Identification of genes that correlate clear cell renal cell carcinoma and obesity and exhibit potential prognostic value

Jiale Zhou#, Zhaolin Yang#, Xiaorong Wu, Jin Zhang, Wei Zhai, Yonghui Chen

Department of Urology, Renji Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai, China

Contributions: (I) Conception and design: J Zhou, W Zhai; (II) Administrative support: Y Chen; (III) Provision of study materials or patients: J Zhang, Y Chen; (IV) Collection and assembly of data: J Zhou, Z Yang, X Wu; (V) Data analysis and interpretation: J Zhou, W Zhai; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

#These authors contributed equally to this work.

Correspondence to: Wei Zhai; Yonghui Chen. Department of Urology, Renji Hospital, School of Medicine, Shanghai Jiaotong University, 160 Pujian Rd, Pudong District, Shanghai 200127, China. Email: jackyzw2007@163.com; cyh1488@163.com.

Background: Renal cell carcinoma (RCC) is a common urologic malignancy. Although the relationship between clear cell RCC (ccRCC) and obesity has been well-established by several large-scale retrospective studies, the molecular mechanisms and genetic characteristics behind this correlation remains unclear. In the current study, several bioinformatics tools were used to identify the key genes in ccRCC related to obesity.

Methods: Microarray data comparing ccRCC with normal renal tissues in patients with and without obesity were downloaded from the GEO database for screening of differentially expressed genes (DEGs). The DEGs were verified with expression level and survival analysis using several online bioinformatics tools.

Results: In the current study, the differential expression of five genes correlated with both ccRCC and obesity; IGHA1 and IGKC as oncogenes, and MAOA, MUC20 and TRPM3 as tumor suppressor genes. These genes were verified by comparing the relationship between the expression levels and survival outcomes from open-source data in The Cancer Genome Atlas (TCGA) dataset.

Conclusions: In conclusion, the five genes differentially expressed in ccRCC and obesity are related to disease progression and prognosis, and therefore could provide prognostic value for patients with ccRCC.

Keywords: Clear cell renal cell carcinoma (ccRCC); obesity; prognosis; biomarker


doi: 10.21037/tau-20-891

View this article at: http://dx.doi.org/10.21037/tau-20-891

Introduction

Renal cell carcinoma (RCC) is a common urologic malignancy with several histopathological subtypes, among which clear cell RCC (ccRCC) remains the most common. According to the latest global cancer statistics in 2018, there were approximately 400,000 cases of newly diagnosed RCC and 175,000 RCC-related deaths annually (1). Although an increasing number of RCC cases were diagnosed at their early stages due to the improvements in imaging techniques, and managed by nephron sparing surgery, nearly 30% of RCCs were diagnosed at the locally advanced stage or with distant metastasis (2). Currently, the first-line therapy for metastatic RCC (mRCC) is either molecular-targeted drugs, such as sorafenib, axitinib, sunitinib, pazopanib and everolimus, or immune checkpoint inhibitors such as nivolumab (3). However, due to the strong side effects of these drugs as well as the individualized sensitivity of patients to these drugs, the overall survival (OS) and quality-of-life of patients with mRCC has not improved much over the past decade. Thus, the main focus for improving OS should be on early screening, diagnosis and intervention.

The etiology of RCC is still not clear; however, the relationship between ccRCC and obesity has been well-established by several large-scale retrospective studies (4-7). Nevertheless, the mechanism of obesity leading to high...
incidence of ccRCC has not been fully studied. Several obesity prognostic biomarkers, including leptin, leptin receptor, adiponectin, visfatin and apelin have been studied. However, the significance of these prognostic biomarkers is controversial. Circulating leptin had a negative correlation with RCC risk in a study by Spyridopoulos et al., but was reported to be associated with a high risk of RCC occurrence by Liao et al. (8,9). The prognostic value of circulating adiponectin may also need further study, since the present studies have come to contradictory conclusions regarding the relationship between circulating adiponectin and RCC prognosis (10,11). Therefore, it is essential to identify the key genes that correlate RCC and obesity so that the potential mechanism behind the obesity/RCC relationship can be further studied.

Currently, with the gradual development and popularization of high-throughput sequencing and gene chips in oncological study, we can identify key genes related to carcinogenesis and prognosis based on open-source platforms and bioinformatics. In this study, we used bioinformatics methods to identify and validate the genes significantly related to both obesity and ccRCC.

