The Prognosis-related Genes Participate in Immunotherapy of Renal Clear Cell Carcinoma Maybe Through Targeting Dendritic Cells

Guodong Fang

Shaanxi Provincial People's Hospital

Xudan Wang (✉ wangxudan_1314@petalmail.com)

Xijing Hospital

Research Article

Keywords: immunotherapy, prognosis, risk score, dendritic cell, clear cell renal cell carcinoma

Posted Date: December 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1164876/v1>

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

[Read Full License](#)

Abstract

Tumor immunotherapy has become one of the hotspot methods of tumor treatment. This study aims to screen and analyze the role of prognostic genes in the diagnosis, prognosis and immunotherapy of clear cell renal cell carcinoma (ccRCC) by bioinformatics methods. Based on the Gene Expression Omnibus Datasets and The Cancer Genome Atlas (TCGA) Datasets, we screened out the nine differentially expressed genes (TYROBP, APOC1, CSTA, LY96, LAPTM5, CD300A, ALOX5, C1QA, C1QB) associated with clinical traits and prognosis by "WGCNA" package in R and Kaplan-Meier Plotter. The ROC curves showed that the combination of these genes had high diagnostic value (AUC=0.990). The COX regression analysis was used to construct the risk signature, and the risk score calculated by the 9 genes was an independent prognostic factor that could predict survival probability for patients at 1 year, 3 years, 5 years. The high-risk group had lower Tumor Immune Dysfunction and Exclusion score and higher tumor mutation burden and PD-1 levels. The immune checkpoint blockade response rates in the high- and low-risk groups were 49.25% and 24.72% ($p \leq 0.001$). The immune infiltration and Pearson correlation analyses suggested that the prognostic genes were negatively correlated with activated dendritic cells in the low-risk group, rather than in the high-risk group. In conclusion, prognostic-related genes have high diagnostic value and can guide prognostic evaluation and immunotherapy for ccRCC, and this process needs to be mediated by the activated dendritic cells.

Introduction

The clear cell renal cell carcinoma (ccRCC) is a malignant tumor of the urinary system, accounting for about 75% of renal cancers[1]. The immunotherapy drugs for ccRCC have made great progress from the initial interleukin and interferon to the emergence of immune checkpoint inhibitors (ICIs) [2, 3]. Immune checkpoint (IC) is an inhibitory pathway of the immune system, its main function is to prevent the overactivation of the immune system from harming normal cells[4]. IC is closely related to immune escape of tumor, and inhibition of key molecules of IC has become the main means of tumor immunotherapy[5].

Common IC molecules include programmed death receptor 1(PD-1)/ programmed cell death 1 ligand 1(PD-L1) and cytotoxic T lymphocytes (CTLA4), in addition, Mismatch repair (MMR)/Microsatellite instability (MSI) and TMB are also used for tumor immunotherapy evaluation [6]. The expression of these indicators varies greatly in different tumors, which seriously restricts the application of immunotherapy. A study comparing the positive rates of PD-L1, TMB and MSI in various solid tumors involving 155 kidney cancer samples found that the positive rates of TMB and MSI in kidney cancer were only 1/155, while the positive rates of PD-L1 were higher (46/155), while the negative rates of all three were as high as 108/155[7]. Another TMB study involving 244 patients with renal transitional cell carcinoma found that high TMB (>20 mutations/Mb) was only 6.3%[8]. Therefore, the lack of molecular markers that can effectively predict treatment outcomes remains a major challenge for clinicians. TIDE scores, a new scoring system that integrates T cell dysfunction and cell rejection markers to simulate tumor immune escape, can predict ICB response rates and has been shown to be superior to PD-L1 and TMB in

predicting immunotherapy accuracy for malignant melanoma[9]. However, there are no clinical studies of TIDE in ccRCC to date. The emergence of a new generation of genetic detection technology opens the door to big data research, and various new molecular markers are constantly emerging. A clinical study conducted genetic testing on 212 high-risk ccRCC samples. Sixteen prognostic genes were selected to successfully construct a recurrence RSMI for ccRCC, and the efficacy of sunitinib was evaluated and predicted in patients with different risk levels. The potential application of this study is related to FDA approval of sunitinib as an adjunctive therapy for patients at high risk of kidney cancer recurrence[10].

In this study, the genetic data and clinical data of ccRCC were downloaded from GEO database and TCGA database, the differentially expressed genes (DEGs) related to prognosis were screened, and the risk scoring model was constructed. The risk stratification of 533 ccRCC samples in TCGA database was carried out. COX regression was used to analyze the relationship between different stratified samples and existing immunotherapy evaluation indexes (PD-1/PD-L1, CTLA4, TMB, MSI, TIDE and ICB response rate), to explore the effect and mechanism of risk score on immunotherapy, and to study the diagnostic value of 9 differential genes in ccRCC by ROC curve.

