Single-Cell Sequencing to Identify Six Heat Shock Protein (HSP) Genes-Mediated Progression Subtypes of Clear Cell Renal Cell Carcinoma

Qinke Li
Maoqing Lu
Zhechuan Zhang
Ronggui Zhang

Department of Urology, The Second Affiliated Hospital of Chongqing Medical University, Chongqing, People’s Republic of China

Background: Heat shock proteins (HSPs) are widely involved in tumor occurrence and development and are prognostic markers for multiple tumors. However, the role of HSPs in clear cell renal cell carcinoma (ccRCC) remains unclear.

Methods: We used Cytoscape to identify hub genes in the ccRCC single-cell sequencing data set from the Gene Expression Omnibus (GEO) data repository. We identified subtypes, C1 and C2, of The Cancer Genome Atlas (TCGA) patients based on the expression of hub genes using unsupervised consensus clustering. Principal component analysis (PCA) was used to verify the clustering differences, and Kaplan–Meier (K-M) estimate was used to verify the survival differences between C1 and C2 patients. We used TIMER 2.0 and CIBERSORT to evaluate the immune cell infiltration of HSP genes and C1 and C2 patients. The R package “pRRophetic” was used to evaluate the sensitivity in C1 and C2 patients to the four first-line treatment drugs.

Results: We identified six hub genes (HSP90AA1, HSPH1, HSPA1B, HSPA8, and HSPA1A) encoding HSP, five of which were significantly downregulated in TCGA group, and four had a protective effect on prognosis (p <0.05). Survival analysis showed that C1 patients had a better overall survival (p <0.001). TIMER 2.0 analysis showed that three HSP genes were significantly correlated with the infiltration of CD4+ T cells and CD4+ Th1 cells (|cor|>0.5, p<0.001). CIBERSORT showed significant differences in multiple infiltrating immune cells between C1 and C2 patients. Meanwhile, the expression of PD1 was significantly lower in C1 patients than in C2 patients, and the expression of PDL1 is the another way around. Drug sensitivity analysis showed that C1 patients were more sensitive to sorafenib, pazopanib, and axitinib (p <0.001).

Conclusion: Our research revealed two molecular subtypes of ccRCC based on 6 HSP genes, and revealed significant differences between the two subtypes in terms of clinical prognosis, immune infiltration, and drug sensitivity.

Keywords: ccRCC, bioinformatics, immune, molecular subtypes, prognosis

Introduction

Kidney cancer affects nearly 300,000 individuals worldwide each year and is responsible for over 100,000 deaths annually.1,2 Clear cell renal cell carcinoma (ccRCC) is the most common RCC subtype, accounting for approximately 80% of all cases.3–5 The mortality rate of kidney cancer depends largely on the disease progression. According to the classification criteria of the eighth edition of the American Joint Committee on Cancer (AJCC), the five-year survival rate for stage
I/II is approximately 93%, that for regional renal cancer (stage III) is approximately 70%, while that for metastatic renal cancer (stage IV) is only 13%. In terms of prognosis assessment, the five-year recurrence-free survival rate of patients with stage I disease is greater than 92%, while the risk of recurrence for patients with stage II and III diseases is as high as 40%. As such, the treatment of advanced renal cancer still faces a huge challenge.

Heat shock proteins (HSPs) are highly conserved chaperone molecules that are synthesized in large quantities to protect cells from damage when the body is in various stress states, such as high temperature, cold, mechanical damage, ischemia, or radiation. According to their relative molecular weights, HSPs are divided into six families-HSP27, HSP40, HSP60, HSP70, HSP90, and the large HSP family (HSP110 and glucose regulatory protein 170, (GRP170)). HSPs are involved in various cellular processes, including protein assembly, secretion, transportation, protein degradation, and transcription factor regulation, mainly to prevent protein misfolding and accelerate protein refolding. Previous studies have shown that the expression of HSPs is elevated in a variety of cancers and is associated with the proliferation, metastasis, apoptosis, and invasion of tumor cells and can also be a biomarker for certain cancers. In terms of immune regulation, HSPs can regulate the production of a series of cytokines, including TNF-α, IL-6, IL-10, and IL-12, which can further mediate the immune regulation of tumors. Extracellular HSPs can bind to the characteristic receptors on dendritic cells, affecting the delicate balance of immune regulation in the tumor microenvironment. HSPs can bind to antigen peptides on the surface of cancer cells, making them viable targets in cancer immunotherapy.

