

Identification and Validation of a Predictive Immune-Related Genes Signature for the Prognosis of Cholangiocarcinoma

Rui Zhang

Department of Hepatobiliary Surgery, The First Affiliated Hospital of Xi'an Jiaotong University, 277 West Yanta Road, Xi'an 710061, Shaanxi, China <https://orcid.org/0000-0002-1727-5524>

Chen Chen

Department of Hepatobiliary Surgery, The First Affiliated Hospital of Xi'an Jiaotong University, 277 West Yanta Road, Xi'an 710061, Shaanxi, China

Qi Li

Department of Hepatobiliary Surgery, The First Affiliated Hospital of Xi'an Jiaotong University, 277 West Yanta Road, Xi'an 710061, Shaanxi, China

Jialu Fu

Department of Hepatobiliary Surgery, The First Affiliated Hospital of Xi'an Jiaotong University, 277 West Yanta Road, Xi'an 710061, Shaanxi, China

Dong Zhang

Department of Hepatobiliary Surgery, The First Affiliated Hospital of Xi'an Jiaotong University, 277 West Yanta Road, Xi'an 710061, Shaanxi, China

Wenzhi Li

Department of Hepatobiliary Surgery, The First Affiliated Hospital of Xi'an Jiaotong University, 277 West Yanta Road, Xi'an 710061, Shaanxi, China

Zhimin Geng (✉ gengzhimin@mail.xjtu.edu.cn)

Department of Hepatobiliary Surgery, The First Affiliated Hospital of Xi'an Jiaotong University, 277 West Yanta Road, Xi'an 710061, Shaanxi, China <https://orcid.org/0000-0003-2645-9808>

Research

Keywords: Cholangiocarcinoma, Immune-related gene, Prognosis, Risk model, Bioinformatics

Posted Date: June 1st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-31517/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Immune-related genes (IRGs) play a crucial role in the initiation and progression of cholangiocarcinoma (CCA). However, immune signatures have rarely been used to predict prognosis of CCA. The aim of this study was to construct a novel model for CCA to predict survival based on IRGs expression data.

Methods: The gene expression profiles and clinical data of CCA patients from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) database were integrated to establish and validate prognostic IRG signatures. Differentially expressed immune-related genes were screened. Univariate and multivariate Cox analysis were performed to identify prognostic IRGs, and the risk model that predicts outcomes was constructed. Furthermore, receiver operating characteristic (ROC) and Kaplan-Meier curve were plotted to examine predictive accuracy of the model, and a nomogram was constructed based on IRGs signature, combining with other clinical characteristics. Finally, CIBERSORT was used to analyze the association of immune cells infiltration with risk score.

Results: We identified that 223 IRGs were significantly dysregulated in patients with CCA, among which five IRGs (AVPR1B, CST4, TDGF1, RAET1E and IL9R) were identified as robust indicators for overall survival (OS), and a prognostic model was built based on the IRGs signature. Meanwhile, patients with high risk had worse OS in training and validation cohort, and the area under the ROC was 0.898 and 0.846, respectively. Nomogram demonstrated that immune risk score contributed much more points than other clinicopathological variables, with a C-index of 0.819 (95% CI, 0.727-0.911). Finally, we found that IRGs signature was positively correlated with the proportion of CD8+ T cells, neurophils and T gamma delta, while negatively with that of CD4+ memory resting T cells.

Conclusions: We established and validated an effective five IRGs-based prediction model for CCA, which could accurately classify patients into groups with low and high risk of poor prognosis.

Background

Cholangiocarcinoma (CCA) remains a highly aggressive malignancy with extremely unsatisfactory prognosis, which ranks as the second most frequently primary liver cancer and its incidence is increasing worldwide[1, 2]. Although surgical resection provides the only opportunity to cure for CCA patients without metastasis, the 5-year survival rate underwent radical resection remains less than 30% owing to metastasis and recurrence[3, 4]. The benefits from adjuvant therapy and targeted therapy appear to be very limited due to the genetic heterogeneity[5]. Hence, prognostic markers are of great significance to clinical decision making and may provide novel insights into underlying mechanisms of CCA. Unfortunately, there are no dependable and effective biomarkers to accurately estimate the prognosis of CCA patients.

An increasing amount of evidence has demonstrated that immune-related genes (IRGs) are closely related to the tumorigenesis and progression of CCA[6–8], and several IRGs have been reported to serve

as promising biomarkers for prognostic evaluation to patients[9–11]. For example, the PD-L1 expression was significantly correlated with TNM stage and overall survival for CCA patients after resection[6, 12]. However, most of prognostic indicators or models only focus on a single IRG, which is not sufficiently rigorous to clinical practice.

Checkpoint inhibitor-based immunotherapy have achieved notable success in the clinical treatment of tumor[13–17]. Although many studies have demonstrated that immunotherapy shows a promising safety and efficacy for patients with CCA, the response rate is disappointing[18–21]. Tumor immune microenvironment (TIME) composed of immune infiltrating cells plays an important role in tumor progression and immunotherapeutic response[22–24]. A recent study by Montal et al. on the molecular classification of CCA identified that immune class had a higher lymphocyte infiltration and overexpression of PD-1/PD-L1, and the patients of the class might become the ideal candidates to receive immunotherapy[25]. These evidences provide a new insight into the role of IRGs and immune cells on identifying high-efficiency biomarkers that predict patients' prognosis and the effectiveness of immunotherapy.

Given the role of immune mechanisms in the pathogenesis of CCA, we conducted a study based on immune genes to establish a predictive model for prognosis of CCA patients. Our study aimed to develop an IRGs signature model to help better predicting survival of CCA patients and reveal potential mechanisms.

Materials And Methods

Data collection and TF genes and IRGs profile mining

We made use of data in the public domain. RNA-seq data from CCA and relevant noncancerous tissues were downloaded from The Cancer Genome Atlas-Cholangiocarcinoma dataset (TCGA-CHOL) (<https://cancergenome.nih.gov/>) and Gene Expression Omnibus dataset (GSE107943) (<https://www.ncbi.nlm.nih.gov/gds/>) [26], the data were normalized by the limma package. The corresponding clinicopathological and survival information were downloaded from open-access resource, which included age, histologic grade, tumor node metastases (TNM) stage, margin status, OS time etc.

IRGs were obtained from the ImmPort database ([https:// www.immport.org/home](https://www.immport.org/home)). The genes of cytokines, cytokine receptors, and those that were related to the signaling pathways of the T-cell receptor, B-cell antigen receptor, natural killer cell cytotoxicity, and antigen processing and presentation were selected. Transcription factor (TF) genes were obtained from the Cistrome Cancer database (<http://cistrome.org/CistromeCancer>).

Identification of differentially expressed TF and IRGs

Differentially expressed genes (DEGs) were detected using Wilcoxon test in limma package. Probe sets without corresponding gene symbols or genes with more than one probe set were removed or averaged. $|\log_2 \text{ fold change}| > 1$ and adjusted $P < 0.05$ were considered as statistically significant. The expression patterns of TF and IRGs were drawn from all DEGs. Volcano plots and heatmaps were plotted respectively by gplots and pheatmap package in R software.

GO and KEGG pathway enrichment analysis

To understand the biological mechanisms underlying IRGs, Gene ontology (GO) annotation and Kyoto encyclopedia of genes and genomes (KEGG) pathway analyses were performed using DAVID (<https://david-d.ncifcrf.gov/>) and ClusterProfiler package in R. GO analysis consists of three categories: biological processes (BP), cellular components (CC) and molecular function (MF). The results were visualized using GO plot package and false discovery rate (FDR) < 0.05 was considered as the threshold.

