

# Identification of biomarkers of clear cell renal cell carcinoma by bioinformatics analysis

Ning Zhang, MD<sup>a</sup>, Wenxin Chen, MD<sup>a</sup>, Zhilu Gan, MD<sup>a</sup>, Alimujiang Abudurexiti, MD<sup>a</sup>, Xiaogang Hu, MD<sup>a</sup>, Wei Sang, PhD, MD<sup>b,\*</sup>

## Abstract

Clear cell renal cell carcinoma (ccRCC) is the most common subtype among renal cancer, and more and more researches find that the occurrence of ccRCC is associated with genetic changes, but the molecular mechanism still remains unclear. The present study aimed to identify aggregation trend of differentially expressed genes (DEGs) in ccRCC, which would be beneficial to the treatment of ccRCC and provide research ideas using a series of bioinformatics approach. Gene ontology (GO) and Kyoto Encyclopedia of Gene and Genomes (KEGG) analysis were used to get the enrichment trend of DEGs of GSE53757 and GSE16449. Draw Venn Diagram was applied for co-expression of DEGs. Cytoscape with the Retrieval of Interacting Gene (STRING) datasets and Molecular Complex Detection (MCODE) were performed protein-protein interaction (PPI) of DEGs. The Kaplan–Meier Plotter analysis of top 15 upregulated and top 15 downregulated were selected in Gene Expression Profiling Interactive Analysis (GEPIA). Then, the expression level of hub genes between normal renal tissue and different pathological stages of ccRCC tissue, which significantly correlated with overall survival in ccRCC patients, were also analyzed by Ualcan based on The Cancer Genome Atlas (TCGA) database. In this study, we got 167 co-expression DEGs, including 72 upregulated DEGs and 95 downregulated DEGs. We identified 11 hub genes had significantly correlated with overall survival in ccRCC patients. Among them, KIF23, APLN, ADCY1, GREB1, TLR4, IRF8, CXCL1, CXCL2, deserved our attention.

**Abbreviations:** BP = biological process, CC = cellular component, ccRCC = clear cell renal cell carcinoma, DAVID = Database for Annotation, Visualization and Integrated Discovery online tool, DEGs = differentially expressed genes, GEO = Gene Expression Omnibus, GEPIA = Gene Expression Profiling Interactive Analysis, GO = gene ontology, HR = hazard ratio, KEGG = Kyoto Encyclopedia of Gene and Genomes, MCODE = Molecular Complex Detection, MF = molecular function, mTOR = mammalian target of rapamycin, PPI = predicted protein-protein interaction, RCC = renal cell carcinoma, STRING = Search Tool for the Retrieval of Interacting Genes, TCGA = The Cancer Genome Atlas, TKI = multi-targeted tyrosine kinase inhibitors, VEGF = vascular endothelial growth factor.

**Keywords:** clear cell renal cell carcinoma, differentially expressed genes, survival prognosis

Editor: Josephine Kam Tai Dermawan.

This work was funded by the Natural Science Foundation of Uygur Autonomous Region of Xinjiang, China (No. 2017D01C312).

The authors have no conflicts of interest to disclose.

Our data analyzed come from the databases of GSE53757 and GSE16449, and the original data of GSE53757 and GSE16449 can be download in <https://www.ncbi.nlm.nih.gov/gds/> and <https://www.ncbi.nlm.nih.gov/geo/geo2/>.

The datasets generated during and/or analyzed during the current study are publicly available.

<sup>a</sup> Surgery Department of Urology, The Third People's Hospital of Xinjiang Uygur Autonomous Region, Urumqi, Xinjiang, <sup>b</sup> The Department of Pathology, The First Affiliated Hospital of Xinjiang Medical University, Urumqi, Xinjiang, PR China.

\* Correspondence: Wei Sang, The Department of Pathology, The First Affiliated Hospital of Xinjiang Medical University, 137 South Liyushan Road, Urumqi, Xinjiang 830054, PR China (e-mail: [qqsangweiqq@163.com](mailto:qqsangweiqq@163.com)).

Copyright © 2020 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

How to cite this article: Zhang N, Chen W, Gan Z, Abudurexiti A, Hu X, Sang W. Identification of biomarkers of clear cell renal cell carcinoma by bioinformatics analysis. *Medicine* 2020;99:21(e20470).

Received: 21 June 2019 / Received in final form: 28 February 2020 / Accepted: 28 April 2020

<http://dx.doi.org/10.1097/MD.00000000000020470>

## 1. Introduction

Renal carcinoma is a common tumor in the urinary system, which accounts for 2% to 3% of adult malignancies.<sup>[1]</sup> About 90% of renal carcinoma are renal cell carcinoma (RCC), the majority of which (70–85%) are clear cell renal cell carcinoma (ccRCC).<sup>[2]</sup> ccRCC can be classified into four pathological stages, including grade I to grade IV. And grade IV has poor prognosis, of which 5-year survival rate is 20%.<sup>[3,4]</sup>

Due to the poor response to conventional radiotherapy and chemotherapy of RCC patients, surgical resection is still the most effective treatment.<sup>[5]</sup> However, once the disease became metastatic, targeted therapies will be the best choice for them.<sup>[6]</sup> Unfortunately, after nearly 2 years of treatment, most of these patients gradually become resistance.<sup>[7]</sup> In this case, more research efforts were prompted into targeted therapeutics, such as multi-targeted tyrosine kinase inhibitors (TKI) and mammalian target of rapamycin (mTOR) inhibitors,<sup>[8]</sup> which have undergone radical breakthrough in the last decade.

The clinical knowledge that ccRCC is a typical hypervascular cancer which tumor cells can promote the growth and progression of tumor through producing pro-angiogenesis factors.<sup>[9,10]</sup> Thus, the vascular endothelial growth factor (VEGF) signaling, especially TKI, played an important role in targeted therapies of ccRCC.<sup>[11]</sup> Furthermore, other therapeutic strategies, such as immune checkpoint inhibitors, holding greater promises. Newer agents,

including varying gene, mRNA, and protein expression signatures were detected within the tumor specimen, and the changes in the dysregulation of mRNA have proved to affect ccRCC pathogenicity.<sup>[12,13]</sup> Beyond predicting prognosis, more effective biomolecular markers and therapeutic targets are urgent needed for the development of systemic therapies in ccRCC.<sup>[14]</sup>

Although many researches have focused on the screening of differentially expressed genes (DEGs) associated with ccRCC progression,<sup>[15–17]</sup> according to gene expression profiles, most of them ignore the high correlations between genes, although genes with similar expression patterns might be functionally related.<sup>[18]</sup> In our study, we downloaded original data of ccRCC from Gene Expression Omnibus (GEO), an online public collection database for different tumors. Two datasets of GSE53757 and GSE16449, were chosen by carefully selected, we compared tumor tissue and normal tissue to get identify DEGs, and analyzed gene ontology (GO) and Kyoto Encyclopedia of Gene and Genomes (KEGG) to get genetic change trend, then obtained genes with clinical guiding significance by survival analysis.

