A Bioinformatic Analysis of Immune-Related Prognostic Genes in Clear Cell Renal Cell Carcinoma Based on TCGA and GEO Databases

Jianpeng Li¹,³,*
Jinlong Cao¹,³,*
Pan Li¹,³
Ran Deng¹,³
Zhiqiang Yao¹,³
Lijun Ying¹,³
Junqiang Tian¹,³

¹Department of Urology, The Second Hospital of Lanzhou University, Lanzhou, People’s Republic of China; ²Key Laboratory of Gansu Province for Urological Diseases, The Second Hospital of Lanzhou University, Lanzhou, People’s Republic of China; ³Clinical Center of Gansu Province for Nephron-Urology, The Second Hospital of Lanzhou University, Lanzhou, People’s Republic of China

*These authors contributed equally to this work

Background: Clear cell renal cell carcinoma (ccRCC) is a commonly occurring tumor. Through a deeper understanding of the immune regulatory mechanisms in the tumor microenvironment, immunotherapy may serve as a potential treatment for cancer patients. This study aimed at identifying the survival-related immune cells and hub genes, which could be potential targets for immunotherapy in ccRCC.

Methods: The gene expression profiles and clinical data of ccRCC patients were extracted from UCSC Xena and Gene Expression Omnibus (GEO) databases. Kaplan–Meier (KM) survival and Least Absolute Shrinkage and Selection Operator (LASSO) regression analyses were utilized to select the survival-related tumor-infiltrating immune cells. Multivariate Cox regression was utilized to develop a signature based on the tumor-infiltrating immune cells (TIICs). Based on the signature, the risk score was calculated, following which the samples were divided into high-risk and low-risk groups. Differentially expressed genes (DEGs) between the two risk groups were identified. Functional enrichment analysis was performed and cytoHubba plug-in of Cytoscape was used to identify the hub genes. Multiple datasets were utilized to validate these hub genes, including the Gene Expression Profiling Interactive Analysis (GEPIA), UALCAN, and the Human Protein Atlas (HPA), and the GEO datasets. Finally, a correlation analysis was performed to evaluate the relationship between the hub genes and TIICs.

Results: Four immune survival-related cells, including T cell CD4 memory-activated, T cell regulatory (Tregs), eosinophils, and mast cell resting were identified. Nine immune-specific hub genes were identified, which included APOE, CASR, CTLA4, CXCL8, EGF, F2, KNG1, MMP9, and IL6. Furthermore, these hub genes were significantly correlated with clinical traits and closely associated with some TIICs.

Conclusion: A total of four survival-related immune cell types and nine hub genes were found to be closely associated with ccRCC. These findings may have implications for the development of novel potential immunotherapeutic targets for ccRCC.

Keywords: ccRCC, tumor-infiltrating immune cells, hub genes, TCGA, tumor microenvironment

Introduction
Kidney cancer is one of the most commonly diagnosed tumors around the globe.¹ According to the statistics from the World Health Organization, annually, there are more than 140,000 RCC-related deaths.² ccRCC is the most typical subtype of kidney cancer and contributes to the majority of kidney cancer-related deaths.³,⁴ Until the recent past, the most widely used therapeutic agents were chemotherapy and cytokine therapy using interferon-alpha (INFα) and IL-2. However, these
therapies produced effective responses only among a limited proportion of patients and adverse side effects among others.\(^5\) Moreover, the overall survival (OS) in advanced ccRCC is less than 30%.\(^6,7\) Therefore, the development of new treatment methods with clinical applicability in ccRCC is important.