Next-generation sequencing (NGS) has promoted in-depth research in the fields of tumor Intratumor heterogeneity (ITH) and tumor somatic mutations in the past ten years. However, traditional RNA sequencing only provides the average gene expression of various types of cells in the tissue, so it is hard to peek the transcriptome characteristics of the cell subsets in the tumor. Single-cell sequencing (scRNA-seq) is a high-throughput experimental technology for quantifying gene expression profiles of cell populations at the single-cell level. Therefore, scRNA-seq can analyze the gene expression or epigenetic modification of the cells that make up the tumor microenvironment with higher sensitivity, and provides a higher resolution scan for the description of ITH.

In this study, we obtained scRNA-seq data related to metastatic ccRCC from the Gene Expression Omnibus (GEO) data repository to determine more precise molecular changes that mediate the progression of ccRCC. Besides, we used an independent patient cohort from The Cancer Genome Atlas (TCGA) to identify the potential functions, clinical prognosis, and therapeutic potential of these molecules. Our research provides new ideas for the mechanism of HSP molecules in ccRCC.

**Materials and Methods**

**Sample Collection**

Single-cell sequencing samples of 46 primary ccRCC (PDX-pRCC) and 36 lung metastatic ccRCC (PDX-mRCC) cases were obtained from the GSE73121 dataset of the GEO (https://www.ncbi.nlm.nih.gov/geo/) data repository. The R “hclust” algorithm was used to cluster the samples hierarchically, and outliers were removed. The principal component analysis (PCA) was used to test the clustering effect. Finally, 44 pRCC and 27 mRCC samples were included for further analysis (Figure S1 shows the PCA results of pRCC and mRCC). The transcriptome expression data (Count) and sample annotation files of 539 patients were obtained from TCGA database (https://portal.gdc.cancer.gov/). We manually screened out the patients who were officially recommended not to be included in the analysis (n = 19, “Do not use” = “TURE”) as well as those without follow-up information (n = 11). Finally, 509 patients with ccRCC were included in the follow-up analysis. Meanwhile, we converted all “Count” expression data into “TPM” to correct the impact of transcription sequencing depth and gene length and used “log2 (TPM + 1)” for clustering and correlation analyses to reduce the bias caused by excessive expression values.

**Difference Analysis and HUB Gene Identification**

The R package “DESeq2” was used to analyze the differentially expressed genes (DEGs) between mRCC and pRCC (screening criteria were |LogFC| >2 and FDR <0.01). The R package “enrichplot” was used to perform GO and KEGG enrichment analyses of the selected DEGs. The online website “STRING” was used to construct a protein-protein interaction (PPI) network for differential protein-coding genes, and the
CytoHubba\textsuperscript{36} plug-in of Cytoscape 3.7.2\textsuperscript{37} was used to screen important nodes and identify hub genes in the network.

**Identification of Molecular Subtypes**
Through the R package “ConsensusClusterPlus”,\textsuperscript{38} based on the expression of six HUB genes, unsupervised consensus clustering (kmax=9) was performed for 509 patients in TCGA cohort, and the k=2 was taken as the best cluster (Figure S2 shows the clustering results and Table 1 shows the clinical information associated with the C1 and C2 cohorts). PCA analysis was performed on the whole transcriptome data of 509 patients (Number of gene = 55,268) to clarify the validity of clustering C1 and C2.

**Prognostic Evaluation**
According to the median value of HSP gene expression as the intercept, the patients were divided into a high expression group and a low expression group. The R package “survival”\textsuperscript{39} was used to assess the overall survival difference between the two groups and between C1 and C2 patients.

