Bioscience Reports (2020) **40** BSR20200492
https://doi.org/10.1042/BSR20200492

Research Article

Profiles of overall survival-related gene expression-based risk signature and their prognostic implications in clear cell renal cell carcinoma

Zihao He[1,2], Tuo Deng[1,2], Xiaolu Duan[1,2] and Guohua Zeng[1,2]

[1]Department of Urology and Guangdong Key Laboratory of Urology, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, Guangdong, China, 510230; [2]Guangzhou Institute of Urology, Guangzhou, China

Correspondence: Guohua Zeng (gzgyzgh@vip.tom.com) or Tuo Deng (gzgy_dengtuo@sina.com)

The present work aimed to evaluate the prognostic value of overall survival (OS)-related genes in clear cell renal cell carcinoma (ccRCC) and to develop a nomogram for clinical use. Transcriptome data from The Cancer Genome Atlas (TCGA) were collected to screen differentially expressed genes (DEGs) between ccRCC patients with OS $>$ 5 years (149 patients) and those with $<$1 year (52 patients). In TCGA training set (265 patients), seven DEGs (cytochrome P450 family 3 subfamily A member 7 (CYP3A7), contactin-associated protein family member 5 (CNTNAP5), adenylate cyclase 2 (ADCY2), TOX high mobility group box family member 3 (TOX3), plasminogen (PLG), enamelin (ENAM), and collagen type VII α 1 chain (COL7A1)) were further selected to build a prognostic risk signature by the least absolute shrinkage and selection operator (LASSO) Cox regression model. Survival analysis confirmed that the OS in the high-risk group was dramatically shorter than their low-risk counterparts. Next, univariate and multivariate Cox regression revealed the seven genes-based risk score, age, and Tumor, lymph Node, and Metastasis staging system (TNM) stage were independent prognostic factors to OS, based on which a novel nomogram was constructed and validated in both TCGA validation set (265 patients) and the International Cancer Genome Consortium cohort (ICGC, 84 patients). A decent predictive performance of the nomogram was observed, the C-indices and corresponding 95% confidence intervals of TCGA training set, validation set, and ICGC cohort were 0.78 (0.74–0.82), 0.75 (0.70–0.80), and 0.70 (0.60–0.80), respectively. Moreover, the calibration plots of 3- and 5 years survival probability indicated favorable curve-fitting performance in the above three groups. In conclusion, the proposed seven genes signature-based nomogram is a promising and robust tool for predicting the OS of ccRCC, which may help tailor individualized therapeutic strategies.

Received: 21 February 2020
Revised: 05 August 2020
Accepted: 06 August 2020

Accepted Manuscript online:
13 August 2020
Version of Record published:
16 September 2020

Introduction

As one of the most common urinary malignancies, renal cell carcinoma (RCC) poses a hidden threat to public health and accounts for approximately 2–3% of adult tumors [1]. The main histologic subtype of RCC is clear cell RCC (ccRCC), which constitutes 75–80% of primary renal malignancies [2]. Reportedly, ccRCC generated 65340 newly diagnosed cases and 14970 deaths in America in 2018 [3].

Compared with other tumors, the prognosis of ccRCC patients remains generally preferable as indicated by the 5-year overall survival (OS) of localized (stage I–III) ccRCC had reached up to 70–90% [4]. Despite this, individual variations should be recognized as patients with similar Tumor, lymph Node, and Metastasis staging system (TNM) stages at diagnosis could end up with significantly different OS. For example, 25–33% of localized ccRCC patients could still progress into recurrence and metastasis even after curative resection and associated with a significantly worse prognosis than other patients with localized ccRCC [5,6]. Besides, the morphologic and genetic heterogeneity of ccRCC was discussed in numerous previous studies [7,8]. Such reports suggested that the prevalent prognostic tools for ccRCC, which were based mainly on pathological and clinical features, had unsatisfactory predictive power. Typical tools of this kind included the TNM staging system, necrosis score, and the University of California Integrated Staging System (UISS) [9–11].

Today, with the development of high-throughput sequencing technology, urologists have turned to identify molecular biomarkers for risk stratification and prognosis prediction. In this regard, prognostic tools from a single gene [12–14], to risk signatures consist of a panel of genes [15–17], and to models integrating gene profiling and clinical features [18–20], have been widely reported. However, there has been no study yet in this field to investigate genes correlated directly to ccRCC patients' OS and to explore their prognostic values in medical practice, and we believe that data mining in this topic could provide new insight into ccRCC progression and help improve therapeutic strategies to a great extent.

Materials and methods

Data acquisition

The transcriptome profiling data and clinical information of ccRCC were obtained from The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/) in December 2019. Expression data as Fragments Per Kilobase per Million (FPKM) files were available for 539 tumor samples, 9 of which lacked survival information. The 530 samples with full expression and clinical data were randomly and evenly assigned to a training set and a validation set via the 'Classification and Regression Training (caret)' package (http://topepo.github.io/caret/) in R (Ver. 3.6.0) for further analysis. Besides, 84 ccRCC patients with full data from the International Cancer Genome Consortium (ICGC, https://icgc.org/) were used as external validation. The OS (or time to death), defined as the time from the start of follow-up (surgery) to death of any cause, was regarded as the target event.

Screening for survival-related genes

In TCGA, 52 patients with OS < 1 year and 149 patients with OS > 5 years were enrolled in the differential expression analysis via the limma package in R to screen out survival-related differentially expressed genes (DEGs) and acquire corresponding fold changes (FCs). Specifically, DEGs with $|\log_2 FC| > 0.60$ and false discovery rate (FDR) < 0.05 were considered to be the hub DEGs. Outcomes were visualized as volcano plot and heatmap using 'ggplot2' and 'pheatmap' packages in R, respectively.

Functional enrichment analysis

After ruling out the non-coding RNAs, the rest DEGs with protein-coding functions were analyzed by the Database for Annotation Visualization and Integrated Discovery (DAVID; https://david.ncifcrf.gov/) online tools to obtain their Gene Oncology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment information. The criteria were set as $P<0.05$ and gene count ≥ 3.

Prognostic risk signature development

The hub DEGs with protein-coding functions were selected as candidates for the analysis of this section. In the training set, the univariable Cox regression analysis using 'survival' package in R was primarily performed to filter insignificant candidates ($P>0.05$). Next, the least absolute shrinkage and selection operator (LASSO) Cox regression method [21] using 'glmnet' and 'survival' packages was performed to select the optimal panel of genes included in the risk score formula. Last, the multivariate Cox regression analysis was used to obtain the coefficients of each included gene. The risk score of each patient was equal to the sum of the products of each gene's expression value (Exp_i) and the corresponding coefficients. Using the median score as the cut-off point, ccRCC patients were divided into a low-risk group and high-risk group. The Kaplan–Meier survival analysis with a log-rank test, and the area under the curve (AUC) of the receiver operating characteristic (ROC) curve were used as an initial evaluation of the risk signature.