Tumor microenvironment refers to the highly heterogeneous and dynamic tumor-associated network, consisting of tumor cells and non-tumor components, including blood vessels, immune cells, adipocytes, and cancer-associated fibroblasts.\(^5,9\) Numerous previous studies show that the tumor microenvironment plays an important role in the progression and prognosis of multiple cancer types. Generally, the von Hippel Lindau (VHL) gene, a tumor suppressor, is typically inactivated in ccRCC, which results in the overexpression of HIF-2a and VEGF and promotes the progression of ccRCC. Therefore, anti-angiogenic agents (AA), targeting the VEGF pathway, such as the VEGF receptor (VEGFR) and tyrosine kinase inhibitors (TKIs), have been recently used for the therapy of patients with ccRCC.\(^10\) Moreover, the immunogenicity of ccRCC has long been acknowledged. Several reports have examined the relationship between the existence of various immune cell types and patient prognosis.\(^11-13\) Giraldo et al show that CD8\(^+\) T cell infiltration is substantially correlated with poor prognosis in patients with ccRCC, unlike in other cancer types.\(^14\) Therefore, some immunotherapies based on immune checkpoint inhibitors (ICI), such as anti-programmed death receptor 1 (anti-PD1), are used for some patients with cancer\(^5\) and produce perfect outcomes. However, targeted therapy, such as TKIs and ICI, produce a complete response in a limited number of patients and potentially exerts adverse effects among some patients with cancers.\(^15\) In several clinical studies, it has been reported that there remains a considerable proportion of patients with no response or resistance to immune checkpoint inhibitors or anti-angiogenic agents,\(^16\) implying that the assessment of several immune checkpoint genes expression alone is insufficient in immunotherapy for most malignancies due to the low accuracy of prediction.\(^17\) Roviello et al performed a meta-analysis and found that the efficacy of immune checkpoint inhibitors in PD-L1 negative expression patients is not satisfactory.\(^18\) Therefore, it is necessary to detect novel immune-related targets for immunotherapy to elevate the efficacy of treatment.

Most previous studies have examined the role of immune cells in the progression and prognosis of cancers, however, they focus on the association between the immune cells and the prognosis of patients with cancers.\(^9,10\) With the advent of gene chips and high-throughput sequencing technologies, it is possible to have an in-depth understanding of the genetic characters of the tumor immune microenvironment through bioinformatic methods. Chen et al have identified prognostic immune genes in endometrioid endometrial adenocarcinoma by analyzing the high-throughput sequencing data using bioinformatic tools.\(^20\) Tian et al have analyzed the tumor-infiltrating immune cells (TIICs) and identified prognostic immune cells in adrenocortical carcinoma using bioinformatics.\(^21\) In this study, we utilized the high-throughput sequencing data and clinical information of ccRCC specimens extracted from the public databases to conduct immune cell infiltration and survival analyses. The survival-related immune cells and immune-specific genes were selected following the survival analysis. Eventually, four categories of survival-related TIICs and nine immune-specific hub genes were screened; these could be potential immunotherapeutic targets in ccRCC therapy.

Materials and Methods

Data Extraction and Processing

The workflow of this study is shown in Figure 1. The gene expression and corresponding clinical data were downloaded from public databases, Gene Expression Omnibus (GEO, [https://www.ncbi.nlm.nih.gov/geo/](https://www.ncbi.nlm.nih.gov/geo/)) and the Cancer Genome Atlas (TCGA). In total, three ccRCC cohorts (TCGA-KIRC, GSE36895, and GSE53757) were analyzed in this study. TCGA-KIRC RNA sequencing data (count data) were downloaded from UCSC Xena ([https://xenabrowser.net/datapages/](https://xenabrowser.net/datapages/)), which consisted of 534 tumor- and 73 adjacent normal tissue samples. GSE36895 included 29 ccRCC tumor samples and 23 normal tissue samples.\(^22\) GSE53757 included 72 ccRCC tumor samples and 72 normal tissue samples.\(^23\) The information on patients from the selected ccRCC cohorts is shown in Table 1. This study design was reviewed by the ethics committee (Lanzhou University Second Hospital Ethics Committee) and deemed exempt for ethical approval.

Calculating the Abundance Ratio Matrix of Immune Cells

To assess immune infiltration in samples, we used the CIBERSORT algorithm, a tool for evaluating the immune cell composition in different organs based on their gene expression patterns.\(^24\) CIBERSORT source code and the LM22 gene signature were downloaded from the CIBERSORT website ([https://cibersort.stanford.edu/down](https://cibersort.stanford.edu/down)) for use in this study.
The abundance ratios for 22 TIICs were calculated from the gene expression profiles of ccRCC by executing the CIBERSORT source code using the R software. The matrix of abundance ratio was visualized using the “barplot” function of R software and normalized using the “scale” function in R. The correlation between these TIICs was calculated using Pearson correlation through the “cor” function and visualized using the R package “corrplot”.

