Identification of Potential Biomarkers Associated With Immune Infiltration in Papillary Renal Cell Carcinoma

Ran Deng

Lanzhou University Second Hospital

Jianpeng Li

Lanzhou University Second Hospital

Hong Zhao

Lanzhou University Second Hospital

Zhirui Zou

Lanzhou University Second Hospital

Jiangwei Man

Lanzhou University Second Hospital

Jinlong Cao

Lanzhou University Second Hospital

Li Yang (✉ ery_yangli@lzu.edu.cn)

Lanzhou University Second Hospital

Research Article

Keywords: papillary renal cell carcinoma, tumor-infiltrating immune cells, hub gene, tumor microenvironment, biomarkers

Posted Date: August 7th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-769170/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Journal of Clinical Laboratory Analysis on October 4th, 2021. See the published version at <https://doi.org/10.1002/jcla.24022>.

Abstract

Background: Immunotherapeutic approaches have recently emerged as effective treatment regimens against various types of cancer. However, the immune-mediated mechanisms surrounding papillary renal cell carcinoma (pRCC) remain unclear. This study aimed to investigate the tumor microenvironment (TME) and identify the potential immune-related biomarkers for pRCC.

Methods: The CIBERSORT algorithm was used to calculate the abundance ratio of immune cells in each pRCC sample downloaded from the database UCSC Xena. Univariate Cox analysis was used to select the prognostic-related tumor-infiltrating immune cells (TIICs). Multivariate Cox regression analysis was performed to develop a signature based on the selected prognostic-related TIICs. Then, these pRCC samples were divided into low- and high-risk groups according to the obtained signature. Analyses using Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Gene Set Enrichment Analysis (GSEA) were performed to investigate the biological function of the DEGs (differentially expressed genes) between the high- and low-risk groups. The hub genes were identified using a Weighted Gene Co-Expression Network Analysis (WGCNA) and a Protein-Protein Interaction (PPI) analysis. The hub genes were subsequently validated using Kaplan-Meier survival analysis, Receiver Operating Characteristic (ROC) analysis, a nomogram prediction model, and via the Gene Expression Omnibus (GEO) database, and the Human Protein Atlas (HPA) database. Finally, we validated the correlation between the nine hub genes and immune cells using the XCELL algorithm.

Results: According to our analyses, nine immune cells play a vital role in the TME of pRCC. Our analyses also obtained nine potential immune-related biomarkers for pRCC, including *TOP2A*, *BUB1B*, *BUB1*, *TPX2*, *PBK*, *CEP55*, *ASPM*, *RRM2*, and *CENPF*.

Conclusion: In this study, our data revealed the crucial TIICs and potential immune-related biomarkers for pRCC and provided compelling insights into the pathogenesis and potential therapeutic targets for pRCC.

1. Introduction

Renal cell carcinoma (RCC) is a common urologic malignancy and has increased incidence in recent years [1]. Aside from clear cell renal cell carcinoma (ccRCC), papillary renal cell carcinoma (pRCC) is the most common subtype of RCC and accounts for 10%-15% of all RCC cases [2]. The histology of pRCC is characterized by a papillary fibrous vascular core of tumor cells [3] and is markedly different from that of ccRCC. There are two main histological subtypes of pRCC, depending on their characteristics. Compared with type 1 tumors, type 2 tumors have a worse prognosis. There was no significant difference in metastatic disease outcomes between pRCC and ccRCC [3]; however, the immune cell infiltration in the TME of pRCC is still unclear, and the elucidation of the different effects of immunotherapy is still in its infancy. Therefore, it is necessary to more comprehensively explore the pathogenesis of pRCC and its potential immune-related targets.

Due to its excellent efficacy, tumor immunotherapy has gradually become the fourth most commonly administered type of tumor treatment after surgery, radiotherapy, and chemotherapy [4]. Effective cancer immunotherapy requires overcoming the immunosuppressive TME [5]. Immune cells constitute a large fraction of the TME, and they function by modulating tumor progression and regulating antitumor immune responses [6]. Therefore, the elucidation of the mechanisms of immune cell infiltration in pRCC is crucial to understand the dynamics of its TME and to identify potential immunotherapeutic targets.

Recently, numerous studies have suggested that the TME components are associated with patient survival outcomes [7, 8]. Specifically, immune cell composition in the TME is associated with survival in patients with RCC [7]. Blass E *et al.* [9] demonstrated that CD8 + T cell responses had been associated with survival in patients with RCC treated with anti-PD-L1 antibodies. However, most studies on RCC are limited to ccRCC, so the specific mechanism of the immune microenvironment in pRCC remains unclear.

In this study, we downloaded transcriptome data for pRCC from the database UCSC Xena. We first performed univariate Cox regression analysis to screen out the prognosis-related immune cells. We then performed multivariable Cox regression to calculate the risk scores of the screened prognosis-related immune cells, using the median value as the threshold to divide the samples into the high- and low-risk groups. Then, the DEGs and the hub genes between the high-risk and low-risk groups were identified. Our study elucidates the mechanisms and dynamics of the TME and the interaction between pRCC and infiltrating immune cells. In addition, this study also provides a new avenue for researchers on immunotherapy to explore the immune-related cellular and genetic targets for pRCC therapy.

2.materials And Methods

2.1 Data sources

Clinical information and the gene expression information from RNA-seq data of patients with pRCC were downloaded from the UCSC Xena database (<https://xenabrowser.net/datapages/>). The data obtained were normalized count data, including information from 291 tumor tissues and 32 adjacent normal tissues. The GSE26574 dataset was downloaded from the GEO database (<https://www.ncbi.nlm.nih.gov/gds/>), which included 34 tumor samples and 15 normal samples.

