

Research Paper

Prognostic Role of the Ubiquitin Proteasome System in Clear Cell Renal Cell Carcinoma: A Bioinformatic Perspective

Hongda Guo^{1,2}, Yan Li^{1,2}, Yaxiao Liu^{1,2}, Lipeng Chen^{1,2}, Zhengdong Gao^{1,2}, Lekai Zhang^{1,2}, Nan Zhou^{1,2}, Hu Guo^{1,2}✉, Benkang Shi^{1,2}✉

1. Department of Urology, Qilu Hospital of Shandong University, 107 Wenhua Road, Jinan, 250012, China.
2. Key Laboratory of Urinary Precision Diagnosis and Treatment in Universities of Shandong, Jinan, P.R. China.

✉ Corresponding authors: Hu Guo, Department of Urology, Qilu Hospital of Shandong University, 107 Wenhua Road, Jinan, 250012, China. E-mail: 18560083911@163.com. Tel: 18560083911; Benkang Shi, Department of Urology, Qilu Hospital of Shandong University, 107 Wenhua Road, Jinan, 250012, China. E-mail: bkang68@sdu.edu.cn. Tel: 18560083917.

© The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>). See <http://ivyspring.com/terms> for full terms and conditions.

Received: 2020.09.26; Accepted: 2021.04.24; Published: 2021.05.13

Abstract

Background: Clear cell renal cell carcinoma (ccRCC) is a common malignant tumor of the urinary system. The ubiquitin proteasome system (UPS) plays an important role in the generation, metabolism and survival of tumor. We are aimed to make a comprehensive exploration of the UPS's role in ccRCC with bioinformatic tools, which may contribute to the understanding of UPS in ccRCC, and give insight for further research.

Methods: The UPS-related genes (UPSs) were collected by an integrative approach. The expression and clinical data were downloaded from TCGA database. R soft was used to perform the differentially expressed UPSs analysis, functional enrichment analysis. We also estimated prognostic value of each UPS with the help of GEPIA database. Two predicting models were constructed with the differentially expressed UPSs and prognosis-related genes, respectively. The correlations of risk score with clinical characteristics were also evaluated. Data of GSE29609 cohort were obtained from GEO database to validate the prognostic models.

Results: We finally identified 91 differentially expressed UPSs, 48 prognosis related genes among them, and constructed a prognostic model with 18 UPSs successfully, the AUC was 0.760. With the help of GEPIA, we found 391 prognosis-related UPSs, accounting for 57.84% of all UPSs. Another prognostic model was constructed with 28 prognosis-related genes of them, and with a better AUC of 0.825. Additionally, our models can also stratify patients into high and low risk groups accurately in GSE29609 cohort. Similar prognostic values of our models were observed in the validated GSE29609 cohort.

Conclusions: UPS is dysregulated in ccRCC. UPS related genes have significant prognostic value in ccRCC. Models constructed with UPSs are effective and applicable. An abnormal ubiquitin proteasome system should play an important role in ccRCC and be worthy of further study.

Key words: the ubiquitin proteasome system (UPS); clear cell renal cell carcinoma (ccRCC); prognosis; bioinformatics

Introduction

Renal cell carcinoma (RCC) is a type of malignant tumor originating from the renal tubular epithelium and accounts for about 3-5% of all adult malignancies [1]. Clear cell renal cell carcinoma (ccRCC) is the most common histological subtype and accounts for approximately 75% of RCC [2]. Surgical resection is the main treatment for ccRCC, but ccRCC is likely to recur and threatens patient's life [3-5]. Therefore, the exploration of new molecular targets in

ccRCC is essential as well as valuable for the development of more accurate diagnosis and more effective treatment. The ubiquitin proteasome system (UPS) seems a promising candidate [6].

The UPS is a specialized proteolysis system which is responsible for protein degradation and plays an essential role in the regulation of abundant biological processes [7, 8]. It consists of a series of key components: ubiquitin, ubiquitin activating enzymes

(E1s), ubiquitin conjugating enzymes (E2s), ubiquitin ligases (E3s), deubiquitinating enzymes (DUBs), and the proteasome [9]. The UPS controls the degradation of more than 80% of cellular proteins and plays an indispensable role in a wide range of biological processes including cell proliferation, development, immune responses, and a variety of human diseases including tumor [10, 11]. Dysregulation of UPS is closely related to tumor and tumor cells are highly dependent on a functional UPS [12]. Considering the current evidence indicates aberrancies in the UPS pathway play an important role in cancer, development of drugs targeting different components of the UPS has been proposed as a promising therapeutic strategy [13-17]. An inspiring example is the success of proteasome inhibitors used in multiple myeloma treatment [18, 19].

Previous researches suggest ccRCC is dominated by inactivating mutations in Von-Hippel Lindau (VHL). The protein encoded by VHL is involved in assembling a protein complex possessing ubiquitin ligase E3 activity. It is responsible for the ubiquitination and degradation of hypoxia-inducible-factor (HIF), which plays a central role in the control of gene expression by oxygen [20]. Besides VHL, there are also some other UPS genes have been reported to be involved in ccRCC such as USP2 [21], USP44 [22], CUL5 [23], SPOP [24] and UHRF1 [25]. Additionally, UPS was also found to be involved in papillary renal cell carcinoma [26]. However, there are few researches focusing on exploring the potential of the whole ubiquitin proteasome system in ccRCC. In this study, we are aimed to make a comprehensive exploration of the ubiquitin proteasome system's role in ccRCC, mainly focus on the prognostic value of UPS-related genes (UPSs) with the assistance of bioinformatic tools.

Materials and Methods

Collection of UPSs

We collected and confirmed the genes encoding the ubiquitin proteasome system (UPS) by an integrative approach. Coding genes of ubiquitin were obtained from the Uniport database, while the gene lists of ubiquitin activating enzymes (E1s) [27], ubiquitin conjugating enzymes (E2s) [28] and deubiquitinating enzymes (DUBs) [29] were collected from the reliable published literatures. There are several online databases for ubiquitin ligases (E3s), among them we select the database named UbiNet, for it have collected abundant members of E3 ligases in human, about 500 genes [30]. As for the proteasome, which is a relatively conservative component of UPS, we get its genes from GSEA

website, by downloading the gene set named KEGG_PROTEASOME [31, 32]. Finally, we get the UPS-related gene set containing 676 UPSs in total. Details of gene set are shown in **Table S1**.

Samples and data retrieval

We acquired both the FPKM-standardized RNA-seq data and the clinical information from the KIRC cohort in the TCGA database. All patients in TCGA-KIRC cohort are involved into our study when doing expression related analyses, while patients with unknown clinical features were excluded in Cox regression analysis. Finally, we extracted an expression matrix of 670 UPSs in 72 normal samples and 539 tumor samples, and obtain a file containing complete clinical information without missing values in 248 patients.

Identification of differentially expressed UPSs

R software (version 4.0.2) and "limma" package were used to make the analysis of differentially expressed UPSs. We set a false discovery rate (FDR) < 0.05 and a $|\log_2 \text{fold change (FC)}| > 1$ as screening criteria to obtain the differentially expressed UPSs. We also visualized them by the heatmap and boxplot.

Enrichment analysis of differentially expressed UPSs

To have a better understanding of differentially expressed UPSs in ccRCC, we carried out a series of gene functional enrichment analyses to determine the major biological attributes, including the gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses. The "Gplot" package was employed to visualize the enrichment terms [33].

Construction of prognostic model (Model 1) with differentially expressed UPSs

We obtained 48 prognostic related UPSs in ccRCC by univariate Cox regression analysis, and these prognosis-related genes were used to construct a prognostic model by multivariate Cox regression with the step function in R. This model was used to calculate the risk score for every ccRCC patient, and the patients were divided into the low-risk group and high-risk group. Survival analysis was made by Kaplan-Meier method. Cox regression analysis was performed to demonstrate whether the UPS related risk score was an independent prognostic factor in ccRCC patients. Heatmap was employed to visualize the expression pattern between two groups. We also made GSEA analysis in GSEA software (version 4.0.1) to analyze which pathways genes are primarily enriched in high- and low-risk groups [32].

Exploration of all 676 UPSs in GEPIA

GEPIA, which is an interactive web server developed for exploring the RNA sequencing expression data from the TCGA and the GTEx projects, was used in this part [34]. The excellent function of patient survival analysis provided by GEPIA was used to estimate each prognostic value of all 676 UPSs. With the log-rank $p < 0.05$ and $|\log_2$ hazard rate (HR) $| > 1$, we got 69 prognosis-related genes with outstanding prognostic value. The correlations of gene expression and tumor stage were also explored with the stage plot function of GEPIA.

Construction of prognostic model (Model 2) with 28 prognosis related UPSs

Considering the extraordinary significance of 69 prognosis-related genes, we constructed another prognostic model with these genes by multivariate Cox regression with the step function in R. 28 UPSs were involved in the Model 2. We also got a risk score formula and made survival analysis between two groups separated by calculated risk score value. Cox regression analysis was performed to support this risk score as an independent prognostic factor.

Estimation and external validation of Model 1 and Model 2

Finally, we estimated these two models by drawing receiver operating characteristic (ROC) curve with the "survivalROC" package. Correlation of risk score with clinical characteristics were also evaluated. Additionally, we downloaded data of GSE29609 from GEO database, GSE29609 provided data of whole genome expression of 39 clear-cell renal cell carcinomas patient as well as survival state and

survival time [35]. This 39 ccRCC patients were used for external validation and their risk scores were calculated with the formulas of Model 1 and Model 2, respectively. Kaplan-Meier curve was plotted with R software.