Data And Methods

2.1 Datasets acquisition

The three gene microarrays with ccRCC (GSE14994, GSE15641, and GSE53757) were obtained from the Gene Expression Omnibus (GEO) Datasets (<https://www.ncbi.nlm.nih.gov/gds/>). Of which GSE14994 included the 90 ccRCC samples and 48 normal renal cortex samples, GSE15641 included the 32 ccRCC samples and 23 normal kidney samples, and GSE53757 included the 72 ccRCC samples and 72 matched normal kidney samples. Besides, the transcriptome data and clinical information of ccRCC were downloaded from The Cancer Genome Atlas (TCGA) Datasets, including 534 tumor samples and 72 normal samples.

2.2 The differentially expressed genes (DEGs) analysis

The GEO2R was used to analyze the DEGs of the three gene microarrays based on a threshold value with $|\log_2 \text{fold change (FC)}| > 1$ and $p\text{-value} < 0.05$. The DEGs were visualized by ggplot2 package (v3.3.3) in R. The intersection genes of each dataset were analyzed using R package “venn”.

2.3 The weighted gene co-expression network analysis (WGCNA)

The WGCNA was analyzed by R package “WGCNA” (v1.69) in tumor samples from TCGA dataset. The clinical traits, such as age, sex, stage and grade, were included in the WGCNA analysis to screen the module that was most closely related to these traits. Then, expression of the genes assigned into the module were visualized by ggplot2 package between tumor samples and normal samples from the TCGA datasets.

2.4 Functional enrichment analysis

Functional enrichment analysis (GO and KEGG analyses) of module genes was conducted by the DAVID online software (<https://david.ncifcrf.gov/>) and visualized by ggplot2 package in R. The correlation of genes and pathway (GO-BP) was analyzed and showed by the ClueGO plug-in of the Cytoscape software (v3.7.2).

2.5 Univariate and multivariate Cox regression analyses

Kaplan-Meier (K-M) Plotter (<http://kmplot.com/analysis/>) was applied to further screen the genes related to the overall survival (OS) of ccRCC, excluding the genes with p-value > 0.05 and getting 9 prognosis-related genes. These prognosis-related genes were included into the univariate and multivariate Cox regression analyses in R package “survival” to obtain regression coefficient of each gene. In order to avoid missing some important factors, we will not do screening at this step. The 9 genes were used to construct the risk signature, and risk scores were calculated by Linear Model and predict.coxph function based on gene expression values experiencing regression coefficient weighting [11]. The risk score formula was:

$$\text{Risk score} = \sum_{i=1}^N \text{Exp}_i \times \beta_i$$

Note: Exp: the expression value of each gene in the sample, β : the regression coefficient of the multivariate Cox regression analysis for each gene.

On the basis of the formula, risk score of each patient from the TCGA datasets was calculated. Then, according to median of the risk score, the 533 samples (excluding a sample with incomplete clinical information) were divided into the high-risk group (> the median, n=266) and low-risk group (< the median, n=267). The K-M curve and receiver operating characteristic (ROC) curve were visualized by survminer package and survivalROC package in R.

Besides, the risk score and clinical traits including age, sex, stage and grade were taken into the multivariate Cox regression analysis to evaluate the independent prognostic factor with a threshold value of p-value < 0.05. The nomogram (Cox) was constructed by rms package (v5.1-4) in R to predict the patient's survival probability with 1, 3 and 5 year. The consistency of nomogram (Cox) was assessed by the calibration curves. The nomogram (Logistic) was constructed by logistic regression model in R to predict the patient's dead risk.

2.6 Evaluation of the accuracy of diagnosis and prognosis prediction

R package “pROC” was used to analyze diagnostic accuracy of each prognosis-related gene, and visualization employed ROC curve with “ggplot2” package. The logistics model was constructed by glm function to evaluate diagnostic accuracy of all prognosis-related genes. Besides, the accuracy of prognosis prediction with each prognosis-related gene were analyzed by R package “timeROC”, and the

prognostic prediction accuracy of the combination of these genes was evaluated by constructing a logistics model.

2.7 Immunotherapy evaluation

Mutation data of ccRCC patients was downloaded from the TCGA datasets, the values of tumor mutation burden (TMB) of each sample were calculated by R package “TCGAmutations”, and the results were visualized by “ggplot2” package.

According to the transcriptome data of ccRCC patients from the TCGA datasets, Tumor Immune Dysfunction and Exclusion (TIDE) of each sample were analyzed by the TIDE algorithm (<http://tide.dfci.harvard.edu/>) [9] and visualized by “ggplot2” package.

The immunophenoscore (IPS) is an immunotherapy evaluation index that only uses the expression of PD-L1 of tumor-related immune cells (lymphocytes, macrophages, etc.) as an evaluation index to distinguish benefiting cohorts [12]. The IPS data of ccRCC patients was downloaded from The Cancer Immunome Atlas (TCIA, <https://tcia.at/home>) datasets and visualized by “ggplot2” package.