Construction and validation of nomogram for OS prediction

To identify independent prognostic factors to OS, parameters including age, gender, TNM stage, history of prior malignancy, and the aforesaid risk score were included in univariate and multivariate Cox regression analyses in the training set. Using the 'rms' package in R, a nomogram incorporating all the significant factors ($P<0.05$) was constructed

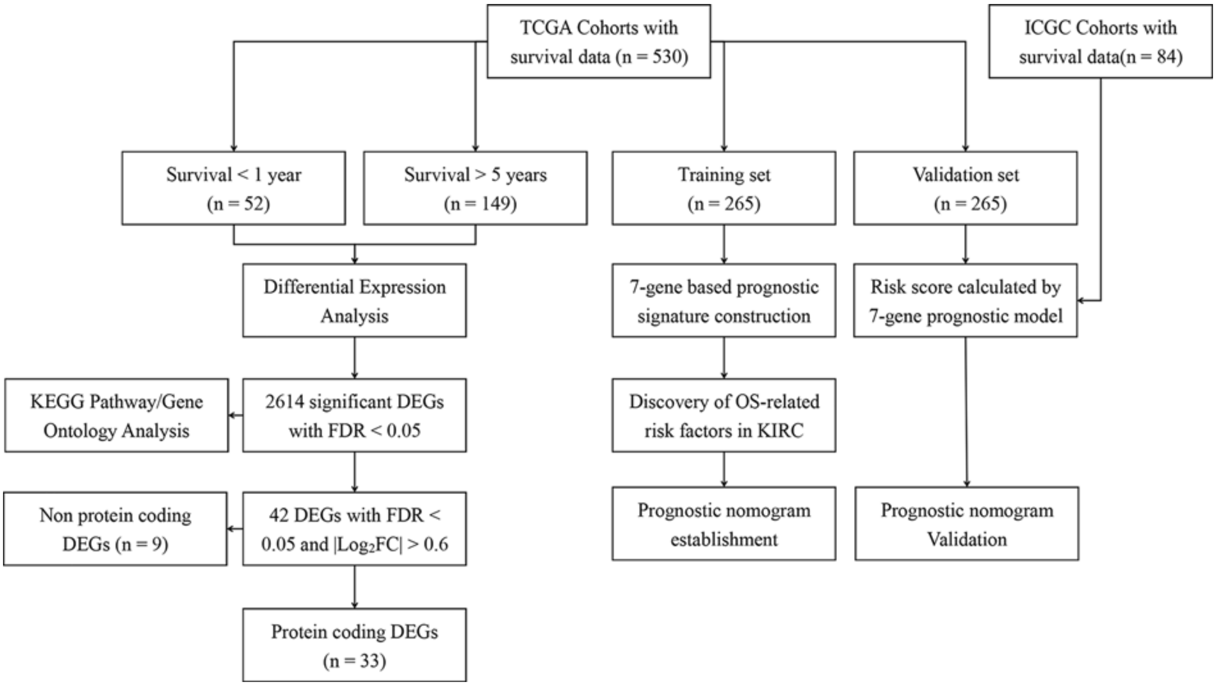
TCGA Cohorts with survival data (n = 530)
ICGC Cohorts with survival data(n = 84)
Survival < 1 year (n = 52)
Survival > 5 years (n = 149)
Training set (n = 265)
Validation set (n = 265)
Differential Expression Analysis
7-gene based prognostic signature construction
Risk score calculated by 7-gene prognostic model
KEGG Pathway/Gene Ontology Analysis
2614 significant DEGs with FDR < 0.05
Discovery of OS-related risk factors in KIRC
Non protein coding DEGs (n = 9)
42 DEGs with FDR < 0.05 and |Log2FC| > 0.6
Prognostic nomogram establishment
Prognostic nomogram Validation
Protein coding DEGs (n = 33)

Figure 1. Flow chart of study design

Abbreviation: KIRC, kidney renal clear cell carcinoma.

to predict the 3- and 5-years OS. For predictive performance assessment, Harrell's concordance index (C-index) and calibration plot were obtained in training and validation sets. Similar to the AUC of ROC curve, C-index using a bootstrap method with 1000 resamplings was calculated to assess the discriminatory ability of nomogram [22,23]. The calibration plot compared the observed and predicted probabilities, and the 45-degree line represents the highest predictive ability.

Results

Patient characteristics

Figure 1 displayed the flow chart of this work. In all, 530 and 84 patients with full expression and clinical data were collected from TCGA and ICGC, respectively. The baseline characteristics of patients in the training set, validation set, and ICGC cohort are collected in Table 1.

DEGs screening and functional enrichment analysis

By comparing 149 samples with OS $>$ 5 years with 52 samples with OS $<$ 1 year, 614 DEGs with criteria set as FDR < 0.05 were identified. GO and KEGG pathway enrichment analyses revealed the functions of these genes. The top 15 significantly enriched GO terms were gathered in Figure 2A and Supplementary Table S1, indicating that DEGs associated with pivotal terms such as the 'oxidation-reduction process' (GO category: biological process), 'cytoplasm' (GO category: cellular component), and 'protein binding' (GO category: molecular function). The top 15 significantly enriched pathways from the KEGG analysis were collected in Figure 2B and Supplementary Table S2, showing that DEGs participated mainly in pathways such as 'valine, leucine and isoleucine degradation', 'fatty acid metabolism', 'PPAR signaling pathway', 'glycolysis/gluconeogenesis', and 'tryptophan metabolism'.

To narrow down the range, 42 hub DEGs with criteria set as FDR < 0.05 and $|\log_2\text{FC}| > 0.60$ were selected. As presented in Supplementary Table S3, ccRCC patients with longer OS were associated with 15 down- and 27 up-regulated hub DEGs. Furthermore, 9 non-coding genes were excluded, leaving 33 candidates for further analysis (volcano plot and heatmap were presented in Figure 2C,D, respectively).

Table 1 Clinical characteristics of ccRCC patients in the TCGA and ICGC datasets

Variables	TCGA cohort (*n*=530)		ICGC cohort (*n*=84) *N* (%)
	Training set (*n*=265) *N* (%)	Validation set (*n*=265) *N* (%)	
Status			
Alive	175 (66.04)	182 (68.68)	56 (66.67)
Dead	90 (33.96)	83 (31.32)	28 (33.33)
Age (years)	60.72 ± 12.84	60.40 ± 11.41	60.86 ± 9.68
Gender			
Male	166 (62.64)	178 (67.17)	45 (53.57)
Female	99 (37.36)	87 (32.83)	39 (46.43)
Stage			
I	133 (50.19)	132 (49.81)	48 (57.14)
II	28 (10.57)	29 (10.94)	12 (14.29)
III	61 (23.02)	62 (23.40)	15 (17.86)
IV	43 (16.22)	39 (14.72)	9 (10.71)
NA	0 (0.00)	3 (1.13)	0 (0.00)
Prior Malignancy			
Yes	37 (13.96)	35 (13.21)	NA
No	228 (86.04)	230 (86.79)	NA

Abbreviation: NA, not available.

Table 2 Outcomes of the multivariate Cox regression analysis of the seven genes identified by the LASSO-penalized model

Genes	Coefficient	HR (95% CI)	*P*-value
CYP3A7	−0.52	0.60 (0.38–0.94)	0.03
CNTNAP5	−0.47	0.62 (0.42–0.92)	0.02
ADCY2	−0.31	0.73 (0.51–1.05)	0.09
TOX3	−0.25	0.78 (0.62–0.98)	0.03
PLG	−0.16	0.85 (0.71–1.02)	0.08
ENAM	0.35	1.42 (1.06–1.91)	0.02
COL7A1	0.61	1.84 (1.46–2.33)	2.65E-07

Abbreviations: CI, confidence interval; HR, hazard ratio.