Statistical Analysis

Statistical analyses were performed with R software (Version 4.0.2). The Wilcoxon signed-rank test was used to compare the expression levels of differentially expressed UPSs between cancer tissues and normal tissues. Cox regression analyses were employed to filter genes to build risk score models. Differences between survival curves generated by the Kaplan-Meier method were defined by log-rank tests. Receiver operating characteristic (ROC) curve analysis was performed with the "survivalROC" package. The Wilcoxon rank sum test was used in estimating the correlations of risk score with clinical characteristics. All statistical tests were bilateral, with $p < 0.05$ being statistically significant.

Results

Differentially expressed UPSs in tumor samples comparing with normal samples

The flowchart of our research was showed in **Figure 1**, detailed clinical parameters of patients were shown in **Table 1**. The expression values of UPSs were extracted from normal and tumor samples. With FDR < 0.05 and $|\log_2$ FC $| > 1$ as the screening criteria, 91 differentially expressed UPSs were obtained. Compared with normal samples, there were 32 downregulated UPSs and 59 upregulated UPSs in tumor samples, details are shown in **Table 2**. Many of

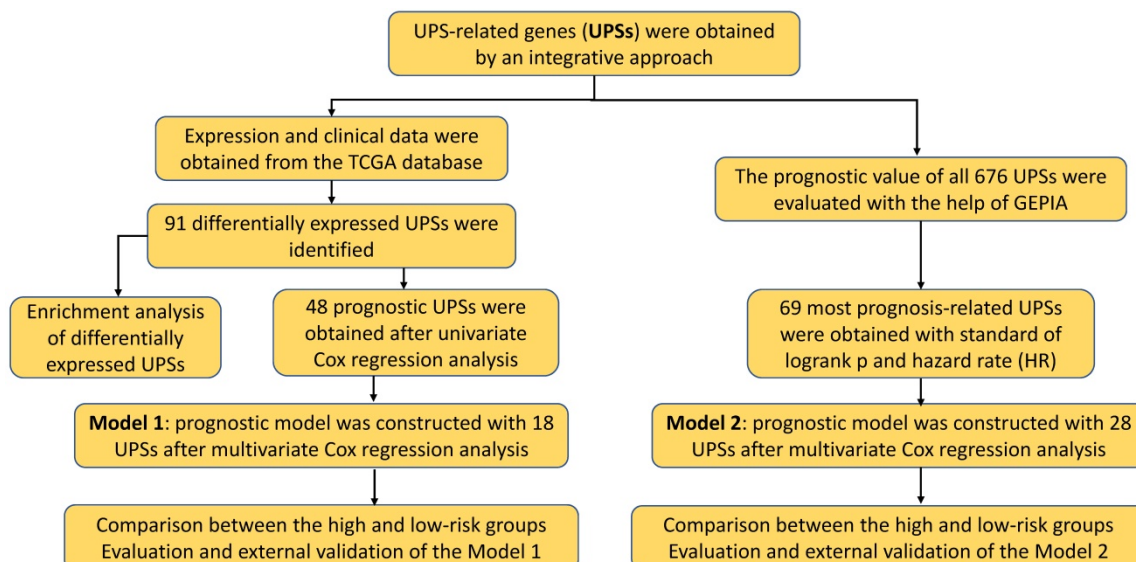


Figure 1. Overview flowchart of this study. Exploration of the prognostic role of UPS in ccRCC from a bioinformatic perspective.

them are not clearly elucidated and waiting for more extensively studies. We also visualized the expression pattern of these genes below (**Figure 2**). It is not difficult to find that there are more upregulated genes than downregulated genes in tumor samples. These findings may show that tumor needs a relatively high activity of UPS to survive and imply these UPSs are potential novel targets for treatment.

Table 1. Clinical parameters of patients in TCGA-KIRC cohorts

Parameters	Cases	Proportion (%)
Age (y)		
≤65	352	65.55
>65	185	34.45
Gender		
Male	346	64.43
Female	191	35.57
Stage		
I	269	50.37
II	57	10.67
III	125	23.41
IV	83	15.55
T stage		
T1	275	51.21
T2	69	12.85
T3	182	33.89
T4	11	2.05
N stage		
N0	240	93.39
N1	17	6.61
M stage		
M0	426	84.36
M1	79	15.64
Grade		
G1	14	2.65
G2	230	43.48
G3	207	39.13
G4	78	14.74

Functional enrichment of the differentially expressed UPSs

Functional enrichment analysis was performed with 91 differentially expressed UPSs. The GO term functional enrichment and the KEGG pathway enrichment analysis of these genes were summarized in **Figure 3**. In the GO terms in biological processes, differentially expressed UPSs were mainly enriched in protein polyubiquitination, proteasome-mediated ubiquitin-dependent protein catabolic process, proteasomal protein catabolic process, protein deubiquitination, protein modification by small protein removal. In the cellular components, differentially expressed UPSs were mainly enriched in ubiquitin ligase complex, cullin-RING ubiquitin ligase complex, proteasome core complex, proteasome core complex (beta-subunit complex), nuclear ubiquitin ligase complex. The top enriched GO terms in the molecular functions were ubiquitin-protein transferase activity, ubiquitin-like protein transferase

activity, ubiquitin protein ligase activity, ubiquitin-like protein ligase activity, thiol-dependent ubiquitin-specific protease activity. In the KEGG pathway enrichment analysis, the UPSs were mainly enriched in Ubiquitin mediated proteolysis, Proteasome, Small cell lung cancer, Necroptosis, IL-17 signaling pathway, NOD-like receptor signaling pathway. The z scores of these pathways were > 0, indicating that the UPSs were upregulated in these pathways, which indicated a relatively high activity of these pathways in tumor samples. More details are shown in **Table S2**.

Table 2. Differentially expressed UPSs in tumor samples comparing with normal samples

Dysregulation	Differentially expressed UPSs
Downregulated	ASB15, HECW1, RNF150, USP44, FBXO2, RNF43, DTX1, KLHL13, MARCH10, CBLC, UCHL1, USP2, TRIM2, TRIM40, CADPS2, RHOBTB3, LNX1, IRF2BPL, AMFR, USP46, RNF182, ZNRF3, PELI2, KLHL21, AREL1, PDZRN3, DCAF11, RNF152, NEDD4L, BRCC3, ASB9, NEURL3
Upregulated	MDM4, KCTD13, TRAF2, PCGF1, TRAF3IP2, WSB1, TRIM31, UBE2L6, FBXL6, LRRC41, CHFR, KLHL17, NEURL1B, RBCK1, TRIM59, SH3RF3, RACK1, PML, SPSB1, RFPL4A, TRIM36, DTX2, TRIM15, TNFAIP3, USP41, RNF149, PRC1, BCL6B, PSMB10, TRIM22, PSMA8, ASB2, MARCH3, HECW2, MYC, CORO7, FBXO17, PSMB8, MARCH1, CDC20, CDCA3, TRIM46, DDB2, TRIM73, DTL, TRIM74, PSMB9, BIRC3, TRIML1, STAMBPL1, UHRF1, TRIML2, RNF113B, MARCH4, UBE2C, TRIM9, RASD2, IFNG, BIRC7

Model 1: construction of a prognostic signature with differentially expressed UPSs

In order to construct a prognostic signature, we first made univariate Cox regression analysis of survival with 91 differentially expressed UPSs, and 48 UPSs were found to be associated with the prognosis. In other words, more than half of the differentially expressed UPSs, up to 52.75%, have statistical significance with ccRCC patients' survival, which have surprised us a lot. After multivariate Cox regression analysis, 18 UPSs were identified and they were used to construct a prognostic signature. We named this prognostic signature "Model 1". The risk score formula, shown in **Table 3**, was used to calculate the risk score of each patient, and patients were classified into the high-risk group (n = 265) and the low-risk group (n = 265) comparing with the median risk score. The different expression pattern of these 18 UPSs in two groups was shown by a heatmap (**Figure 4B**). It is obvious that the expression pattern proves essentially different in two groups. Particularly, we can easily find that the CDC20 and UBE2C are highly expressed in the high-risk group, while in the low-risk group, USP2 is highly expressed.