2.8 Immune infiltration analysis

To further explore the mechanism of prognosis-related genes, the ESTIMATE algorithm (<https://bioinformatics.mdanderson.org/estimate/>) was employed to analyze the relationship between the risk score and the Stromal and Immune cells. The tumor-infiltrating immune-cell fraction of each ccRCC patient and normal patient was calculated by the CIBERSORT algorithm(<https://cibersort.stanford.edu/index.php>) and visualized by “ggplot2” package. Excluding the immune cells that were not significantly different between ccRCC samples and normal samples, the correlation of the prognosis-related genes and the differential immune cells were calculated by the Pearson correlation analysis and visualized by R package “ggcorrplot”. The molecular markers of the activated dendritic cells were downloaded from the TCIA datasets, the correlation of the markers and the prognosis-related genes in expression was calculated by the Pearson correlation analysis and visualized by the Cytoscape software.

2.9 Statistical analysis

All statistics were analyzed by R software (v4.1.0). The statistical significance of continuous variables was analyzed by the Wilcoxon rank sum test. The correlation was analyzed by the Pearson correlation analysis. A two-sided p-value < 0.05 was considered as statistically significant.

Results

3.1 Screening of the DEGs in ccRCC

According to a threshold value with $|\log_2 \text{FC}| > 1$ and $p\text{-value} < 0.05$, we obtained 458, 547 and 1289 DEGs between ccRCC samples and normal samples from GSE14994 (Figure 1a), GSE15641 (Figure 1b) and GSE53757 (Figure 1c), respectively. Then, we used the venn diagram to intersect the down-regulated DEGs and up-regulated DEGs of the three gene microarrays, obtaining 51 common down-regulated DEGs (Figure 1d) and 86 common up-regulated DEGs (Figure 1e). Further, we intersected these genes with those in the TCGA database, obtaining 47 down-regulated DEGs and 81 up-regulated DEGs (Figure 1f). These 128 DEGs were shared by the three gene microarrays and the TCGA database.

3.2 Identification of clinical traits-related genes

A gene co-expression network of above-mentioned 128 genes was constructed by the R package “WGCNA” to identify clinical traits-related genes. After excluding samples with incomplete clinical traits, the 533 samples from TCGA database were included into the gene co-expression network. A sample clustering tree were shown as Figure 2a. In this study, we chose the soft-threshold $b = 9$ (scale free $R^2 = 0.8$) to construct a scale-free network (Figure 2b). The four modules were identified according to dynamic tree clipping and average hierarchical clustering (Figure 2c). The Pearson correlation analysis of the modules and the clinical traits suggested that the brown module was the most closely related to clinical stage and grade (Figure 2d), so this module was considered as a clinical traits-related module, which included 20 gene. These genes were all up-regulated in the ccRCC tissue samples from the TCGA database (Figure 2e). The genes included in each module were presented in Table S1.

3.3 Functional enrichment analysis of clinical traits-related genes

The gene ontology (GO) analysis mainly included the cell component (CC), biological process (BP) and molecular function (MF)) analyses. The CC analysis of the 20 clinical traits-related genes showed that these genes were mainly involved in the composition of multiple membranes, such as Plasma membrane, Endocytic vesicle membrane, Extracellular exosome, and Lysosomal membrane. The MF analysis showed that these genes mainly acted as binding proteins, including Peptide antigen binding, MHC class II receptor activity, and Cysteine-type endopeptidase inhibitor activity. The BP analysis showed that these genes were mainly enriched into the immune-related pathways, such as T cell receptor signaling pathway, Immune response, Innate immune response, and Regulation of immune response (Figure 3a). Besides, the Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis suggested that these genes were primarily involved in a number of infection and inflammatory pathways (Figure 3b). Further, the analysis of relationship of these genes and pathways suggested that 6 of 20 clinical traits-related genes were closely associated with the immune-related pathways, like regulation of B cell proliferation, macrophage activation, and regulation of myeloid leukocyte mediated immunity, but also in the inflammation-related pathways, like leukocyte activation involved in inflammatory response, and neuroinflammatory response (Figure 3c). In follow-up analysis, we mainly focus on immune-related pathways.

3.4 Identification of hub prognosis-related genes

Kaplan-Meier survival analysis was performed to further identify the genes closely related to OS of ccRCC. According to a threshold value of $p\text{-value} < 0.05$, we obtained 9 prognostic-related genes, which were ALOX5, APOC1, C1QA, C1QB, CD300A, CSTA, LAPTM5, LY96, and TYROBP (Figure 3d).

