BMC Cancer

RESEARCH

Open Access

Identification and validation of molecular subtypes and prognostic models in patients with kidney cancer based on differential genes based on B cells: a multiomics analysis

Jiaao Sun[1†], Shiyan Song[1†], Qiancheng Ma[1], Feng Chen[1*], Xiaochi Chen[1*] and Guangzhen Wu[1*]

Abstract

Background B cells play a variety of complex roles in cancer, both promoting cancer progression and enhancing anti-tumor immune responses, but their mechanism of action in kidney cancer has not been elucidated.

Results We collected kidney cancer sample data from the GEO database and TCGA database, mapped the single-cell landscape inside kidney cancer tissue, identified 25 B-cell-related genes, and based on this, identified related molecular subtypes of kidney cancer patients, and explored their internal microenvironment characteristics. Finally, we constructed a 6-gene biological prognostic model that can be used to predict survival in patients with renal cancer, and we further validated the predictive performance of the model based on imaging omics. It is worth mentioning that the structural patterns and functional sites of 6 model gene transcription proteins were also mined.

Conclusions Overall, we explored for the first time the profound role of B cells in kidney cancer and developed a bio-predictive model based on B cell-related genes, providing scientific guidance for personalized treatment of kidney cancer patients.

Keywords Renal carcinoma, B cell, Immune microenvironment, Multiomics, Prognostic model

Background

Renal cell carcinoma (RCC) is a highly aggressive malignant tumor originating from tubular epithelial cells of the kidney, characterized by intrinsic resistance to cytotoxic chemotherapy, and is one of the three major malignant tumors of the urinary system. Renal clear cell carcinoma (ccRCC) is the most common pathological tissue subtype, accounting for about 70% of RCC [1, 2]. At present, in addition to surgery and radiotherapy, targeted therapy drugs such as tyrosine kinase inhibitors (TKI) and mTOR have become first-line drugs [2]. With the rise of immunotherapy, the application of monotherapy of immune checkpoint inhibitors (ICI) and the combination of immunotherapy and targeted therapy can also improve the survival of RCC patients to a certain extent [3, 4]. However, although the treatment of RCC has made great progress in recent years, due to reasons such as delayed diagnosis, only some patients benefit from immunotherapy, and the 5-year survival rate of patients with metastatic RCC is still low [5]. Therefore, it is necessary to develop effective biomarkers and novel prediction

† Jiaao Sun, Shiyan Song and Qiancheng Ma contributed to this study equally.

*Correspondence:
Feng Chen
dmuchenfeng@163.com
Xiaochi Chen
chenxiaochi1216@163.com
Guangzhen Wu
wuguang0613@hotmail.com
[1] First Affiliated Hospital of Dalian Medical University, Dalian, China

models for early diagnosis, effective treatment, and prognosis prediction of RCC patients.

As an important subgroup of antigen-presenting cells, B cells play an important role in anti-tumor immunity. Specifically, the B cell receptor (BCR) has a high affinity with antigens, which can effectively bind and present antigens, and then display antigens on the cell surface through MHC Class II molecules to activate homologous CD4+T cells. This induces the production of antibodies [6–8]. As an important part of the tumor microenvironment, B cells participate in the formation of tertiary lymphoid structures and play a role in promoting tumor immunity. Regulatory B cells also affect cancer progression by producing cytokines such as IL-10, IL-35, and TGF-β [7, 9]. There has been evidence that B-cell infiltration is associated with positive prognosis in patients with various cancers (such as non-small cell lung cancer, breast cancer, and colorectal cancer, etc.) [10–12], and active B-cell immune infiltration has also been detected in RCC [13], but the detailed mechanism of B-cell action in RCC has not been clarified. Given the important role of B cells in cancer immune infiltration, we speculate that B cells may be the key to distinguishing between different RCC molecular subtypes and become an important factor affecting patient prognosis.

Compared to conventional bulk RNA sequencing, single-cell sequencing techniques reveal the cellular heterogeneity of the tumor microenvironment. Starting with single-cell transcriptomics, our study analyzed the scRNA-seq data of 7 RCC samples and 6 normal kidney samples and revealed the important role of B cells in the RCC tumor microenvironment through quasi-time series analysis and intercellular communication. After that, we screened the relevant differential genes based on B cells, identified and explored the molecular subtypes of RCC, and finally constructed a 6-gene biological prediction model, and verified its prediction performance by means of imaging omics. Our research process r is shown in Fig. 1 (Fig. 1).

Results

Establish a single-cell landscape of RCC and normal renal tissue

This study involved 7 RCC patient samples and 6 normal tissue samples based on the GSE159115 dataset. The criteria for quality control are described in the Materials and methods. After filtration, 24,793 cells were obtained, including 14,274 cells from RCC and 10,519 cells from normal tissues, and 23,541 genes were obtained. T-SNE and UMAP were used to visually display the dimensionality reduction results between RCC and normal tissues (Fig. 2C, D). We identified the first 2,000 highly variable genes between cells for subsequent analysis, and Harmony analysis was used to remove batch effects from the original single-cell data. Through Elbow Plot and JackStraw analysis (Fig. 2A, B), combined with the heat map of the first 20 principal components (Supplementary Fig. 1), we selected the top 15 PCS with the most explanatory power for subsequent cluster analysis. After that, we use the "clustree" package to visualize the division relationship between cell populations at different resolutions. As resolution is turned up, we can clearly see which cell populations continue to fission into multiple subpopulations. When resolution was 1.5, we observed that the cell clusters were relatively stable, and 32 cell clusters were obtained (Supplementary Fig. 2). The RCC samples and normal samples were visualized by tSNE and UMAP methods. The clustering of different samples in each cell group was shown (Fig. 2E, F).

Based on the 11 cell types that may appear in RCC, we selected their specific marker genes. Epithelial cell (KRT8, KRT18, CD24, EPCAM), Myeloid cell (CD68, CD14, CSF-1R), Fibroblast cell (TAGLN, RGS5, MYL9, ACTA2), Myeloid cell (CD68, CD14, CSF-1R), T cell (CD2, CD3D, CD3E, CD3G), Endothelial cell (ENG, PECAM1, VWF, CDH5), NK cell (NKG7, GNLY, KLRD1, CCL5), B cell (MS4A1, CD79A, CD79B), Macrophage (APOE, C1QA, C1QB, CD14), DC(CLEC10A, CD1C, CD1E, FCER1A), Monocyte (APOBEC3A FCN1, S100A8, S100A9), Erythrocyte (HBG1, HBD, HBM), combines the condition of each specific marker gene expression in Dotplot (Supplementary Fig. 3). We identified all cells from the 13 samples as six cell subtypes: Epithelial cell (cluster0, 1,2,9,12,13,19,20,21,22,23,26), Fibroblast cell (8,11,24,31 cluster), T/NK cell(cluster 7,18,25), Endothelial cell(cluster 4,10,14,15,16,27,29), B cell(cluster 17), Myeloid cell(cluster 3,5,6,28,30). tSNE and UMAP were used to annotate the samples for cell populations, visualizing the distribution and clustering of the 6 types of cells in the sample (Fig. 2G, H).

The difference in cell subsets between RCC and normal renal tissues was visualized by tSNE and UMAP (Fig. 3A, B). The proportion of myeloid cells and T/NK cells in RCC was lower than that in normal tissues. The proportion of Epithelial cells, Endothelial cells, B cells, and Fibroblast cells is relatively high. The bar chart visually shows the composition ratio of the number of 6 types of cells in RCC and para-cancerous tissues (Fig. 3C). The bubble diagram shows the marker genes used to identify cell types (Fig. 3D). Then all the genes of the B cells as a whole are compared with all other cells to screen out the differential genes of the B cells (Supplementary Table 1). We selected genes with log2FC > 1 as the b-cell-related gene set.