Identification of Survival-Related Immune Cells and Grouping
To identify the survival-related immune cells from the 21 cell types calculated through CIBERSORT, we performed the survival analysis. The R package, “survival”, was used to select the survival-related tumor-infiltrating cells from 534 samples following the Kaplan–Meier (KM) analysis. To filter the immune cells, the LASSO regression analysis was performed to select the significant survival-related TIICs using the R package, “glmnet”. Next, the selected survival related TIICs were used to perform multivariate Cox regression analysis and the risk score for each sample was calculated using the relative ratio of TIICs multiplied by the regression coefficient. Samples were divided into high-risk and low-risk groups determined using the median score as a cut-off. The Nomogram based on multiple Cox regression for TIICs and clinical traits was constructed using the R package, “rms”. Subsequently, a calibration plot and time-dependent ROC were used to evaluate the prognostic value of the TIICs using the R packages, “rms” and “timeROC”, respectively.

Examining the Relationship Between Tumor-Infiltrating Cells and Clinical Characteristics
To understand the role of the immune cell types in affecting the clinical characteristics, we compared the immune infiltration score for different clinical parameters. The relationship...
between the abundance ratio of immune cells and clinical characteristics, including tumor or normal tissue types, tumor grade, and clinical stage were analyzed using the Wilcoxon test. Next, we plotted the immune cells for various clinical characters using the “boxplot” package in R software.

Identification of Differentially Expressed Genes and Enrichment Analysis
To explore the differences between the high-risk group and the low-risk group, we performed the analysis of differentially expressed genes (DEGs) and evaluated their functional enrichment. DEGs between the high-risk and low-risk groups were identified using the R package, “DESeq2”, with the cut-off set at $P$-value < 0.05 and |foldchange| > 1. Next, Gene ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were performed using the R packages, “clusterProfiler” and “org.Hs.eg.db”, with $P$-value < 0.05 set as the cut-off. Gene set enrichment analysis (GSEA) was performed using the GSEA function of the R package, “clusterProfiler”. The immunological signature gene set (C7 gene set) was selected as the reference for GSEA (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp).

Construction of Protein–Protein Interaction Network and Identification of Hub Genes
To analyze the correlation among DEGs at the protein level, we constructed a protein–protein interaction (PPI) network for an in-depth understanding of these genes. The PPI network for the DEGs was constructed using the “STRING” plug-in in the Cytoscape software with a mutual score greater than 0.4 set as the threshold. Next, the “cytoHubba” plug-in in Cytoscape v3.8.2 was utilized to identify the top 30 genes in the network as the central genes based on four algorithms, namely Edge Percolated Component (EPC), betweenness, closeness, and degree. The hub genes were obtained through the intersection of DEGs among the four algorithms.

Multidimensional Validation of Hub Genes
To examine the relationship between hub genes and clinical characteristics, we analyzed their expression in different clinical grades and pathological stages. The outcome was visualized using the R package, “ggpur”. The correlation between hub genes and the OS was calculated using the Log rank test using the R package, “survival”. Subsequently, a multidimensional validation was performed to reduce the false-positive rates for these hub genes using Gene Expression Profiling Interactive Analysis (GEPIA) (http://gepia2.cancer-pku.cn/), UALCAN (http://ualcan.path.uab.edu/index.html), and the Human Protein Atlas (HPA) (https://www.proteinatlas.org/). Additionally, GSE36895 and GSE53757 served as the independent external datasets for the transcript-level validation of these hub genes. Finally, univariate and multivariate Cox analyses for these hub genes were used to examine their correlation with the patient prognosis in ccRCC.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>The Information of Clinical Traits</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TCGA Dataset</td>
</tr>
<tr>
<td></td>
<td>Case</td>
</tr>
<tr>
<td>Normal</td>
<td>73</td>
</tr>
<tr>
<td>Tumor</td>
<td>534</td>
</tr>
<tr>
<td>Grade</td>
<td></td>
</tr>
<tr>
<td>G1</td>
<td>14</td>
</tr>
<tr>
<td>G2</td>
<td>229</td>
</tr>
<tr>
<td>G3</td>
<td>207</td>
</tr>
<tr>
<td>G4</td>
<td>76</td>
</tr>
<tr>
<td>Gx</td>
<td>5</td>
</tr>
<tr>
<td>Unknown</td>
<td>3</td>
</tr>
<tr>
<td>Stage</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>268</td>
</tr>
<tr>
<td>II</td>
<td>57</td>
</tr>
<tr>
<td>III</td>
<td>123</td>
</tr>
<tr>
<td>IV</td>
<td>84</td>
</tr>
<tr>
<td>Unknown</td>
<td>2</td>
</tr>
</tbody>
</table>
Relationship Between Immune Cells and Hub Genes