## 2. Methods

### 2.1. Data collection

We downloaded the gene expression profiles of GSE53757 and GSE16449 from GEO database (<http://www.ncbi.nlm.nih.gov/geo/>). The GSE53757 dataset included 144 samples, containing 72 ccRCC samples and 72 normal tissues. The GSE16449 dataset included 70 samples, containing 52 ccRCC samples and 18 normal tissues.

### 2.2. Data processing of DEGs

The analysis was carried out by using GEO2R (<https://www.ncbi.nlm.nih.gov/geo/geo2r/>), an interactive online tool allowing users to compare two or more groups of samples in a GEO series and it analyzed most GEO series with gene symbol. The adjusted *P* values were used to decrease the false positive rate using Benjamini and Hochberg false discovery rate method by default. The adjusted *P* value < .05 and  $|\log FC| > 1$  were set as the cutoff criterion. Then Draw Venn Diagram was used for co-expression of DEGs of GSE53757 and GSE16449.

### 2.3. Gene ontology and KEGG pathway analysis of DEGs

The Database for Annotation, Visualization and Integrated Discovery online tool (DAVID, <https://david.ncifcrf.gov/>) was used to perform GO functional and KEGG pathway enrichment analysis for the DEGs. The GO analysis, including biological process (BP), molecular function (MF), and cellular component (CC), was used to annotate genes and gene products, and also identify characteristic biological attributing to genomic or transcriptomic data. Kyoto Encyclopedia of Genes and Genomes was used to handle genomes, biological pathways, diseases, chemical substances and drugs. The *P* < .05 was considered to have significant differences.

### 2.4. PPI network and module analysis

The Search Tool for the Retrieval of Interacting Genes (STRING) was an online tool designed to predicted protein-protein interaction (PPI) information. The STRING database was used to analyze the PPIs among the proteins encoded by the DEGs with a combined score  $\geq 0.4$ , and maximum number of interactors = 0

were set as the cutoff criterion. Then the PPI networks for the upregulated and the downregulated genes were separately visualized by Cytoscape software (version 3.6.1).

In addition, the Molecular Complex Detection (MCODE) app was utilized to screen modules of the PPI network in Cytoscape with degree cutoff=2, node score cutoff=0.2, k-core=2, and max. depth=100.

### 2.5. Survival analysis of hub genes

Hub genes were a series of genes with the highest degree of connectivity in a gene module and determined the characteristics of a module. The Kaplan–Meier Plotter was used to measure hub genes. In the meantime, 30 genes, including 15 top upregulated genes and 15 top downregulated genes, were also inserted into Gene Expression Profiling Interactive Analysis (GEPIA) (<http://gepia.cancer-pku.cn/index.html>) with confidence score  $\geq 0.4$  and maximum number of interactors=0. The hazard ratio (HR) with 95% confidence intervals and log rank *P* value were calculated and displayed on the plot.

### 2.6. Comparison of the hub gene expression level

In order to evaluate the expression level of hub genes, which significantly correlated with overall survival in ccRCC patients, we used boxplot by Ualcan (<http://ualcan.path.uab.edu/index.html>) based on The Cancer Genome Atlas (TCGA) database (<https://genome-cancer.ucsc.edu/>). The integrated data of 72 normal and 533 primary tumor in TCGA samples were enrolled for analyses. Demographic, clinical and tumor pathological features of ccRCC patients are listed in Table 1. The expression level of hub genes between normal renal tissue and different pathological stages of ccRCC tissue were also analyzed.

### 2.7. Ethics statement

All analyses were based on the public GEO and TCGA databases,<sup>[19]</sup> we did not need the informed consent of the

**Table 1**  
Patient and tumor characteristics of ccRCC subtype cohorts in TCGA.

Variable	KIRC
Sample (n)	533
Gender	
Male	345
Female	188
Race	
Caucasian	462
African-American	56
Asian	8
Clinical stage	
Stage I	267
Stage II	57
Stage III	123
Stage IV	84
Tumor grade	
Grade 1	14
Grade 2	229
Grade 3	206
Grade 4	76

KIRC = kidney clear cell carcinoma, RCC subtype.