2.2 Screening prognostic-related immune cells

CIBERSORT is an approach that primarily characterizes the immune cell composition of diverse tissue according to their gene expression profiles [10]. The CIBERSORT R source code and the LM22 signature matrix file, both of which obtain information from a total of 22 immune cell types, were downloaded from a website (<https://cibersort.stanford.edu/>). The R CIBERSORT algorithm was used to determine the abundance ratio of 22 TIICs (including 22 immune cells) in the 291 tumor samples obtained from the UCSC Xena database. The “barplot” and “pheatmap” [11] functions of R software were used to visualize the abundance ratio matrix from the data obtained in the previous algorithm. The Pearson correlation coefficients among these TIICs were then calculated using the “cor” function of the R package and was

visualized using the “corrplot” package [12], also in R. Finally, univariate Cox analysis was performed to screen the prognostic related TIICs, with the threshold set at $P < 0.05$. The “forestplot” R package [13] was used to visualize the prognostic-related immune cells.

2.3 Identifying the relationship between prognostic-related TIICs and clinical traits

A Wilcoxon test with a P -value < 0.05 was used to analyze the relationship between prognostic-related TIICs and clinical traits. The “boxplot” function of R software was used to visualize the data obtained from this analysis. Kaplan-Meier (KM) survival analysis was then implemented using the R package “survival” [14] to show the relationship between the prognostic related TIICs and OS (overall survival) in patients with pRCC.

2.4 Defining the high- and low-risk groups based on the prognostic-related TIICs

The risk score for each pRCC sample was calculated using the regression coefficients derived from the multivariate Cox regression analysis for the prognostic-related TIICs. These samples were separated into high- and low-risk groups using the median risk score as a cut-off. A KM analysis via the log-rank test was performed to assess the survival differences between the high-risk and low-risk groups using the “survival” [14] package in R. A nomogram prediction model was then constructed for these TIICs using the R package “rms” [15]. A time-dependent ROC curve and a corresponding calibration curve were derived to assess the predictive significance of the model by using the “rms” [15] and “timeROC” [16] packages in R.

2.5 Identification and functional annotation of the differentially expressed genes (DEGs) between the high- and low-risk groups

Differential expression analysis between the high-risk and low-risk groups was conducted using the “DESeq2” [17] package in R, with P -values < 0.05 and $|\text{fold change}| > 1$ as the filters to determine if these genes were significantly differentially expressed. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were performed using the “clusterProfiler” [18] R package and “org.Hs.eg.db” to reveal the potential functions of these DEGs, with a p -value < 0.05 set as the cut-off. Gene Set Enrichment Analysis (GSEA) was then performed, also using the “clusterProfiler” [18] R package and immunologic signature gene set (C7 gene sets) obtained from the Molecular Signatures Database (MSigDB, <https://www.gsea-msigdb.org/>).

2.6 Construction of a weighted gene co-expression network

A weighted gene co-expression network analysis (WGCNA) is a tool typically used for constructing gene co-expression networks. It aims to explore the correlation between gene networks and clinical phenotypes and identify the core genes in the network. We employed the R package “WGCNA” using a soft-thresholding power of 4 and a minimum module size of 30 [19] to construct a co-expression network for

the DEGs obtained. We then calculated the correlation between the modules obtained and clinical traits to identify the core modules

2.7 Constructing a protein-protein interaction (PPI) network and determining the hub genes

The STRING database (<https://string-db.org/>) was used to construct the PPI network of the DEGs in the core modules, those with an interaction score > 0.4. The five topological analysis methods in the “cytoHubba” app of Cytoscape v3.8.2, including Closeness, Maximal Clique Centrality (MCC), Maximum Neighborhood Component (MNC), Degree, and Edge Percolated Component (EPC), were utilized to screen the top 30 genes in the PPI network constructed. The hub genes were then identified by intersecting the top 30 genes obtained after running the various algorithms in the topological analyses mentioned above.

2.8 Verification of the clinical significance of the hub genes

To further validate the clinical significance of the obtained hub genes, we analyzed the relationship between their expression profile and the pathological stage of the sample using the “ggpubr” [20] R package. A KM survival analysis and the log-rank test using the R package “survival” were employed to analyze the relationship between the hub genes and OS. An ROC curve was obtained to evaluate the diagnostic value of the hub genes using the “pROC” package in R [21]. Moreover, a nomogram prediction model for the hub genes was constructed using the “rms” [15] package in R. A time-dependent ROC and calibration curve were also derived to assess the predictive significance of the model using the “rms” [15] and “timeROC” packages [22] in R. Finally, to confirm these results, the protein levels of these hub genes were verified using immunohistochemistry (IHC) data obtained from the Human Protein Atlas (HPA) database (<https://www.proteinatlas.org/>). The GSE26574 dataset served as the external validation dataset for these hub genes.

2.9 Validation of the correlation between the hub genes and infiltrating immune cells

Spearman correlations between the infiltration of all 22 types of immune cells and the expression of hub genes were calculated via the “psych” [23] R package, and the results were visualized using the “pheatmap” [11] package in R. Meanwhile, The XCELL algorithm in Tumor Immune Estimation Resource 2.0 (TIMER2.0) [24], was also used to perform comprehensive correlation analyses between the immune cell signatures and the obtained hub genes.