Herein, we performed a correlation analysis to assess the relationship between the hub genes and immune cells, which can further understand these genes in immune infiltration. The Spearman correlation between the hub genes and 22 immune cell types was calculated using the R package, “psych”. The R package, “pheateam”, was utilized to visualize the results of the correlation analysis. The Tumor Immune Estimation Resource 2.0 (TIMER2.0) (http://timer.comp-genomics.org/) tool was utilized to validate the association between hub genes and immune cells.

Statistical Analysis

In this study, groups of boxplots were analyzed using the Wilcoxon test. In the KM survival analysis, the P-value was estimated using Log rank tests, and the hazard ratio (HR) was calculated through the univariate Cox proportional hazards regression analysis. All analyses were performed in the R software version 4.0.3 (The R Foundation for Statistical Computing, 2020). All statistical tests were two-tailed; P-value < 0.05 was considered statistically significant.

Results

Infiltration Landscape of TIICs in ccRCC

The workflow of this research is shown in Figure 1. The information on these datasets is shown in Table 1. The abundance ratio matrix of 22 tumor-infiltrating cell types was calculated for 534 samples using CIBERSORT locally and is shown in Figure 2A and B. The T cells and Macrophages exhibited high infiltration in ccRCC. Moreover, we found that the abundance ratios of several immune cell types had a significant correlation with each other (Figure 2C). Mast cell resting was significantly positively correlated with T cell CD4 memory resting (Figure S1A) and Monocytes (Figure S1B); T cell regulatory (Tregs) were negatively correlated with Mast cells (Figure S1C) and T cell CD4 memory resting (Figure S1D).

Identification of Survival Related TIICs

The results of the KM survival analysis indicated that there were eight immune cell types related to the OS of ccRCC, including Dendritic cells resting, Eosinophils, Monocytes, Mast cells resting, T cell CD4 memory resting, T cell CD4 memory activated, T cell regulatory (Tregs), and Mast cells activated (Figure 2D–K). Before multivariate Cox regression analysis, these eight immune cell types were identified through the LASSO regression analysis (Figure 3A and B). Following this, four types of immune cells, including T cell regulatory (Tregs), T cell CD4 memory activated, mast cells resting, and Eosinophils, were found to be possibly related to survival. The risk scores for each sample were calculated using the coefficients of the prior multivariate Cox regression analysis, which was as follows:

Risk score = \sum_{i=1}^{4} \beta_i \times Frac_i

where Frac is the abundance ratio of selected immune cells, and \( \beta \) is the coefficient of the Cox regression.

Subsequently, based on the median risk score as the cut-off, the patients were divided into the high-risk and low-risk groups (Figure 3C). Figure 3D shows the survival status of the patients. Figure 3E shows the heatmap for the fraction of the four immune cell types. As shown, Tregs had high infiltration in the high-risk group, which was in line with the findings of survival analysis. The KM survival analysis revealed that the high-risk group had a worse OS than the low-risk group, which confirmed that this classification was reasonable (Figure 3F). Then the nomogram based on the multivariate Cox regression was constructed for the prediction of overall survival in ccRCC (Figure 3G). The calibration curve plot (Figure 3H) and time-dependent ROC analysis for the nomogram (Figure 3I) showed satisfactory discrimination and calibration of the signature, which further indicated that these four immune cell types were strongly associated with OS in ccRCC.