**Immune Correlation Assessment**
TIMER 2.0\textsuperscript{40} was used to evaluate the correlation between the expression level of six HSP genes and the abundance immune cell infiltration in ccRCC patients. CIBERSOFT,\textsuperscript{41} a deconvolution algorithm based on gene expression, which can infer 22 types of human immune cell, was used to analyze immune cell infiltration in C1 and C2 patients (p<0.05 is considered a valid result). The T test was used to evaluate the differences in the expression of immune checkpoint inhibitors PD1, PDL1 and CTLA4 in C1 and C2 patients.

**Drug Susceptibility Prediction**
The R package “pRRophetic”\textsuperscript{42} was used to evaluate the sensitivity of C1 and C2 patients to four advanced ccRCC first-line drugs, and the algorithm evaluated the maximum inhibitory concentration (IC50) and prediction accuracy through ridge regression and 10-fold cross-validation of the training set of the Dependent Cancer Drug Sensitivity Genomics (GDSC) database (https://www.cancerrxgene.org/).

**Results**

**Gene Enrichment Analysis and HUB Gene Identification**
We used the R package “DEseq2” to evaluate the differences in single-cell transcriptome levels between 46 cases of PDX-pRCC and 36 cases of lung metastasis PDX-mRCC,
and finally identified 494 DEGs, of which 173 were downregulated and 321 were upregulated (The specific results are shown in Table S1). Subsequently, we evaluated the possible pathways involved in the expression of the DEGs through GO and KEGG enrichment analyses. Figure 1A shows the results of GO enrichment analysis (adj. p < 0.05). Interestingly, multiple pathways related to protein folding were significantly enriched. Among them, biological processes (BP) are mainly enriched in protein folding, response to unfolded protein and chaperone-mediated protein folding.

Figure 1A Gene Ontology (GO) analysis of differentially expressed genes (DEGs) in pRCC and mRCC samples. (B) Circle graphs of the first five pathways and the DEGs involved in the biological process of GO. (C) The six HSP-related HUB genes and their interaction network with DEGs were obtained by Cytoscape. Abbreviations: BP, biological process; CC, cellular component; MF, molecular function.
pathways. The cellular component (CC) is mainly enriched in inclusion body, ficolin−1−rich granule, U2−type precatalytic spliceosome pathways. Molecular function (MF) is mainly enriched in protein folding chaperone, MHC protein complex binding, unfolded protein binding and misfolded protein binding pathways. This suggests that in metastatic ccRCC cells, protein folding-related pathways may be widely activated or silenced. Figure 1B shows the association of DEGs in the five most significant enrichment pathways (lysosomal transport, positive regulation of cytokine production, protein folding, response to unfolded protein and vacuolar transport).

Subsequently, we conducted PPI analysis on the protein-coding genes of DEGs using STRING, and we used the various algorithms built into Cytoscape to identify the most closely linked key genes (hub genes). Interestingly, the enrichment results of multiple algorithms showed that the 6 genes (HSPA1A, HSPA1B, HSPA4L, HSPA8, HSP90AA1 and HSPH1) of the HPS family all showed significant connections with other DEGs. These genes are widely involved in the biological processes related to protein folding. As such, we speculate that these HPS genes may play an essential role in ccRCC. Figure 1C shows the interaction network between six HPS family genes and DEGs.

Prognosis Analysis of the Six HSP Genes
To further confirm the role of these six HSP genes in ccRCC, we analyzed the expression data of 509 ccRCC patients in the TCGA cohort. Compared with those in the normal group, HSP90AA1, HSPH1, HSPA1B, HSPA8 and HSPA1A genes were significantly downregulated in the tumor group (p<0.01, Figure 2A). Furthermore, the expression of HSP90AA1, HSPA4L and HSPA8 were significantly downregulated in patients at an advanced stage (Stage III and IV) than that in patients at an early stage (stage I and II) (p <0.01, Figure 2B). Meanwhile, the K-M survival analysis showed that low expression of HSP90AA1, HSPA8, HSPA1A and HSPA4L were associated with poor prognosis (p<0.05, Figure 2C).

