Bioinformatics Analysis of C3 and CXCR4 act as Potential Prognostic Biomarkers in Clear Cell Renal Cell Carcinoma (ccRCC)

Jing Quan

Zhejiang Provincial People's Hospital

YUCHEN BAI

YUCHEN BAI <https://orcid.org/0000-0001-7117-1534>

YunBei Yang

Zhejiang Provincial People's Hospital

ErLei Han

Zhejiang Provincial People's Hospital

Hong Bai

Zhejiang Provincial People's Hospital

Qi Zhang

Zhejiang Provincial People's Hospital

DaHong Zhang (✉ urology@zju.edu.cn)

Zhejiang Provincial People's Hospital, People's Hospital of Hangzhou Medical College <https://orcid.org/0000-0003-3051-884X>

Primary research

Keywords: Gene, biomarker, ccRCC, bioinformatics analysis

Posted Date: April 12th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-319333/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: The molecular pathogenesis of ccRCC was still unknown. Hence, the ccRCC-associated genes needs to explored.

Methods: Three ccRCC expression microarray datasets (GSE14762, GSE66270 and GSE53757) downloaded from gene expression omnibus (GEO) database. The distinguish of expressed genes (DEGs) between ccRCC and normal tissue was discuss and explored. the function of our DEGs was analyzed by Gene Ontology (GO) ,Kyoto Encyclopedia of Genes and Genomes (KEGG) .Then the protein-protein interaction network (PPI) was established in order to screen the hub genes. Then the expressions of hub genes were identified by oncomine database.The prognostic values of hub genes were analyzed by GEPIA database in ccRCC patients.

Result: A total of 137 DREs were analyzed, which including 63 upregulated genes and 74 downregulated genes. According to our result,137 DREs were mainly enriched in 82 functional terms and 24 pathways. 14 highest-scoring genes were screened as hub gene in the PPI network which including 12 upregulated candidate genes and 2 downregulated candidate genes. The result reveals that patients with higher C3 expression related to poor OS, while patients with high expression of CTSS and TLR3 related to better OS. Patients with high C3 and CXCR4 expression had a poor DFS, while ccRCC patients with high expression of TLR3 had better DFS. At last, C3 and CXCR4 were selected to detect the prognosis of patients with ccRCC.

Conclusion: The result identified the C3 and CXCR4 as candidate biomarkers and potential therapeutic targets in the molecular mechanism and individual treatment of ccRCC.

1. Introduction

Renal cell carcinoma (RCC) is the most common kidney malignancies, which originates in the renal tubular epithelium[1]. Among of them, ccRCC was the most important histological type, accounting for ~80% of kidney cancers [2]. The vast majority of kidney cancers are discovered by accident. Less than 5% of renal cancers are detected by the classic triad (gross hematuria, flank pain and abdominal mass) and are often advanced. Due to resistant to radiotherapy and chemotherapy, surgical resection is still the best treatment for renal cancer [2]. Although the emergence of immunotherapy and targeted therapy has diversified the treatment of renal cancer, the prognosis of patients with renal cancer who have lost the opportunity of surgery is extremely poor [3]. Therefore, it is particularly important to understand the pathogenesis of renal cancer and to explore biomarkers which could support early diagnosis and prediction of prognosis.

In recent years, bioinformatics analysis of gene expression microarrays as an new method for identifying the potential target genes of diseases and providing the molecular characteristics, regulatory pathways and cellular networks of diseases [4]. The gene expression omnibus (GEO, www.ncbi.nlm.nih.gov/geo/) database is an international public functional genomics database that stores common array and sequence data. In the past decades, more and more studies had reported that tumor-related genes have been discovered by mining GEO databases. For instance, Guo et al found that 31 mostly changed hub genes were significant enriched in several pathways by integrated bioinformatical analysis, which mainly associated with cell cycle process, chemokines and G-protein coupled receptor signaling pathways in colorectal cancer [5]. Besides, Liang's results indicated that BCL2, CCND1 and COL1A1 may be key genes in thyroid papillary carcinoma through bioinformatics analysis [6].

Based on the above theories, we excavated and analysis the ccRCC gene expression profile from the GEO database, and then further analyzed the data to find valuable hub genes for the next research.

2. Materials And Methods

2.1 Access to public resources

Three expression profiling datasets (GSE14762[7], GSE66270[8] and GSE53757[9]) associated ccRCC are acquired from the GEO database. The GSE14762 dataset included 11 tumor tissue samples and matched normal tissue samples. GSE66270 dataset included 14 normal tissue samples and 14 tumor tissue samples. GSE53757 dataset included 72 tumor tissue samples and adjacent tissue samples. There is microarray data of GSE14762 was underlying GPL4866 Platforms. The microarray data of GSE66270 and GSE53757 were underlying GPL570 Platforms. Platform and series matrix files were downloaded as TXT files. Details for GEO ccRCC data were shown in Table 1.

Table 1
Details for GEO ccRCC data

Reference	GEO	Platform	Sample	
			normal	tumor
Furge K[1]	GSE14762	GPL4866	11	11
Jung K[2]	GSE66270	GPL570	14	14
Von Roemeling CA[3]	GSE53757	GPL570	72	72
1. Renal Cell Carcinoma: Hypoxia and Endocytosis				
2. Genome-wide analysis of gene expression patterns in human kidney cancer [patients without metastasis]				
3. Gene array analysis of clear cell renal cell carcinoma tissue versus matched normal kidney tissue				

2.2 Screening of DEGs

The R language software (version 3.5.0; <https://www.r-project.org/>) and annotation package used to handle downloaded data files. Probe name in the downloaded data files was changed to international standard name which call (gene symbol). The package in the Bioconductor package (<http://www.bioconductor.org/>) used in gene distinguish expression analysis. Robust Multi-array Average (RMA) algorithm used for the gene expression profile data preprocessing. And quantile normalization was performed to normalize the above data. $P < 0.05$ and $|\log_2 FC| \geq 2$ were regard as the threshold for DEGs screening. The Robust Rank Aggreg (RRA) analysis (<http://cran.r-project.org/>) was performed to list the up-regulated and down-regulated genes. DEGs of three datasets were represented by volcano map and hierarchical clustering heat map.

