Screening of renal clear cell carcinoma prognostic marker genes based on TCGA and GTEx chip data and construction of transcription factor-related regulatory networks

Wei Zhu, Lingfeng Wu, Wenhua Xie, Gaoyue Zhang, Yanqin Gu, Yansong Hou, Yi He [*]

Department of Urology, The First Hospital of Jiaxing & The Affiliated Hospital of Jiaxing University, Jiaxing 314000, PR China

A R T I C L E   I N F O

A B S T R A C T

This study aimed to identify prognostic marker genes for renal clear cell carcinoma (RCCC) and construct a regulatory network of transcription factors and prognostic marker genes. Three hundred eighty-six genes were significantly differentially expressed in RCCC, with functional enrichment analysis suggesting a relationship between these genes and kidney function and development. Cox and Lasso regression analyses revealed 10 prognostic marker genes (RNASET2, MSC, DPEP1, FGF1, ATP1A1, CLDN10, PLG, SLC44A1, PCSK1N, and LGI4) that accurately predicted RCCC patient prognosis. Upstream transcription factors of these genes were also identified, and in vitro experiments suggested that ATP1A1 may play a key role in RCCC patient prognosis. The findings of this study provide important insights into the molecular mechanisms of RCCC and may have implications for personalized treatment strategies.

1. Introduction

Renal clear cell carcinoma (RCCC) is the most common type of kidney cancer, accounting for approximately 70–80% of all kidney cancers [1]. Its diagnosis mainly relies on histological and imaging examinations, including renal ultrasound, CT, and MRI [2]. For early-stage RCCC patients, partial or total nephrectomy is the main treatment option, while molecular targeted therapy becomes an important treatment approach for advanced or metastatic RCCC patients [3]. The incidence and mortality of RCCC are increasing globally, severely affecting people's life and health [4].

The clinical symptoms of RCCC are not specific and are often ignored or misdiagnosed [5]. Therefore, it is significant to find biomarkers that can accurately predict the prognosis of RCCC patients [6]. Recently, with the continuous development of genomics and bioinformatics technologies, many studies have shown that prognostic biomarkers can provide clinicians with the basis for prognosis evaluation, thereby guiding patient treatment options and prognosis management [7–9]. Therefore, it is clinically significant to identify and study prognostic biomarkers of RCCC [10].

Researchers have recently identified a series of molecular markers related to RCCC prognosis through genomics and transcriptomics methods [11]. Ki67, a proliferation marker, is important in RCCC prognosis evaluation [12]. Previous studies have found

---

* Corresponding author. Department of Urology, The First Hospital of Jiaxing & The Affiliated Hospital of Jiaxing University, No. 1882, Zhonghuan South Road, Nanhu District, Jiaxing 314000, Zhejiang Province, PR China.

E-mail address: heyi@zjxu.edu.cn (Y. He).

that high Ki67 expression levels are associated with pathological grade, tumor size, lymph node metastasis, and survival rate in RCCC [13]. The VHL gene is one of the most common mutated genes in RCCC and an important regulatory factor in RCCC pathogenesis [14]. Studies have shown that VHL gene mutations are closely related to RCCC prognosis, and its mutation leads to disease progression and poor prognosis [15]. HIF-1α is an important regulatory gene that regulates tumor metabolism and angiogenesis [16]. Multiple studies have found that high expression of HIF-1α is associated with high-grade and malignant pathological features of RCCC and is also related to poor prognosis [17–19]. Survivin is an important anti-apoptotic factor that participates in the growth and proliferation of tumor cells [20]. Many studies have shown that high expression of Survivin is associated with a poor prognosis of RCCC [21–23].

Despite the discovery of many molecular biomarkers associated with the prognosis of renal clear cell carcinoma (RCCC) patients, the accuracy and precision of predicting RCCC patient prognosis remain limited [24–26]. It is because the biological characteristics of RCCC are complex, and its prognosis is not only influenced by tumor molecular features but also by several other factors, such as patient age, gender, disease course, and treatment methods [27]. Additionally, different studies have varied in selecting prognostic markers and analysis methods, leading to inconsistent results [28]. Therefore, more large-scale studies are needed to identify the optimal combination of prognostic markers to improve the accurate prediction of RCCC patient prognosis [29].

This study aimed to screen the prognostic markers for RCCC using gene differential expression analysis, functional enrichment analysis, single-factor Cox analysis, Lasso regression analysis, and multiple-factor Cox analysis. Additionally, the study aimed to construct a transcription factor-prognostic marker gene regulatory network and validate the effect of key markers on the biological characteristics of RCCC cells. This study's scientific and clinical significance lies in identifying potential key prognostic markers that may provide new targets for the individualized treatment of RCCC. Furthermore, by constructing the transcription factor-prognostic marker gene regulatory network, this study has provided insights into the molecular mechanism and interaction of prognostic markers in RCCC, which may open up new directions and ideas for future research. Additionally, the prognostic risk evaluation model developed in this study can provide clinical physicians with valuable references to better predict the prognosis of RCCC patients and guide clinical decision-making.

2. Materials and methods

2.1. Data download

TCGA-RCCC and GTEx expression data were downloaded from the UCSC Xena database (https://xena.ucsc.edu/). Additionally, TCGARCCC sample phenotype data and clinical prognosis data were downloaded. The TCGA data comprised 72 normal kidney tissue samples and 535 RCCC samples, while the GTEx data included 28 normal kidney tissue samples. All data types were FPKM data. Chip annotation information was obtained from the Gencode database (https://www.gencodegenes.org/human/). Perl was used to perform ID conversion on the TCGA and GTEx data and merged the two datasets, retaining only genes annotated in both TCGA and GTEx. After merging, 100 normal kidney tissue samples and 535 RCCC samples were included.

2.2. Differential gene analysis

Differential analysis was performed on the merged TCGA and GTEx data. Using normal kidney tissue samples as a control, differential analysis was performed using the "limma" package in R, with FDR correction applied to the differential p-values. Significant differentially expressed genes were identified using a filtering threshold of $|logFC|>2$ and $FDR<0.05$. The expression values of significant differentially expressed genes were extracted, and a differential gene expression heatmap was constructed using the "pheatmap" package in R. Next, the "clusterprofiler" package in R was used to perform GO and KEGG functional enrichment analysis on the differentially expressed genes [30].

2.3. Univariate Cox analysis

The expression values of all differentially expressed genes were extracted and merged with RCCC patient survival time and survival status data. Univariate Cox analysis was performed using the "Survival" package in R, and the hazard ratio and significant p-value were calculated for each gene. A dendrogram of univariate Cox analysis was constructed for the top 50 genes with $p < 0.05$ [31].

2.4. Lasso regression analysis

The lasso regression model was constructed using the "glmnet" package in R (https://CRAN.R-project.org/package=glmnet) with the glmnet (.) and cv.glmnet (.) functions and the "Survival" package Surv (.) function. Each gene was repeatedly calculated during this process, and the number of genes was progressively reduced to decrease the error value. The coefficients of each gene were calculated at different $\log(\lambda)$ values, and when the value was 0, the gene was excluded. By repeating the calculation, the candidate gene number and corresponding $\log(\lambda)$ value with the minimum error were obtained [32].

2.5. Survival analysis

Perl was used to extract the expression data of significant differentially expressed genes in RCCC samples and merge it with TCGARCCC clinical prognosis data for survival analysis. To improve the accuracy of survival analysis, samples with survival times less

than 90 days were excluded. A total of 497 samples were retained for survival analysis. The "Survival" package in R was used to construct survival curves and calculate p-values, with $p < 0.05$ considered significant [31].