Relationship Between TIICs and Clinical Characteristics

We examined the relationship between immune cells and clinical traits, including tumor or normal tissue type, clinical stage, and grade. The levels of T cells CD4 memory activated, T cells gamma delta, Macrophages M2, Dendritic cells resting, Mast cells activated, and Eosinophils were not significantly different in their infiltration in TCGA dataset (Figure 4A), and the finding was relatively consistent with those in the GSE53757 dataset (Figure 4B). The relationship between immune cells and grade in TCGA dataset is shown in Figure S2A; the relationship between immune cells and stage in TCGA...
Figure 2 (A) The overview of tumor-infiltrating immune cells; (B) Heatmap of tumor-infiltrating immune cell abundance ratio; (C) The coefficient of relationship between the different specific immune cells; (D–K) The KM survival analysis for abundance ratios of different tumor-immune infiltrating cells.
dataset is shown in Figure S2B. The GSE53757 dataset was utilized to validate the association between immune cells and the clinical stage, and the findings were consistent with the results obtained in TCGA dataset (Figure S2C). The relationship between immune cells and clinical characters showed that the majority of immune cells exhibited differential infiltration in different clinical traits, in particular, several types of T cells showed significantly different infiltration across different stages and grades. These results demonstrated that the tumor immune microenvironment was associated with the progression of ccRCC.

Identification and Enrichment of DEGs
The samples were separated into low-risk and high-risk groups to identify the survival-related DEGs in ccRCC,
with the median risk score set as a cut-off value. In total, 707 DEGs were identified with $P$-value < 0.05 and $|\text{foldchange}|$ > 1 set as the threshold criteria. The detailed information of these DEGs is shown in Table S1. The volcano plot of DEGs is shown in Figure 5A. The results of GO enrichment and KEGG pathway analyses for these upregulated genes are shown in Figure 5B–E (Tables S2 and S3). Some enriched functions related to immune reactions and cytokines included T cell activation, cytokine-cytokine receptor interaction, and humoral immune response; these suggested that immune response and immune microenvironment may contribute substantially to the classification of risk group. For further confirmation of the strong correlation of DEGs with immunological characteristics, gene set enrichment analysis (GSEA) for the immunological signature (C7 gene set) was performed and the immune functions of upregulated DEGs were analyzed; A total of 574 gene sets were significantly enriched ($P < 0.05$). The top 10 gene sets are shown in Figure 5F (Table S4). These gene sets were mainly correlated with the T cell type infiltration, which implied that these DEGs were associated with immune infiltration.

**Protein–Protein Interaction Network and Hub Genes**

To further understand these DEGs, we constructed a PPI network using the "STRING" app in Cytoscape software. From the four algorithms in the cytoHubba plug-in, the top 30 DEGs in each algorithm were used to create the protein interaction network (Figure 6A–D). These genes in the protein interaction network played essential roles in the occurrence and development of the disease. The intersection
of genes obtained from the four algorithms consisted of nine hub genes (Figure 6E), including APOE, CASR, CTLA4, CXCL8, EGF, F2, IL6, KNG1, and MMP9. The information of these nine hub genes is summarized in Table S5. The detailed results of the four algorithms are shown in Table S6. The hub genes consisted of several immune checkpoint inhibitor genes and cytokine genes, such as CTLA4, CXCL8, and IL6, which suggested that the method of...
obtaining the immune-related hub genes was reasonable and these hub genes warranted further evaluation.

**Validation of Hub Genes in Public Datasets**

To further verify the correlation between the hub genes and clinical characteristics, their expressions according to different clinical traits were analyzed. **CASR**, **CTLA4**, **EGF**, **F2**, **KNG1**, and **MMP9** were significantly differentially expressed between normal and cancer tissue types (Figure 7A), which was validated in both the GEO datasets, GSE36895 and GSE53757 (Figure 7B and Figure S3C). The expressions of hub genes in GSE36895 were consistent with the results obtained in TCGA cohort. The expression levels of **APOE**, **CTLA4**, **F2**, **IL6**, and **MMP9** increased with the increase in clinical grade and pathological stage (Figure S3A and B), and this was validated in the GSE55375 dataset (Figure S3D). The Kaplan–Meier survival analysis showed that five hub genes, **APOE**, **CTLA4**, **F2**, **IL6**, and **MMP9**, were significantly associated with the OS (Figure 7C). GEPIA indicated that the expressions of these hub genes at the genomic level were consistent with our results (Figure S4). The Human Protein Atlas showed immunohistochemically verified expressions of **APOE**, **CXCL8**, **F2**, **IL6**, and **KNG1** in renal carcinoma (Figure 7D). UALCAN analysis for these hub genes at the protein level is shown in Figure S5, and the findings were consistent at the mRNA level. These results indicated that hub genes were associated with clinical characters and prognosis both at the mRNA and protein levels, which suggested that these genes contributed to the development or progression of ccRCC.