Identification of ccRCC Molecular Subtypes
To identify the presence of molecular subtypes related to HSP in ccRCC patients, we performed unsupervised consensus clustering of 509 ccRCC patients based on the expression of the six HSP genes using the R package “ConsensusClusterPlus”, and successfully divided the patients into two Subtypes-C1 and C2 (k = 2, Figure 3A). Significantly, the expression of HSP90AA1, HSPA8, HSPA1A, and HSPA4L genes in C2 patients were remarkably downregulated, while HSPA1A and HSPA1B seemed to have a low contribution to clustering (Figure 3B). Besides, we conducted PCA on patients with expression information at the whole gene level (n = 55,268) to further evaluate the effectiveness of clustering. The results revealed that the transcriptome expression patterns of patients with C1 and C2 showed significant differences, indicating that clustering based on the six genes was feasible and effective (Figure 3C). We used a heat map to evaluate the difference between the C1 and C2 patients with the expression levels of the top 5000 genes, and found that the transcription profiles of the two types of patients were significantly different (Figure S3). Subsequently, we conducted a survival analysis of the patients with C1 and C2. The prognosis of the two subtypes was significantly different (p<0.0001), with a risk ratio of 1.909 (95% CI = 1.351–2.697) for C2/C1 (Figure 3D).

Evaluation of Immune Response of Six HSP Genes and ccRCC Subtypes
Given the close relationship between HSPs and immune responses, we further evaluated the correlation between the six HSP genes and immune cell infiltration. Using TIMER 2.0, we found that the expression of the six HSP genes in ccRCC was significantly related to CD4+ T cells, CD8+ T cells, B cells, and macrophages (Figure 4A). Remarkably, HSP90AA1, HSPA4L, and HSPA8 were significantly positively correlated with CD4+ T cells (Pho >0.4, p <0.0001), and significantly negatively correlated with CD4+ Th1 cells (Pho <−0.4, p <0.0001) (Figure 4B). We used CIBERSORT to evaluate the differences in immune cell infiltration between C1 and C2 patients. Plasma cells, CD8+ T cells, follicular helper T cells, Tregs, and M0 macrophages showed lower infiltration levels in C1 patients than in C2 patients (p <0.05), while M1 and M2 macrophages, resting dendritic cells, and resting mast cells showed higher infiltration levels (p <0.05) (Figure 4C). Furthermore, we evaluated the differences in the expression of immune checkpoint protein programmed cell death 1 (PD1/PDCD1), programmed death-ligand 1 (PD_L1/CD274), and cytotoxic T-lymphocyte associated protein 4 (CTLA4) in C1 and C2 patients. The expression of PD1 in C1 patients was significantly lower than that in C1 patients, while the content of PD_L1 was significantly higher than that in
C2 patients \((p < 0.001)\) (Figure 4D). In summary, the subtypes of kidney cancer mediated by six HSP genes showed significant differences in tumor immune response.

**Prediction of Drug Sensitivity for ccRCC Subtypes**

Given that members of the HSP family are widely involved in resistance to chemotherapeutics, we evaluated the sensitivity of C1 and C2 patients to the first-line treatment drugs sorafenib, sunitinib, pazopanib, and axitinib, using the R package “pRRophetic”. Interestingly, C2 patients were more sensitive to sunitinib (Figure 5A), and C1 patients were more sensitive to sorafenib, pazopanib, and axitinib (Figure 5B–D). These results further indicated that patients of different subtypes may have more significant differences in sensitivity to different chemotherapeutic drugs.

**Discussion**

The accurate definition of tumor subtypes based on sequencing results is very important for precise treatment.
of patients. Previous studies have identified multiple tumor subtypes of ccRCC, such as 3 subtypes based on tumor suppressor genes, or 4 stratification based on transcription factor activity. These subtypes reveal part of the transcriptome characteristics of ccRCC and have guiding significance for the treatment of patients. It is worth noting that the tumor tissues derived from sequencing include many different types of cells, including immune cells. Therefore, traditional transcriptome sequencing is difficult to describe the evolutionary characteristics of different subgroups of tumor cells. In this study, we used published scRNA-seq data to obtain a higher-resolution understanding of molecular