2.6. Prognostic risk evaluation model construction

Multivariate Cox analysis was performed on the candidate genes retained after Lasso regression analysis. The candidate gene expression values, sample survival time, and survival status were merged, and samples with a survival time greater than 90 days were retained, resulting in 497 samples. The "Survival" package in R was then used to perform a multivariate Cox analysis and calculate the risk value of each RCCC patient, constructing the prognostic risk evaluation model. The "Survival" and "Survminer" packages in R were used to plot the survival curve of the prognostic risk evaluation model and calculate the p-value. The "Survival" package was also used to perform ROC analysis on the prognostic risk evaluation model and calculate the AUC value. ROC analysis judged the potential of candidate genes as marker genes, with a higher AUC value indicating a higher potential for the gene as a marker gene [33].

2.7. Transcription factor-prognostic marker gene regulatory network prediction

Human transcription factor information was extracted from the Cistrome database (http://cistrome.org/). Then, the transcription factors in all significantly differentially expressed genes were extracted to obtain the expression information of differentially expressed transcription factor genes in RCCC. Based on the expression levels of transcription factors and candidate genes, the correlation between each transcription factor and each immune gene was analyzed, and the Pearson correlation analysis was used as the screening standard, setting the correlation coefficient cor >0.1 and the correlation $p < 0.05$. Transcription factors negatively correlated with immune genes were considered negative regulators of immune genes, while those positively correlated were considered positive regulators. The transcription factor-prognostic marker gene regulatory network was constructed based on the correlation analysis results using Cytoscape v3.7.1 software [34,35].

2.8. Cell transfection

Human normal renal proximal tubule epithelial cells (RPTEC) were purchased from ATCC (PCS-400-010, ATCC, USA), and human RCCC cell lines (Caki-1 and 786-O) were purchased with catalog numbers HTB-46 and CRL-1932, respectively, from ATCC (USA). The cell lines were cultured in RPMI 1640 medium (GIBCO, Carlsbad, USA) containing 10% fetal bovine serum and incubated in a 37 °C, 5% CO_2 cell culture incubator with a mixture of penicillin (100U/ml) and streptomycin (0.1 mg/ml).

Logarithmically growing Caki-1 and 786-O cells were collected and seeded in 96-well plates with 1×10^5 cells per well. After 24 h of routine culture, the cells were transfected according to the instructions of Lipofectamine 3000 (Invitrogen, USA) when the fusion rate reached about 75%. The cell groups were as follows: si-NC (5′-AAGACTGCGACTGAGGCA-3′), si-RNASET2-1 (5′-CTTGCTTAGT-GAGGCACAGTTC-3′), and si-RNASET2-2 (5′-TGCTTCCAGAAGCGGCTGCGA-3′). After 6 h of incubation, the medium was replaced with a complete culture medium, and the cells were incubated for 48 h. Transfection efficiency was detected by qPCR.

2.9. RT-qPCR

Total RNA was extracted from the cells using Trizol reagent (15596026, Invitrogen, USA) and then reverse transcribed into cDNA using the PrimeScript RT reagent Kit (RR047A, Takara, Japan) according to the manufacturer's instructions. The RNA was quantitatively analyzed using the Fast SYBR Green PCR kit (Applied Biosystems) and ABI PRISM 7300 PCR system (Applied Biosystems). Each sample was repeated in triplicate, and GAPDH was used as an internal reference gene. The relative expression level of the gene was analyzed using the $2-\Delta\Delta Ct$ method, where $\Delta Ct = CT$ (target) - CT (interference), $\Delta\Delta Ct = \Delta Ct$ (test) - ΔCt (control), and the average of three repeated experiments was taken. All primers were purchased from Shanghai Sangon Biotech, and their sequences are shown in Supplementary Table 1.

2.10. CCK8 assay

Cell proliferation of RCCC cells was detected using the CCK-8 assay kit (CCK-8; Dojindo, Kyushu Island, Japan). Cells in the logarithmic growth phase were seeded at a density of 1×10^4 cells/well in a 96-well plate and pre-cultured for 24 h before transfection. The cells were transfected for 48 h and then treated with CCK-8 detection solution (10 μL/well) at 0h, 24h, 48h, and 72h after transfection. The plate was incubated at 37 °C for 3 h, and the absorbance at 450 nm was measured using a microplate reader to plot the growth curve [36,37].

2.11. Scratch assay

Cells were seeded into 6-well plates at a 2.5×10^4 cells/cm2 density and allowed to attach overnight. After that, a scratch was made in each well using a 200 μL pipette tip, and the cells were then cultured in RPMI 1640 containing 5% FBS for 24 h. Images of the scratches were captured at 0 and 24 h using an inverted microscope. Each group was repeated three times. The width of each scratch was measured using Image J software, and the healing rate was calculated as follows: healing rate = (width of scratch at 0h - width of scratch at 24h)/width of scratch at 0h $\times$ 100% [37].

2.12. Transwell assay

A total of 600 μL of RPMI 1640 medium containing 20% FBS was added to the lower compartment of the Transwell chamber (8 μm pore size) coated with Matrigel, followed by incubation at 37 °C for 1 h. After transfection for 48 h, the cells were resuspended in serum-free RPMI 1640 medium and seeded in the upper compartment at a density of 1×10^6 cells/mL. The cells were then incubated at 37 °C with 5% CO2 for 24 h. The Transwell chamber was removed, and the cells were washed twice with PBS and fixed with 0.1% crystal violet. After washing with PBS, the cells on the upper surface were removed using a cotton swab. Finally, five random fields were selected, the cells were observed under an inverted fluorescence microscope, and the average was calculated [37].
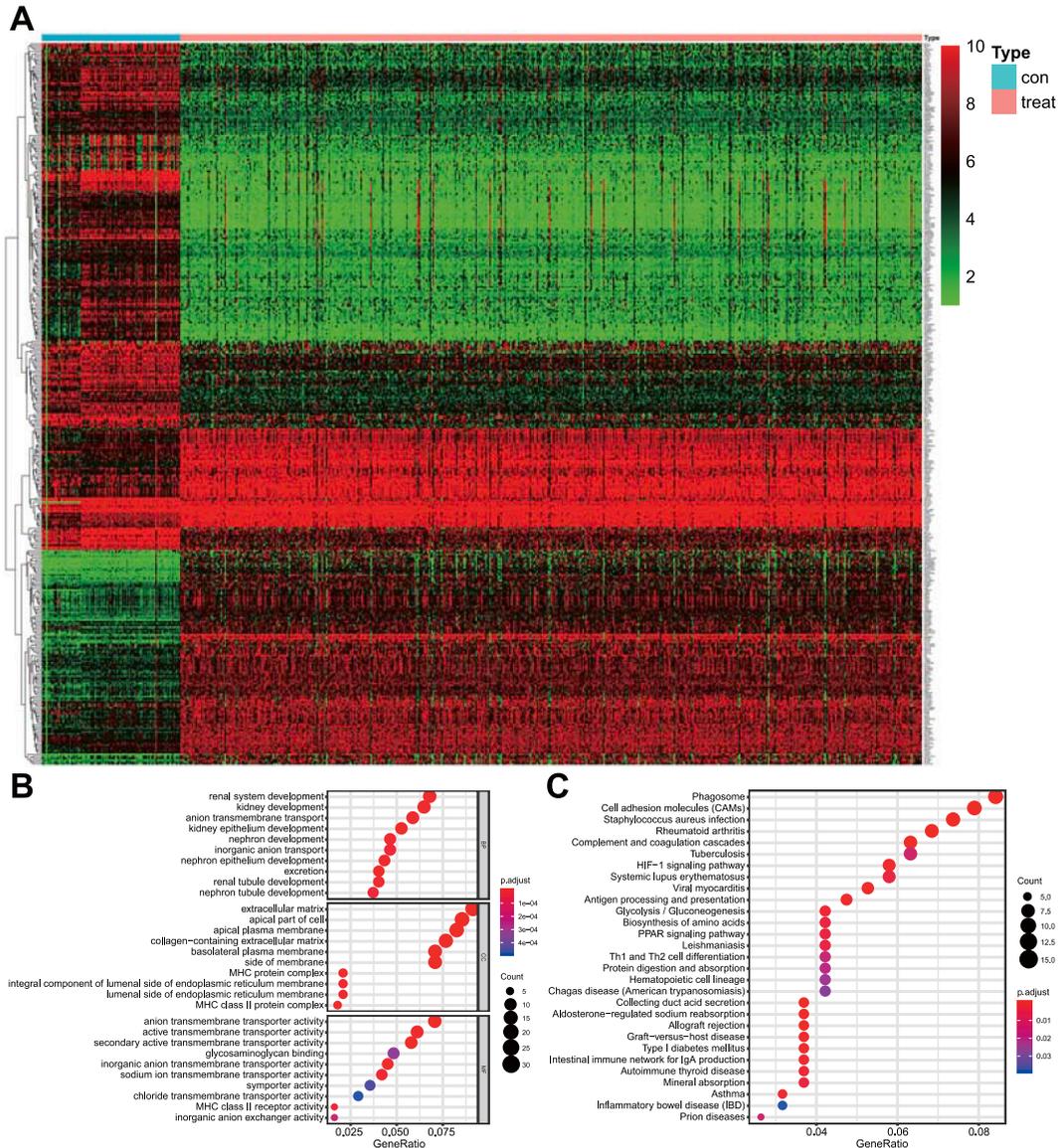
Fig. 1. Differential gene expression heatmap and functional enrichment analysis of RCCC-related expression data in TCGA and GTEx databases. Note: (A) Left panel shows the hierarchical clustering of differentially expressed genes, while the histogram on the upper right side represents the color scale from green to red, indicating the expression value of the gene from low to high. The blue on the top represents normal kidney tissue samples, while the red indicates RCCC samples. Each small square in the figure represents the expression status of a gene in a sample (N: Normal = 100, T: RCC = 535). (B) GO functional enrichment analysis of 386 differentially expressed genes. The horizontal axis represents the GeneRatio, the vertical axis represents the GO functional entry, and the three parts under the GO entry are shown on the right side. The histogram on the right side represents the color scale, the one on the top indicates the p-value and the different color ranges of p-values, and the one on the bottom indicates the counts data. The larger the circle, the more genes are enriched in that entry. (C) KEGG pathway enrichment analysis of 386 differentially expressed genes. The horizontal axis represents the GeneRatio, the vertical axis represents the KEGG pathway entry, and the histogram on the right side represents the color scale.