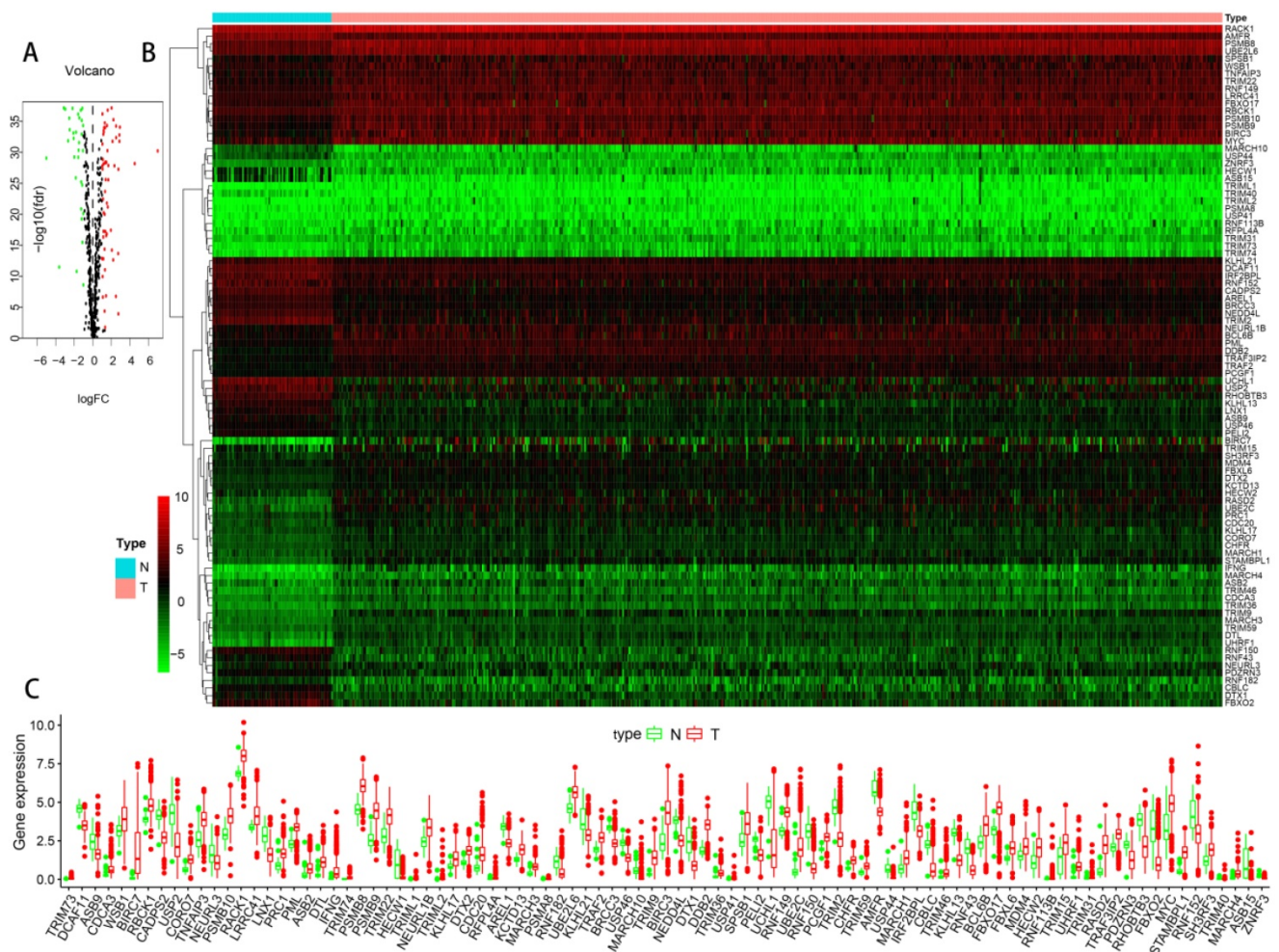


Figure 2. The expression profiles of UPSs between tumor samples and normal samples in TCGA cohort of ccRCC. (A) Volcano plot of 676 UPSs. The vertical axis indicates the $-\log_{10}$ False Discovery Rate (FDR), and the horizontal axis indicates the \log_2 fold change (FC). The red dots and the green dots represent up- and down-regulated genes, respectively (P -value < 0.05 and $|\log_2(\text{FC})| > 1$). (B) Heatmap of 91 differentially expressed UPSs. Red and green indicate higher expression and lower expression, respectively. (C) Box plot of the expression of 91 differentially expressed UPSs between tumor and normal tissues, tumor in red and normal in green.

Kaplan-Meier analysis was performed to evaluate the overall survival (OS) in the two groups. As is shown, a significant difference in the survival rate between the high- and low-risk groups was observed ($P = 3.775e-15$, **Figure 4A**). Patients in the high-risk group had a worse OS than those in the low-risk group. The survival rates of 1, 3, 5 years in low-risk group are about 95.0%, 87.2%, 81.2%, while in the high-risk group, they are significantly reduced to only about 83.7%, 62.6%, 37.8%. We also visualized the relationship between risk score and patients' survival state. As is shown in the dot plot, there is an increasing number of patients died with the increase of risk score (**Figure 4D**). These results support the presume that the risk score accurately reflect the survival of patients. In order to determine whether the UPS-related signature for OS is an independent prognostic factor, univariate and multivariate Cox regression analyses were performed. As the results of univariate Cox regression analysis suggested, age, stage, grade, T stage, N stage, M stage and risk score

were all significantly associated with OS in ccRCC patients. Multivariate Cox regression analysis showed that age and risk score were significantly associated with OS. These results support risk score as an independent prognostic factor in ccRCC patients.

In order to have a more comprehensive understanding of the difference between high and low-risk groups, we performed GSEA analysis. Results suggested that the pathways enriched in the two groups were really different (**Figure 5**). The following pathways were mostly enriched in the low-risk group: proximal tubule bicarbonate reclamation, vasopressin regulated water reabsorption, propanoate metabolism, pyruvate metabolism, tight junction. And in the high-risk group, the following pathways were mostly enriched: intestinal immune network for IgA production, cytokine-cytokine receptor interaction, homologous recombination, primary immunodeficiency, asthma. We can roughly conclude that the tumor tissues in low-risk group may have a better histologic

differentiation and a lower malignancy phenotype, for its main enrichment pathways are more likely to occur in normal renal cells which may mean the low-risk group tumor samples are more like normal tissues. On the contrary, the pathways enriched in the high-risk group are more related to immune response which may mean the tumor tissues are worse differentiated and have a higher malignancy which have induced a relatively more violently immune response against the tumor in the body.

Prognostic value of all UPSs

Apart from constructing a prognostic signature with differentially expressed UPSs, in order to have a more comprehensive understanding of the UPSs' role in ccRCC, we also analyze the prognostic value of all 676 UPSs with the help of GEPIA (Figure 6). With log-rank $p < 0.05$, we found 391 prognosis-related UPSs in total among 676 UPSs, which means 57.84% UPSs had a significant relationship with the overall

survival of ccRCC patients in TCGA cohort. Considering the fact that such a large proportion of UPSs had a significant correlation with the prognosis of ccRCC, we infer that the dysregulation of ubiquitin proteasome system should play a vital role in the occurrence and development of ccRCC.

With the convenience of GEPIA, we also made an exploration of the correlation of the expression of 69 most prognosis related genes and tumor stage. To our surprise, most of them have a significance relationship with tumor stage. If we see genes as "good genes" when high expression followed with better survival, and "bad genes" when high expression comes with worse survival. Then we can find there are 65 "good genes" among them, and most of 65 "good genes" are negatively correlate with tumor stage, while the only 4 "bad genes" (CDC20, CDCA3, FBXL6, UBE2C) are positively correlate with tumor stage (Figure 7).

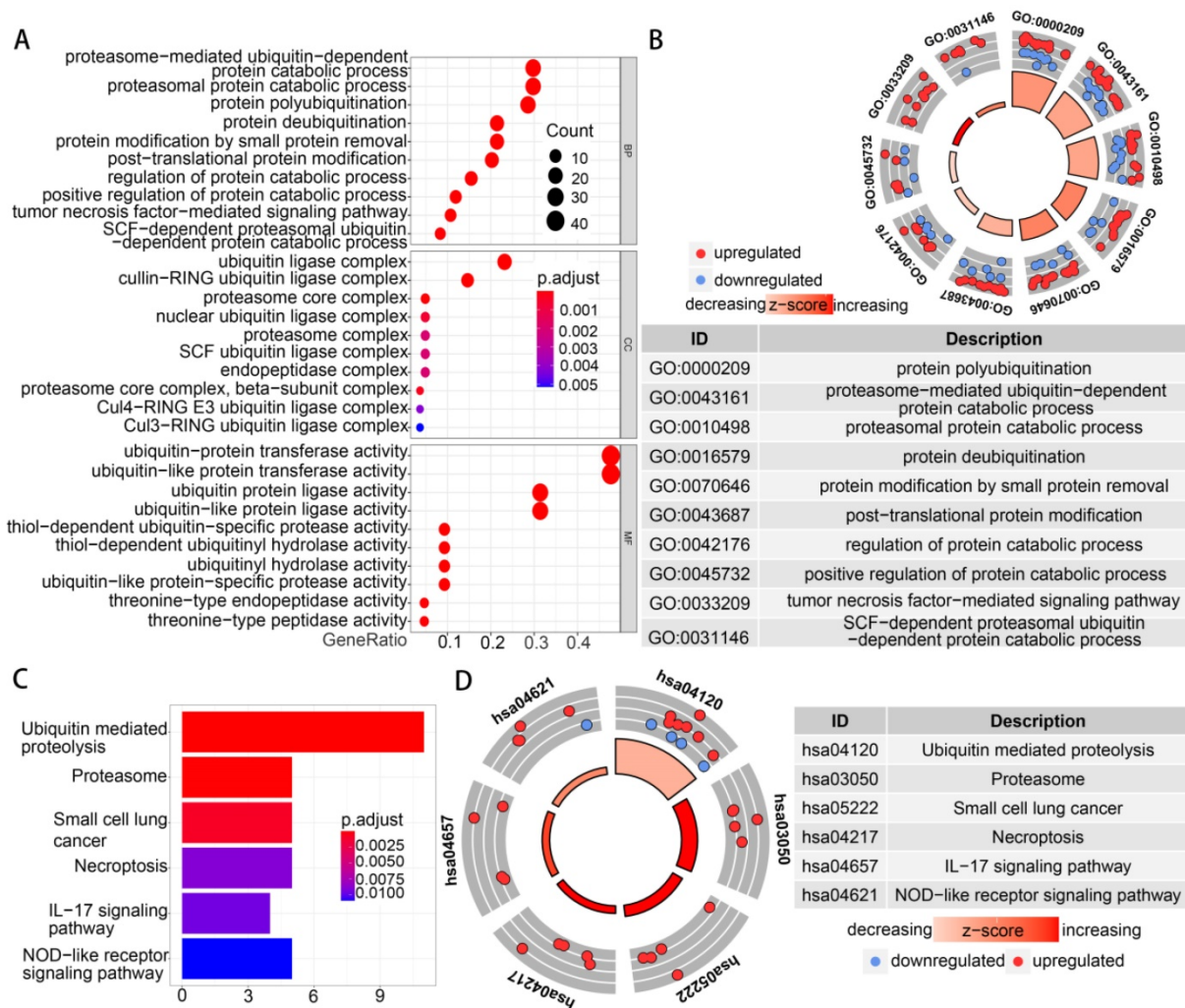


Figure 3. Functional enrichment of the differentially expressed UPSs. (A) The top 30 significant terms of GO function enrichment. BP biological process, CC cellular component, MF molecular function. (B) The GO circle shows the scatter map of the log FC of the specified gene. (C) The terms of KEGG analysis with statistical significance. (D) The KEGG circle shows the scatter map of the log FC of the specified gene. The higher the z-score value indicated, the higher expression of the enriched pathway.