3.5 Construction of risk signature

To further explore the predictive value of the 9 prognosis-related genes, these genes were included into the univariate and multivariate Cox regression analyses (Figure 4a-b). In order to avoid missing some important factors, this step did not exclude genes, just to obtain the regression coefficient of each gene. Then, according to the regression coefficient and expression data of each gene, we calculated the risk score of each sample (Table S2). And based on the median value of the risk score, 533 samples were divided into the high-risk group and low-risk group. We found that the number of deaths in high-risk group was higher than that in low-risk group, and the 9 prognosis-related genes were all high-expressed in high-risk group (Figure 4c). Besides, risk score had higher accuracy in prognostic prediction of the ccRCC patients ($AUC_{1\text{-year}}=0.69$, $AUC_{3\text{-year}}=0.68$, $AUC_{5\text{-year}}=0.66$, Figure 4d). The patients in high-risk group had a poorer prognosis than who in low-risk group ($p < 0.0001$, $HR=2.72$, $95\%CI=1.97-3.75$, Figure 4e). These results suggested that the risk signature constructed by the 9 prognosis-related genes had better value of prognostic prediction.

3.6 Relationship of risk score and clinical traits

We employed the multivariate Cox regression analysis to estimate the relationship of risk score and clinical traits. Age, Sex, Stage and Grade were included into the analysis. We found that Age, Stage, Grade and risk score were significantly associated with unfavorable prognosis of ccRCC patients ($p < 0.001$, Figure 5a), suggesting that risk score may use as an independent prognostic factor for ccRCC patients. Then, we used these significant factors to construct nomograms based on Cox regression analysis or Logistic regression analysis. The greater the value of these four factors, the smaller the probability of survival for patients 1, 3 and 5 years (Figure 5b). Similarly, the greater the value of these four factors, the greater the dead risk (Figure 5c). The calibration curve was used to evaluate the consistency of the nomogram. The calibration curves for 1, 3 and 5 years were almost diagonal, indicating that the nomogram was reliable (Figure 5d-f). These results suggested that risk score could not only be used as an independent prognostic factor for the patients, but also could predict the survival probability and dead risk of ccRCC patients.

3.7 The diagnostic and prognostic value of 9 prognosis-related genes

ROC curves were used to evaluate the diagnostic and prognostic value of 9 prognosis-related genes. These genes had excellent diagnostic value with $AUC_{TYROBP}=0.963$, $AUC_{APOC1}=0.984$, $AUC_{CSTA}=0.928$, $AUC_{LY96}=0.877$, $AUC_{LAPTM5}=0.962$, $AUC_{CD300A}=0.963$, $AUC_{ALOX5}=0.877$, $AUC_{C1QB}=0.953$, and $AUC_{C1QA}=0.944$ (Figure 6a). Then, we used the 9 prognosis-related genes to build a Logistic regression

model to evaluate the diagnostic value of their combination. Results indicated that the diagnostic value of the combination was superior to that of gene alone ($AUC_{model}=0.990$, Figure 6b).

Besides, we analyzed the value of the 9 prognosis-related genes in prognostic prediction of ccRCC patients. The $AUC_{1,3,5\text{ years}}$ respectively were: $AUC_{TYROBP} > 0.572$, $AUC_{APOC1} > 0.571$, $AUC_{CSTA} > 0.589$, $AUC_{LY96} > 0.616$, $AUC_{LAPTM5} > 0.536$, $AUC_{CD300A} > 0.555$, $AUC_{ALOX5} > 0.594$, $AUC_{C1QB} > 0.558$, and $AUC_{C1QA} > 0.570$ (Figure 6c-k). Similarly, the prognostic prediction value of the combination was superior to that of gene alone ($AUC_{model} > 0.653$, Figure 6l). These results suggested that combination of the 9 prognosis-related genes possessed higher diagnostic and prognostic value, especially diagnostic value.

3.8 The relationship between risk score and immunotherapy

According to previous functional enrichment analysis, these genes may be involved in immune regulation. Thus, we firstly explored the relationship of risk score and immunotherapy, and found that the response ratio to immune checkpoint blocker (ICB) in high-risk group was significantly higher than that in low-risk group (49.25% vs 24.72%, $c^2 = 34.409$, $p \leq 0.001$, Table 1). Similarly, the TMB score ($p = 0.044$, Figure 7a), IFNG ($p = 1.2e-06$, Figure 7e), CD8 ($p = 6.5e-08$, Figure 7f), TAM_M2 ($p = 1.1e-08$, Figure 7g), and CAF ($p = 5.6e-08$, Figure 7h) in high-risk group was significantly higher than that in low-risk group. The TIDE score in high-risk group was significantly lower than that in low-risk group ($p = 0.024$, Figure 7b). PD-L1 ($p = 0.13$, Figure 7c), MSI_Exp_Sig ($p = 0.15$, Figure 7d), and MDSC ($p = 0.15$, Figure 7i) were not significantly different between the two groups. These results suggested that the patients of high-risk group might be more likely to benefit from immunotherapy.