2.3 Gene ontology (GO) and KEGG Enrichment Analyses

To explore the biological processes (BP), molecular functions (MF), and cellular components (CC) of DEGs, two online biological tools were applied. Online website g:Profiler (<https://biit.cs.ut.ee/gprofiler/gost>) was used for Go analysis. And DAVID 6.8 (<https://david.ncifcrf.gov/>) was used for KEGG analysis. $P < 0.05$ was considered as the threshold for Go and KEGG pathway analysis.

2.4 PPI network construction

Online database STRING (<http://string-db.org>) and Cytoscape software (Version 3.6.1, <http://www.cytoscape.org/>) were applied to generate the PPI network of DREs and find hub genes. In addition, the Molecular Complex Detection (MCODE) plug-in in cytoscape software was used to analyze clustered sub-networks of highly intraconnected nodes from the above PPI network. The default parameters of MCODE plug-in were as follows: Degree cutoff ≥ 2 , Node score cutoff ≥ 0.2 , K-core ≥ 2 , and Max depth = 100.

2.5 Expression and survival analysis of hub genes

To better validate the expression level of hub genes, the meta-analysis function on the oncomine database (<https://www.oncomine.org/>) was used. Besides, Online database GEPIA (<http://gepia.cancer-pku.cn/detail.php>) is interactive web server for analyzing the expression of tumor and normal gene. In this study, it was applied to analyze the relationship between the expression and survival analysis (include overall survival (OS) and disease free survival (DFS)) of hub genes.

3. Results

3.1 Microarray data information

The RCC expression microarray datasets GSE14762, GSE66270 and GSE53757 were standardized by RMA algorithm, and the results were displayed in Fig. 1. By screening criteria ($P < 0.05$ and $|\log_2 FC| \geq 2$), 381 DREs were obtained from GSE14762, 870 DREs were obtained from GSE66270, and 1324 DREs were obtained from GSE53757. The DREs from the two groups of sample data included in each of the three databases were shown by volcano plot (Fig. 2). The cluster heatmaps of the top 100 DEGs from the three microarrays were displayed in Fig. 3.

3.2 Identification of DEGs in ccRCC

The three microarray databases of kidney cancer were analyzed and sorted by the limma package (threshold: $P < 0.05$ and $|\log_2 FC| \geq 2$), and then further analyzed by RRA method. As a result, 137 DREs were identified, contained 63 overexpressed genes and 74 under-expressed genes (Table 2). Figure 4 revealed the heatmap of the top 20 overexpressed and under-expressed genes by R-heatmap software.

Table 2
The genes differentially expressed both in GEO database were ider

Gene names	
Upregulated DEGs	EGLN3,CA9,ANGPTL4,IGFBP3,ENO2,NDUFA4L2,SPAG4,HK2,CXCR4,APOC1,NOL3,LAPTM5,LPCAT1,PSMB9,CTSS,TYROBP,NETO2,RSKC16A3,C3,COL23A1,C1QA,CSTA,CAV1,ITGB2,SEMA5B,PLOD2,C1QB,TRIB3,MS4A6A,PDK1,BIRC3,DDDB2,ENTPD1,TREM2,EVI2A,P2ZNF395,ST8SIA4,CTHRC1,PRKCSBP,ENPP3,ISG20,MNDA,SLC16A3,ZNF395,FCER1G,PLK2,TNFSF13B,FCGR3A /// FCGR3B,RGS1,TL
Downregulated DEGs	KCNJ1,KNG1,CLCNKB,FGF9,DMRT2,CALB1,RHCG,CLDN8,ATP6V0A4,SFRP1,ATP6V1G3,NPHS2,HS6ST2,ABAT,ATP6V1B1,AQP2,ALD999E24.3,ALDH4A1 SERPINA5,XPNPEP2,DCXR,TMEM52B,ACOX2,TMEM213,LPPR1,HEPACAM2,GPR110,TFCP2L1,FXYP4,HRG,GGT6,ERP27,SLC12A3,T TCF21,EFHD1,FBP1,HPD,TMEM30B,SLC13A3,SLC22A7,AFM,ACSF2,PCK2,PLG,FABP1,LOC155006,SUCNR1,LINC01187,CRYAA,CHL1

3.3 Go and KEGG analysis of DREs

DAVID database was performed to further understand the function of hub genes, including BP, CC and MF. Significant results of the GO enrichment analysis of DEGs in ccRCC are shown in Tables 3. As shown in Fig. 5A and 5B, Go analysis (threshold: P < 0.05 and count ≥ 2) demonstrated that hub genes of ccRCC were mainly enriched in 50 terms in BP group, such as response to hypoxia, oxidation-reduction process and proteolysis. In CC group, DREs were enriched in 21 terms, such as extracellular exosome, plasma membrane and integral component of membrane. Similarly in MF group, DREs were enriched in 11 terms, such as identical protein binding, receptor binding and heparin binding. Shown in Fig. 5C, the result illustrated the relationship between the different functions by cytoscape software.