3. Results

3.1 Identifying the prognosis-related immune cells

The infiltration levels of 22 immune cells in the 291 pRCC samples obtained from patients are shown in Figs. 2A and 2B. Macrophages M2, T cells CD4 memory resting, and Mast cells resting were of higher abundance than other immune cells. The correlation among these immune cells is shown in Fig. 2C.

Additionally, we screened the nine immune cells associated with OS via univariate Cox analysis, and the results were shown in Fig. 2D. The immune cells associated with OS were T cells follicular helper, Macrophages M1, Dendritic cells activated, T cells regulatory (Tregs), B cells memory, T cells CD8, Macrophages M2, B cells naïve, and T cells CD4 memory activated.

3.2 Clinical significance of prognostic-related immune cells

We also analyzed the association of the nine prognostic-related immune cells with various clinical traits, including TNM stage, tumor type, pathological stage, and OS. The results of this analysis are shown in Fig. 3. The degree of infiltration of Macrophage M1, Macrophage M2, T cells CD8, and T cells regulatory (Tregs) differed between type 1 and type 2 pRCC (Figs. 3A-3D). Macrophage M1, Macrophage M2, and T cells regulatory (Tregs) were closely related to the pathological stage (Fig. 3E-3G). Macrophage M1, Macrophage M2, and T cells regulatory (Tregs) were associated with the TNM stage (Figs. 3H-3L). B cells naïve, Macrophage M1, and T cells CD4 memory activated were related to overall survival (Figs. 3M-3O).

3.3 Classification of patient data into low- and high-risk groups and construction of a nomogram prediction model

To further investigate immune cell infiltration in pRCC, we constructed a risk grouping and prediction model based on the prognostic related TIICs obtained. The risk score of each patient was calculated using the regression coefficients of the multivariate Cox regression of the nine prognosis-related TIICs. Furthermore, the median risk score cut-off for the patients was divided into a low-risk group and a high-risk group (Fig. 4A). The analysis showed that more patients died in the high-risk group than in the low-risk group (Fig. 4B). The degree of infiltration of the nine prognosis-related TIICs is shown in Fig. 4C. The KM survival curves demonstrated that the OS of the high-risk group was shorter than that of the low-risk group ($P=0.0028$), indicating that the grouping based on the risk score was reasonable (Fig. 4D). A nomogram was then constructed based on a multivariate Cox regression analysis of the prognosis-related TIICs (Fig. 4E). The calibration curve (Fig. 4F) and the time-dependent ROC analysis (Fig. 4G) of the nomogram were used to estimate the accuracy of the actual observed rates with the predicted survival probability. Our results show that the nomogram has a good prognostic value, suggesting that the nine TIICs are closely associated with OS in pRCC.

3.4 Identification and functional annotation of the DEGs

In this study, 1157 DEGs (1097 upregulated and 60 downregulated genes) were identified between the high-risk and low-risk groups. A volcano plot was obtained, and the results of the differential analysis are shown in Fig. 5A. GO and KEGG enrichment analysis was then performed to investigate the biological functions associated with the obtained DEGs. The significantly enriched KEGG pathways are shown in Fig. 5B (Supplementary Table). The pathways related to cytokine-cytokine receptor interactions and the cAMP signaling pathway, which play a key role in pRCC, were significantly enriched. The significant GO enrichment terms of the upregulated and downregulated DEGs are shown in Fig. 5D and Fig. 5E,

respectively. Enriched biological functions related to immunity include the humoral immune response, IgG receptor activity, and T cell chemotaxis.

Further, GSEA was used to clarify the relationship between these DEGs and immunity. The results of the GSEA are shown in Fig. 5E (Supplementary Table) and showed that regulatory T cells and CD8 T cells, which play a crucial role in cancer-related immunity, were significantly enriched.

3.5 Identification of the core modules via WGCNA analysis

We also performed a complete WGCNA analysis on the 1157 DEGs, including screening the soft thresholds, constructing a dynamic tree cut, and plotting a network heatmap (Figs. 6A-6C). The results showed that the 1157 DEGs could be divided into six modules (blue, brown, green, grey, turquoise, yellow). There were 496 DEGs in the turquoise module, 313 in the blue module, 167 in the brown module, 135 in the yellow module, 95 in the green module, and 40 in the gray module (Fig. 6B). Subsequently, we investigated the relationship between the six modules and clinical traits (pathologic TNM, pathologic stage, and tumor-type) and found that the turquoise and blue modules were strongly associated with pathological stage and T stage (Fig. 6D). Both gene significance and module membership were plotted for the blue module (Figs. 6E-6F), and the results indicated that this module was significantly related to pathological stage ($\text{cor} = 0.91$, $p = 6\text{e-}121$) and T stage ($\text{cor} = 0.89$, $p = 4.3\text{e-}108$). Likewise, the gene significance of the turquoise module was also related to pathological stage ($\text{cor} = 0.83$, $p = 2.4\text{e-}127$) and T stage ($\text{cor} = 0.81$, $p = 1.4\text{e-}116$) (Figs. 6G-6H). Hence, the blue and turquoise modules were used as core modules for further analyses.

3.6 Identification of hub genes via multiple algorithms

The 496 genes in the turquoise module and the 313 genes in the blue module were analyzed using the STRING database, and a PPI network containing 706 nodes and 4261 edges was obtained. The top 30 genes in the PPI network were screened using DMNC, MCC, MNC, Degree, and EPC (Fig. 7A-7E). Based on the five algorithms, the PPI network contained nine common hub genes (*TOP2A*, *BUB1B*, *BUB1*, *TPX2*, *PBK*, *CEP55*, *ASPM*, *RRM2*, and *CENPF*) (Fig. 7F). The detailed information regarding these nine hub genes is shown in Table 1.