Seven-gene signature development

Based on univariate analysis, all 33 hub DEGs were significantly associated with ccRCC patients' OS in the training set ($P<0.05$, Supplementary Table S4). Nine genes, including collagen type VII α 1 chain (COL7A1), plasminogen (PLG), inositol-trisphosphate 3-kinase A (ITPKA), adenylate cyclase 2 (ADCY2), solute carrier family 16 member 12 (SLC16A12), cytochrome P450 family 3 subfamily A member 7 (CYP3A7), TOX high mobility group box family member 3 (TOX3), contactin-associated protein family member 5 (CNTNAP5), and enamelin (ENAM), were identified as the most effective combination with the least components by LASSO-penalized Cox analysis (Figure 3A,B). Two genes (ITPKA and SLC16A12) were excluded based on the multivariate Cox regression model, and thus, a seven-gene prognostic signature was finally established (Figure 3C and Table 2). The risk score was calculated as follows:

$$\text{Risk score} = -0.52 \times \text{Exp(CYP3A7)} - 0.47 \times \text{Exp(CNTNAP5)} - 0.31 \times \text{Exp(ADCY2)}$$
$$- 0.25 \times \text{Exp(TOX3)} - 0.16 \times \text{Exp(PLG)} + 0.35 \times \text{Exp(ENAM)} + 0.61 \times \text{Exp(COL7A1)}$$

A higher risk score predicted worse survival. The distribution of risk scores and survival status of patients in the training set was exhibited in Figure 4A,B, respectively. Using the median score as the cut-off value, patients were classified into low-risk and high-risk groups. The Kaplan–Meier curves confirmed significantly better survival for low-risk groups than their high-risk counterparts (log-rank tests $P<0.05$, Figure 4C). Moreover, this advantage remained stable in both stage I/II and III/IV subgroups (Figure 4D,E). The ROC curves were plotted to assess the prognostic value of the seven-gene signature. The AUCs for 3- and 5-year OS predictions in the training set were 0.76 and 0.81 (Figure

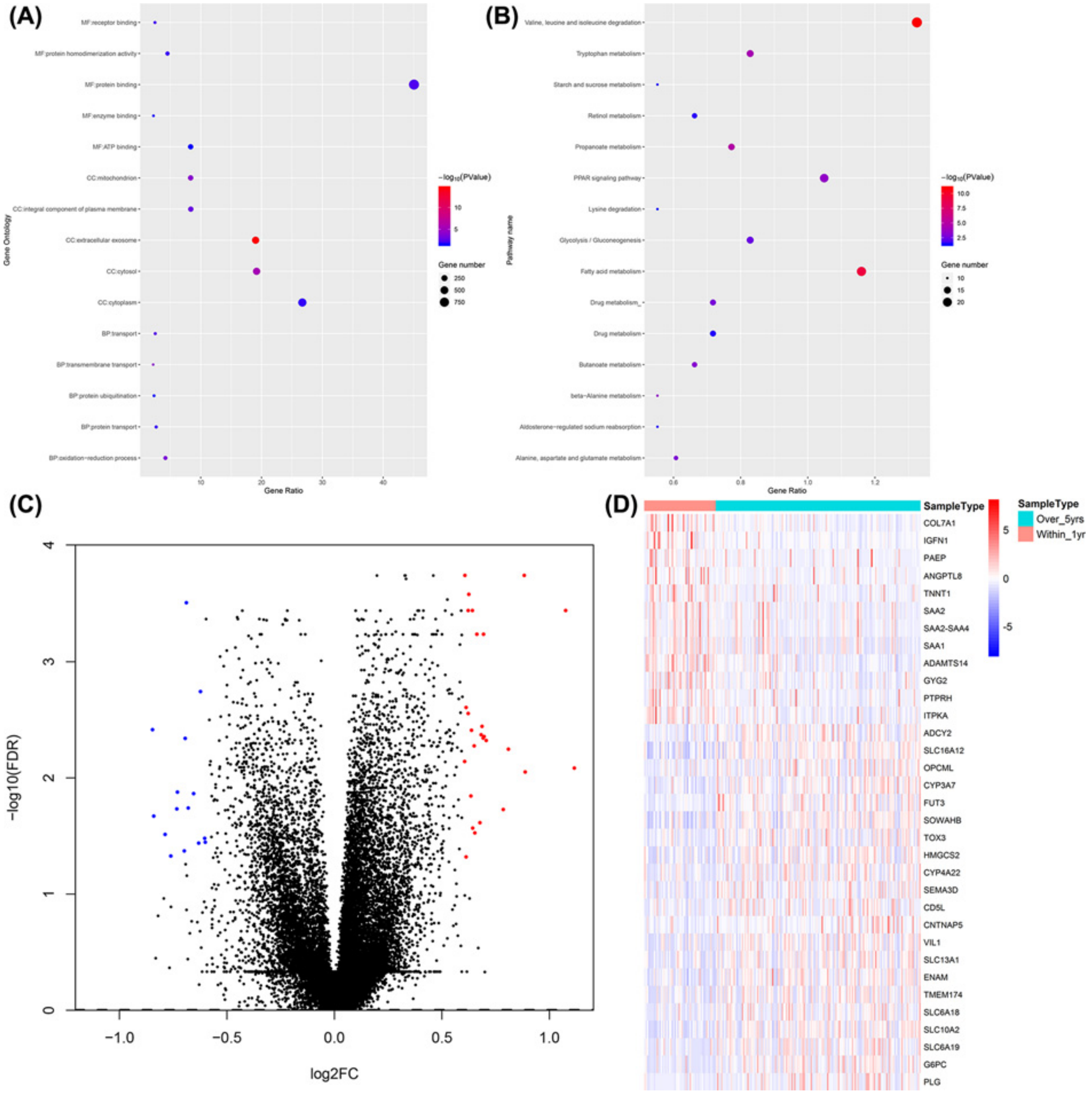

Figure 2. Identification and function enrichment analyses of the survival-related DEGs in the TCGA ccRCC cohort

(**A**) Top 15 enriched GO terms of DEGs. (**B**) Top 15 enriched KEGG pathways of DEGs. (**C**) Volcano plot of DEGs: the abscissa represents |log2FC| and the ordinate represents −log10(FDR). The blue and red spots represent significantly down-regulated and up-regulated hub DEGs, respectively. (**D**) Cluster heatmap of the 33 hub DEGs.

4F,G). At the discovery stage, the preliminary result indicated the seven-gene signature achieved good performance in predicting OS using training set data.

