.4.4)[53], containing Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) ($p < 0.05$). The enrichment results were visualized by drawing a treemap via treemap package (v 2.4-4).

Confirmation of prognostic genes and survival analysis

Candidate genes were sequentially integrated into univariate Cox regression analysis ($p < 0.05$), proportional hazards (PH) hypothesis test ($p > 0.05$), and multivariate Cox stepwise regression analysis ($p < 0.05$) to screen for prognostic genes conjunction with BRCA. Subsequently, the diagnostic performance of prognostic genes was evaluated by generating receiver operating characteristic (ROC) curves using timeROC package (v 0.4)[54]. In the same time, BRCA patients were categorised into two groups, high and low expression, based on the optimal threshold of gene expression levels that predict prognosis. To establish the difference in survival rates between these groups, we employed the survminer package (v 0.4.9)[55] to conduct Kaplan–Meier (K–M) survival curve analysis.

Development and validation of a risk model

In the TCGA-BRCA cohort, prognostic genes found in the previously stated analysis were used to build a risk model with risk scores calculated as follows: riskScore = $\sum_{i=1}^{n}(\beta_i \times Expr_i)$. Among them, βi and Expri represented the multivariate Cox stepwise regression coefficient and relative expression of prognostic gene i, respectively. Based on the risk score's median value, patients with BRCA were categorized into two distinct groups, a high-risk group and a low-risk group, based on their level of risk. The risk curve was employed to analyze the survival outcomes of patients across different risk groups. Afterwords, K–M survival curve analysis was implemented to investigate the survival differences between two risk groups. Meanwhile, ROC curves were presented through timeROC package, which facilitated the calculation of the area under curve (AUC) at 1-, 3-, and 5-year. A higher AUC value indicated a stronger predictive power of the risk model for BRCA. Furthermore, we implemented the validation of the risk model in GSE42568 cohort through the same analysis as in TCGA-BRCA cohort.

Clinical characteristics analysis

To further assess the correlation between the expression of prognostic genes and clinical features (including age, N stage, T stage, M stage, and stage), the expression of prognostic genes was analyzed between different groups with different clinical features in the TCGA-BRCA cohort ($p < 0.05$).

Construction and evaluation of nomogram

In the TCGA-BRCA cohort, to investigate independent prognostic factors, we performed univariate and multivariate Cox regression analyses ($p < 0.05$) by sequentially incorporating risk scores and various clinical features. Additionally, a PH hypothesis test was carried out to assess the proportional hazards assumption ($p > 0.05$). Based on the independent prognostic factors obtained from the above screening, a nomogram was constructed through rms package (v 6.3-0)[56]. A score was assigned to each independent prognostic factor, which was then added to the corresponding total score. Survival rates for BRCA patients at 1-, 3-, and 5-year intervals were estimated using the cumulative score. A higher score indicated a lower survival rate. Furthermore, the calibration curve was used to assess the precision of the nomogram's predictive ability.

Immune microenvironment analysis

The CIBERSORT algorithm was used to identify the infiltration frequency of 22 immune cells in different risk groups in the TCGA-BRCA cohort. A stacking histogram was plotted to present the results. Relevant risk groups were compared using the Wilcoxon test to assess variances in levels of immune cell infiltration ($p < 0.05$). What's more, Spearman correlation analysis was implemented to evaluate the correlation between prognostic genes and differential immune cell. At the same time, we also compared the differences in the expression of the immune checkpoints. (CTLA4, TIGIT, PVR, CD28, CD226, and IDO1)[57,58] between different risk groups ($p < 0.05$). Similarly, Spearman correlation analysis was implemented between prognostic genes and immune checkpoints. In addition, to assess the relationship between risk scores and immunotherapy, an IPS analysis was performed[59]. The IPS was a good predictor of response to checkpoint blockers, and the analysis included both PD1 and CTLA4 and was divided into four sections: ips_ctla4_neg_pd1_neg (CTLA4-negative, PD1-negative response), ips_ctla4_neg_pd1_pos (CTLA4-negative, PD1-positive reaction), ips_ctla4_pos_pd1_neg (CTLA4-positive, PD1-negative reaction) and ips_ctla4_pos_pd1_pos (CTLA4-positive, PD1-positive reaction). Differences in distribution of these four sections between high and low risk groups were analysed by Wilcoxon.