**Validation of Hub Genes in TCGA Dataset Using the Prognostic Signature**

To verify the relationship between these candidate genes and OS in patients with ccRCC, both univariate and multivariate Cox analyses were performed. The univariate Cox regression analysis showed that all gene expressions were significantly correlated with the OS of patients with ccRCC, except for two genes, **EGF** and **KNG1** (Figure 8A). The multivariate Cox regression analysis for the seven genes showed that these genes were correlated with OS, and **CASR** \( P < 0.05, \text{HR} = 0.9322 \) and **CTLA4**
Figure 7 The expression of nine hub genes between normal tissue and tumor tissues in TCGA dataset (A) and GSE36895 dataset (B). (C) KM survival analysis of the nine hub genes. The expression of CTLA4, CXCL8, F2, IL6, and MMP9 is significantly correlated with overall survival. (D) Immunohistochemistry (IHC) of five genes in renal cell carcinoma and normal kidney tissue. ****P < 0.0001; *P < 0.05.

**Abbreviation:** ns, not significant.
Figure 8 The results of prognostic signature based on hub genes and clinical traits. (A) The forest plot of univariate Cox regression analysis for hub genes. (B) The forest plot of multivariate Cox regression analysis for hub genes. (C) Nomogram integrated nine immune-related genes and clinical traits. (D) The calibration plot indicates an agreement test between predictions of OS for 1, 3, and 5 years and actual OS outcomes in the TCGA dataset. (E–G) The time-dependent ROC curves for the nomogram and clinical trait model in TCGA dataset. (H–J) Decision curve analysis for the prediction of OS in 1-year, 3-year, and 5-year using the nomogram. The blue line indicates that there are no high-risk patients. The green line indicates that all patients are considered to be high-risk. The red line indicates that the nomogram can provide higher net benefits for accurately forecasting patient survival. **P < 0.01, *P < 0.05.
(P < 0.01, HR = 1.1692) were independent prognostic factors for ccRCC patients (Figure 8B). The nomogram for the candidate genes and clinical traits, based on the multivariate Cox regression, was developed for prognostic prediction of ccRCC (Figure 8C), and the calibration curve showed strong discrimination and calibration power of the model (Figure 8D). Time-dependent ROC analysis demonstrated that the AUC values of the nomogram were 0.853, 0.806, and 0.777 for 1-year, 3-year, and 5-year, respectively (Figure 8E–G); this indicated that the nomogram based on genes and clinical traits had a higher accuracy than the stage- and grade-based models. Decision curve analysis showed that the nomogram based on genes and clinical traits had greater benefits than the stage- and grade-based models when the threshold probabilities were set greater than 0.2 (Figure 8H–J). The evaluation of the Nomogram based on hub genes showed that they were well capable of predicting the prognosis in ccRCC, which indicated that these hub genes had a prognostic role in the ccRCC and deserve further analysis.

Relationship Between TIICs and Hub Genes

The relationship between hub genes and the TIICs is shown in Figure 9A, which demonstrated that the proportion of TIICs was linked to several hub genes. For instance, the expression of CTLA4 was positively related to the infiltration level of T cell CD8 and T cell follicular helper, while negatively related to the abundance ratio of mast cell resting. The expression of MMP9 was positively related to the proportion of Macrophages M0. These findings suggested that the hub genes may play a significant role in the function of the TIICs and are potential immunotherapy targets. Next, TIMER 2.0 was used to validate the correlation between hub genes and immune cells based on the XCELL algorithm (Figure 9B). The correlation estimated by XCELL was consistent with the result obtained using the CIBERSORT algorithm. The relationship between immune cells and hub genes in the GSE53757 dataset was evaluated and it was consistent with the findings in TCGA dataset (Figure 8E).