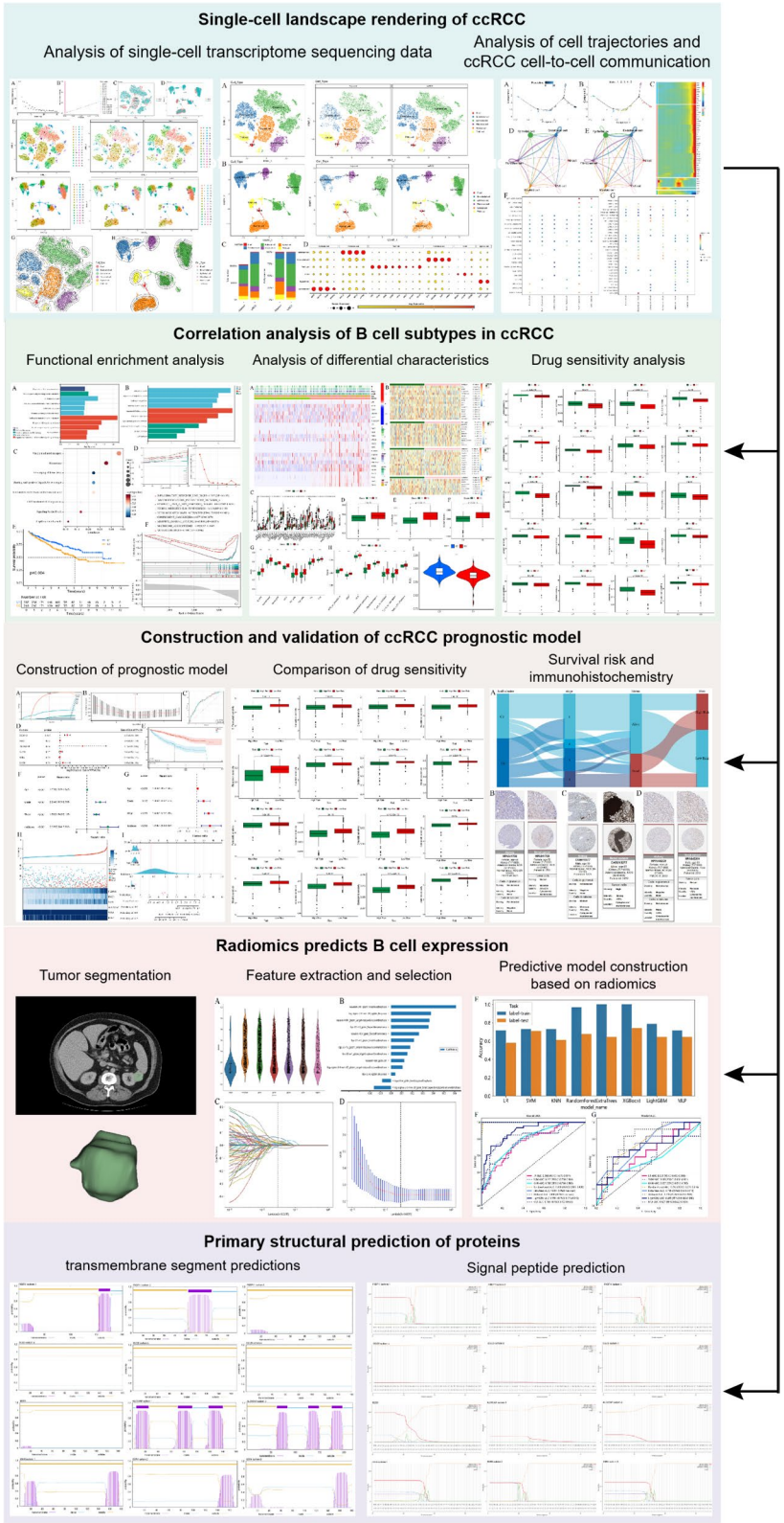

Fig. 1 Study flow chart

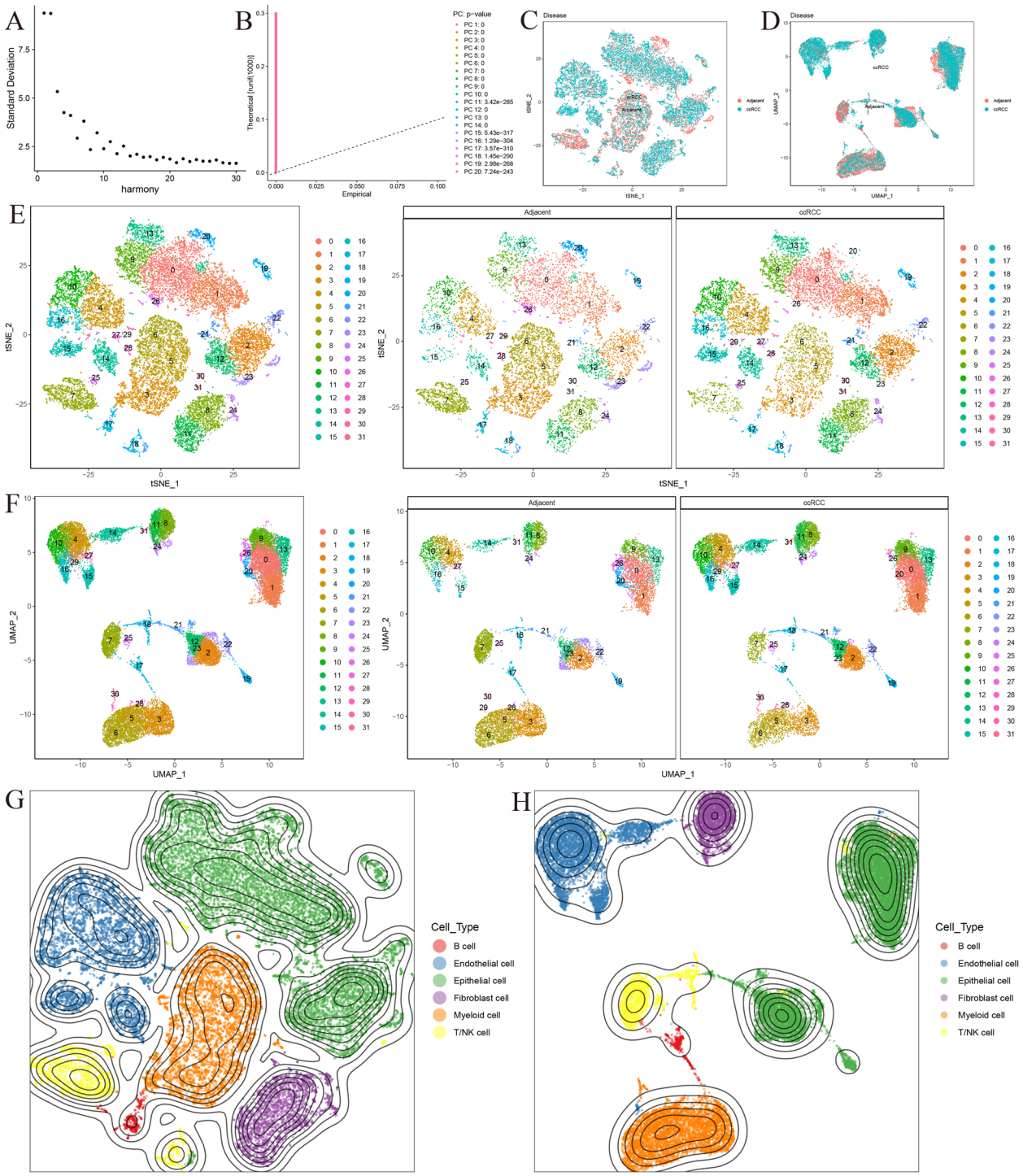

Fig. 2 **A** Elbow plot showing optimal cluster number. **B** JackStraw analysis was used to select significant principal components. **C** 24,793 high-quality cell t-SNE maps after QC were used to reflect the dimensionality reduction between RCC and normal tissue samples. **D** 24,793 high-quality cell UMAP after QC were used to reflect the dimensionality reduction between RCC and normal tissue samples. **E** t-SNE color-coded visualization of RCC with cluster cells of normal tissue samples with a total of 32 cell clusters at a resolution of 1.5. **F** UMAP color-coded visualization of RCC and normal tissue samples of clustered cells with a total of 32 cell clusters at a resolution of 1.5. **G** Six major cell types identified by the t-SNE map. **H** Six main cell types identified by UMAP

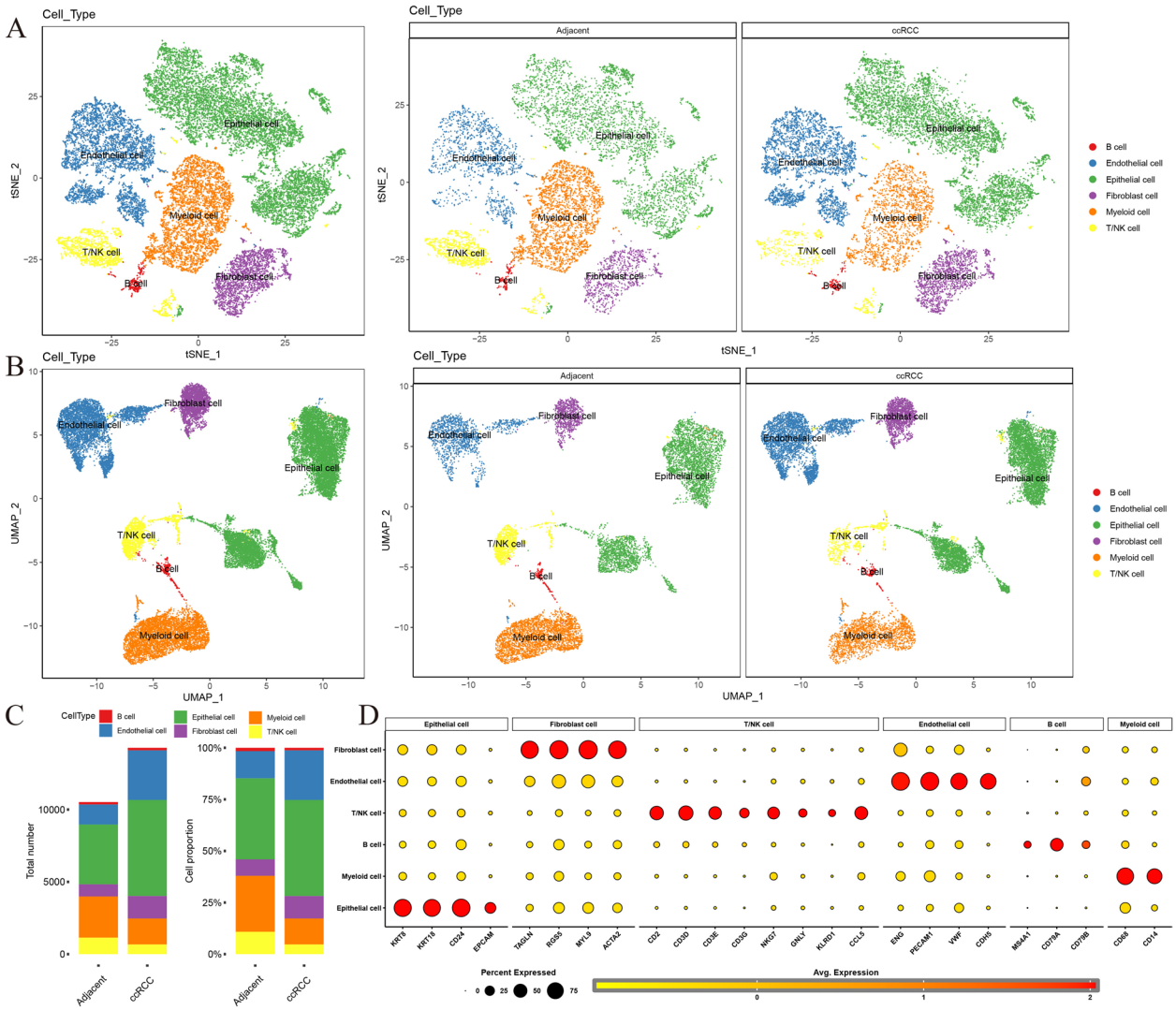

Fig. 3 **A** t-SNE illustrates the results of annotating cell subpopulations in RCC versus normal tissue. **B** UMAP illustrates the results of annotating cell subpopulations in RCC versus normal tissues. **C** The bar chart visually shows the ratio of the number of 6 types of cells in RCC and para-cancerous tissue. **D** Marker genes for the six cell types identified

B cell occupies an important position in the TME of RCC

Cell trajectory analysis is used to reconstruct the dynamic changes of cell state and reveal the cell differentiation process and evolution law from the single cell level. Specifically, according to the sequential gene expression of each cell, it is arranged on the corresponding trajectory according to time to generate an intuitive lineage development tree, and then predict the cell differentiation and development trajectory. By using Monocle, we delineate the differentiation trajectory of B cells in ccRCC. Although the proportion of B cells in RCC is low, it still shows a clear development track in the tree diagram, reflecting the dynamic change process of B cells in RCC (Fig. 4A, B). We also drew heat maps to show the expression of related genes in B cells at different developmental stages (Fig. 4C). As can be seen from the figure, AREG, CD63, NFKBIA, and CD83 genes are highly expressed in the late stage of B cell evolution, IGLL5, IGHG4, IGHG1, and IGHM are highly expressed in the middle stage of evolution, and the expressions of KLRB1, IGLC2, and IGKC gradually decrease with cell changes. To identify the signal transmission pathway and interaction between different cell types in RCC, we used CellChat to calculate the interaction network of 6 cell components in RCC. In the process of signal transmission, the number and weight of communication between B cell and Myeloid cell are more significant than those of other cells. In the process of signal introduction, the

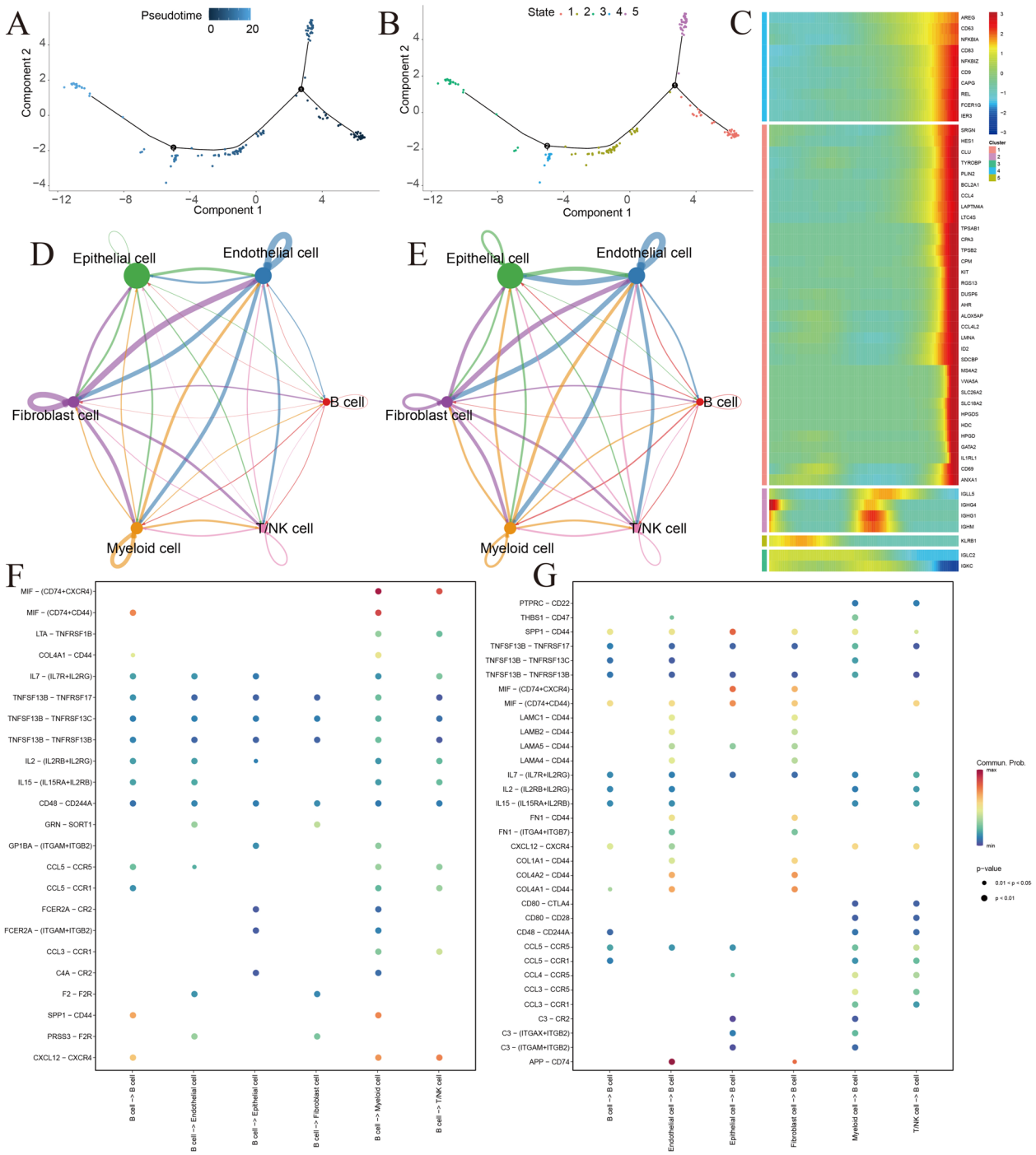

Fig. 4 **A** Trajectory analysis results of the B cell. Different shades of blue represent different times. **B** Five developmental states divided the developmental trajectories of B cells. **C** Heat maps of differentially expressed genes in each branch of RCC with pseudo time and hierarchical clustering. **D** The number of cell communications between the six-cell components in the RCC tumor microenvironment. **E** The strength of cell communication between the six-cell components in the RCC tumor microenvironment. **F** The bubble map shows the intensity with which B cells regulate other cell types through ligand-receptor interactions. **G** The bubble map demonstrates the intensity of regulation of B cells by other cell types through ligand-receptor interactions

B cell is highly regulated by the T/NK cell and Epithelial cell (Fig. 4D, E). We drew the bubble map to show the ligand-receptor pairs with strong interactions between various cell types and B cells (Fig. 4F, G).