Drug prediction and mutation analysis

To explore the possibility of mitigating or treating the disease by targeting prognostic genes, prognostic genes were imported into the Comparative Toxicogenomics Database (CTD) database (https://ctdbase.org), drugs interacting with them were predicted, and prognostic genes-drug network was constructed by Cytoscape software (v 3.10.2)[60]. Subsequently, to assess the difference in response to chemotherapy between patients in high and low risk groups, 198 common chemotherapeutic drugs were obtained from the Genomics of Cancer Drug Sensitivity (GDSC) database (http://www.cancerrxgene.org/). And the pRRophetic R package (v 0.5)[61] was used to calculate the Inhibitory Concentration 50% (IC50) values of common chemotherapy and molecular targeted drugs in the TCGA-BRCA dataset. Then, a rank-sum test was performed to compare the IC50 differences between the high-risk and low-risk groups. In addition, in order to explore the mutation status between the high- and low-risk groups, the TCGAplot R package (v 0.3.0) was used to download the somatic mutation information of TCGA-BRCA, and the waterfall plots of the top 20 genes with the highest mutation frequencies were drawn. At the same time, to understand the mutation load of the tumours in the high- and low-risk groups, the TMB scores of each patient were calculated using the tmb function of the maftools R package (v 2.14.0)[62], and the differences in TMB scores between the two groups were compared by the Wilcoxon test.

Single-cell analysis

The single-cell dataset GSE176078 underwent rigorous data quality control using the Seurat R package (v 5.0.1)[63] to ensure compliance with the defined criteria. Cells were eliminated based on specific criteria: cells with fewer than 200 genes or fewer than 3 genes covered by cells, cells with fewer than 200 or more than 5000 expressed genes, genes with counts less than 500 or greater than 30,000, and cells with mitochondrial proportions exceeding 15%. The data were then normalised using the NormalizeData function in the Seurat R package (v 5.0.1)[63], and the genes with high variability were filtered by Analysis of Variance (ANOVA) using the FindVariableFeatures function. Finally, the LabelPoints function in the plotrix R package (v 3.8.2)[64] was used to visualise the results and indicate the top 10 genes with the highest variability in expression levels. After identifying highly variable genes, PCA was performed using the RunPCA function to reduce the dimensionality of the Single-Cell RNA Sequencing (scRNA-seq) data ($p < 0.05$) and the results of the dimensionality reduction were visualised by the ElbowPlot function. To further identify cell clusters, the UMAP algorithm (dims = 1:3, resolution = 0.1) was applied to perform a full dimensionality reduction analysis of the filtered principal components.

Subsequently, the FindAllMarkers function in the Seurat R package (v 5.0.1)[63] was used to identify marker genes specific to each cell cluster (logfc.threshold = 0.5, min.pct = 0.25, return. 0.25, return.thresh = 0.01). Next, these marker genes were used for type classification of cell clusters using SingleR R package (v 1.4.1)[65] and the CellMarker database (http://yikedaxue.slwshop.cn/). Finally, UMAP plot was created using the umap R package (v 0.2.10.0)[66] to show the annotated cell types. The expression of marker genes in different cell clusters was visualised by bubble plot. In addition, in order to explore the biological pathways of different cellular roles, the annotated cells were functionally enriched in the single-cell dataset using the ReactomeGSA R package (v 1.12.0)[67] ($p < 0.05$). To obtain key cells, the distribution of prognostic genes expression in each cell type was demonstrated using the UMAP method and the result was visualised using the theme function of the ggplot2 R package (v 3.3.6)[68]. Subsequently, the communication network between cells was analysed using the CellChat R package (v 1.6.1)[69]($p < 0.05$), and cell communication maps were created. More, bubble diagram was drawn to demonstrate the receptor-ligand interactions between different cells, and cell trajectories were constructed using the Monocle R package (v 2.30.1)[70] to analyse the proposed time series of key cells. Next, the Branched Expression Analysis Modeling (BEAM) method from the Monocle R package (v 2.30.1)[70] was used to analyze the pseudotime-ordered cell data and specified nodes. The plot_pseudo_time_heatmap function was then employed to generate a dynamic heatmap, illustrating the expression changes of prognostic genes across different pseudotimes. Finally, the relative expression profiles of prognostic genes in key cells along the pseudotime trajectory were plotted.