Table 3
Significant results of the GO enrichment analysis of DEGs

function	Term	Count	PValue	Genes
biological processes	GO:0007588 ~ excretion	7	2.64E-07	NPHS2, CLCNKB, UMOD, ATP6V1B1, ATP6V0A4, KCNJ1, AQP2
	GO:0001666 ~ response to hypoxia	11	5.42E-07	CAV1, NOL3, CA9, PLOD2, CXCR4, EGLN3, TLR2, HSD11B2, ABAT, CASP1, ANGPTL4
	GO:0090383 ~ phagosome acidification	4	9.87E-04	ATP6V1G3, ATP6V1B1, ATP6V0A4, ATP6V0D2
	GO:0033572 ~ transferrin transport	4	0.002115	ATP6V1G3, ATP6V1B1, ATP6V0A4, ATP6V0D2
	GO:0008286 ~ insulin receptor signaling pathway	5	0.002575	CAV2, ATP6V1G3, ATP6V1B1, ATP6V0A4, ATP6V0D2
	GO:0032755 ~ positive regulation of interleukin-6 production	4	0.004349	P2RX7, TLR2, FCER1G, TLR3
	GO:0034220 ~ ion transmembrane transport	7	0.004506	FXYD4, CLCNKB, ATP6V1G3, ATP6V1B1, ATP6V0A4, ATP6V0D2, AQP2
	GO:0015695 ~ organic cation transport	3	0.00525	RHCG, SLC7A8, RHBG
	GO:0006885 ~ regulation of pH	3	0.005972	RHCG, ATP6V1B1, ATP6V0A4
	GO:0090090 ~ negative regulation of canonical Wnt signaling pathway	6	0.006989	CTHRC1, CAV1, SOST, GPC3, SFRP1, PSMB9
	GO:0042493 ~ response to drug	8	0.007086	P2RX7, CA9, SFRP1, LGALS1, SLC34A1, HSD11B2, ABAT, NNMT
	GO:0008152 ~ metabolic process	6	0.007913	ENPP6, LPCAT1, SUCLG1, MAN1C1, ACSF2, ACAA1
	GO:0032092 ~ positive regulation of protein binding	4	0.010136	CTHRC1, CAV1, PLK2, TRIB3
	GO:0055074 ~ calcium ion homeostasis	3	0.011171	CAV1, ATP6V1B1, CALB1
	GO:0006508 ~ proteolysis	10	0.011293	C1QA, C1QB, GGT6, SFRP1, C3, CTSS, CASP1, PLG, DPEP1, XPNPEP2
	GO:0055114 ~ oxidation-reduction process	11	0.011711	ALDH6A1, TYRP1, PLOD2, NDUFA4L2, RRM2, EGLN3, HSD11B2, ALDH4A1, DIO1, DCXR, HPD
	GO:0010951 ~ negative regulation of endopeptidase activity	5	0.012067	KNG1, C3, SERPINA5, HRG, CSTA
	GO:0050900 ~ leukocyte migration	5	0.012407	SLC16A3, CAV1, SLC7A8, FCER1G, ITGB2
	GO:0045880 ~ positive regulation of smoothened signaling pathway	3	0.013221	GPC3, SFRP1, FGF9
	GO:0001798 ~ positive regulation of type IIa hypersensitivity	2	0.014597	C3, FCER1G
	GO:2000054 ~ negative regulation of Wnt signaling pathway involved in dorsal/ventral axis specification	2	0.014597	SOST, SFRP1

function	Term	Count	PValue	Genes
	GO:0061621 ~ canonical glycolysis	3	0.01542	ENO2, PFKF, HK2
	GO:0001503 ~ ossification	4	0.020939	SOST, SLC34A1, ATP6V1B1, ATP6V0A4
	GO:0006954 ~ inflammatory response	8	0.021562	KNG1, TNFAIP6, P2RX7, CXCR4, C3, TLR2, TLR3, ITGB2
	GO:0050717 ~ positive regulation of interleukin-1 alpha secretion	2	0.021815	P2RX7, CASP1
	GO:2000116 ~ regulation of cysteine-type endopeptidase activity	2	0.021815	BIRC3, PSMB9
	GO:0070634 ~ transepithelial ammonium transport	2	0.021815	RHCG, RHBG
	GO:0019065 ~ receptor-mediated endocytosis of virus by host cell	2	0.021815	CAV2, CAV1
	GO:0015991 ~ ATP hydrolysis coupled proton transport	3	0.022874	ATP6V1B1, ATP6V0A4, ATP6V0D2
	GO:0051480 ~ regulation of cytosolic calcium ion concentration	3	0.022874	CAV1, PVALB, CALB1
	GO:0001525 ~ angiogenesis	6	0.024169	CAV1, FGF9, TGFBI, HRG, FGF1, ANGPTL4
	GO:0002931 ~ response to ischemia	3	0.024234	CAV1, NOL3, HK2
	GO:0072221 ~ metanephric distal convoluted tubule development	2	0.028982	UMOD, CALB1
	GO:0002283 ~ neutrophil activation involved in immune response	2	0.028982	FCER1G, TYROBP
	GO:0007162 ~ negative regulation of cell adhesion	3	0.029996	KNG1, TGFBI, HRG
	GO:0006955 ~ immune response	8	0.03533	RGS1, TNFSF13B, C3, ENPP3, TLR2, CTSS, FCGR3A, FCGR3B
	GO:0070836 ~ caveola assembly	2	0.036096	CAV2, CAV1
	GO:0015696 ~ ammonium transport	2	0.036096	RHCG, RHBG
	GO:0006873 ~ cellular ion homeostasis	2	0.036096	RHCG, SLC4A1
	GO:0019740 ~ nitrogen utilization	2	0.036096	RHCG, RHBG
	GO:0010543 ~ regulation of platelet activation	2	0.036096	FCER1G, HRG
	GO:0051005 ~ negative regulation of lipoprotein lipase activity	2	0.036096	APOC1, ANGPTL4

