Identification of Biomarkers Related to Tumor Associated Macrophages in Clear Cell Renal Cell Carcinoma with Weighted gene co-expression network analysis

Yunzhong Jiang
the First Affiliated Hospital of Xi’an Jiaotong University

Mengzhao Zhang
the First Affiliated Hospital of Xi’an Jiaotong University

Lu Zhang
the First Affiliated Hospital of Xi’an Jiaotong University

Lu Wang
the First Affiliated Hospital of Xi’an Jiaotong University

Minghai Ma
the First Affiliated Hospital of Xi’an Jiaotong University

Minxuan Jing
the First Affiliated Hospital of Xi’an Jiaotong University

Jianpeng Li
the First Affiliated Hospital of Xi’an Jiaotong University

Rundong Song
the First Affiliated Hospital of Xi’an Jiaotong University

Yuanquan Zhang
the First Affiliated Hospital of Xi’an Jiaotong University

Zezhong Yang
the First Affiliated Hospital of Xi’an Jiaotong University

Yaodong Zhang
the First Affiliated Hospital of Xi’an Jiaotong University

Yuanchun Pu
the First Affiliated Hospital of Xi’an Jiaotong University

Jinhai Fan

Research Article
**Keywords:** Clear cell renal carcinoma, WGCNA, Tumor associated macrophage, Tumor microenvironment, Prognostic model, Bioinformatics

**Posted Date:** October 19th, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-2160942/v1

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

Background

Tumor associated macrophages (TAMs) play an exceedingly important role in tumor microenvironment (TME). However, few studies has reported the specific biomakers and TAM related gene signature in clear cell renal carcinoma (RCC). Herein, our research aims to pick out specific biomakers and construct a TAM related gene prognostic signature in clear cell renal carcinoma.

Methods

We downloaded clear cell renal carcinoma patients sequence data from The Cancer Genome Atlas (TCGA). Then, we used CIBERSORT and QuanTIseq algorithm to estimate the percentage of immune cell infiltration. Weighted gene co-expression network analysis (WGCNA) could sort the specific genes related to TAM(M2 macrophage) in clear cell renal carcinoma. In addition, Least Absolute Shrinkage and Selection Operator (LASSO), univariate and multivariate Cox regression were performed to establish a TAM related gene prognostic model in clear cell renal carcinoma patients. At last, single cell sequence from Gene Expression Omnibus (GEO) was also verify the expression of TAM related genes.

Results

We found that TAM (M2 macrophage) infiltration was higher than normal tissue, which suggested that TAM may play an vital role in clear cell renal carcinoma progression. Twenty-seven TAM related genes was picked out by using WGCNA. In addition, the six gene prognostic signature which contained: TNFSF8, CD300FL, C3AR1, LAIR1, LY86, RNASE6 was established. We concluded that the riskscore and tumor stage were independent prognostic factors through univariate and multivariate Cox regression analysis. Single cell sequence results show that the six hub genes were all expressed in TAMs.

Conclusions

In summary, our research discovered that the TAM related genes and constructed a six gene prognostic signature for clear cell renal carcinoma, which may provide a novel insight for targeting therapy in clear cell renal carcinoma.

1. Background

Renal cell carcinoma accounts for 3% of all cancers, with the highest incidence in Western countries[1, 2]. It is reported that renal cell carcinoma is the most common solid lesion within the kidney and accounts for approximately 90% of all kidney malignancies[3, 4]. Clear cell renal carcinoma is one of the most common pathological type in renal cell carcinoma[5]. Tumor associated macrophages (TAMs) play an
important role in tumor progression and tumorigenesis, which is regarded as a major components of the
tumor environment [6, 7]. In recent years, many studies suggested tumor associated macrophage may
promote the tumor cell proliferation, invasion, and metastasis[8, 9]. Furthermore, the high TAMs
infiltration was reported to have a correlation with the poor prognosis in patients with cancer[10].
Therefore, Targeting TAM may become a potential method for treatment of cancer. However, the specific
biomarkers in clear cell renal carcinoma was less reported. It is imperative to construct a TAM related gene
prognostic model for clear cell renal carcinoma. In our study, we discovered the specific TAM genes and
construct a novel six gene prognostic model based on TCGA data, which may have a potential
significance for immunotherapy in renal cell carcinoma.