Construction and validation of the nomogram

In the training set, the univariate analysis indicated that age, TNM stage, and the seven-gene risk score impacted significantly on OS, whereas the gender and history of prior malignancy were found to be insignificant parameters (Table 3). The following multivariate analysis confirmed they were independent risk factors to OS of ccRCC patients (*P*-value for age, stage, and risk score were 1.91E-04, 1.49E-07, and 8.01E-08, respectively). The hazard ratios with 95% confidence intervals of age (elder versus young), stage (III/IV versus I/II), and risk score (high versus low) were 1.04 (1.02–1.06), 3.46 (2.18–5.49), and 1.15 (1.09–1.21), respectively. Subsequently, a nomogram predicting 3- and 5

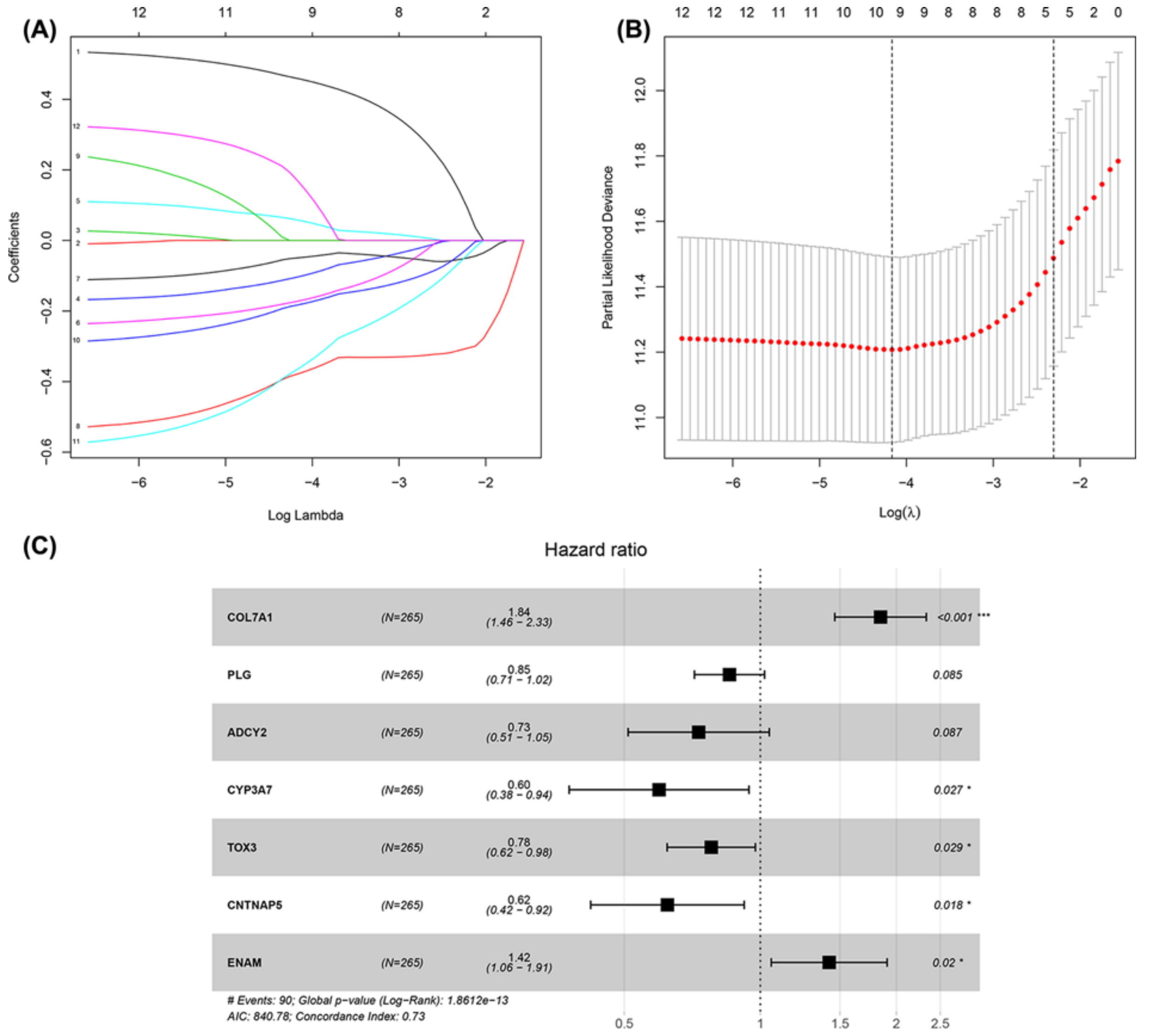

Figure 3. Risk score formula construction based on a seven-gene signature

(**A**) The LASSO coefficient profiles of the 33 hub DEGs selected by Univariate Cox regression analysis. (**B**) Partial likelihood deviance for the LASSO coefficient profiles. (**C**) Forest plot based on Multivariate Cox regression results displays the HRs with corresponding 95% CIs of the seven genes selected by the LASSO model.

Table 3 Univariate and multivariate analyses of OS in the training set

Variables	Univariate analysis		Multivariate analysis	
	HR (95% CI)	***P*-value**	**HR (95% CI)**	***P*-value**
Age (years)	1.03 (1.02–1.05)	1.44E-04	1.04 (1.02–1.06)	1.91E-04
Gender				
Female	1			
Male	0.87 (0.57–1.33)	0.52	0.80 (0.52–1.24)	0.32
Stage				
I/II	1			
III/IV	3.72 (2.41–5.75)	3.46E-09	3.46 (2.18–5.49)	1.49E-07
Prior malignancy				
No	1			
Yes	0.85 (0.46–1.55)	0.59	1.01 (0.54–1.90)	0.98
Risk score	1.21 (1.15–1.26)	3.22E-15	1.15 (1.09–1.21)	8.01E-08

Abbreviations: CI, confidence interval; HR, hazard ratio.

