Key genes associated with prognosis and metastasis of clear cell renal cell carcinoma

Tingting Zhong, Zeying Jiang, Xiangdong Wang, Honglei Wang, Meiyi Song, Wenfang Chen and Shicong Yang

Department of Pathology, The First Affiliated Hospital of Sun Yat-Sen University, Guangzhou, China

ABSTRACT

Background. Clear cell renal cell carcinoma (ccRCC) is a tumor that frequently shows the hematogenous pathway and tends to be resistant to radiotherapy and chemotherapy. However, the exact mechanism of ccRCC metastasis remains unknown.

Methods. Differentially expressed genes (DEGs) of three gene expression profiles (GSE85258, GSE105288 and GSE22541) downloaded from the Gene Expression Omnibus (GEO) database were analyzed by GEO2R analysis, and co-expressed DEGs among the datasets were identified using a Venn drawing tool. The co-expressed DEGs were investigated using Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis, and hub genes were determined based on the protein-protein interaction network established by STRING. After survival analysis performed on UALCAN website, possible key genes were selected and verified in ccRCC cell lines and ccRCC tissues ($n = 44$). Statistical analysis was conducted using GraphPad Prism (Version 8.1.1).

Results. A total of 104 co-expressed DEGs were identified in the three datasets. Pathway analysis revealed that these genes were enriched in the extracellular matrix (ECM)–receptor interaction, protein digestion and absorption and focal adhesion. Survival analysis on 17 hub genes revealed that four key genes with a significant impact on survival: procollagen C-endopeptidase enhancer (*PCOLCE*), prolyl 4-hydroxylase subunit beta (*P4HB*), collagen type VI alpha 2 (*COL6A2*) and collagen type VI alpha 3 (*COL6A3*). Patients with higher expression of these key genes had worse survival than those with lower expression. *In vitro* experiments revealed that the mRNA expression levels of *PCOLCE*, *P4HB* and *COL6A2* were three times higher and that of *COL6A3* mRNA was 16 times higher in the metastatic ccRCC cell line Caki-1 than the corresponding primary cell line Caki-2. Immunohistochemistry revealed higher expression of the proteins encoded by these four genes in metastatic ccRCC compared with tumors from the corresponding primary sites, with statistical significance.

Conclusion. PCOLCE, P4HB, COL6A2 and COL6A3 are upregulated in metastatic ccRCC and might be related to poor prognosis and distant metastases.

Corresponding author
Shicong Yang, yang-shc@mail.sysu.edu.cn

INTRODUCTION

Clear cell renal cell carcinoma (ccRCC) is a highly aggressive malignancy that accounts for 85%–90% of all renal cell carcinoma cases (*Chen et al., 2017*). Approximately 17%–30% of ccRCC patients show distant metastasis at the time of diagnosis (*Motzer, Bander & Nanus, 1996*). Since ccRCC is not sensitive to radiotherapy and chemotherapy, radical or partial nephrectomy is still the main treatment, and 20% to 40% of cases that receive radical/partial nephrectomy still show recurrence or metastasis (*Teng et al., 2014*). Through the in-depth studies of ccRCC driving genes, studies have shown promising findings for molecular targeted treatments such as those targeting hypoxia-inducible factor 2α (*HIF2α*), indicating the potential success with new targeted therapies (*Chen et al., 2016*). However, treatments for patients with metastasis are still limited and the mechanism underlying ccRCC metastasis remains unclear. Hence, clarifying the mechanism on ccRCC metastasis is critical for of effective therapeutic targets to ccRCC patients with distant metastasis.

Microarray technology and bioinformatics analysis have facilitated the exploration of genetic alterations in various cancers. Bioinformatics analyses have helped researchers to identify genes specifically related with distant metastasis in a variety of tumors. For example, bioinformatics studies revealed that the expression of the sodium-potassium-chloride cotransporter 1 (NKCC1) gene was significantly higher in gastric cancer tissue than in normal gastric tissue. NKCC1 was shown to promote the proliferation, invasion, migration and epithelial to mesenchymal transition (EMT) of gastric cancer cells (*Wang et al., 2021*). Another study reported that U three protein 14a (UTP14A) stimulates the proliferation and metastasis of esophageal squamous cell carcinoma cells by regulating the PERK/eIF2a signaling pathway (*Li et al., 2021*).

With the aim to elucidate the key molecular events involved in metastatic ccRCC, we downloaded and analyzed three gene expression microarray datasets on ccRCC from GEO. We performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment of differentially expressed genes (DEGs) to investigate pathways involved in ccRCC metastasis. We also identified key factors associated with ccRCC prognosis and metastasis through protein-protein interaction (PPI) network construction, survival analysis and *in vitro* verification.

MATERIALS AND METHODS

Microarray data collection

Three gene expression datasets, GSE85258 (*Ho et al., 2017*), GSE105288 (*Nam et al., 2019*) and GSE22541 (*Wuttig et al., 2012*), were selected and downloaded from GEO (http://www.ncbi.nlm.nih.gov/geo) (*Edgar, Domrachev & Lash, 2002*). GSE85258 includes 14 primary ccRCC samples and 14 patient-matched pulmonary metastatic ccRCC samples; GSE105288 includes nine primary ccRCC samples and 26 metastatic ccRCC samples; and GSE22541 includes 24 primary ccRCC samples and 24 pulmonary metastatic ccRCC samples. The samples above are from the primary and metastatic ccRCC sources. The GPL10558 Illumina HumanHT-12 V4.0 expression beadchip (GSE85258) and GPL570

[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array (GSE105288 and GSE22541) contain the microarray data of the three gene datasets.

Identification of DEGs

GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r), an interactive web tool for DEG analysis of GEO series across experimental conditions, was used to analyze the DEGs between primary ccRCC and metastatic ccRCC in the three datasets. To screen out more DEGs of these datasets, gene expression with |log2 fold change| > 0.5 and $P < 0.05$ were considered statistically significant. An online Venn drawing tool (http://bioinformatics.psb.ugent.be/webtools/Venn/) was used to obtain the intersecting co-expressed DEGs among the three datasets.

Bioinformatic analyses of DEGs, hub gene selection and analysis

GO dataset (biological process (BP), cellular component (CC), molecular function (MF)) (*Ashburner et al., 2000*) and KEGG pathway (*Kanehisa et al., 2010*) functional enrichment analyses of DEGs were conducted by the Database for Annotation, Visualization and Integrated Discovery (DAVID, http://david.ncifcrf.gov/, version 6.8) (*Dennis et al., 2003*). False discovery rate (FDR) < 0.05 was considered statistically significant. The bubble plots of functional enrichment analysis were performed by bubble tools in Hiplot (https://hiplot.com.cn), a comprehensive web platform for scientific data visualization.