Development and validation of prognostic IRGs signature for CCA

Univariate and multivariate Cox analysis were utilized to assess the relationship between IRGs expression and overall survival (OS). To identify robust independent risk prognostic immune-related genes (PIR-DEGs), the final IRGs to build prognostic prediction model were based on the threshold of $P < 0.05$. We also used GSE107943 as validation cohort to test reliability and stability of the model. To confirm reliability of the detected immune-related genes, a risk score was designed to identify a signature for prognostic analysis. We applied a linear model for prognosis risk score prediction:

$$\text{Risk score} = \beta_{\text{gene}(1)} \times \text{expr}_{\text{gene}(1)} + \beta_{\text{gene}(2)} \times \text{expr}_{\text{gene}(2)} + \dots + \beta_{\text{gene}(n)} \times \text{expr}_{\text{gene}(n)}$$

The risk score was calculated based on combination of the expression level of immune-related genes weighted by the regression coefficient (β). Expr_{gene} refers to the expression of IRGs in the sample, and β is the regression coefficient generated from multivariate Cox analysis in training set. The median risk score was used to divide samples into low-risk and high-risk group. The survival outcomes of the two groups were compared by KM curve. The predictive accuracy of the prognostic model was assessed by ROC curve (survival ROC package). And area under the curve (AUC) showed the variation between the predicted and observed outcomes.

Construction of TF and IRGs network

Correlations test between TF and PIR-DEGs was investigated by Pearson correlation coefficient (r), and the cut-off criteria were set as correlation coefficient > 0.4 and $P < 0.05$. The regulatory network was further visualized using Cytoscape version 3.7.2.

Construction and validation of the nomogram

A nomogram was visualized using “rms” package in the R platform based on risk scores and clinicopathological parameters. The nomogram was subjected to 1000 bootstrap resamples for internal validation. Discrimination and calibration were used to assess its predictive accuracy. Calibration of the nomogram for 3-, and 5-year OS was performed by comparing the actual survival with the predicted survival. Further, the discriminatory ability of the nomogram was evaluated by calculating the concordance index (C-index). All statistical tests were two-sided and $P < 0.05$ were considered statistically significant. Data compilations and descriptive statistics were performed using the SPSS24.0 (Chicago, IL, USA), along with rms package.

Estimation of tumor-infiltrating immune cells

To explore the relationships between risk scores derived from the prognostic model and immune cells infiltration, we employed CIBERSORT, a useful resource to comprehensively profile tumor infiltrating immune cells. CIBERSORT can precisely estimate the proportion of 22 immune cells in tissues by utilizing a deconvolution algorithm[27]. We uploaded the normalized gene expression data with standard annotation files to the CIBERSORT web portal (<http://cibersort.stanford.edu/>), and the algorithm was determined by 1000 permutations and LM22 gene signature as described in previous literatures. Correlation matrix for all 22 immune cell proportions were displayed by the corheatmap package.

Statistical analysis

Primary end point was OS, which was defined as the interval between the date of diagnosis and that of the most recent follow-up or of death from any cause. Categorical data were analyzed using Chi-square test or Fisher’s exact test, and continuous variables were analyzed using ANOVA or Kruskal-Wallis H test for variables with an abnormal distribution. Survival curves were plotted using Kaplan-Meier. Multivariate Cox proportional hazards regression model was used to identify independent predictors of OS. Unless otherwise noted, $P < 0.05$ was considered to be significant.

Results

Differential expressions of TF genes and IRGs in CCA

We used RNA-seq data and clinical follow-up information collected from 45 samples, including 36 CCA and 9 non-tumor tissues in TCGA. Wilcoxon rank test in R was performed to preprocess and analyze the gene differential expression of microarray data. Totally, 2657 DEGs were identified, which consist of 1203 significantly down-regulated DEGs and 1454 significantly up-regulated DEGs, for the subsequent bioinformatics analysis (Fig. 1a, d). To explore potential function and pathway of these DEGs identified above, GO and KEGG pathway enrichment analysis were performed (Supplementary Fig. 1).

The mRNA levels of 2498 IRGs (ImmPort database) and 318 TF genes (Cistrome Cancer database) were examined. This analysis eventually revealed 223 immune-related genes (130 downregulated and 93 upregulated) and 32 TFs (13 downregulated and 19 upregulated) with the thresholds of $|\log_2$ fold change $| > 1$ and adjusted $P < 0.05$, which were identified as differentially expressed in CCA tissues compared with normal tissues. Then, heatmaps and volcano plots were visualized to show the expression pattern of differentially expressed IRGs and TFs between CCA and non-tumor tissues (Fig. 1b, c, e, f).

Screening of prognostic immune-related genes and construction of TF regulatory network

Considering the association IR-DEGs with OS, we further conducted a Cox regression hazards analysis to screen the IR-DEGs. To prevent overfitting of the model, TCGA-CHOL dataset was utilized as training group. We found 11 prognostic IRGs (RAET1E, CST4, CCL24, CCK, CGB, GUCA2A, TDGF1, TDGF3, THPO, AVPR1B and IL9R) that were significantly related to the survival of CCA patients in the training group (Fig. 2a)

To further explore potential mechanisms behind deregulation of IR-DEGs expression in CCA, we analyzed the correlation between TF genes and IRGs expression. We investigated the correlations between the TF-DEGs and prognostic IR-DEGs by the method of Pearson correlation coefficient. Based on the threshold criteria, 24 TF genes were prominently associated with the PIR-DEGs ($P < 0.05$). We utilized Cytoscape software to build and visualize the TF-based regulatory network, as showed in Fig. 2b. This TF regulatory network revealed relationships among these immune-related genes. Furthermore, GO annotation analyses of IR-DEGs led to the identification of top significantly GO terms, and it indicated that the differentially expressed immune-related genes were mainly involved in the regulation of immune effector process and cell killing (Fig. 2c).

Construction of the prognostic model in the training group

To investigate significance of risk genes in estimating prognosis, the 11 survival-related IRGs were further submitted to a multivariate Cox proportional hazards regression analysis (with forward selection and backward selection). Finally, five candidate PIR-DEGs (RAET1E, CST4, TDGF1, AVPR1B and IL9R) which may serve as significant predictors of the prognosis were obtained for inclusion in the risk model. Among these genes, RAET1E, CST4 and TDGF1 were identified as high-risk genes (serving as risk factors), while AVPR1B and IL9R were identified as low-risk genes (serving as protective factors) (Table 1). In the TCGA training cohort, the expression distribution of five IRGs and Kaplan-Meier curves for OS were shown in Supplementary Fig. 2.

IRGs	Coef	HR	HR.95L	HR.95H	P-value
RAET1E	0.563	1.756	1.104	2.794	0.017
CST4	0.415	1.514	1.094	2.094	0.012
TDGF1	0.394	1.483	1.018	2.161	0.040
AVPR1B	-1.248	0.287	0.117	0.704	0.006
IL9R	-1.488	0.226	0.085	0.598	0.003

Table 1

The multivariate Cox regression analysis of key immune-related genes

A prognostic index was established based on the mRNA levels and estimated regression coefficients of the risk genes. The formula of prognostic model containing five PIR-DEGs is described as follows:

Risk score = (0.5631 × expression of RAET1E) + (0.4147 × expression of CST4) + (0.3941 × expression of TDGF1) + (−1.2478 × expression of AVPR1B) + (−1.4883 × expression of IL9R).