Figure 3 (A) Unsupervised consistent clustering analysis to obtain the best clustering results for patients in TCGA cohort (k = 2). (B) Heat map of expression of the six genes in Cluster 1 (C1) and Cluster 2 (C2). (C) The PCA clustering results of patients in TCGA cohort are colored according to the consistent clustering results (PCA analysis is based on genome-wide genes, and the mean value >0.5 was used as the criterion for gene inclusion). (D) Survival analysis of C1 and C2 patients.
Figure 4 (A) TIMER 2.0 was used to evaluate the association between the six HSP genes and CD4+ T cells, CD8+ T cells, macrophages, and B cells in ccRCC. (B) The correlation between HSP90AA1, HSPA4L, and HSPA8 and CD4+ T cells and TH1 CD4+ T cells. (C) The difference in immune cell infiltration between C1 and C2 patients using CIBERSORT (p < 0.05). Differences in the expression of three immune checkpoints (ICPs) in C1 and C2 patients.
changes inside metastatic ccRCC cells. Through differential and enrichment analyses, we identified six hub genes encoding heat shock proteins. Among them, HSPA1A, HSPA1B, and HSPA8 encode HSP70. HSP90AA1 encodes HSP90, and HSPA4L and HSPH1 encode HSP110.13

Specifically, HSPA1A and HSPA1B are paralogous genes located in the major histocompatibility complex (MHC) class III region on the short arm of chromosome 6, and code for the classical HSP70 protein.48,49 Previous studies have shown that HSP70 is overexpressed in liver, prostate, colorectal, lung, and cervical cancers.50 In human acute leukemia cells, Hsp70 can bind to death receptors 4 and 5 (DR4 and DR5), inhibit Apo-2L/TRAIL-induced assembly and death-inducing signaling complex (DISC) activity, further inhibiting apoptosis.51 In addition, Gabai et al demonstrated that HSP70 (HSPA1) could inhibit oncogene induction through two separate signal transduction pathways (the PI3K-mediated p53-dependent pathway or RAS/ERK-mediated p53-independent pathway) and cell senescence.52 However, HSP70 is downregulated in kidney cancer cells.53 Ramp et al analyzed the expression of HSP70 in 145 patients with renal cell carcinoma. In contrast to healthy cells, HSP70 in renal cancer cells was

Figure 5 Differences in response sensitivity of C1 and C2 patients to sunitinib (A), sorafenib (B), pazopanib (C), and axitinib (D).
significantly downregulated in both the cytoplasm and nucleus \( (p = 0.0176) \), both in well-differentiated \((G1)\) and poorly differentiated cells.\(^{53}\)

The HSPA8 gene, located on chromosome 11q23.3–25, encodes the structural HSP70 protein HSC70, which has 85% homology with HSPA1A/B.\(^{54,55}\) HSC70 regulates cell signal transduction in normal cells and can mediate the migration and angiogenesis of endothelial cells induced by VEGF-induced Akt phosphorylation.\(^{56,57}\) Previous studies have shown that HSC70 is overexpressed in human colon cancer and glioma, and may be a poor prognostic indicator.\(^{58,59}\) Azuma et al found that HSC70 can induce cytotoxic T lymphocyte (CTL) responses by binding and carrying anti-cancer peptides.\(^{60}\) In contrast, Yehiely et al showed that HSC70 could bind to mutants of the tumor suppressor protein p53 and inhibit oncogene-mediated transformation.\(^{61}\)

HSP90AA1 is located on the complement chain of chromosome 14q32.33, and encodes two different mRNA transcripts, Hsp90α TV1 and TV2.\(^{62,63}\) Elevated Hsp90α levels have previously been observed in leukemia, breast cancer, and pancreatic cancer.\(^{64,65}\) Teng et al found that the proto-oncogene MYC can bind to the transcription start site \((TSS)\) at the proximal end of the HSP90AA1 gene and induce its expression.\(^{66}\) Perotti et al observed that in breast cancer cells, the growth hormone prolactin induces the expression of HSP90AA1 through STAT5.