Table 1 The response ratio to immune checkpoint blocker

	High-risk group		Low-risk group		χ^2	P
	n	%	n	%		
Total	266		267		34.409	≤ 0.001
R	131	49.25	66	24.72		
NR	135	50.75	201	75.28		

Note: R: response. NR: non-response.

Besides, IPS score also showed that IPS scores with PD1_pos (both CTLA4_neg and CTLA4_pos) in high-risk group were all significantly higher than that in low-risk group (CTLA4_neg_PD1_pos: 7.84 vs 7.70, $p = 0.045$, Figure 7l; CTLA4_pos_PD1_pos: 7.42 vs 7.23, $p = 0.018$, Figure 7n). But, total IPS score ($p = 0.3$, Figure 7j), and IPS scores with PD1_neg [both CTLA4_neg ($p = 0.24$, Figure 7k) and CTLA4_pos ($p = 0.64$, Figure 7m)] had no significant difference between the two groups. These results further suggested that the risk score, especially high-risk group, might be associated with PD-1-related immunotherapy.

3.9 Exploration of the mechanism of 9 prognostic-related genes

To further explore and illustrate the mechanism of 9 prognostic-related genes on immune regulation, we firstly analyze the relationship between these genes and the immune microenvironment. Estimate analysis indicated that the Immune score ($p = 9.9\text{e-}16$, Figure 8a), Stromal score ($p = 9.6\text{e-}09$, Figure 8b) and Estimate score ($p = 8.4\text{e-}16$, Figure 8c) of high-risk group were significantly increased compared with low-risk group. Besides, further functional enrichment analysis of 9 prognostic-related genes showed that 5 of 9 genes were related to defense response, of which 3 genes were related to innate immune response, and of which 2 genes were related to B cell mediated immunity (Figure 8d, Table S3). These results indicated that the 9 prognostic-related genes were indeed involved in immune regulation.

To further explore the relationship between these genes and immune cells, we first used the CIBERSORT algorithm to analyze the fractions of 22 immune cells in the high- and low-risk groups (Figure 8e). Excluding immune cells without significant difference between the two groups, we calculated the Pearson correlation between 9 prognostic-related genes and significant immune cells respectively in the high-risk group and low-risk group. Results indicated that that in the high-risk group, the activated dendritic cells were only significantly negatively correlated with 3 genes (Figure 8f), while in the low-risk group, the activated dendritic cells were significantly negatively correlated with all 9 genes (Figure 8g). This suggested that the immune regulation of 9 prognostic-related genes might be related to activated dendritic cells.

To further analyze the correlation of 9 prognostic-related genes and the activated dendritic cells, we downloaded the molecular markers of the activated dendritic cells and analyzed the correlation of the markers and the prognosis-related genes in expression respectively in the high-risk group and low-risk group. Consistent with the above results, there were more markers closely related to 9 genes in the low-risk group (Figure 8i, Table S5) than in the high-risk group (Figure 8h, Table S4). These results further suggested that the 9 prognostic-related genes seemed to be closely related to activated dendritic cells in the low-risk group, rather than in the high-risk group.

Discussion

The ccRCC is a malignant tumor lacking specific diagnostic markers and therapeutic targets. To solve this problem, we analyzed transcriptome data and clinical data of ccRCC from TCGA database, screened prognostic genes, and verified their role in the diagnosis and immunotherapy of ccRCC. Surprisingly, 9 genes (TYROBP, APOC1, CSTA, LY96, LAPTM5, CD300A, ALOX5, C1QA, and C1QB) not only have high diagnostic value, Moreover, the risk scoring model based on them was significantly correlated with immunotherapy indexes such as TIDE score, TMB, PD-1 and ICB response rate.

At present, the diagnosis of ccRCC is mainly based on histomorphology and immunohistochemistry. With the development and need of precision medicine, more and more molecules are proving to be helpful in the diagnosis of tumors. Cui et al. demonstrated that APOC1 is a novel diagnostic indicator of ccRCC at mRNA and protein levels in 32 patients and 2 human renal cancer cell lines (KAKI-2 and ACNH) [13]. Wu

et al. performed immunohistochemical staining on 15 cases of ccRCC and its adjacent tissues and found that TYROBP was positive in cancer tissues, suggesting its potential diagnostic indicator in ccRCC [14]. Through the diagnostic value of LAPTM5 in ccRCC by ROC curve, Sui et al. found that LAPTM5 has high diagnostic value (AUC = 94.808, cut off value = 12.983, Sensitivity = 85.158, Specificity = 94.444)[15]. Our study not only demonstrated the diagnostic value of APOC1, TYROBP, LAPTM5, but also six other genes (LY96, CSTA, CD300A, ALOX5, C1QA, and C1QB) in ccRCC, and found that the combined diagnostic value of these nine genes was higher than that of individual genes.