According to median risk score, patients were divided into high-risk (n = 18) and low-risk group (n = 18) in the training cohort. We ranked the risk scores of patients and analyzed their distribution (Fig. 3a). The survival status of each patient was marked on dot plot, which showed that mortality rate in the high-risk group was significantly higher than that of the low-risk group (Fig. 3b). Heatmap revealed that the expression level of five PIR-DEGs varied with risk scores in two groups (Fig. 3c). Among patients with higher risk scores in the training cohort, three risk genes (RAET1E, CST4 and TDGF1) were upregulated, while two protective genes (AVPR1B and IL9R) were downregulated. In patients with lower risk scores, these genes displayed an opposite expression pattern. Kaplan-Meier plot and ROC curve were used to evaluate the performance of five PIR-DEGs signature in predicting the outcome of patients. The prognosis was significantly worse in high-risk group than low-risk group (P<0.001, Fig. 3d). The OS rates at three years and five years for high-risk group in the training cohort were 19.9% and 9.9%, respectively, while the corresponding rates for low-risk group were 83.9% and 69.9%, respectively. Then we used ROC curves to examine predictive accuracy of the prognostic model at three years. The AUC value of ROC was 0.898 (Fig. 3e).

Validation of the prognostic model by GEO cohort

The GSE107943 cohort including 30 CCA samples was used for validation of the prognostic model. The risk score distribution, survival status and risk gene expression in the GEO cohort are displayed in Fig. 4a-c. In agreement with results of TCGA cohort, the Kaplan-Meier survival curves demonstrated that the OS was significantly poorer in high-risk group than low-risk group (P<0.05; Fig. 4d), and the survival rates at three and five years in high-risk group were 30.9% and 0.0%, respectively, while the corresponding rates in low-risk group were 75.0% and 40.0%, respectively. The AUC for three years was 0.846 (Fig. 4e). Taking

together, these results indicated a moderate sensitivity and specificity of the prognostic model. We also analyzed the expression distribution of five IRGs and Kaplan-Meier curves for OS, the results were shown in Supplementary Fig. 3.

Independent prognostic value of the risk model in the TCGA cohort

Next, we performed univariate and multivariate Cox regression analyses to assess the performance of model with the clinicopathological factors of CCA, such as age, gender, histological grade, T stage, N stage, M stage, AJCC stage and margin status. The univariate analysis indicated that N stage (HR = 4.148, 95%CI = 1.517–11.336, P = 0.006), margin status (HR = 2.754, 95%CI = 1.013–7.491, P = 0.047) and risk score (HR = 2.255, 95%CI = 1.409–3.611, P<0.001) were associated with the prognosis of CCA patients (Fig. 5a). Meanwhile, the multivariate analysis revealed that risk score was an independently risk factor in the TCGA cohort (HR = 3.009, 95%CI = 1.540–5.880, P = 0.001) (Fig. 5b). These results demonstrated that the prognostic model showed a good performance to predict prognosis.

Nomogram and ROC curves for the prediction of prognosis in the training cohort.

To better estimate the prognosis of CCA patients, we established a nomogram that integrated prognostic risk score and clinical variables (age, gender, N stage, M stage, AJCC stage, margin status and risk score) (Fig. 5c). The nomogram could accurately predict OS, and demonstrated that risk score contributed much more risk points than other variables. C-index for OS was 0.819 (95% CI, 0.727-0.911), and calibration plot for probability of survival at 3- or 5-year showed satisfactory agreements between the predicted and actual values (Fig. 5d, e).

The correlation between the risk score and tumor-infiltrating immune cells (TIICs) in the TCGA cohort

To determine whether our model could reflect the status of the tumor immune microenvironment in CCA, we investigated the correlation between risk score and immune cell infiltration in the TCGA cohort. The tumor-infiltrating immune cells were estimated by CIBERSORT algorithm. We plotted bar plots to demonstrate the proportion of immune cells of CCA and non-tumor tissues. As shown in Fig. 6a-b, the 22 TIIC proportions between two groups were significantly different. Heatmap was used to show the correlation between immune cells subpopulations, which indicated the fractions of different immune cells were weakly to moderately correlated in tumor tissues in the TCGA cohort. The three immune cells showed strong and moderate positive correlation with T gamma delta cells, CD8+ T cells and resting memory CD4+ T cells (Fig. 6c).

To analyze whether infiltration of immune cells might contribute to risk score, we analyzed correlation between risk score and tumor-immune cell proportions. As the risk score increased, the content of CD8+ T cells ($P<0.05$), neutrophils ($P<0.05$) and T gamma delta cells ($P<0.001$) in CCA tissues also increased, while the content of resting memory CD4+ T cells decreased ($P<0.05$) (Fig. 6d).

Discussion

Activation of immune system has been widely proven to be a decisive factor during tumorigenesis and metastasis[28–30]. With the development of tumor immunotherapy in recent years, the function of the immune system in the cancer growth and metastasis has been addressed, in which immune cells inhibit or promote malignant cells by up- or down-regulating expression of IRGs[31, 32]. Hence, IRGs expression may be an important predictor of prognosis and progression in CCA. Up to date, considerable efforts have been made to identify prognostic model based on differentially expressed genes, but the report on prognostic IRGs signature of CCA is lacking[33, 34]. In our study, we identified a new prognostic IRGs signature by screening the differentially expressed genes, and established a dependable model to predict survival of CCA patients based on this immune signature.

In the current study, we analyzed expression of 2498 IRGs, and screened out 223 IR-DEGs from TCGA-CHOL, including 93 upregulated and 130 downregulated genes. Eleven IR-DEGs were detected to show significant correlation with prognosis. The results revealed that IRGs were vital contributors to prognosis of CCA patients. Our functional analysis suggested that the IR-DEGs were widely involved in the tumor immunological process. To excavate the underlying molecular mechanisms behind the aberrant expression of PIR-DEGs, a TF regulatory network was built to display the relationship between TFs and PIR-DEGs.

Then we obtained five robust PIR-DEGs of interest (AVPR1B, CST4, TDGF1, RAET1E and IL9R) through multivariate Cox regression analyses, and utilized them to establish a novel prognostic IRGs model. The model showed a prominent performance by using TCGA-CHOL cohort. After adjusting to other clinical factors, the risk scores generated by the model were demonstrated to be an independent prognostic factor, and had better discriminating power for risk stratification. The reliability of the model was further validated based on GSE107943 cohort. Our results indicated that the model could act as an effective marker for survival prediction of CCA patients. Based on immune risk scores combining with other clinical factors (age, gender, AJCC stage, N stage, M stage and margin status), a nomogram of survival prediction was generated and exhibited satisfactory agreements between the predicted and actual values, which increased the predictive capacity. Thus, our study demonstrated the potential clinical application for survival prediction of CCA patients.

Recently, some studies have demonstrated that immune infiltration was a crucial determinant of responsiveness to immunotherapy and prognosis prediction of CCA[35–37]. Our immune-related prognostic model displayed positive correlation with CD8+ T cells, neutrophils and T gamma delta cells, while negative with CD4+ memory resting T cells. A study about ICC revealed that increased CD8+ tumor-

infiltrating lymphocytes (TILs) in tumors was associated with better tumor differentiation[38]. The infiltration of tumor-associated neutrophils (TANs) is recognized to facilitate tumor progression in CCA[39, 40], which is consistent with our findings. Several studies confirmed that high levels of CD4+ TILs were associated with better prognosis[41, 42]. In summary, the immune milieu in tumor microenvironment may account to the difference of survival outcome between low- and high- risk group.