Classification of tumor cells in the epithelium into high and low risk groups based on risk scores

Next, epithelial tumor cells were classified into high-risk and low-risk groups based on the risk score, in order to explore the biological processes and signaling pathways that the differentially expressed genes between the two groups might be involved in. Firstly, epithelial and endothelial cells from the breast cancer single-cell dataset GSE176078 were extracted, and endothelial cells were used as a reference to calculate the copy number variation of epithelial cells. The CNV Score of the genes was calculated by reading the result file of infercnv (run.final.infercnv_obj) and based on the copy number results of each gene on each cell. Specifically, copy number quantification was performed based on the following formula:

$$\text{down} = \text{mean}(\text{rowMeans}(\text{tmp})) - 2 * \text{mean}(\text{apply}(\text{tmp}, 1, \text{sd}))$$

$$\text{up} = \text{mean}(\text{rowMeans}(\text{tmp})) + 2 * \text{mean}(\text{apply}(\text{tmp}, 1, \text{sd}))$$

where rowMeans represents the mean of the number of copies of each gene in all cells, tmp represents the result file of infercnv, rows represent genes, and columns represent samples. Next, the significance of the difference between epithelial cell clusters and endothelial cells on the CNV Score was calculated using the Wilcox test ($p < 0.05$). A corresponding risk score was calculated for each cell based on the riskScore formula: riskScore = (− 0.308) * SIAH2 expression level + (− 0.203) * PIGR expression level + (− 0.295), and epithelial cells were categorized into high-risk and low-risk groups by median. Subsequently, the FindMarkers function was used

to calculate the DEGs 1 between the high and low risk groups (avg_log_2FC > 0.5, adj $p < 0.05$). Finally, GO and KEGG enrichment analyses were performed on the DEGs 1.

Expression evaluation of prognostic genes

To investigate prognostic gene expression, we collected 10 paired BRCA and NC samples from Shandong Provincial Third Hospital. This study had the approval of the Ethics Committee of the Third Hospital of Shandong Province. Informed consent forms were completed by all participants. Total RNA from 20 samples was isolated with TRIzol (Ambion, Austin, USA) according to the manufacturer's instructions. Total RNA was reverse transcribed to cDNA via the SureScript First strand cDNA synthesis kit (Servicebio, Wuhan, China) according to the manufacturer's instructions. Then, qPCR was performed using the 2xUniversal Blue SYBR Green qPCR Master Mix (Servicebio, Wuhan, China) according to the instructions in the manual. PCR primer sequences were shown in (Supplementary Table S1). Expression was in accordance with the internal reference GAPDH and calculated using the 2-ΔΔCT method[71].

Ethics declarations

The research study has been approved by the relevant ethics committee or institutional review board. The approval number and date of approval are as follows: [KYLL-2024067] and [2024-8-19].

Consent to participate/consent to publish

Informed consent forms were completed by all participants.

Statistical analysis

Data processing and analyses were performed using R software (version 4.2.1) and differences between groups were examined using the Wilcoxon test. $P < 0.05$ was defined as statistically significant.

Data availability

The datasets analysed during the current study are available in the [TCGA] repository, [https://portal.gdc.cancer.gov/projects/TCGA-BRCA] and from GEO (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE42568; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE176078) database.