3.7 Verification of the clinical significance of the hub genes

The nine hub genes obtained were overexpressed in tumor tissues compared to normal tissues (Fig. 8A). The expression profile of these hub genes varied across different pathological stages and tended to increase with an increasing pathological stage (Fig. 8B). In addition, we analyzed the relationship between the nine hub genes and the OS and found that the group with a lower expression of these genes had a better prognosis than the group with a high expression of these genes (Fig. 8C). ROC curves were obtained to evaluate the diagnostic value of these genes in pRCC tumors and different tumor types (Figs. 8D-8E). Results of the analysis indicate that all these genes had a certain accuracy in differentiating between type 1 from type 2 pRCC, except *TPOX2* (area under the curve [AUC] < 0.5)

(Fig. 8D). The AUCs of all the hub genes were greater than 0.8, suggesting that they have good diagnostic value in differentiating normal tissues from pRCC (Fig. 8E).

The results of univariate and multivariate Cox regression analyses of the hub genes are shown in Figs. 9A and 9B. The univariate Cox regression analysis results showed that these genes were both significantly associated with a shorter OS (Fig. 9A). In the multivariate Cox regression analysis, *ASPM*, *BUB1B*, and *TPX2* were found to be independent prognostic factors of patients with pRCC (Fig. 9B). Based on the results of both the univariate and multivariate Cox regression analyses, we further constructed a nomogram model of these genes and assessed the predictive significance of the model via a time-dependent ROC and calibration curve (Figs. 9C-9E). The nomogram achieved a good area under the ROC curve, between 0.787 and 0.893, and fitted well with the obtained calibration curves, indicating that the model provided good predictability.

The protein expression levels of the hub genes were verified through the HPA database, and the results are shown in Figs. 10A-10G. *TPX2*, *TOP2A*, *CEP55*, and *CEBPF* were confirmed to be expressed in renal tumors. External validation for these hub genes found that they were consistent with the analyses in TCGA (Fig. 10H).

3.8 Validation of the correlation between the hub genes and the TIICs

The correlation of the nine hub genes with 22 TIICs was analyzed, and the results are shown in Fig. 11A. It was observed that Macrophages M2 and Macrophages M1 were related to OS and they were also significantly associated with the nine hub genes. Moreover, the XCELL algorithm was used to verify the correlation between the hub genes and immune infiltration of macrophages. The results of the correlation analysis as calculated via XCELL were consistent with the results obtained using the CIBERSORT algorithm (Fig. 11B).

Discussion

Tumor-infiltrating immune cells are a critical part of the tumor microenvironment. These cells regulate tumor growth, invasion, and metastasis by altering the immune status of tumor cells. In this study, we evaluated the status of the nine TIICs related to OS in 291 patients with pRCC and found that patients with high-risk scores had a shorter overall survival time. Nine immune-related markers for pRCC were screened out via multiple analyses by comparing the differences between the high-risk and low-risk groups.

In this study, nine TIICs related to OS, including T cells follicular helper, Macrophages M1, Dendritic cells activated, T cells regulatory (Tregs), B cells memory, T cells CD8, Macrophages M2, B cells naïve, T cells CD4 memory activated, and T cells CD4 memory resting, were identified via univariate Cox analysis. Previous studies have shown that macrophages are one of the most abundant cell types in the tumor

microenvironment, contributing to tumor progression [25, 26]. In our study, correlation analysis with clinical traits showed that macrophages affected the tumor type, pathological stage, TNM stage, and prognosis of patients with pRCC. Moreover, we found that macrophages were also the most abundant cell types in pRCC and were significantly associated with the nine hub genes we obtained (*TOP2A*, *BUB1B*, *BUB1*, *TPX2*, *PBK*, *CEP55*, *ASPM*, *RRM2*, and *CENPF*). Therefore, the results of previous studies are consistent with the results obtained in this study [25, 26].

Functional enrichment analysis showed that the DEGs in the high- and low-risk groups were significantly associated with tumor cell infiltration-related pathways. Both the cytokine-cytokine receptor and cAMP signaling pathways were significantly enriched, based on our analyses. Cytokines play an important role in cancer-related immune responses and promote tumor angiogenesis, tumor cell invasion, and tumor cell metastasis [27, 28]. IFN- γ is one of the cytokines necessary in immunomodulation and anticancer immunity, which also induces the expression of PD-L1 in most tumor cells [29]. The cAMP signaling pathway is a critical modulator of specific tumor cell properties such as proliferation, differentiation, and migration [30]. Dou A-X *et al.* showed that regulatory T cells might suppress the antitumor immune response through the intercellular transport of cAMP and the activation of the cAMP-protein kinase A signaling pathway [31]. Furthermore, the GSEA found that regulatory T cells were significantly enriched in pRCC, which was consistent with the KEGG enrichment results.