2.13. Statistical analysis

Experimental data were analyzed using GraphPad Prism 8.0 software (GraphPad Software, San Diego, USA). Data are presented as mean ± standard deviation (SD) and were compared by analysis of variance (ANOVA). A p-value <0.05 was considered statistically significant. The Pearson correlation coefficient was used to analyze the correlation between prognostic marker genes, and the Kaplan-Meier curve was plotted for survival analysis. In all analyses, p < 0.05 was considered the statistical significance level.

3. Results

3.1. 386 differentially expressed genes involved in renal function development and may affect the occurrence and development of RCC

We obtained expression data of RCCC and corresponding normal kidney tissue samples from the TCGA database, including 72 normal kidney tissue samples and 535 RCCC samples (Supplementary Table 2). To increase the sample size and improve the reliability of the analysis results, we further obtained expression data from 28 normal kidney tissue samples from the GTEx database. After correcting the expression data and conducting differential analysis using normal kidney tissue samples as controls, we finally identified 386 differentially expressed genes (Fig. 1A, Supplementary Table 3). Compared to normal kidney tissue samples, these 386 differentially expressed genes showed significant differential expression in RCC, including 165 significantly up-regulated and 221 genes significantly down-regulated in RCC.

We conducted GO and KEGG pathway enrichment analyses to further understand the functions of these 386 differentially expressed genes. GO enrichment analysis included three parts: biological process (BP), cellular component (CC), and molecular function (MF). The results of the GO enrichment analysis (Fig. 1B, Supplementary Table 4) showed that these differentially expressed genes were
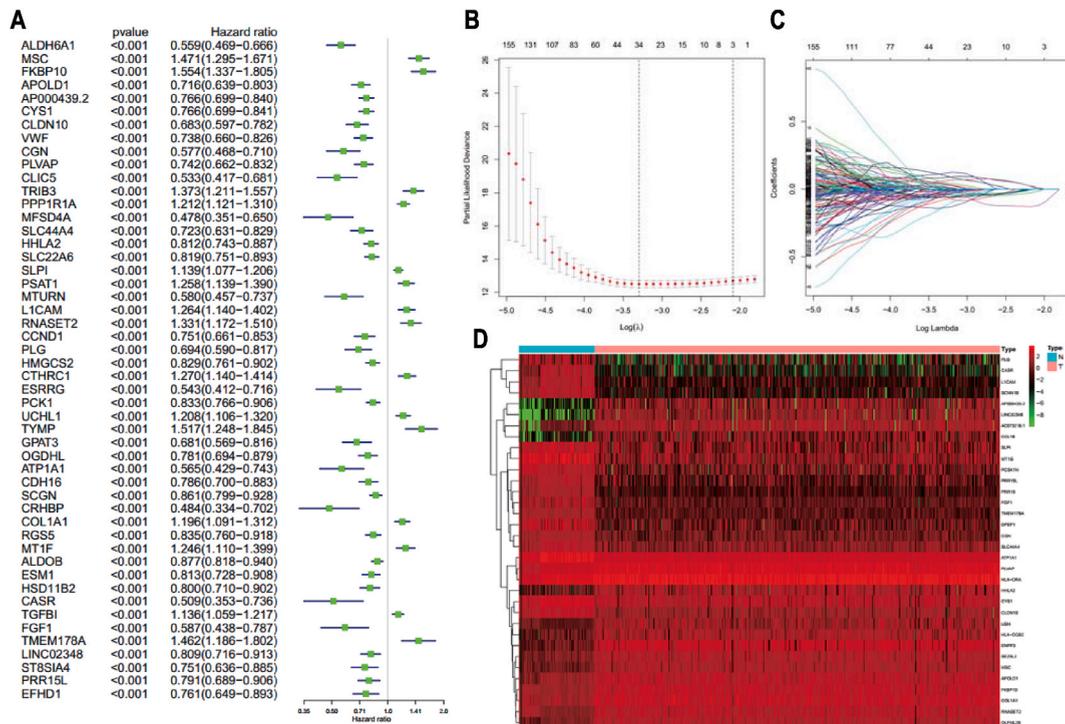
Fig. 2. Key differentially expressed genes associated with RCCC patient prognosis identified through bioinformatics analysis. Note: (A) Univariate COX analysis of RCCC-associated prognosis-related genes. The left panel represents the gene names. The middle panel shows the p-values, and the hazard ratio represents the risk rate. A hazard ratio greater than 1 indicates a high risk for that gene, while a hazard ratio less than 1 indicates a low risk. The right panel shows the gene's risk rate distribution, with the left side indicating low risk and the right side indicating high risk. (B) Lasso regression analysis to determine the key genes associated with RCCC prognosis. The horizontal axis represents the log(λ) value, while the vertical axis represents the Partial Likelihood of Deviance. The top panel indicates the number of genes retained when the corresponding log(λ) value was used for calculation, while the dashed line indicates the log(λ) value corresponding to the optimal Partial Likelihood Deviance and the number of genes retained. (C) Lasso regression analysis to determine the key genes associated with RCCC prognosis. The horizontal axis represents the log(λ) value, while the vertical axis represents the coefficient value. The different colored lines represent the genes used for calculation, and the top axis indicates the number of genes. (D) Heatmap of candidate genes retained after Lasso regression. The gene names are shown on the right side, and the dendrogram on the left represents the hierarchical gene expression level clustering. The histogram on the upper right side represents the color scale, indicating the expression value of the gene from low to high. The blue on the top represents normal kidney tissue samples, while the red indicates RCCC samples (N: Normal = 100, T: RCC = 535).

mainly enriched in "renal system development" and "kidney development" at the BP level and in "extracellular matrix" and "anion transmembrane transporter activity" at the CC and MF levels, respectively.