The Search Tool for the Retrieval of Interacting Genes (STRING, http://string-db.org, version 11.0) was used to predict the PPI network of DEGs (*Szklarczyk et al., 2015*). An interaction with a combined score > 0.4 among DEGs was recognized as statistical significance. The Molecular Complex Detection (MCODE) app of Cytoscape version 3.7.1 (*Smoot et al., 2011*) was used to cluster the PPI network of DEGs based on topology and to identify the closely connected area (*Bandettini et al., 2012*). The criteria were as follows: degree cut-off = 2, node score cut-off = 0.2, max depth = 100 and k-score = 2. A degree ≥ 10 was set for hub gene selection.

Survival analysis of hub genes was conducted on UALCAN (http://ualcan.path.uab.edu/index.html). $P < 0.01$ was interpreted to be statistical significance to enroll crucial genes that are related to patients' survival as many as possible. The UALCAN, which mainly based on the TCGA database, is also used to analyze the expression level of genes-of-shortlist between normal renal tissue and primary ccRCC. $P < 0.05$ was considered statistically significant. (*Chandrashekar et al., 2017*).

Cell culture

Human ccRCC cell lines Caki-1 and Caki-2 were purchased from the BIOWING company in Shanghai. Cells were cultured in DMEM medium with 10% fetal bovine serum (FBS) and maintained at 37 °C in a 5% CO_2 incubator.

qRT-PCR

TRIzol was used to extract total RNA from cells. The Roche Transcriptor cDNA synthesis kit was used to reverse transcribe RNA to cDNA. qRT-PCR was performed on a Bio-Rad CFX-96 system (Bio-Rad), and reactions were performed in triplicate. The qRT-PCR primer sequences are as follows:

Beta-Actin (F): 5′-CACCATTGGCAATGAGCGGTTC-3′
Beta-Actin (R): 5′- AGGTCTTTGCGGATGTCCACGT-3′
PCOLCE (F): 5′-GACACCTACTGCCGCTATGACT-3′
PCOLCE (R): 5′- GACGAGGAGTTCATTCCCTTCG-3′
P4HB (F): 5′-TCACCAAGGAGAACCTGCTGGA-3′
P4HB (R): 5′-GGCAAGAACAGCAGGATGTGAG-3′
COL6A2 (F): 5′-CGTGGAGACTCAGGACAGCCA-3′
COL6A2 (R): 5′-CCTTTCAAGCCAAAGTCGCCTC-3′
COL6A3 (F): 5′-CCTGGTGTAACTGATGCTGCCA-3′
COL6A3 (R): 5′-AAGATGGCGTCCACCTTGGACT-3′

Immunohistochemistry

Paraffin-embedded formalin-fixed sections and clinical information of 22 paired primary ccRCCs and their metastatic counterparts were collected at the First Affiliated Hospital of Sun Yat-sen University. The metastatic sites included bone (10 cases), adrenal gland (3 cases), pancreas (2 cases), liver (2 cases), lung (1 case), brain (1 case), abdominal wall (1 case), nasal cavity (1 case) and chest wall (1 case). The study was approved by the institutional ethics committee of the First Affiliated Hospital of Sun Yat-sen University.

For section preparation, xylene and ethanol were used for deparaffinization and rehydration. The antigen retrieval procedure was performed with EDTA buffer in a microwave oven at 700 W for 5 minutes and then 600 W for 20 minutes. Slides were incubated in 3% hydrogen peroxide for 10 minutes and blocked in 5% bovine serum albumin (BSA) for 10 minutes. Slides were incubated with primary antibodies overnight at 4 °C. The primary antibodies included antibodies against PCOLCE (1:100; ab154261, Abcam), P4HB (1:200; ab137110, Abcam), COL6A2 (1:100; YT1035, Immunoway) and COL6A3 (1:100; YT1036, Immunoway). Slides were then incubated with SignalStain Boost IHC Detection Reagent (HRP, Rabbit, #8114, CST) as secondary antibody for 30 minutes at room temperature. The SignalStain DAB Substrate Kit (#8059, CST) was used for staining.

Slides were scanned with a Digital Pathological Section Scanner (KF-PRO-005) and three images were randomly selected in 200X magnification as the areas of interest (AOI), so as to minimize the effect of heterogeneity. The integral optical density (IOD) of each image was measured and calculated as the pixel area with Image Pro Plus for quantification of protein expression. Samples with a higher mean IOD value of AOIs were considered to have a stronger expression. After the quantification, 2 pathologists reviewed all slides and confirmed that the expression trends in the slides were consistent with the quantification results.

Statistical analysis

Statistical analyses were conducted in GraphPad Prism (Version 8.1.1). The results of qRT-PCR were analyzed with unpaired t-test, and the results of immunohistochemistry quantification were analyzed with paired t-test. The error bars of qRT-PCR represent $\pm$ standard deviation (SD) of samples. $P < 0.05$ was considered statistically significant.
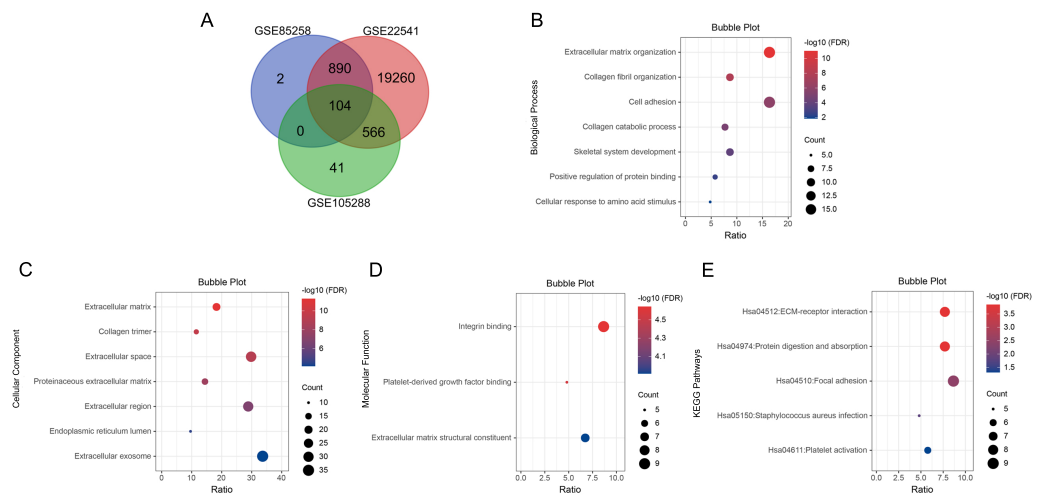
Figure 1 **Venn diagram, GO enrichment analysis and KEGG pathway analysis (FDR < 0.05).** (A) DEGs were selected with $P < 0.05$ and |log2 FoldChange| > 0.5 among the datasets GSE85258, GSE105288 and GSE22541. Venn diagram showed an overlap of 104 genes. (B) All BP results of 104 DEGs. (C) All CC results of 104 DEGs. (D) All MF results of 104 DEGs. (E) All KEGG pathway enrichment results of the 104 DEGs. The bubble plots were performed in Hiplot website.