Currently, researchers have reported that hub genes such as *TPX2*, *TOP2A*, and *BUB1* are closely associated with pRCC progression [32–34]. A study based on computer modeling showed that *TPX2* was a potential risk biomarker involved in cell proliferation in pRCC [34]. Further, bioinformatics analyses showed that both *TOP2A* and *BUB1* were predicted to be related to the occurrence and development of pRCC [32, 33]. In addition, some studies have also reported the effect of the expression of these genes in other tumors [35, 36]. *TOP2A* cleavage is a broad DNA damage mechanism found in oncogenic translocations [37]. MiR-139-5 upregulates the expression of its target gene *TOP2A* and promotes the progression of ccRCC [35]. Downregulation of *TPX2* inhibits the proliferation and invasion of endometrial cancer cells and promotes the apoptosis of EC-derived cells [36]. It also suppresses the growth of liver cancer by regulating the PI3K/AKT signaling pathway [38]. A dysregulated expression of *BUB1* leads to aberrant chromosomal replication and aneuploidy, contributing to the development of tumors [39]. Previous studies have demonstrated that these three genes play an important role in tumorigenesis and cancer development, and verified the role of these genes in pRCC through computer models and bioinformatics analyses.

The underlying mechanism of the other hub genes such as *BUB1B*, *PBK*, *CEP55*, *ASPM*, *RRM2*, and *CENPF* remains unclear in pRCC. However, we found that these hub genes strongly correlated with macrophages and had clinical significance in pathological staging, survival, and diagnostic value. A nomogram obtained containing these hub genes had good predictability, suggesting that these genes are significantly associated with prognosis. Therefore, these hub genes are potential immune-related biomarkers for pRCC. Additionally, our conclusions are supported by the results of previous studies. Mechanistically, it modulates the transcriptional activation of the mitotic checkpoint kinase *BUB1B*, which

also promotes tumor growth and chemoresistance, leading to poor outcomes for patients with lung adenocarcinoma [40]. An *in vitro* study showed that the overexpression of BUB1B enhanced the proliferation, migration, and invasion of prostate cancer cell lines, while the removal of BUB1B did not affect these cell functions [41]. PBK is a novel serine-threonine kinase related to the mitogen-activated protein kinase (MAPK) family. It is an important link in many carcinogenic signaling pathways, including p38, extracellular signal-regulated kinase 1/2 (ERK1/2), and the FAK/Src-MMP signal pathways [42]. CEP55 is a key regulatory factor of the cytoplasmic split and is associated with genomic instability; genomic instability is a hallmark of cancer [43]. Carcinogenic CEP55 regulates the proliferation, migration, and invasion of tumor cells as mediated by the PI3K/AKT/mTOR pathway and is related to poor prognosis [44]. The potential role of *ASPM* on pRCC is still unclear; however, it was highly expressed in various cancers such as endometrial cancer and lung squamous cell carcinoma and is related to poor clinical prognosis and an increased risk of cancer recurrence [45, 46]. RRM2 is essential for DNA synthesis and repair and is frequently overexpressed in various cancers [47]. The oncogenic role of RRM2 has been linked to the promotion of epithelial-mesenchymal transition (EMT) and angiogenesis [47]. CENPF is a critical regulator of cancer metabolism, and the silencing of CENPF has been shown to increase the expression of inactive forms of pyruvate kinase M2, a rate-limiting enzyme indispensable for an irreversible reaction in glycolysis and reduces global bio-energetic capacity, acetyl-CoA production, histone acetylation, and lipid metabolism [48]. The above analysis mainly reveals that the hub genes obtained play an important role in tumor development and may be potential therapeutic targets in the treatment of pRCC.

In summary, nine prognostic-related immune cells and nine hub genes were identified, providing compelling insights into the pathogenesis of pRCC and may serve as potential therapeutic targets for pRCC. These nine types of immune cells may also provide important clues so we can better understand the immune microenvironment in pRCC. The hub genes obtained can be considered as biomarkers for the prognosis of pRCC and may also serve as key targets for immunotherapy in pRCC. However, it is important to point out that the evidence presented in this study was obtained indirectly from bioinformatics analyses, which is its major limitation.

Abbreviations

ccRCC:clear cell renal cell carcinoma; RCC:renal cell carcinoma; TME:tumor microenvironment; pRCC: papillary renal cell carcinoma; KIRP: Kidney renal papillary cell carcinoma. TIICs: tumor-infiltrating immune cells. DEGs: differentially expressed genes; GEO: Gene Expression Omnibus; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes;GSEA: Gene Set Enrichment Analysis; IHC: immunohistochemical; PPI: Protein-Protein Interaction; ROC: Receiver Operating Characteristic; TCGA: The Cancer Genome Atlas; WGCNA: Weighted Gene Co-Expression Network Analysis; HPA:Human Protein Atlas HPA.

Declarations

Author contributions

Ran Deng and Jianpeng Li contributed this article equally

The study conception and design were performed by RD, PL, and LY. RD, PL, HZ, ZZ, JM, and JC analyzed the data, authored or reviewed drafts of the paper, approved the final draft.

Data availability statement

All original data were loaded from public database. These data are publicly available.

Funding

This study was supported by grants from the Gansu Provincial Natural Science Foundation (grant no. 17JR5RA237).

Acknowledgment

We are grateful for TCGA, UCSC, and GEO database.

Declarations

Conflict of interest No potential conflict of interest was reported by the author(s).

Ethics statement All data used in this work are publicly available.

