Identification of 4-Genes Model in Papillary Renal Cell Tumor Microenvironment Based on Comprehensive Analysis

Liang Luo (luoliang6@mail2.sysu.edu.cn)  
Third Affiliated Hospital of Sun Yat-sen University

Haiyi Zhou  
Shanxi Academy of Traditional Chinese Medicine

Hao Su  
Third Affiliated Hospital of Sun Yat-sen University

Research Article

Keywords: papillary renal cell carcinoma, tumor microenvironment, hub genes

Posted Date: January 7th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-138186/v1

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

**Background:** The tumor microenvironment acts a pivotal part in the occurrence and development of tumor. However, there are few studies on the microenvironment of papillary renal cell carcinoma (PRCC). Our study aims to explore prognostic genes related to tumor microenvironment in PRCC.

**Methods:** PRCC expression profiles and clinical data were extracted from The Cancer Gene Atlas (TCGA) database. Immune/stromal scores were performed utilizing the ESTIMATE algorithm. 323 samples were split into two groups on the basis of median immune/stromal score, and comparison of gene expression were conducted. Cross genes were obtained by Venn diagrams. Hub genes were selected through protein-protein interaction (PPI) network construction, and relevant functional analysis was conducted by DAVID. We used Kaplan–Meier analysis to identify the correlations between genes and overall survival (OS). Finally, univariate and multivariate cox regression analysis were employed to construct survival model and predict prognosis.

**Results:** We found immune/stromal score was correlated with T pathological grade and PRCC subtypes. 989 differentially expressed genes (DEGs) and 1169 DEGs were identified respectively on the basis of immune and stromal score. Venn diagrams indicated that 763 co-upregulated genes and 4 co-downregulated genes were identified. Kaplan-Meier analysis revealed that 120 genes were involved in tumor prognosis. Then PPI network analysis identified 22 hub genes, and four of which were significantly related to OS in patients with PRCC confirmed by cox regression analysis.

**Conclusions:** Four tumor microenvironment-related genes (CD79A, CXCL13, IL6 and CCL19) were identified as biomarkers for PRCC prognosis.

Introduction

The incidence of renal cell carcinoma (RCC) is approximately 403,000 in 2018, accounting for 2% of common malignant neoplasm on a global scale [1]. Clear cell, papillary, and chromophobe renal cell carcinoma are three main histologic subtypes of RCC. Study reported papillary renal cell carcinoma (PRCC) is the second most common type [2]. Clinicopathological features are the main diagnostic criteria for PRCC, but their predictive outcome is not accurate because of the lack of consistent standards [3]. Surgery is the primary choice of treatment for early stage PRCC, and comprehensive therapeutic approaches which include surgery, chemotherapy, and immunotherapy are applied for advanced stage patients [4]. However, the prognosis for PRCC is various on account of tumor metastasis and complications. The World Health Organization (WHO) classification (2016), according to different clinicopathological and immunohistochemical features, split PRCC into two subtypes: type 1 and type 2 [5]. In general, the prognosis of type 2 PRCC was poorer than type 1 [6].

Tumor microenvironment (TME) is an intricate system comprised of tumor cells, surrounding immune and inflammatory cells, tumor-related fibroblasts and nearby stromal tissues, and varieties of cytokines and chemokines [7]. Relevant studies have indicated macrophages play a crucial part in tumor initiation
and invasion of adjacent tissues [8]. Tumor-related fibroblasts, a kind of stromal cell, affect cancer progression partly by interacting with tumor cells and immune regulation [9]. Studies have demonstrated the highly enriched CD8+ T cells in PRCC are significantly connected with tumor development, progression, and mortality [10]. Yoshihara et al. put forward the ESTIMATE algorithm that utilizes gene expression patterns to calculate the immune/stromal scores in different tumor tissues [11]. Accumulating evidence showed ESTIMATE helped to clarify the importance of TME in numerous cancers [12–14].