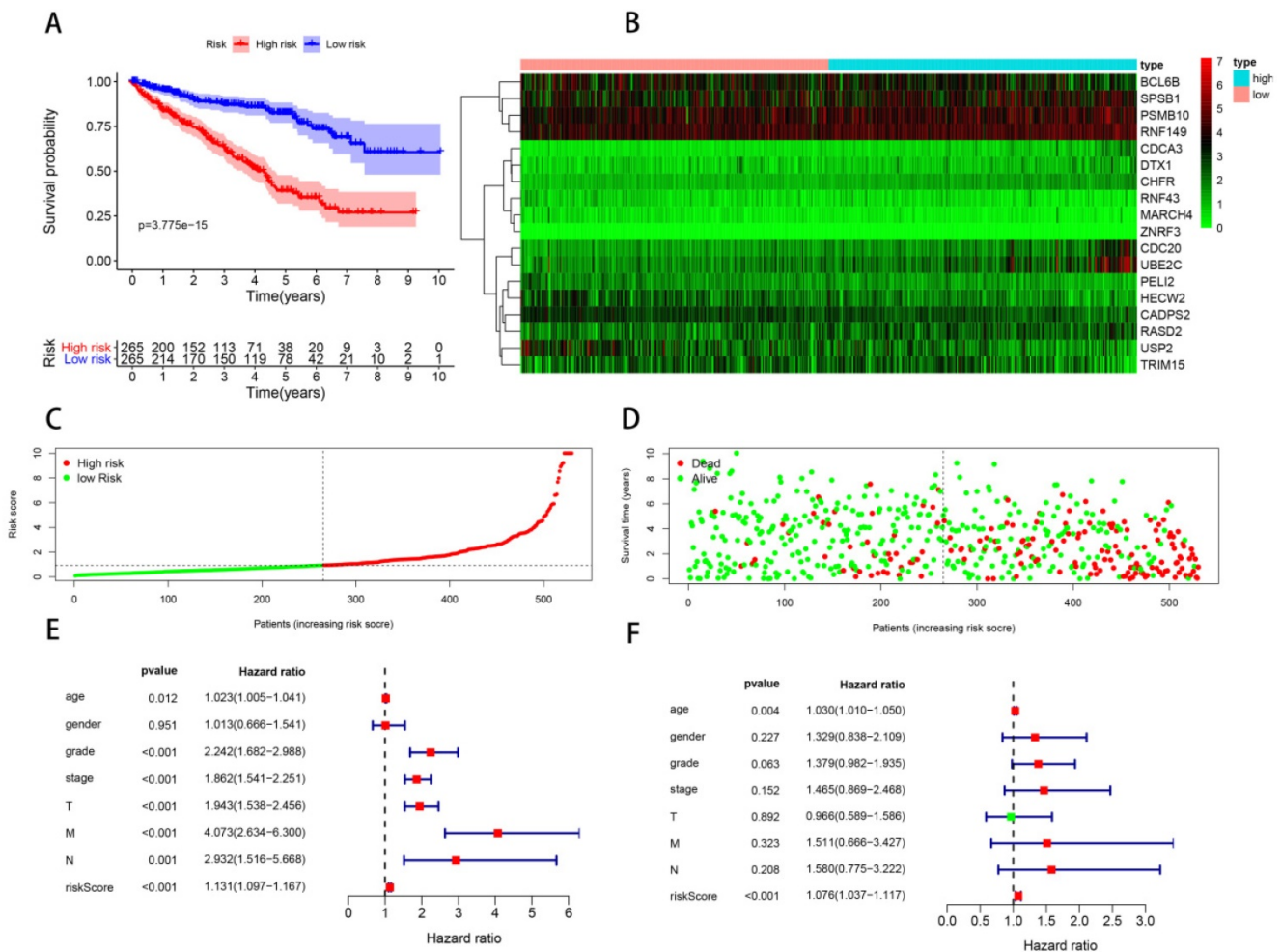


Figure 4. Model 1: prognostic signature constructed with 18 differentially expressed UPSs. (A) Kaplan-Meier curves of OS in the high- and low-risk groups stratified by the median risk score. (B) Heatmap of the expression profile of the model genes in two groups. (C) Distribution of the risk scores of ccRCC patients. (D) Survival status of patients in different groups, red dots denote patients that are dead, and green dots denote patients that are alive. (E) A forest plot of univariate Cox regression analysis in the cohorts. (F) A forest plot of multivariate Cox regression analysis in the cohorts.

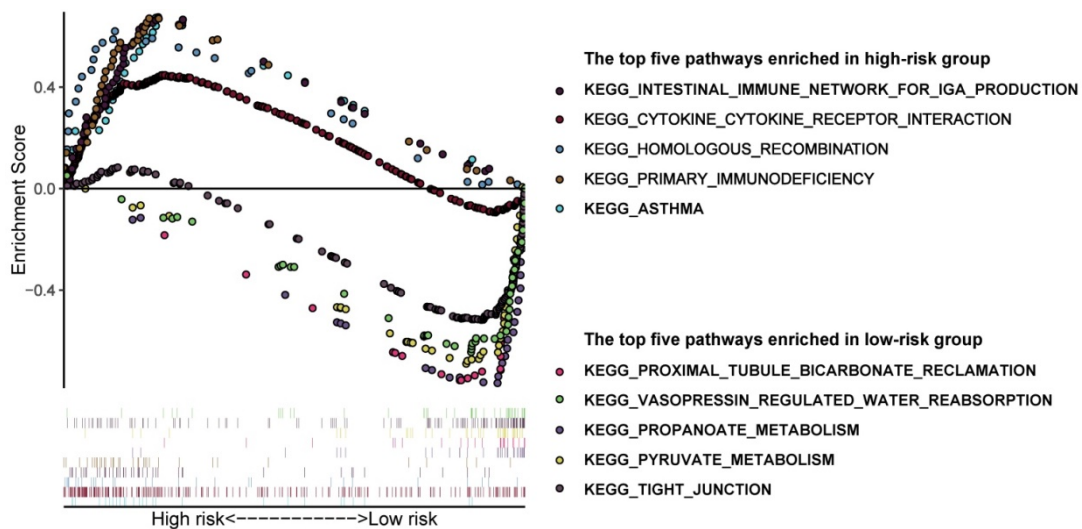


Figure 5. Gene set enrichment analysis in different groups identified by Model 1. The top 5 pathways enriched in the high-risk group and low-risk group, respectively.

Model 2: construction of prognostic model with 28 prognosis-related UPSs

Making a prognostic model with genes selected

from differentially expressed UPSs may reject some candidates which have outstanding prognostic value but without differentially expressed significance. Considering this fact, we assume that a better

prognostic predicting model may be constructed with these 391 prognosis-related genes recognized above without considering whether there are differentially expressed significances or not. With the standard log-rank $p < 0.05$ and $|\log_2 \text{ hazard rate (HR)}| > 1$, we get 69 most prognosis-related genes with the help of GEPIA (Figure 6B). Our new prognostic model was formed with 28 genes of these 69 most prognosis-related UPSs by Cox regression analysis, we named it "Model 2", shown in Table 3, and patients were classified into the high-risk group ($n = 265$) and the low-risk group ($n = 265$), survival analysis show that the survival rate of 1, 3, 5 years in low-risk group are about 96.7%, 89.8%, 83.9%, while in the high-risk group, the survival rate of 1, 3, 5 years are significantly reduced to only about 81.8%, 60.1%, 37.8%. The difference of survival rate is more obvious between two groups comparing with Model 1. We also visualized the relationship between risk score and patients' survival state as what we had done with Model 1, Cox regression analysis was also made in Model 2 (Figure 8).