Further KEGG pathway enrichment analysis of all differentially expressed genes (Fig. 1C, Supplementary Table 5) revealed that these 386 differentially expressed genes were mainly enriched in pathways such as "phagosome," "Cell adhesion molecules (CAMs)," and "*Staphylococcus aureus* infection."

These results indicate that these 386 differentially expressed genes are mainly involved in the development of renal function, and their abnormal expression may cause renal disease and potentially affect the occurrence and development of RCC.

3.2. Thirty-four RCCC prognosis-related genes were identified through univariate Cox analysis

To investigate the correlation between the 386 differentially expressed genes and the prognosis of RCCC patients, we obtained the corresponding prognostic data of RCCC patients from the TCGA database. The expression of differentially expressed genes was then merged with the prognostic data of RCCC patients. Univariate Cox analysis showed that out of the 386 differentially expressed genes, 137 were closely related to RCCC patient prognosis (See Supplementary Table 6). Fig. 2A shows the top 50 differentially expressed genes with significant univariate Cox analysis results.

To obtain key differentially expressed genes associated with RCCC patient prognosis, we performed Lasso regression analysis on the 137 candidates differentially expressed genes identified from the univariate Cox analysis. As the number of genes decreased, the error
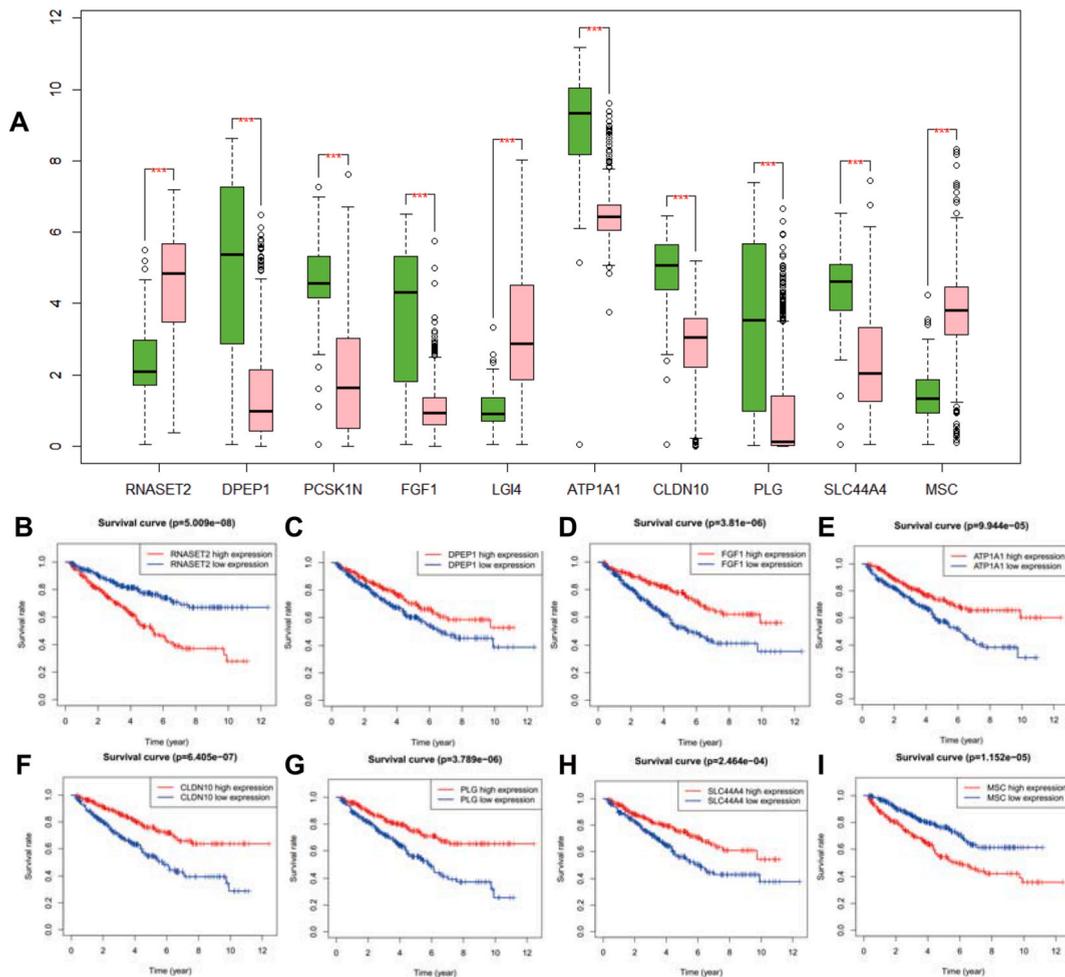
Fig. 3. Multivariate COX analysis and survival analysis of 10 prognostic signature genes. Note: (A) Differential expression of 10 prognostic signature genes in normal kidney tissue and RCCC samples. The x-axis represents gene names, and the y-axis represents gene expression values (corrected FPKM values). The green boxplot represents normal kidney tissue samples (n = 100), and the red boxplot represents RCCC samples (n = 535) (***: FDR <0.001); (B–I): Survival analysis of 8 prognostic signature genes (RNASET2, DPEP1, FGF1, ATP1A1, CLDN10, PLG, SLC44A1, and MSC). The x-axis of the graph represents survival time, and the y-axis represents survival rate. The blue line represents low-expression samples, while the red line represents high-expression samples. The gene name is located in the upper right corner of the survival curve, and the p-value is the significance value of the survival analysis.

values gradually decreased. Subsequently, coefficients were calculated for each gene under a different log (λ) values and genes with coefficients equal to zero were excluded, resulting in 34 genes (Fig. 2B–C, Supplementary Table 7). Fig. 2D shows the differential expression heatmap of these 34 genes in normal kidney tissue samples and RCCC samples.