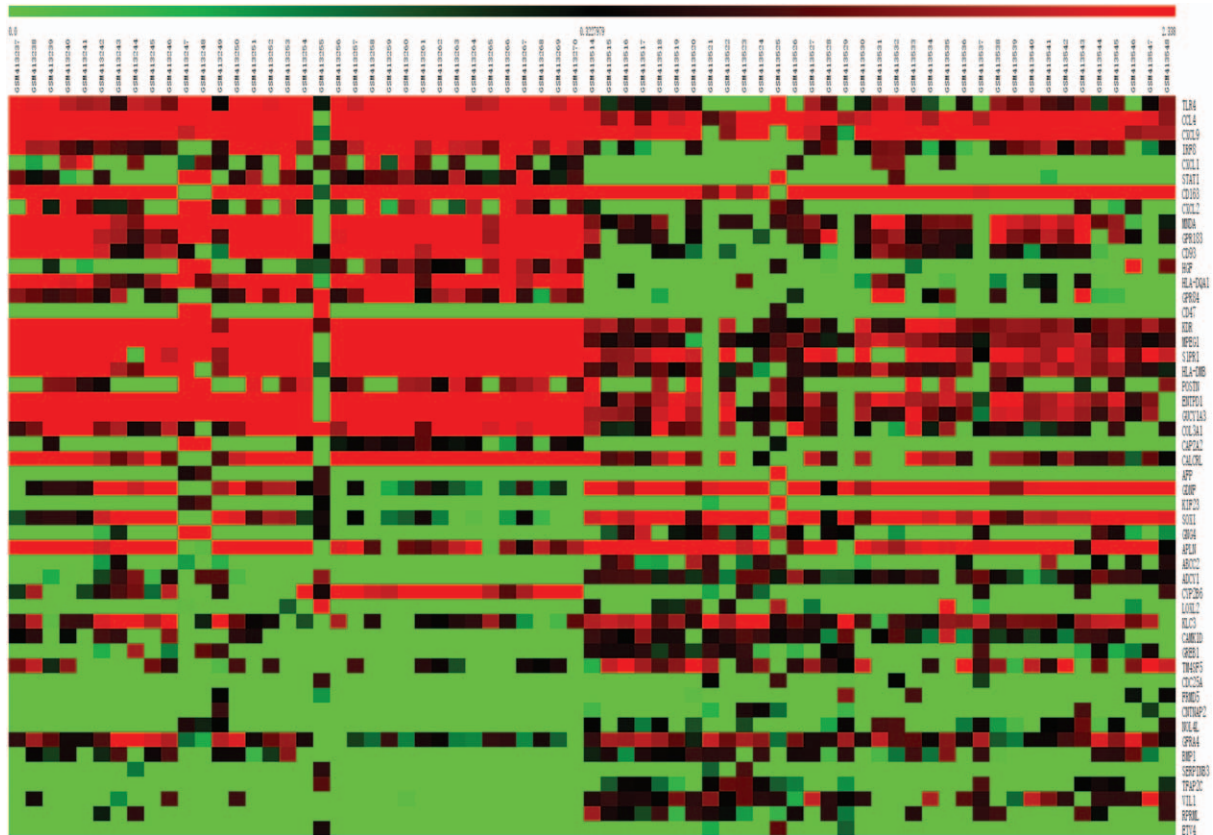


Figure 1. Heat map of the top upregulated 25 genes and the top downregulated 25 genes (red: upregulated, green: downregulated).

patients, thus no ethical approval and patient consent are required.

### 3. Results

#### 3.1. Identification of DEGs

GSE53757 selected 4542 DEGs, including 2441 upregulated DEGs and 2101 downregulated DEGs. GSE16449 selected 5308 DEGs, including 2219 upregulated DEGs and 3089 downregulated DEGs. A total of 167 co-expression of DEGs between GSE53757 and GSE16449 were detected by Draw Venn Diagram analysis, including 72 upregulated DEGs and 95 downregulated DEGs. The heat map, including 25 top upregulated genes and 25 top downregulated genes, was shown in Figure 1. The Volcano plot showed all co-expression genes of GSE53757 and GSE16449 (Fig. 2).

#### 3.2. GO function and KEGG pathway analysis of DEGs

The top five enriched terms of upregulated and downregulated DEGs were selected in Table 2 (Fig. 3), according to the *P* values. The DEGs were mainly enriched in BP, including cellular response to glucagon stimulus, negative regulation of phosphatase activity, ureteric bud formation, axo-dendritic transport and receptor-mediated endocytosis for upregulated DEGs, and for downregulated DEGs including immune response, chemokine-mediated signaling pathway, inflammatory response, chemotaxis, and cell chemotaxis. In MF, the upregulated DEGs were particularly enriched in microtubule binding, RNA polymerase II

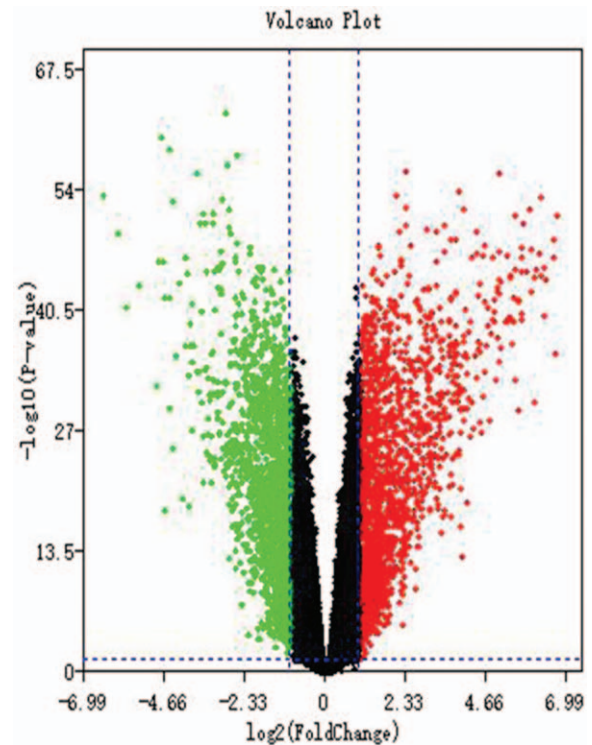


Figure 2. Volcano plot filtering of coexpression genes of GSE16449 and GSE57357 (red: up-regulated gene, green: down-regulated gene).

**Table 2****Gene ontology analysis of differentially expressed genes associated with ccRCC.**

Expression	Category	Term	Count	Ratio (%)	P value	FDR	
Upregulated	GOTERM_BP_DIRECT	GO:0071377~cellular response to glucagon stimulus	3	4.166666667	.008975164	12.01345286	
	GOTERM_BP_DIRECT	GO:0010923~negative regulation of phosphatase activity	3	4.166666667	.014307411	18.50058775	
	GOTERM_BP_DIRECT	GO:0060676~ureteric bud formation	2	2.777777778	.017740531	22.43892707	
	GOTERM_BP_DIRECT	GO:0008088~axo-dendritic transport	2	2.777777778	.017740531	22.43892707	
	GOTERM_BP_DIRECT	GO:0006898~receptor-mediated endocytosis	4	5.555555556	.028862714	34.01634845	
	GOTERM_MF_DIRECT	GO:0008017~microtubule binding	4	5.555555556	.041026449	40.42265788	
	GOTERM_MF_DIRECT	GO:0000978~RNA polymerase II core promoter proximal region sequence-specific DNA binding	5	6.944444444	.041361864	40.67976106	
	GOTERM_MF_DIRECT	GO:0019841~retinol binding	2	2.777777778	.046724112	44.65399992	
	GOTERM_MF_DIRECT	GO:0015179~L-amino acid transmembrane transporter activity	2	2.777777778	.046724112	44.65399992	
	GOTERM_MF_DIRECT	GO:0001077~transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding	4	5.555555556	.055903263	50.89373927	
	GOTERM_CC_DIRECT	GO:0032432~actin filament bundle	2	2.777777778	.032612849	31.37193962	
	GOTERM_CC_DIRECT	GO:0008021~synaptic vesicle	3	4.166666667	.045100525	40.78478735	
	Downregulated	GOTERM_BP_DIRECT	GO:0006955~immune response	13	14.13043478	2.43E-07	3.56E-04
		GOTERM_BP_DIRECT	GO:0070098~chemokine-mediated signaling pathway	5	5.434782609	2.56E-04	0.373813862
		GOTERM_BP_DIRECT	GO:0006954~inflammatory response	8	8.695652174	.001279643	1.853567282
		GOTERM_BP_DIRECT	GO:0006935~chemotaxis	5	5.434782609	.001961362	2.827923415
		GOTERM_BP_DIRECT	GO:0060326~cell chemotaxis	4	4.347826087	.002838318	4.068074672
GOTERM_MF_DIRECT		GO:0008009~chemokine activity	4	4.347826087	.001340009	1.620974265	
GOTERM_MF_DIRECT		GO:0005044~scavenger receptor activity	3	3.260869565	.019255957	21.09885484	
GOTERM_MF_DIRECT		GO:0003677~DNA binding	6	6.52173913	.877022802	100	
GOTERM_CC_DIRECT		GO:0005576~extracellular region	11	11.95652174	.195572641	91.98459912	
GOTERM_CC_DIRECT		GO:0042613~MHC class II protein complex	3	3.260869565	.004462142	5.054156031	
GOTERM_CC_DIRECT		GO:0005765~lysosomal membrane	6	6.52173913	.008294677	9.20763831	
GOTERM_CC_DIRECT		GO:0030666~endocytic vesicle membrane	4	4.347826087	.003390511	3.862118369	
GOTERM_CC_DIRECT		GO:0005634~nucleus	10	10.86956522	.999990054	100	