Stage and pathological grade are commonly used prognostic indicators of tumor[16, 17]. The RSM and the nomograph constructed in this study showed that the risk score had an independent prognostic function, and the nomograph could predict the 1/3/5-year survival probability of patients. The nine genes that construct the risk score are associated with poor prognosis of ccRCC, and they play an important role in the occurrence, progression, and metastasis of various tumors. The expression of APOC1 is closely related to the patient's age, clinical stage and pathological grade, indicating a shorter survival time and a poor prognosis[18]. Downregulation of LAPTM5 expression can inhibit the proliferation and survival of bladder cancer cells, resulting in cell cycle arrest[19]. CD300A promotes the progression of acute myeloid leukemia through the AKT signaling pathway[20]. Fang et al. used a variety of clinical data and molecular markers to construct a RSM and a nomograph to predict the overall survival rate and progress-free survival of patients with colorectal cancer, and achieved good results[21]. Differently, this study selected the nine genes associated with prognosis to build RSM, further screening independent prognostic factor of Cox regression analysis, which has the remarkable characteristics of factors (age, stage, pathological classification and risk score) into the nomogram to predict survival in patients with possibilities, provide guidance for clinical prognosis of the patients with forecast.

These indicators of immunotherapy (PD-1/PD-L1, TMB, MSI/MMR) are both independent and interrelated, what they have in common is that each indicator is an independent prediction of the efficacy of immunotherapy[7]. Except MMR, the higher the expression level of each indicator above the cut-off value, the higher the predictive value of immunotherapy[6, 22]. The protein expression of MMR was lost, and the consistency rate between MMR and MSI was up to 91.92%[23]. TIDE score is a predictor of tumor ICB response rate. A higher TIDE score corresponds to a lower ICB response rate, and a lower TIDE score corresponds to a higher ICB response rate[9]. This was consistent with the results of this study, namely, patients in the high-risk group had lower TIDE score and higher TMB and PD-1 levels, and the ICB response rate in the high-risk group was higher than that in the low-risk group, and the difference was significant. The difference was that the risk score in this study was only applicable to ccRCC. In this study, there were no significant differences in the expression levels of MSI, PD-L1 and CTLA4 between the high and low risk groups, which may be related to tumor heterogeneity and the relative independence of each indicator. Among all solid tumors, the proportion of both high TMB, high PD-L1 and MSI-H is only 73/11348[7]. The expression level of PD-1 in the high-risk group was higher than that in the low-risk group, while there was no difference in the expression of PD-L1. Why is this? PD-1 is an important immunosuppressive molecule, and PD-L1 is one of the ligands of PD-1. Under normal circumstances, the combination of the two can avoid the body's own immune response. Inflammation and tumors can lead

to upregulation of PD-1 and/or PD-L1 to escape the body's immune cells [24]. In addition, there is no expression data of MMR in this study, so it cannot be analyzed. However, all of these did not affect our conclusion that the RSM based on these 9 genes could effectively predict the efficacy of immunotherapy and might become a new immunotherapy evaluation index for ccRCC.

There was a positive correlation between immune score and risk score (high risk group > low risk group, $P=9.9E-16$). Immune infiltrating cells in tumor microenvironment were significantly different between high and low risk groups, especially activated dendritic cells. The expression of nine genes was lower in the low-risk group and negatively correlated with activated dendritic cells, but not in the high-risk group. The results showed that the low-risk group had low expression of 9 genes and high expression of activated dendritic cells. In the high-risk group, 9 genes were highly expressed and the number of activated dendritic cells did not change significantly. It suggests that patients in the low-risk group have higher immunity, while those in the high-risk group have weaker immunity. The complex relationship between tumor and immune system goes through three stages of clearance, balance and escape, accompanied by the process of immune system from strong to weak[25]. Immunotherapy involves either boosting the body's own immunity to eliminate tumor cells, such as the application of dendritic cell vaccine, or to reduce the immunity of tumor cells to enhance the recognition of the body's immune system, eliminate the immune escape of tumor, and achieve the purpose of eliminating tumor cells, such as the application of ICIs[26, 27]. Most studies have shown that high TMB level indicates that tumor cells have a higher level of neoantigen and a higher possibility of being recognized by the body's immune system[8]. Low TIDE score, high PD-1 expression and high ICB response rate are all indicators of the potential efficacy of ICIs, which, together with high TMB, predict increased chances of tumor cells being recognized by the immune system.

In conclusion, 9 genes associated with poor prognosis of ccRCC in 3 gene microarrays and TCGA database were selected in this study to construct a risk scoring model and a nomograph to predict the survival rate and immunotherapy efficacy of patients. Clinical data in TCGA database were used for verification. The roles of 9 genes in the diagnosis, prognosis and immunotherapy of ccRCC were demonstrated and the underlying mechanisms were explored. In particular, it proved the guiding value of risk score in immunotherapy, enriched the existing therapeutic efficacy prediction indicators of immunotherapy, and pointed out the direction for our further clinical research.

Declarations

Acknowledgments

No applicable.

Funding

No applicable.

