Biomarker Screening and Prognostic Significance Analysis for Renal Cell Carcinoma

Xiangui Meng1–3,*
Hongwei Yuan1–3,*
Weiquan Li1–3,*
Wen Xiao1–3
Xiaoping Zhang1–3

1Department of Urology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, 430022, People’s Republic of China; 2Institute of Urology, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, 430022, People’s Republic of China
*These authors contributed equally to this work

Background: Studies report that conventional treatment of clear cell renal cell carcinoma (ccRCC) is effective, but several advanced patients present with poor prognosis. The current study explored potential new tumor markers and therapeutic targets in advanced ccRCC.

Methods: Biomarker gene expression of ccRCC was retrieved from GEO database and the Cancer Genome Atlas Kidney Clear Cell Carcinoma (TCGA-KIRC) database. Gene ontology (GO) analysis and protein–protein interaction (PPI) networks of biomarker genes were constructed using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) tool. Kaplan–Meier analysis and receiver operating characteristic curve (ROC) analysis were performed to explore the prognostic and diagnostic roles of these genes. Gene set enrichment analysis (GSEA) analysis was used to determine hallmark functions of the biomarker genes.

Results: A total of 21 upregulated genes were identified between advanced ccRCC and early ccRCC (grade III+IV vs grade I+II). Gene ontology analysis showed that the 21 upregulated genes were mainly implicated in biological processes including metabolic and lipid transport. The findings showed that 7 out of the 21 genes were significantly upregulated in 72-paired samples retrieved from the TCGA-KIRC. High expression of 5 genes indicated a poor prognosis of overall survival and disease-free survival in KIRC. Three genes effectively distinguished renal cancer tissue and adjacent renal tissues in a total of 533 ccRCC samples. GSEA showed that the 3 biomarkers were significantly enriched in epithelial–mesenchymal transition, G2M checkpoint, and angiogenesis. The results of qRT-PCR showed that STEAP3, IBSP, and AQPI had a significant identification effect in ccRCC.

Conclusion: The findings showed that 3 biomarkers were significantly upregulated in advanced ccRCC and could be used for diagnosis, prediction, and potential novel therapeutic targets for progression of ccRCC.

Keywords: advanced ccRCC, biomarker, progression, differentially expressed genes

Introduction

Clear cell renal cell carcinoma (ccRCC) is the most common pathological type of renal cancer (RCC).1,2 RCC is characterized by high incidence in the United States, and approximately 14,830 new deaths and 73,750 estimated new cases were reported in 2020 in the United States.3 New RCC cases in 2018 were approximately 403,262 (2.2% of all types of tumors), and new deaths were approximately 175,098 (1.8%) globally.4 Previous studies report that ccRCC is a highly fatal tumor disease and contributes to the highest renal cancer-related deaths.5 Timely diagnosis and confirmation of the cancer pathological type and pathological grade are important for effective management and treatment of patients. Treatment is adjusted based on...
the results of expression of clinical diagnostic markers, which has a positive effect on the prognosis of patients. Currently, no effective tumor markers have been identified for renal cancer and diagnosis is mainly performed through non-targeted imaging. Some patients present with metastasized tumors at the time of diagnosis with ccRCC. Therefore, determining effective biomarkers is important for abrogating the progression of renal cancer and improving the prognosis of patients.6,7

The main therapy for early-stage renal cancer is mainly local or radical nephrectomy, however, this approach is associated with a high risk of local recurrence or distant metastasis.8,9 Most patients die of cancer metastasis and invasion, implying that these processes affect tumor progression and treatment of ccRCC patients.10,11 Studies report that tyrosine kinase inhibitor (TKI) can effectively treat advanced renal cancer; however, the survival prognosis of many patients is still poor.12,13 Cell cycle regulation and cell adhesion maintenance can promote the occurrence and development of tumor cells.14 Phenotypic plasticity promoted by epithelial–mesenchymal transition (EMT) has an important clinical significance for cancer progression and drug resistance.15,16 TKIs inhibit the tumor cell cycle and EMT to affect tumor growth and metastasis.17,18 However, TKI resistance has been reported to be associated with poor progression of ccRCC.19 Therefore, there is an urgent need to explore more effective prognostic biomarkers and determine new drug targets for novel ccRCC therapies.