function	Term	Count	PValue	Genes
	GO:0034123 ~ positive regulation of toll-like receptor signaling pathway	2	0.036096	TLR2, TLR3
	GO:0031623 ~ receptor internalization	3	0.039536	CAV1, FCER1G, ITGB2
	GO:0006094 ~ gluconeogenesis	3	0.041225	ENO2, FBP1, PCK2
	GO:0050776 ~ regulation of immune response	5	0.041952	C3, ITGB2, FCGR3A, TREM2, TYROBP
	GO:0030514 ~ negative regulation of BMP signaling pathway	3	0.04294	CAV1, SOST, SFRP1
	GO:0000187 ~ activation of MAPK activity	4	0.043944	P2RX7, CXCR4, FGF1, DUSP9
	GO:0007596 ~ blood coagulation	5	0.046407	P2RX7, SERPINA5, FCER1G, ENTPD1, PLG
	GO:0032760 ~ positive regulation of tumor necrosis factor production	3	0.046448	TLR2, FCER1G, TLR3
cell composition	GO:0070062 ~ extracellular exosome	61	4.70E-17	FGF9, SLC7A8, CALB1, AQP2, EFHD1, GPC3, PVALB, CXCR4, PLOD2, SERPINA5, TMEM52B, TGFB1, SLC4A1, FCGR3A, ATP6V0D2, FCGR3B, DPEP1, HPD, KNG1, ALDH6A1, CRYAA, SUCLG1, SLC22A8, PFKP, FBP1, C1QA, C1QB, RHCG, MNDA, ABAT, CSTA, CHL1, SH3GL2, ENPP6, C3, ENPP3, APOC1, ITGB2, UMOD, ATP6V1B1, ENO2, HS6ST2, HRG, SUCNR1, ENTPD1, SLC12A3, LGALS1, PCK2, MAN1C1, PLG, PSMB9, XPNPEP2, AFM, GGT6, SFRP1, NPHS2, SLC13A3, FABP1, ATP6V0A4, IGFBP3, DCXR
	GO:0016323 ~ basolateral plasma membrane	12	6.23E-08	CLDN8, CAV1, RHCG, CA9, SLC22A7, SLC22A8, SLC7A8, RHBG, UMOD, SLC4A1, ATP6V1B1, AQP2
	GO:0005886 ~ plasma membrane	53	4.79E-06	CLDN8, TLR2, SLC7A8, AQP2, GPC3, CXCR4, TGFB1, SLC4A1, FCGR3A, FCGR3B, DPEP1, KNG1, COL23A1, SLC22A7, SUCLG1, SLC22A8, SLC34A1, TNFSF13B, RHCG, CA9, ATP6V1G3, TREM2, SH3GL2, CHL1, ENPP6, CAV2, CAV1, FXYD4, C3, RHBG, TRIB3, CLCNKB, ITGB2, KCNJ1, ENO2, TMEM30B, FCER1G, HRG, SUCNR1, ENTPD1, TYROBP, SLC12A3, PLG, XPNPEP2, SLC16A3, P2RX7, RGS1, SFRP1, NPHS2, SLC13A3, DIO1, ATP6V0A4, DCXR
	GO:0005887 ~ integral component of plasma membrane	25	5.00E-05	CAV2, CAV1, FXYD4, SLC12A3, SLC22A7, ENPP3, SLC22A8, TLR2, RHBG, SLC34A1, SLC7A8, TLR3, CLCNKB, AQP2, SLC16A3, P2RX7, LAPTM5, GPC3, RHCG, NPHS2, FCER1G, SLC13A3, SLC4A1, ENTPD1, TYROBP
	GO:0016471 ~ vacuolar proton-transporting V-type ATPase complex	4	7.28E-05	ATP6V1G3, ATP6V1B1, ATP6V0A4, ATP6V0D2
	GO:0072562 ~ blood microparticle	8	1.04E-04	KNG1, C1QB, AFM, C3, HRG, SLC4A1, PLG, ANGPTL4
	GO:0005578 ~ proteinaceous extracellular matrix	10	1.24E-04	CTHRC1, SOST, GPC3, SFRP1, LGALS1, TGFB1, UMOD, FGF1, CHL1, ANGPTL4
	GO:0016324 ~ apical plasma membrane	10	2.30E-04	CAV1, RHCG, SLC12A3, SLC34A1, UMOD, ATP6V1B1, ATP6V0A4, ATP6V0D2, DPEP1, AQP2
	GO:0005615 ~ extracellular space	22	4.94E-04	KNG1, CTHRC1, C3, FGF9, LGALS1, HILPDA, CTSS, PLG, TNFAIP6, AFM, GPC3, SOST, TNFSF13B, SFRP1, SERPINA5, TGFB1, ENO2, CSTA, FGF1, IGFBP3, DPEP1, ANGPTL4
	GO:0031225 ~ anchored component of membrane	6	0.001254	ENPP6, GPC3, UMOD, FCGR3B, DPEP1, XPNPEP2
	GO:0016021 ~ integral component of membrane	53	0.002201	CLDN8, TLR2, SLC7A8, TLR3, AQP2, CXCR4, EVI2A, TMEM52B, SMIM5, SLC4A1, FCGR3A, FCGR3B, COL23A1, SLC22A7, SLC22A8, SLC34A1, DMRT2, HEPACAM2, TNFSF13B, CA9, RHCG, SPAG4, HSD11B2, TREM2, CHL1, NETO2, CAV2, TYRP1, CAV1, TMEM213, ENPP3, RHBG, CLCNKB, UMOD, KCNJ1, SEMA5B, LPCAT1, TMEM30B, FCER1G, HS6ST2, SUCNR1, ENTPD1, MS4A6A, TYROBP, TMEM45A, SLC12A3, NDUFA4L2, HILPDA, GGT6, SFRP1, SLC13A3, DIO1, ATP6V0A4