3.3. Ten RCCC prognosis markers were identified through multivariate Cox analysis

To obtain independent prognostic genes associated with RCCC patients, we matched the expression of the 34 key genes with the Lasso risk status of RCCC patients. Fourteen genes with inconsistent risk status and expression were removed, leaving 20 key genes. Multivariate Cox analysis of these 20 key genes identified 10 prognostic markers (RNASET2, MSC, DPEP1, FGF1, ATP1A1, CLDN10, PLG, SLC44A1, PCSK1N, and LGI4), suggesting that these 10 markers may be independent risk genes associated with RCCC patient prognosis (Fig. 3A). RNASET2, LGI4, and MSC were significantly upregulated in RCCC samples, whereas DPEP1, FGF1, ATP1A1,
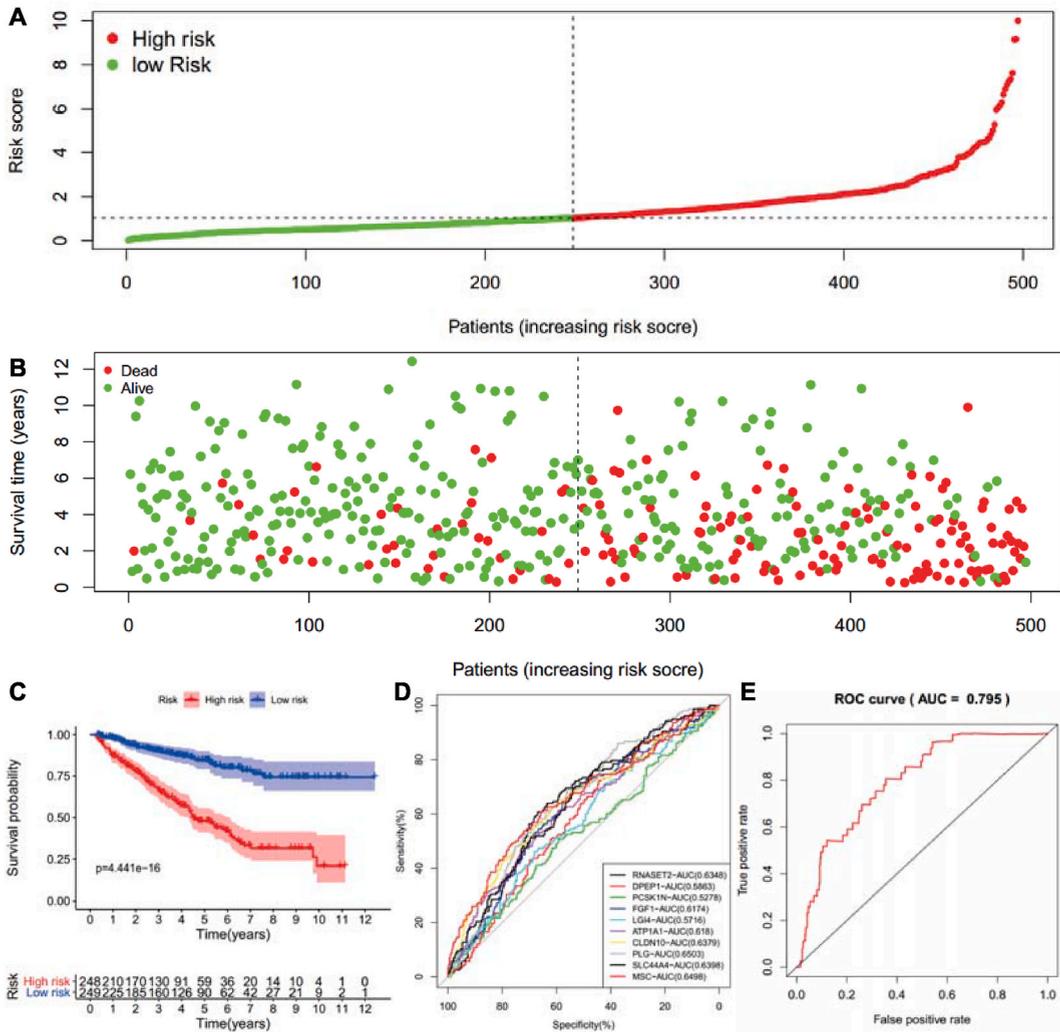
Fig. 4. Construction of a prognostic risk assessment model based on 10 prognostic signature genes in RCCC patients, survival analysis, and ROC analysis. Note: (A) Risk value chart of the prognostic risk assessment model. The x-axis represents patient numbers (from left to right, the risk values increase gradually), and the y-axis represents risk values. The green color represents low-risk patients, and the red represents high-risk patients; (B) Survival status chart of the prognostic risk assessment model. The x-axis represents patient numbers (from left to right, the risk values increase gradually), and the y-axis represents survival time. The green dot represents the surviving patients, and the red dot represents the deceased patients; (C) Survival analysis of the risk patients. The upper graph is a survival curve, the x-axis represents survival time, and the y-axis represents survival rate. The blue color represents the low-risk group, while the red color represents the high-risk group. The p-value is 4.441e-16. The lower graph is the sample risk grouping, the x-axis represents survival time, and the upper numbers represent the number of patients in the high-risk group at each time point, while the lower numbers represent the number of patients in the low-risk group at each time point. (D) ROC analysis of 10 prognostic signature genes individually, and the AUC value for each gene is shown in the lower right corner; (E) ROC analysis of the prognostic risk assessment model constructed using the 10 prognostic signature genes as a gene set. The AUC value is 0.795.

CLDN10, PLG, PCSK1N, and SLC44A1 were downregulated in RCCC samples.

Survival analysis showed that PCSK1N and LGI4 were not associated with RCCC patient survival (See Supplementary Figs. 1A–B), while the other eight genes were closely related to RCCC patient survival (Fig. 3B–I). Patients with high expression of RNASET2 and MSC had a poorer prognosis, whereas patients with low expression of DPEP1, FGF1, ATP1A1, CLDN10, PLG, and SLC44A1 had a better prognosis.

In summary, eight out of the ten identified prognostic markers were significantly associated with the prognosis of RCCC patients. High expression of RNASET2 and MSC was associated with poor prognosis, whereas high expression of DPEP1, FGF1, ATP1A1, CLDN10, PLG, and SLC44A1 was associated with a better prognosis.

3.4. A prognostic risk assessment model based on 10 prognostic marker genes accurately predicts the prognosis of RCCC patients

Through multivariate COX analysis, 10 prognostic marker genes were ultimately obtained, the risk value of tumor patients was calculated based on these genes, and a prognostic risk assessment model was constructed. Using the median risk value as the dividing line, RCCC patients were divided into the high-risk group (N = 248) and low-risk group (N = 249) through multivariate COX risk calculation (Fig. 4A). At the same time, each patient's survival status, survival time, and risk values in the high-risk and low-risk groups were statistically analyzed, and a survival status graph was plotted (Fig. 4B). From the survival status graph, it can be seen that the number of deceased patients in the high-risk group is higher than that in the low-risk group, and the survival time of patients in the high-risk group is shorter than that in the low-risk group, especially in the sample group with higher risk values on the right side, where the number of deceased patients is more concentrated.

Furthermore, survival time, survival status, and risk value data of RCCC patients were extracted, survival analysis was performed on high-risk and low-risk groups of patients, and a Kaplan-Meier survival curve was constructed (Fig. 4C). The results showed that the survival rate of patients in the high-risk group was significantly lower than that in the low-risk group. ROC curve analysis showed that, among these 10 prognostic marker genes, except for PCSK1N and LGI4, the AUC values of the other genes were greater than 0.6, and the largest was the MSC gene (AUC = 0.6498). Subsequently, these 10 prognostic marker genes were used as a gene set for ROC curve
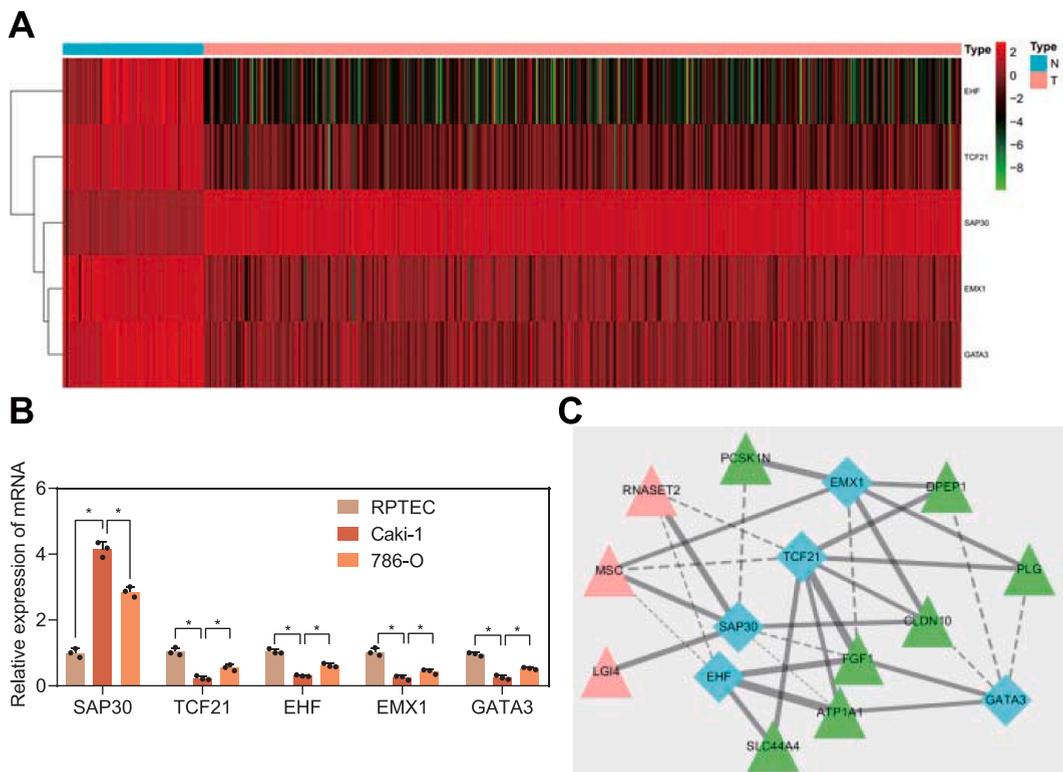
Fig. 5. **Shows the selection and predicted regulatory relationships of upstream transcription factors (TFs) for the 10 prognostic signature genes.** Panel (A) presents a heatmap of TF expression levels for the normal kidney and RCCC samples, with TF names on the right and clustering of TF expression levels on the left. The color bar at the top indicates the expression values from low (green) to high (red). Panel (B) shows the relative expression levels of the target genes in normal kidney cell line RPTEC and RCCC cell lines Caki-1 and 786-O, as detected by RT-qPCR. The asterisk indicates a statistically significant difference between the two groups (p < 0.05). Panel (C) presents a predicted regulatory network diagram of TFs and the 10 prognostic signature genes. The diamonds represent the TFs, and the triangles represent the predicted prognostic signature genes. The green and red colors indicate low- and high-risk genes, respectively. The lines connecting the TFs and genes indicate their regulatory relationships, with solid lines indicating positive regulation and dotted lines indicating negative regulation.