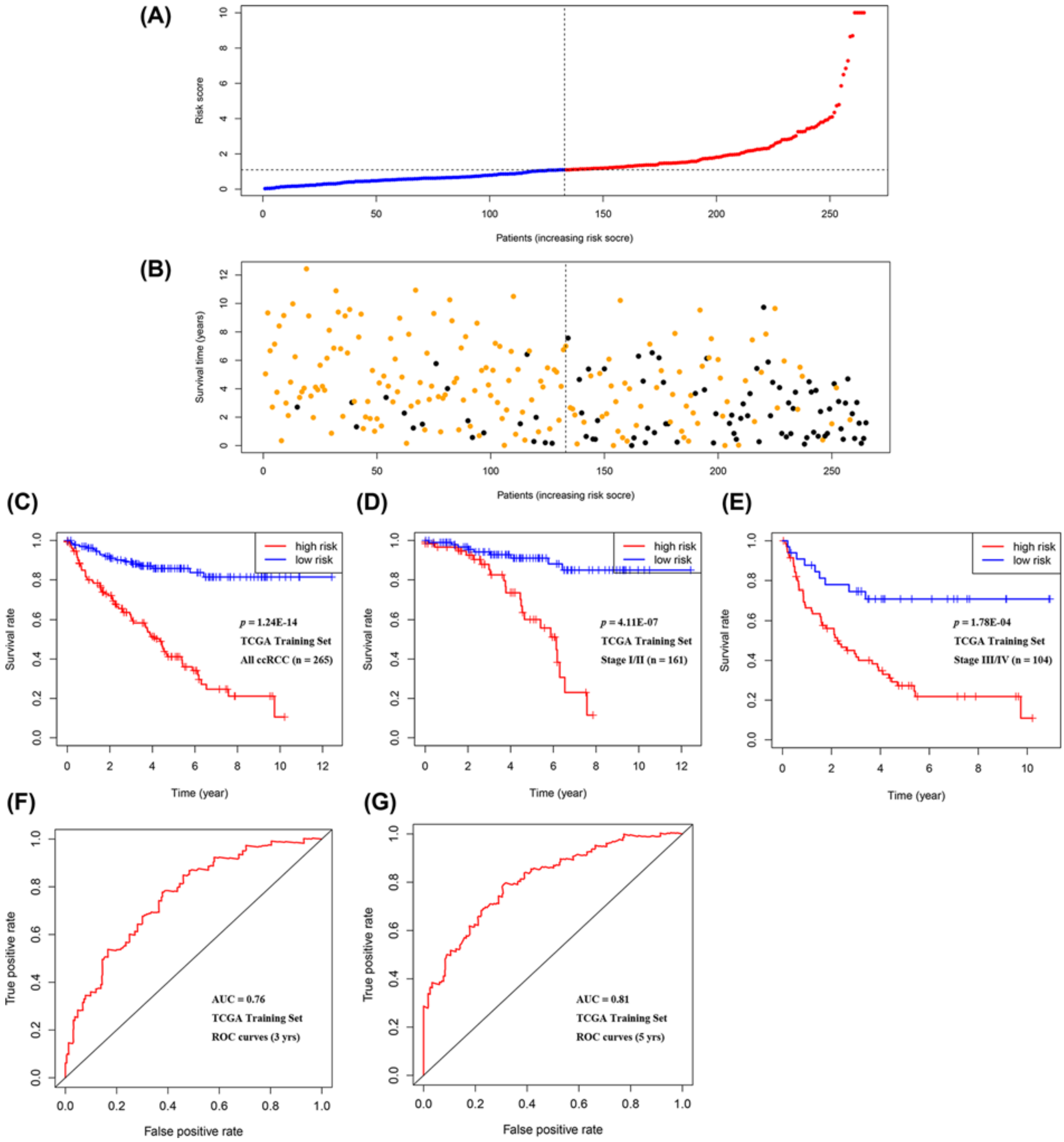

Figure 4. Preliminary evaluation of the predictive ability of the seven-gene signature in the training set

(**A**) The seven-gene-based risk score distribution: using the median risk score as a cut-off point, patients were divided into a low-risk group (blue spots) and high-risk group (red spots). (**B**) The vital status of 265 patients: yellow and black spots represent alive and dead patients, respectively. (**C**–**E**) K–M survival curves of all patients' OS (*n*=265), Stage I/II patients' OS (*n*=161), Stage III/IV patients' OS (*n*=104), respectively. (**F**,**G**) ROC curves for OS prediction based on the seven-gene signature within 3- and 5-years, respectively.

years OS of ccRCC patients was constructed according to the multivariate analysis results of the training set (Figure 5A). The C-index for OS prediction of the nomogram was 0.78 (95% CI: 0.74–0.82). Internal validation using data from validation set revealed that a C-index of 0.75 (95% CI: 0.70–0.80). For external validation, C-index calculated using ICGC data was 0.70 (95% CI: 0.60–0.80). Besides, the calibration plots displaying the probability of 3- and 5

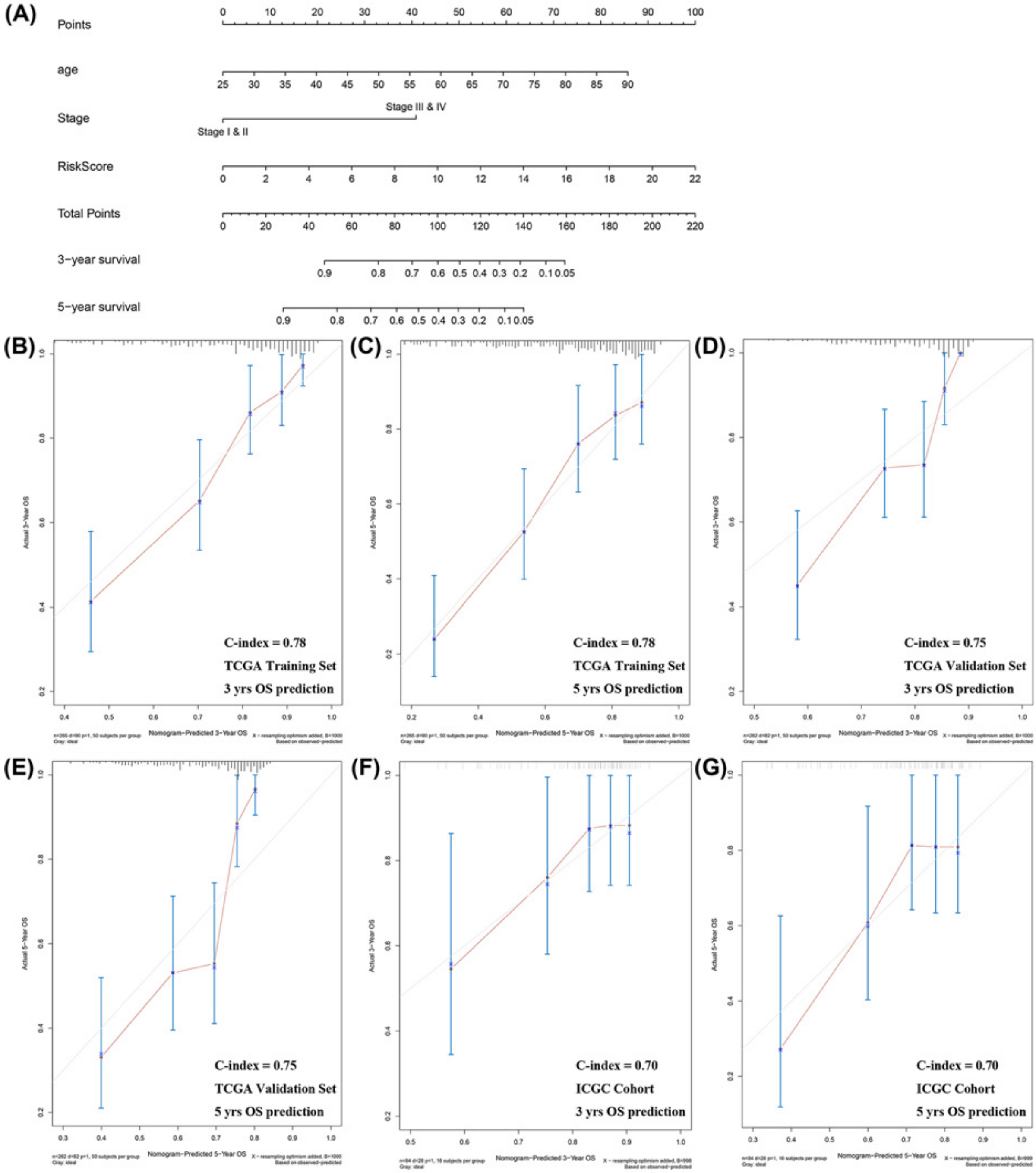

Figure 5. The establishment and assessment of a novel nomogram

(**A**) A nomogram integrating clinical features with a seven-gene risk score for predicting of 3- and 5- years OS in patients with ccRCC. Calibration plots of the nomogram for 3- and 5- years OS prediction in the training set (**B**,**C**), internal validation set (**D**,**E**), and ICGC cohort (**F**,**G**), respectively. The abscissa represents the nomogram-predicted survival probability and the ordinate represents the actual survival.

years survival indicated favorable curve-fitting between the nomogram-predicted outcomes and actual observation in the training set, validation set, and ICGC cohort, respectively (Figure 5B–G).