We present the following article in accordance with the STREGA reporting checklist (available at http://dx.doi.org/10.21037/tau-20-891).

**Methods**

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and the Harmonized Tripartite Guideline for Good Clinical Practice from the International Conference on Harmonization. The study was approved by the research medical ethics committee of Shanghai Renji Hospital [Ethics board approval number: Renji Ethics [2017]091].

**Gene expression profile data**

GSE46699, the gene expression data and related clinical data of ccRCC, was obtained from Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/). GSE46699 contains 60 patients with ccRCC divided into two groups; 12 obese patients compared with 13 non-obese patients, and 16 smoking patients compared with 19 non-smoking patients. The tissue samples included 60 tumor and 56 normal tissues. GSE46699 was used to screen for genes differentially expressed in both ccRCC and obesity. Hub genes were verified with sequencing data and clinical information from The Cancer Genome Atlas (TCGA) data portal (https://portal.gdc.cancer.gov/).

**Data standardization and screening of differentially expressed genes (DEG)**

Raw biological data in GSE46699 downloaded from GEO dataset was first preprocessed and standardized. The data platform for GSE46699 was GPL19109 (HG-U133_Plus_2) Affymetrix Human Genome U133Plus 2.0 Array. The Robust Multi-array Average method was used to eliminate noise and guarantee integrity.

DEGs between ccRCC vs. normal kidney tissues and between obese vs. non-obese cases were detected using the “limma” (linear models for microarray data) R package. The screening criteria for DEGs was based on fold change and t-test. The criteria for statistical significance included either absolute value of log2FC (fold change) >1.00 and t-test P value <0.05. DEGs with log FC <0 were defined as downregulated genes, while DEGs with log FC >0 were defined as upregulated genes.

**KEGG and GO enrichment analyses of DEGs**

To explore the functional annotations of the DEGs including the associated biological processes, molecular functions, and cellular components, Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) enrichment analyses were conducted. Based on the GO database, GO analysis obtains all functions involving a gene and then uses Fisher’s exact test and multiple comparison tests to calculate the significance level (P value) and false discovery rate (FDR) of each function, so as to determine the significant functions embodied by the DEGs (12). KEGG is a database that systematically analyzes the relationships between genes (and their coding products), gene functions, and genomic information (13). Fisher’s exact test and chi-square test were used in KEGG pathway analysis to identify significant pathways. A P value <0.05 was set as the cutoff criteria for GO and KEGG analysis. Hierarchical clustering was also conducted.

**Validation and survival analysis of DEGs**

The expression of DEGs in ccRCC was validated using the Gene Expression Profiling Interactive Analysis (GEPIA) web server, an online tool that provides gene expression profiling of cancer and normal tissues, as well as interactive
analyses based on TCGA and Genotype-Tissue Expression (GTEx) data portals (14). The survival analysis of the DEGs was also conducted through GEPIA, with the OS as the primary end point. The expression of the DEGs in different stages of ccRCC were also validated based on the TCGA data in GEPIA.

Statistical analysis

R 4.0.0 was used for the analysis of DEGs. χ² test was used to analyze the statistical significance of GO and KEGG pathway analysis. Differences with P<0.05 were considered statistically significant.

Results

Screening of DEGs in ccRCC

After searching the GEO datasets with the key words ‘Obesity’ and ‘Clear Cell Renal Cell Carcinoma’ (ccRCC), the GSE46699 dataset which includes smoking- and obesity-related molecular alterations in ccRCC was used for further analysis. Gene expression was compared between tumor samples from obese vs. non-obese patients to identify tumor genes that correlate with obesity. Additionally, we compared gene expression in tumor vs. para-cancer tissue samples from patients with both ccRCC and obesity to identify genes that correlate with ccRCC. The DEGs obtained from these two analyses are shown as heat maps and volcano plots in Figure 1. Twelve genes were differentially expressed in both analyses, including 8 upregulated and 4 downregulated genes.

Functional annotation and KEGG pathway enrichment of DEGs

The GO analysis and KEGG pathway enrichment were performed on the two groups of DEGs to further explore potential biological processes in ccRCC (Figure 2). GO analysis indicated that the DEGs play roles in the hydrogen peroxide catabolic process, extracellular structure organism construction, and leukocyte migration (Figure 2A,B). Based on KEGG pathway enrichment analysis, the most significant pathway was “Phagosome”; other significant pathways included “HIF-1 signaling pathway”, and “Cell adhesion molecules” (CAMs) (Figure 2C,D). Hierarchical clustering showed that the DEGs can distinguish RCC tissue from normal tissue (Figure 3).