Functional enrichment analysis of differential gene sets in B cells and identification of RCC molecular subtypes

After screening out the significant differential genes of B cells, to determine their functional enrichment status in vivo, We carried out functional enrichment analysis of KEGG (Fig. 5A), GO (Fig. 5B), and Reactome (Fig. 5C) successively, and showed the related pathways with high enrichment. We found that in addition to Leukocyte transendothelial migration, Fc epsilon RI signaling pathway, IgA immunoglobulin complex, and other immune-related functions, B cell differential genes are also associated with the activity of cancer and intercellular signaling pathways, including colorectal cancer and vesicle transport. Then, the ConsensusClusterPlus software package was used for cluster analysis, and the optimal

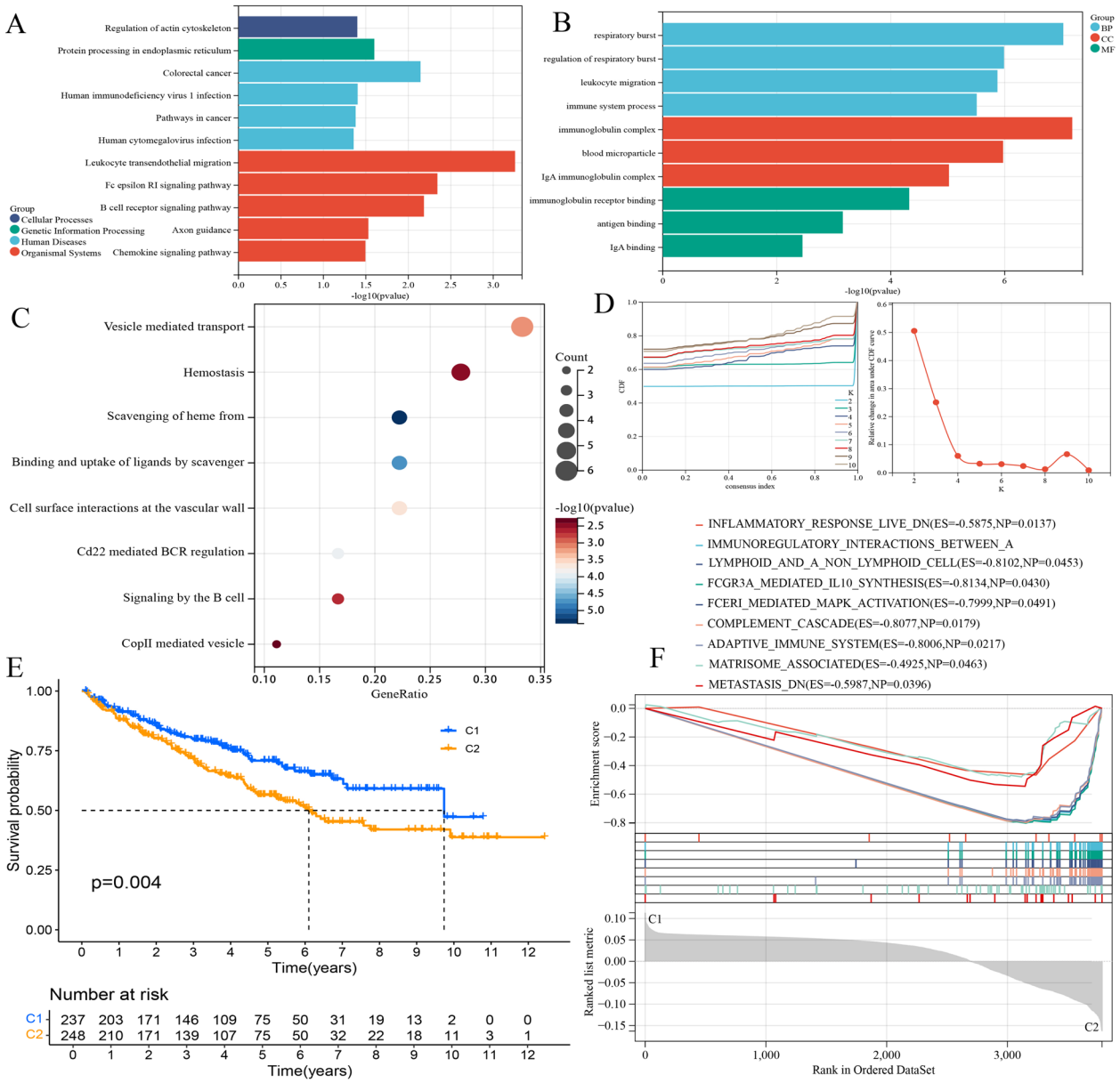

Fig. 5 **A** KEGG enrichment analysis of B cell related gene sets. **B** GO enrichment analysis of B cell related gene set. **C** Reactome enrichment analysis of B cell related gene set. **D** Cluster analysis results show that the optimal cluster number k = 2 is determined by the cumulative experience distribution function graph. **E** Comparison of survival of C1 and C2 subtypes. **F** Pathway enrichment of C1 and C2 subtypes was analyzed by GSEA

cluster number k = 2 was determined by the cumulative experience distribution function graph (Fig. 5D). Therefore, we divided 485 RCC patients into C1 (N = 237) and C2 (N = 248) subtypes based on the expression levels of 25 B cell differential genes, in which C1 had a lower expression level of B cell differential genes and C2 had a higher expression level of B cell differential genes. The heat map details the differences in expression levels of 25 B cell differential genes between the two subtypes and their relationship to clinical data (Fig. 6A). Using the R software packages "survival" and "survminer", we plotted the survival difference between the two subtypes (Fig. 5E). As shown in the figure, the C1 group had a better survival state, suggesting that a lower level of B cell expression predicted a better prognosis for RCC patients. Then, to further analyze the functional enrichment of the two subtypes, we performed a GSEA analysis to show statistically significant and consistent differences in nine mechanisms or biological functions between the C1 and C2 groups. Including inflammatory response,

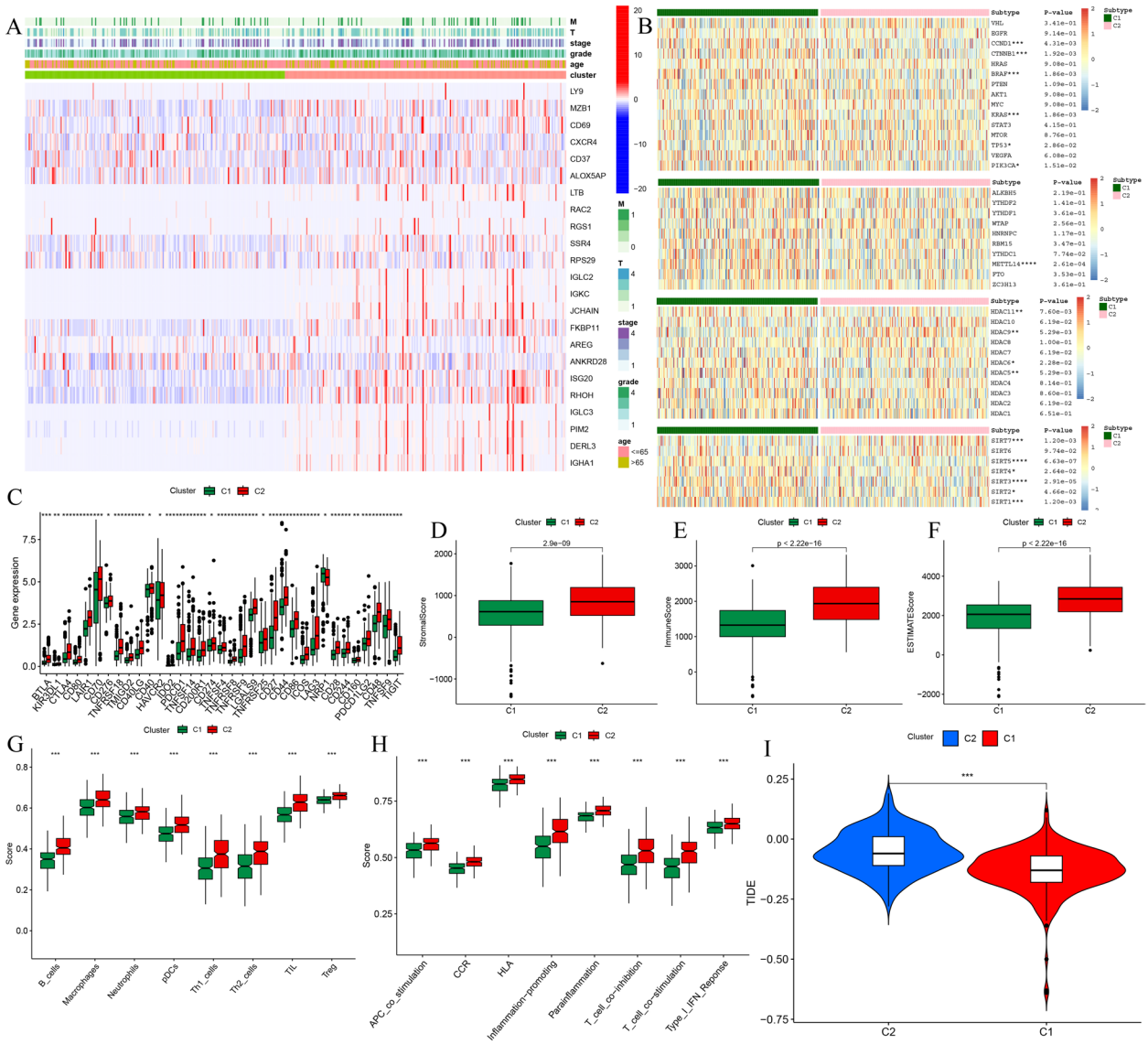

Fig. 6 **A** Heat maps showing clinical data for C1 and C2 subtypes and B-cell-associated gene expression levels. **B** Expression differences of deacetylation-related genes (HDAC and SITR), RNA methylation-related genes (ALKBH5, WTAP, FTO, etc.), and proto-oncogenes (VHL, TP53, MTOR, etc.) between C1 and C2 subtypes. **C** Expression comparison of immune checkpoint-related genes of C1 and C2 subtypes. The ESTIMATE analysis showed stromal scores (**D**), immune scores (**E**), and tumor purity scores (**F**) for C1 and C2 subtypes. **G** Comparison of the degree of immune cell infiltration between C1 and C2 subtypes. **H** Comparison of immune activity between C1 and C2 subtypes. **I** Comparison of immune escape levels between C1 and C2 subtypes

immunoregulatory interactions, FCGR3A mediated IL-10 synthesis, FCERI mediated MAPK activation, complement cascade, adaptive immune system, matrisome associated, and metastasis (Fig. 5F).