RESULTS

Identification of DEGs in metastatic ccRCC

We searched for DEGs in metastatic ccRCC samples compared with primary ccRCC samples in three datasets. A total of 996 DEGs in GSE85258, 711 DEGs in GSE105288 and 20,820 DEGs in GSE2254 were obtained. Venn diagram analysis revealed 104 co-expressed DEGs among the three datasets (Fig. 1A).

Bioinformatic analyses of co-expressed DEGs

GO and KEGG pathway enrichment analyses were applied for the 104 DEGs using DAVID. The GO results revealed that in BP, DEGs were mainly enriched in extracellular matrix organization, collagen fibril organization, cell adhesion, collagen catabolic process, skeletal system development, positive regulation of protein binding, cellular response to amino acid stimulus and collagen biosynthetic process (Fig. 1B). In CC, DEGs were enriched in extracellular matrix, collagen trimer, extracellular space, proteinaceous extracellular matrix, extracellular region, endoplasmic reticulum lumen, extracellular exosome and basement membrane (Fig. 1C). In MF, DEGs were predominantly enriched in integrin binding, platelet-derived growth factor binding and extracellular matrix structural constituent (Fig. 1D). KEGG pathway analysis indicated that the DEGs were enriched in ECM-receptor interaction, protein digestion and absorption, focal adhesion, Staphylococcus aureus infection and platelet activation (Fig. 1E).

PPI network construction, hub gene selection and key gene identification

Analysis using the STRING online database demonstrated the PPI network of the DEGs. The most significant module containing 17 hub genes was established by the MCODE app of Cytoscape (Figs. 2A, 2B). GO and KEGG pathway enrichment analyses of the 17 hub genes are shown in Table 1. GO enrichment results revealed that the 17 hub genes were mainly enriched in extracellular matrix organization, collagen fibril organization, collagen catabolic process, skeletal system development cellular response to amino acid, cell adhesion and protein heterotrimerization in BP; extracellular matrix, collagen trimer, endoplasmic reticulum lumen, proteinaceous extracellular matrix, extracellular region 14, extracellular space and extracellular exosome in CC; and extracellular matrix structural constituent, platelet-derived growth factor binding, integrin binding, heparin binding, SMAD binding 3, collagen binding and peptidase activator activity in MF. KEGG pathways were mainly enriched in protein digestion and absorption, ECM-receptor interaction, focal adhesion, PI3K-Akt signaling pathway, Amoebiasis and Platelet activation (Table 1).

Four genes, *PCOLCE*, *P4HB*, *COL6A2* and *COL6A3* were sorted out with *P*-value < 0.01 among the 17 hub genes after survival analysis performed on the UALCAN website. Data showed that patients with high expression of *PCOLCE*, *P4HB*, *COL6A2* and *COL6A3* had worse survival than those with low expression of each gene (Figs. 2A, 2B). Also, these four genes were found to be higher expressed in primary ccRCC compared to normal kidney tissue (Fig. S1).

In vitro validation of the four key genes in metastatic ccRCC

qRT-PCR showed that *PCOLCE*, *P4HB*, *COL6A2* and *COL6A3* mRNAs were all upregulated in the Caki-1 cell line compared with its primary counterpart, the Caki-2 cell line (*P* < 0.05) (Fig. 3A).

We further examined the expressions of the proteins encoded by the four genes in 22 pairs of primary ccRCC and their corresponding metastatic counterpart tissue with immunohistochemistry. The TNM staging information of 12 cases of primary ccRCC at first diagnosis is presented in Table S1. Higher expressions of all four proteins were observed in most metastatic ccRCC tissues compared with the primary sites, as shown by mean IOD results (*P* < 0.05). Higher expressions of COL6A2 and PCOLCE in metastatic counterparts were observed in 19 out of 22 pairs, and higher expressions of COL6A3 and P4HB expression were detected in 17 out of 22 pairs (Fig. 3B).