There have been multiple studies for the value of IRG models to predict prognosis in various cancers, such as renal papillary cell carcinoma[43], clear cell renal cell carcinoma[44], hepatocellular carcinoma[45] and cervical cancer[46]. Compared with the previous studies, our study had some advantages. First, we explored the relationship between IRGs expression signature and prognosis in CCA, and identified some prognostic IRGs. Second, our prognostic model based on IRGs and the nomogram showed outstanding performance in survival prediction. Inevitably, some limitations should be addressed. The included cohort in our study has limited sample size, which needs a larger number of CCA samples to verify the model accuracy. Next, the potential molecular mechanisms of IRGs impact on CCA are not fully elucidated, which require further investigation.

Conclusions

In summary, we established a novel five IRGs-based prognostic model that accurately predicted the prognosis in CCA, and further confirmed the prognostic performance of this model using GEO database. Our finding elucidated the association between immune risk score and immune cells infiltration, proving its key role in tumor immune microenvironment. These findings provide new insight into the role of IRGs and tumor immune cells infiltration on CCA, and could be useful in the future as a theoretical foundation for guiding immunotherapy.

Abbreviations

CCA: Cholangiocarcinoma; IRGs: Immune-related genes; TCGA: The Cancer Genome Atlas; GEO: Gene Expression Omnibus; TF: Transcription factor; ROC: Receiver operating characteristic; AUC: The area under the ROC curve; TNM: Tumor-node-metastasis, the most widely used cancer staging system constructed by the American Joint Committee on Cancer and the Union for International Cancer Control; ImmPort: The Immunology Database and Analysis Portal; DEGs: Differentially expressed genes; OS: Overall survival; FDR: False discovery rate; TIME: Tumor immune microenvironment; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; TILs: Tumor-infiltrating lymphocytes.

Declaration

Acknowledgements

None.

Authors' contributions

RZ and ZG conceived and designed this work; RZ, CC, QL, DZ and WL participated in data acquisition, interpretation and statistical analysis; RZ and CC were responsible for writing the manuscript; ZG and JF revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China, No. 81572420, and the Key Research and Development Program of Shaanxi Province, No. 2017ZDXM-SF-055.

Availability of data and materials

All data used in this study were acquired from The Cancer Genome Atlas (TCGA) portal and Gene Expression Omnibus (GEO) dataset.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Akinyemiju T, Abera S, Ahmed M, Alam N, Alemayohu MA, Allen C, Al-Raddadi R, Alvis-Guzman N, Amoako Y, Artaman A, et al: The Burden of Primary Liver Cancer and Underlying Etiologies From 1990 to 2015 at the Global, Regional, and National Level Results From the Global Burden of Disease Study 2015. *Jama Oncol* 2017, 3:1683-1691.
2. Banales JM, Cardinale V, Carpino G, Marzioni M, Andersen JB, Invernizzi P, Lind GE, Folseraas T, Forbes SJ, Fouassier L, et al: Expert consensus document: Cholangiocarcinoma: current knowledge and future perspectives consensus statement from the European Network for the Study of Cholangiocarcinoma (ENS-CCA). *Nat Rev Gastroenterol Hepatol* 2016, 13:261-280.

3. Hyder O, Hatzaras I, Sotiropoulos GC, Paul A, Alexandrescu S, Marques H, Pulitano C, Barroso E, Clary BM, Aldrighetti L, et al: Recurrence after operative management of intrahepatic cholangiocarcinoma. *Surgery* 2013, 153:811-818.
4. Spolverato G, Kim Y, Alexandrescu S, Marques HP, Lamelas J, Aldrighetti L, Clark Gamblin T, Maithel SK, Pulitano C, Bauer TW, et al: Management and Outcomes of Patients with Recurrent Intrahepatic Cholangiocarcinoma Following Previous Curative-Intent Surgical Resection. *Ann Surg Oncol* 2016, 23:235-243.
5. Lamarca A, Hubner RA, David Ryder W, Valle JW: Second-line chemotherapy in advanced biliary cancer: a systematic review. *Ann Oncol* 2014, 25:2328-2338.
6. Xu G, Sun L, Li Y, Xie F, Zhou X, Yang H, Du S, Xu H, Mao Y: The Clinicopathological and Prognostic Value of PD-L1 Expression in Cholangiocarcinoma: A Meta-Analysis. *Front Oncol* 2019, 9:897.
7. Ma K, Wei X, Dong D, Wu Y, Geng Q, Li E: PD-L1 and PD-1 expression correlate with prognosis in extrahepatic cholangiocarcinoma. *Oncol Lett* 2017, 14:250-256.
8. Fukuda Y, Asaoka T, Eguchi H, Yokota Y, Kubo M, Kinoshita M, Urakawa S, Iwagami Y, Tomimaru Y, Akita H, et al: Endogenous CXCL9 affects prognosis by regulating tumor-infiltrating NK cells in intrahepatic cholangiocarcinoma. *Cancer Sci* 2019.
9. Dong Z, Liao B, Shen W, Sui C, Yang J: Expression of Programmed Death Ligand 1 Is Associated with the Prognosis of Intrahepatic Cholangiocarcinoma. *Dig Dis Sci* 2019.
10. Jing CY, Fu YP, Yi Y, Zhang MX, Zheng SS, Huang JL, Gan W, Xu X, Lin JJ, Zhang J, et al: HHLA2 in intrahepatic cholangiocarcinoma: an immune checkpoint with prognostic significance and wider expression compared with PD-L1. *J Immunother Cancer* 2019, 7:77.
11. Miyata T, Yamashita YI, Yoshizumi T, Shiraishi M, Ohta M, Eguchi S, Aishima S, Fujioka H, Baba H: CXCL12 expression in intrahepatic cholangiocarcinoma is associated with metastasis and poor prognosis. *Cancer Sci* 2019, 110:3197-3203.
12. Lu JC, Zeng HY, Sun QM, Meng QN, Huang XY, Zhang PF, Yang X, Peng R, Gao C, Wei CY, et al: Distinct PD-L1/PD1 Profiles and Clinical Implications in Intrahepatic Cholangiocarcinoma Patients with Different Risk Factors. *Theranostics* 2019, 9:4678-4687.
13. Rimassa L, Personeni N, Aghemo A, Lleo A: The immune milieu of cholangiocarcinoma: From molecular pathogenesis to precision medicine. *J Autoimmun* 2019, 100:17-26.
14. Kaufman HL, Kirkwood JM, Hodi FS, Agarwala S, Amatruda T, Bines SD, Clark JI, Curti B, Ernstoff MS, Gajewski T, et al: The Society for Immunotherapy of Cancer consensus statement on tumour immunotherapy for the treatment of cutaneous melanoma. *Nat Rev Clin Oncol* 2013, 10:588-598.
15. Camidge DR, Doebele RC, Kerr KM: Comparing and contrasting predictive biomarkers for immunotherapy and targeted therapy of NSCLC. *Nat Rev Clin Oncol* 2019, 16:341-355.
16. Havel JJ, Chowell D, Chan TA: The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat Rev Cancer* 2019, 19:133-150.
17. Rizvi S, Gores GJ: Emerging molecular therapeutic targets for cholangiocarcinoma. *J Hepatol* 2017, 67:632-644.