function	Term	Count	PValue	Genes
	GO:0043234 ~ protein complex	10	0.002667	CAV2, CAV1, SOST, PVALB, SERPINA5, NPHS2, DDB2, FABP1, PRKCDBP, BIRC3
	GO:0005576 ~ extracellular region	22	0.004463	KNG1, ENPP6, FGF9, C3, APOC1, UMOD, CTSS, PLG, C1QA, C1QB, AFM, SOST, TNFSF13B, SFRP1, SERPINA5, TGFB1, HRG, FGF1, TREM2, CASP1, IGFBP3, ANGPTL4
	GO:0009986 ~ cell surface	11	0.005205	SFRP1, CXCR4, LGALS1, TLR2, SLC34A1, FCER1G, TLR3, HILPDA, ITGB2, PLG, TYROBP
	GO:0002080 ~ acrosomal membrane	3	0.007858	CAV2, CAV1, SERPINA5
	GO:0005602 ~ complement component C1 complex	2	0.014107	C1QA, C1QB
	GO:0045121 ~ membrane raft	6	0.015662	CAV2, CAV1, NPHS2, TLR2, SLC34A1, BIRC3
	GO:0000139 ~ Golgi membrane	10	0.024687	CAV2, CAV1, LPCAT1, ST8SIA4, TLR3, HS6ST2, HEPACAM2, MAN1C1, SH3GL2, HPD
	GO:0005581 ~ collagen trimer	4	0.027574	C1QA, CTHRC1, C1QB, COL23A1
	GO:0005759 ~ mitochondrial matrix	7	0.028942	PDK1, ALDH6A1, SUCLG1, ALDH4A1, ABAT, PCK2, ACSF2
	GO:0005782 ~ peroxisomal matrix	3	0.042028	ACOX2, FABP1, ACAA1
molecular function	GO:0019864 ~ IgG binding	4	6.11E-05	FCER1G, UMOD, FCGR3A, FCGR3B
	GO:0008201 ~ heparin binding	7	0.001175	KNG1, SOST, SFRP1, FGF9, SERPINA5, HRG, FGF1
	GO:0005102 ~ receptor binding	10	0.001189	KNG1, ACOX2, P2RX7, CAV1, TNFSF13B, C3, HRG, HILPDA, PLG, TYROBP
	GO:0030506 ~ ankyrin binding	3	0.009327	RHCG, RHBG, SLC4A1
	GO:0001530 ~ lipopolysaccharide binding	3	0.011232	P2RX7, TLR2, TREM2
	GO:0043027 ~ cysteine-type endopeptidase inhibitor activity involved in apoptotic process	3	0.012243	NOL3, BIRC3, DPEP1
	GO:0015301 ~ anion:anion antiporter activity	3	0.012243	SLC22A7, SLC22A8, SLC4A1
	GO:0051117 ~ ATPase binding	4	0.017177	CAV1, FXD4, ATP6V1G3, ATP6V0A4
	GO:0042802 ~ identical protein binding	12	0.022021	CLDN8, CAV1, NOL3, SFRP1, CRYAA, FBP1, ALDH4A1, TLR3, SH3GL2, DCXR, TYROBP, ANGPTL4
	GO:0015078 ~ hydrogen ion transmembrane transporter activity	3	0.022995	ATP6V1B1, ATP6V0A4, ATP6V0D2
	GO:0004869 ~ cysteine-type endopeptidase inhibitor activity	3	0.025762	KNG1, HRG, CSTA

To further analyze the above DREs, the significantly enriched pathways were submitted to KEGG analysis. Table 4 and Fig. 5D indicated the significant pathway enrichment of DEGs by KEGG analysis. These DEGs were enriched in 24 pathways, which mainly related to metabolic pathways, phagosome and other pathways are involved.

Table 4
KEGG pathway analysis of DEGs associated with ccRCC

Term	Count	P Value	Genes
hsa04966:Collecting duct acid secretion	6	1.77E-05	CLCNKB, SLC4A1, ATP6V1G3, ATP6V1B1, ATP6V0A4, ATP6V0D2
hsa05150:Staphylococcus aureus infection	7	5.12E-05	C1QA, C1QB, C3, ITGB2, FCGR3A, FCGR3B, PLG
hsa04145:Phagosome	10	9.64E-05	C3, TLR2, ITGB2, CTSS, ATP6V1G3, FCGR3A, ATP6V1B1, ATP6V0A4, FCGR3B, ATP6V0D2
hsa05323:Rheumatoid arthritis	7	7.65E-04	TNFSF13B, TLR2, ITGB2, ATP6V1G3, ATP6V1B1, ATP6V0A4, ATP6V0D2
hsa05152:Tuberculosis	9	0.001573	C3, TLR2, FCER1G, ITGB2, CTSS, FCGR3A, ATP6V0A4, FCGR3B, ATP6V0D2
hsa04610:Complement and coagulation cascades	6	0.001629	KNG1, C1QA, C1QB, C3, SERPINA5, PLG
hsa01100:Metabolic pathways	26	0.006292	ACOX2, TYRP1, ENPP3, HK2, UPP2, ATP6V1B1, LPCAT1, ENO2, ALDH4A1, ATP6V0D2, HPD, ALDH6A1, NDUFA4L2, SUCLG1, FBP1, PFKP, PCK2, MAN1C1, GGT6, RRM2, ABAT, ATP6V1G3, ATP6V0A4, DCXR, ACAA1, NNMT
hsa03320:PPAR signaling pathway	5	0.009598	ACOX2, FABP1, PCK2, ACAA1, ANGPTL4
hsa00010:Glycolysis / Gluconeogenesis	5	0.009598	ENO2, FBP1, PFKP, HK2, PCK2
hsa05140:Leishmaniasis	5	0.011719	C3, TLR2, ITGB2, FCGR3A, FCGR3B
hsa01200:Carbon metabolism	6	0.013278	ALDH6A1, SUCLG1, ENO2, FBP1, PFKP, HK2
hsa05133:Pertussis	5	0.01412	C1QA, C1QB, C3, ITGB2, CASP1
hsa05110:Vibrio cholerae infection	4	0.026859	ATP6V1G3, ATP6V1B1, ATP6V0A4, ATP6V0D2
hsa05134:Legionellosis	4	0.029616	C3, TLR2, ITGB2, CASP1
hsa04721:Synaptic vesicle cycle	4	0.043794	ATP6V1G3, ATP6V1B1, ATP6V0A4, ATP6V0D2
hsa05230:Central carbon metabolism in cancer	4	0.045546	SLC16A3, PDK1, PFKP, HK2
hsa00640:Propanoate metabolism	3	0.047335	ALDH6A1, SUCLG1, ABAT
hsa01130:Biosynthesis of antibiotics	7	0.04889	SUCLG1, ENO2, FBP1, PFKP, HK2, PCK2, ACAA1
hsa05120:Epithelial cell signaling in Helicobacter pylori infection	4	0.051005	ATP6V1G3, ATP6V1B1, ATP6V0A4, ATP6V0D2
hsa00051:Fructose and mannose metabolism	3	0.060164	FBP1, PFKP, HK2
hsa04650:Natural killer cell mediated cytotoxicity	5	0.065949	FCER1G, ITGB2, FCGR3A, FCGR3B, TYROBP
hsa00190:Oxidative phosphorylation	5	0.084557	NDUFA4L2, ATP6V1G3, ATP6V1B1, ATP6V0A4, ATP6V0D2
hsa04960:Aldosterone-regulated sodium reabsorption	3	0.085035	FXYP4, HSD11B2, KCNJ1
hsa05322:Systemic lupus erythematosus	5	0.086365	C1QA, C1QB, C3, FCGR3A, FCGR3B