Validation of DEGs

The 12 DEGs were further validated using the patients’ data from GEPIA, an online, interactive, gene expression profiling analysis tool. The expression levels of the 12 DEGs were all significantly different in ccRCC tissues compared with normal para-cancerous tissues (Figure 4). Furthermore, among the 12 DEGs, there were 5 genes that statistically correlated with OS; IGHA1 and IGKC were negatively correlated with OS, and MAOA, MUC20 and TRPM3 were positively correlated with OS (Figure 5). The expression level of these five DEGs at different stages of ccRCC showed similar trends in that the expression of both IGHA1 and IGKC were elevated in the later stages (stage III–IV) of ccRCC during which the survival rate drops dramatically, while MAOA, MUC20 and TRPM3 expression decreased in the later stages of ccRCC (Figure 6).

Discussion

Currently, mRCC is detected in nearly 30% of newly diagnosed RCC cases. Even though the application of molecular-targeted drugs and immune checkpoint inhibitors has improved the survival of patients with mRCC, the outcome of these interventions is far from ideal.

Obesity is well recognized as a risk factor for many diseases, including type II diabetes mellitus, cardiovascular disease, hypertension, etc. (15). Epidemiological studies have also revealed the relationship between obesity and various types of human malignancies, including RCC (7,16). Approximately 3.6% of all newly diagnosed cancer cases have been connected with an increased body mass index (17), and obesity likely increases cancer mortality since a previous study showed that weight reduction reduces the risk (18). The mechanisms by which obesity causes carcinogenesis and tumor progression remain unknown. One possible reason is that adipose tissue, now viewed as one of the largest endocrine organs in human body, is believed to cause a chronic state of inflammation and to have mitogenic, proangiogenic and antiapoptotic effects (19,20). After reviewing previous studies, we realized that the mechanism and the genes involved have not been clearly defined. There is still controversy regarding the biological function of obesity-related genes in the progression and prognosis of ccRCC.

In the current study, bioinformatics methods were used to identify five genes whose expression correlate, either positively or negatively, with both ccRCC and obesity,
Figure 1 Differentially expressed genes (DEGs) and commonly DEGs in the GSE46699 dataset. Heatmaps of differential gene expression in (A) tumor samples from obese vs. non-obese patients and (B) tissue samples from patients with vs. without ccRCC. Volcano plots of gene expression in (C) tumor samples from obese vs. non-obese patients with ccRCC and (D) tissue samples from patients with vs. without ccRCC. (E) Venn diagram of commonly DEGs in the GSE46699 dataset. ccRCC, clear cell renal cell carcinoma.
Figure 2 GO analysis and KEGG pathway enrichment. GO analysis of (A) obesity-related DEGs and (B) ccRCC-related DEGs. KEGG pathway enrichment of (C) obesity-related and (D) ccRCC-related DEGs. GO, Gene Ontology; DEGs, differentially expressed genes; KEGG, Kyoto Encyclopedia of Genes and Genomes; ccRCC, clear cell renal cell carcinoma.
Figure 3 Hierarchical clustering analysis of ccRCC-related DEGs. Hierarchical clustering of ccRCC-related DEG expression in relation to (A) haptoglobin binding, (B) haptoglobin-hemoglobin complex, (C) hemoglobin complex, (D) hydrogen peroxide catabolic process, and (E) oxygen transport. ccRCC, clear cell renal cell carcinoma; DEGs, differentially expressed genes.