Received: 1 July 2024; Accepted: 22 May 2025
Published online: 29 May 2025

References

1. Sung, H. et al. Global Cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
2. Franzoi, M. A., Romano, E. & Piccart, M. Immunotherapy for early breast cancer: too soon, too superficial, or just right? *Ann. Oncol.* **32**, 323–336 (2021).
3. Chung, W. P. et al. PI3K inhibitors in trastuzumab-resistant HER2-positive breast cancer cells with PI3K pathway alterations. *Am. J. Cancer Res.* **12**, 3067–3082 (2022).
4. Gobel, A. et al. High serum levels of leucine-rich alpha-2 glycoprotein 1 (LRG-1) are associated with poor survival in patients with early breast cancer. *Arch. Gynecol. Obstet.* **309**, 2789–2798 (2024).
5. Lu, Y. et al. Downregulation of BUBR1 regulates the proliferation and cell cycle of breast cancer cells and increases the sensitivity of cells to cisplatin. *Vitro Cell. Dev. Biol. Anim.* **59**, 778–789 (2023).
6. Van Wagoner, C. M. et al. Antibody-mediated phagocytosis in cancer immunotherapy. *Immunol. Rev.* **319**, 128–141 (2023).
7. Tsao, L. C., Force, J. & Hartman, Z. C. Mechanisms of therapeutic antitumor monoclonal antibodies. *Cancer Res.* **81**, 4641–4651 (2021).
8. Wang, S. S. et al. Tumor-infiltrating B cells: their role and application in anti-tumor immunity in lung cancer. *Cell. Mol. Immunol.* **16**, 6–18 (2019).
9. Yuen, G. J., Demissie, E. & Pillai, S. B lymphocytes and cancer: a love-hate relationship. *Trends Cancer.* **2**, 747–757 (2016).
10. Kuroda, H. et al. Prognostic value of tumor-infiltrating B lymphocytes and plasma cells in triple-negative breast cancer. *Breast Cancer.* **28**, 904–914 (2021).
11. Garaud, S. et al. Tumor infiltrating B-cells signal functional humoral immune responses in breast cancer. *JCI Insight.* **5** (2019).
12. Mohammed, Z. M. et al. The relationship between lymphocyte subsets and clinico-pathological determinants of survival in patients with primary operable invasive ductal breast cancer. *Br. J. Cancer.* **109**, 1676–1684 (2013).
13. Li, K., Li, J., Ye, M. & Jin, X. The role of Siah2 in tumorigenesis and cancer therapy. *Gene.* **809**, 146028 (2022).
14. Qi, J. et al. The E3 ubiquitin ligase Siah2 contributes to castration-resistant prostate cancer by regulation of androgen receptor transcriptional activity. *Cancer Cell.* **23**, 332–346 (2013).
15. Ahmed, A. U. et al. Effect of disrupting seven-in-absentia homolog 2 function on lung cancer cell growth. *J. Natl. Cancer Inst.* **100**, 1606–1629 (2008).
16. Malz, M. et al. Nuclear accumulation of seven in absentia homologue-2 supports motility and proliferation of liver cancer cells. *Int. J. Cancer.* **131**, 2016–2026 (2012).
17. Wu, A. et al. YTHDF1 enhances stemness and chemoresistance in triple-negative breast cancer cells by upregulating SIAH2. *Mol. Carcinog.* **63**, 417–429 (2024).
18. Ma, B. et al. The SIAH2-NRF1 axis spatially regulates tumor microenvironment remodeling for tumor progression. *Nat. Commun.* **10**, 1034 (2019).
19. Wong, C. S. et al. Vascular normalization by loss of Siah2 results in increased chemotherapeutic efficacy. *Cancer Res.* **72**, 1694–1704 (2012).
20. Eirew, P. et al. Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature* **518**, 422–426 (2015).
21. Ma, B. et al. Hypoxia regulates Hippo signalling through the SIAH2 ubiquitin E3 ligase. *Nat. Cell. Biol.* **17**, 95–103 (2015).
22. Adam, M. G. et al. SIAH ubiquitin ligases regulate breast cancer cell migration and invasion independent of the oxygen status. *Cell. Cycle.* **14**, 3734–3747 (2015).
23. Xu, J. et al. Delving into the heterogeneity of different breast Cancer subtypes and the prognostic models utilizing scRNA-Seq and bulk RNA-Seq. *Int J. Mol. Sci.* **23** (2022).