BP=biological process, CC=cell component, ccRCC=clear cell renal cell carcinoma, FDR=false discovery rate, GO=gene ontology, MF=molecular function.

core promoter proximal region sequence-specific DNA binding, retinol binding, L-amino acid transmembrane transporter activity and transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding, and the downregulated DEGs including chemokine activity, scavenger receptor activity, and DNA binding. In addition, the cell component (CC) analysis also displayed that the upregulated DEGs were significantly enriched in actin filament bundle and synaptic vesicle, and downregulated DEGs were mainly enriched in extracellular region, MHC class II protein complex, lysosomal membrane, endocytic vesicle membrane, and nucleus.

Table 3 shows the most significantly enriched KEGG pathway of upregulated and downregulated DEGs (Fig. 3). The upregulated DEGs were enriched in Dilated cardiomyopathy, GABAergic synapse and Glutamatergic synapse, and the downregulated DEGs were enriched in Phagosome, Inflammatory bowel disease (IBD), Asthma, Leishmaniasis, and Salmonella infection.

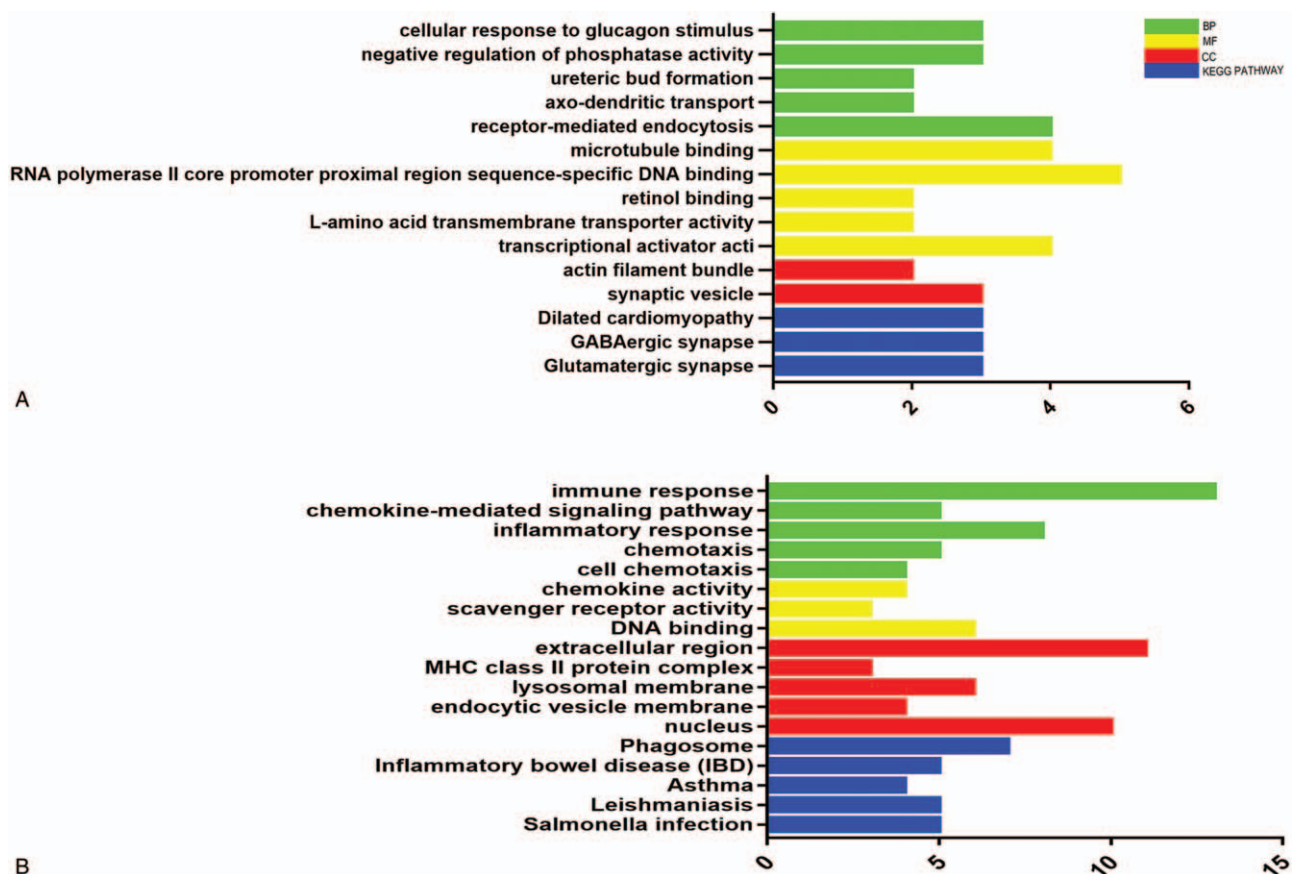
### 3.3. Hub genes and module screening from PPI network

Based on the information in the STRING protein relationship, we made the PPI network of the co-expression DEGs (Fig. 4a),

including 72 upregulated DEGs and 95 downregulated DEGs. The upregulated DEGs mainly had two modules by using the MCODE plug-in (Fig. 4b and c), including module 1 (AFP, ABCC2, CYP2B6), module 2 (GNG4, APLN, ADCY1). The downregulated DEGs mainly had four modules by using the MCODE plug-in (Fig. 4d–g), including module 3 (HLA-DMB, HLA-DQA1, CAPZA2, HLA-DOA), module 4 (TLR4, IRF8, CD163, GPR183, CXCL1, CXCL2, STAT1, HGF, S1PR1), module 5 (GPR20, CALCRL, GPR84, PTGDR), and module 6 (GUCY1A3, GUCY1B3, ENTPD1).