In the current study, differentially expressed genes (DEGs) were analyzed between advanced ccRCC and early ccRCC (grade III+IV vs grade I+II) using GSE150404 dataset. GO analysis showed that the metabolic and lipid transport is the main biological process associated with the 21 upregulated genes. The interaction of 21 upregulated genes was explored using the PPI network. The findings showed that 7 genes were significantly upregulated in paired 72-paired samples in TCGA-KIRC dataset. High expression of 5 genes indicated a poor prognosis of overall survival (OS) and disease-free survival (DFS) in ccRCC. Further analysis using ROC curves showed that 3 genes effectively discriminated between renal cancer tissue and adjacent tissues in a total of 533 ccRCC samples. Gene set enrichment analysis (GSEA) was used for prediction of the hallmark functions of the three upregulated genes.

In summary, the high expression level of the STEAP3, IBSP, and AQP9 indicated the patients were more likely to progress into advanced renal cancer and implied a poor prognosis. The biomarker genes were mainly enriched in EMT, G2M checkpoint, and angiogenesis biological processes. Further study should be conducted to explore the mechanism of action of STEAP3, IBSP and AQP9 in the occurrence and progression of advanced clear cell renal cell carcinoma.

Materials and Methods
Gene Expression Data of Patients’ Samples
GSE150404 gene expression data were retrieved from GEO database. GSE150404 dataset comprised four groups (total 60 samples) of samples for four grades identified by pathologists with 15 ccRCC samples in each group. mRNA expression data from the TCGA-KIRC database was retrieved from Xena tool (https://xenabrowser.net/heatmap),20 which included 72-paired cancer tissues and a total of 533 cancer tissues.

Analysis of Differentially Expressed Genes
Differentially expressed genes (DEGs) with log-Foldchange (logFC) ≥ 1.5 (upregulated) or logFC ≤ −1.5 (downregulated) between advanced ccRCC (grade III+IV) and early ccRCC (grade I+II) samples were identified using GEO2R tool at a p value < 0.05. The GEO2R tool was also used to identify the fold change of 21 upregulated DEGs between ccRCC and normal tissues.

Gene Ontology and PPI Network Analysis
Gene ontology (GO) analysis of DEGs was conducted using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) webserver (https://david.ncifcrf.gov/). Protein–protein interaction (PPI) networks of DEGs were constructed using the Search Tool for the Retrieval of Interacting Genes (STRING) (https://string-db.org/).

Kaplan Meier Analysis and ROC Curves Analysis
Overall survival (OS) and disease-free survival (DFS) of KIRC were determined using GEPIA tool (http://gepia.cancer-pku.cn/). The potential clinical diagnostic value of candidate genes was explored by determining expression levels of the genes with the receiver operating characteristic curve (ROC).21,22
Gene Set Enrichment Analysis
All cases of KIRC were divided into two groups (high expression vs low expression) based on the median value of biomarker genes. Gene set enrichment analysis (GSEA) analysis was then performed to determine hallmark functions of biomarker genes.

RNA Isolation and qRT-PCR
Twenty-four pairs of ccRCC tissue samples were obtained from the Department of Urology, Union Hospital, Tongji Medical College, Wuhan, China. The study was compliant with all relevant ethical regulations regarding research involving human participants. Total RNA was extracted from samples by the TRIzol reagent (Thermo, Massachusetts, USA) and then reverse-transcribed. The resulting cDNA was then subjected to real-time PCR analysis with SYBR Green mix (Thermo, Massachusetts, USA). Results were normalized to GAPDH for each sample. Gene primers are obtained from Sangon Biotech (Shanghai) and list as follows:

GAPDH
Forward 5′-GAGTCAACGGATTTGGGTAG-3′
Reverse 5′-GACAAGCTTCCCGTTCTCAG-3′

AQP9
Forward 5′-CCACCAGAAGAGCAGATATCAGCAG-3′
Reverse 5′-CGGTCTCTTCATTTACACAGCAGATG-3′

IBSP
Forward 5′-GCACCAGTACCAACAGCACAGAG-3′
Reverse 5′-GGCTCCAGTGACACTTTCTTCTTCC-3′

STEAP3
Forward 5′-CGGAGGTCATCTTTGTGGCTGTG-3′
Reverse 5′-GCTGAAGGTGCTCTTGCTCTTAG-3′

Statistical Analysis
SPSS Statistical software version 22.0 (IBM SPSS, Chicago, IL) was used for statistical analysis. All clinical-pathological gene expression data were presented as mean and SEM. Data on paired cases were analyzed using a paired sample t-test, while analysis of unpaired cases was performed using t-test. P-value < 0.05 was considered statistically significant.