Table 4 Comparison of predicting performance with other reported prognostic tools

Study	Source	Stage	Size	Gene signature	Clinical feature in nomogram	C-index (95% CI)
Xiong et al., 2020	Hospital	All	101	IGPI consisting 17 gene pairs	Histologic grade, TNM stage	0.76
				-	UISS risk model only	0.72
Wu et al., 2019	TCGA	III	122	ATP6V1C2, PCSK1N, PREX1, ANK3, HLA-DRA, SELENBP1, TYRP1, GABRA2, SERPINA5	Age, ISUP grade, pN stage	0.79 (0.75–0.84)
Qu et al., 2018	TCGA	I–III	444	ENSG00000255774, ENSG00000248323, ENSG00000260911, ENSG00000231666	TNM stage	0.73 (0.65–0.81)
Develasco et al., 2017	TCGA	IV	54	ClearCode34	-	0.63 (0.51–0.75)
Tang et al., 2019	Hospital	All	140	-	TNM stage only	0.65 (0.56–0.74)
				-	Fuhrman grade only	0.61 (0.52–0.70)
The present study	TCGA	All	530	CYP3A7, CNTNAP5, ADCY2, TOX3, PLG, ENAM, COL7A1	Age, TNM stage	0.78 (0.74–0.82)

Abbreviations: ClearCode34, a 34-gene signature model; IGPI, immune-related gene pair index.

Comparison with previously reported prognostic tools

The predictive performance of our nomogram was compared with several reported prognostic tools, which were retrieved from PubMed database using '*(overall survival) AND (((c-index) AND signature) AND ((((clear cell renal cell carcinoma) OR renal clear cell carcinoma) OR clear cell carcinoma) OR KIRC))*' as search terms. As presented in Table 4, clinical features such as TNM stage [12], Fuhrman grade [12], and the UISS risk model [24] alone seemed to be less competitive in terms of discrimination (c-indices were 0.65, 0.61, and 0.72, respectively). Similarly, using only gene signature such as the ClearCode34, a 34-gene signature model, was hardly satisfying when predicting OS in stage IV ccRCC patients [25]. The c-index of our nomogram was 0.78, second only to a 9-gene-signature-based nomogram reported by Wu et al. [26], which focused solely on stage III ccRCC (c-index: 0.79). Xiong et al. [24] reported a tool combining IGPI (immune-related 17 gene pairs index) with histologic grade and TNM stage for all ccRCC patients, with a c-index of 0.76. For localized ccRCC (stage I–III), Qu et al. [27] reported a tool combining four lncRNAs with the TNM stage, with a c-index of 0.73. In summary, by comprising only seven genes and two clinical features, our nomogram was economic and applicable to all stages of ccRCC without compromising the prognostic ability.

Discussion

In this work, we developed an OS-related seven-gene signature in ccRCC, namely CYP3A7, CNTNAP5, ADCY2, TOX3, PLG, ENAM, and COL7A1, from TCGA training set. The ensuing univariate and multivariate Cox regression indicated that the patient's age, TNM stage, and the seven-gene risk score were independent prognostic factors to OS and a nomogram was then constructed. Subsequently, C-indices and the curve-fitting calibration plots of the training set, internal validation set, and ICGC ccRCC cohort demonstrated the decent predictive performance of the nomogram.

To the best of our knowledge, this is the first study that revealed the putative protective role of CYP3A7 in the prognosis of ccRCC. Members of the cytochrome P450 superfamily are a group of metalloproteins that involve in metabolic biotransformation of endogenous and exogenous substrates, including carcinogens [28]. Conversely, overexpression of CYP3A7 was witnessed in hepatocellular carcinoma [29,30], suggesting that it might exert varied functions among different types or stages of tumor. Belonging to the neurexin family, the product of CNTNAP5 functioned as cell adhesion molecules in the nervous system and participated in diseases such as autism, Alzheimer's disease, and schizophrenia [31–33]. The expression level of CNTNAP5 in the kidney is relatively low but still detectable, and only one study reported a SH3KBP1–CNTNAP5 fusion in upper tract urothelial carcinoma [34]. A higher level of ADCY2 was shown to connect to longer OS in our study, but few studies reported on its role in tumor progression. Reportedly, ADCY suppressed migration and invasion of pancreatic tumor cells by increasing the level of second messenger cyclic adenosine monophosphate (cAMP), however, ADCY2 was found to be down-regulated in pancreatic tumor tissues [35]. TOX3 has been newly identified as a ccRCC suppressor gene as it inhibited tumor cell migration and invasion by repressing the SNAIL members SNAI1 and SNAI2 at the transcriptional level [36]. This was consistent with our results that TOX3 exerted protective influence in ccRCC. A lower level of PLG in ccRCC patients with shorter survival, higher stages and grades were described in several studies [37–39], consistent with our finding that PLG served as a protective factor. Our outcomes illustrated that patients with OS > 5 years had significantly up-regulated ENAM when compared with those with OS < 1 year. Similarly, Bhalla et al. [40] reported a lower expression of ENAM in late-stage ccRCC when compared with those in the early stage. When extending to patients with all lengths of OS in the training set, however, multivariate Cox regression yielded the opposite conclusion that ENAM was a risk factor to the OS of ccRCC. Thus, ENAM may have a more complex role in the progression of ccRCC. Coding for type IV collagen, COL7A1 was also a risk factor to ccRCC survival according to our results, supported by a previous study revealing high expression of COL7A1 was associated with tumor invasion and shorter survival in several types of squamous cell cancer [41–43].

Prevalence of targeted and individualized therapy calls for novel prognostic tools integrating genetic signatures with clinical features to improve risk assessment and stratification. The comparison of c-indices revealed that the present nomogram based on an OS-related seven-gene risk score, age, and TNM stage has better predictive accuracy than the traditional prognostic tools such as the TNM staging system, Fuhrman grading system, UISS risk model, ClearCode34. Furthermore, the present nomogram had undergone both internal and external validation and showed good reproducibility. In terms of clinical significance, the present work suggested that patients with higher points calculated according to our nomogram might benefit from more active surveillance as well as adjuvant treatments such as tyrosine kinase inhibitors [44] and immunotherapy [45].

Limitations of the present work should be notified. The first is the retrospective design of the present study. Second, the seven genes were less reported in ccRCC. Third, the external validation cohort in the present study comprising merely 84 samples. Hence, further experiments are warranted to elucidate the roles of the seven genes in ccRCC development and progression and to validate the predictive ability of the nomogram in prospective studies with a larger population.

Conclusions

In the present study, we excavated seven novel OS-related genes (CYP3A7, CNTNAP5, ADCY2, TOX3, PLG, ENAM, and COL7A1) from TCGA and used them to build a formula for risk score calculation. Besides, by integrating the seven-gene signature and clinical features (age and TNM stage), we proposed and validated a nomogram for OS prediction in ccRCC which might have promising application prospects.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China [grant numbers 1670643, 81870483, 81802821]; the Collaborative Innovation Project of Guangzhou Education Bureau [grant number 1201620011]; the Guangzhou Science Technology and Innovation Commission [grant number 201704020193]; and the Science and Technology Planning Project of Guangdong Province [grant number 2017B030314108].

Author Contribution

Conception and design: T.D., Z.H. Acquisition of data: Z.H. Analysis and interpretation of data: Z.H., T.D. Manuscript drafting: Z.H. Manuscript revision: T.D., X.D., G.Z. All authors read and approved the final version of the manuscript.