18. Paillet J, Kroemer G, Pol JG: Immune contexture of cholangiocarcinoma. *Curr Opin Gastroenterol* 2020, 36:70-76.
19. Fabris L, Perugorria MJ, Mertens J, Bjorkstrom NK, Cramer T, Lleo A, Solinas A, Sanger H, Lukacs-Kornek V, Moncsek A, et al: The tumour microenvironment and immune milieu of cholangiocarcinoma. *Liver Int* 2019, 39 Suppl 1:63-78.
20. Arkenau HT, Martin-Liberal J, Calvo E, Penel N, Krebs MG, Herbst RS, Walgren RA, Widau RC, Mi G, Jin J, et al: Ramucirumab Plus Pembrolizumab in Patients with Previously Treated Advanced or Metastatic Biliary Tract Cancer: Nonrandomized, Open-Label, Phase I Trial (JVDF). *Oncologist* 2018, 23:1407-e1136.
21. Bang YJ, Doi T, Braud FD, Piha-Paul S, Hollebecque A, Razak ARA, Lin CC, Ott PA, He AR, Yuan SS, et al: 525 Safety and efficacy of pembrolizumab (MK-3475) in patients (pts) with advanced biliary tract cancer: Interim results of KEYNOTE-028. *Eur J Cancer* 2015, 51:S112.
22. Fridman WH, Zitvogel L, Sautes-Fridman C, Kroemer G: The immune contexture in cancer prognosis and treatment. *Nat Rev Clin Oncol* 2017, 14:717-734.
23. Albini A, Bruno A, Noonan DM, Mortara L: Contribution to Tumor Angiogenesis From Innate Immune Cells Within the Tumor Microenvironment: Implications for Immunotherapy. *Front Immunol* 2018, 9:527.
24. da Silva JL, Dos Santos ALS, Nunes NCC, de Moraes Lino da Silva F, Ferreira CGM, de Melo AC: Cancer immunotherapy: the art of targeting the tumor immune microenvironment. *Cancer Chemother Pharmacol* 2019, 84:227-240.
25. Montal R, Sia D, Montironi C, Leow WQ, Esteban-Fabro R, Pinyol R, Torres-Martin M, Bassaganyas L, Moeini A, Peix J, et al: Molecular classification and therapeutic targets in extrahepatic cholangiocarcinoma. *J Hepatol* 2020.
26. Ahn KS, O'Brien D, Kang YN, Mounajjed T, Kim YH, Kim TS, Kocher JA, Allotey LK, Borad MJ, Roberts LR, Kang KJ: Prognostic subclass of intrahepatic cholangiocarcinoma by integrative molecular-clinical analysis and potential targeted approach. *Hepatol Int* 2019, 13:490-500.
27. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA: Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 2015, 12:453-457.
28. Ren B, Cui M, Yang G, Wang H, Feng M, You L, Zhao Y: Tumor microenvironment participates in metastasis of pancreatic cancer. *Mol Cancer* 2018, 17:108.
29. Kitamura T, Qian BZ, Pollard JW: Immune cell promotion of metastasis. *Nat Rev Immunol* 2015, 15:73-86.
30. Upadhyay S, Sharma N, Gupta KB, Dhiman M: Role of immune system in tumor progression and carcinogenesis. *J Cell Biochem* 2018, 119:5028-5042.
31. Kalbasi A, Ribas A: Tumour-intrinsic resistance to immune checkpoint blockade. *Nat Rev Immunol* 2020, 20:25-39.
32. Qin S, Xu L, Yi M, Yu S, Wu K, Luo S: Novel immune checkpoint targets: moving beyond PD-1 and CTLA-4. *Mol Cancer* 2019, 18:155.

33. Guo H, Cai J, Wang X, Wang B, Wang F, Li X, Qu X, Kong X, Gao Y, Wu H, et al: Prognostic values of a novel multi-mRNA signature for predicting relapse of cholangiocarcinoma. *Int J Biol Sci* 2020, 16:869-881.
34. Da Z, Gao L, Su G, Yao J, Fu W, Zhang J, Zhang X, Pei Z, Yue P, Bai B, et al: Bioinformatics combined with quantitative proteomics analyses and identification of potential biomarkers in cholangiocarcinoma. *Cancer Cell Int* 2020, 20:130.
35. Sabbatino F, Villani V, Yearley JH, Deshpande V, Cai L, Konstantinidis IT, Moon C, Nota S, Wang Y, Al-Sukaini A, et al: PD-L1 and HLA Class I Antigen Expression and Clinical Course of the Disease in Intrahepatic Cholangiocarcinoma. *Clin Cancer Res* 2016, 22:470-478.
36. Gani F, Nagarajan N, Kim Y, Zhu Q, Luan L, Bhaijje F, Anders RA, Pawlik TM: Program Death 1 Immune Checkpoint and Tumor Microenvironment: Implications for Patients With Intrahepatic Cholangiocarcinoma. *Ann Surg Oncol* 2016, 23:2610-2617.
37. Zhu Y, Wang XY, Zhang Y, Xu D, Dong J, Zhang Z, Yi CH, Jia HL, Yang X: Programmed death ligand 1 expression in human intrahepatic cholangiocarcinoma and its association with prognosis and CD8(+) T-cell immune responses. *Cancer Manag Res* 2018, 10:4113-4123.
38. Ye Y, Zhou L, Xie X, Jiang G, Xie H, Zheng S: Interaction of B7-H1 on intrahepatic cholangiocarcinoma cells with PD-1 on tumor-infiltrating T cells as a mechanism of immune evasion. *J Surg Oncol* 2009, 100:500-504.
39. Kitano Y, Okabe H, Yamashita YI, Nakagawa S, Saito Y, Umezaki N, Tsukamoto M, Yamao T, Yamamura K, Arima K, et al: Tumour-infiltrating inflammatory and immune cells in patients with extrahepatic cholangiocarcinoma. *Br J Cancer* 2018, 118:171-180.
40. Mao ZY, Zhu GQ, Xiong M, Ren L, Bai L: Prognostic value of neutrophil distribution in cholangiocarcinoma. *World J Gastroenterol* 2015, 21:4961-4968.
41. Goeppert B, Frauenschuh L, Zucknick M, Stenzinger A, Andrulis M, Klauschen F, Joehrens K, Warth A, Renner M, Mehrabi A, et al: Prognostic impact of tumour-infiltrating immune cells on biliary tract cancer. *Br J Cancer* 2013, 109:2665-2674.
42. Zhou G, Sprengers D, Mancham S, Erkens R, Boor PPC, van Beek AA, Doukas M, Noordam L, Campos Carrascosa L, de Ruiter V, et al: Reduction of immunosuppressive tumor microenvironment in cholangiocarcinoma by ex vivo targeting immune checkpoint molecules. *J Hepatol* 2019, 71:753-762.
43. Wang Z, Song Q, Yang Z, Chen J, Shang J, Ju W: Construction of immune-related risk signature for renal papillary cell carcinoma. *Cancer Med* 2019, 8:289-304.
44. Yin Z, Dong C, Jiang K, Xu Z, Li R, Guo K, Shao S, Wang L: Heterogeneity of cancer-associated fibroblasts and roles in the progression, prognosis, and therapy of hepatocellular carcinoma. *J Hematol Oncol*. 2019, 12:101.
45. Carone C, Olivani A, Dalla Valle R, Manuguerra R, Silini EM, Trenti T, Missale G, Cariani E: Immune Gene Expression Profile in Hepatocellular Carcinoma and Surrounding Tissue Predicts Time to Tumor Recurrence. *Liver Cancer* 2018, 7:277-294.