Results
Differentially Expressed Genes (DEGs) in Advanced ccRCC
A total of 60 samples, comprising four groups of four grades, were confirmed by hospital pathologists and included in the current study. Each group comprised 15 ccRCC samples. DEGs between advanced ccRCC (grade III+IV) and early ccRCC (grade I+II) samples were identified using GEO2R tool. P-value < 0.05, logFC ≥ 1.5 or logFC ≤ −1.5 was used as the cut-off criteria. A total of 131 DEGs were identified, with 21 upregulated DEGs and 110 downregulated DEGs (Figure 1A and B).

Functional Enrichment Analysis Showed Enrichment of Pathways Implicated in Cancer and a PPI Network Was Constructed Based on DEGs
GO function analysis of upregulated genes was performed with the DAVID tool. The findings showed that enriched biological processes included acute inflammatory response, lipid transport, innate immune response, cellular protein metabolic process, protein secretion, positive regulation of the multicellular organismal process, wound healing, cell adhesion, and protein metabolic process (Figure 2A). Significantly enriched cellular components included extracellular matrix, external side of the cell surface, plasma membrane, extracellular exosome, vesicle, extracellular space and extracellular region part (Figure 2B). Analysis showed that enriched molecular functions included growth factor binding, enzyme inhibitor activity, receptor binding, enzyme regulator activity and molecular function regulator (Figure 2C). PPI network of the upregulated genes was constructed using the STRING tool (Figure 2D).

Differentially Up-Regulated Genes Were Identified in ccRCC
Differential expression of these 21 genes, in tumor and normal tissues, was further explored by TCGA-KIRC database. Analysis showed that most of the 21 genes were also upregulated in KIRC cancer tissues compared with the level in adjacent normal renal tissue (Figure 3A). Setting the cut-off criterion as p<0.05 and [logFC] > 1.5, we identified 7 upregulated genes as common to the two independent datasets by employing integrated bioinformatics analysis (Figure 3B). The 7 upregulated genes included: six-transmembrane epithelial antigen of prostate 3 (STEAP3), insulin-like growth factor-binding protein 1 (IGFBP1), lipopolysaccharide-binding protein (LBP), CUB domain-containing protein 1 (CDCP1), interleukin 2 receptor subunit alpha (IL2RA), integrin-binding sialoprotein (IBSP) and aquaporin 9 (AQP9) (Figure 3C and D). In addition,
the correlation between the expression levels of 7 DEGs and pathological grade was also validated by TCGA database. The resulting data showed a higher mRNA expression in advanced ccRCC (grade III+IV) than early ccRCC (grade I +II) patients (Supplementary Figure 1A–G).

Selected Genes are Prognostic Markers in ccRCC

The prognostic value of the 7 upregulated genes in ccRCC was explored. Patients were divided into two groups based on the median expression level of selected genes. Kaplan–Meier analysis results showed that patients with a high expression level of 5 genes (STEAP3, LBP, IGFBP1, IBSP, and AQP9) had a longer OS (Figure 4A–E and Supplementary Figure 2A, B) and DFS compared with those with low expression levels of these genes (Figure 5A–E and Supplementary Figure 2C, D).

Selected Genes Showed Differential Expression Between TCGA-KIRC and Normal Samples

To further explore the difference in the expression level of the 5 selected genes in tumor samples, the expression levels of the genes in all samples, including 72 normal tissues and 533 tumor tissues, were determined (Figure 6A). The findings showed that expression of the 5 genes was significantly differentially upregulated in 72 paired tissues (Figure 6B), whereas expression of 4 genes was significantly upregulated in 533 tumor samples (Figure 6C).