Microenvironment analysis of RCC molecular subtypes

Studies have shown that histone deacetylation modifications, abnormal methylation of tumor-associated gene promoter regions, and mutations in proto-oncogenes are associated with the development of RCC. To uncover such differences between C1 and C2 subtypes, Heat maps were drawn to clearly show the expression differences of deacetylation-related genes (HDAC and SITR), RNA methylation-related genes (ALKBH5, WTAP, FTO, etc.) and proto-oncogenes (VHL, TP53, MTOR, etc.) among subtypes (Fig. 6B), as shown in the figure. The C1 subtype had higher expression of CTNNB1, BRAF, METTL14, and SIRT3 genes, while the C2 subtype had higher expression of TP53, HDAC9, and SIRT7 genes.

The immune microenvironment of cancer is a complex system that plays an important role in cancer progression and patient response to immunotherapies such as immune checkpoint-blocking therapy. A series of additional analyses were performed to reveal the differences in the immune microenvironment between C1 and C2 subtypes. First, we mapped the differences in the expression levels of immune checkpoint-related genes between the two subtypes. As shown in the Figure, the C2 subtype generally had a high expression of immune checkpoint-related genes (Fig. 6C). The C2 subtype also has an active immune function (macrophages, neutrophils, Treg, pDCs, etc.) (Fig. 6G) and immune cell infiltration (inflammation−promoting, Parainflammation and Type I) IFN response et al.) (Fig. 6H). In addition, the ESTIMATE algorithm used the pre-screened matrix/immune-related gene set to predict the level of matrix/immune cell infiltration in tumor tissues using gene expression data. Based on the ESTIMATE algorithm, we inferred the differences in tumor purity and tumor immune status between C1 and C2 subtypes (Fig. 6D-F). The C2 subtype has a higher stromal score and tumor purity, which also explains the poor prognosis of patients with the C2 subtype shown in previous results, while the C2 subtype has a higher expression of B-cell-associated genes and therefore a higher immune infiltration score.

Significance of RCC molecular subtypes in drug therapy

There are two different immune escape mechanisms in cancer: on the one hand, some immunosuppressive factors can prevent the invasion of immune cells; On the other hand, some immune cells are in a state of functional inhibition. The TIDE online tool can predict a tumor's immune escape ability by comprehensively evaluating the activity of these two mechanisms, and a higher TIDE score is associated with higher immune resistance and poorer immune checkpoint suppression efficacy. We evaluated TIDE scores for the C1 and C2 subtypes, as shown in the Figure. The C2 subtype had a higher TIDE score, suggesting a higher immune escape capacity and a poorer immunotherapy response (Fig. 6I).

Then, to further explore the potential value of B-cell-related gene typing in clinical drug therapy for RCC patients, we predicted the response of 485 RCC patients to multiple targeted drug therapy based on the "oncoPredict" software package, analyzed the potential beneficial drugs for different subtypes of RCC patients, and drew a box diagram. The effects of B cell-related genes on the sensitivity of 20 common RCC-targeted drugs were determined. Specifically, patients with subtype C1 showed increased response to Cisplatin, Gefitinib, Lapatinib, Nilotinib, Vorinostat, and other 20 RCC potent drugs showed high sensitivity (Fig. 7). These results will provide scientific guidance for the development of individualized clinical treatment for RCC patients.

Construction of RCC prognostic model

Based on the important role of B-cell-related genes in the RCC microenvironment, we further screened prognostic-related genes for the construction of prognostic models. Specifically, we used the R software package "glmnet" to integrate survival time, survival state, and gene expression data, and utilized the lasso-cox method for regression analysis to obtain the optimal model. Finally, a genetic model composed of six genes FKBP11, ISG20, MZB1, ALOX5AP, SSR4 and IGKC was obtained (Fig. 8A, B, Supplementary Table 1). At the same time, we plotted the forest map to show the Hazard Ratio of each model gene, where SSR4 and IGKC are protective genes of RCC, while FKBP11, ISG20, MZB1 and ALOX5AP are risk genes of RCC (Fig. 8D). From this, we obtained the formula for calculating the risk score of RCC patients (RiskScore = 0.105985123365355*FKBP11 + 0.0223755912241455*ISG20 + 0.0118787615362657*MZB1 + 0.337174027637544*ALOX5AP-0.00261656598176431*SSR4-0.000107083273338916*IGKC). After that, the R software package "maxstat" was used to calculate the optimal cutoff value of RiskScore, and the patients were divided into two groups of high and low risk. The survfit function of the R software package "survival" was further used to analyze the difference in survival state between the two groups (Fig. 8E). It can be seen from the figure that high-risk patients have a poor prognosis.

Validation and follow-up analysis of the RCC prognostic model

In order to evaluate the accuracy of the prognosis model constructed by us, ROC curves were drawn for RCC

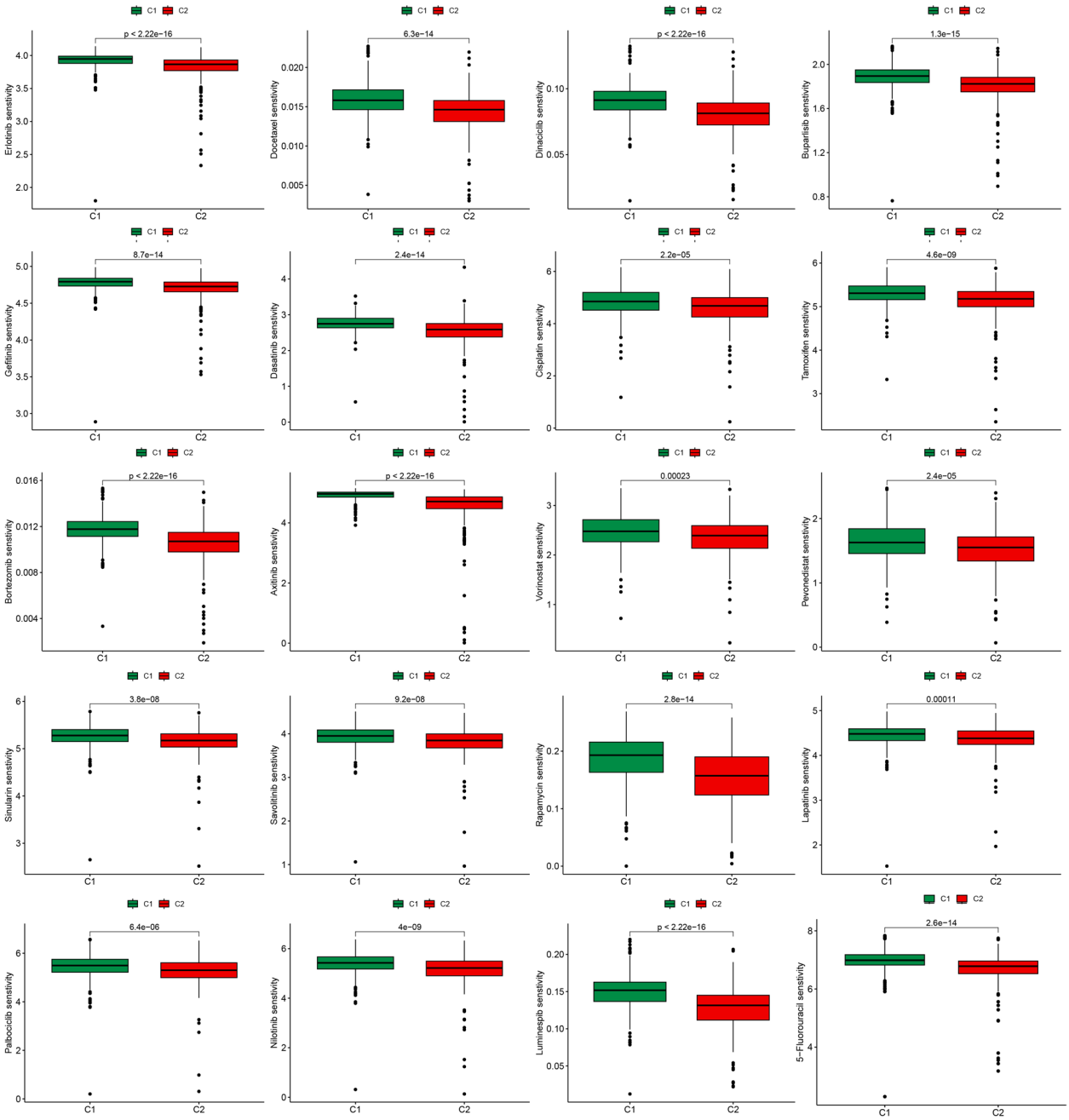

Fig. 7 Comparison of the sensitivity of C1 and C2 subtypes to 20 RCC-targeting drugs

patients, and the UC (area under ROC curve) of 1, 3, and 5 years were 0.71, 0.70, and 0.71 respectively, indicating high accuracy of the model (Fig. 8C). In addition, univariate prognostic analysis showed that RiskScore was correlated with the prognosis of RCC patients ($p < 0.001$) (Fig. 8F). Multivariate independent prognostic analysis showed that RiskScore could be used as a prognostic predictor of RCC patients independently of age, stage, and grade ($p < 0.001$) (Fig. 8G). Risk and expression heat maps were also drawn. As shown in the Figure, with the increase of RiskScore, the death rate of patients increased, and the expression level of model genes also changed (Fig. 8H). Finally, we used the R software package "rms" to integrate the RiskScore data and three clinical features to establish a nomogram by the Cox method. On the one hand, we evaluated the prognostic

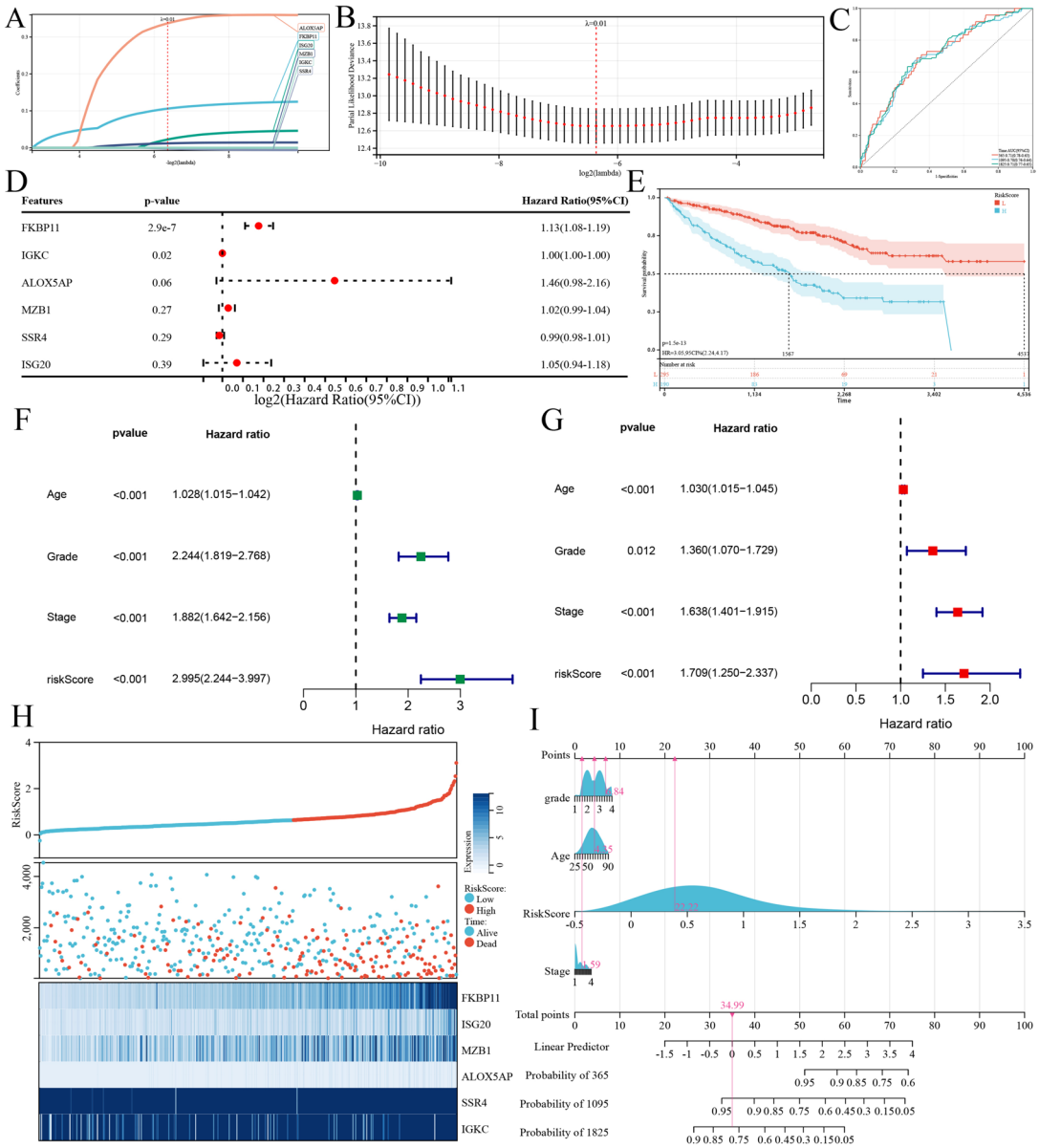

Fig. 8 **A**, **B** Regression analysis was performed by the lasso-cox method to obtain the optimal model, and the gene model composed of six genes (FKBP11, ISG20, MZB1, ALOX5AP, SSR4, and IGKC) was obtained. **C** ROC curves for RCC patients at 1, 3, and 5 years. **D** Forest maps showing the statistical significance and Hazard Ratio of each model gene. **E** Comparison of survival in high—and low-risk groups. **F** Univariate independent prognostic analysis. **G** Multivariate independent prognostic analysis. **H** Heat maps showed changes in survival and model gene expression levels of RCC patients as Riskscore increased. **I** We draw a nomogram to accurately predict the survival probability of RCC patients for 1, 3, and 5 years