### 3.4. The Kaplan–Meier Plotter of the top 15 upregulated and top 15 downregulated DEGs with high degree of connectivity

The top 15 upregulated genes including AFP, GDNF, KIF23, GNG4, APLN, ABCC2, ADCY1, CYP2B6, LOXL2, KLC3, CASC5, GREB1, TM4SF5, CDC25A, FRMD5, while the top 15 downregulated genes including TLR4, CCL4, CXCL9, IRF8, CXCL1, STAT1, CD163, CXCL2, MNDA, GPR183, CD93, HGF, HLA-DQA1, GPR84, and CD47. The GEPIA showed that 11 hub genes, including ADCY1, APLN, FRMD5, GNG4,



**Figure 3.** The top GO terms and KEGG pathways enriched separately for the up-regulated DEGs (a) and the down-regulated DEGs (b). The horizontal axis represents the count of enriched DEGs. The vertical axis represents the enriched GO terms and KEGG pathway. BP=biological processes, CC=cellular component, DEGs=differentially expressed genes, GO=gene ontology, KEGG=Kyoto Encyclopedia of Gene and Genomes, MF=molecular function.

GREB1, KIF23, CXCL1, CXCL2, GPR84, IRF8, and TLR4, had significantly correlated with overall survival in ccRCC patients (Fig. 5a and b).

**3.5. Analysis of hub gene expression level**

After analyzed the TCGA database, we found KIF23 and APLN were highly expressed in ccRCC tissue compared normal renal tissue. However, the expressions of other nine genes have opposite results (Fig. 6a and b).

**4. Discussion**

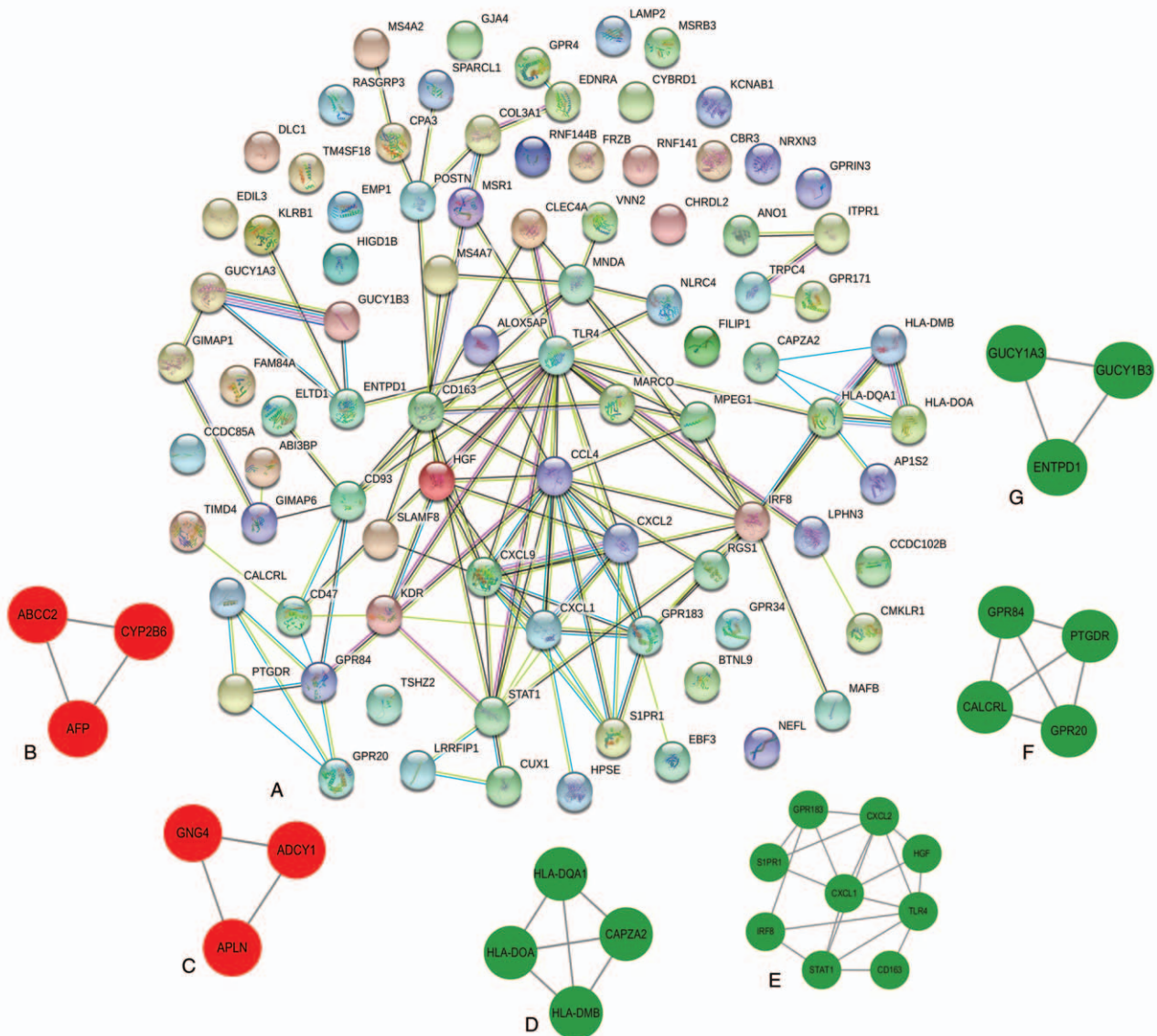
Although the use of targeted therapies, such as multi-kinase inhibitors, anti-VEGF antibodies and mTOR, has improved the median progression-free and overall survival of ccRCC patients for nearly 2 years,<sup>[6]</sup> the development of drug resistance is still one of the main causes of treatment failure, and the molecular mechanism of drug resistance phenomenon still remains unclear.<sup>[20]</sup> In this study, we identified DEGs between ccRCC and normal samples and used a series of bioinformatics analyses to screen key gene and pathways associated with cancer. To

**Table 3**

**KEGG pathway analysis of differentially expressed genes associated with ccRCC.**

Expression	Term	Count	Ratio (%)	P value	Genes	FDR
Upregulated	hsa05414:Dilated cardiomyopathy	3	4.166666667	.039622266	ADCY1, SGCD, CACNG4	35.64459489
	hsa04727:GABAergic synapse	3	4.166666667	.040484109	ADCY1, SLC38A1, GNG4	36.27141844
	hsa04724:Glutamatergic synapse	3	4.166666667	.068358036	ADCY1, SLC38A1, GNG4	53.78791249
Downregulated	hsa04145:Phagosome	7	7.608695652	1.29E-04	MARCO, LAMP2, MSR1, TLR4, HLA-DMB, HLA-DOA, HLA-DQA1	0.143747593
	hsa05321:Inflammatory bowel disease (IBD)	5	5.434782609	3.57E-04	TLR4, HLA-DMB, HLA-DOA, STAT1, HLA-DQA1	0.396788712
	hsa05310:Asthma	4	4.347826087	5.26E-04	MS4A2, HLA-DMB, HLA-DOA, HLA-DQA1	0.583680397
	hsa05140:Leishmaniasis	5	5.434782609	5.32E-04	TLR4, HLA-DMB, HLA-DOA, STAT1, HLA-DQA1	0.590181434
	hsa05132:Salmonella infection	5	5.434782609	9.61E-04	CXCL1, NLR4, CXCL2, TLR4, CCL4	1.063686392

ccRCC=clear cell renal cell carcinoma, KEGG=Kyoto Encyclopedia of Genes and Genomes, FDR=false discovery rate.