Receiver Operating Characteristic Curve (ROC) Analysis Showed That DEGs Had Prognostic Value

The diagnostic value of the 5 DEGs was explored using ROC curves to verify the diagnostic capability of ccRCC. The findings showed that only STEAP3, IBSP, AQP9 could effectively distinguish 72 normal renal tissues from 72-paired cancer tissues or a total of 533 cancer tissues, with an area under the curve above 0.5 (Figure 7A–F, and Supplementary Figure 2E–H). Furthermore, the aberrant expressions of STEAP3, IBSP and AQP9 mRNA, in ccRCC clinical samples, were also evaluated by real-time PCR analysis. As expected, STEAP3, IBSP and AQP9 mRNA expressions were significantly upregulated in ccRCC tissues (Supplementary Figure 3A–C). And ROC results from clinical data, with a high AUC, further supported that these biomarkers can effectively distinguish ccRCC from adjacent renal tissues, consistent with the previous results (Supplementary Figure 3D–F). These findings indicate that the selected genes are potentially effective diagnostic biomarkers for ccRCC.
Hallmark Functions of Biomarker Genes

GSEA analysis was performed to predict the hallmark functions of the biomarker genes. Expressions of STEAP3, IBSP, and AQP9 were significantly enriched in EMT, G2M checkpoint, and angiogenesis (Figure 8A–C). These hallmark functions are implicated in tumor metastasis, cell cycle, and angiogenesis, which leads to the progression of the tumor. These genes can thus be used as diagnostic biomarkers or treatment targets for advanced renal cancer.

Discussion

RCC is a common malignant tumor in the urinary system. CcRCC is a pathological type of renal cancer characterized by high incidence and high fatality rates. Cancer treatment has been significantly improved by targeting the expression of tumor markers, however, few effective tumor markers have been identified for renal cancer. Therefore, determining an effective biomarker is important to reduce mortality of renal cancer. Local or radical nephrectomy is the main treatment approach for early-stage renal cancer; however, it is characterized by a high risk of local recurrence or distant metastasis after the procedure. Application of tyrosine kinase inhibitor (TKI) is highly effective in the treatment of advanced renal cancer, however, patients with advanced stages present with poor survival prognosis. EMT plays a significant role in cancer progression and drug resistance of TKI. There is a need to explore more effective prognostic
Figure 3 Gene expression level of selected upregulated genes in ccRCC. (A) Volcano map of gene expression level of selected 21 upregulated genes in 72 matched normal and adjacent renal cancer tissues retrieved from TCGA-KIRC. (B) A Venn diagram of two independent datasets from GSE150404 dataset and TCGA-KIRC database. (C) Heat map showing expression level of the 7 genes in ccRCC from GSE150404 dataset. (D) Heat map showing expression level of the 7 genes in 72 matched KIRC samples.
Figure 4 Overall survival analysis of selected upregulated genes in ccRCC. (A) STEAP3, (B) LBP, (C) IGFBP1, (D) IBSP and (E) AQP9.
Figure 5 Disease-free survival analysis of selected upregulated genes in ccRCC. (A) STEAP3, (B) IGFBP1, (C) LBP, (D) IBSP and (E) AQP9.
biomarkers and identify novel drug targets for the development of ccRCC therapies.

Specific biomarkers were identified in the current study through microarray expression profile from the GSE150404 dataset for advanced ccRCC. The dataset comprised a total of 60 samples, containing four groups of four grades confirmed by hospital pathologists with 15 ccRCC samples in each group. The four grades were divided into two groups, namely early and advanced renal cancer groups. A total of 21 up-regulated genes were identified and GO analysis was performed to explore biological pathways and gene functions of the upregulated genes using DAVID tool. The findings showed that upregulated genes were enriched in the biological process including acute-phase response, acute inflammatory response, lipid transport, innate immune response, cellular protein metabolic process, protein secretion, wound healing, cell adhesion, and protein metabolic process.