46. Yang S, Wu Y, Deng Y, Zhou L, Yang P, Zheng Y, Zhang D, Zhai Z, Li N, Hao Q, et al: Identification of a prognostic immune signature for cervical cancer to predict survival and response to immune checkpoint inhibitors. *Oncoimmunology* 2019, 8:e1659094.

Figures

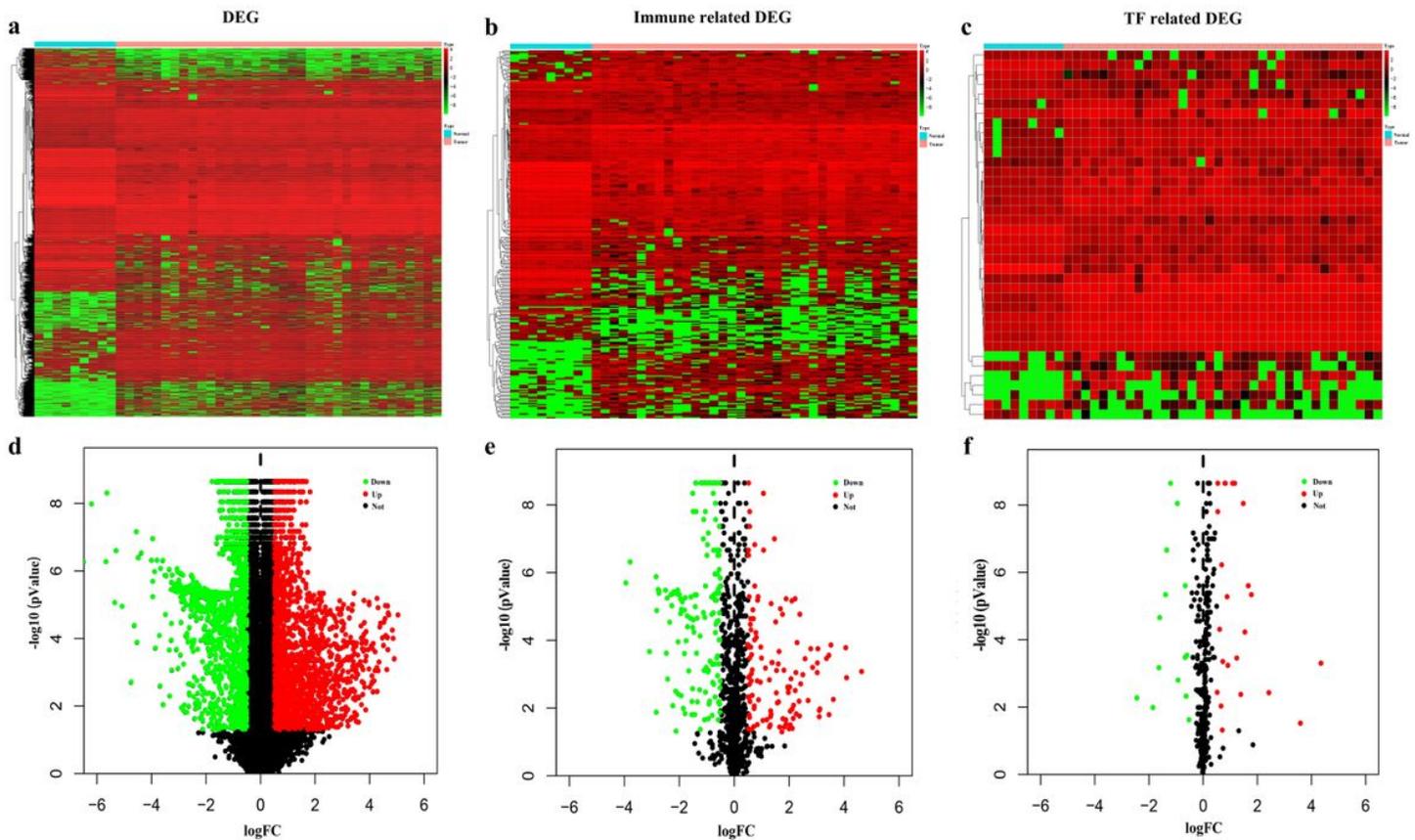


Figure 1

Differential expression of immune-related genes and TF genes in CCA. Upper panel: Heatmaps of the differential expressed genes (DEGs) between CCA and non-tumor samples, including a All differential expressed genes, b IRGs, and c TF genes. DEGs are represented in rows, and samples are represented in columns. The expression value for each row was normalized by the z-score. The color from green to red represents the progression from low expression to high expression. Blue bar represents non-tumor, while red one represents tumor samples. Bottom panel: Volcano plots of the DEGs between CCA and non-tumor samples, including d All differential expression of DEGs, e IRGs, and f TF genes. Colored dots represent differentially expressed genes and black dots represent no differentially expressed genes. Red dots represent upregulation while green dots represent downregulation.