**Figure 4.** Protein-Protein interaction network for the up-regulated and down-regulated DEG (a), the Module 1 (b), the Module 2 (c), the Module 3 (d), the Module 4 (e), the Module 5 (f), and the Module 6 (g) (red: up-regulated Module, green: down-regulated Module). DEG=differentially expressed gene.

obtain an in-depth understanding of these DEGs, we performed the GO function and KEGG pathway analysis of them.

In order to screen the key candidate proteins in ccRCC, we conducted the GEO database of GSE53757 and GSE16449, and 167 co-expression genes, including 72 upregulated DEGs and 95 downregulated DEGs, were identified finally. After further investigating and validating Kaplan–Meier Plotter, the top 15 upregulated and top 15 downregulated DEGs were identified as the key candidate proteins with high degree of connectivity.

After analyzing the TCGA database, six genes in upregulated, including KIF23, GNG4, APLN, ADCY1, GREB1, and FRMD5, and five genes in downregulated, including TLR4, IRF8, CXCL1, CXCL2, and GPR84, may affect the survival time of ccRCC patients. Some of them deserved our attention and discussion. Kinesin family member 23 (KIF23), was a member of kinesin-like protein family. Previous study found that KIF23 was associated

with prognostic factors in patients with breast cancer,<sup>[21]</sup> and tightly related to progression of pancreatic ductal adenocarcinoma (PDAC),<sup>[22]</sup> but others found that the expression of KIF23 showed negative relation to the prognosis of pulmonary adenocarcinoma patients.<sup>[23]</sup> APLN, was encoded a peptide that functions as an endogenous ligand for the G-protein coupled apelin receptor. Yang L et al,<sup>[24]</sup> found that APLN might be a therapeutic potential biomarker in muscle-invasive bladder cancer patients, and also highly expressed in breast cancer, and associated with lymph node metastasis and TNM stage.<sup>[25]</sup> Adenylate cyclase 1 (ADCY1), was encoded a member of the of adenylate cyclase gene family that is primarily expressed in the brain. Hua Y et al,<sup>[26]</sup> found that ADCY1 might play essential roles in the metastasis process of RAC through pancreatic secretion and cell adhesion molecules pathways. Growth regulating estrogen receptor binding 1 (GREB1), was an estrogen-responsive gene that is an early response gene in the

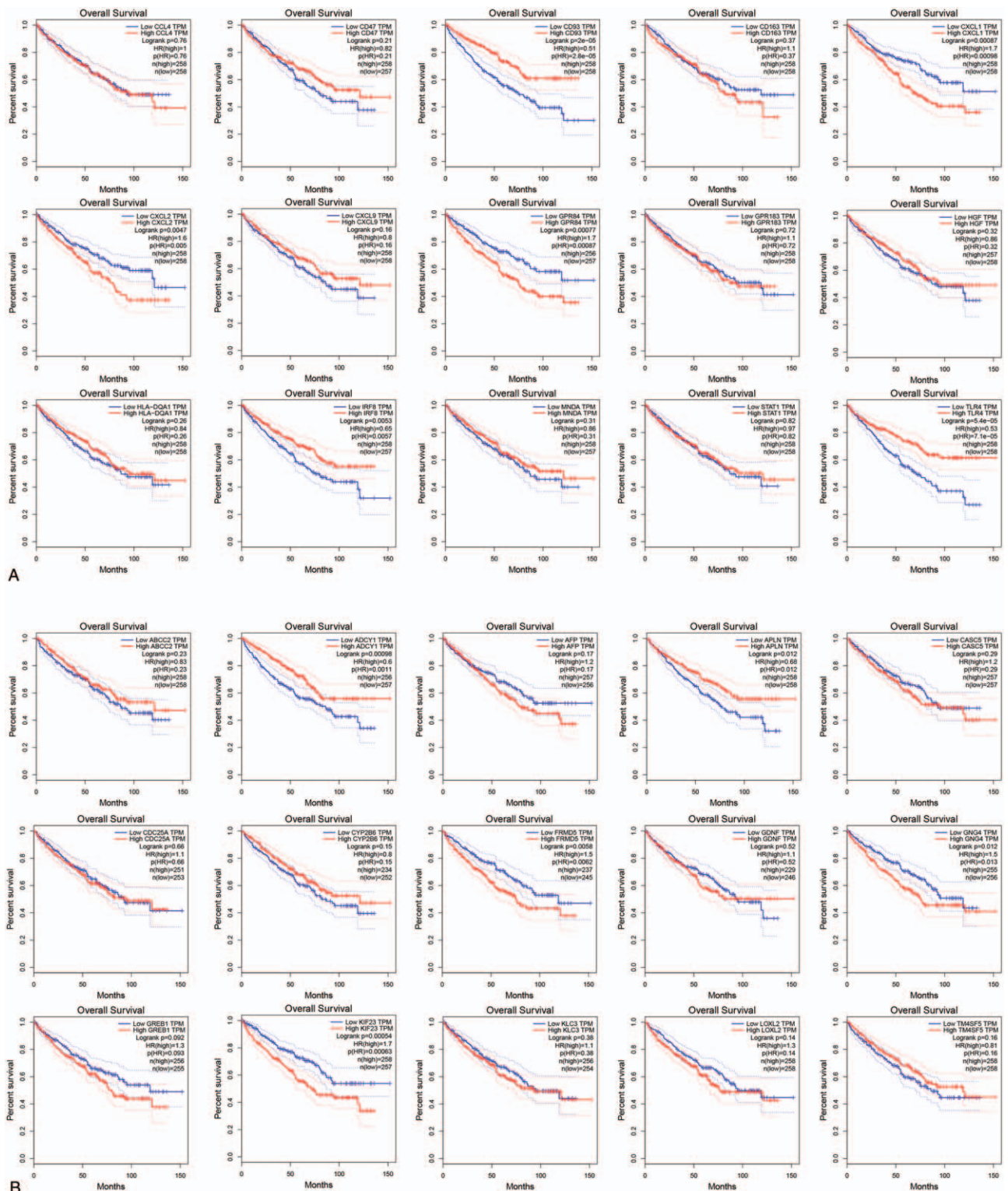
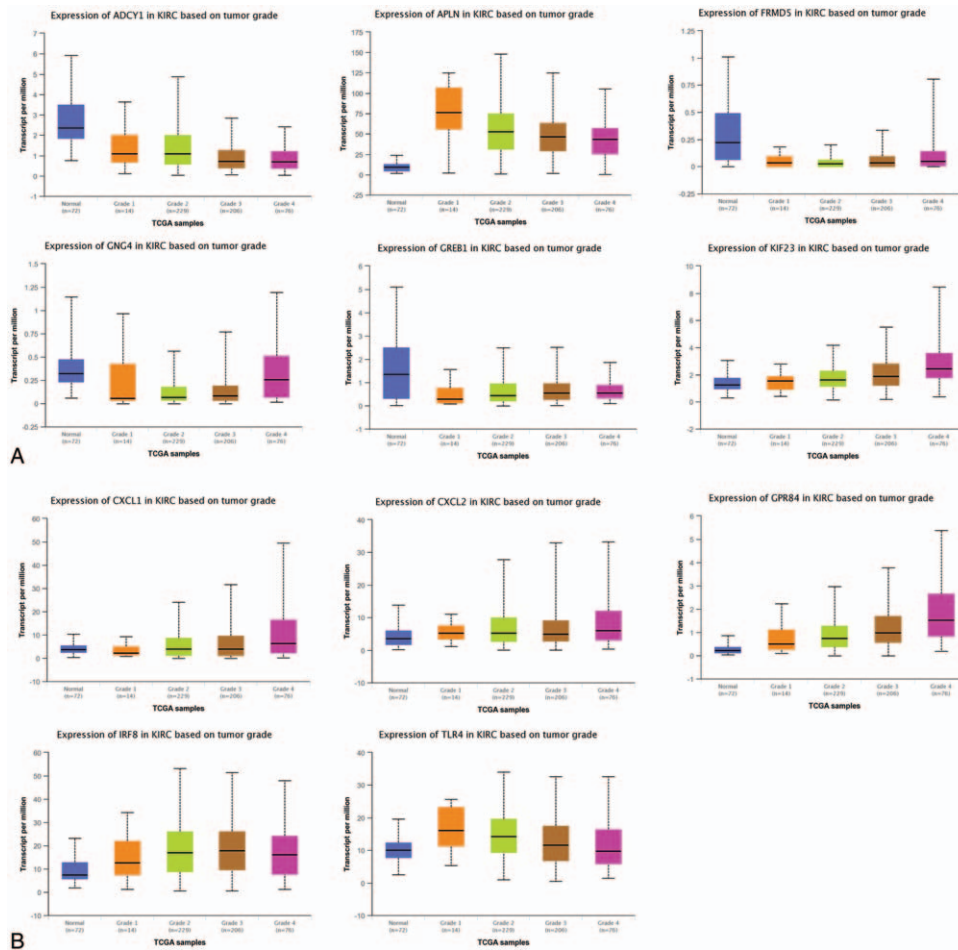