In this study, we extracted PRCC datasets from TCGA and obtained relevant immune and stromal score calculated by ESTIMATE. After a series of bioinformatics analyses, several tumor microenvironment-related genes in connection with the prognosis of PRCC patients were selected and validated.

**Materials And Methods**

1. **Data preparation**

   Level_3 gene expression patterns for PRCC was downloaded from TCGA via UCSC Xena (http://xena.ucsc.edu/). Validation datasets (GSE2748) were obtained from the GEO database. And corresponding clinical information including overall survival (OS) time and status was extracted from the ONCOMINE database (www.oncomine.org/). Immune and stromal scores were calculated by ESTIMATE (https://bioinformatics.mdanderson.org/estimate/).

2. **Screening of differentially expressed genes (DEGs)**

   PRCC samples were split into low and high groups on the basis of median immune/stromal score, limma package of R (v3.6.3) was utilized to compare DEGs, |Log2FC|>1 and FDR < 0.05 were deemed significant. Volcano plot was drawn by the ggplot2 package of R.

3. **Survival analysis**

   We have adopted Kaplan–Meier analysis to evaluate the relations between DEGs and OS. And Log-Rank test was applied for comparing the differences. P < 0.05 was considered statistically significant.

4. **PPI construction and hub genes selection**

   The String database (https://string-db.org/) was used to analyze the molecular interactions of DEGs and construct a PPI network visualized by Cytoscape (v3.7.2). Then we used the MCODE tool to find densely connected hub genes based on topology.

5. **Functional enrichment analysis**

   Gene ontology (GO) analysis was performed by DAVID (http://david.abcc.ncifcrf.gov), comprising three aspects: biological process, cellular component, and molecular function. And Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis was applied for searching functional pathways of hub genes. P < 0.05 was deemed statistically significant.
6. Cox proportional hazard regression analysis

Based on accessible clinical data, we used univariate cox regression analysis to identify the impact of expression level of certain hub genes on patients’ overall survival. Then, multivariate cox regression analysis was applied to coefficients analysis of selected genes. Cox regression analysis was conducted by survival packages of R.

Results

1. Immune/stromal score were significantly associated with T pathological stage and PRCC subtypes

We downloaded 323 PRCC samples from the TCGA database. Gene expression patterns and clinical characteristics were included. Based on the ESTIMATE algorithm, immune score (-1952.04 to 3372.51) and stromal score (-1821.04 to 1469.55) of 291 samples were calculated. We attempted to analyze the relations between immune/stromal score and TNM, stage and subtype (Fig. 1A-D). Results showed the differences in the stromal score of T pathological stages were significant (T3 > T2 > T1, T4, P < 0.05). Since only two T4 samples are not representative, we considered that high T pathological grade correlated with high stromal score. And the average immune scores of type 2 were higher than type 1 (P < 0.05). Similarly, type 2 had higher stromal scores compared to type 1 (P < 0.05, Fig. 1E, F). Finally, we divided the samples into two groups on the basis of median immune/stromal score, but no significant differences were found in OS via Kaplan–Meier analysis (Fig. 1G, H).

2. Identification of DEGs with immune/stromal score

To determine whether all genes were associated with immune/stromal score, gene expression profile of 291 patients were analyzed. We mapped volcano plots with the cutoffs of |Log2FC|>1 and FDR < 0.05 (Fig. 2A, B). 983 upregulated and 6 downregulated genes were identified according to immune score. 1142 upregulated and 27 downregulated genes were extracted based on stromal score. Venn diagrams identified 763 upregulated common genes and 4 downregulated common genes (Fig. 2C, D, Supplementary Table S1). GO analysis showed upregulated genes were mostly associated with immune and inflammatory response, chemokine activity, IgG binding and cell adhesion molecule binding (Fig. 2E-G).

3. Association of upregulated common genes with OS

In order to obtain the relationships between each gene and OS, Kaplan–Meier was applied for survival analysis of 763 upregulated genes, and the impact of 120 genes (Supplementary Table S2) on OS was statistically significant.