References

1. Liu X, Mao Y-H, He X-M et al (2017) Analysis on Inpatient Health Expenditures of Renal Cell Carcinoma in a Grade-A Tertiary Hospital in Beijing. *Chin Med J* 130:2447–2452. <https://doi.org/10.4103/0366-6999.216412>
2. Huang J, Huang D, Yan J et al (2020) Comprehensive subgroup analyses of survival outcomes between clear cell renal cell adenocarcinoma and papillary renal cell adenocarcinoma. *Cancer Med* 9:9409–9418. <https://doi.org/10.1002/cam4.3563>
3. Mendhiratta N, Muraki P, Sisk AE, Shuch B (2021) Papillary renal cell carcinoma. *Review Urologic Oncology: Seminars Original Investigations* 39:327–337. <https://doi.org/10.1016/j.urolonc.2021.04.013>
4. Zhao B, Zhang W, Yu D et al (2017) Adoptive immunotherapy shows encouraging benefit on non-small cell lung cancer: a systematic review and meta-analysis. *Oncotarget* 8:113105–113119. <https://doi.org/10.18632/oncotarget.19373>
5. Marigo I, Zilio S, Desantis G et al (2016) T Cell Cancer Therapy Requires CD40-CD40L Activation of Tumor Necrosis Factor and Inducible Nitric-Oxide-Synthase-Producing Dendritic Cells. *Cancer Cell* 30:377–390. <https://doi.org/10.1016/j.ccell.2016.08.004>

6. Shirota H, Klinman DM, Ito S et al (2017) IL4 from T Follicular Helper Cells Downregulates Antitumor Immunity. *Cancer Immunol Res* 5:61–71. <https://doi.org/10.1158/2326-6066.CIR-16-0113>
7. Chevrier S, Levine JH, Zanutelli VRT et al (2017) An Immune Atlas of Clear Cell Renal Cell Carcinoma. *Cell* 169:736–749.e18. <https://doi.org/10.1016/j.cell.2017.04.016>
8. Yang L, Fu J, Zhou Y (2018) Circular RNAs and Their Emerging Roles in Immune Regulation. *Front Immunol* 9:2977. <https://doi.org/10.3389/fimmu.2018.02977>
9. Blass E, Ott PA (2021) Advances in the development of personalized neoantigen-based therapeutic cancer vaccines. *Nat Rev Clin Oncol* 18:215–229. <https://doi.org/10.1038/s41571-020-00460-2>
10. Newman AM, Steen CB, Liu CL et al (2019) Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* 37:773–782. <https://doi.org/10.1038/s41587-019-0114-2>
11. Raivo K pheatmap: Pretty Heatmaps
12. Wei T, Simko V (2017) R package “corrplot”. Visualization of a Correlation Matrix
13. Gordon M, Thomas L (2020) forestplot: Advanced Forest Plot Using “grid” Graphics
14. Terry M, Patricia MA Package for Survival Analysis in R. R package version 3.1–12
15. Frank E (2020) rms: Regression Modeling Strategies
16. Blanche P, Dartigues J-F, Jacqmin-Gadda H (2013) Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Statist Med* 32:5381–5397. <https://doi.org/10.1002/sim.5958>
17. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550. <https://doi.org/10.1186/s13059-014-0550-8>
18. Yu G, Wang L-G, Han Y, He Q-Y (2012) clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS* 16:284–287. <https://doi.org/10.1089/omi.2011.0118>
19. Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. <https://doi.org/10.1186/1471-2105-9-559>
20. Alboukadel K (2020) ggpubr: “ggplot2”. Based Publication Ready Plots
21. Robin X, Turck N, Hainard A et al (2011) pROC: an open-source package for R and S + to analyze and compare ROC curves. *BMC Bioinformatics* 12:77. <https://doi.org/10.1186/1471-2105-12-77>
22. Paul B (2013) Estimating and Comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Statistics in Medicine* 5381–5397
23. William R (2020) psych: Procedures for Psychological, Psychometric, and Personality Research
24. Li T, Fu J, Zeng Z et al (2020) TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res* 48:W509–W514. <https://doi.org/10.1093/nar/gkaa407>
25. Terashima Y, Toda E, Itakura M et al (2020) Targeting FROUNT with disulfiram suppresses macrophage accumulation and its tumor-promoting properties. *Nat Commun* 11:609. <https://doi.org/10.1038/s41467-020-14338-5>