Availability of data and materials

The three gene microarrays with ccRCC (GSE14994, GSE15641, and GSE53757) were obtained from the Gene Expression Omnibus (GEO) Datasets (<https://www.ncbi.nlm.nih.gov/gds/>). The transcriptome data and clinical information of ccRCC were downloaded from The Cancer Genome Atlas (TCGA) Datasets. Data in our study allows researchers to verify the results of the analysis and conduct secondary analyses.

Authors' contributions

Conceptualization, XW and GF; Bioinformatics analysis, XW and GF; Supervision, XW; Writing—original draft, XW and GF; Writing—review and editing, XW. All authors read and approved the final manuscript.

Ethics approval and consent to participate

No applicable.

Patient consent for publication

No applicable.

Competing interests

The authors declare no conflicts of interest.

References

1. Deleuze A, Saout J, Dugay F, Peyronnet B, Mathieu R, Verhoest G, et al. Immunotherapy in Renal Cell Carcinoma: The Future Is Now. *Int J Mol Sci.* 2020; 21.
2. Motzer R J, Escudier B, George S, Hammers H J, Srinivas S, Tykodi S S, et al. Nivolumab versus everolimus in patients with advanced renal cell carcinoma: Updated results with long-term follow-up of the randomized, open-label, phase 3 CheckMate 025 trial. *Cancer.* 2020; 126: 4156-67.
3. Brown L C, Desai K, Zhang T, Ornstein M C. The Immunotherapy Landscape in Renal Cell Carcinoma. *BioDrugs.* 2020; 34: 733-48.
4. Li B, Chan H L, Chen P. Immune Checkpoint Inhibitors: Basics and Challenges. *Curr Med Chem.* 2019; 26: 3009-25.
5. Petitprez F, Meylan M, de Reynies A, Sautes-Fridman C, Fridman W H. The Tumor Microenvironment in the Response to Immune Checkpoint Blockade Therapies. *Front Immunol.* 2020; 11: 784.
6. Zhu J, Armstrong A J, Friedlander T W, Kim W, Pal S K, George D J, et al. Biomarkers of immunotherapy in urothelial and renal cell carcinoma: PD-L1, tumor mutational burden, and beyond. *J Immunother Cancer.* 2018; 6: 4.
7. Vanderwalde A, Spetzler D, Xiao N, Gatalica Z, Marshall J. Microsatellite instability status determined by next-generation sequencing and compared with PD-L1 and tumor mutational burden in 11,348

- patients. *Cancer Med.* 2018; 7: 746-56.
8. Chalmers Z R, Connelly C F, Fabrizio D, Gay L, Ali S M, Ennis R, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med.* 2017; 9: 34.
 9. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med.* 2018; 24: 1550-8.
 10. Rini B I, Escudier B, Martini J F, Magheli A, Svedman C, Lopatin M, et al. Validation of the 16-Gene Recurrence Score in Patients with Locoregional, High-Risk Renal Cell Carcinoma from a Phase III Trial of Adjuvant Sunitinib. *Clin Cancer Res.* 2018; 24: 4407-15.
 11. Liu Y, Zhu D, Xing H, Hou Y, Sun Y. A 6gene risk score system constructed for predicting the clinical prognosis of pancreatic adenocarcinoma patients. *Oncol Rep.* 2019; 41: 1521-30.
 12. Liu X, Chen J, Lu W, Zeng Z, Li J, Jiang X, et al. Systematic Profiling of Immune Risk Model to Predict Survival and Immunotherapy Response in Head and Neck Squamous Cell Carcinoma. *Front Genet.* 2020; 11: 576566.
 13. Cui Y, Miao C, Hou C, Wang Z, Liu B. Apolipoprotein C1 (APOC1): A Novel Diagnostic and Prognostic Biomarker for Clear Cell Renal Cell Carcinoma. *Front Oncol.* 2020; 10: 1436.
 14. Wu P, Xiang T, Wang J, Lv R, Wu G. TYROBP is a potential prognostic biomarker of clear cell renal cell carcinoma. *FEBS Open Bio.* 2020; 10: 2588-604.
 15. Sui Y, Lu K, Fu L. Prediction and analysis of novel key genes ITGAX, LAPTM5, SERPINE1 in clear cell renal cell carcinoma through bioinformatics analysis. *PeerJ.* 2021; 9: e11272.
 16. Suzuki T, Sadahiro S, Tanaka A, Okada K, Saito G, Miyakita H, et al. A Modified Classification of Prognostic Factors Based on Pathological Stage and Tumor Regression Grade in Patients with Rectal Cancer Who Receive Preoperative Chemoradiotherapy. *Oncology.* 2017; 93: 287-94.
 17. Sun Y S, Zhao Z, Yang Z N, Xu F, Lu H J, Zhu Z Y, et al. Risk Factors and Preventions of Breast Cancer. *Int J Biol Sci.* 2017; 13: 1387-97.
 18. Xiao H, Xu Y. Overexpression of Apolipoprotein C1 (APOC1) in Clear Cell Renal Cell Carcinoma and Its Prognostic Significance. *Med Sci Monit.* 2021; 27: e929347.
 19. Chen L, Wang G, Luo Y, Wang Y, Xie C, Jiang W, et al. Downregulation of LAPTM5 suppresses cell proliferation and viability inducing cell cycle arrest at G0/G1 phase of bladder cancer cells. *Int J Oncol.* 2017; 50: 263-71.
 20. Sun X, Huang S, Wang X, Zhang X, Wang X. CD300A promotes tumor progression by PECAM1, ADCY7 and AKT pathway in acute myeloid leukemia. *Oncotarget.* 2018; 9: 27574-84.
 21. Wei F Z, Mei S W, Chen J N, Wang Z J, Shen H Y, Li J, et al. Nomograms and risk score models for predicting survival in rectal cancer patients with neoadjuvant therapy. *World J Gastroenterol.* 2020; 26: 6638-57.
 22. Tucker M D, Rini B I. Predicting Response to Immunotherapy in Metastatic Renal Cell Carcinoma. *Cancers (Basel).* 2020; 12.