significance of these features and, on the other hand, based on the relevant information, Survival rates at 1, 3, and 5 years can be effectively predicted (Fig. 8I). For example, as we show in the figure, based on the total points calculated by the patient's age, stage, grade, and RiskScore, the probability of 1, 3, and 5 years of survival will be calculated accordingly.

Then, based on the oncoPredict software package, we predicted the response of RCC patients in the high-low-risk group to multiple targeted drug therapies and analyzed the potential beneficial drugs for RCC patients at different risks. Specifically, Patients in the low-risk group showed higher sensitivity to 16 RCC-potent drugs, including Axitinib, Bortezomib, Gefitinib, Savolitinib, and Tamoxifen (Fig. 9). Mulberry map clearly shows the aggregation relationship of B-cell molecular subtypes, stage, survival status, and risk subtypes in RCC patients (Fig. 10A). It can be seen

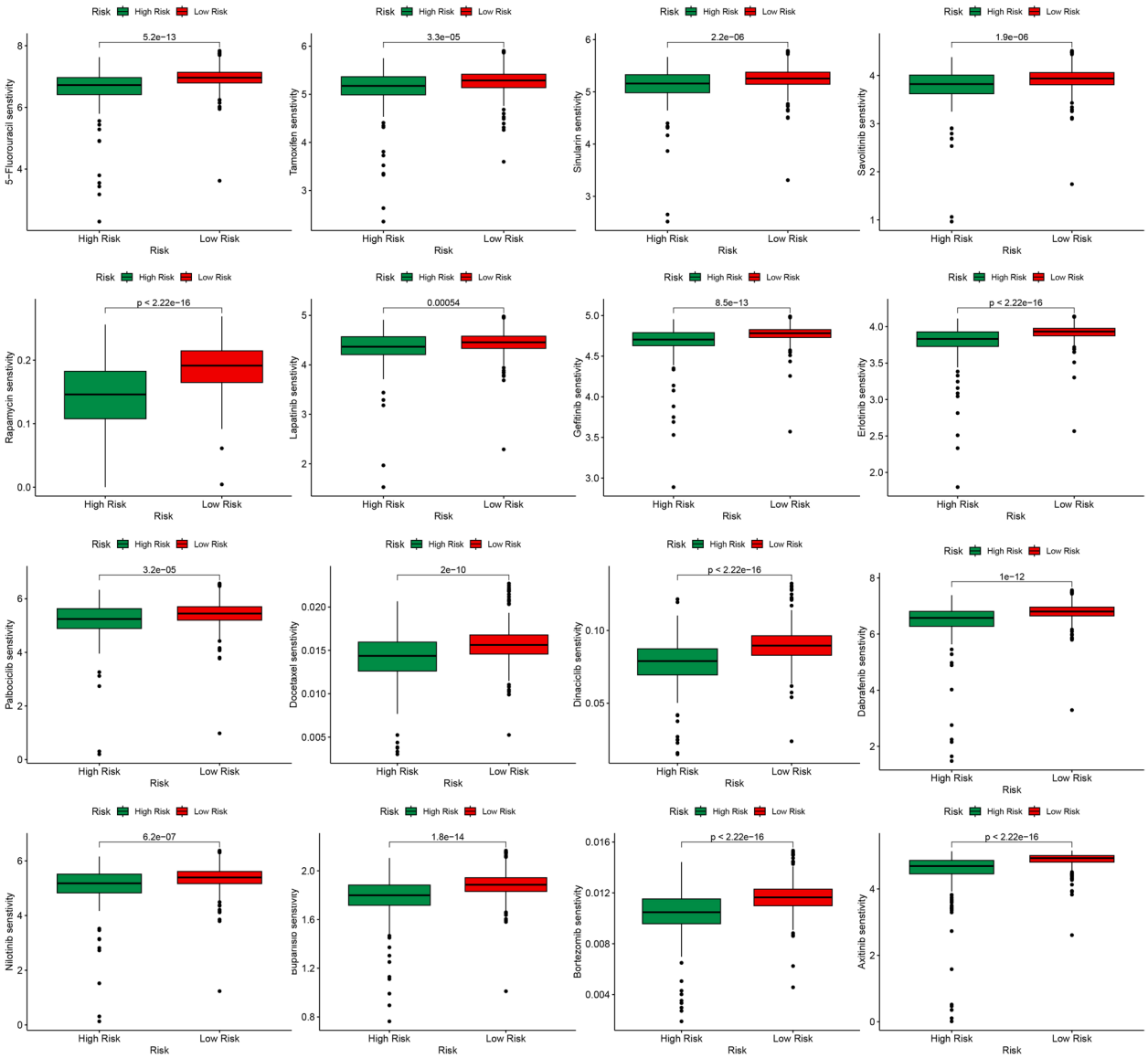

Fig. 9 Comparison of the sensitivity of patients in high and low-risk groups to 16 RCC-targeting drugs

that the C2 subtype corresponds to late stage and poor prognosis, which also corresponds to high-risk types. Finally, based on the HPA website, we downloaded the immunohistochemical comparison results of FKBP11 (Fig. 10B), IGKC (Figure), and SSR4 (Fig. 10D) in RCC and adjacent tissues, and the expression of the three genes in RCC tissues was higher than that in normal tissues.

Further validation of the prognostic model based on imaging omics

We built a radiomic model to predict the level of B-cell-related gene expression in RCC. After 1830 features were extracted using the pyradiomicv3.0 program in Python 3.8 (FIG. 10A), the Pearson correlation coefficient was used for feature screening ($p < 0.05$), and a total of 302 features were preliminarily screened. In addition, the correlation coefficient matrix of 302 features was drawn (Supplementary Fig. 4), and the features were filtered according to the correlation coefficient. One of the two features with a correlation coefficient > 0.9 was selected, totaling 87 features. We then used LASSO to select 13 high-weighted features (Fig. 11B, C, D). We used eight machine-learning algorithms to create a radiomic model. We drew the confusion matrix (Supplementary Fig. 5) and ROC curve (Fig. 11F, G) to evaluate

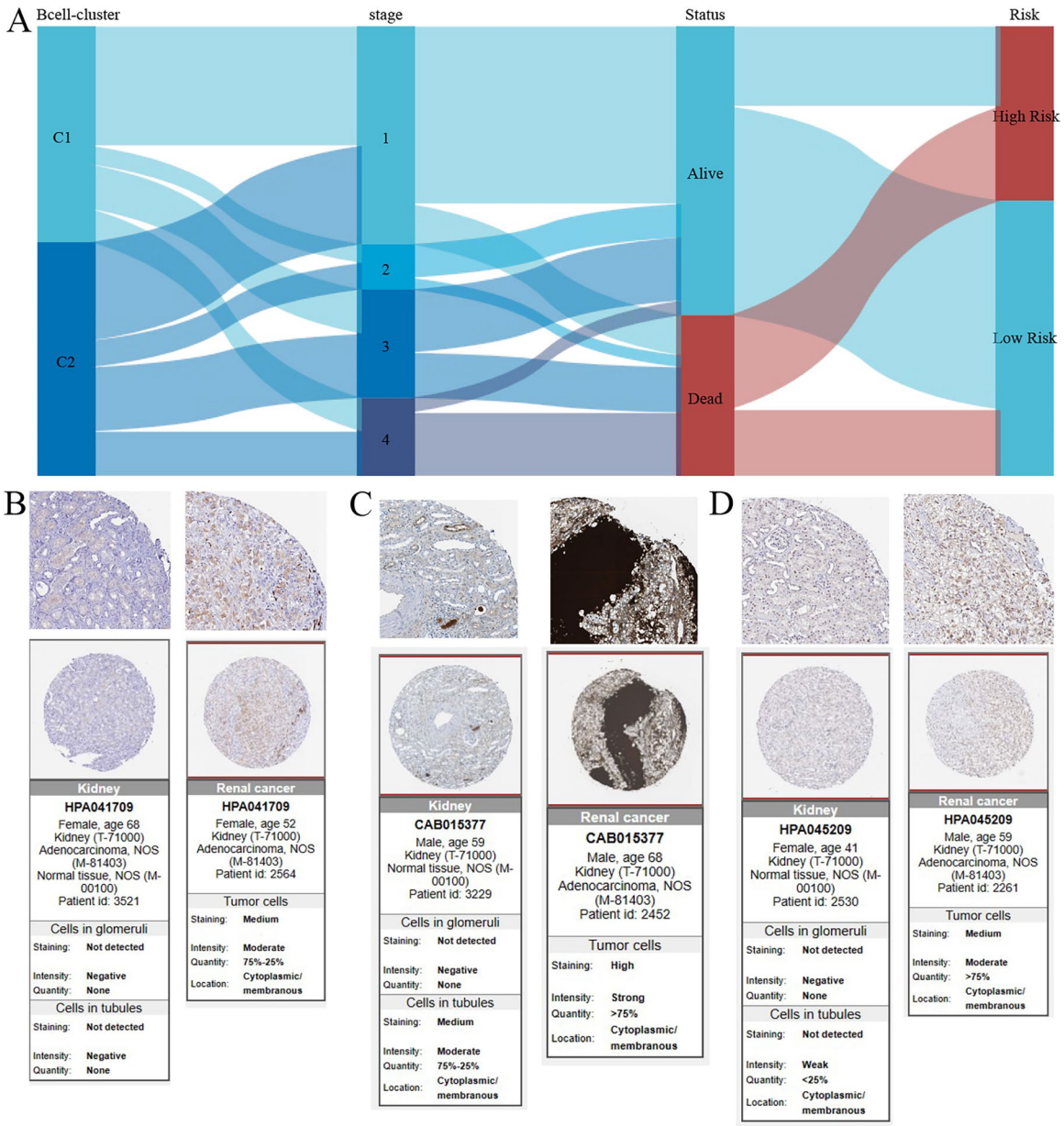

Fig. 10 **A** Mulberry diagram shows the polymerization relationship of B cell molecular subtypes, stage, survival state, and risk subtypes. Immunohistochemical results of FKBP11 (**B**), IGKC (**C**), and SSR4 (**D**) in RCC and paracancer tissues

the diagnostic performance of the predictive model. AUC, diagnostic accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were analyzed. The results are shown in Supplementary Table 3, FIG. 10E. In the test set, the accuracy of XGBoost and SVM was higher, 0.742 and 0.710 respectively (Supplementary Table 3, FIG. 10E). The sensitivity of the XGBoost model is the highest (0.667). SVM, RandomForest, LightGBM, and MLP models produced the highest specificity of 0.895, 0.895, 0.842, and 0.842, respectively. The SVM model produced the highest PPV (0.714) and XGBoost produced the highest NPV (0.789). On this basis, we pay attention to the Recall of the predicted results of the eight models, and it can be seen that XGBoost produces the highest Recall, which is 0.667 (Supplementary Table 3). As shown in FIG. 10F, G, among the 8 kinds of machine learning classifiers, XGBoost, and ExtraTrees produce the highest training set ROC AUC, which is 1.000; XGBoost, RandomForest, and ExtraTrees produce the highest training set ROC