Table 3. Risk score formula of Model 1 and Model 2

Model 1		Model 2	
Genes	Coefficient	Genes	Coefficient
CDCA3	0.647733	CDCA3	0.813077
CADPS2	-0.34575	RHOBTB1	-0.24355
USP2	-0.16609	FBXL5	-0.52466
PSMB10	-0.32733	PJA2	0.480908
CDC20	0.572892	FBXO7	0.57878
DTX1	0.257956	KLHL9	-1.31442
SPSB1	-0.19703	RNF20	-1.46314
PELI2	-0.40024	CDC20	0.596902
RNF149	0.506406	SOCS6	-0.57623
UBE2C	-0.43403	UBE2D3	-1.08593
CHFR	0.886869	OTUD7A	-1.42335
RNF43	-0.44588	PDE4D	-0.61161
BCL6B	0.293386	PELI2	-1.01604
HECW2	-0.49501	TOPORS	0.660042
RASD2	0.34123	PEX12	0.770969
MARCH4	-1.06908	UBE2C	-0.62467
ZNRF3	3.390968	TRIM2	-0.5344
TRIM15	-0.14047	EP300	-1.08857
		BFAR	-0.61988
		RAG1	-1.53854
		RNF38	0.751749
		MYLIP	0.50646
		KLHL8	1.201221
		TBLIX	0.309252
		TRAF6	0.762172
		FBXL4	0.962905
		FBXW2	1.622785
		ZNRF3	4.939872

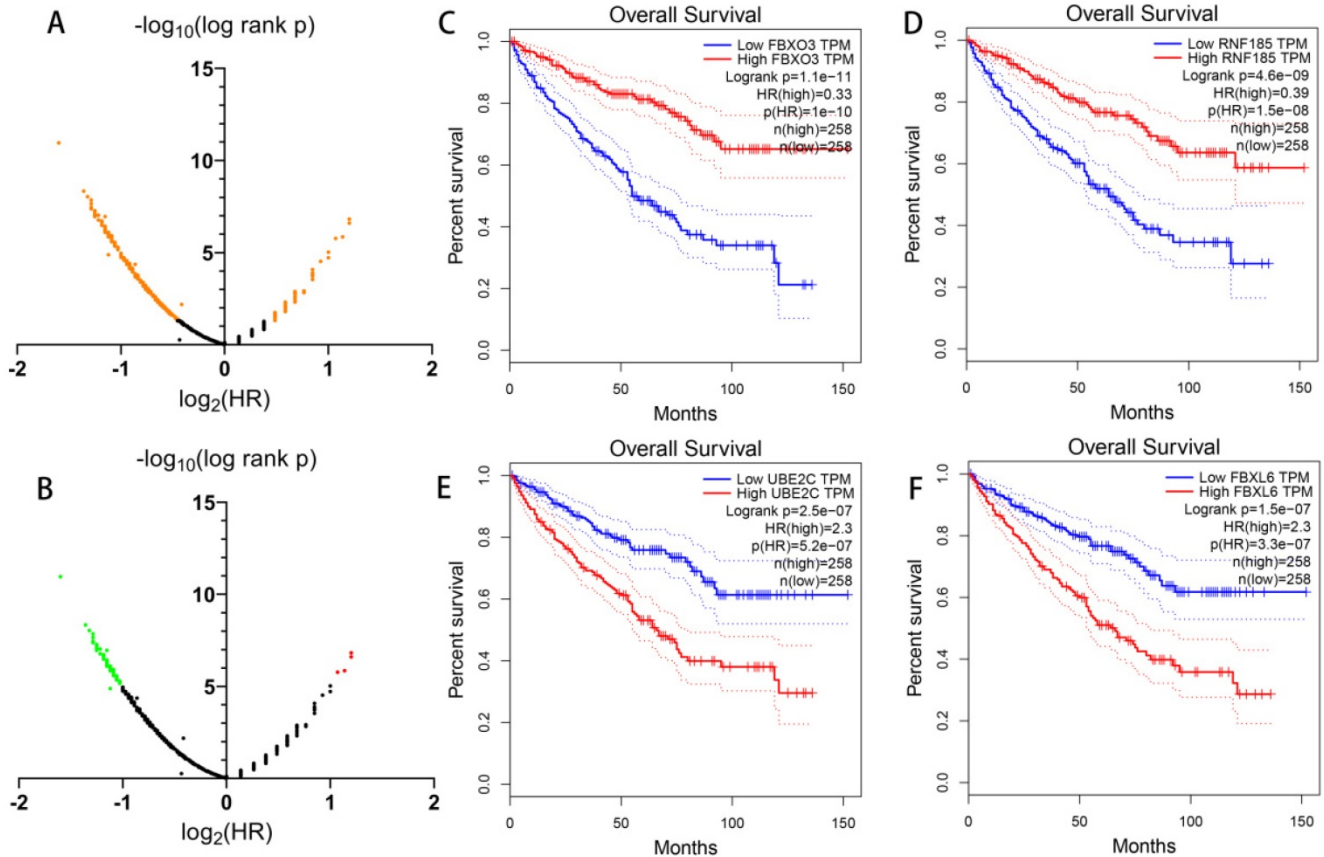


Figure 6. Prognostic value of all UPS-related genes exploring with GEPIA. (A) Dot plot to show the large proportion of UPSs with significant prognostic value, significant genes are shown in orange. (B) Dot plot to show the UPSs with outstanding prognostic value, red dots represent genes when $\log_2 \text{ hazard rate (HR)} > 1$, green dots represent genes when $\log_2 \text{ hazard rate (HR)} < -1$. (C-F) Kaplan-Meier curves of OS of representative genes with outstanding prognostic value.

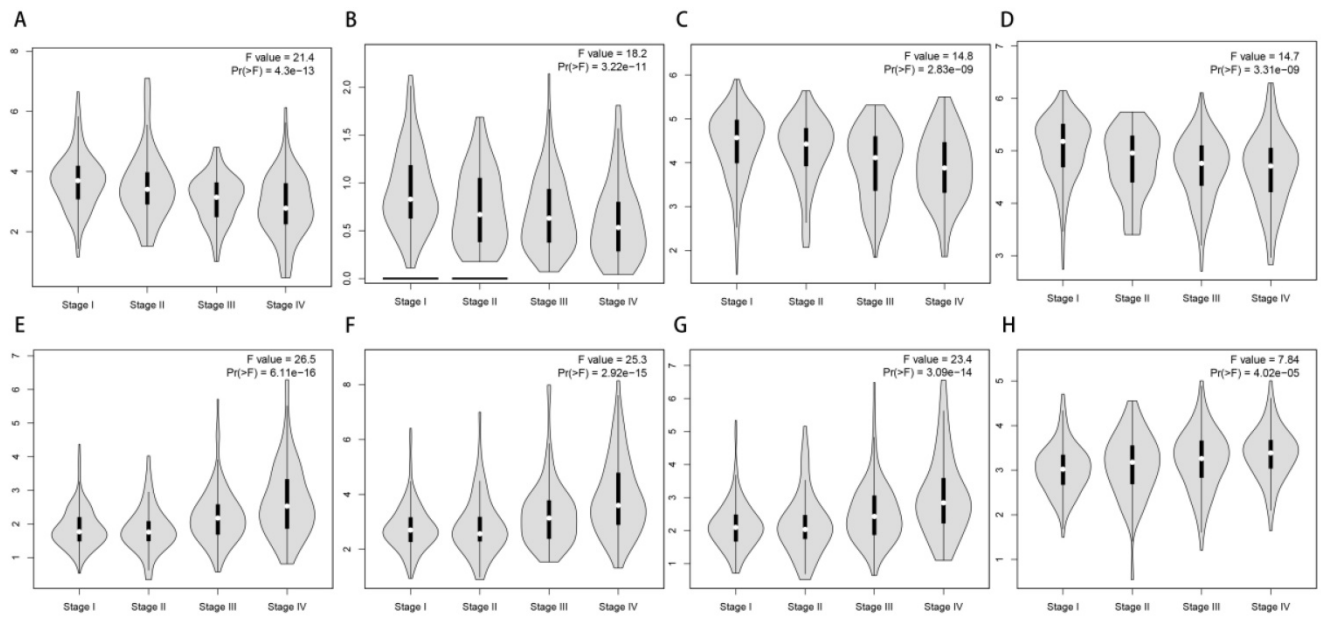


Figure 7. Correlation of the expression of UPSs with tumor stages. (A-D) 4 representative genes that have negative correlations with tumor stages. They are TRIM21 (A), OTUD7A (B), RCHY1 (C), DCAF11 (D). (E-H) 4 representative genes that have positive correlations with tumor stages. They are CDCA3 (E), UBE2C (F), CDC20 (G), FBXL6 (H).