- ## Figures

Screening for DEGs. (a) Screening of DEGs for GSE14994. (b) Screening of DEGs for GSE15641. (c) Screening of DEGs for GSE53757. (d) The venn diagram presents common down-regulated DEGs for the three gene microarrays. (e) The venn diagram presents common up-regulated DEGs for the three gene microarrays. (f) The venn diagram presents the down-regulated and up-regulated DEG common to the three gene microarrays and the TCGA database.

Figure 2

Figure 2

Identification of clinical traits-related genes. (a) Clustering dendrogram of all samples. (b) Analyses of scale-free index and mean connectivity for multiple soft-threshold powers. (c) Cluster dendrogram of all genes. (d) Correlation heatmap of the module eigengenes and clinical traits. (e) The expression of the 20 clinical traits-related genes based on the TCGA database.

Figure 3

Figure 3

Functional enrichment analysis of clinical traits-related genes and the identification of hub prognosis-related genes. (a) GO analysis, (b) KEGG analysis, (c) the analysis of relationship of the genes and pathways using the ClueGO plug-in of the Cytoscape software. The colors represent different clusters, and the thickness of the lines represents the strength of the correlation. (d) Kaplan-Meier survival analysis. The p-value < 0.05 was considered as a threshold value.

Figure 4

Figure 4

Construction of risk signature with 9 prognosis-related gene in 533 ccRCC patients. (a) the univariate Cox regression analysis. (b) the multivariate Cox regression analysis. (c) The distribution and survival rules of risk score, and the expression pattern of the 9 prognosis-related genes in high- and low-risk group. (d) Receiver operating characteristic (ROC) curve of 1 year, 3 years and 5 years overall survival (OS). (e) K-M survival curve of the OS in the high- and low-risk groups.

Figure 5

Figure 5

Correlation between risk score and clinical traits and their predictive value. (a) The multivariate Cox regression analysis for predicting the independent prognostic factor. (b) The nomogram based on Cox regression analysis. (c) The nomogram based on Logistic regression analysis. (d) The calibration curve for 1 year. (e) The calibration curve for 3 years. (f) The calibration curve for 5 years.

Figure 6

Figure 6

The evaluation of 9 prognosis-related genes in diagnostic and prognostic value of ccRCC patients. (a) ROC curves of each gene in evaluation of diagnostic value. (b) ROC curves of the model in evaluation of diagnostic value. (c-k) ROC curves of each gene in evaluation of prognostic value. (l) ROC curves of the model in evaluation of prognostic value.

Figure 7

Figure 7

The relationship between risk score and immunotherapy. (a) TMB score. (b) TIDE score. (c) PD-L1. (d) MSI_Exp_Sig. (e) IFNG. (f) CD8. (g) TAM_M2. (h) CAF. (i) MDSC. (j) Total IPS score. (k) IPS score with CTLA4_neg_PD1_neg. (l) IPS score with CTLA4_neg_PD1_pos. (m) IPS score with CTLA4_pos_PD1_neg. (n) IPS score with CTLA4_pos_PD1_pos.

Abbreviation: TMB: tumor mutational burden, TIDE: tumor immune dysfunction and exclusion, PD-L1: programmed cell death-ligand 1, MSI: microsatellite instability, IFNG: interferon gamma, TAM: tumor-associated macrophages, CAF: carcinoma-associated fibroblasts, MDSC: myeloid-derived suppressor cells, CTLA4: cytotoxic T - lymphocyte antigen 4, PD-1: programmed cell death 1.

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

- [TableS1.xlsx](#)
- [TableS2.xlsx](#)
- [TableS3.xlsx](#)
- [TableS4.xlsx](#)
- [TableS5.xlsx](#)