Abbreviations

ADCY2, adenylate cyclase 2; AUC, area under the curve; ccRCC, clear cell renal cell carcinoma; CNTNAP5, contactin-associated protein family member 5; COL7A1, collagen type VII α 1 chain; CYP3A7, cytochrome P450 family 3 subfamily A member 7; C-index, Harrell's concordance index; DEG, differentially expressed gene; ENAM, enamelin; FC, fold change; FDR, false discovery rate; GO, Gene Oncology; ICGC, International Cancer Genome Consortium; ITPKA, inositol-trisphosphate 3-kinase A; KEGG, Kyoto Encyclopedia of Genes and Genomes; LASSO, least absolute shrinkage and selection operator; OS, overall survival; PLG, plasminogen; RCC, renal cell carcinoma; ROC, receiver operating characteristic; SLC16A12, solute carrier family 16 member 12; TCGA, The Cancer Genome Atlas; TNM, Tumor, lymph Node, and Metastasis staging system; TOX3, TOX high mobility group box family member 3; UISS, University of California Integrated Staging System.

References

1 Ferlay, J., Shin, H.R., Bray, F., Forman, D., Mathers, C. and Parkin, D.M. (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer* **127**, 2893–2917, https://doi.org/10.1002/ijc.25516

2 Rini, B.I., Campbell, S.C. and Escudier, B. (2009) Renal cell carcinoma. *Lancet* **373**, 1119–1132, https://doi.org/10.1016/S0140-6736(09)60229-4

3 Siegel, R.L., Miller, K.D. and Jemal, A. (2018) Cancer statistics, 2018. *CA Cancer J. Clin.* **68**, 7–30

4 Malouf, G.G., Zhang, J., Yuan, Y., Comperat, E., Roupret, M., Cussenot, O. et al. (2015) Characterization of long non-coding RNA transcriptome in clear-cell renal cell carcinoma by next-generation deep sequencing. *Mol. Oncol.* **9**, 32–43, https://doi.org/10.1016/j.molonc.2014.07.007

5 Finelli, A., Ismaila, N., Bro, B., Durack, J., Eggener, S., Evans, A. et al. (2017) Management of small renal masses: American Society of Clinical Oncology Clinical Practice Guideline. *J. Clin. Oncol.* **35**, 668–680, https://doi.org/10.1200/JCO.2016.69.9645

6 Gupta, K., Miller, J.D., Li, J.Z., Russell, M.W. and Charbonneau, C. (2008) Epidemiologic and socioeconomic burden of metastatic renal cell carcinoma (mRCC): a literature review. *Cancer Treat. Rev.* **34**, 193–205, https://doi.org/10.1016/j.ctrv.2007.12.001

7 Linehan, W.M. and Zbar, B. (2004) Focus on kidney cancer. *Cancer Cell* **6**, 223–228, https://doi.org/10.1016/j.ccr.2004.09.006

8 Yao, M., Tabuchi, H., Nagashima, Y., Baba, M., Nakaigawa, N., Ishiguro, H. et al. (2005) Gene expression analysis of renal carcinoma: adipose differentiation-related protein as a potential diagnostic and prognostic biomarker for clear-cell renal carcinoma. *J. Pathol.* **205**, 377–387, https://doi.org/10.1002/path.1693

9 Ljungberg, B., Bensalah, K., Canfield, S., Dabestani, S., Hofmann, F., Hora, M. et al. (2015) EAU guidelines on renal cell carcinoma: 2014 update. *Eur. Urol.* **67**, 913–924, https://doi.org/10.1016/j.eururo.2015.01.005

10 Rioux-Leclercq, N., Karakiewicz, P.I., Trinh, Q.D., Ficarra, V., Cindolo, L., de la Taille, A. et al. (2007) Prognostic ability of simplified nuclear grading of renal cell carcinoma. *Cancer* **109**, 868–874, https://doi.org/10.1002/cncr.22463

11 Leibovich, B.C., Cheville, J.C., Lohse, C.M., Zincke, H., Frank, I., Kwon, E.D. et al. (2005) A scoring algorithm to predict survival for patients with metastatic clear cell renal cell carcinoma: a stratification tool for prospective clinical trials. *J. Urol.* **174**, 1759–1763, discussion 63, https://doi.org/10.1097/01.ju.0000177487.64651.3a

12 Tang, M., Cao, X., Li, Y., Li, G.Q., He, Q.H., Li, S.J. et al. (2019) High expression of herpes virus entry mediator is associated with poor prognosis in clear cell renal cell carcinoma. *Am. J. Cancer Res.* **9**, 975–987

13 Ha, M., Son, Y.R., Kim, J., Park, S.M., Hong, C.M., Choi, D. et al. (2019) TEK is a novel prognostic marker for clear cell renal cell carcinoma. *Eur. Rev. Med. Pharmacol. Sci.* **23**, 1451–1458

14 Ha, M., Jeong, H., Roh, J.S., Lee, B., Lee, D., Han, M.E. et al. (2019) VNN3 is a potential novel biomarker for predicting prognosis in clear cell renal cell carcinoma. *Anim. Cells Syst.* **23**, 112–117, https://doi.org/10.1080/19768354.2019.1583126

15 Luo, Y., Shen, D., Chen, L., Wang, G., Liu, X., Qian, K. et al. (2019) Identification of 9 key genes and small molecule drugs in clear cell renal cell carcinoma. *Aging* **11**, 6029–6052, https://doi.org/10.18632/aging.102161

16 Pan, Q., Wang, L., Zhang, H., Liang, C. and Li, B. (2019) Identification of a 5-gene signature predicting progression and prognosis of clear cell renal cell carcinoma. *Med. Sci. Monit.* **25**, 4401–4413, https://doi.org/10.12659/MSM.917399

17 Zeng, M.H., Qiu, J.G., Xu, Y. and Zhang, X.H. (2019) IDUA, NDST1, SAP30L, CRYBA4, and SI as novel prognostic signatures clear cell renal cell carcinoma. *J. Cell. Physiol.*, https://doi.org/10.1002/jcp.28297

18 Chen, Y., Jiang, S., Lu, Z., Xue, D., Xia, L., Lu, J. et al. (2020) Development and verification of a nomogram for prediction of recurrence-free survival in clear cell renal cell carcinoma. *J. Cell. Mol. Med.* **24**, 1245–1255, https://doi.org/10.1111/jcmm.14748

19 Jiang, W., Guo, Q., Wang, C. and Zhu, Y. (2019) A nomogram based on 9-lncRNAs signature for improving prognostic prediction of clear cell renal cell carcinoma. *Cancer Cell Int.* **19**, 208, https://doi.org/10.1186/s12935-019-0928-5

20 Zhang, C., He, H., Hu, X., Liu, A., Huang, D., Xu, Y. et al. (2019) Development and validation of a metastasis-associated prognostic signature based on single-cell RNA-seq in clear cell renal cell carcinoma. *Aging* **11**, 10183–10202, https://doi.org/10.18632/aging.102434

21 Fontanarosa, J.B. and Dai, Y. (2011) Using LASSO regression to detect predictive aggregate effects in genetic studies. *BMC Proc.* **5**, S69, https://doi.org/10.1186/1753-6561-5-S9-S69

22 Harrell, Jr, F.E., Califf, R.M., Pryor, D.B., Lee, K.L. and Rosati, R.A. (1982) Evaluating the yield of medical tests. *JAMA* **247**, 2543–2546, https://doi.org/10.1001/jama.1982.03320430047030

23 Pencina, M.J. and D'Agostino, R.B. (2004) Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat. Med.* **23**, 2109–2123, https://doi.org/10.1002/sim.1802