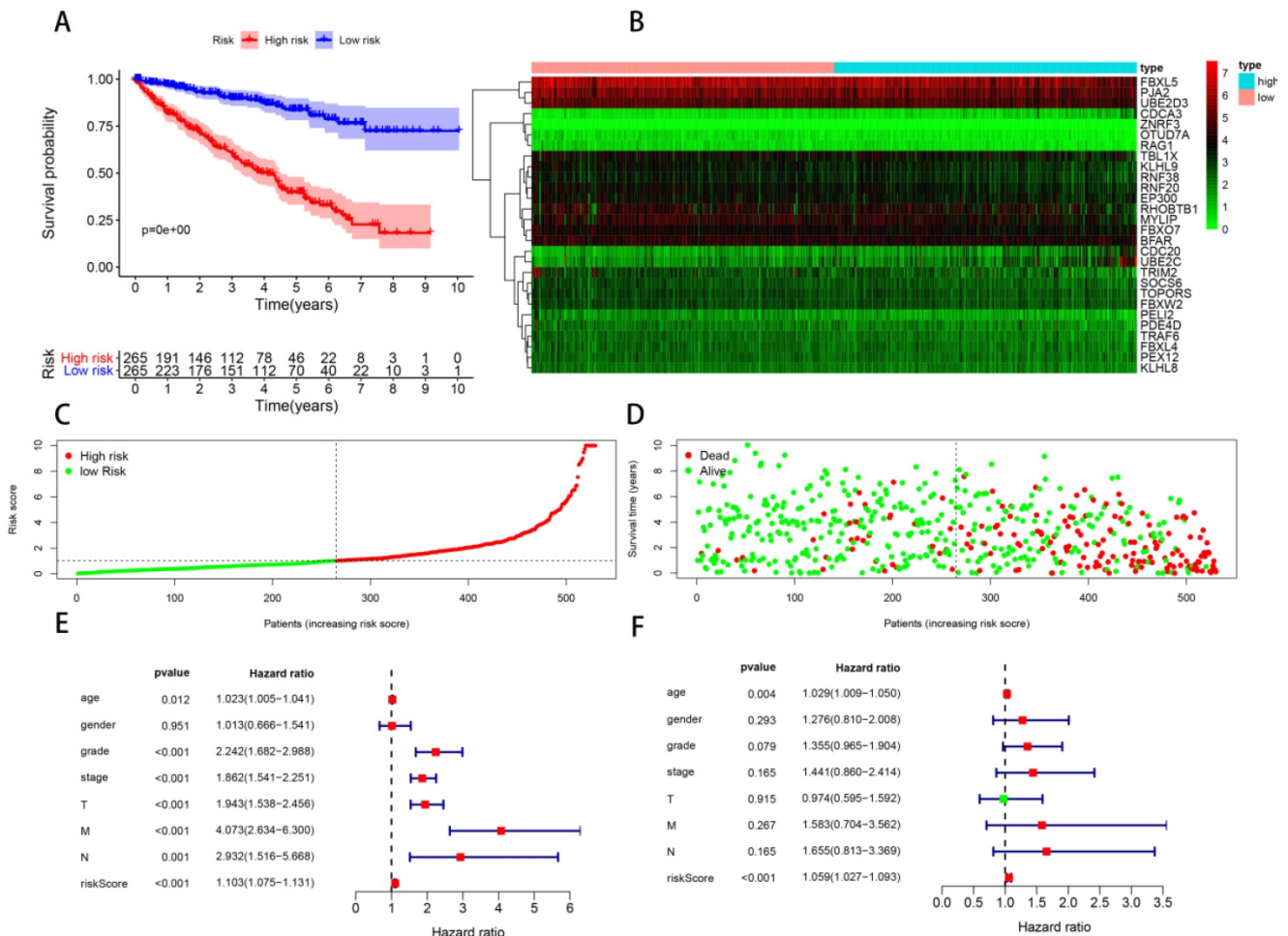


Figure 8. Model 2: prognostic signature constructed with 28 prognosis related UPSs. (A) Kaplan-Meier curves of OS in the high- and low-risk groups stratified by the median risk score. (B) Heatmap of the expression profile of the model genes in two groups. (C) Distribution of the risk scores of ccRCC patients. (D) Survival status of patients in different groups, red dots denote patients that are dead, and green dots denote patients that are alive. (E) A forest plot of univariate Cox regression analysis in the cohorts. (F) A forest plot of multivariate Cox regression analysis in the cohorts.

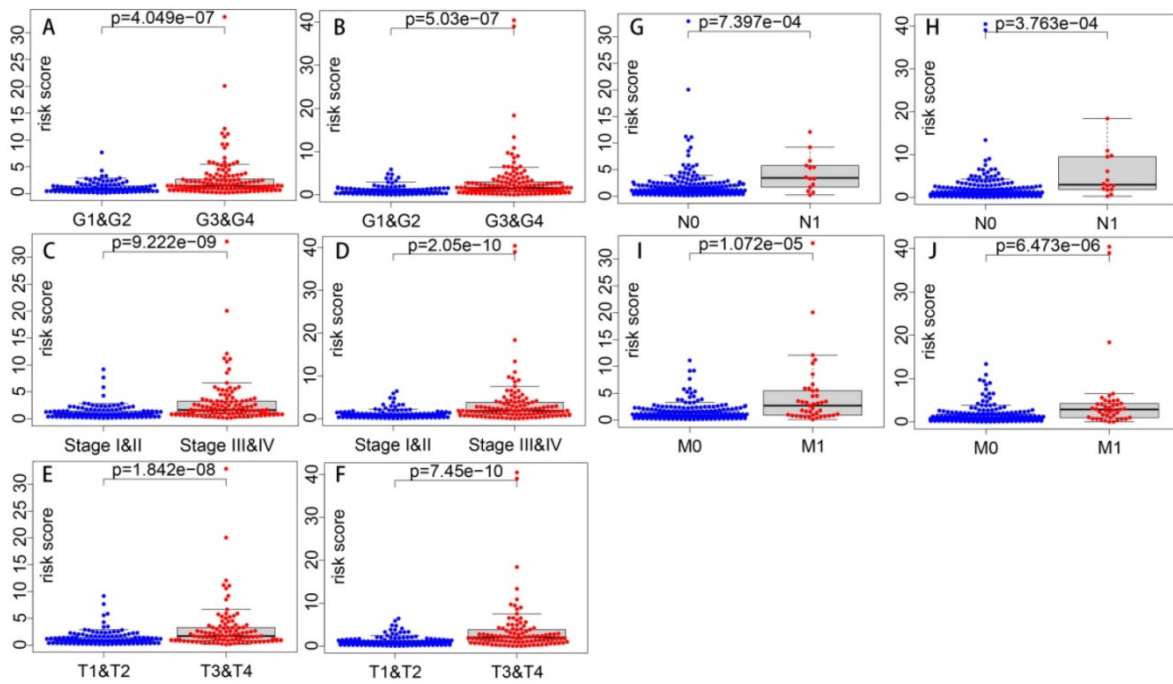


Figure 9. Correlations of risk score with clinical parameters with statistical significance. Correlations of risk score with grade in Model 1 (A) and Model 2 (B). Correlations of risk score with stage in Model 1 (C) and Model 2 (D). Correlations of risk score with T stage in Model 1 (E) and Model 2 (F). Correlations of risk score with N stage in Model 1 (G) and Model 2 (H). Correlations of risk score with M stage in Model 1 (I) and Model 2 (J).

Table 4. Correlations of risk score with clinical parameters (p value)

Clinical parameters	Models	
	Model 1	Model 2
Age	0.208	0.262
Gender	0.105	0.23
Grade	4.049e-07	5.03e-07
Stage	9.222e-09	2.05e-10
T	1.842e-08	7.45e-10
N	7.397e-04	3.763e-04
M	1.072e-05	6.473e-06

Estimation and external validation of Model 1 and Model 2

Correlations of the risk score with clinical parameters were shown in Table 4, and correlations with statistical significance were shown in Figure 9. Results show that the risk score is strongly correlated with tumor grade, tumor stage, T stage, N stage and M stage. And the higher the risk score is, the higher the tumor grade and the later the tumor stage is likely to be.

ROC curves were constructed to determine the predictive accuracy of the two different models. The area under the curve (AUC) for OS was 0.760 in Model 1, and 0.825 in Model 2, indicating both of them had good predictive accuracy, and Model 2 may be better (Figure 10). In order to further validate the prognostic predicting ability of our models, we downloaded expression and clinical data of 39 ccRCC patients in GSE29609 from GEO database, and

calculated their risk scores with the formulas of Model 1 and Model 2, respectively. Patients in GSE29609 were classified into high and low risk groups with the median risk score of Model 1 and Model 2, respectively. As Kaplan–Meier curves in Figure 10 show, the p values are as small as 3.069e-06 and 1.079e-04, indicating that our models remain effectively in this external cohort. All in all, the success of construction and validation of prognostic model with UPSs by two different approaches has strengthened our belief that the ubiquitin proteasome system (UPS) is of great importance in ccRCC, and models made with UPSs could predict the prognosis of patients effectively.

Discussion

As early as 1991, Kanayama had explored the changes in expressions of ubiquitin and proteasome genes in renal cancer cells by Northern blot as well as immunochemical analysis, and they drew a conclusion that the ubiquitin and proteasome system should play a role in the renal cancer [36]. In the past few decades, researchers have made great efforts to uncover the underlying mechanism of the UPS in the development of ccRCC [37-41]. It is indisputable that we have made tremendous achievements in this field, however, on the other hand, it is unanimously agreed that there remain enormous appealing mechanisms waiting us to uncover. At the same time, there are some attempts trying to translate experimental findings of ccRCC into clinical applications. For

instance, some UPS components are identified as pharmaceutical target to inhibit. Given bortezomib as an example, which is a proteasome inhibitor, the clinical trials have been done, but results are not very satisfactory [42, 43]. There is no doubt that there is still a long way to go. Our results suggest there are more upregulated UPSs in tumor and UPS related pathways are highly enriched, while prognosis analysis implies there are more UPSs whose high expression followed with better survival. This seems paradoxical and indicates us UPS is definitively complicated and we should keep open-minded with UPS in searching potential targets and solutions [44, 45]. Our differentially expressed analysis has identified some novel UPSs, for instance, RNF150, TRIM40, IRF2BPL, AREL1, NEURL3, FBXL6, LRRC41, KLHL17, RFPL4A, USP41, RNF149, PSMA8, CORO7, TRIM73, TRIM74, TRIM11, TRIM12, RNF113B and MARCH4. These UPSs have not been intensively studied and should be worth of further research.

When it comes to the prognostic value of UPS in ccRCC, there are researches focusing on specific UPS related genes, such as USP2 [21] and SMURF1 [46]. But there are less systematically estimations of the

whole ubiquitin proteasome system. With the development of the high-throughput sequencing and emergence of bioinformatic methods, such an exploration is practicable as well as attractive. Our study initially was aimed to draw a vivid picture of UPS in ccRCC from a bioinformatic perspective. Firstly, we conducted bioinformatic analysis of differential expression, functional enrichment and constructed Model 1 in a traditional way. 18 UPSs were involved in this prognostic predicting model. Some genes in Model 1 have been extensively studied such as CDC20 and ZNRF3. CDC20 was found to promote tumor cell migration and invasion for it was involved in the degradation of SMAR1, and SMAR1 is a tumor suppressor [47]. The coefficient of CDC20 in our model is positive which means the higher CDC20 expression is, the higher the risk score will be, which is consistent with the previous researches. As for ZNRF3, a higher expression of ZNRF3 was found to be related with a better survival in colorectal carcinoma [48], while its role in ccRCC has not been elucidated. As its coefficient in our model is not only positive but also large, ZNRF3 may also play an important role in ccRCC.