Figure 6 Verification of expression level of 5 selected upregulated genes in total cases of ccRCC. (A) Heat map of expression level of 5 selected upregulated genes in total cases of TCGA-KIRC. (B) Expression level of 5 selected upregulated genes in 72 matched normal and cancer tissues in TCGA-KIRC. (C) Expression level of 5 selected upregulated genes in 72 normal and 533 cancer tissues in TCGA-KIRC. *P<0.05, **P<0.01 and ****P<0.0001.
Moreover, upregulated genes were used for the construction of a protein–protein interaction (PPI) network. Previous studies report that lipids that accumulate in kidney cancer cells are mainly cholesterol and triglycerides.\textsuperscript{1,31} Lipid storage can promote the integrity of endoplasmic reticulum, thus maintaining the homeostasis of tumor cells and promoting tumor development.\textsuperscript{32} Findings from previous studies showed that reducing lipid accumulation can inhibit growth and metastasis of tumor cells.\textsuperscript{33,34} Further studies should be conducted to explore the potential role of identified genes in lipid transport or metabolic process in ccRCC.

To further determine the most critical genes that affect the progression of renal cancer, the expression levels of the upregulated genes in KIRC were determined using 72 renal cancer tissues and adjacent normal renal tissues. The findings showed that 7 were significantly upregulated in renal cancer tissues compared with the levels in adjacent normal renal tissues. Kaplan–Meier curves were used to explore the prognostic role of 7 genes by evaluating their correlation with OS and DFS in KIRC. The findings showed that high expression levels of STEAP3, LBP, IGFBP1, IBSP, and AQP9 were correlated with a poor prognosis of ccRCC patients. Further analysis showed that STEAP3, IBSP, and AQP9 were significantly highly expressed in cancer samples and effectively distinguished tumors from normal tissues in a dataset of 72 normal tissues and 533 tumor tissues. These findings indicate that these genes can be used as biomarkers for renal cancer.

GSEA findings showed that high expression of STEAP3, IBSP and AQP9 were enriched in EMT, G2M checkpoint, and angiogenesis. High expression of these genes may play a role in the progression of the tumor to enter the advanced stage or may be implicated in TKI drug resistance. Previous studies report that AQP9 is upregulated in breast cancer\textsuperscript{35} and ccRCC.\textsuperscript{36} However, a previous study reported that overexpression of AQP9 inhibits

Figure 7 Diagnostic value of the selected upregulated genes in ccRCC. (A–C) ROC curve of STEAP3, IBSP, AQP9 between 72 matched normal and cancer tissues in TCGA-KIRC. (D–F) ROC curve of STEAP3, IBSP, AQP9 between 72 normal and 533 cancer tissues in TCGA-KIRC.
growth and metastasis of hepatocellular carcinoma. STEAP3 plays a key role in cellular iron uptake and homeostasis functions and thus serves as an oncogene in glioma progression. In addition, STEAP3 is a key gene for bone metastasis in breast cancer. IBSP encodes a specific regulator of matrix metalloproteinase, and high expression of IBSP is correlated with poor prognosis of esophageal carcinoma. However, the role and expression levels of STEAP3 or IBSP have not been explored in renal cell carcinoma. The current study conducted bioinformatics analysis, and the findings showed that STEAP3 and IBSP are highly expressed in advanced renal cell carcinoma and are implicated in EMT, cell cycle regulation, and angiogenesis. The high expression levels of STEAP3 and IBSP can thus be used as biomarkers for advanced renal cancer.

In summary, the current study used comprehensive bioinformatics analysis to identify 3 biomarkers for advanced ccRCC. Low expression of these genes showed a good prognosis, whereas a higher expression level of 3 biomarkers is associated with a poor prognosis of advanced renal cancer. These findings provide a basis for exploring the mechanism of occurrence and progression of renal cancer. However, further in vitro and in vivo studies should be conducted to verify these findings and explore the mechanism of action of these genes in renal cancer.

Ethics
This investigation was approved by the Ethics Committee of Huazhong University of Science and Technology.
Author Contributions
All authors made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; took part in drafting the article or revising it critically for important intellectual content; agreed to submit to the current journal; gave final approval for the version to be published; and agreed to be accountable for all aspects of the work.

Funding
This work was supported by grants from the National Natural Science Foundation of China (81902588), the National Key Scientific Instrument Development Project (81927807), and the National Key R&D Program of China (2017YFB1303100), and the project funded by the China Postdoctoral Science Foundation (2020T103229, 2020M672368), Wuhan Science and Technology Plan Application Foundation Frontier Project (2020020601012247).

Disclosure
The authors declare that they have no competing interests.

References