DISCUSSION

CcRCC is believed to originate from proximal tubular epithelium with aggressive behavior that was unparallel with its bland appearance. After surgery, the median survival time of ccRCC patients with metastasis is only 13 months, and the 5-year survival rate is 12.3% (*Babaian et al., 2014*). As one of the main genetic events contributing to ccRCC pathogenesis, the mutant *VHL* gene pathway has become a target for treatment towards ccRCC (*Chen et al., 2016*). With the ongoing effort in bioinformatic data management, novel candidate key genes may be identified as new therapeutic targets for ccRCC patients
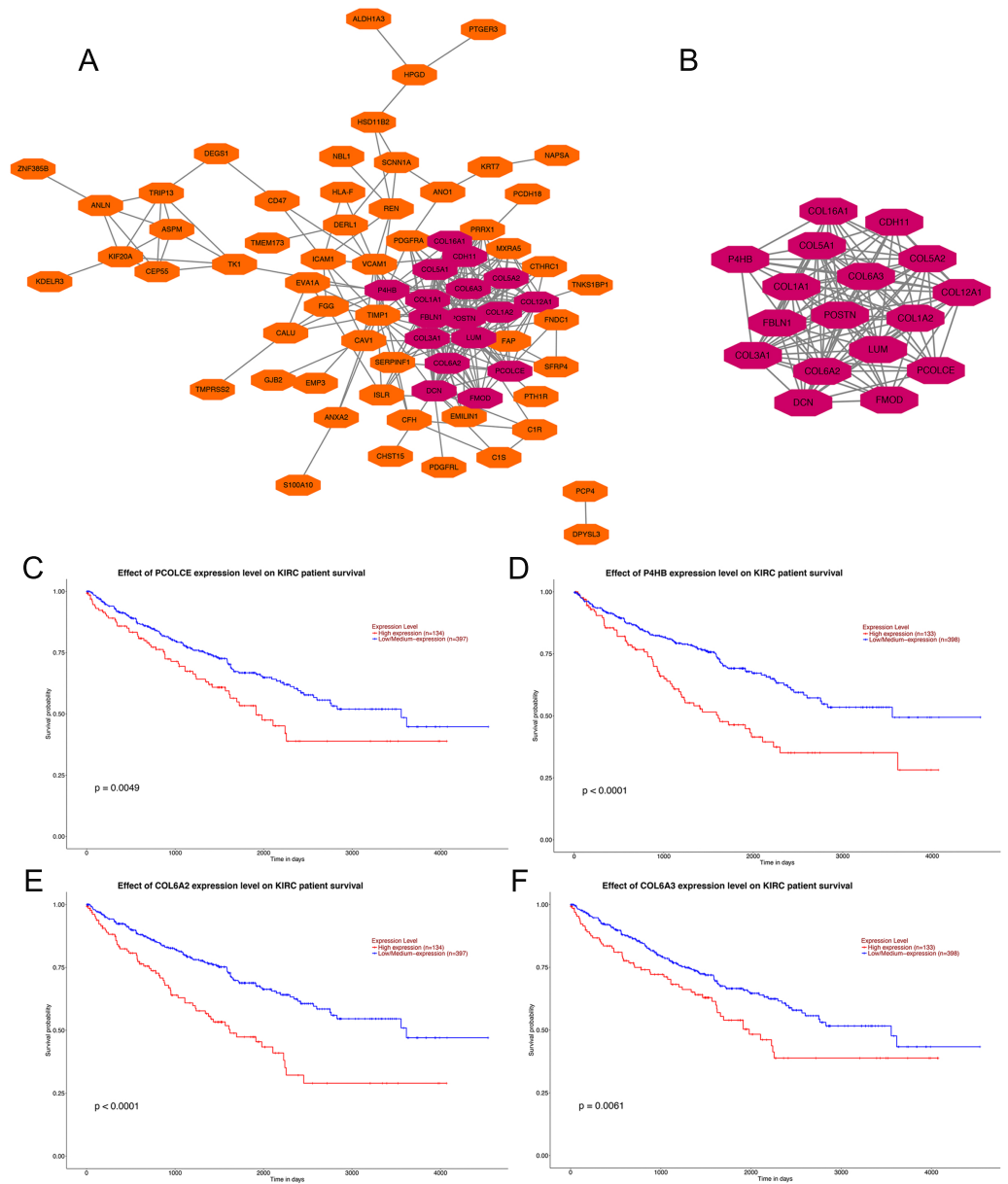
Figure 2 PPI network of DEGs, hub genes selection, and survival analysis of four key genes. (A) PPI network of 104 DEGs, the representation of nodes and edges are protein and the interaction of the proteins. (B) Seventeen hub genes of DEGs were obtained with module analysis using the MCODE app. ($P < 0.01$, KIRC: kidney renal clear cell carcinoma). (C) Survival analysis of *PCOLCE*. (D) Survival analysis of *P4HB*. (E) Survival analysis of *COL6A2*. (F) Survival analysis of *COL6A3*.

with metastasis. In order to obtain datasets with a long time span and sufficient samples from different institutions, we downloaded 3 datasets from GEO websites to screen out the intersection of differential expression genes.

In our study, we first obtained 104 co-expressed DEGs from three datasets. In the course of searching for key genes involved in ccRCC metastasis, we conducted bioinformatic

Table 1 GO and KEGG pathway enrichment analysis of 17 hub DEGs.

Term	Description	Gene count	Gene ratio (%)	FDR
GOTERM_BP_DIRECT				
GO:0030198	extracellular matrix organization	12	70.59	2.39E−16
GO:0030199	collagen fibril organization	8	47.06	1.62E−13
GO:0030574	collagen catabolic process	8	47.06	4.33E−12
GO:0001501	skeletal system development	7	41.18	6.92E−08
GO:0071230	cellular response to amino acid stimulus	5	29.41	2.35E−06
GO:0007155	cell adhesion	8	47.06	2.35E−06
GO:0070208	protein heterotrimerization	3	17.65	0.00154
GO:0043588	skin development	3	17.65	0.00821
GO:0001568	blood vessel development	3	17.65	0.00912
GO:0009612	response to mechanical stimulus	3	17.65	0.01975
GO:1903225	negative regulation of endodermal cell differentiation	2	11.76	0.02424
GO:0048592	eye morphogenesis	2	11.76	0.03332
GO:0007229	integrin-mediated signaling pathway	3	17.65	0.04214
GOTERM_CC_DIRECT				
GO:0031012	extracellular matrix	15	88.24	2.13E−22
GO:0005581	collagen trimer	9	52.94	5.36E−14
GO:0005788	endoplasmic reticulum lumen	10	58.82	1.33E−13
GO:0005578	proteinaceous extracellular matrix	10	58.82	2.06E−12
GO:0005576	extracellular region 14	14	82.35	4.64E−11
GO:0005615	extracellular space	12	70.59	4.99E−09
GO:0070062	extracellular exosome	10	58.82	7.93E−04
GO:1903561	extracellular vesicle	3	17.65	0.00302
GO:0005584	collagen type I trimer	2	11.76	0.00546
GO:0043202	lysosomal lumen	3	17.65	0.00614
GO:0005589	collagen type VI trimer	2	11.76	0.00614
GO:0005588	collagen type V trimer	2	11.76	0.00614
GO:0005796	Golgi lumen	3	17.65	0.00676
GOTERM_MF_DIRECT				
GO:0005201	extracellular matrix structural constituent	7	41.18	7.00E−10
GO:0048407	platelet-derived growth factor binding	4	23.53	1.66E−06
GO:0005178	integrin binding	5	29.41	2.35E−05
GO:0008201	heparin binding	4	23.53	0.00310
GO:0046332	SMAD binding 3	3	17.65	0.00431
GO:0005518	collagen binding	3	17.65	0.00698
GO:0016504	peptidase activator activity	2	11.76	0.03911
KEGG_PATHWAY				
hsa04974	Protein digestion and absorption	8	47.06	2.56E−11
hsa04512	ECM-receptor interaction	7	41.18	1.73E−09
hsa04510	Focal adhesion	7	41.18	2.12E−07
hsa04151	PI3K-Akt signaling pathway	7	41.18	3.37E−06
hsa05146	Amoebiasis	5	29.41	1.04E−05
hsa04611	Platelet activation	5	29.41	2.34E−05

Notes.