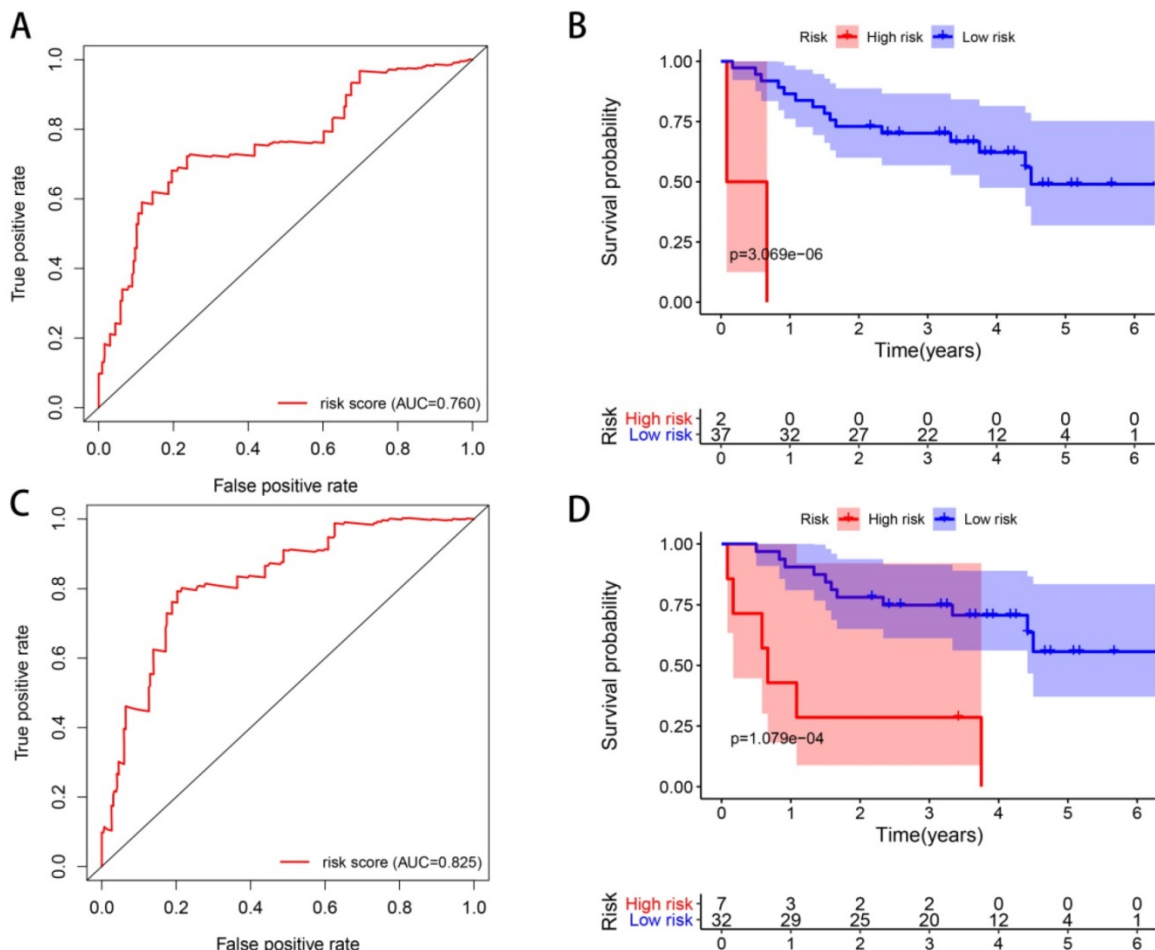


Figure 10. ROC curves and external validations of Model 1 and Model 2. ROC curves of Model 1 (A) and Model 2 (C). Kaplan–Meier curves of OS in the high- and low-risk groups stratified by the median risk score of Model 1 (B) and Model 2 (D) in GSE29609 cohort.

With the help of GEPIA, we made a comprehensive analysis of the prognostic value of all 676 UPSs. Results show that 391 of them, which accounts for 57.84%, have significant difference between two groups separated by the expression value. It is so amazing to find such a big proportion of UPSs is strongly related with prognosis. This may at least suggest that there is an extensive dysregulation of UPS in ccRCC tissues. Additionally, we found many UPSs also have a strong relationship with tumor stage. These results indicate that the prognosis difference of UPSs may be partially explained by tumor stage, and some UPSs indeed have a good prediction of tumor stage. Furtherly, with the 69 most prognosis-related genes found in GEPIA, we constructed another prognostic model which was named Model 2. Model 2 is more exciting for it has a better predicting ability and with AUC up to 0.825. Model 2 is not only a better model maybe useful for further clinical diagnosis, but also strong evidence implying that the dysregulation of UPS must have something to do in ccRCC. Recently, research suggests that ccRCC is a metabolic disease with metabolic reprogramming covering a wide range of biological processes, such as fatty acid metabolism, aerobic glycolysis and amino acid metabolism [49]. As a key pathway for protein degradation, UPS is also reported to be an essential modulator of cancer metabolism [50]. Obviously, there requires more studies and efforts focusing on this field to develop novel diagnostic and therapeutic methods.

Finally, we estimated the relationship of the risk score with clinical parameters, and risk score was shown to have positive correlations with traditional parameters that can indicate the tumor malignancy. And further external validation with GSE29609 cohort successfully showed similar prognostic values of our models. However, our research also has some limitations. Firstly, our models seem to be a little complex for there are many genes involved in them. Secondly, further experiments *in vivo* and *in vitro* are still needed to validate the diagnostic and therapeutic value of these genes and our models. Additionally, there are some good prognosis models for ccRCC already [51-55]. However, our models are very different from each other for the focus of our researches varies, and all signatures are treasures for they have enhanced our understanding of ccRCC. And in this study, we mainly focus on the ubiquitin proteasome system. If the small steps of our exploration do benefit patients suffering from ccRCC, we will feel gratified.

Conclusion

In conclusion, we made a comprehensive exploration of the prognostic role of UPS in ccRCC from a bioinformatic perspective. UPS is dysregulated in ccRCC. UPS related genes have significant prognostic value in ccRCC. Models constructed with UPSs are effective and applicable. An abnormal ubiquitin proteasome system should play an important role in ccRCC. The ubiquitin proteasome system is a promising target for ccRCC and deserves further study.

Abbreviations

AUC: area under the curve; ccRCC: clear cell renal cell carcinoma; DUBs: deubiquitinating enzymes; E1s: ubiquitin activating enzymes; E2s: ubiquitin conjugating enzymes; E3s: ubiquitin ligases; FC: fold change; FDR: false discovery rate; GO: gene ontology; GSEA: Gene Set Enrichment Analysis; HIF: hypoxia-inducible-factor; HR: hazard rate; KEGG: Kyoto Encyclopedia of Genes and Genomes; KIRC: Kidney renal clear cell carcinoma; OS: overall survival; RCC: renal cell carcinoma; ROC: receiver operating characteristic; TCGA: the cancer genome atlas; UPS: ubiquitin proteasome system; UPSs: UPS-related genes; VHL: Von-Hippel Lindau.

Supplementary Material

Supplementary table.

<http://www.jcancer.org/v12p4134s1.pdf>

Acknowledgements

Funding

This work was supported by the National Natural Science Foundation of China (Grant No. 81970661 and NO. 81670687), the Tai Shan Scholar Foundation to Benkang Shi.

Availability of data and materials

The results <published or shown> here are in whole or part based upon data generated by the TCGA Research Network: <https://www.cancer.gov/tcga>.

Author Contributions

HG and HG designed the study, conducted the data analysis and wrote the manuscript. LC, ZG, LZ and NZ conducted the data process. YL and YL revised and edited the manuscript. BS reviewed the manuscript and supervised the project. All authors have read and concurred with the content of the manuscript, and this manuscript has been approved by all authors for publication.

Competing Interests

The authors have declared that no competing interest exists.