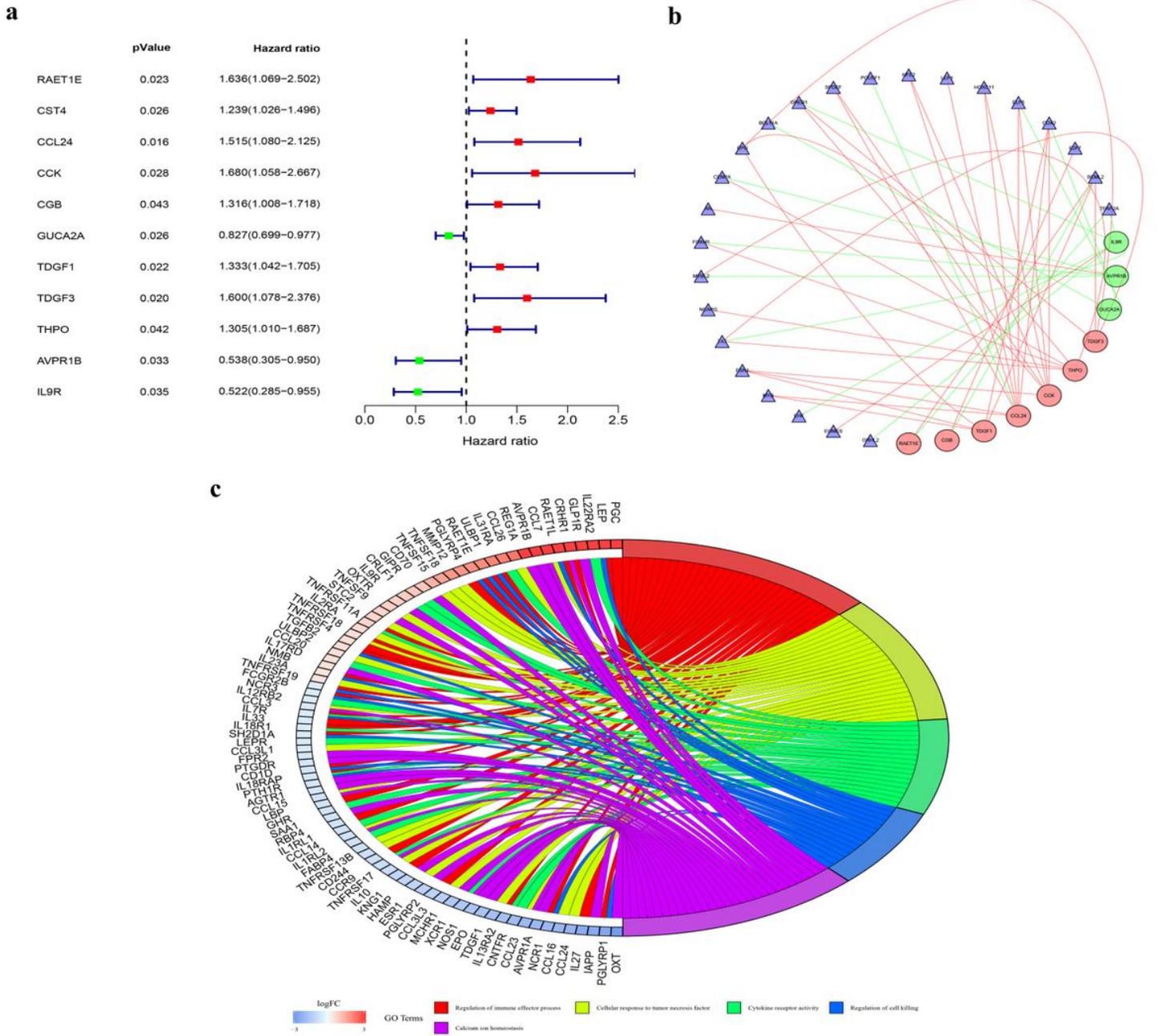


Figure 2

Screening of prognostic immune-related genes and construction of TF regulatory network. a The forest plot of 11 prognostic immune-related DEGs screened out by univariate Cox proportional hazards regression. b The main regulatory network constructed based on prognostic immune-related genes and TFs. The blue triangles in the network represent TFs and the circles represented IRGs that red is up-regulated and green is down-regulated. The color of lines represents regulatory relationship, which red is positive regulation and green is negative regulation. c GO annotation analyses of the IR-DEGs. Gene ontology terms which significantly enriched were showed.

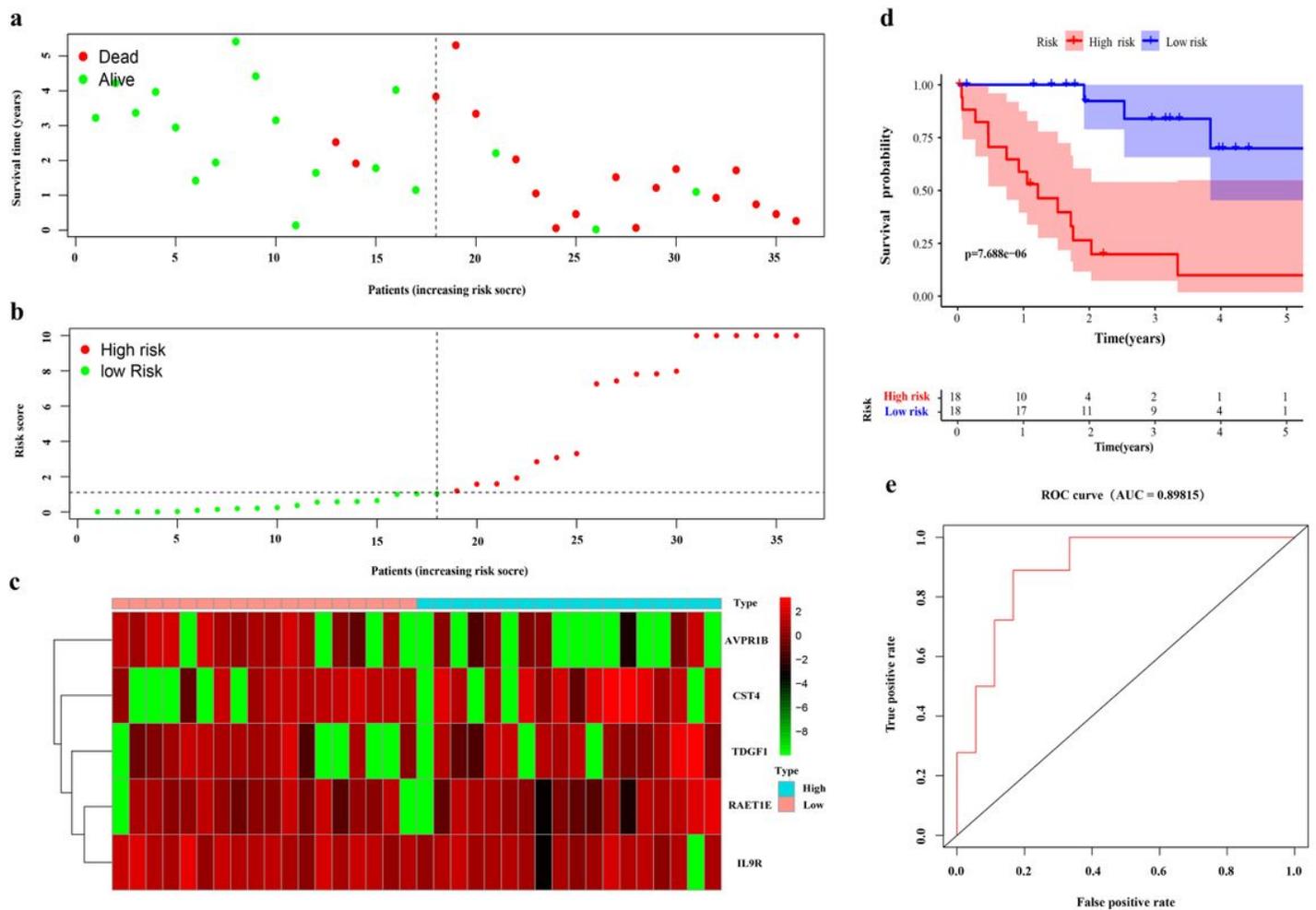


Figure 3

Construction of the prognostic model in the training group Risk score analysis, survival and ROC curves for the IRGs signature in training group. a Risk score distribution of patients with different risks (low, green; high, red). b Scatterplots of patients with different survival status. c Heatmap of AVPR1B, CST4, TDGF1, RAET1E and IL9R expression. The color from green to red indicates the expression level from low to high. The dotted line indicates the individual inflection point of the risk score curve, by which the patients were categorized into high-and low-risk group. d Kaplan-Meier survival curves showing overall survival outcomes according to high- and low-risk group. e ROC analysis of the model consisted of IRGs signature. ROC, Receiver operating characteristic; AUC, Area under curve.