24. Jansen, M. P. et al. Downregulation of SIAH2, an ubiquitin E3 ligase, is associated with resistance to endocrine therapy in breast cancer. *Breast Cancer Res. Treat.* **116**, 263–271 (2009).
25. Anhao et al. YTHDF1 enhances stemness and chemoresistance in triple-negative breast cancer cells by upregulating SIAH2. *Mol. Carcinog.* **63**, 417–429 (2024).
26. Jie et al. Overexpression of seven in absentia homolog 2 protein in human breast cancer tissues is associated with the promotion of tumor cell malignant behavior in in vitro. *Oncol. Rep.* **36**, 1301–1312 (2016).
27. Rodrigo, B. I. et al. Seven in absentia homolog 2 (SIAH2) downregulation is associated with Tamoxifen resistance in MCF-7 breast cancer cells. *J. Surg. Res.* **190**, 203–209 (2014).
28. Turula, H. & Wobus, C. E. The role of the polymeric immunoglobulin receptor and secretory immunoglobulins during mucosal infection and immunity. *Viruses*. **10** (2018).
29. Yue, X. et al. Polymeric immunoglobulin receptor promotes tumor growth in hepatocellular carcinoma. *Hepatology*. **65**, 1948–1962 (2017).
30. Liu, F. et al. COLORECTAL polymeric Immunoglobulin receptor expression is correlated with hepatic metastasis and poor prognosis in Colon carcinoma patients with hepatic metastasis. *Hepatogastroenterology*. **61**, 652–659 (2014).
31. Ocak, S. et al. Loss of polymeric immunoglobulin receptor expression is associated with lung tumourigenesis. *Eur. Respir. J.* **39**, 1171–1180 (2012).
32. Berntsson, J. et al. Expression and prognostic significance of the polymeric Immunoglobulin receptor in epithelial ovarian cancer. *J. Ovarian Res.* **7**, 26 (2014).
33. Zhou, M. et al. Expression of polymeric Immunoglobulin receptor and its biological function in endometrial adenocarcinoma. *J. Cancer Res. Ther.* **15**, 420–425 (2019).
34. Asanprakit, W., Lobo, D. N., Eremin, O. & Bennett, A. J. M1 macrophages evoke an increase in polymeric Immunoglobulin receptor (PIGR) expression in MDA-MB468 breast cancer cells through secretion of interleukin-1beta. *Sci. Rep.* **12**, 16842 (2022).
35. Asanprakit, W., Lobo, D. N., Eremin, O. & Bennett, A. J. Expression of polymeric Immunoglobulin receptor (PIGR) and the effect of PIGR overexpression on breast cancer cells. *Sci. Rep.* **13**, 16606 (2023).
36. Yulong Bao. Transcriptome profiling revealed multiple genes and ECM-receptor interaction pathways that May be associated with breast cancer. *Cell. Mol. Biol. Lett*. **6**, 24:38 (2019).
37. Yiming Cao. et al. PANoptosis-related molecular clustering and prognostic signature associated with the immune landscape and therapy response in breast cancer. *Med. (Baltim)*. 103e39511 (2024).
38. Huan Song. Construction of a circRNA-related CeRNA prognostic regulatory network in breast cancer. *Onco Targets Ther.* **13**, 8347–8358 (2020).
39. Jiajia et al. DSMT-Net: dual self-supervised multi-operator transformation for multi-source endoscopic ultrasound diagnosis. *IEEE Trans. Med. Imaging*. **43**, 64–75 (2024).
40. Jacqueline, C. S. et al. A least absolute shrinkage and selection operator-derived predictive model for postoperative respiratory failure in a heterogeneous adult elective surgery patient population. *CHEST Crit. Care*. **1**,100025 (2023).
41. Ruyi et al. Prediction model for ocular metastasis of breast cancer:machine learning model development and interpretation study. *BMC Cancer*. **24**,1472 (2024).
42. Xu, J. et al. Delving into the heterogeneity of different breast Cancer subtypes and the prognostic models utilizing scRNA-Seq and bulk RNA-Seq. *Int. J. Mol. Sci.* **23**, 9936 (2022).
43. Andrews, J. M. et al. Loss of synergistic transcriptional feedback loops drives diverse B-cell cancers. *EBioMedicine* **71**, 103559 (2021).
44. Scortegagna, M. et al. Siah2 control of T-regulatory cells limits anti-tumor immunity. *Nat. Commun.* **11**, 99 (2020).
45. Huntington, N. D., Cursons, J. & Rautela, J. The cancer-natural killer cell immunity cycle. *Nat. Rev. Cancer*. **20**, 437–454 (2020).
46. Lee, B. R. et al. Combination of PD-L1 and PVR determines sensitivity to PD-1 Blockade. *JCI Insight*. **5** (2020).