References

- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin.* 2019; 69: 7-34.
- Lopez-Beltran A, Carrasco JC, Cheng L, Scarpelli M, Kirkali Z, Montironi R. 2009 update on the classification of renal epithelial tumors in adults. *Int J Urol.* 2009; 16: 432-43.
- Athar U, Gentile TC. Treatment options for metastatic renal cell carcinoma: a review. *Can J Urol.* 2008; 15: 3954-66.
- Rossi SH, Blick C, Handforth C, Brown JE, Stewart GD, Renal Cancer Gap Analysis C. Essential Research Priorities in Renal Cancer: A Modified Delphi Consensus Statement. *Eur Urol Focus.* 2020; 6: 991-8.
- Hawkins R, Fife K, Hurst M, Wang M, Naicker N, Nolasco S, et al. Treatment patterns and health outcomes in metastatic renal cell carcinoma patients treated with targeted systemic therapies in the UK. *BMC Cancer.* 2020; 20: 670.
- Chen D, Dou QP. The ubiquitin-proteasome system as a prospective molecular target for cancer treatment and prevention. *Curr Protein Pept Sci.* 2010; 11: 459-70.
- Hershko A, Ciechanover A. The ubiquitin system. *Annu Rev Biochem.* 1998; 67: 425-79.
- Cartel M, Didier C. Regulation of CHK1 by the Ubiquitin-Proteasome System. *FEBS J.* 2020; 287: 1982-4.
- Glickman MH, Ciechanover A. The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. *Physiol Rev.* 2002; 82: 373-428.
- Ishizawa J, Yoshida S, Oya M, Mizuno R, Shinjima T, Marumo K, et al. Inhibition of the ubiquitin-proteasome pathway activates stress kinases and induces apoptosis in renal cancer cells. *Int J Oncol.* 2004; 25: 697-702.
- Zhang X, Linder S, Bazzaro M. Drug Development Targeting the Ubiquitin-Proteasome System (UPS) for the Treatment of Human Cancers. *Cancers (Basel).* 2020; 12: 902.
- Mani A, Gelmann EP. The ubiquitin-proteasome pathway and its role in cancer. *J Clin Oncol.* 2005; 23: 4776-89.
- Dalla Via L, Nardon C, Fregona D. Targeting the ubiquitin-proteasome pathway with inorganic compounds to fight cancer: a challenge for the future. *Future Med Chem.* 2012; 4: 525-43.
- Johnson DE. The ubiquitin-proteasome system: opportunities for therapeutic intervention in solid tumors. *Endocr Relat Cancer.* 2015; 22: T1-17.
- Di Costanzo A, Del Gaudio N, Conte L, Altucci L. The Ubiquitin Proteasome System in Hematological Malignancies: New Insight into Its Functional Role and Therapeutic Options. *Cancers (Basel).* 2020; 12: 1898.
- Antao AM, Tyagi A, Kim KS, Ramakrishna S. Advances in Deubiquitinating Enzyme Inhibition and Applications in Cancer Therapeutics. *Cancers (Basel).* 2020; 12: 1579.
- Tundo GR, Sbardella D, Santoro AM, Coletta A, Oddone F, Grasso G, et al. The proteasome as a druggable target with multiple therapeutic potentialities: Cutting and non-cutting edges. *Pharmacol Ther.* 2020; 213: 107579.
- Guerrero-Garcia TA, Gandolfi S, Laubach JP, Hideshima T, Chauhan D, Mitsiades C, et al. The power of proteasome inhibition in multiple myeloma. *Expert Rev Proteomics.* 2018; 15: 1033-52.
- Saavedra-Garcia P, Martini F, Auner HW. Proteasome inhibition in multiple myeloma: lessons for other cancers. *Am J Physiol Cell Physiol.* 2020; 318: C451-C62.
- Keefe SM, Nathanson KL, Rathmell WK. The molecular biology of renal cell carcinoma. *Semin Oncol.* 2013; 40: 421-8.
- Meng X, Xiong Z, Xiao W, Yuan C, Wang C, Huang Y, et al. Downregulation of ubiquitin-specific protease 2 possesses prognostic and diagnostic value and promotes the clear cell renal cell carcinoma progression. *Ann Transl Med.* 2020; 8: 319.
- Zhou J, Wang T, Qiu T, Chen Z, Ma X, Zhang L, et al. Ubiquitin-specific protease-44 inhibits the proliferation and migration of cells via inhibition of JNK pathway in clear cell renal cell carcinoma. *BMC Cancer.* 2020; 20: 214.
- Tapia-Laliena MA, Korzeniewski N, Pena-Llopis S, Scholl C, Frohling S, Hohenfellner M, et al. Cullin 5 is a novel candidate tumor suppressor in renal cell carcinoma involved in the maintenance of genome stability. *Oncogenesis.* 2019; 8: 4.
- Zhao W, Zhou J, Deng Z, Gao Y, Cheng Y. SPOP promotes tumor progression via activation of beta-catenin/TCF4 complex in clear cell renal cell carcinoma. *Int J Oncol.* 2016; 49: 1001-8.
- Jiao D, Huan Y, Zheng J, Wei M, Zheng G, Han D, et al. UHRF1 promotes renal cell carcinoma progression through epigenetic regulation of TXNIP. *Oncogene.* 2019; 38: 5686-99.
- Roos FC, Evans AJ, Brenner W, Wondergem B, Klomp J, Heir P, et al. Deregulation of E2-EPF ubiquitin carrier protein in papillary renal cell carcinoma. *Am J Pathol.* 2011; 178: 853-60.
- Jin J, Li X, Gygi SP, Harper JW. Dual E1 activation systems for ubiquitin differentially regulate E2 enzyme charging. *Nature.* 2007; 447: 1135-8.
- Stewart MD, Ritterhoff T, Klevit RE, Brzovic PS. E2 enzymes: more than just middle men. *Cell Res.* 2016; 26: 423-40.
- Nijman SM, Luna-Vargas MP, Velds A, Brummelkamp TR, Dirac AM, Sixma TK, et al. A genomic and functional inventory of deubiquitinating enzymes. *Cell.* 2005; 123: 773-86.
- Nguyen VN, Huang KY, Weng JT, Lai KR, Lee TY. UbiNet: an online resource for exploring the functional associations and regulatory networks of protein ubiquitylation. *Database (Oxford).* 2016; 2016: baw054.
- Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet.* 2003; 34: 267-73.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005; 102: 15545-50.
- Walter W, Sanchez-Cabo F, Ricote M. GPlot: an R package for visually combining expression data with functional analysis. *Bioinformatics.* 2015; 31: 2912-4.
- Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* 2017; 45: W98-W102.
- Edeline J, Mottier S, Vigneau C, Jouan F, Perrin C, Zerrouki S, et al. Humiliation of 2 angiogenic phenotypes in clear cell renal cell carcinoma. *Histo Pathol.* 2012; 43: 1982-90.
- Kanayama H, Tanaka K, Aki M, Kagawa S, Miyaji H, Satoh M, et al. Changes in expressions of proteasome and ubiquitin genes in human renal cancer cells. *Cancer Res.* 1991; 51: 6677-85.
- Xiao J, Zhang R, Peng J, Yang Z. BAP1 maintains chromosome stability by stabilizing DIDO1 in renal cell carcinoma. *Am J Cancer Res.* 2020; 10: 1455-66.
- Hong K, Hu L, Liu X, Simon JM, Ptacek TS, Zheng X, et al. USP37 promotes deubiquitination of HIF2alpha in kidney cancer. *Proc Natl Acad Sci U S A.* 2020; 117: 13023-32.
- Li H, Qu L, Zhou R, Wu Y, Zhou S, Zhang Y, et al. TRIM13 inhibits cell migration and invasion in clear-cell renal cell carcinoma. *Nutr Cancer.* 2020; 72: 1115-24.
- Fukasawa H, Yamamoto T, Fujigaki Y, Misaki T, Ohashi N, Takayama T, et al. Reduction of transforming growth factor-beta type II receptor is caused by the enhanced ubiquitin-dependent degradation in human renal cell carcinoma. *Int J Cancer.* 2010; 127: 1517-25.
- Yu S, Dai J, Ma M, Xu T, Kong Y, Cui C, et al. RBCK1 promotes p53 degradation via ubiquitination in renal cell carcinoma. *Cell Death Dis.* 2019; 10: 254.
- Corn PG. Role of the ubiquitin proteasome system in renal cell carcinoma. *BMC Biochem.* 2007; 8 Suppl 1: S4.
- Dou QP, Zonder JA. Overview of proteasome inhibitor-based anti-cancer therapies: perspective on bortezomib and second generation proteasome inhibitors versus future generation inhibitors of ubiquitin-proteasome system. *Curr Cancer Drug Targets.* 2014; 14: 517-36.
- Park J, Cho J, Song EJ. Ubiquitin-proteasome system (UPS) as a target for anticancer treatment. *Arch Pharm Res.* 2020; 43: 1144-61.
- Celebi G, Kesim H, Ozer E, Kutlu O. The Effect of Dysfunctional Ubiquitin Enzymes in the Pathogenesis of Most Common Diseases. *Int J Mol Sci.* 2020; 21: 6335.
- Ke M, Mo L, Li W, Zhang X, Li F, Yu H. Ubiquitin ligase SMURF1 functions as a prognostic marker and promotes growth and metastasis of clear cell renal cell carcinoma. *FEBS Open Bio.* 2017; 7: 577-86.
- Paul D, Ghorai S, Dinesh US, Shetty P, Chattopadhyay S, Santra MK. Cdc20 directs proteasome-mediated degradation of the tumor suppressor SMAR1 in higher grades of cancer through the anaphase promoting complex. *Cell Death Dis.* 2017; 8: e2882.
- Yu N, Zhu H, Tao Y, Huang Y, Song X, Zhou Y, et al. Association between prognostic survival of human colorectal carcinoma and ZNRF3 expression. *Onco Targets Ther.* 2016; 9: 6679-87.
- Lucarelli G, Loizzo D, Franzin R, Battaglia S, Ferro M, Cantiello F, et al. Metabolomic insights into pathophysiological mechanisms and biomarker discovery in clear cell renal cell carcinoma. *Expert Rev Mol Diagn.* 2019; 19: 397-407.

50. Sun T, Liu Z, Yang Q. The role of ubiquitination and deubiquitination in cancer metabolism. *Mol Cancer*. 2020; 19: 146.
51. Wang S, Zhang L, Yu Z, Chai K, Chen J. Identification of a Glucose Metabolism-related Signature for prediction of Clinical Prognosis in Clear Cell Renal Cell Carcinoma. *J Cancer*. 2020; 11: 4996-5006.
52. Li R, Wang L, Wang X, Geng RX, Li N, Liu XH. Identification of hub genes associated with outcome of clear cell renal cell carcinoma. *Oncol Lett*. 2020; 19: 2846-60.
53. Pan Q, Wang L, Zhang H, Liang C, Li B. Identification of a 5-Gene Signature Predicting Progression and Prognosis of Clear Cell Renal Cell Carcinoma. *Med Sci Monit*. 2019; 25: 4401-13.
54. Shen C, Liu J, Wang J, Zhong X, Dong D, Yang X, et al. Development and validation of a prognostic immune-associated gene signature in clear cell renal cell carcinoma. *Int Immunopharmacol*. 2020; 81: 106274.
55. Hu M, Xie J, Hou H, Liu M, Wang J. Prognostic Value of DNA Methylation-Driven Genes in Clear Cell Renal Cell Carcinoma: A Study Based on Methylation and Transcriptome Analyses. *Dis Markers*. 2020; 2020: 8817652.