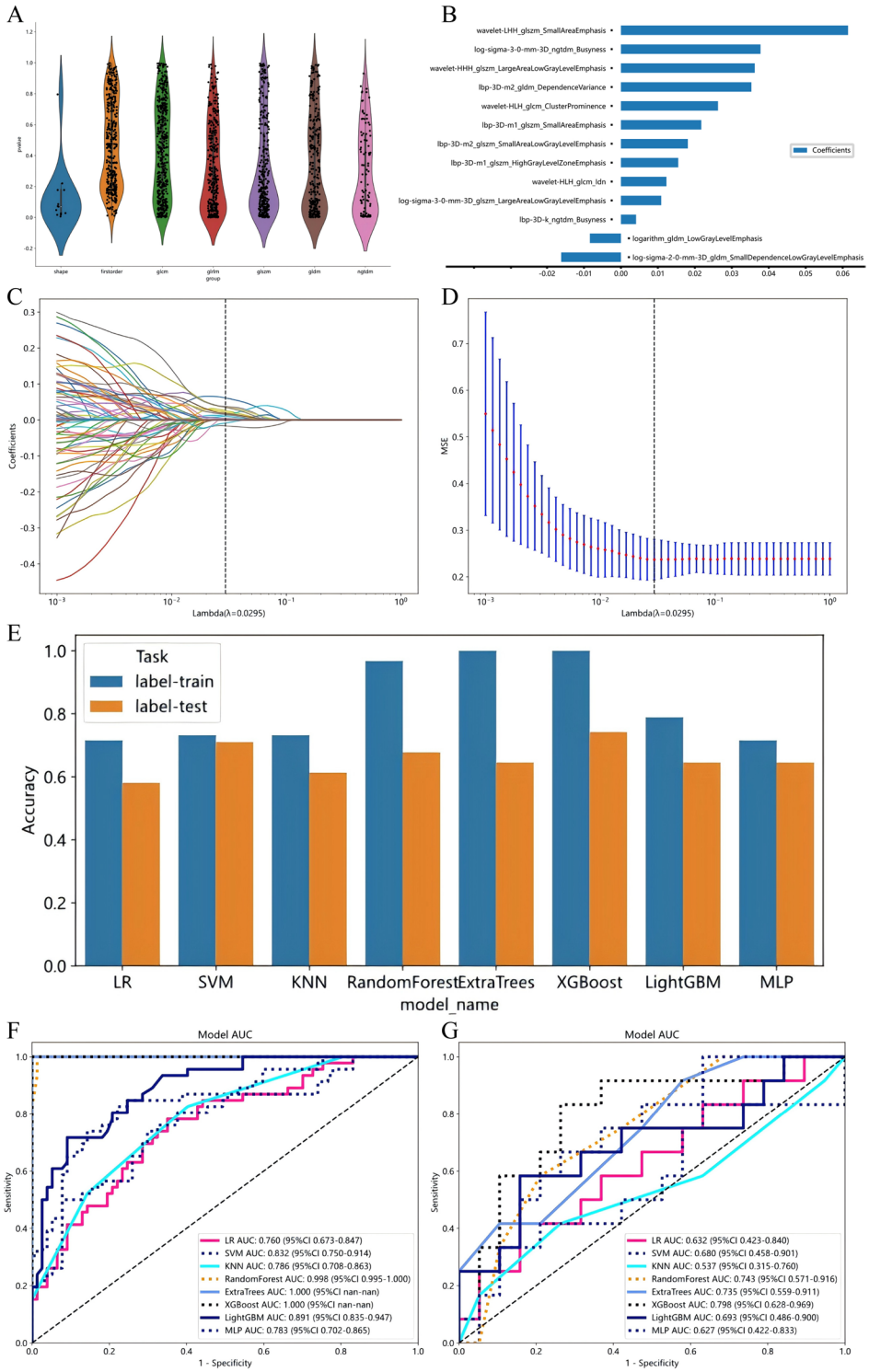

Fig. 11 **A** Feature distribution after feature extraction by pyradiomicv3.0 toolkit on the collected ROI, totaling 1830 features. **B** Use the LASSO method to select the weight of each feature with 13 high weights. **C**, **D** 13 high-weight features were selected by the LASSO method to construct our model. **E** Histogram of the accuracy of predicted results of 8 models in the training set and test set. **F** ROC curves of 8 models predicting the high and low groups of B-cell-related gene expression in the training set. **G** Eight models predicted ROC curves of high and low groups of B-cell-related gene expression in the test set

AUC, which is 0.798 respectively. 0.743 and 0.735. It can be seen from the results that XGBoost, RandomForest, and ExtraTrees algorithms have good predictive performance in predicting tumor heterogeneity. At the same time, we plotted decision curve analysis (DCA) to assess the clinical utility of the predictive model (Supplementary Fig. 6). XGBoost, RandomForest, and ExtraTrees models produce high net benefits. In summary, XGBoost, RandomForest, and ExtraTrees models all show strong comprehensive forecasting ability.

The primary structural prediction of the transcribed proteins of the model genes was performed

We searched the protein sequences of 6 genes, FKBP11, ISG20, MZB1, ALOX5AP, SSR4, and IGKC. Except for IGKC6, we found the amino acid sequences of each protein subtype after transcription. Among them, FKBP11 includes three subtypes (subtype 1, subtype 2, and subtype 3). ISG20 includes three subtypes (subtypes a, b, and c), ALOX5AP includes three subtypes (subtypes 1 and 2), and SSR4 includes three subtypes (subtypes 1, 2, and 3). We then predicted the transmembrane segment of proteins FKBP11, ISG20, MZB1, ALOX5AP, and SSR4. Figure 12 shows whether each protein subtype is an intracellular segment, transmembrane segment, or extracellular segment in each amino acid interval. For example, amino acids 1–52 of FKBP11 subtype 2 are extracellular segments and amino acids 53–75 are transmembrane segments, amino acids after 76 are intracellular segments. ISG20 subtype a subtype b, and subtype c are extracellular segments in amino acids 1–180, indicating that the protein is an extracellular protein (Fig. 12). After that, we predicted the signal peptides for all subtypes of the five proteins. If the signal peptides were present, the proteins would be cut at the amino acid nodes corresponding to the signal peptides, such as the two subtypes of FKBP11 (isoforms 1 and 3), MZB1, MZB1, and MZB1. Three subtypes of SSR4 (subtype 1, subtype 2, and subtype 3) have obvious line intersection and green line peaks, and they have obvious signal peptide sequences (Fig. 13).

Discussion

RCC is one of the most common urogenital malignancies, and according to the statistics of the American Cancer Society, the mortality rate is as high as 30% to 40% [14], but its pathogenesis remains elusive. Traditional TNM staging and imaging techniques have

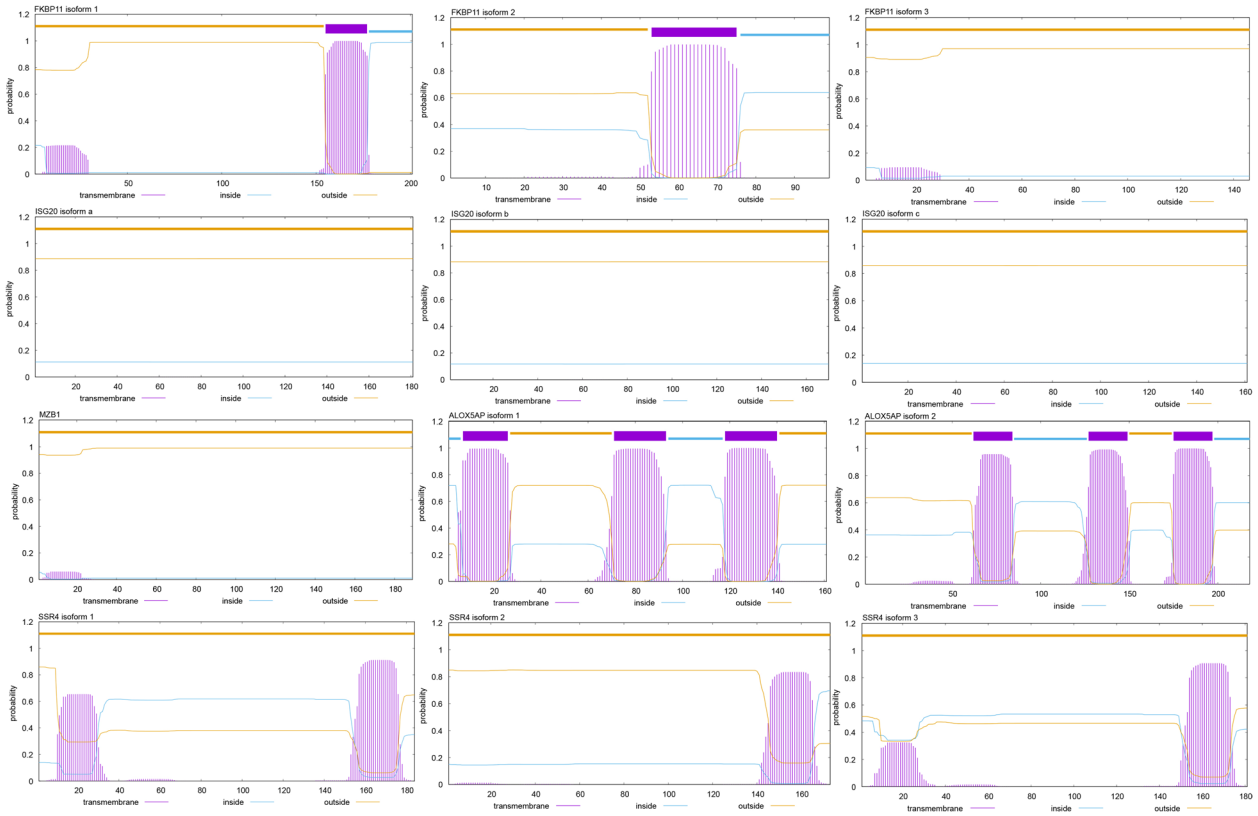

Fig. 12 shows transmembrane segment predictions for all subtypes of FKBP11, ISG20, MZB1, ALOX5AP, and SSR4 proteins

Fig. 13 shows the prediction of signal peptides for all subtypes of FKBP11, ISG20, MZB1, ALOX5AP, and SSR4 proteins

limitations in the early diagnosis and prognosis prediction of RCC patients, so there is a great need for innovative biomarkers and prognostic models. It is well known that the tumor microenvironment of RCC contains extensively infiltrated immune cells, pathologically activated fibroblasts, dysfunctional vascular endothelial cells, and a large number of extracellular matrix [15, 16], and their abnormal evolutionary patterns and constant interactions are the key to the occurrence and development of RCC, among which B cells are one of the main immune cells. The available evidence suggests that B cells are closely associated with hepatocellular carcinoma [17], lymphoma [18] gastric cancer [19], and nasopharyngeal cancer [20], but unfortunately little is known about their biological function in RCC. With the development of single-cell genome sequencing technology, researchers have gradually gained a clear understanding of the single-cell gene expression pattern of cancer, the clonal origin of tumor cells, and the direction of selective evolution [21]. Therefore, our study, starting with single-cell transcriptomics, reveals the potential mechanism of action of B cells in the RCC tumor microenvironment, to aid clinical decision-making in patients with RCC.