26. Li R, Serrano JC, Xing H et al (2018) Interstitial flow promotes macrophage polarization toward an M2 phenotype. *MBoC* 29:1927–1940. <https://doi.org/10.1091/mbc.E18-03-0164>
27. Qin F, Liu X, Chen J et al (2020) Anti-TGF- β attenuates tumor growth via polarization of tumor associated neutrophils towards an anti-tumor phenotype in colorectal cancer. *J Cancer* 11:2580–2592. <https://doi.org/10.7150/jca.38179>
28. Xu R, Li Y, Liu Y et al (2020) How are MCP1 and cytokines mutually regulated in cancer-related immunity? *Protein Cell* 11:881–893. <https://doi.org/10.1007/s13238-020-00739-1>
29. Nishino M, Ramaiya NH, Hatabu H, Hodi FS (2017) Monitoring immune-checkpoint blockade: response evaluation and biomarker development. *Nat Rev Clin Oncol* 14:655–668. <https://doi.org/10.1038/nrclinonc.2017.88>
30. Schönrath K, Pan W, Klein-Szanto AJ, Braunewell K-H (2011) Involvement of VILIP-1 (visinin-like protein) and opposite roles of cyclic AMP and GMP signaling in in vitro cell migration of murine skin squamous cell carcinoma. *Mol Carcinog* 50:319–333. <https://doi.org/10.1002/mc.20707>
31. Dou A-X, Feng L-L, Liu X-Q, Wang X (2012) Cyclic adenosine monophosphate involvement in low-dose cyclophosphamide-reversed immune evasion in a mouse lymphoma model. *Cell Mol Immunol* 9:482–488. <https://doi.org/10.1038/cmi.2012.34>
32. Gao Z, Zhang D, Duan Y et al (2019) A five-gene signature predicts overall survival of patients with papillary renal cell carcinoma. *PLoS ONE* 14:e0211491. <https://doi.org/10.1371/journal.pone.0211491>
33. Ye M, He Z, Dai W et al (2018) A TOP2A-derived cancer panel drives cancer progression in papillary renal cell carcinoma. *Oncol Lett*. <https://doi.org/10.3892/ol.2018.9179>
34. Wang Y, Yan K, Lin J et al (2020) Three-gene risk model in papillary renal cell carcinoma: a robust likelihood-based survival analysis. *Aging* 12:21854–21873. <https://doi.org/10.18632/aging.104001>
35. Zhang C, Qu Y, Xiao H et al (2019) LncRNA SNHG3 promotes clear cell renal cell carcinoma proliferation and migration by upregulating TOP2A. *Exp Cell Res* 384:111595. <https://doi.org/10.1016/j.yexcr.2019.111595>
36. Jiang T, Sui D, You D et al (2018) MiR-29a-5p inhibits proliferation and invasion and induces apoptosis in endometrial carcinoma via targeting TPX2. *Cell Cycle* 17:1268–1278. <https://doi.org/10.1080/15384101.2018.1475829>
37. Yu X, Davenport JW, Urtishak KA et al (2017) Genome-wide TOP2A DNA cleavage is biased toward translocated and highly transcribed loci. *Genome Res* 27:1238–1249. <https://doi.org/10.1101/gr.211615.116>
38. Huang D, Jian J, Li S et al (2019) TPX2 silencing exerts anti-tumor effects on hepatocellular carcinoma by regulating the PI3K/AKT signaling pathway. *Int J Mol Med*. <https://doi.org/10.3892/ijmm.2019.4371>
39. Sun Z, Xiao B, Jha HC et al (2014) Kaposi's Sarcoma-Associated Herpesvirus-Encoded LANA Can Induce Chromosomal Instability through Targeted Degradation of the Mitotic Checkpoint Kinase Bub1. *J Virol* 88:7367–7378. <https://doi.org/10.1128/JVI.00554-14>

40. Shih J-H, Chen H-Y, Lin S-C et al (2020) Integrative analyses of noncoding RNAs reveal the potential mechanisms augmenting tumor malignancy in lung adenocarcinoma. *Nucleic Acids Res* 48:1175–1191. <https://doi.org/10.1093/nar/gkz1149>
41. Zhong W, Fu X, Chen G et al (2016) Overexpression of BUB1B contributes to progression of prostate cancer and predicts poor outcome in patients with prostate cancer. *OTT* 2211. <https://doi.org/10.2147/OTT.S101994>
42. Mao P, Bao G, Wang Y-C et al (2020) PDZ-Binding Kinase-Dependent Transcriptional Regulation of CCNB2 Promotes Tumorigenesis and Radio-Resistance in Glioblastoma. *Translational Oncology* 13:287–294. <https://doi.org/10.1016/j.tranon.2019.09.011>
43. Kalimutho M, Sinha D, Jeffery J et al (2018) CEP55 is a determinant of cell fate during perturbed mitosis in breast cancer. *EMBO Mol Med* 10:e8566. <https://doi.org/10.15252/emmm.201708566>
44. Jia Y, Xiao Z, Gongsun X et al (2018) CEP55 promotes the proliferation, migration and invasion of esophageal squamous cell carcinoma via the PI3K/Akt pathway. *OTT Volume* 11:4221–4232. <https://doi.org/10.2147/OTT.S168861>
45. Yuan Y-J, Sun Y, Gao R et al (2020) Abnormal spindle-like microcephaly-associated protein (ASPM) contributes to the progression of Lung Squamous Cell Carcinoma (LSCC) by regulating CDK4. *J Cancer* 11:5413–5423. <https://doi.org/10.7150/jca.39760>
46. Zhou J-W, Wang H, Sun W et al (2020) ASPM is a predictor of overall survival and has therapeutic potential in endometrial cancer. *Am J Transl Res* 12:1942–1953
47. Mazzu YZ, Armenia J, Chakraborty G et al (2019) A Novel Mechanism Driving Poor-Prognosis Prostate Cancer: Overexpression of the DNA Repair Gene, Ribonucleotide Reductase Small Subunit M2 (RRM2). *Clin Cancer Res* 25:4480–4492. <https://doi.org/10.1158/1078-0432.CCR-18-4046>
48. Shahid M, Lee MY, Piplani H et al (2018) Centromere protein F (CENPF), a microtubule binding protein, modulates cancer metabolism by regulating pyruvate kinase M2 phosphorylation signaling. *Cell Cycle* 17:2802–2818. <https://doi.org/10.1080/15384101.2018.1557496>

Table

Table 1: The information of nine hub genes

Gene	Gene ID	Full name
<i>TPX2</i>	22974	<i>TPX2</i> microtubule nucleation factor
<i>TOP2A</i>	7153	DNA topoisomerase II alpha
<i>BUB1B</i>	701	<i>BUB1</i> mitotic checkpoint serine/threonine kinase B
<i>BUB1</i>	699	<i>BUB1</i> mitotic checkpoint serine/threonine kinase
<i>CEP55</i>	55156	centrosomal protein 55
<i>RRM2</i>	6241	ribonucleotide reductase regulatory subunit M2
<i>ASPM</i>	259266	assembly factor for spindle microtubules
<i>PBK</i>	55872	PDZ binding kinase
<i>CENPF</i>	1063	centromere protein F