Figure 5. Prognostic value of top up-regulated 15 DEGs (a) and top down-regulated 15 DEGs (b) in the patient of clear cell renal cell carcinoma. HR = hazard ratio.

estrogen receptor-regulated pathway, whose expression was correlated with estrogen levels in breast cancer patients.<sup>[27]</sup> Toll like receptor 4 (TLR4), was a member of the Toll-like receptor (TLR) family which played a fundamental role in pathogen recognition and activation of innate immunity. It was thought to

play an important role in hormone-responsive tissues and cancer. Kusuvara Y et al,<sup>[28]</sup> concluded that low TLR4 expression was correlated with tumor progression, and the expression of TLR4 was inversely associated with prognosis of patients with invasive bladder cancer (BCa), and depletion of TLR4 significantly



**Figure 6.** Boxplots of up-regulated six DEGs (KIF23, GNG4, APLN, ADCY1, GREB1, and FRMD5) (a) and down-regulated five DEGs (TLR4, IRF8, CXCL1, CXCL2, and GPR84) (b) across normal and different pathological stages of clear cell renal cell carcinoma in the TCGA data set.

enhanced the invasive capability of BCa cells. Interferon regulatory factor 8 (IRF8), acted as a tumor suppressor gene through the transcriptional repression of  $\beta$ -catenin-TCF/LEF in NSCLC.<sup>[29]</sup> IRF8 methylation may serve as a potential biomarker in NSCLC prognosis. C-X-C motif chemokine ligand 1, C-X-C motif chemokine ligand 2 (CXCL1, CXCL2), were parts of chemokine superfamily that encoded secreted proteins involved in immunoregulatory and inflammatory processes.<sup>[30]</sup> Aberrant expression of these protein were associated with the growth and progression of certain tumors,<sup>[31]</sup> and may suppressed hematopoietic progenitor cell proliferation.

One thing should be noted was that 11 hub genes had significantly correlated with overall survival in ccRCC patients, as we showed in Figure 5a and b, based on Kaplan–Meier Plotter. However, after analyzed the TCGA database, only KIF23 and APLN were highly expressed in ccRCC tissue compared normal renal tissue, while the expression of other nine genes have opposite results. As shown in Figure 6a, KIF23 had high expression in tumor tissues compared in normal tissues. Moreover, with the increase of tumor grade, the expression of KIF23 was also enhanced. Interestingly, APLN also had high expression in tumor tissues. While the expression of APLN was decreased with the increase of tumor grade. The reason was mainly due to the choice of different databases. Thus in-depth

experimental research studies should be performed to confirm this finding.

## 5. Conclusions

We presumed 167 co-expression DEGs by a series of bioinformatics analyses between ccRCC samples and normal tissues, probably related to the development of ccRCC. Besides, these hub genes might be prognostic and recurrence genes for ccRCC patients as validated from two independent datasets of GSE53757 and GSE16449. These identified genes and pathways provide a more detailed molecular mechanism for underlying ccRCC initiation and development. According to the study, upregulation of KIF23, GNG4, APLN, ADCY1, GREB1, FRMD5, and downregulation of TLR4, IRF8, CXCL1, CXCL2, GPR84 maybe considered as biomarkers or therapeutic targets for ccRCC. However, further molecular and biological experiments are required to validate our findings, which will be our subsequent research work.