2. Results

2.1 Immune Cell Infiltration in Clear cell Renal Carcinoma
TME

CIBERSORT was used to estimate the immune cells’ infiltration in clear cell renal carcinoma (Supplement
Fig. 1A). It is clearly found that the proportion of M2 macrophage in the tumor tissue was higher than that
in normal tissue (Fig. 1B). Furthermore, we also found that the proportion of M2 macrophage increased
with the higher tumor stage and higher tumor grade (Fig. 1C,D), which indicated that the M2 macrophage
may play an important role in tumor progression (Fig. 1A). Similarly, quanTIseq was also used to analyze
different immune cells’ infiltration in clear cell renal carcinoma. We calculated the average fraction of
various immune cells, which was showed in pie chart. The proportion of M2 macrophage in the normal
tissue was lower than that in tumor tissue (Fig. 1E).

2.2 WGCNA can sort the TAM related genes

539 patients were clustered using the R software (Supplement Fig. 1B). Then, we chose β = 8 as an
appropriate soft threshold to construct a scale-free network (no scale R2 = 0.85) (Fig. 2A). The dynamic
cut tree was made after merging similar gene modules (Fig. 2B). Among the 13 gene module, we found
that the black module had a close relationship with M2 macrophage fraction features (Fig. 2C).
Additionally, we identified the key genes in the black gene module under the condition of GS > 0.7 and
MM > 0.8 (Fig. 2D). Furthermore, we performed a GO enrichment analysis for the black module. The genes
in the black model were correlated to immune receptor activity, amid binding, MHC protein binding (Fig.
3).

2.3 Construction of the PPI network and Six-gene
prognostic model

Twenty-seven key genes were picked out under the conditions: GS > 0.7 and MM > 0.8. Then, we
performed the protein-protein interaction (PPI) network analysis of 27 genes using String Database
(http://string-db.org/) [15]. The Cytoscape software was used to visualize the above results (Fig. 3C). 520
patients were randomly divided into two groups with the ratio 1:1. Then we used Lasso regression to filter
genes (Fig. 3C,D) in training cohort. From the analysis, only six genes were screened out, which contained: TNFSF8, CD300LF, C3AR1, LAIR1, LY86, RNASE6. Their coefficients were -0.785, -0.286, -0.125, 1.593, -0.208, and -0.226, respectively. Furthermore, we calculated the risk score for 396 patients according to the formula: risk score = TNFSF8*-0.785 + CD300LF*-0.286 + C3AR1*-0.125 + LAIR1*1.593 + LY86*-0.208 + RNASE6*-0.226.

2.4 Validation and Evaluation Of Six-Gene Signature

Patients in training cohorts were divided into high risk and low risk group according the median risk score. By using survival analysis, we concluded that the patients in the high-risk group had a lower survival time than that in the low-risk group (Fig. 4A). Moreover, a time-dependent ROC and calibration curves were drawn to verify our model's accuracy (Figs. 4B and 4C). The area under ROC curve in the train was nearly 0.7, and the calibration curve demonstrated the difference between patient actual survival status and predictive status in 1, 3, 5 years. The risk curve and survival status plot are shown in Fig. 5A, B. It was demonstrated that the patient’s risk had a positive correlation with their risk score. Furthermore, we regraded risk score as an independent variable and combined it with other clinical features to perform univariate and multivariate Cox analyses (Fig. 5C,D). We concluded that the risk score and tumor stage, is an independent prognostic factor in clear cell renal patients. The relationship between risk score based on six gene and clinical features were showned in Fig. 6 (A, B, C).

2.5 The Expression Of Key Genes In ScRNA Sequence Data

We downloaded the single cell data (GSE152938) from GEO(www.ncbi.nlm.nih.gov/). Then we used seurat R package to control data quality and filter genes (Supplement Fig. 2A B). Then we used harmony R package to eliminate the batch effect in two clear cell renal carcinoma samples (Supplement Fig. 3A). Seurat R package was also used to perform dimensionality reduction and cell clustering (Supplement Fig. 3B). Then, we make cell annotation according the results of difference analysis results(Supplement Fig. 3C, Fig. 7A), the expression of six key genes were visualized by umap function in seurat package (Fig. 7C). Furthermore, we used our six gene signature as a geneset to calculate the the degree of activity for our gene signature in every single cell. We found that our TAM related gene signature was higer expressed in TAM cells, which verified our results powerfully(Fig. 7B, D).

2.6 Construction and validation of Nomogram

Nomogram was established to visualize our results of multivariate Cox (Fig. 8A). It also provided a reference model to predict the 1-year, 3-year, and 5-year death rates of clear cell renal carcinoma patients. Then, ROC curve was made to verify the accuracy of our prediction model. The area under the ROC curve was approximately 0.8(Fig. 8B). The calibration curve was also plotted to observe the difference between the survival rate predicted by the nomogram and the actual survival rate (Fig. 8C). At last, DCA curve suggested that our nomogram can benefit in predicting the survival rate compared to risk score alone (Fig. 8D).