Although HSPA4L (heat shock protein family A \((\text{Hsp70})\) member 4 like) is homologous to HSPA4, it is a member of HSP110. HSPA4L was found to be expressed mainly in human testicular germ cells and sperm.\(^{67}\) HSPA4L was found to be highly expressed in leukemia cells. Hosaka et al showed that HSPA4L might be involved in the resistance of cancer cells to apoptosis.\(^{58}\)

HSPH1 is a member of the HSP110 family, and the encoded protein acts as the nucleotide exchange factor for HSPA1A/B, promoting the release of ADP from HSPA1A/B.\(^{59}\) Zappasodi et al showed that HSPH1 expression is elevated in aggressive human B-cell non-Hodgkin lymphoma \((\text{B-NHL})\) and positively correlated with c-Myc or Bcl-6 expression.\(^{70}\) Zhang et al confirmed that elevated HSPH1 levels might mediate tamoxifen resistance.\(^{71}\)

Although these six HSP genes and their encoded proteins have been shown to promote tumor development in many studies, their role in ccRCC is unknown, and their direct impact on ccRCC is still unclear.

In this study, besides HSPA4L, the other five genes were seen to be downregulated in ccRCC tissues. In addition, HSPA4L, HSP90AA1, and HSPA8 were downregulated in patients with stages III and IV, suggesting that they may be involved in tumor progression. Based on the expression of HSP genes, obtaining molecular subtypes through consistent clustering seems to be effective. The expression of HSPA8, HSP90AA1, HSPA4L, and HSPH1 in C1 patients was much higher than that in C2 patients, and their survival was also better. As such, our research seems to strengthen the evidence that HSP is downregulated in kidney cancer. In addition, 6 HSPs are not only significantly associated with CD4+ T cells and CD4+ Th1 cells, but the expression of immune checkpoints \((\text{ICPs, PD1 and PDL1})\) in C1 and C2 subtypes is also significantly different, suggesting that HSPs may be potential immunotherapy maker. Furthermore, we used drug sensitivity prediction and found that C2 patients were less sensitive to a wider range of drugs. This may be the underlying reason for the poor prognosis. However, it is worthwhile to further study the specific mechanisms of HSPs directly or indirectly responding to ccRCC immunotherapy and drug sensitivity. Although our study obtained a considerable amount of positive results, it still has limitations, including the lack of in vitro and in vivo studies and external verification of larger sample sizes. Our work requires independent, clinical prospective studies.

**Conclusion**

In summary, we identified six heat shock protein-encoding genes that may have protective functions in ccRCC using single-cell sequencing data. We identified two ccRCC subtypes based on the expression of six genes, and there were significant differences in the prognosis, immune cell infiltration, and drug sensitivity between the two subtypes. Our research provides a potential direction for studying the mechanism of these six HSP molecules in mediating immunity and drug sensitivity in ccRCC.

**Abbreviations**

AJCC, American Joint Committee on Cancer; B-NHL, B-cell non-Hodgkin lymphoma; BP, biological processes; CC, cellular component; ccRCC, clear cell renal cell carcinoma; CTL, cytotoxic T lymphocyte; CTLA4, cytotoxic T-lymphocyte associated protein 4; DISC, death-inducing signaling complex; DR4, death receptors 4; DR5, death receptors 5; GEO, Gene Expression Omnibus; HSP, Heat shock protein; MF, Molecular function; MHC, major histocompatibility complex; PCA, Principal component analysis; PD1, programmed cell death 1; PD_L1, programmed...
death-ligand 1; TCGA, The Cancer Genome Atlas; TSS, transcription start site.

**Data Sharing Statement**
All data can be obtained from the corresponding author in a reasonable reasons.

**Ethics**
The original data required for this research were collected from public databases, so no medical ethics review is required.

**Consent for Publication**
All authors approved publication.

**Acknowledgments**
We are grateful to the uploader of the original data and the public database.

**Author Contributions**
All authors contributed to data analysis, drafting or revising the article, gave final approval of the version to be published, and agree to be accountable for all aspects of the work.

**Funding**
The study was sponsored by Research on the Frontiers and Application of Chongqing Science and Technology, Commission No. cstc2015jcyjA10030.

**Disclosure**
The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**References**