24 Xiong, Y., Liu, L., Bai, Q., Xia, Y., Qu, Y., Wang, J. et al. (2020) Individualized immune-related gene signature predicts immune status and oncologic outcomes in clear cell renal cell carcinoma patients. *Urol. Oncol.* **38**, 7.e1–7.e8, https://doi.org/10.1016/j.urolonc.2019.09.014

25 de Velasco, G., Culhane, A.C., Fay, A.P., Hakimi, A.A., Voss, M.H., Tannir, N.M. et al. (2017) Molecular subtypes improve prognostic value of international metastatic renal cell carcinoma database consortium prognostic model. *Oncologist* **22**, 286–292, https://doi.org/10.1634/theoncologist.2016-0078

26 Wu, J., Jin, S., Gu, W., Wan, F., Zhang, H., Shi, G. et al. (2019) Construction and validation of a 9-gene signature for predicting prognosis in stage III clear cell renal cell carcinoma. *Front. Oncol.* **9**, 152, https://doi.org/10.3389/fonc.2019.00152

27 Qu, L., Wang, Z.L., Chen, Q., Li, Y.M., He, H.W., Hsieh, J.J. et al. (2018) Prognostic value of a long non-coding RNA signature in localized clear cell renal cell carcinoma. *Eur. Urol.* **74**, 756–763, https://doi.org/10.1016/j.eururo.2018.07.032

28 Eun, H.S., Cho, S.Y., Lee, B.S., Seong, I.O. and Kim, K.H. (2018) Profiling cytochrome P450 family 4 gene expression in human hepatocellular carcinoma. *Mol. Med. Rep.* **18**, 4865–4876

29 Kondoh, N., Wakatsuki, T., Ryo, A., Hada, A., Aihara, T., Horiuchi, S. et al. (1999) Identification and characterization of genes associated with human hepatocellular carcinogenesis. *Cancer Res.* **59**, 4990–4996

30 Neunzig, I., Dragan, C.A., Widjaja, M., Schwaninger, A.E., Peters, F.T., Maurer, H.H. et al. (2011) Whole-cell biotransformation assay for investigation of the human drug metabolizing enzyme CYP3A7. *Biochim. Biophys. Acta* **1814**, 161–167, https://doi.org/10.1016/j.bbapap.2010.07.011

31 Pagnamenta, A.T., Bacchelli, E., de Jonge, M.V., Mirza, G., Scerri, T.S., Minopoli, F. et al. (2010) Characterization of a family with rare deletions in CNTNAP5 and DOCK4 suggests novel risk loci for autism and dyslexia. *Biol. Psychiatry* **68**, 320–328, https://doi.org/10.1016/j.biopsych.2010.02.002

32 Schott, J.M., Crutch, S.J., Carrasquillo, M.M., Uphill, J., Shakespeare, T.J., Ryan, N.S. et al. (2016) Genetic risk factors for the posterior cortical atrophy variant of Alzheimer's disease. *Alzheimers Dement.* **12**, 862–871, https://doi.org/10.1016/j.jalz.2016.01.010

33 Yu, H., Yan, H., Wang, L., Li, J., Tan, L., Deng, W. et al. (2018) Five novel loci associated with antipsychotic treatment response in patients with schizophrenia: a genome-wide association study. *Lancet Psychiatry* **5**, 327–338, https://doi.org/10.1016/S2215-0366(18)30049-X

34 Moss, T.J., Qi, Y., Xi, L., Peng, B., Kim, T.B., Ezzedine, N.E. et al. (2017) Comprehensive genomic characterization of upper tract urothelial carcinoma. *Eur. Urol.* **72**, 641–649, https://doi.org/10.1016/j.eururo.2017.05.048

35 Quinn, S.N., Graves, S.H., Dains-McGahee, C., Friedman, E.M., Hassan, H., Witkowski, P. et al. (2017) Adenylyl cyclase 3/adenylyl cyclase-associated protein 1 (CAP1) complex mediates the anti-migratory effect of forskolin in pancreatic cancer cells. *Mol. Carcinog.* **56**, 1344–1360, https://doi.org/10.1002/mc.22598

36 Jiang, B., Chen, W., Qin, H., Diao, W., Li, B., Cao, W. et al. (2019) TOX3 inhibits cancer cell migration and invasion via transcriptional regulation of SNAI1 and SNAI2 in clear cell renal cell carcinoma. *Cancer Lett.* **449**, 76–86, https://doi.org/10.1016/j.canlet.2019.02.020

37 Luo, T., Chen, X., Zeng, S., Guan, B., Hu, B., Meng, Y. et al. (2018) Bioinformatic identification of key genes and analysis of prognostic values in clear cell renal cell carcinoma. *Oncol. Lett.* **16**, 1747–1757

38 Nouhaud, F.X., Blanchard, F., Sesboue, R., Flaman, J.M., Sabourin, J.C., Pfister, C. et al. (2018) Clinical relevance of gene copy number variation in metastatic clear cell renal cell carcinoma. *Clin. Genitourin. Cancer* **16**, e795–e805, https://doi.org/10.1016/j.clgc.2018.02.013

39 Schrodter, S., Braun, M., Syring, I., Klumper, N., Deng, M., Schmidt, D. et al. (2016) Identification of the dopamine transporter SLC6A3 as a biomarker for patients with renal cell carcinoma. *Mol. Cancer* **15**, 10, https://doi.org/10.1186/s12943-016-0495-5

40 Bhalla, S., Chaudhary, K., Kumar, R., Sehgal, M., Kaur, H., Sharma, S. et al. (2017) Gene expression-based biomarkers for discriminating early and late stage of clear cell renal cancer. *Sci. Rep.* **7**, 44997, https://doi.org/10.1038/srep44997

41 Kita, Y., Mimori, K., Tanaka, F., Matsumoto, T., Haraguchi, N., Ishikawa, K. et al. (2009) Clinical significance of LAMB3 and COL7A1 mRNA in esophageal squamous cell carcinoma. *Eur. J. Surg. Oncol.* **35**, 52–58, https://doi.org/10.1016/j.ejso.2008.01.025

42 Martins, V.L., Vyas, J.J., Chen, M., Purdie, K., Mein, C.A., South, A.P. et al. (2009) Increased invasive behaviour in cutaneous squamous cell carcinoma with loss of basement-membrane type VII collagen. *J. Cell Sci.* **122**, 1788–1799, https://doi.org/10.1242/jcs.042895

43 Pourreyron, C., Chen, M., McGrath, J.A., Salas-Alanis, J.C., South, A.P. and Leigh, I.M. (2014) High levels of type VII collagen expression in recessive dystrophic epidermolysis bullosa cutaneous squamous cell carcinoma keratinocytes increases PI3K and MAPK signalling, cell migration and invasion. *Br. J. Dermatol.* **170**, 1256–1265, https://doi.org/10.1111/bjd.12715

44 Sun, M., Marconi, L., Eisen, T., Escudier, B., Giles, R.H., Haas, N.B. et al. (2018) Adjuvant vascular endothelial growth factor-targeted therapy in renal cell carcinoma: a systematic review and pooled analysis. *Eur. Urol.* **74**, 611–620, https://doi.org/10.1016/j.eururo.2018.05.002

45 Lenis, A.T., Donin, N.M., Johnson, D.C., Faiena, I., Salmasi, A., Drakaki, A. et al. (2018) Adjuvant therapy for high risk localized kidney cancer: emerging evidence and future clinical trials. *J. Urol.* **199**, 43–52, https://doi.org/10.1016/j.juro.2017.04.092