analysis, and the AUC value was 0.795 (Fig. 4D and E).

These results indicate that the prognostic risk assessment model based on 10 prognostic marker genes has good accuracy and can effectively distinguish the prognosis of RCCC patients. They can be used as a gene set to predict the prognosis of RCCC patients.

3.5. A transcription factor-prognostic marker gene regulatory network was predicted and constructed

Five transcription factors were significantly differentially expressed in RCCC based on differential gene expression data from the TCGA and GTEx databases (Fig. 5A). Subsequently, RT-qPCR was used to detect the expression levels of relevant genes in RPTEC, Caki-1, and 598RCC. The results showed that compared with the normal human kidney transparent cell line RPTEC, the transcription factors TCF21, EHF, EMX1, and GATA3 were significantly downregulated, while the transcription factor SAP30 was significantly up-regulated in the human RCCC cell lines Caki-1 and 786-O (Fig. 5B).

Subsequently, the expression values of these 5 transcription factors and 10 prognostic marker genes in RCCC were extracted, their correlation and regulatory relationships were predicted, and a regulatory network diagram was constructed. It was found that multiple transcription factors can regulate these 10 prognostic marker genes. As shown in Fig. 5C, based on the expression trends of these 10 prognostic marker genes and 5 transcription factors in RCCC samples and normal kidney tissue samples, it was found that TGF21 and EHF have a more significant regulatory effect on these 10 prognostic marker genes than the other 3 transcription factors.

The above results suggest that the regulation of the 10 prognostic marker genes by the 5 transcription factors may not be solely through direct transcriptional regulation, and their regulatory mechanisms may be highly complex, involving both direct and indirect transcriptional regulation of different genes. Therefore, further in-depth research is needed to elucidate the specific regulatory mechanisms between transcription factors and prognostic marker genes.

3.6. ATP1A1 may be a key prognostic marker gene affecting RCCC patient outcomes

We performed RT-qPCR to detect the expression levels of the 10 prognostic marker genes in the human normal kidney cell line RPTEC and two human RCCC cell lines (Caki-1 and 786-O). The results showed that compared to RPTEC, Caki-1, and 786-O cells exhibited significantly higher expression of RNASET2, LGI4, and MSC, while DPEP1, FGF1, ATP1A1, CLDN10, PLG, PCSK1N, and
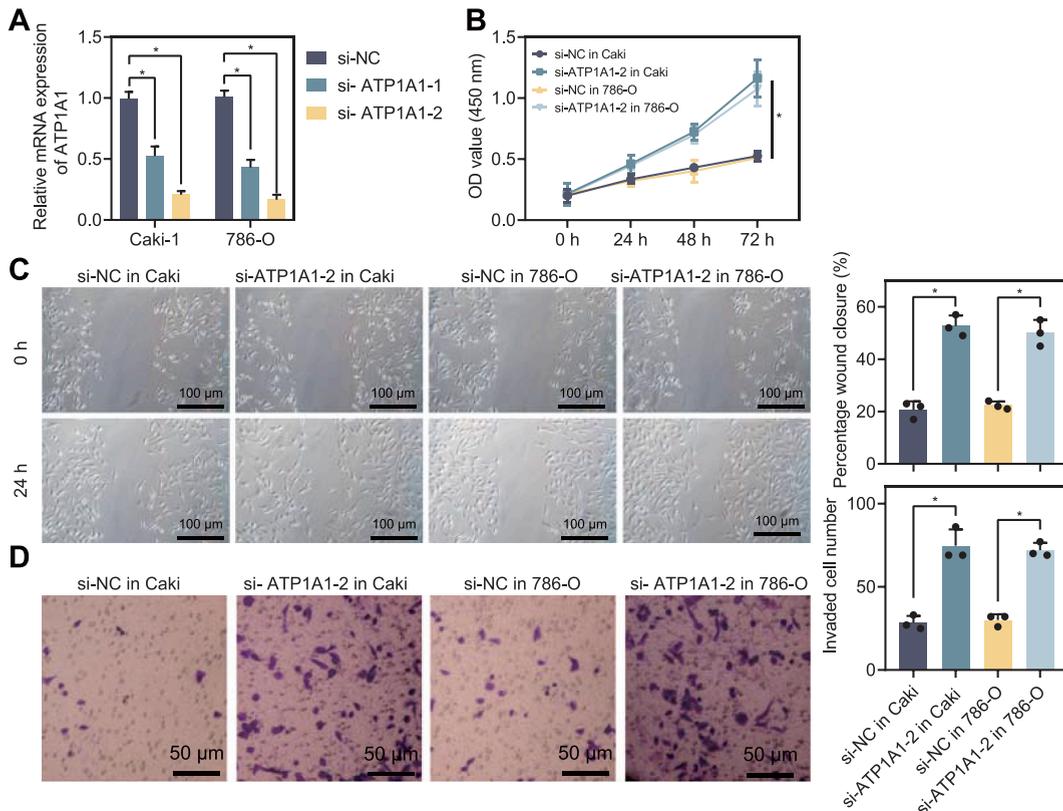
Fig. 6. Shows the effects of ATP1A1 gene knockdown on RCCC cell proliferation, migration, and invasion. Panel (A) shows the efficiency of ATP1A1 knockdown in Caki-1 and 786-O RCCC cell lines, as detected by RT-qPCR. Panels (B)–(D) show the CCK-8 assay, scratch test, and Transwell assay results, respectively, for cell proliferation, migration, and invasion in the RCCC cell lines with ATP1A1 knockdown. The asterisk indicates a statistically significant difference between the two groups (p < 0.05). The experiments were repeated three times.

SLC44A1 were significantly downregulated (Supplementary Fig. 2A). According to Fig. 3A, DPEP1, FGF1, ATP1A1, CLDN10, PLG, and SLC44A1 were downregulated in RCCC samples, suggesting that they may act as protective genes in the development and progression of RCCC. Therefore, we further analyzed the differential expression of these 6 downregulated prognostic marker genes (DPEP1, FGF1, ATP1A1, CLDN10, PLG, and SLC44A1) using TCGA RCCC expression profile data. Supplementary Fig. 2B shows that ATP1A1 had the highest expression level among these genes. Thus, we selected ATP1A1 as the target gene for subsequent in vitro cell experiments to validate its function.

We used siRNA to knock down ATP1A1 and measured the proliferation, migration, and invasion abilities of Caki-1 and 786-O cells. The results showed that both si-ATP1A1#1 and si-ATP1A1#2 significantly reduced ATP1A1 expression in Caki-1 and 786-O cells, with si-ATP1A1#2 having the best knockdown effect (Fig. 6A). Therefore, we used si-ATP1A1#2 to establish a stable cell line for subsequent experiments.