3. Discussion
Tumor microenvironment (TME) becomes an attractive topic in recent years. For example, the relationship between tumor and tumor microenvironment is like the relationship between seeds and soil. As the major component of TME, tumor associated macrophages (TAMs) was thought to be a complicated functional immune cells. Not only can it inhibit the tumor growth, but it also can promote the tumor growth, invasion, and metastasis[19]. Several studies verify that TAMs can differentiate into two major cells contained M1 and M2 macrophages. The process of cell differentiation was regulated by cytokines and chemokines in TME.[20, 21]. Accumulated research evidences indicated that M2 TAMs could promote tumor growth and invasion such as breast cancer, bladder cancer, gastric cancer[22, 23]. Therefore, targeting TAMs may become a potential treatment for various cancer. It is imperative to find some specific biomarkers related to TAMs. Fernando et.al performed the transcriptome profiling that revealed the key genes associated with human monocyte-to macrophage differentiation and polarization activation. Simultaneously, they also summarized TAM related well known biomarkers such as CD163, MRC1, MS4A4A, MS4A6A [24]. A TAM-related gene signature was constructed by Cassetta et al in order to predict the survival rate in patients with breast cancer.[25]. Shan et.al constructed a nine gene signature based on TAMs in glioma patients [26]. However, few studies reported the TAM related gene signature in clear cell renal carcinoma. Herein, we used WGCNA method to screen out TAM related gene signature. Simultaneously, a TAM related gene signature was also established to predict the patients’ prognosis. Furthermore, we used to single cell sequence data to verify the specific expression of our gene signature in TAMs. Moreover, we used to single cell data to verify our gene signature that was specific expressed in TAMs, which is different from other studies. Our gene signature contained six genes: TNFSF8, CD300LF, C3AR1, LAIR1, LY86,RNASE6. TNFSF8 is a member of tumor necrosis factor receptor superfamily. Several studies reported the TNF or TNF receptor has close correlation to TAMs. Duan et.al reported that the deficiency of CD30 ligand may accelerate the glioma progression and related to TAMs infiltration[27]. C3AR1 was a prognostic biomarker in various cancer. Yao et.al also established a TAM related gene signature contained C3AR1 to predict the prognosis of patients with esophageal squamous cell carcinoma[28]. Chen et.al found that C3AR1 was one of the key genes related to macrophage infiltration in gastric cancer[29]. C3AR1 was considered a TAM biomarkers in colorectal cancer-associated macrophage, which was reported by Cui et al. [30]. We can concluded that C3AR1 was a key gene related to TAMs infiltration in various cancer. Accumulated studies showed that LAIR1 was also a vital gene in the process of regulating TAMs. Keerthivasan et.al thought that the deficiency of LAIR1 was related to poor prognosis in human metastatic melanoma[31]. Furthermore, LAIR1 play an important role in regulating monocytes and macrophages. Ho et.al found that LAIR1 was specific TAMs (M2 macrophage) related biomarkers in hepatocellular carcinoma tumorenvironment. They analysed TCGA data to indicate the expression pattern was simliar to M2 macrophage biomakers such as CD163. Simultaneously, LAIR1 can mediate immunosuppression in M2 macrophages, which demonstrated that the expression of LAIR1 was related to prognosis of patients with hepatocellular carcinoma[32]. Few studies reported that TAM related gene signature in clear cell renal carcinoma. In our research, we found TAM related gene in clear cell renal carcinoma and riskscore based on our gene signature was a independent prognosis factor in patients with renal cancer, which can give clinical doctor a novel insight to treat cancer. However, there were limitations in our research. Our gene signature should be verified in basic and clinical experiment.
4. Conclusion

Our study firstly discovered that the TAM related genes and constructed a six gene prognostic signature for patients with clear cell renal carcinoma. Our gene signature was specific expressed in TAMs, not only could it provide potential biomarkers, but it could also a novel insight for targeting therapy in clear cell renal carcinoma.

5. Methods

5.1 Data collection and extraction

The gene expression data and clinical information were downloaded from TCGA database (http://portal.gdc.cancer.gov/) on March 24, 2022. The former contained 539 tumor tissue and 72 normal tissue samples. The latter included 527 cases. Patients who had missing follow-up data were excluded. Then, 520 patients were included in our research, and we used R 4.0 software to extract the gene expression matrix. In addition, single-cell sequence data were downloaded from Gene Expression Omnibus (GEO) database, with the data being derived from a study (GSE152938) focus on revealing potential tumor specific makers in renal cell carcinoma [11].