4. Selection of molecular complexes from protein interaction networks

To study protein interactions of 120 genes, String tool was performed to structure a PPI network comprised of 84 nodes and 297 edges. Then we used MCODE to select prominent modules: module 1,
composed of 17 hub genes (CXCR5, CD19, PDCD1, IL21R, GZMB, TNFRSF9, CD38, CXCL10, SELL, LAG3, CD44, CD80, CCL21, ITGA4, CCL19, IL6, and CXCL13, Fig. 3A) and module 2, including 5 hub genes (BLK, CD79A, FCRLA, MS4A1 and POU2AF, Fig. 3B), which were strongly associated with other genes, suggesting that they might play an important role in PRCC.

5. Functional enrichment analysis of valuable genes

To understand whether 22 genes have an impact on the TME, we performed KEGG analysis by DAVID, and the results showed that these genes mainly participated in the following pathways: cytokine-cytokine receptor interaction, chemokine signaling pathway and primary immunodeficiency (Fig. 3C).

6. Construction and validation of a survival model to predict prognosis of PRCC patients

323 PRCC samples from TCGA and 34 samples from GEO were used as train and test group. Univariate cox regression analysis has approved all of 22 genes were significantly with OS (Supplementary Table S3). Then, we applied multivariate cox regression analysis to evaluate the coefficients between 22 hub genes from train group and OS. 4 genes (CD79A, CXCL13, IL6 and CCL19) significantly with patients’ prognosis were confirmed (Supplementary Table S3). Risk score was calculated based on the formula (risk score= (-0.24020) *CD79A + 0.22031* CXCL13 + 0.12025* IL6 + 0.17997* CCL19). According to median risk score, 323 PRCC samples and 34 samples were divided into high and low risk group, Kaplan–Meier survival curves both indicated patients’ prognosis in high risk group were worse than that in low risk group (p = 0.00326, p = 0.02537, Fig. 4A, B). Receiver operating characteristic (ROC) analysis was performed, and area under curve (AUC) of train group and test group was 0.76 and 0.708 (Fig. 4C, D).

Discussion

Recently, TME has been identified to have profound significance and effect on the occurrence, progression, and treatment of tumors, and it is extremely necessary to study its related mechanism. Study has shown changes in mesenchymal stromal cell differentiation promoted the progression of multiple tumors by altering the tumor microenvironment [15]. Maeda-Otsuka et al demonstrated hypoxia could accelerate the proliferation and migration of angiosarcoma by regulating tumor microenvironment [16]. Bader et al. reported that TME was closely linked to immunotherapy, and the understanding of TME could better guide treatment and achieve precision therapy [17]. Meanwhile, study indicated that immunotherapy was one of the promising options for advanced stage PRCC patients [18]. This study aimed to identify genes in connection with tumor microenvironment and significantly with OS in PRCC patients.

Firstly, we discovered that the differences in the stromal score of T pathological grades were statistically significant, indicating that stromal cells of TME were involved in tumor invasion to surrounding tissues. Relevant study has demonstrated that changes in stromal components, such as the presence of myofibroblasts, initiated tumor invasion, and metastasis [19]. Type 1 PRCC inclined to be localized in kidney whereas type 2 generally tended to invade the surrounding organs and had a poor prognosis [20].
Consistent with previous studies, results showed the mean immune/stromal scores of type 2 PRCC cases were significantly higher than type 1. GO analysis of genes co-upregulated in the immune/stromal groups indicated that they were involved in immune and inflammatory responses, suggesting that most of the genes were related to TME. Then, Kaplan–Meier analysis of 763 upregulation genes showed that 120 genes were associated with PRCC prognosis. Combined with PPI network construction and significant modules extraction, we obtained 22 hub genes. Then KEGG enrichment analysis showed that they were active in immune-related pathways, such as cytokine-cytokine receptor interaction, chemokine signaling pathway and primary immunodeficiency. Finally, univariate and multivariate regression analysis put forward a four-genes (CD79A, CXCL13, IL6 and CCL19) survival model based on 323 PRCC samples from the TCGA database, of which AUC was 0.76, and confirmed by a set of GEO data, with the AUC being 0.708.