Discussion

ccRCC is the most common type of kidney renal cell carcinoma and accounts for the majority of kidney cancer-related deaths. In recent years, several studies show that immunotherapy is beneficial for patients with ccRCC, however, the OS of patients with ccRCC remains unsatisfactory. Therefore, pursuing novel immune-related genes as therapeutic targets is critical for the development of effective therapy for ccRCC. The purpose of this study was to identify the immune cell types and immune-related genes correlated with the prognosis of ccRCC. These findings provided potential novel immunotherapeutic targets for ccRCC.

In this study, four hub immune cells were identified, including T cells regulatory (Tregs), T cells CD4 memory activated, Mast cells resting, and Eosinophils. Tregs are responsible for regulating the basic balance of the adaptive immune response, adjusting function, and maintaining homeostasis in response to diverse tissue contexts and immunological circumstances. Giraldo et al show that ICOS+ Tregs can be used as a prognostic marker in localized ccRCC. Some previous studies also show that CD4+ T cells are upregulated and are significantly correlated with OS in RCC. These previous results were consistent with our findings as a higher abundance ratio of Treg cells was identified in patients with worse OS. Nakanishi et al showed that mast cells played an important role in tumor growth, progression, and malignant aggressiveness of RCC tissues, which was consistent with our findings that the larger quantity of mast cells resting was associated with a better prognosis. These studies show, to some extent, that these four immune cell subgroups may be predictors of ccRCC prognosis. However, the abundance ratio of eosinophils was low in the majority of samples, and therefore the conclusion for eosinophils should be analyzed with caution.

Functional enrichment was performed for the DEGs between the high-risk and low-risk groups. The biological processes associated with upregulated DEGs were enriched in T cell activation and humoral immune response. These results implied that immune-related upregulated genes may be associated with the function of several T cell subtypes, which was also consistent with the findings in risk grouping based on immune cells. KEGG enrichment analysis demonstrated that most upregulated DEGs were enriched in cytokine-cytokine receptor interaction, chemokine signaling pathway, IL-17 signaling pathway, T cell receptor signaling pathway, and NF-kB signaling pathway. According to the findings, these signaling pathways played a crucial role in ccRCC. Previous studies have shown that cytokine-cytokine receptor interaction, chemokine signaling pathway, and IL-17 signaling pathway play significant roles in the proliferation,
epithelial-mesenchymal transition, immune evasion, and metastases of several cancer types. Therefore, the regulation of these DEGs and tumor microenvironment could impact the tumor progression and prognosis in ccRCC.

Nine hub genes were selected through four algorithms in the cytoHubba plug-in in Cytoscape, including IL6, EGF, APOE, KNG1, MMP9, F2, CTLA4, CASR, and CXCL8. In the early stages of the immune responses, CTLA4 is expressed on the surface of CD4+ and CD8+ cells. CTLA4 binding reduces IL2 production and T cell proliferation. Some studies show that tumor cells can use the pathways regulated by CTLA4, including checkpoint protein signaling, to create a microenvironment that permits tumor cell growth. Anti-CTLA4 antibodies (ipilimumab, tremelimumab) are now being tested in different clinical trials as putative immune checkpoint inhibitors. In 2007, the first clinical study on anti-CTLA4 agents in ccRCC was undertaken, which demonstrated the efficacy of CTLA-4 inhibition in some patients with metastatic RCC. Thus, to a certain extent, the
analysis of identification of hub genes is justified and other hub genes are also valuable to an in-depth investigation of the tumor immune microenvironment. **IL6** is a cytokine produced by the immune and some tumor cells and is related to tumor proliferation and metastasis.\(^59,60\) These results were consistent with our findings as **IL6** was highly expressed in grade 4 and stage 4. **IL6** is a key factor in the differentiation of CD4+ T cell subsets and is required for the formation of T follicular helper cells.\(^61\) The IL6 receptor antibody, tocilizumab, is used for the treatment of rheumatoid arthritis.\(^35\) These studies indicate that **IL6** correlates with immune cells and processes, which were also validated by IHC of renal cancer samples. The Epidermal growth factor, **EGF**, belongs to the epidermal growth factor superfamily. This protein is a robust mitogenic factor that plays a significant role in the growth, proliferation, and differentiation of numerous cell types.\(^52\) Polimeno et al show that **EGF** is a marker of host immunity in patients with renal cell carcinoma,\(^50\) which was also consistent with our findings. **APOE** is essential for the normal catabolism of triglyceride-rich lipoprotein constituents. Ostendorf et al demonstrate that **APOE** genotype is a biomarker for the outcome and therapeutic response in patients with melanoma.\(^63\) Tavazoie et al show that the **APOE**-associated axis regulates innate immune response suppression and plays an important role in enhancing the efficacy of cancer immunotherapy.\(^64\)