47. Li, J., Huang, Q. & Mo, L. Comprehensive analysis of phagocytosis-related regulators to aid prognostic prediction and immunotherapy in patients with low-grade glioma. *Dis. Mark.* **2022**, 4142684 (2022).
48. Kamber, R. A. et al. Inter-cellular CRISPR screens reveal regulators of cancer cell phagocytosis. *Nature* **597**, 549–554 (2021).
49. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
50. Gustavsson, E. K. et al. Ggtranscript: an R package for the visualization and interpretation of transcript isoforms using ggplot2. *Bioinformatics* **38**, 3844–3846 (2022).
51. Gu, Z. & Hubschmann, D. Make interactive complex heatmaps in R. *Bioinformatics* **38**, 1460–1462 (2022).
52. Chen, H. & Boutros, P. C. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinform.* **12**, 35 (2011).
53. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
54. Lu, L. et al. Overexpression of OAS1 is correlated with poor prognosis in pancreatic Cancer. *Front. Oncol.* **12**, 944194 (2022).
55. Liu, T. T. et al. Identification of CDK2-related immune forecast model and CeRNA in lung adenocarcinoma, a pan-cancer analysis. *Front. Cell. Dev. Biol.* **9**, 682002 (2021).
56. Sachs, M. C. & plotROC A tool for plotting ROC curves. *J Stat. Softw.* **79** (2017).
57. Fang, J. et al. Prognostic value of immune checkpoint molecules in breast cancer. *Biosci. Rep.* **40** (2020).
58. Luo, Y. et al. Prognosis stratification in breast cancer and characterization of immunosuppressive microenvironment through a pyrimidine metabolism-related signature. *Front. Immunol.* **13**, 1056680 (2022).
59. Xinyue et al. Construction of a novel damage-associated molecular-pattern-related signature to assess lung adenocarcinoma's prognosis and immune landscape. *Biomolecules*. **14**, 108 (2024).
60. Zhuang, Y. & Li, X. Osteosarcoma biomarker analysis and drug targeting prediction based on pyroptosis-related genes. *Med. (Baltim)*. 104e40240 (2025).
61. Geeleher, P. et al. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS One*. **9**, e107468 (2014).
62. Anand, M. et al. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* **28**, 1747–1756 (2018).
63. Yuhan Hao. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).
64. Lemon, J. et al. Plotrix: a package in the red light district of R. *R-News*. **6**, 8–12 (2006).
65. Li, X. et al. A comprehensive analysis to reveal the underlying molecular mechanisms of natural killer cell in thyroid carcinoma based on single-cell RNA sequencing data. *Discov. Oncol.* **16**, 44 (2025).
66. Hanjie, L. et al. Exploring the molecular characterization of PANoptosis-related genes with features of immune dysregulation in Alzheimer's disease based on bulk and single-nuclei RNA sequencing. *Metab. Brain Dis.* **40**, 09 (2025).
67. Johannes, G. et al. ReactomeGSA—efficient multi-omics comparative pathway analysis. *Mol. Cell. Proteom.* **19**, 2115–2125 (2020).
68. Cong Xiao. CYFIP2: potential pancreatic cancer biomarker and immunotherapeutic target. *Discov. Oncol.* **15**, 847 (2024).
69. Lei et al. From single-cell to Spatial transcriptomics: decoding the glioma stem cell niche and its clinical implications. *Front. Immunol.* **15**, 1475235 (2024).
70. Cole et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).

71. Shang, Y. et al. Decreased E2F2 expression correlates with poor prognosis and immune infiltrates in patients with colorectal cancer. *J. Cancer*. **13**, 653–668 (2022).

Acknowledgements
We would like to express our sincere gratitude to all individuals and organizations who supported and assisted us throughout this research. Without your support, this research would not have been possible.

Author contributions
X.J.: Conceptualization, data curation, validation, visualization, writing-original draft, writing-review and editing. J.Z.: Data curation, validation, visualization, writing-review and editing. X.L.: Conceptualization, project administration, supervision, writing-review and editing. This work currently described has not been published, is not being considered for publication elsewhere, and its publication was approved by all authors. All authors agree to be accountable for the content of the work.

Funding
This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-03825-8.

Correspondence and requests for materials should be addressed to X.L.