Among the 4 genes, Certain studies have demonstrated that CCL19 could act as an immunomodulator by activating dendritic cells, T and B cells in secondary lymphoid tissue to modulate primary (or secondary) adaptive immune responses [21]. Besides, overexpression of CCL19 was discovered to be implicated in tumor progression in cervical cancer, but might be contribute to anti-vascular treatment in colorectal cancer through inhibiting angiogenesis [22, 23]. Numerous studies have shown that CXCL13, as a chemokine secreted by stromal cells was implicated in the occurrence, invasion, and lymph node metastasis of tumors by binding to its receptor CXCR5 [24–27]. Rachana et al. reported it could be a novel and compelling target for prostate cancer therapy. [28] Several studies indicated that CD79A was connected with aggressive hematological malignancy, could be a popular marker in the detection and treatment of hematopathy such as Hodgkin’s lymphoma and B-cell lymphoma. [29–31] IL6, a proinflammatory cytokine, has been reported to be influenced by cancer-associated fibroblasts (CAFs) in TME, and participated in the process of human tumor immunity improvement, tumor metastasis and colonization. [32–34]

The interaction of PRCC and its TME affected the whole process from tumor occurrence and progress to metastasis and recurrence, which created plenty of opportunities for diagnosis, treatment and prognosis of PRCC patients. In present study, we attempted to find tumor microenvironment-genes associated with PRCC patients’ survival. Our results may provide some related data to future research on the relations of PRCC and TME. However, the mechanism of PRCC and its microenvironment could be extremely complex, our analysis based on the TCGA and GEO database and bioinformatic tools is only one part of it. Further research and analysis based on large samples will be essential.

**Conclusion**

By analyzing the TCGA and GEO database and applying ESTIMATE algorithm and a series of bioinformatic methods, we obtained tumor microenvironment associated genes (CD79A, CXCL13, IL6 and CCL19), which were related to the clinical outcome of PRCC patients. These genes could be used as biomarkers for predicting PRCC prognosis.
### Abbreviations

PRCC  

### Declarations

**Ethics approval and consent to participate**

All data of the study were acquired from public database so that no more ethical approval was needed.

**Consent for publication**

Not applicable.

**Availability of data and material**


**Competing interests**

The authors declare that they have no competing interests.

**Funding**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Author contributions**

LL and HZ analyzed and interpreted the data regarding the papillary renal cell tumor. LL and HS prepared the manuscript and revised important intellectual content. All authors read and approved the final manuscript.

**Acknowledgements**

Not applicable.

**Authors' information**

1 Department of Urology, The Third Affiliated Hospital, Sun Yat-sen University, Guangzhou, 510630, China
References


29. Hoeller S, Zihler D, Zlobec I, Obermann EC, Pileri SA, Dirnhofer S, Tzankov A: **BOB.1, CD79a and cyclin E are the most appropriate markers to discriminate classical Hodgkin’s lymphoma from primary mediastinal large B-cell lymphoma.** *Histopathology* 2010, **56**(2):217–228.


**Figures**
Figure 1

The correlations between the stromal score and clinical characteristics, and OS analysis. A–D. The relationship between stromal score and T stage, N stage, M stage and clinical stage. E-F. The relationship between immune/stromal score and PRCC subtypes. G. Kaplan-Meier analysis of immune score with papillary renal cell carcinoma patients. H. Kaplan-Meier analysis of stromal score with PRCC patients
Figure 2

Figure 3

PPI networks of significant module. A. module1, B. module2. The node color changes gradually from grey to black indicating the ascending order of the degree of the genes. C KEGG pathway analysis for 17 hub genes
Figure 4

Cox regression analysis. Kaplan-Meier survival curves of TCGA samples (A) and GEO samples (B), Receiver operating characteristic curves of TCGA samples (C) and GEO samples (D).

Supplementary Files

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

- Additionalfile1.docx
- Additionalfile2.docx
- Additionalfile3.docx