i.e., *IGHA1*, *IGKC*, *MAOA*, *MUC20* and *TRPM3*. *IGHA1* (immunoglobulin heavy constant alpha 1) encodes the N-terminal sequence of the constant region the IgA1 isoform of immunoglobulin IgA. In previous studies, *IGHA1* was reported to be a prognostic biomarker for better outcomes in patients with triple-negative breast cancer (21,22). In another study, it was reported to also have a relationship with altered adipose tissue homeostasis in prostate cancer and could predict high-risk cases (23). *IGKC* (immunoglobulin κ constant) is a polymorphic gene which is characterized by the segregation of three Km alleles—Km1, Km2, and Km3—on chromosome 2p12 (24).
**Figure 4** Expression levels of the DEGs in normal para-cancerous kidney and ccRCC tumor-associated tissues in GEPIA database. Gene expression levels of (A) CCL18, (B) COL8A1, (C) HS3ST2, (D) IGHAI, (E) IGKC, (F) IGSF6, (G) MS4A4A and (H) S100A8 were significantly upregulated in ccRCC compared with normal tissues. Gene expression levels of (I) MAOA, (J) MUC20, (K) PP7080 and (L) TRPM3 were significantly downregulated in ccRCC compared with normal tissues. GEPIA, Gene Expression Profiling Interactive Analysis. DEGs, differentially expressed genes; ccRCC, clear cell renal cell carcinoma. *, P<0.05.
Figure 5 Overall survival analysis of key genes in ccRCC (based on TCGA data in GEPIA). Expression levels of (A) IGHA1, (B) IGKC, (C) MAOA, (D) MUC20 and (E) TRPM3 were significantly correlated, either negatively or positively, with the overall survival of patients with ccRCC. GEPIA, Gene Expression Profiling Interactive Analysis; ccRCC, clear cell renal cell carcinoma; TCGA, The Cancer Genome Atlas.
Figure 6 Gene expression levels of key genes at progressive stages of ccRCC (based on TCGA data in GEPIA) Expression levels of (A) IGHA1, (B) IGKC, (C) MAOA, (D) TRPM3 and (E) MUC20 were significantly correlated, either positively or negatively, to specific stages of ccRCC. GEPIA, Gene Expression Profiling Interactive Analysis.
IGKC has been recognized as an effective prognostic biomarker for several human malignancies (25), and most notably for breast cancer (26,27). MAOA (monoamine oxidase A) is a mitochondrial enzyme, which degrades monoamine neurotransmitters and dietary amines, and produces H₂O₂ (28). Its biological roles in prostate cancer have been studied, including its participation in epithelial-mesenchymal transition, tumor hypoxia, and oxidative stress, which are related to high-grade Gleason score prostate cancer as well as progression and metastasis (29-31). Additionally, MAOA is involved in mediating neuroendocrine differentiation of prostate cancer cells, a feature associated with hormone-refractory prostate cancer (32). Mucin-20, encoded by MUC20, is a member of the mucin family, a group of heavily glycosylated proteins expressed by a variety of epithelial cells. Previous literature reported that the mucin family plays a critical role in a variety of malignancies, and family members have also been viewed as potential diagnostic markers and therapeutic targets. Chen et al. reported that Mucin-20 promotes aggressive and metastatic behavior of epithelial ovarian cancer by activating the integrin beta 1 signaling pathway (33). A similar result was found in colorectal cancer by Xiao et al. and in pancreatic duct carcinoma by Chen et al. (34,35). The biological function of TRPM3 (transient receptor potential melastatin 3) in ccRCC has been previously studied by Hall et al. who showed that an increase in TRPM3 expression in ccRCC correlated with Ca²⁺ influx, activation of CAMKK2, AMPK, and ULK1, and phagophore formation, which ultimately leads to tumor progression (36). However, in our present study, TRPM3 appears to be a tumor suppressor gene.

The present study identified five key genes involved in ccRCC progression that are also related to obesity, i.e., IGHA1 and IGKC as oncogenes and MAOA, MUC20 and TRPM3 as tumor suppressor genes. This study has the following limitations. First, the data used were downloaded from a publicly available database, and we did not validate the findings with in vivo or in vitro experiments. Second, the study was a prospective study, and we did not explore the underling mechanisms of the genes linking obesity and ccRCC. Thus, a multicenter prospective study will be conducted to evaluate the possible applications of these key genes to predict disease outcomes, and further study will focus on the related mechanisms of these genes.

In conclusion, with the help of several bioinformatics tools, the current study identified five obesity-associated genes that correlate with the progression and prognosis of ccRCC and could provide prognostic value in the treatment of ccRCC.

Acknowledgments

Funding: This study was supported by the Three-Year-Action-Project (No. 16CR3062B) of Shanghai Shen Kang Hospital Development Center, the Medical-Engineering cross fund of Shanghai Jiao Tong University (No. YG2017QN46), and the National Natural Science Foundation of China (No. 81972369).

Footnote

Reporting Checklist: The authors have completed the STREGA reporting checklist. Available at http://dx.doi.org/10.21037/tau-20-891

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/tau-20-891). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and the Harmonized Tripartite Guideline for Good Clinical Practice from the International Conference on Harmonization. The study was approved by the research medical ethics committee of Shanghai Renji Hospital [Ethics board approval number: Renji Ethics [2017]091].

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References