5.2 Fraction of immune cells in TME

CIBERSORT is a powerful bioinformatics tool that can predict the proportion of 22 immune cells infiltrated in TME [12]. By using this algorithm, we can observe infiltration of various immune cells in renal cancer tissues. The results with $p < 0.05$ were preserved. Then, we used ggplot2 R packages to show the TAM proportion in normal and tumor tissues. QuanTIseq is also an algorithm that can quantify tumor immunization according to human RNA-seq data [13]. The ImmuneDeconv R package was used to calculate TAM proportions in the samples by using the quanTIseq algorithm.

5.3 WGCNA

WGCNA can combine the gene expression and clinical information and sort the specific gene from our interest clinical modules. The WGCNA R package was used to establish a gene co-expression network[14]. Genes with an upper 25 median deviation were included and those with no significant expression were excluded. Pearson correlation values were calculated to construct a gene matrix. Then, we chose $\beta = 8$ as an appropriate soft threshold to construct a scale-free network with high connectivity. Additionally, a topological overlap matrix was produced to estimate the network's connectivity. Simultaneously, a hierarchical clustering dendrogram of the TOM matrix was established with a minimum size of 30 for every gene dendrogram. We set a cutoff threshold of $< 0.25$ to merge modules with high similarity. Then, the quanTIseq analysis results were combined with the different module eigengenes (MEs). Gene significance (GS) that shows the correlation of gene and immune cell fraction and module membership (MM). We set the condition that GS > 0.7 and MM > 0.8 to pick out 27 hub genes related to M2 macrophages.
5.4 Functional Enrichment analysis and Construction of protein-protein interaction (PPI) network

We used the clusterProfiler, ggpolt2 R packages to perform GO enrichment for key gene modules. To investigate the interactions between key genes, we performed the PPI network analysis of key genes using String Database (http://string-db.org/) [15]. Cytoscape software was used to show the interaction with TAM related genes.

5.5 Construction and validation of gene prognostic model and nomogram

520 patients were randomly divided into training cohorts and validation cohorts at a ratio of 1:1 according to R software. Then, we used the Lasso regression method to filter genes. From the analysis, only six genes were filtered out, and we extracted their coefficients to calculate the risk score for every patient. The risk score formula was as follows: risk score = \sum \text{Coefficient (mRNAi)} \times \text{Expression (mRNAi)}. Validation cohorts were used to verify this formula. We divided validation cohorts into high-risk and low-risk groups based on the median risk score in train cohorts. Survival, ROC curve analysis and DCA curve were used to evaluate the model performance and effectiveness. Finally, we regarded the risk score as an independent variable to perform univariate and multivariate Cox regression analysis, which was showned in a nomogram. The correlation between key genes and clinical features was visualized by using the pheatmap R package. Finally, we analyzed survival status according to the expression of key genes by using GEPIA (http://gepia.cancer-pku.cn/)[16].

5.6 Expression of Key Genes In Clear cell renal carcinoma Single-Cell Sequence Data

We downloaded the single-cell sequence data (GSE152938) from the GEO database (www.ncbi.nlm.nih.gov/). Then, we used the Seurat R package to control data quality, reduce data dimension. GSM4300628 and GSM4300629 were choose to perform analysis. Cells with < 200 and > 5,000 genes and a mitochondrial gene percentage of > 10% were filtered. Then, we used Harmony R package to eliminate the batch effect between two samples[17]. Furthermore, we combined two samples to make cell annotation according to the results of difference analysis results. We verified the expression of the gene signature in single-cell data. At last, we regarded our TAM related gene signature as a new geneset. AUCell R package was used to estimate the degree of activity for our gene signature in every single cell[18].

5.7 Statistical analysis

R 4.1.2 software and Graphpad Prism 9 were used to make calculations and statistical analyses. The Wilcoxon rank-sum test was used for comparison of two groups, and the Kruskal-Wallis test was used for comparison of more than two groups. Survival analysis was performed using Kaplan-Meier method with
the log-rank test. LASSO and univariate and multivariate Cox regressions were used to construct a prognostic model for clear cell renal patients. \( P < 0.05 \) was considered as significant.