GO, Gene Ontology; MF, Molecular Function; CC, cellular component; BP, biological process; KEGG, Kyoto Encyclopedia of Genes and Genomes.
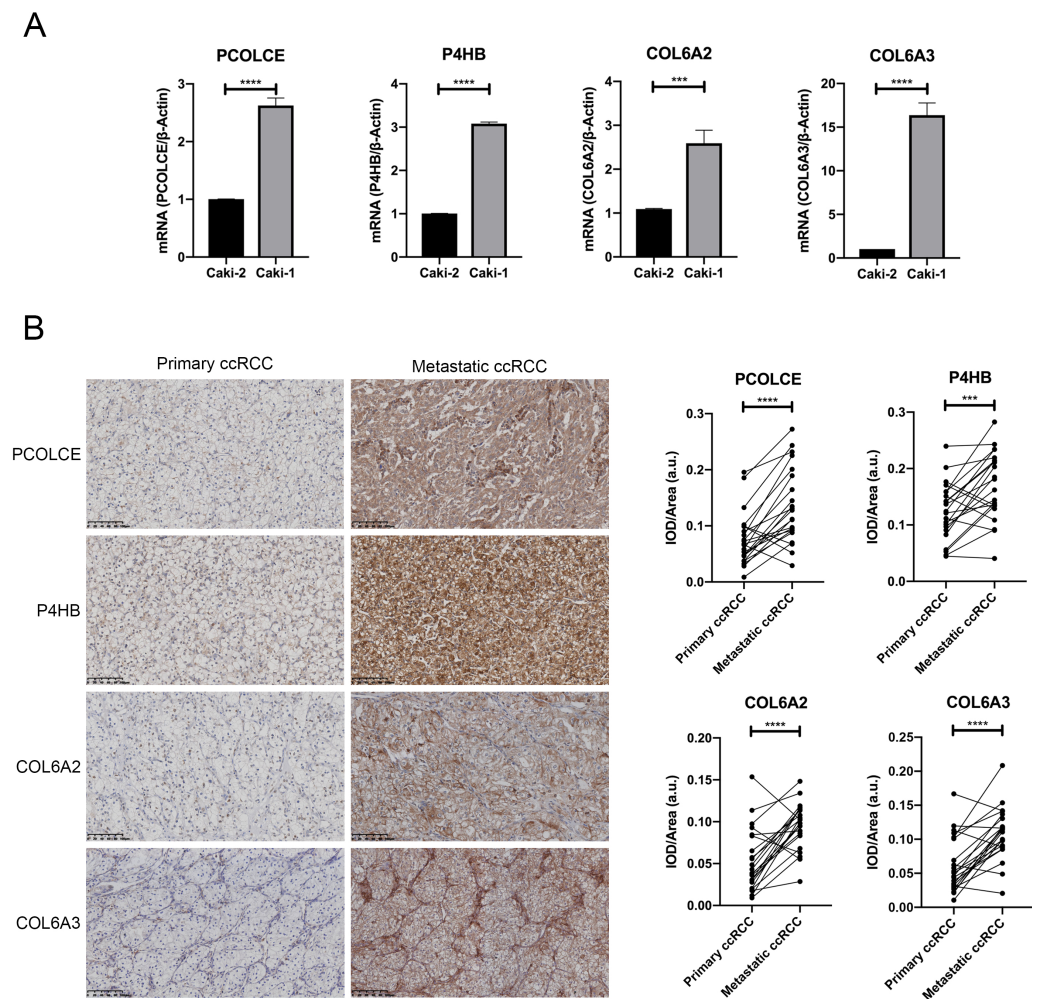
Figure 3 Expression of key genes in primary ccRCC and the corresponding metastatic ccRCC. (A) the qRT-PCR analysis showed the expression of *PCOLCE*, *P4HB*, *COL6A2* and *COL6A3* in primary ccRCC cell line Caki-2 and its corresponding metastatic ccRCC cell line Caki-1. (***$P < 0.001$, ****$P < 0.0001$) (B) Representative IHC staining images of *PCOLCE*, *P4HB*, *COL6A2* and *COL6A3* in human primary ccRCC and corresponding metastatic ccRCC (22 paired cases) are on the left, and the quantitative results on the right. 200x. (***$P < 0.001$, ****$P < 0.0001$).

analysis on 17 hub genes which were screened from the PPI network analysis to identify specific pathways that might be related to ccRCC metastasis. GO analysis showed the DEGs were mainly distributed in extracellular matrix, extracellular space, extracellular exosome and Golgi lumen of CC, and were mainly enriched in extracellular matrix organization, collagen fibril organization, cell adhesion and blood vessel development of BP, and extracellular matrix structural constituents, platelet-derived growth factor binding, integrin binding and collagen binding of MF. KEGG pathway analysis revealed that these DEGs were enriched in protein digestion and absorption, ECM-receptor interaction, focal adhesion and PI3K-Akt signaling pathway.

Exosomes, which functioned in carrying protein, ceramide, mRNA or micro-RNA as cargo, can induce metastasis through participation in EMT, mesenchymal-epithelial reverting transition and neoangiogenesis (*Steinbichler et al., 2017*). For example, in ccRCC cell lines, exosomes containing miR-19b-3p could initiate EMT and accelerate tumor metastasis by reducing PTEN expression (*Wang et al., 2019a*; *Wang et al., 2019b*). EMT was essential for tumor metastasis through inducing processes such as loss of cell polarity, disconnection with the basement membrane, acquiring plasticity, invasion and anti-apoptosis activities (*Talbot, Bhattacharya & Kuo, 2012*). Additionally, cell adhesion to ECM participates in multiple steps of tumor metastasis through ECM receptor interaction (*Hamidi & Ivaska, 2018*). For example, through the several binding sites from matricellular proteins members of the CCN family, integrin could help tumor cells resist anoikis and promote tumor cells metastasize (*Lau, 2016*; *Simpson, Anyiwe & Schimmer, 2008*). Studies have shown that the PI3K-Akt signaling pathway might induce the EMT process directly or through crosstalk with the Wnt/β-catenin signaling pathway, TGF-β or NF-kB and initiate distant metastasis (*Xu, Yang & Lu, 2015*). The phosphorylated AKT, a key kinase of the PI3K signaling pathway, was reported to be related to EMT in ccRCC (*Wu et al., 2011*). Our results also showed that DEGs in metastatic ccRCC were enriched in the PI3K signaling pathway, suggesting that PI3K signaling pathway–mediated EMT was likely to participate in the metastasis of ccRCC. Together, our results have revealed various processes that might be involved in ccRCC metastasis.

To explore the relationship between 17 hub genes and patient survival, the survival analyses were conducted. From survival analyses, *PCOLCE*, *P4HB*, *COL6A2* and *COL6A3* were found to be related to ccRCC patient survival, and these candidate genes were further tested with qRT-PCR and immunohistochemical staining in ccRCC cell lines and tissues. We found that the four genes showed higher expression in the metastatic counterpart of ccRCC both in cell lines and paired ccRCC tissues. *P4HB* and *COL6A3* have been previously reported to be associated with ccRCC. And our study presents the linkage of *PCOLCE* and *COL6A2* with ccRCC.