Taken together, **APOE** influences the prognosis of cancer patients and correlates with immune therapy. However, there have been only a few studies on the association of **APOE** as a novel therapeutic target and renal cell carcinoma. A previous study also shows that **MMP9** and **CASR** are hub genes significantly correlated with OS in ccRCC.\(^6\) In summary, among these hub genes, several genes have already been investigated in clinical trials, for example **CTLA4**, indicating these genes are associated with cancer therapy. Therefore, some of hub genes were novel genes that deserve in-depth investigation in the tumor immune environment.

However, the present study has certain limitations. First, the data of high-throughput sequencing of samples were downloaded from UCSC Xena, which is a repository for TCGA. Immune infiltration investigations were not considered at the time that the samples were acquired. Consequently, patients with immune system illnesses were enrolled, which may have resulted in selection bias at the beginning of the trial. Second, these survival-related immune cells and hub genes were identified from TCGA database, where the majority of patients are Americans or white. Considering the genetic heterogeneity, these immune cells and hub genes need to be validated in more public databases. Finally, this study lacks relevant in vivo and in vitro experiments to validate the functional roles of the identified hub genes.

**Conclusion**

In conclusion, four survival-related immune cell types and nine immune-related hub genes were identified using the CIBERSORT algorithm and regression analyses. Furthermore, these candidate hub genes were validated in clinical datasets and may also have implications as novel potential targets for ccRCC immunotherapy.

**Abbreviations**

ccRCC, clear cell renal cell carcinoma; KIRC, kidney renal clear cell carcinoma; KM, Kaplan–Meier; TIICs, tumor-infiltrating immune cells; IC, immune checkpoint; LASSO, Least Absolute Shrinkage and Selection Operator; HR, hazard ratio; DEGs, differentially expressed genes; GEPIA, Gene Expression Profiling Interactive Analysis; HPA, Human Protein Atlas; OS, overall survival; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

**Data Sharing Statement**

All high-throughput sequencing data and clinical data can be obtained from UCSC Xena (https://xenabrowser.net/data pages/). The datasets, GSE36895 and GSE53757, can be obtained from the Gene Expression Omnibus (http://www. ncbi.nlm.nih.gov/geo/) database. The Immunohistochemistry (IHC) images of hub genes can be obtained from Human Protein Atlas (HPA) (https://www.proteinatlas.org/). The protein expression of genes can be obtained from UALCAN (http://ualcan.path.uab.edu/index.html).

**Ethics Approval**

Not applicable. All datasets used in this work are obtained from public databases and are freely available. This work did not include any experiments on humans or animals. The patients involved in the public databases have been enrolled after ethical approval. Additionally, this study was reviewed by the ethics committee (Lanzhou University Second Hospital Ethics Committee) and deemed exempt for ethical approval.

**Acknowledgments**

The information for this study was obtained from UCSC Xena, Gene Expression Omnibus (GEO). We are grateful
Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This study was supported by grants from the (1) Science and Technology Project of Chengguan District, Lanzhou City, Gansu Province Science and Technology Bureau (Project number: 2017KJGG0052); (2) the Fundamental Research Funds for the Central Universities (Project number: 561219007); (3) Cuiying Graduate Supervisor Applicant Training Program of Lanzhou University Second Hospital (Project number: 201704); (4) Cuiying Scientific and Technological Innovation Program of Lanzhou University Second Hospital (Project number: CY2017-BJ16); (5) Lanzhou City Talent Innovation and Entrepreneurship Project (Project number: 2019-RC-37), (6) Gansu Provincial Department of Education: Outstanding Postgraduate “Innovation Star” Project (Project number: 2021CXZX-154), and (7) Gansu Health Industry Research Project (Project number: GSWSKY2017-10).

Disclosure

The authors report no potential conflicts of interest.

References