3.4 PPI network and module analysis

String database was used to generate PPI networks of DREs in RCC. Figure 6A showed the relationship between 137 candidate hub genes. Besides, MCODE application was applied to screen out the highest-scoring nodes. And Fig. 6B displayed the module with the highest score (score = 10, node = 11, edges = 50). As a result, MCODE application select the 14 nodes with the highest score, including 12 upregulated candidate genes(C1QA, C1QB, C3, CTSS, CXCR4, FCER1G, ITGB2, TLR2, TLR3 and TYROBP) and 2 downregulated candidate genes(AQP2, PLG).

3.5 Expression and survival analysis of hub genes

To further explore the expression and prognosis of the above screened genes, the oncomine database and GEPIA database were applied. Six analyses were obtained from the oncomine database (Fig. 7). The result of meta-analysis suggested that the expression of 10 genes was significant ($P < 0.05$). Figure 8 indicated the OS and DFS of 10 genes. And the result demonstrated that ccRCC patients with high C3 expression had a poor OS, while ccRCC patients with

high expression of CTSS and TLR3 had a good OS. Besides, in ccRCC patients, high C3 and CXCR4 expression indicated a poor DFS, while high expression of TLR3 indicated a good DFS. Finally, C3 and CXCR4 were selected to detect the prognosis of patients with ccRCC.

4. Discussion

Kidney cancer accounts for about 2–3% of adult malignant tumors, and 80–90% of adult renal malignancies. In the Worldwide, kidney cancer ranks 13th among common tumors. In 2012, about 338 000 cases of kidney cancer were newly discovered, accounting for 24% of all tumors, and the death toll was 144 000 cases, accounting for 17% of all tumors.[11] However, the National Cancer Registry conducted a long-term survey on the incidence and mortality of kidney cancer across the country and found that the incidence of kidney cancer in China was nearly to the world average and is increasing year by year.[12] Renal cell carcinoma (RCC) is the most common kidney malignancies. The early symptoms of RCC are not obviously, and most patients are diagnosed with advanced stage or metastasis. [13] RCC was characteristic of easy recurrence and metastasis due to the complexity of the causes and pathogenesis. Moreover, it was insensitivity to traditional chemoradiotherapy. Those reasons caused RCC usually leads to poor clinical outcomes. Hence, it is helpful to improve the diagnosis, treatment and prognosis of renal cancer by increasing the understanding of the biological molecular mechanism of renal cancer.

With develop of sequencing technology and bioinformatics, the collect and analyzed of previous data will support to explored the pathogenesis of RCC and discover possible diagnosis and treatment biomarkers. [14] Bioinformatics research to analyze biological data, propose on the development of related gene or group of disease, then performed experimental to prove the result, which is a highly efficient research pathway. In the present, Bioinformatics has been widely performed at all areas of medical research, the design of the discover disease-related genes, clinical diagnosis of disease, individualized treatment of diseases and new molecular targets for drug discovery. Bioinformatics was play very important role on those area[15].

In this study, we downloaded three chips from the GEO database. By RRA method, 137 DREs were identified, including 63 overexpressed genes and 74 under-expressed genes. By Go analysis, 137 DREs were mainly enriched in 82 terms, such as extracellular exosome, basolateral plasma membrane, excretion and IgG binding. By KEGG analysis, these DEGs were principally enriched in 24 pathways, including metabolic pathways, phagosome and other pathways are involved. Through PPI network, 14 highest-scoring genes were screened as hub gene, including 12 upregulated candidate genes (C1QA, C1QB, C3, CTSS, CXCR4, FCER1G, ITGB2, TLR2, TLR3 and TYROBP) and 2 downregulated candidate genes (AQP2, PLG). In order to further verify the association with 14 hub genes and ccRCC, the oncomine platform was used. As a result, 10 hub genes were significantly abnormally expressed by meta-analysis, and all of above genes were highly expressed in ccRCC. Based on the above result, we used the GEPIA platform to predict the prognosis of 10 hub genes. Finally, we found that ccRCC patients with high C3 expression had a poor OS, while ccRCC patients with high expression of CTSS and TLR3 had a better OS. Besides, in ccRCC patients, high C3 and CXCR4 expression indicated a poor DFS, while high expression of TLR3 indicated a good DFS.