To the best of our knowledge, this is the first report that showed the relevance between *PCOLCE* and ccRCC. PCOLCE is a secreted glycoprotein that participates in the maturation of procollagen and ECM reconstruction by enhancing the activity of bone morphogenetic protein-1 (BMP-1) (*Kessler, Mould & Hulmes, 1990*; *Moali et al., 2005*; *Zhu et al., 2019a*; *Zhu et al., 2019b*). *PCOLCE* was overexpressed in osteosarcoma and promoted lung metastasis via twist family bHLH transcription factor 1 (TWIST1). Reducing *PCOLCE* expression could prevent the migration, invasion and lung metastasis of osteosarcoma cells (*Wang et al., 2019a*; *Wang et al., 2019b*). *BMP-1* was reported to be up-regulated in gastric cancer and associated with poor survival and distant metastasis (*Hsieh et al., 2018*). Notably, a previous study showed that BMP-1 was elevated in ccRCC and promoted proliferation, migration and invasion in ccRCC cell lines (*Xiao et al., 2020*). Our study showed that the expression of *PCOLCE*, which functions as an enhancer of *BMP-1*, was higher in metastatic ccRCC. This finding raised the hypothesis that PCOLCE might be involved in ccRCC metastasis through regulation of *BMP-1*. Therefore, further research is needed to clarify the function and potential mechanism of PCOLCE in ccRCC metastasis.

P4HB, a beta subunit of prolyl 4-hydroxylase, helps cancer cells survive from apoptosis induced by endoplasmic reticulum stress (*Lovat et al., 2008*; *Pajunen et al., 1991*). P4HB promoted proliferation, invasion, migration and angiogenesis in glioma through the mitogen-activated protein kinase (MAPK) signaling pathway (*Sun et al., 2017*). In gastric cancer, P4HB was shown to play an important role in regulating invasion and migration in the HIF1α pathway (*Zhang et al., 2019*). Previous studies showed that HIF1α was activated in renal clear cell carcinoma and associated with a more aggressive phenotype (*Maxwell et al., 1999*; *Na et al., 2003*). And the overexpression of *P4HB* in ccRCC was associated with poor outcome (*Zhu et al., 2019a*; *Zhu et al., 2019b*). In our study, we found that *P4HB* was specifically elevated in metastatic ccRCC. Hence, these findings suggest that P4HB might play a key role in modulating ccRCC metastasis through HIF1α pathway.

In this study, we showed that *COL6A2* and *COL6A3* were related to ccRCC progression. COL6A1–3 are three subunits of collagen VI (*Hessle & Engvall, 1984*; *Weil et al., 1988*). Strong evidence has demonstrated the role of COL6A1–3 in promoting the progression of tumors (*Chen, Cescon & Bonaldo, 2013*). In serous ovarian cancer, *COL6A2* was dysregulated and associated with poor prognosis through the TGFβ1 pathway (*Cheon et al., 2014*; *Svoronos et al., 2020*). COL6A3 functions in tumorigenesis and progression of cholangiocarcinoma through the E2F1/LMCD1-AS1/miR-345-5p/COL6A3 axis and serves as prognostic factor for pancreatic cancer (*Yu et al., 2019*). Up-regulation of *COL6A1* in ccRCC was associated with poor prognosis of patients and promoted tumor growth *in vivo* (*Wan et al., 2015*). *COL6A3* was reported to be overexpressed in metastatic renal cell carcinoma and associated with poor survival (*Li et al., 2019*). To date, no report has described the relation between *COL6A2* and ccRCC. As *COL6A2* and *COL6A3* were identified as key genes in our results, further research should be carried out to examine their mechanisms in ccRCC metastasis.

This study has several limitations. First, we only examined 22 paired ccRCC samples and 2 cell lines to verify the bio-informatic findings. More samples need to be examined to confirm our results. Second, the mechanisms of the four key genes in regulating ccRCC cell metastasis are still unclear. Further exploration is required to clarify the function and potential mechanism of the four key proteins both *in vitro* and *in vivo*.

CONCLUSION

Our studies identified four key genes (*PCOLCE*, *P4HB*, *COL6A2* and *COL6A3*) that were related to ccRCC patient survival and showed potential involvement in ccRCC metastasis. The proteins encoded by the four genes may be involved in the ECM-receptor interaction, focal adhesion and the PI3K-Akt signaling pathway. In vitro experiments confirmed increased expression of the candidate genes in metastatic ccRCC, both in cell lines and tissues. Future studies are required to evaluate the significance and mechanism of these key proteins in ccRCC.

ACKNOWLEDGEMENTS

We thank the Nephropathy Laboratory of the First Affiliated Hospital of Sun Yat-sen University for providing instruments for our experiments *in vitro*. And we thank Gabrielle White Wolf, PhD, from Liwen Bianji, Edanz Editing China, for editing the English text of a draft of this manuscript.

Funding

Our study is funded by National Natural Science Foundation of China (No. 81772701). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
National Natural Science Foundation of China: 81772701.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Tingting Zhong conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Zeying Jiang and Honglei Wang analyzed the data, prepared figures and/or tables, and approved the final draft.
- Xiangdong Wang and Meiyi Song performed the experiments, prepared figures and/or tables, and approved the final draft.
- Wenfang Chen and Shicong Yang conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

Animal Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

The study was approved by the institutional ethics committee of the First Affiliated Hospital of Sun Yat-sen University ([2017]233).

Data Availability

The raw measurements are available in the Supplementary File.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.12493#supplemental-information.

REFERENCES

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene ontology: tool for the unification of biology. *The Gene Ontology Consortium. Nature Genetics* **25**:25–29 DOI 10.1038/75556.

Babaian KN, Merrill MM, Matin S, Tamboli P, Tannir NM, Jonasch E, Wood CG, Karam JA. 2014. Partial nephrectomy in the setting of metastatic renal cell carcinoma. *Journal of Urology* **192**:36–42 DOI 10.1016/j.juro.2014.01.086.

Bandettini WP, Kellman P, Mancini C, Booker OJ, Vasu S, Leung SW, Wilson JR, Shanbhag SM, Chen MY, Arai AE. 2012. MultiContrast Delayed Enhancement (MCODE) improves detection of subendocardial myocardial infarction by late gadolinium enhancement cardiovascular magnetic resonance: a clinical validation study. *Journal of Cardiovascular Magnetic Resonance* **14**:83 DOI 10.1186/1532-429x-14-83.

Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, Varambally S. 2017. UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia* **19**:649–658 DOI 10.1016/j.neo.2017.05.002.

Chen P, Cescon M, Bonaldo P. 2013. Collagen VI in cancer and its biological mechanisms. *Trends in Molecular Medicine* **19**:410–417 DOI 10.1016/j.molmed.2013.04.001.

Chen W, Hill H, Christie A, Kim MS, Holloman E, Pavia-Jimenez A, Homayoun F, Ma Y, Patel N, Yell P, Hao G, Yousuf Q, Joyce A, Pedrosa I, Geiger H, Zhang H, Chang J, Gardner KH, Bruick RK, Reeves C, Hwang TH, Courtney K, Frenkel E, Sun X, Zojwalla N, Wong T, Rizzi JP, Wallace EM, Josey JA, Xie Y, Xie XJ, Kapur P, McKay RM, Brugarolas J. 2016. Targeting renal cell carcinoma with a HIF-2 antagonist. *Nature* **539**:112–117 DOI 10.1038/nature19796.

Chen J, Shu Y, Yu Q, Shen W. 2017. MicroRNA-645 promotes cell metastasis and proliferation of renal clear cell carcinoma by targeting GK5. *European Review for Medical and Pharmacological Sciences* **21**:4557–4565.

Cheon DJ, Tong Y, Sim MS, Dering J, Berel D, Cui X, Lester J, Beach JA, Tighiouart M, Walts AE, Karlan BY, Orsulic S. 2014. A collagen-remodeling gene signature regulated by TGF-β signaling is associated with metastasis and poor survival in serous ovarian cancer. *Clinical Cancer Research* **20**:711–723 DOI 10.1158/1078-0432.Ccr-13-1256.

Dennis Jr G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. 2003. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biology* **4**(**P3**).

Edgar R, Domrachev M, Lash AE. 2002. Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research* **30**:207–210 DOI 10.1093/nar/30.1.207.
Hamidi H, Ivaska J. 2018. Every step of the way: integrins in cancer progression and metastasis. *Nature Reviews Cancer* **18**:533–548 DOI 10.1038/s41568-018-0038-z.

Hessle H, Engvall E. 1984. Type VI collagen, studies on its localization, structure, and biosynthetic form with monoclonal antibodies. *Journal of Biological Chemistry* **259**:3955–3961 DOI 10.1016/S0021-9258(17)43189-9.

Ho TH, Serie DJ, Parasramka M, Cheville JC, Bot BM, Tan W, Wang L, Joseph RW, Hilton T, Leibovich BC, Parker AS, Eckel-Passow JE. 2017. Differential gene expression profiling of matched primary renal cell carcinoma and metastases reveals upregulation of extracellular matrix genes. *Annals of Oncology* **28**:604–610 DOI 10.1093/annonc/mdw652.

Hsieh YY, Tung SY, Pan HY, Yen CW, Xu HW, Deng YF, Lin YJ, Hsu WT, Wu CS, Li C. 2018. Upregulation of bone morphogenetic protein 1 is associated with poor prognosis of late-stage gastric Cancer patients. *BMC Cancer* **18**:508 DOI 10.1186/s12885-018-4383-9.

Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. 2010. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Research* **38**:D355–D360 DOI 10.1093/nar/gkp896.

Kessler E, Mould AP, Hulmes DJ. 1990. Procollagen type I C-proteinase enhancer is a naturally occurring connective tissue glycoprotein. *Biochemical and Biophysical Research Communications* **173**:81–86 DOI 10.1016/s0006-291x(05)81024-1.

Lau LF. 2016. Cell surface receptors for CCN proteins. *Journal of Cell Communication and Signaling* **10**:121–127 DOI 10.1007/s12079-016-0324-z.

Li KK, Mao CY, Ma Q, Bao T, Wang YJ, Guo W, Zhao XL. 2021. U three protein 14a (UTP14A) promotes tumour proliferation and metastasis via the PERK/eIF2a/GRP78 signalling pathway in oesophageal squamous cell carcinoma. *Journal of Cancer* **12**:134–140 DOI 10.7150/jca.44649.

Li C, Shao T, Bao G, Gao Z, Zhang Y, Ding H, Zhang W, Liu F, Guo C. 2019. Identification of potential core genes in metastatic renal cell carcinoma using bioinformatics analysis. *American Journal of Translational Research* **11**:6812–6825.

Lovat PE, Corazzari M, Armstrong JL, Martin S, Pagliarini V, Hill D, Brown AM, Piacentini M, Birch-Machin MA, Redfern CP. 2008. Increasing melanoma cell death using inhibitors of protein disulfide isomerases to abrogate survival responses to endoplasmic reticulum stress. *Cancer Research* **68**:5363–5369 DOI 10.1158/0008-5472.Can-08-0035.

Maxwell PH, Wiesener MS, Chang GW, Clifford SC, Vaux EC, Cockman ME, Wykoff CC, Pugh CW, Maher ER, Ratcliffe PJ. 1999. The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature* **399**:271–275 DOI 10.1038/20459.

Moali C, Font B, Ruggiero F, Eichenberger D, Rousselle P, François V, Oldberg A, Bruckner-Tuderman L, Hulmes DJ. 2005. Substrate-specific modulation of a multisubstrate proteinase, C-terminal processing of fibrillar procollagens is the only

BMP-1-dependent activity to be enhanced by PCPE-1. *Journal of Biological Chemistry* **280**:24188–24194 DOI 10.1074/jbc.M501486200.

Motzer RJ, Bander NH, Nanus DM. 1996. Renal-cell carcinoma. *New England Journal of Medicine* **335**:865–875 DOI 10.1056/nejm199609193351207.

Na X, Wu G, Ryan CK, Schoen SR, di'Santagnese PA, Messing EM. 2003. Overproduction of vascular endothelial growth factor related to von Hippel-Lindau tumor suppressor gene mutations and hypoxia-inducible factor-1 alpha expression in renal cell carcinomas. *Journal of Urology* **170**:588–592 DOI 10.1097/01.ju.0000074870.54671.98.

Nam HY, Chandrashekar DS, Kundu A, Shelar S, Kho EY, Sonpavde G, Naik G, Ghatalia P, Livi CB, Varambally S, Sudarshan S. 2019. Integrative epigenetic and gene expression analysis of renal tumor progression to metastasis. *Molecular Cancer Research* **17**:84–96 DOI 10.1158/1541-7786.Mcr-17-0636.