## Author contributions

**Data curation:** Ning Zhang, Lu Zhi Gan, Alimujiang Abudurxiti, Gang Xiao Hu, Wei Sang.



**Formal analysis:** Ning Zhang, Xin Wen Chen, Lu Zhi Gan, Alimujiang Abudurexiti, Gang Xiao Hu, Wei Sang.  
**Funding acquisition:** Wei Sang.  
**Investigation:** Ning Zhang.  
**Methodology:** Ning Zhang.  
**Resources:** Ning Zhang, Wei Sang.  
**Software:** Ning Zhang.  
**Supervision:** Xin Wen Chen.  
**Writing – original draft:** Ning Zhang.  
**Writing – review & editing:** Ning Zhang.

## References

- [1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. *CA Cancer J Clin* 2017;67:7–30.
- [2] Cairns P. Renal cell carcinoma. *Cancer Biomark* 2010;9:461–73.
- [3] Hsieh JJ, Purdue MP, Signoretti S, et al. Renal cell carcinoma. *Nat Rev Dis Primers* 2017;3:17009.
- [4] Motzer RJ, Jonasch E, Agarwal N, et al. Kidney cancer, version 3.2015. *J Natl Compr Canc Netw* 2015;13:151–9.
- [5] Motzer RJ, Bander NH, Nanus DM. Renal-cell carcinoma. *N Engl J Med* 1996;335:865–75.
- [6] Vera-Badillo FE, Templeton AJ, Duran I, et al. Systemic therapy for non-clear cell renal cell carcinomas: a systematic review and meta-analysis. *Eur Urol* 2015;67:740–9.
- [7] Coppin C, Kollmannsberger C, Le L, et al. Targeted therapy for advanced renal cell cancer (RCC): a Cochrane systematic review of published randomised trials. *BJU Int* 2011;108:1556–63.
- [8] Goyal R, Gersbach E, Yang XJ, et al. Differential diagnosis of renal tumors with clear cytoplasm: clinical relevance of renal tumor subclassification in the era of targeted therapies and personalized medicine. *Arch Pathol Lab Med* 2013;137:467–80.
- [9] Sosman JA, Puzanov I, Atkins MB. Opportunities and obstacles to combination targeted therapy in renal cell cancer. *Clin Cancer Res* 2007;13:764s–9s.
- [10] Wang JH, Min PQ, Wang PJ, et al. Dynamic CT evaluation of tumor vascularity in renal cell carcinoma. *AJR Am J Roentgenol* 2006;186:1423–30.
- [11] Motzer RJ, McCann L, Deen K. Pazopanib versus sunitinib in renal cancer. *N Engl J Med* 2013;369:1970.
- [12] Gowrishankar B, Ibragimova I, Zhou Y, et al. MicroRNA expression signatures of stage, grade, and progression in clear cell RCC. *Cancer Biol Ther* 2014;15:329–41.
- [13] Luo T, Chen X, Zeng S, et al. Bioinformatic identification of key genes and analysis of prognostic values in clear cell renal cell carcinoma. *Oncol Lett* 2018;16:1747–57.
- [14] Xu T, Ruan H, Song Z, et al. Identification of CXCL13 as a potential biomarker in clear cell renal cell carcinoma via comprehensive bioinformatics analysis. *Biomed Pharmacother* 2019;118:109264.
- [15] Dahinden C, Ingold B, Wild P, et al. Mining tissue microarray data to uncover combinations of biomarker expression patterns that improve intermediate staging and grading of clear cell renal cell cancer. *Clin Cancer Res* 2010;16:88–98.
- [16] Gerlinger M, Horswell S, Larkin J, et al. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat Genet* 2014;46:225–33.
- [17] Eckl J, Buchner A, Prinz PU, et al. Transcript signature predicts tissue NK cell content and defines renal cell carcinoma subgroups independent of TNM staging. *J Mol Med (Berl)* 2012;90:55–66.
- [18] Tavazoie S, Hughes JD, Campbell MJ, et al. Systematic determination of genetic network architecture. *Nat Genet* 1999;22:281–5.
- [19] Xie L, Wang Q, Dang Y, et al. OSkirc: a web tool for identifying prognostic biomarkers in kidney renal clear cell carcinoma. *Future Oncol* 2019;15:3103–10.
- [20] Barata PC, Rini BI. Treatment of renal cell carcinoma: current status and future directions. *CA Cancer J Clin* 2017;67:507–24.
- [21] Song X, Zhang T, Wang X, et al. Distinct diagnostic and prognostic values of kinesin family member genes expression in patients with breast cancer. *Med Sci Monit* 2018;24:9442–64.
- [22] Zhou Z, Cheng Y, Jiang Y, et al. Ten hub genes associated with progression and prognosis of pancreatic carcinoma identified by co-expression analysis. *Int J Biol Sci* 2018;14:124–36.
- [23] Ye L, Li H, Zhang F, et al. Expression of KIF23 and its prognostic role in non-small cell lung cancer: analysis based on the data-mining of oncomine. *Zhongguo Fei Ai Za Zhi* 2017;20:822–6.
- [24] Yang L, Li YL, Li XQ, et al. High apelin level indicates a poor prognostic factor in muscle-invasive bladder cancer. *Dis Markers* 2019;2019:4586405.
- [25] Hu D, Zhu WF, Shen WC, et al. Expression of apelin and snail protein in breast cancer and their prognostic significance. *Zhonghua Bing Li Xue Za Zhi* 2018;47:743–6.
- [26] Hua Y, Ma X, Liu X, et al. Identification of the potential biomarkers for the metastasis of rectal adenocarcinoma. *APMIS* 2017;125:93–100.
- [27] Cheng M, Michalski S, Kommagani R. Role for growth regulation by estrogen in breast cancer 1 (GREB1) in hormone-dependent cancers. *Int J Mol Sci* 2018;19:
- [28] Kusuhara Y, Daizumoto K, Kawai K, et al. Low expression of toll-like receptor 4 is associated with poor prognosis in bladder cancer. *Anticancer Res* 2019;39:703–11.
- [29] Ye L, Xiang T, Zhu J, et al. Interferon consensus sequence-binding protein 8, a tumor suppressor, suppresses tumor growth and invasion of non-small cell lung cancer by interacting with the wnt/beta-catenin pathway. *Cell Physiol Biochem* 2018;51:961–78.
- [30] Hsu YL, Chen YJ, Chang WA, et al. Interaction between tumor-associated dendritic cells and colon cancer cells contributes to tumor progression via CXCL1. *Int J Mol Sci* 2018;19:
- [31] Subat S, Mogushi K, Yasen M, et al. Identification of genes and pathways, including the CXCL2 axis, altered by DNA methylation in hepatocellular carcinoma. *J Cancer Res Clin Oncol* 2019;145:675–84.