Figures

Figure 1

The workflow of the selection process for the eligible studies in the analysis. KIRP: Kidney renal papillary cell carcinoma. CIBERSORT: an algorithm for analysis of immune cell composition of complex tissue from their gene expression. TIICs: tumor-infiltrating immune cells. DEGs: differentially expressed genes;

GEO: Gene Expression Omnibus; GSEA: Gene Set Enrichment Analysis; IHC: immunohistochemical; PPI: Protein-Protein Interaction; ROC: Receiver Operating Characteristic; TCGA: The Cancer Genome Atlas; WGCNA: Weighted Gene Co-Expression Network Analysis.

Figure 2

Identification of 9 prognostic immune cells. (A) The landscape of tumor-infiltrating immune cells. (B) Heatmap of the fraction of tumor-infiltrating immune cells. (C) The correlation coefficient between the abundance ratios of the distinct immune cells. (D) The nine prognostic-related immune cells.

Figure 3

The relationship between the nine prognostic-related immune cells and clinical traits. (A-D) The relationship between immune cells and tumor type. (E-G) The relationship between immune cells and pathological stage. (H-L) The relationship between immune cells and TNM stage. (M-O) The relationship between immune cells and overall survival.

Figure 4

Constructing high and low-risk groups and nomogram based on nine immune cells. (A) The curve of risk score. (B) Survival status of the patients. More dead patients in the high group. (C) Heatmap of the expression profiles of the nine prognostic immune cells in low- and high-risk groups. (D) Kaplan-Meier survival analysis of the low- and high-risk groups. (E) Nomogram integrated nine survival-related cells. (F) The calibration plot of the nomogram for agreement test between 1-, 3- and 5-year OS prediction and actual outcome in TCGA dataset. (G) The time-dependent ROC curves of the nomogram. TCGA: The Cancer Genome Atlas; ROC: Receiver operating characteristic.

Figure 5

Identification and functional annotation of DEGs. (A) Volcano plot of the DEGs between low- and high-risk groups. Red nodes represent the significantly up-regulated genes, and blue nodes represent the significantly down-regulated genes. (B) Top 10 of KEGG enrichment analysis of DEGs. (C) The significant GO terms of up-regulated genes. (D) The significant GO terms of down-regulated genes. (E) Top 5 of

GSEA enrichment analysis of DEGs. DEGs: differentially expressed genes; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; GSEA: gene set enrichment analysis.

Figure 6

Identification of the core modules via constructing a weighted gene co-expression network. (A) Analysis of the scale-free topology model fit index for soft threshold powers (β). (B) A cluster dendrogram was built based on the dissimilarity of the topological overlap, which presents six gene co-expression modules in these DEGs, the grey module indicates none co-expression between the genes. (C) Heatmap of the weighted gene co-expression network. (D) Heatmap of the correlation between module eigengenes and clinical traits of pRCC. (E-F) The scatterplot of GS for pathologic stage and T stage vs MM in the blue module. (G-H) The scatterplot of GS for pathologic stage and T stage vs MM in the turquoise module. DEGs: differentially expressed genes; pRCC: papillary renal cell carcinoma; GS: gene significance; MM: module membership.

Figure 7

Identification of hub genes via multiple algorithms. (A-E) Screening the top 30 genes in the PPI network of DEGs as central genes using five algorithms including Closeness, MCC, MNC, Degree, and EPC. (F) Multiple algorithms obtain a Venn diagram of common genes. PPI: protein-protein interaction; DEGs: differentially expressed genes; MCC: Maximal Clique Centrality; MNC: Maximum Neighborhood Component; EPC: Edge Percolated Component.

Figure 8

Construction of the prognostic signature based on hub genes. (A) The forest plot of the nine hub genes via univariate Cox regression analysis. (B) The forest plot of the nine hub genes multivariate Cox regression analysis. (C) Nomogram integrated the nine immune-related hub genes. (D) The calibration plot of the nomogram for agreement test between 1-, 3- and 5-year OS prediction and actual outcome in

TCGA dataset. (E) The time-dependent ROC curves of the nomogram in TCGA dataset. TCGA: The Cancer Genome Atlas; ROC: Receiver operating characteristic.

Figure 9

Validation of the hub genes in different clinical traits. (A) The expression difference of the nine hub genes between normal and tumor. (B) The relationship between the expression of the nine hub genes and pathologic stage. (C) Kaplan-Meier survival analysis of the nine hub genes. (D) Diagnostic value of the nine hub genes in different tumor types. (E) Diagnostic value of the nine hub genes in the tumor.

Figure 10

IHC of hub genes and validation of the hub genes in GEO dataset. (A-G) IHC of TPX2, TOP2A, RRM2, PBK, CEP55, CEBPF, and ASPM in renal carcinoma. (H) Independent dataset validation of the nine hub genes in GSE26574. IHC: Immunohistochemistry.

Figure 11

The correlation between the nine hub genes and immune cells. (A) The heatmap plot of the Spearman correlation coefficients between the nine hub genes and immune cells. (B) validation of the correlation between the nine hub genes and macrophages via the XCELL algorithm.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS1.csv](#)
- [TableS2.csv](#)
- [TableS3.csv](#)

- [TableS4.csv](#)
- [TableS5.csv](#)
- [TableS6.csv](#)