Pajunen L, Jones TA, Goddard A, Sheer D, Solomon E, Pihlajaniemi T, Kivirikko KI. 1991. Regional assignment of the human gene coding for a multifunctional polypeptide (P4HB) acting as the beta-subunit of prolyl 4-hydroxylase and the enzyme protein disulfide isomerase to 17q25. *Cytogenetic and Genome Research* **56**:165–168 DOI 10.1159/000133078.

Simpson CD, Anyiwe K, Schimmer AD. 2008. Anoikis resistance and tumor metastasis. *Cancer Letters* **272**:177–185 DOI 10.1016/j.canlet.2008.05.029.

Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. 2011. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27**:431–432 DOI 10.1093/bioinformatics/btq675.

Steinbichler TB, Dudás J, Riechelmann H, Skvortsova II. 2017. The role of exosomes in cancer metastasis. *Seminars in Cancer Biology* **44**:170–181 DOI 10.1016/j.semcancer.2017.02.006.

Sun S, Kiang KMY, Ho ASW, Lee D, Poon MW, Xu FF, Pu JKS, Kan ANC, Lee NPY, Liu XB, Man K, Day PJR, Lui WM, Fung CF, Leung GKK. 2017. Endoplasmic reticulum chaperone prolyl 4-hydroxylase, beta polypeptide (P4HB) promotes malignant phenotypes in glioma via MAPK signaling. *Oncotarget* **8**:71911–71923 DOI 10.18632/oncotarget.18026.

Svoronos C, Tsoulfas G, Souvatzi M, Chatzitheoklitos E. 2020. Prognostic value of COL6A3 in pancreatic adenocarcinoma. *Annals of Hepato-Biliary-Pancreatic Surgery* **24**:52–56 DOI 10.14701/ahbps.2020.24.1.52.

Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C. 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research* **43**:D447–D452 DOI 10.1093/nar/gku1003.

Talbot LJ, Bhattacharya SD, Kuo PC. 2012. Epithelial-mesenchymal transition, the tumor microenvironment, and metastatic behavior of epithelial malignancies. *International Journal of Biochemistry and Molecular Biology* **3**:117–136.

Teng J, Gao Y, Chen M, Wang K, Cui X, Liu Y, Xu D. 2014. Prognostic value of clinical and pathological factors for surgically treated localized clear cell renal cell carcinoma. *Chinese Medical Journal* **127**:1640–1644.

Wan F, Wang H, Shen Y, Zhang H, Shi G, Zhu Y, Dai B, Ye D. 2015. Upregulation of COL6A1 is predictive of poor prognosis in clear cell renal cell carcinoma patients. *Oncotarget* **6**:27378–27387 DOI 10.18632/oncotarget.4860.

Wang L, Yang G, Zhao D, Wang J, Bai Y, Peng Q, Wang H, Fang R, Chen G, Wang Z, Wang K, Li G, Yang Y, Wang Z, Guo P, Peng L, Hou D, Xu W. 2019a. CD103-positive CSC exosome promotes EMT of clear cell renal cell carcinoma: role of remote MiR-19b-3p. *Molecular Cancer* **18**:86 DOI 10.1186/s12943-019-0997-z.

Wang JF, Zhao K, Chen YY, Qiu Y, Zhu JH, Li BP, Wang Z, Chen JQ. 2021. NKCC1 promotes proliferation, invasion and migration in human gastric cancer cells via activation of the MAPK-JNK/EMT signaling pathway. *Journal of Cancer* **12**:253–263 DOI 10.7150/jca.49709.

Wang S, Zhong L, Li Y, Xiao D, Zhang R, Liao D, Lv D, Wang X, Wang J, Xie X, Chen J, Wu Y, Kang T. 2019b. Up-regulation of PCOLCE by TWIST1 promotes metastasis in Osteosarcoma. *Theranostics* **9**:4342–4353 DOI 10.7150/thno.34090.

Weil D, Mattei MG, Passage E, N'Guyen VC, Pribula-Conway D, Mann K, Deutzmann R, Timpl R, Chu ML. 1988. Cloning and chromosomal localization of human genes encoding the three chains of type VI collagen. *American Journal of Human Genetics* **42**:435–445.

Wu ST, Sun GH, Hsu CY, Huang CS, Wu YH, Wang HH, Sun KH. 2011. Tumor necrosis factor-α induces epithelial-mesenchymal transition of renal cell carcinoma cells via a nuclear factor kappa B-independent mechanism. *Experimental Biology and Medicine* **236**:1022–1029 DOI 10.1258/ebm.2011.011058.

Wuttig D, Zastrow S, Füssel S, Toma MI, Meinhardt M, Kalman K, Junker K, Sanjmyatav J, Boll K, Hackermüller J, Rolle A, Grimm MO, Wirth MP. 2012. CD31, EDNRB and TSPAN7 are promising prognostic markers in clear-cell renal cell carcinoma revealed by genome-wide expression analyses of primary tumors and metastases. *International Journal of Cancer* **131**:E693–E704 DOI 10.1002/ijc.27419.

Xiao W, Wang X, Wang T, Xing J. 2020. Overexpression of BMP1 reflects poor prognosis in clear cell renal cell carcinoma. *Cancer Gene Therapy* **27**:330–340 DOI 10.1038/s41417-019-0107-9.

Xu W, Yang Z, Lu N. 2015. A new role for the PI3K/Akt signaling pathway in the epithelial-mesenchymal transition. *Cell Adhesion & Migration* **9**:317–324 DOI 10.1080/19336918.2015.1016686.

Yu J, Zhang B, Zhang H, Qi Y, Wang Y, Wang W, Wang Y, Wang Y. 2019. E2F1-induced upregulation of long non-coding RNA LMCD1-AS1 facilitates cholangiocarcinoma cell progression by regulating miR-345-5p/COL6A3 pathway. *Biochemical and Biophysical Research Communications* **512**:150–155 DOI 10.1016/j.bbrc.2019.03.054.

Zhang J, Guo S, Wu Y, Zheng ZC, Wang Y, Zhao Y. 2019. P4HB, a novel hypoxia target gene related to gastric cancer invasion and metastasis. *BioMed Research International* **2019**:9749751 DOI 10.1155/2019/9749751.

Zhu Q, Guo W, Zhang S, Feng Y, Wang X, Zhou L, Huang GR. 2019a. Synergistic effect of PCPE1 and sFRP2 on the processing of procollagens via BMP1. *FEBS Letters* **593**:119–127 DOI 10.1002/1873-3468.13291.

Zhu Z, He A, Lv T, Xu C, Lin L, Lin J. 2019b. Overexpression of P4HB is correlated with poor prognosis in human clear cell renal cell carcinoma. *Cancer Biomark* **26**:431–439 DOI 10.3233/cbm-190450.