First, based on the characteristic genes of the cells, we annotated 6 types of cells in RCC tissue and adjacent tissues. Epithelial cells, Fibroblast cells, T/NK cells, Endothelial cells, B cells, and Myeloid cells, respectively, are the major cell components in the RCC tumor microenvironment [22]. Our results showed that there were abundant Myeloid cells, T/NK cells, and B cells in the adjacent tissues. However, RCC tissues were enriched with Epithelial cells, Fibroblast cells, and Endothelial cells. This is consistent with previous findings that RCC has strong crosstalk between endothelial tissue and extracellular matrix, which is why RCC is one of the most vascularized and aggressive solid tumors [23, 24], and our results once again prove this. In addition, most cancers, including RCC, can prevent the immune system from eradicating tumor cells by changing the way antigens are presented, activating immune checkpoint pathways, and attracting immunosuppressive cells, thus establishing a suppressive immune environment [25], which is closely related to a variety of immune cells. Compared with T cells, NK cells, and Tumor-associated macrophages, B cells seem to lack anticancer properties, but there is speculation about their potential function in the immune microenvironment of RCC [26]. Therefore, we identified

B cell differential genes and tried to elucidate the relationship between B cell and RCC based on significant B cell characteristic genes. It is also worth mentioning that, given that one of the characteristics of RCC is obvious tumor heterogeneity, we have analyzed the complex interactions among the internal cell components, the differentiation process of B cells, and the differential expression of genes during this process, which will provide an important reference value for exploring the role of B cells in RCC.

After the selection of B cell characteristic gene collection, we performed the analysis of functional pathway enrichment. Besides the obvious immune pathway, the correlation between B cell and cancer occurrence and cell–cell interaction was confirmed again. Subsequently, we identified the subtypes of RCC patients (C1 and C2) based on the expression of B cell characteristic genes, in which C1 represents a low level of B cell-related genes and C2 represents a high level of B cell-related genes. During our study, C1 and C2 showed significant differences in various aspects. It includes survival status, clinical features, active functional pathways, expression differences of acetylation-related genes, RNA methylation modification-related genes, proto-oncogenes, and so on. In addition, due to the important role of the immune microenvironment in the process of RCC, we focused on exploring the differences in immune-related indicators between C1 and C2, and the results showed that, as expected, the C2 subtype had active immune cell infiltration and immune function up-regulation, but also accompanied by increased levels of immune checkpoint related genes. This also explains the poor survival of patients with subtype C2. At the same time, the ESTIMATE analysis showed that the C2 subtype had higher tumor purity and higher immune score, so we believe that the abnormally active B cell may bring about a disturbed immune microenvironment, which will be conducive to the development and deterioration of RCC. In terms of immune escape, we evaluated TIDE scores for the C1 and C2 subtypes, and TIDE scores for the C2 subtype were also higher, suggesting higher immune escape capacity and poor immunotherapy efficacy, which is also consistent with our conclusions.

Then, to obtain an effective biological model, we constructed and verified a RCC survival prognosis model containing 6 genes based on lasso-cox regression analysis, ROC curve establishment, univariate and multivariate independent prognostic analysis, etc., which could be used to accurately predict the survival probability of RCC patients for 1, 3, and 5 years. The genes included in the model were FKBP11, ISG20, MZB1, ALOX5AP, SSR4, and IGKC, among which SSR4 and IGKC were protective genes of RCC, while FKBP11, ISG20, MZB1 and ALOX5AP were risk genes of RCC. In previous studies, these genes have been confirmed to be closely related to RCC. For example, FKBP is a family of intracellular receptors. Sun et al. found that FKBP11 is mainly related to immune-related biological processes and autophagy, and knocking down FKBP11 can inhibit the proliferation, migration and invasion of RCC cell lines [27]; Xu et al. found that ISG20 can promote the proliferation and metastasis of RCC cells by regulating the expression of MMP9/CCND1, which can be used as a potential biomarker and therapeutic target of RCC, and high expression of ISG20 is associated with poor survival [28]; Efferth et al. found that ALOX5AP was significantly overexpressed in RCC and other tumors and was related to oxidative stress response, which may be the target of RCC as a patent drug [29]; Zhou et al. identified a correlation between IGKC and the disease progression and prognosis of RCC[30]. Although the role of SSR4 in RCC has not been confirmed, its role as a protective factor in colorectal cancer has been elucidated, which can be used to predict better overall survival and treatment effects in colorectal cancer patients[31], and our study also revealed that SSR4 is an independent predictor of protection in RCC patients. It is worth mentioning that, based on the "oncoPredict" software package, we also predicted the response of RCC patients to multiple targeted drug therapies, and analyzed the potential beneficial drugs for RCC patients in the C1 and C2 subtypes and high and low-risk groups. As expected, patients in the C1 subtype and low risk groups were highly sensitive to most RCC targeted drugs. This will provide scientific reference value for guiding drug use in RCC patients.

Radiomics creates characteristic models by analyzing data to assist in predicting relevant outcomes [32]. At present, most researchers focus on combining radiomics with other related "omics" to explore the relationship between image data and disease genes, pathways, tumor microenvironment, and immune correlation [33]. Computed tomography (CT) is convenient to operate. Objective diagnosis is often used to detect kidney tumors. Here, we developed a radiomics based on CT images to predict the level of B-cell-associated gene expression in RCC, thereby achieving prognosis and responsiveness to immunotherapy in patients with RCC. From the results, we found that 8 kinds of machine learning models, XGBoost, RandomForest, and ExtraTrees models all showed strong comprehensive prediction ability, which indicated that the image omics machine learning model based on plain scan CT could accurately predict the level of B-cell related gene expression in RCC patients. And it has good clinical practicability.

Despite our new findings, this study has certain limitations in combining radiomics studies, such as the

overall small dataset, which may lead to selection bias and underfitting of results. Therefore, future studies should have a larger sample size and external test sets. In addition, considering the universality of application, we choose plain CT as the image source for the study of imaging omics, which is inferior to enhanced CT in terms of diagnostic accuracy. Therefore, in the future, enhanced Ct-based radionics combined with other "omics" can be used to predict relevant information.

As the main executor of life activities, protein is one of the important targets of life science research. However, the diversity of protein structure and the change of function make it difficult to be the focus of disease research. Understanding the transmembrane mode of the target protein and the specific site of the signal peptide is very helpful for us in predicting the role of the protein. In our study, after constructing the RCC prognostic model of 6 genes, we explored the spatial structure of their transcription proteins. If a transmembrane protein is present with a signal peptide, it will cut at the signal peptide, producing an exocyclic protein, which can help us predict how the protein will act. Our results show that three subtypes of FKBP11 subtype 1 and SSR4 have both transmembrane segments and signaling peptides, indicating that they are exocrine proteins, which is a preliminary prediction of the function and structure of the target protein. In the future, their detailed exploration may be a new direction to uncover the pathological mechanism of RCC.

Conclusion

Overall, this study reveals for the first time the important role B cells play in RCC. Based on single-cell transcriptome analysis, we mapped the single-cell landscape of RCC and screened out B-cell-related genes. We found that an abnormal immune environment and immune escape caused by a high level of B cell signal is one of the important reasons for the poor prognosis of RCC. In addition, a biological model composed of six B cell-related genes independently and effectively predicted disease outcomes in RCC patients. Our findings will provide precise and powerful guidance for addressing the clinical dilemma of RCC.

Materials and methods

Data collection

Based on the GEO database (https://www.ncbi.nlm.nih.gov/geo/), we downloaded the RCC single-celled GSE159115 sequencing data, including 13 data sets, including 7 samples of renal clear cell carcinoma, 6 samples for cancer. Based on the TCGA database (https://portal.gdc.cancer.gov/), we downloaded the clinical data of patients with RCC and gene expression matrix, a total of 485 samples were used for subsequent analysis.

Analysis of single-cell transcriptome sequencing data

We used the Seurat R software package (version 5.1.0) for data processing, first through rigorous quality control (QC), we screened high-quality cells suitable for analysis. Cells with more than 20% mitochondrial gene expression were screened, while cells with RNA expression greater than 1000 and characteristic gene numbers between 300 and 6000 were screened for subsequent analysis. LogNormalize is used to standardize the data, and FindVariableFeatures is used to identify the first 2000 highly variable genes between cells. Then ScaleData is used to normalize the data. The "RunPCA" function was used to conduct principal component analysis (PCA) on the selected high-variable genes, and extract the main sources of variation in the data, and the "harmony" method was used to integrate the data from multiple samples and remove the batch effect. In order to determine the number of principal components, the JackStraw test and Elbow Plot analysis were performed on the data. By visualizing significant principal components from the results of the DimHeatmap, "ElbowPlot", and "JackStrawPlot" functions, we obtained 15 PCs for subsequent analysis. Based on the first 15 significant principal components, cell proximity maps were constructed using FindNeighbors, and the diversity of cell subpopulations was explored by multi-resolution clustering using FindClusters at 0.01 to 2 resolutions. Clustree was then used to generate a cluster tree to visually display the clustering relationship at different resolutions. We determined that the optimal resolution for downstream analysis was 1.5, on this basis, a total of 31 clusters were screened, and the FindAllMarkers function was used to screen for feature genes. According to classical marker genes, DotPlot was used to show the expression of these genes in each cell population, and they were classified into 6 types of cells, which were as follows: Epithelial cell, Fibroblast cell, T/NK cell, Endothelial cell, B cell, and Myeloid cell. The clusters are then mapped to the tsne and umap diagrams to make their distribution clearer.

Analysis of B cell trajectory and RCC intercellular communication

In many biological processes, the development of cells is not perfectly synchronized. In response to various stimuli, cells transition from one functional state to another and the transition process of cells in different states is accompanied by the silencing and activation of a large number of genes. The R package "monocle 2" was

applied to conduct a quasi-time series analysis on B cells of RCC patients to reveal the transformation of cell state [34]. Apply the DDRTree function to reduce the size of the default setting and use the differentialGeneTest function in Monocle2 to identify DEGs that vary with pseudo time. In addition, the mode of intercellular information transfer is mainly in the form of cell surface ligand-receptors, and we used CellChat (v 1.6.1) to quantitatively infer and analyze cellular communication networks from ccRCC scRNA-seq data, in order to obtain the continuous state and interaction relationship on the cell development trajectory [35].

Functional enrichment analysis and identification of molecular subtypes

We used the R software package "org.Hs.eg.db" for GO annotation, KEGG rest API to obtain KEGG Pathway for KEGG annotation, and "Molecular Signatures Database" for reactome annotation. The gene was mapped to the background set, and the R software package "clusterProfiler" was used for enrichment analysis to obtain the enrichment results of the gene set. We used the "ConsensusClusterPlus" software package for cluster analysis and used the cumulative experience distribution function graph to determine the optimal cluster number ($k=2$) for subsequent analysis. For GSEA analysis, we obtained the GSEA software from the GSEA website, divided the samples into two groups according to the results of cluster analysis, and downloaded the KEGG sub-set from the Molecular Signatures Database to evaluate the relevant pathways and molecular mechanisms, and grouped them based on gene expression profiles and phenotypes. The minimum gene set was set to 5, the maximum gene set to 5000, and the P value of <0.05 was considered statistically significant for 1000 resamples.

methylation-related genes, acetylation-related genes, and proto-oncogenes of the two subtypes. Using the R software packages "limma", "reshape2", "ggplot2" and "ggpubr", we describe the differences in the expression of immune checkpoint-related genes between the two subtypes and create a box map to visualize them. Using the R software package "estimate", we performed an analysis of tumor microenvironment differences between the two subtypes and drew a box diagram to visualize them. Then we further carried out ssGSEA analysis, using the R software packages "limma", "reshape2", and "ggplot2", we described the analysis of immune cells and functional differences between the two subtypes and drew box diagrams for visualization. Finally, we use TIDE online tools (http://tide.dfci.harvard.edu), analyze the samples from the TCGA database immune escape and immune therapy, and use the R package "limma" and "ggplot2" map box with visualization.

Drug sensitivity analysis

Based on the R software package "oncoPredict", we predicted the therapeutic response of 485 RCC patients to 184 targeted drugs, displayed the commonly used RCC targeted drugs in the clinic and analyzed the potential beneficial drugs for RCC patients with two subtypes and two risks. The comparative analysis between the two groups was based on the Wilcox.test function.