As a protein coding gene, complement component 3 (C3) is involved in the occurrence and development of many diseases, including C3 deficiency, Autosomal Recessive and Hemolytic Uremic Syndrome, Atypical 5. [16] Among its related pathways are Immune response Lectin induced complement pathway and Signaling by GPCR. And past reports was proved that C3 was a potential prognostic marker for non-small cell lung cancer and may be a new immune marker to judge the prognosis of patients with non-small cell lung cancer. Because of C3 expression was decreased significantly in stage III if you compared with stage I of NSCLC.[17,18] Our result reveals that C3 was increased express in RCC, and the increase the expression was related to worse prognosis. CTSS (Cathepsin S) is a Protein Coding gene, previous articles of papillary thyroid carcinoma reveals that CTSS was highly expressed and related to transformation. [19] Those results reveals CTSS highly expression was associated with poor prognosis and lymph node metastasis. Hence, the CTSS contributes to the genes of oncogenes was justified.[20] However, we analyzed our data that CTSS was indeed highly expressed in RCC, but overexpression was associated with better prognosis. The prognosis of patients with highly expression was even better. which is on opposite effect between expression and prognosis. we consider of that might be degraded after binding to other target genes or inhibited in the subsequent functioning pathways which still needs Explore in the further. As a member of the Toll-like receptor (TLR) family, past studies has proved that TLR3 is abnormally expressed in a variety of tumors, including breast, ovarian and prostate tumors. But TLR3 was associated with either better or poor clinical result in various cancers. [21,22] Hence, we were as well analyzed TLR3 expression in RCC which we realized that it expression as CTSS. TLR3 was highly expression in RCC but it was related to result with better prognosis. chemokine receptor-4 (CXCR4) is a gene which belongs to the super-family of the 7-transmembrane domain, heterotrimeric G-protein-coupled receptors and is associated with cell proliferation, migration, invasion and survival. In the previous reports, It has been demonstrated that CXCR4 is upregulated in sporadic Vestibular schwannomas (VS) as well as in neurofibromatosis type 2 (NF2) tumors. [23–25] According to our result that we explored CXCR4 was overexpressed in RCC as well as the higher expression related with poor prognosis.

5. Conclusion

In summary, through bioinformatics analysis, 2 ccRCC-associated candidate genes (C3 and CXCR4) were identified using three expression profile datasets from the GEO database. Besides, CTSS and TLR3 are abnormally expressed in ccRCC and are associated with ccRCC prognosis. However, their expression is contrary to the prognosis. The above novel biomarkers may have important clinical significance for the diagnosis and prognosis of renal carcinoma, but the detailed mechanism of their action in the development of renal carcinoma needs to be further explored.

Declarations

Data Available

The data used to support the findings of this study are included within the article. The gene expression data can be accessed on Gene Expression Omnibus (GEO). Oncomine database was used to get the expression profile of hub genes. The overall survival and disease free survival analysis of genes were acquired from GEPIA database.

Conflicts of Interest

The authors report no conflicts of interest in this work.

Funding:

This study was supported by the General Project Funds from the Health Department of Zhejiang Province (Grant no 2021439557).

Ethics Statement:

The study was approved by the Ethics Committee of Zhejiang Provincial Peoples' Hospital, Peoples' Hospital of Hangzhou Medical College.

Author Contribution Statement:

Protocol/project development—DaHong Zhang; Data collection or management JingQuan,Yuchen Bai, and ErLei Han; Data analysis—YunBei Yang,Hong Bai and Qi Zhang; Manuscript writing/editing—YuChen Bai and Jing Quan;

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. 2018;68(1):7–30.
2. Escudier B, et al. Renal cell carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Ann Oncol*. 2019;30(5):706–20.
3. Li F, et al. Identification of Key Biomarkers and Potential Molecular Mechanisms in Renal Cell Carcinoma by Bioinformatics Analysis. *J Comput Biol*. 2019;26(11):1278–95.
4. Alizadeh M, Valizadeh R, Rahimi MM. Immediate successful renal autotransplantation after proximal ureteral avulsion following ureteroscopy: a case report. *Journal of Surgical Case Reports*, 2017. 2017(2).
5. Guo Y, et al., Identification of Key Candidate Genes and Pathways in Colorectal Cancer by Integrated Bioinformatical Analysis. *Int J Mol Sci*, 2017. 18(4).
6. Liang W, Sun F. Identification of key genes of papillary thyroid cancer using integrated bioinformatics analysis. *J Endocrinol Invest*. 2018;41(10):1237–45.
7. Wang Y, et al. Regulation of endocytosis via the oxygen-sensing pathway. *Nat Med*. 2009;15(3):319–24.
8. Wotschovsky Z, et al. Integrated microRNA and mRNA Signature Associated with the Transition from the Locally Confined to the Metastasized Clear Cell Renal Cell Carcinoma Exemplified by miR-146-5p. *PLoS One*. 2016;11(2):e0148746.
9. von Roemeling CA, et al. Neuronal pentraxin 2 supports clear cell renal cell carcinoma by activating the AMPA-selective glutamate receptor-4. *Cancer Res*. 2014;74(17):4796–810.
10. Tan J, et al. Integrated bioinformatics analysis reveals that the expression of cathepsin S is associated with lymph node metastasis and poor prognosis in papillary thyroid cancer. *Oncol Rep*. 2018;40(1):111–22.
11. Jacques F, et al. Cancer Incidence and Mortality Worldwide: Sources, Methods and Major Patterns in GLOBOCAN 2012. *Int J Cancer*. 2015;136(5 (March 1):6. <https://doi.org/10.1002/ijc.29210>.
12. Howlader N, Noone AM, Krapcho M, et al., Cancer Statistics Review SEER. 1975–2008, National Cancer Institute,Bethesda, Md, USA, 2011, based on November 2010 SEER data submission, posted to the SEER web site, http://seer.cancer.gov/csr/1975_2008/.
13. Ingrid ME, Desai, et al. Sorafenib Reduces the Percentage of Tumour Infiltrating Regulatory T Cells in Renal Cell Carcinoma Patients. *Int J Cancer*. July 15, 2011;129(2):1. <https://doi.org/10.1002/ijc.25674>.
14. Lv J, Zhu Y, Zhang Q. An increased level of MiR-222-3p is associated with TMP2 suppression, ERK activation and is associated with metastasis and a poor prognosis in renal clear cell carcinoma[J]. *Cancer Biomark A*. 2020;28(1):1–9.
15. Lijun, Cheng B, et al. A bioinformatics approach for precision medicine off-label drug drug selection among triple negative breast cancer patients.[J]. *Journal of the American Medical Informatics Association Jamia*, 2016.
16. Assirelli E, et al. "Complement Expression and Activation in Osteoarthritis Joint Compartments. " *Front Immunol*. 2020;11:535010.
17. Xu J, et al. (2020). "Using Machine Learning Modeling to Explore New Immune-Related Prognostic Markers in Non-Small Cell Lung Cancer." *Frontiers in Oncology* 10.
18. Fayad Christelle,Audi Hassib,Khnayzer. Rony S, et al. The anti-cancer effect of. series of strained photoactivatable Ru(II) polypyridyl complexes on non-small-cell lung cancer and triple negative breast cancer cells.[J].*J Biol Inorg Chem*, 2020, undefined: undefined.
19. Juan T, Xiaoxiao Q, Bin S, et al. Integrated bioinformatics analysis reveals. that the expression of cathepsin S is associated with lymph node metastasis and poor prognosis in papillary thyroid cancer[J]. *Oncol Rep*. 2018;40:111–22.
20. Lyu Weifeng,Jia Bo,Liu. Wei, et al. [Cathepsin S (CTSS) is highly expressed in. temozolomide-resistant glioblastoma T98G cells and associated with poor prognosis].[J].*Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi*, 2020, 36: 924–929.
21. Huang Y,Li Xing,Pan Chunlian et al. The intervention mechanism of emodin on TLR3 pathway in the process of central nervous system injury caused by herpes virus infection.[J].*Neurol Res*, 2020, undefined: 1–7.
22. Muresan XM, et al. (2020). "Toll-Like Receptor 3 in Solid Cancer and Therapy Resistance." *Cancers (Basel)* 12(11).
23. Fan G, Wei J. Identification of potential novel biomarkers and therapeutic targets involved in human atrial fibrillation based on bioinformatics analysis. *Kardiol Pol*. 2020;78(7–8):694–702.
24. Lassaletta L, et al. "Biomarkers in Vestibular Schwannoma-Associated Hearing Loss. " *Front Neurol*. 2019;10:978.