The CCK8 assay showed that compared to si-NC, the proliferation of Caki-1 and 786-O cells in the si-ATP1A1-2 group was significantly increased (Fig. 6B). The scratch assay showed that the migration ability of Caki-1 and 786-O cells in the si-ATP1A1-2 group was significantly higher than that in the si-NC group (Fig. 6C). The Transwell assay showed that the invasion ability of Caki-1 and 786-O cells in the si-ATP1A1-2 group was significantly higher than that in the si-NC group (Fig. 6D).

These results suggest that ATP1A1 may be a key prognostic marker gene affecting RCCC patient outcomes, and knockdown of ATP1A1 can promote RCCC cell proliferation, migration, and invasion.

4. Discussion

Renal clear cell carcinoma (RCCC) is one of the most common subtypes of kidney cancer and is associated with high heterogeneity and uncertain prognosis [38]. In this study, we conducted differential gene analysis and screening of prognostic markers in RCCC patients based on TCGA and GTEx databases and constructed a transcription factor-prognostic marker gene regulatory network. Our study identified ATP1A1 as a potential key prognostic marker gene in RCCC, and the knockdown of ATP1A1 in vitro promoted RCCC cell proliferation, migration, and invasion.

Our study made important contributions to selecting RCCC prognostic marker genes and constructing a prognostic risk evaluation model. Firstly, we identified 386 genes significantly differentially expressed in RCCC through differential gene analysis and functional enrichment analysis, with 165 genes up-regulated, and 221 genes downregulated. These differentially expressed genes were mainly related to kidney function and development. Secondly, we selected 34 key genes that were associated with RCCC prognosis using single-factor COX analysis and Lasso regression analysis and constructed a prognostic risk evaluation model based on 10 prognostic marker genes (RNASET2, MSC, DPEP1, FGF1, ATP1A1, CLDN10, PLG, SLC44A1, PCSK1N, and LGI4) using multiple-factor COX analysis. ROC curve analysis showed that this prognostic risk evaluation model could accurately predict RCCC patient prognosis. These 10 prognostic marker genes have been previously found to be associated with RCCC tumorigenesis and development. For example, MSC was found to promote proliferation, migration, and invasion of RCCC cells and was associated with prognosis [39]. DPEP1 was highly expressed in RCCC cells and was associated with tumor grading and prognosis [40]. CLDN10, PCSK1N, and LGI4 also play important roles in RCCC tumorigenesis and development [41].

Furthermore, we constructed a transcription factor-prognostic marker gene regulatory network to explore the transcription factor regulation network in RCCC tumorigenesis and development. Our results showed that these 10 prognostic marker genes were regulated by five different transcription factors (EMX1, TCF21, SAP30, EHF, and GATA3), which were involved in RCCC tumorigenesis and development. This transcription factor-prognostic marker gene regulatory network provides a new direction for studying the molecular mechanisms of RCCC tumorigenesis and development. By predicting the upstream transcription factors of these 10 prognostic marker genes, we found that different transcription factors regulated these genes, and EMX1, TCF21, SAP30, EHF, and GATA3 were transcription factors that regulate multiple marker genes. These transcription factors may regulate the transcription network of RCCC, affecting cell proliferation, apoptosis, invasion, and other biological processes.

ATP1A1 was a key prognostic marker gene in RCCC patients in this study. ATP1A1 is an important cell membrane Na+/K + -ATPase that transports sodium and potassium ions and regulates the balance of extracellular and intracellular fluids [42]. Previous studies have shown that ATP1A1 plays an important role in the occurrence and development of RCCC and is associated with prognosis [43]. ATP1A1 is overexpressed in RCCC and is associated with malignancy, prognosis, and treatment response [43]. Another study has shown that the downregulation of ATP1A1 promotes RCCC cell apoptosis and slows tumor growth [43]. In this study, it was found through in vitro experiments that knocking down ATP1A1 can promote RCCC cell proliferation, migration, and invasion, further demonstrating the importance of ATP1A1 [43].

In summary, this study identified 10 RCCC prognostic signature genes (RNASET2, MSC, DPEP1, FGF1, ATP1A1, CLDN10, PLG, SLC44A1, PCSK1N, and LGI4), among which ATP1A1 may be a key prognostic marker for RCCC patients. The prognostic risk assessment model based on these 10 prognostic signature genes can accurately predict the prognosis of RCCC patients. These findings provide important references and guidance for diagnosing and treating RCCC and are expected to provide better strategies for evaluating patient prognosis and personalized treatment. In addition, this study also constructed a transcription factor-prognostic signature gene regulatory network, revealing the relationships between multiple transcription factors and RCCC prognostic signature genes, providing new ideas and directions for further research on the occurrence and development of RCCC. The results of this study can also provide useful references and inspirations for the screening of prognostic signature genes and the study of molecular mechanisms in other types of tumors.

Although this study analyzed data from the TCGA and GTEx public databases and combined in vitro cell experiments for validation, some limitations remain. Firstly, the samples in these databases may be affected by some biases, such as sample sources and sample

sizes, which may impact the reliability and generalizability of the research results. Secondly, this study used in vitro cell experiments for validation, rather than clinical samples or animal experiments, which cannot fully reflect the complexity of the intrinsic biological characteristics of RCCC patients. In addition, the samples in this study only included RCCC, so the results of this study may not apply to other types of kidney cancer or other malignant tumors. Finally, this study did not conduct functional experiments to validate the mechanism of action of ATP1A1, so further research is needed to elucidate the role and mechanism of ATP1A1 in RCCC.

Funding

Author contributions

Wei Zhu: Conceived and designed the experiments; Performed the experiments; Wrote the paper.
Lingfeng Wu: Conceived and designed the experiments; Performed the experiments; Wrote the paper.
Wenhua Xie: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.
Gaoyue Zhang: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.
Yanqin Gu: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.
Yansong Hou: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.
Yi He: Conceived and designed the experiments; Conceived and designed the experiments; Wrote the paper.

Availability of data and materials

The data supporting this study's findings are available on request from the corresponding author upon reasonable request.

Ethics statement

This article contains no studies with human participants or animals performed by authors.

Consent for publication

Not applicable.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e18870.

References

[1] E. Zhao, X. Li, B. You, J. Wang, W. Hou, Q. Wu, Identification of a five-miRNA signature for diagnosis of kidney renal clear cell carcinoma, Front. Genet. 13 (2022), 857411.
[2] R. Restrepo, D. Zahrah, L. Pelaez, H.T. Temple, J.W. Murakami, Update on aneurysmal bone cyst: pathophysiology, histology, imaging and treatment, Pediatr. Radiol. 52 (9) (2022) 1601–1614.
[3] S. Bahadoram, M. Davoodi, S. Hassanzadeh, M. Bahadoram, M. Barahman, L. Mafakher, Renal cell carcinoma: an overview of the epidemiology, diagnosis, and treatment, G. Ital. Nefrol. 39 (3) (2022).
[4] A.L. Oliver, Lung cancer: epidemiology and screening, Surg. Clin. 102 (3) (2022) 335–344.
[5] J.A. Pena, et al., Dose-efficient assessment of trabecular microstructure using ultra-high-resolution photon-counting CT, Z. Med. Phys. 32 (4) (2022) 403–416.
[6] D.X. Dong, Z.G. Ji, H.Z. Li, W.G. Yan, Y.S. Zhang, Preliminary application of WCX magnetic bead-based matrix-assisted laser desorption ionization time-of-flight mass spectrometry in analyzing the urine of renal clear cell carcinoma, Chin. Med. Sci. J. 32 (4) (2017) 248–252.
[7] S.W. Brady, et al., The genomic landscape of pediatric acute lymphoblastic leukemia, Nat. Genet. 54 (9) (2022) 1376–1389.
[8] P. Horak, et al., Standards for the classification of pathogenicity of somatic variants in cancer (oncogenicity): joint recommendations of clinical genome resource (ClinGen), cancer genomics consortium (CGC), and variant interpretation for cancer consortium (VICC), Genet. Med. 24 (5) (2022) 986–998.
[9] X. Wang, et al., Comprehensive assessment of cellular senescence in the tumor microenvironment, Briefings Bioinf. 23 (3) (2022).