6. Abbreviations

TAM  Tumor associated marcophage
TME  Tumor microenvironment
WGCNA  Weighted gene co-expression net work analysis
TCGA  The Cancer Genome Atlas
GEO  Gene Expression Omnibus
LASSO  Least Absolute Shrinkage and Selection Operator
PPI  Protein-Protein interaction
RCC  Renal cell carcinoma
TNFSF8  Tumor Necrosis Factor (Ligand) Superfamily, Member 8
CD300FL  CD300 Antigen-Like Family Member F
C3AR1  C3a Anaphylatoxin Chemotactic Receptor
LAIR1  Leukocyte-Associated Immunoglobulin-Like Receptor
LY86  Lymphocyte Antigen 86
RNASE6  Ribonuclease, RNase A Family, K6 2 3

Declarations

Author Contributions

JYZ, ZMZ and FJH had wrote the main manuscript text. SRD, ZYQ, PYC, ZYD, YZZ and JMX take part in the process of acquisition of data, analysis and interpretation of data. YZZ, WL, ZL, LJP prepared figure1-figure8. All authors reviewed the manuscript.

Funding

There is no funding on this article

Data Availability
The data in our study was downloaded from TCGA (http://portal.gdc.cancer.gov/) and GEO (www.ncbi.nlm.nih.gov/) under the accession number GSE152938

**Ethics approval and consent to participate**

Not applicable

**Consent for publication**

Not applicable

**Conflicts of Interest**

The authors declare that they have no competing interests.

**Acknowledgments**

We would like to sincerely appreciate all colleagues from the Department of Urology, The First Affiliated Hospital of Xi’an Jiaotong University, for their support. Figure1A was created with Biorender.com

**References**


Figures
Figure 1

Tumor associated macrophages (M2) infiltration in clear cell renal cancer tissue. (A) TAM mediated cancer progression. Figure 1A was created with Biorender.com. (B) M2 macrophages infiltration in various tissue (C) M2 macrophages infiltration in different tumor grade (D) M2 macrophages infiltration in different tumor stage. (E) The average M2 macrophage fraction in differ tissue type based on QuanTIseq algorithm.
Figure 2

Weighted gene co-expression network analysis. (A) Screening for suitable soft thresholds and scale-free network validation. The soft threshold is selected as 8, the distribution curve and network connectivity k, which represented a satisfactory scale-free network. (B) The cluster dendrogram with the gene modules and module merging. (C) The correlation between gene modules and immune cell fraction. (D) The correlation between GS and MM shonwn in scatter plot in black modules.
Figure 3

Molecular functional enrichment analysis in black module genes and identification of six TAM gene signature. (A) Molecular functional enrichment enrichment in black module genes. (B) The network of 27 TAM related genes filtered by the condition (GS>0.7, MM>0.8) (C) Lasso regression analysis was used to identify the six genes signature. (D) The cross-validation in the lasso model.
Figure 4

Validation of Evaluation of Six gene signature. (A) The Survival curve based on Kaplan Meier with log-rank test in train cohort, test cohort and entire cohort. (B) Time-dependent ROC curve means the accuracy of our model in predicting the 1 year, 3 year and 5 year survival rate. (C) Calibration curve that predicted 1 year, 3 year and 5 year survival probability.
Figure 5

Independent prognostic analysis and the correlation between survival status and risk score. (A) The risk score curve in the train cohort, test cohort, and entire cohort. (B) Patients status as risk score increases in the train cohort, test cohort, and entire cohort. (C) Univariate Cox analysis in the train cohort, test cohort, and entire cohort. (D) Multivariate Cox analysis in in train cohort, test cohort, and entire cohort.
Figure 6

The correlation between riskscore and clinical features. (A) The correlation between riskscore and different tumor grade. (B) The correlation between riskscore and tumor stage. (C) The heatmap shows that the correlation between riskscore and clinical features.
Figure 7

Single cell sequence data analysis. (A) Umap plots of single cell sequence data in GSE152938. (B) Violin plot of AUC distribute in different cell clusters. (C) The expression of six gene in TAMs (D) Umap plots of different cells according to AUC value.
Figure 8

Evaluation of Validation of Nomogram. (A) The nomogram was established to predict the prognosis of patients. (B) The time-dependent ROC curve was shown the accuracy of our nomogram. (C) The calibration curve of nomogram that predict the 1, 3, 5 year survival rate. (D) DCA curve was shown that the nomogram was better than risksore on predicting survival rate of patients.
Supplementary Files

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

- Supplemenstrokefigure1.pdf
- supplementfigure2.pdf
- Supplementfigure3.pdf
- supplementfigure4.pdf