Figures

Figure 1

Standardization of gene expression by boxplot. The GSE14762 data (A), GSE66270 data (B) and GSE55757 data (C) was standardized.

Figure 2

Differential expressions of data between two sets of samples by volcano plot. Three figures show the volcano plot of GSE14762 data (A), GSE66270 data (B) and GSE55757 data (C). The red points represent overexpressed genes (threshold: $P < 0.05$ and $|\log_2 \text{FC}| \geq 2$). The green points represent under-expressed genes (threshold: $P < 0.05$ and $|\log_2 \text{FC}| \geq 2$). The black points represent undifferentiated genes.

Figure 3

Clustering heatmap of DEGs. Three figures show the heatmap of GSE14762 data (A), GSE66270 data (B) and GSE55757 data (C). Red grid shows that the expression of genes is uoverexpressed, green grid shows that the expression of genes is under-expressed, and black grid shows no significant difference; gray grid shows that genes are too weak to be detected..

Figure 4

RRA analyses. The figure shows the top 20 overexpressed and under-expressed genes by RRA analyses. Red grid indicates that the expression of genes is upregulated, blue grid indicates that the expression of genes is downregulated, and white grid indicates no detected gene expression.

Figure 5

Go and KEGG analysis of DREs in ccRCC. (A) DEGs were divided into three functional groups by GO analysis, including molecular function, biological processes and cell composition. (B) GO enrichment significance items of DEGs in three functional groups. (C) The relationship between the different functions. (D) Significant pathway enrichment of DEGs. Purple represents the signaling pathway, red represents overexpressed genes and green represents under-expressed genes.

Figure 6

Protein-protein interaction network and MCODE application. (A) PPI network. (B) Top 13 degree genes by MCODE application.

Comparison of 13 Genes Across 6 Analyses

Median Rank	p-Value	Gene	1	2	3	4	5	6
12607.0	1.000	AQP2						
482.0	3.65E-13	C1QA						
236.0	3.31E-17	C1QB						
529.0	6.69E-6	C3						
1430.0	4.14E-4	CTSS						
435.0	1.65E-4	CXCR4						
436.0	2.97E-6	FCER1G						
326.0	6.43E-7	ITGB2						
12544.0	1.000	KNG1						
16998.0	1.000	PLG						
325.0	6.75E-7	TLR2						
1256.0	0.002	TLR3						
337.0	3.75E-15	TYROBP						

Legend

1. Clear Cell Renal Cell Carcinoma vs. Normal
Gumz Renal, Clin Cancer Res, 2007
2. Clear Cell Renal Cell Carcinoma vs. Normal
Higgins Renal, Am J Pathol, 2003
3. Clear Cell Renal Cell Carcinoma vs. Normal
Jones Renal, Clin Cancer Res, 2005
4. Clear Cell Renal Cell Carcinoma vs. Normal
Lenburg Renal, BMC Cancer, 2003
5. Clear Cell Renal Cell Carcinoma vs. Normal
TCGA Renal 2, No Associated Paper, 2011
6. Clear Cell Renal Cell Carcinoma vs. Normal
Yusenko Renal, BMC Cancer, 2009

The rank for a gene is the median rank for that gene across each of the analyses.
The p-Value for a gene is its p-Value for the median-ranked analysis.

Figure 7

The expression level of 13 hub genes among 6 different analysis datasets by the ONCOMINE database.

Figure 8

The OS and DFS of 10 candidate genes in ccRCC patients by GEPIA database. (OS: A-J, DFS: K-T).