[10] Z. Bian, R. Fan, L. Xie, A novel cuproptosis-related prognostic gene signature and validation of differential expression in clear cell renal cell carcinoma, Genes 13 (5) (2022).

[11] X. Fang, R. Miao, J. Wei, H. Wu, J. Tian, Advances in multi-omics study of biomarkers of glycolipid metabolism disorder, Comput. Struct. Biotechnol. J. 20 (2022) 5935–5951.

[12] C.S. Sheela Devi, S. Suchitha, R.S. Veerendrasagar, Evaluation of nuclear morphometry and ki-67 index in clear cell renal cell carcinomas: a five-year study, Iran J. Pathol. 12 (2) (2017) 150–157.

[13] A. Zhang, X. Wang, C. Fan, X. Mao, The role of Ki67 in evaluating neoadjuvant endocrine therapy of hormone receptor-positive breast cancer, Front. Endocrinol. 12 (2021), 687244.

[14] D. Coco, S. Leanza, Von hippel-lindau syndrome: medical syndrome or surgical syndrome? A surgical perspective, J. Kid. Canc.VHL 9 (1) (2022) 27–32.

[15] S. Nolting, et al., Personalized management of pheochromocytoma and paraganglioma, Endocr. Rev. 43 (2) (2022) 199–239.

[16] N. Elzakra, Y. Kim, HIF-1alpha metabolic pathways in human cancer, Adv. Exp. Med. Biol. 1280 (2021) 243–260.

[17] J. Zheng, et al., HIF-1alpha in myocardial ischemia-reperfusion injury, Mol. Med. Rep. 23 (5) (2021) (Review).

[18] M. Rashid, et al., Up-down regulation of HIF-1alpha in cancer progression, Gene 798 (2021), 145796.

[19] Q. He, et al., Biological functions and regulatory mechanisms of hypoxia-inducible factor-1alpha in ischemic stroke, Front. Immunol. 12 (2021), 801985.

[20] A. Putra, I. Riwanto, S.T. Putra, I. Wijaya, Typhonium flagelliforme extract induce apoptosis in breast cancer stem cells by suppressing survivin, J. Cancer Res. Therapeut. 16 (6) (2020) 1302–1308.

[21] M. Watanabe, Y. Nishikawaji, H. Kawakami, K.I. Kosai, Adenovirus biology, recombinant adenovirus, and adenovirus usage in gene therapy, Viruses 13 (12) (2021).

[22] A. Faldt Beding, P. Larsson, K. Helou, Z. Einbeigi, T.Z. Parris, Pan-cancer analysis identifies BIRC5 as a prognostic biomarker, BMC Cancer 22 (1) (2022) 322.

[23] W. Wu, et al., Drivers and suppressors of triple-negative breast cancer, Proc. Natl. Acad. Sci. U. S. A. 118 (33) (2021).

[24] E. Jonasch, C.L. Walker, W.K. Rathmell, Clear cell renal cell carcinoma ontogeny and mechanisms of lethality, Nat. Rev. Nephrol. 17 (4) (2021) 245–261.

[25] T.O. Bui, V.T. Dao, V.T. Nguyen, J.P. Feugeas, F. Pamoukdjian, G. Bousquet, Genomics of clear-cell renal cell carcinoma: a systematic review and meta-analysis, Eur. Urol. 81 (4) (2022) 349–361.

[26] N. Chowdhury, C.G. Drake, Kidney Cancer, An overview of current therapeutic approaches, Urol. Clin. 47 (4) (2020) 419–431.

[27] Y. Li, R.Y. Wang, Y.J. Deng, S.H. Wu, X. Sun, H. Mu, Molecular characteristics, clinical significance, and cancer immune interactions of cuproptosis and ferroptosis-associated genes in colorectal cancer, Front. Oncol. 12 (2022), 975859.

[28] J. Singh, S. Luqman, A. Meena, Emerging role of phytochemicals in targeting predictive, prognostic, and diagnostic biomarkers of lung cancer, Food Chem. Toxicol. 144 (2020), 111592.

[29] C. Pierrakos, D. Velissaris, M. Bisdorff, J.C. Marshall, J.L. Vincent, Biomarkers of sepsis: time for a reappraisal, Crit. Care 24 (1) (2020) 287.

[30] G. Li, Y. Wang, L. Cai, L. Zhou, Screening for genes and subnetworks associated with atypical teratoid/rhabdoid tumors using bioinformatics analysis, Int. J. Neurosci. 131 (4) (2021) 319–326.

[31] C. Tao, K. Huang, J. Shi, Q. Hu, K. Li, X. Zhu, Genomics and prognosis analysis of epithelial-mesenchymal transition in glioma, Front. Oncol. 10 (2020) 183.

[32] X. Yan, X. Fu, Z.X. Guo, X.P. Liu, T.Z. Liu, S. Li, Construction and validation of an eight-gene signature with great prognostic value in bladder cancer, J. Cancer 11 (7) (2020) 1768–1779.

[33] M. Zhao, X. Tang, D. Gong, P. Xia, F. Wang, S. Xu, Bungeanum improves cognitive dysfunction and neurological deficits in D-galactose-induced aging mice via activating PI3K/Akt/Nrf2 signaling pathway, Front. Pharmacol. 11 (2020) 71.

[34] H. Cao, Y. Zhang, Z. Chu, B. Zhao, H. Wang, L. An, MAP-1B, PACS-2 and AHCYL1 are regulated by miR-34A/B/C and miR-449 in neuroplasticity following traumatic spinal cord injury in rats: preliminary explorative results from microarray data, Mol. Med. Rep. 20 (4) (2019) 3011–3018.

[35] J. Yang, Identification of novel biomarkers, MUC5AC, MUC1, KRT7, GAPDH, CD44 for gastric cancer, Med. Oncol. 37 (5) (2020) 34.

[36] M.G. Jaarsma-Coes, T.A. Goncalves Ferreira, G.R. van Haren, M. Marinkovic, J.M. Beenakker, MRI enables accurate diagnosis and follow-up in uveal melanoma patients after vitrectomy, Melanoma Res. 29 (6) (2019) 655–659.

[37] C. Yue, J. Chen, Z. Li, L. Li, J. Chen, Y. Guo, microRNA-96 promotes occurrence and progression of colorectal cancer via regulation of the AMPKalpha2-FTO-m6A/MYC axis, J. Exp. Clin. Cancer Res. 39 (1) (2020) 240.

[38] C.P. Miller, et al., Immunohistochemical detection of 5T4 in renal cell carcinoma, Appl. Immunohistochem. Mol. Morphol. 31 (3) (2023) 135–144.

[39] J. Korbecki, et al., The role of CXCL16 in the pathogenesis of cancer and other diseases, Int. J. Mol. Sci. 22 (7) (2021).

[40] A. Lau, et al., Dipeptidase-1 governs renal inflammation during ischemia reperfusion injury, Sci. Adv. 8 (5) (2022), eabm0142.

[41] Y. Hou, L. Fan, H. Li, Oncogenic miR-27a delivered by exosomes binds to SFRP1 and promotes angiogenesis in renal clear cell carcinoma, Mol. Ther. Nucleic Acids 24 (2021) 92–103.

[42] K.P. Schlingmann, et al., Germline de novo mutations in ATP1A1 cause renal hypomagnesemia, refractory seizures, and intellectual disability, Am. J. Hum. Genet. 103 (5) (2018) 808–816.

[43] A. Megarbane, et al., A 20-year clinical and genetic neuromuscular cohort analysis in Lebanon: an international effort, J. Neuromuscul. Dis. 9 (1) (2022) 193–210.