Construction of prognostic model

We used the R software package "glmnet" to integrate survival time, survival state, and gene expression data, and performed regression analysis using the lasso-cox method. In addition, we also set up tenfold cross-validation to obtain the optimal model. We set the Lambda value to 0.0121610482551112 and ended up with 6 genes. The formula of the constructed model is as follows:

$$\begin{aligned}\text{RiskScore} = {} & 0.105985123365355 * \text{FKBP11} + 0.0223755912241455 * \text{ISG20} + 0.0118787615362657 * \text{MZB1} \\ & + 0.337174027637544 * \text{ALOX5AP} - 0.00261656598176431 * \text{SSR4} - 0.000107083273338916 * \text{IGKC}\end{aligned}$$

Analysis of differential characteristics of B cell molecular subtypes

We used the R software packages "survival" and "survminer" to integrate survival time, survival status, and gene expression data to evaluate the prognostic differences between the two molecular subtypes. We integrate the clinical data and expression matrix of RCC samples from the TCGA database and use the R software package "pheatmap" to draw them into heat maps for display. Using the R package "stringr", "Pheatmap" and "gplots", we plotted heat maps to show the expression profiles of

We used the R software package "maxstat" to calculate the optimal truncation value of RiskScore, and set the minimum sample size to be greater than 25% and the maximum sample size to be less than 75%. Finally, the optimal truncation value was 0.63656875080553. Based on this, the patients were divided into two groups, high and low risk, and then we further analyzed the prognostic difference between the two groups using the survfit function of the R software package "survival", and used the logrank test method to evaluate the prognostic difference between the different groups.

Validation of the prognostic model

We used the R software package "pheatmap" to visualize the relationship between different risk scores and patient follow-up time, events, and changes in the expression of individual genes. We performed ROC analysis using the R software package "pROC" to obtain the AUC, roc analysis using the ROC function for 1, 3, and 5-year time points and evaluated the AUC and confidence intervals using the ci function to obtain the final AUC results. We used the R software package "survival" to integrate data on survival time, survival state, and six prognostic gene features, and evaluated the prognostic significance of these features in 485 samples using Cox methods. Using the R software package "rms", we established a nomogram to evaluate the prognostic significance of these features by integrating the data of survival time, survival status, and four features. FKBP11, IGKC, and SSR4 immunohistochemical data download in HPA online database (https://www.proteinatlas.org/).

Image omics analysis

Source of CT image and Tumor segmentation

Plain scan CT data from patients with ccRCC from the Cancer Imaging Archive (TCIA), a total of 154 samples were obtained after screening for exclusion and included in this study. Exclusion criteria were: (1) low-quality TCIA CT image data not covered by TCGA transcriptome data, and (3) preoperative plain scan CT in non-CCRCC patients. CT images of these patients were then used to map tumor areas, also known as areas of interest. First, after the image voxel is resampled to $1\times1\times1$mm3, the image signal intensity is normalized, and then the gray level is standardized by z-score. Image segmentation was performed independently by two urologists with extensive experience in imaging diagnosis of kidney tumors, using the software 3D Slicer 4.11 to manually plot the ROI in the CT image layer by layer.

Radiomics feature extraction and selection

After drawing ROI, we used the pyradiomicv3.0 toolkit in Python 3.4 to extract features from the collected ROI, including first-order features, geometric features, and texture features. Please refer to the attached materials for the yaml file used for extracting features. After feature extraction, Z-score was used to normalize the feature data (mean and standard deviation were 0 and 1, respectively), and then Pearson correlation coefficient was used for feature screening ($p<0.05$). Features were filtered according to correlation coefficient and one of the two features with correlation coefficient > 0.9 was selected. Then build the training set (50%) and the test set (50%).

Construction of predictive model based on imagomics

In order to investigate the correlation between radiomics and tumor heterogeneity and disease outcome and immunotherapy, we used LASSO regression analysis for further feature screening (coefficient > 0), based on the selected tumor radiomics characteristics as independent variables, and the level of B-cell-related gene expression as dependent variables. We then constructed eight machine learning models to predict the groupings, followed by cross-validation, each using 80% training and random 20% testing to select a high level of classification performance. Machine learning algorithms include: ExtraTrees, k-Nearest Neighbor(KNN), LightGBM, Logistic Regression (LR), Multi-LayerPerceptron (MLP), Random Forest, Support vector machines (SVM) and XGBoost perform diagnostic analysis on the test set. Receiver operating characteristic (ROC) curves were plotted to evaluate the diagnostic performance of the predictive model and to analyze the corresponding area under the curve (AUC), diagnostic accuracy, sensitivity, specificity, etc. A decision curve analysis (DCA) was drawn to assess the clinical utility of the predictive model.

Primary structural prediction of proteins

In order to predict the protein function of gene expression, we can predict the expression protein structure of selected genes. First of all, we through the National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/) for FKBP11, ISG20, MZB1, ALOX5AP, SSR4, and IGKC 6 genes searched for their protein sequences. Then, we through the website http://www.cbs.dtu.dk/services/TMHMM/ will import these sequences, the transmembrane segment prediction, and, Through the website http://www.cbs.dtu.dk/services/SignalP/ transmembrane protein signal peptide prediction.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-025-13923-5.

Supplementary Material 1: Supplementary Fig. 1: Heat maps of gene expression differences in 1000 random cells in the first 20 PCS

Supplementary Material 2: Supplementary Fig. 2: clustertree of cell clustering results at different resolutions

Supplementary Material 3: Supplementary Fig. 3: Dotplot of expression of 11 possible cell type-specific genes in 32 clusters

Supplementary Material 4: Supplementary Fig. 4: Correlation coefficient matrix of 302 features after feature screening using Pearson correlation coefficient

Supplementary Material 5: Supplementary Fig. 5: Eight models were used to predict the confusion matrix of the results of the training set and the test set

Supplementary Material 6: Supplementary Fig. 6: DCA evaluated the clinical application value of 8 predictive models

Supplementary Material 7: Supplementary Table 1: A collection of selected B cell related genes

Supplementary Material 8: Supplementary Table 2: A collection of 6 genes used to construct the model

Supplementary Material 9: Supplementary Table 3: Prediction results of the eight machine learning models in the test set

Acknowledgements
Not applicable.

Authors' contributions
Jiaao Sun, Shiyan Song, Qiancheng Ma were the first author, to collect data, drafting and revising the manuscript. Feng Chen, Xiaochi Chen and Guangzhen Wu design research directions provide writing guidance and participate in the manuscript revision and supplement. All authors have agreed to the version of the manuscript for this release and have agreed to work on their respective aspects.

Funding
This study was funded by the Dalian Life and Health Field Guidance Plan (2024ZDJH01PT068).

Data availability
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 16 November 2024 Accepted: 12 March 2025
Published online: 28 March 2025

References

1. Jonasch E, Walker CL, Rathmell WK. Clear cell renal cell carcinoma ontogeny and mechanisms of lethality. Nat Rev Nephrol. 2021;17(4):245–61.
2. González-Garza R, et al. Biomarkers for evaluating the clinical response to immune checkpoint inhibitors in renal cell carcinoma (Review). Oncol Rep. 2024;52(6):1–0.
3. Kase AM, George DJ, Ramalingam S. Clear Cell Renal Cell Carcinoma: From Biology to Treatment. Cancers (Basel). 2023;15(3):665.
4. Chakiryan NH, et al. Real-world survival outcomes associated with first-line immunotherapy, targeted therapy, and combination therapy for metastatic clear cell renal cell carcinoma. JAMA Netw Open. 2021;4(5): e2111329.
5. Chipuc S, et al. Immunotherapeutic innovations in clear cell renal cell carcinoma: current strategies and future directions. Cancer Diagn Progn. 2024;4(5):558–62.
6. Cyster JG, Allen CDC. B cell responses: cell interaction dynamics and decisions. Cell. 2019;177(3):524–40.
7. Shi Y. PLAN B for immunotherapy: promoting and leveraging anti-tumor B cell immunity. J Control Release. 2021;339:156–63.
8. Xue D, et al. Tumor-infiltrating B cells: their dual mechanistic roles in the tumor microenvironment. Biomed Pharmacother. 2024;179: 117436.
9. Downs-Canner SM, et al. B cell function in the tumor microenvironment. Annu Rev Immunol. 2022;40:169–93.
10. Wouters MCA, Nelson BH. Prognostic significance of tumor-infiltrating B cells and plasma cells in human cancer. Clin Cancer Res. 2018;24(24):6125–35.
11. Li X, et al. Construction of a B cell-related gene pairs signature for predicting prognosis and immunotherapeutic response in non-small cell lung cancer. Front Immunol. 2022;13: 989968.
12. Wang Q, et al. Single-cell transcriptome sequencing of B-cell heterogeneity and tertiary lymphoid structure predicts breast cancer prognosis and neoadjuvant therapy efficacy. Clin Transl Med. 2023;13(8): e1346.
13. Chevrier S, et al. An immune atlas of clear cell renal cell carcinoma. Cell. 2017;169(4):736–749.e18.
14. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin. 2024;74(1):12–49.
15. Shen M, Du Y, Ye Y. Tumor-associated macrophages, dendritic cells, and neutrophils: biological roles, crosstalk, and therapeutic relevance. Med Rev (2021). 2021;1(2):222–43.
16. Deng J, Fleming JB. Inflammation and myeloid cells in cancer progression and metastasis. Front Cell Dev Biol. 2021;9: 759691.
17. Xu KQ, et al. B-cell-specific signatures reveal novel immunophenotyping and therapeutic targets for hepatocellular carcinoma. World J Gastroenterol. 2024;30(34):3894–925.
18. Martínez LE, et al. Characterization of unique B-cell populations in the circulation of people living with HIV prior to non-Hodgkin lymphoma diagnosis. Front Immunol. 2024;15:1441994.
19. Cai X, et al. Re-analysis of single cell and spatial transcriptomics data reveals B cell landscape in gastric cancer microenvironment and its potential crosstalk with tumor cells for clinical prognosis. J Transl Med. 2024;22(1):807.
20. Ge Y, et al. Immunoinhibitory effects of hypoxia-driven reprogramming of EGR1(hi) and EGR3 positive B cells in the nasopharyngeal carcinoma microenvironment. Oral Oncol. 2024;158: 106999.
21. Cao YH, et al. Deciphering cell-cell interactions and communication in the tumor microenvironment and unraveling intratumoral genetic heterogeneity via single-cell genomic sequencing. Bioengineered. 2022;13(7–12):14974–86.
22. Lasorsa F, et al. Cellular and molecular players in the tumor microenvironment of renal cell carcinoma. J Clin Med. 2023;12(12):3888.
23. Monjaras-Avila CU, et al. The tumor immune microenvironment in clear cell renal cell carcinoma. Int J Mol Sci. 2023;24(9):7946.
24. Sjöberg E. Molecular mechanisms and clinical relevance of endothelial cell cross-talk in clear cell renal cell carcinoma. Ups J Med Sci. 2024;129:10.
25. Díaz-Montero CM, Rini BI, Finke JH. The immunology of renal cell carcinoma. Nat Rev Nephrol. 2020;16(12):721–35.
26. Sweis RF, Galsky MD. Emerging role of immunotherapy in urothelial carcinoma-Immunobiology/biomarkers. Urol Oncol. 2016;34(12):556–65.
27. Sun Z, et al. Multi-omics analysis of the expression and prognosis for FKBP gene family in renal cancer. Front Oncol. 2021;11: 697534.
28. Xu T, et al. ISG20 serves as a potential biomarker and drives tumor progression in clear cell renal cell carcinoma. Aging (Albany NY). 2020;12(2):1808–27.
29. Özenver N, Efferth T. Identification of prognostic and predictive biomarkers and druggable targets among 205 antioxidant genes in 21 different tumor types via data-mining. Pharmaceutics. 2023;15(2):427.
30. Zhou J, et al. Identification of genes that correlate clear cell renal cell carcinoma and obesity and exhibit potential prognostic value. Transl Androl Urol. 2021;10(2):680–91.
31. He W, et al. SSR4 as a prognostic biomarker and related with immune infiltration cells in colon adenocarcinoma. Expert Rev Mol Diagn. 2022;22(2):223–31.
32. Gardin I, et al. Radiomics: principles and radiotherapy applications. Crit Rev Oncol Hematol. 2019;138:44–50.
33. Wu J, et al. Heterogeneous enhancement patterns of tumor-adjacent parenchyma at MR imaging are associated with dysregulated signaling pathways and poor survival in breast cancer. Radiology. 2017;285(2):401–13.
34. Qiu X, et al. Single-cell mRNA quantification and differential analysis with census. Nat Methods. 2017;14(3):309–15.
35. Jin S, et al. Inference and analysis of cell-cell communication using cell chat. Nat Commun. 2021;12(1):1088.

Publisher's Note