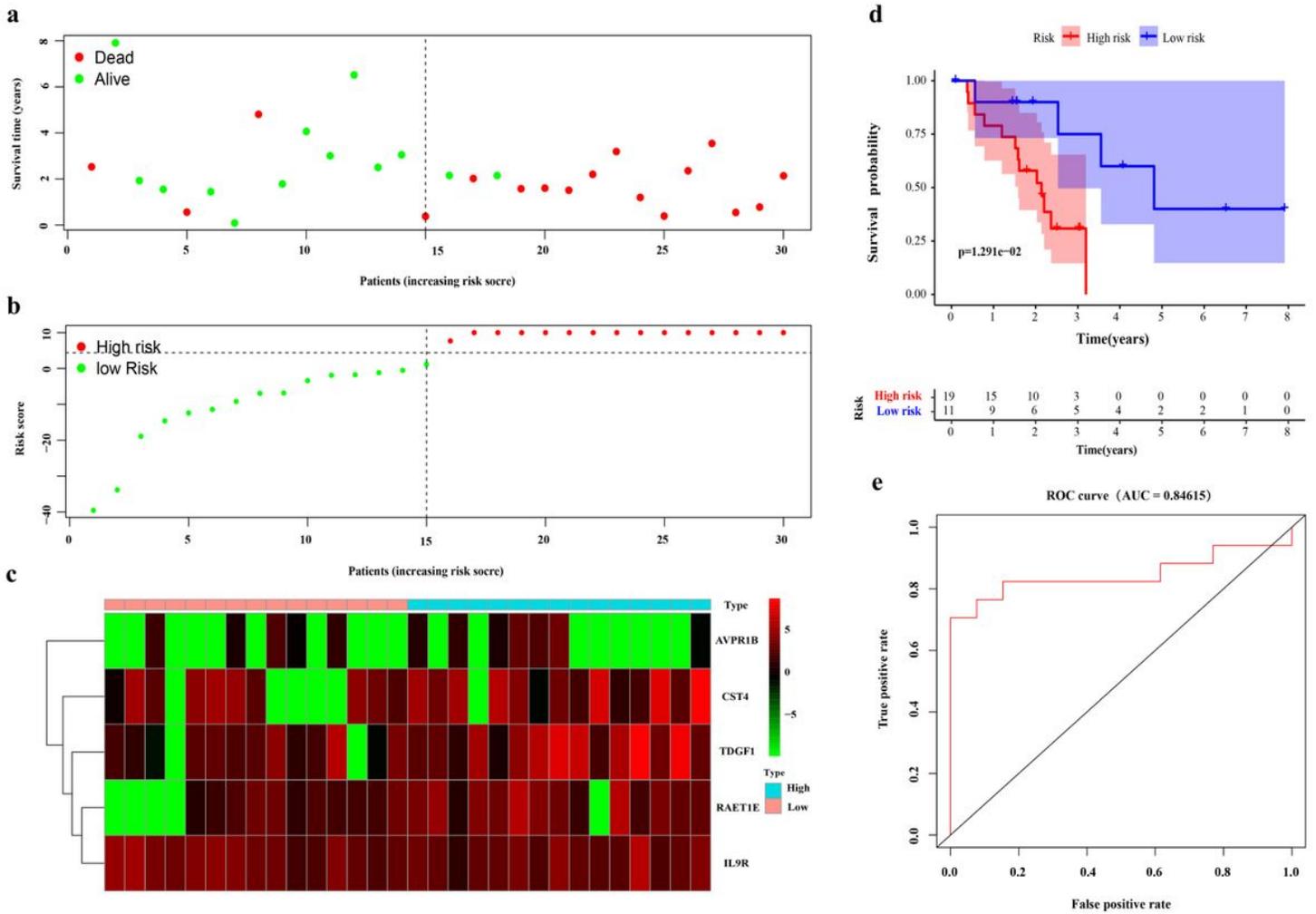


Figure 4

Validation of the prognostic model in the GEO cohort. Risk score analysis, survival and ROC curves for the IRGs signature in GEO validation cohort. a Risk score distribution of patients with different risks. b Scatterplots of patients with different survival status. c Heatmap of five IRGs expression. d Kaplan-Meier survival curves. e ROC analysis of the model.

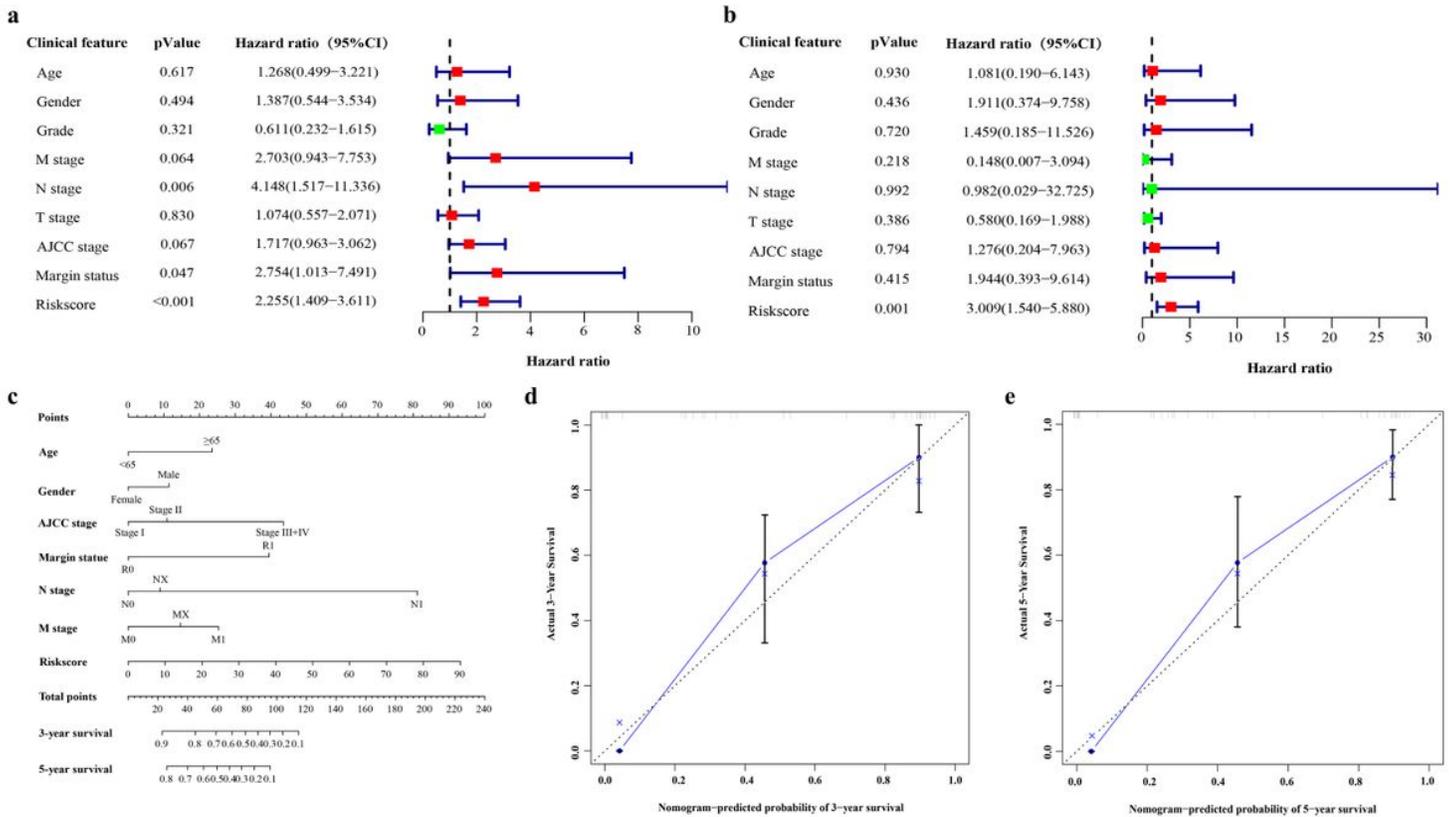


Figure 5

Independent prognostic value of the risk model, nomogram and ROC curves for the prediction of prognosis in the TCGA cohort. a Prognostic effect analysis of immune-related gene signature and commonly used prognostic factors using univariate Cox regression model. b Independent prognostic effect analysis of immune-related gene signature and commonly used prognostic factors using multivariate Cox regression model. c The nomogram based on the IRGs signature and clinical information to predict the 3- and 5-year survival of patients. d, e Calibration curves evaluating the predictive accuracy of nomogram at 3- and 5- year.

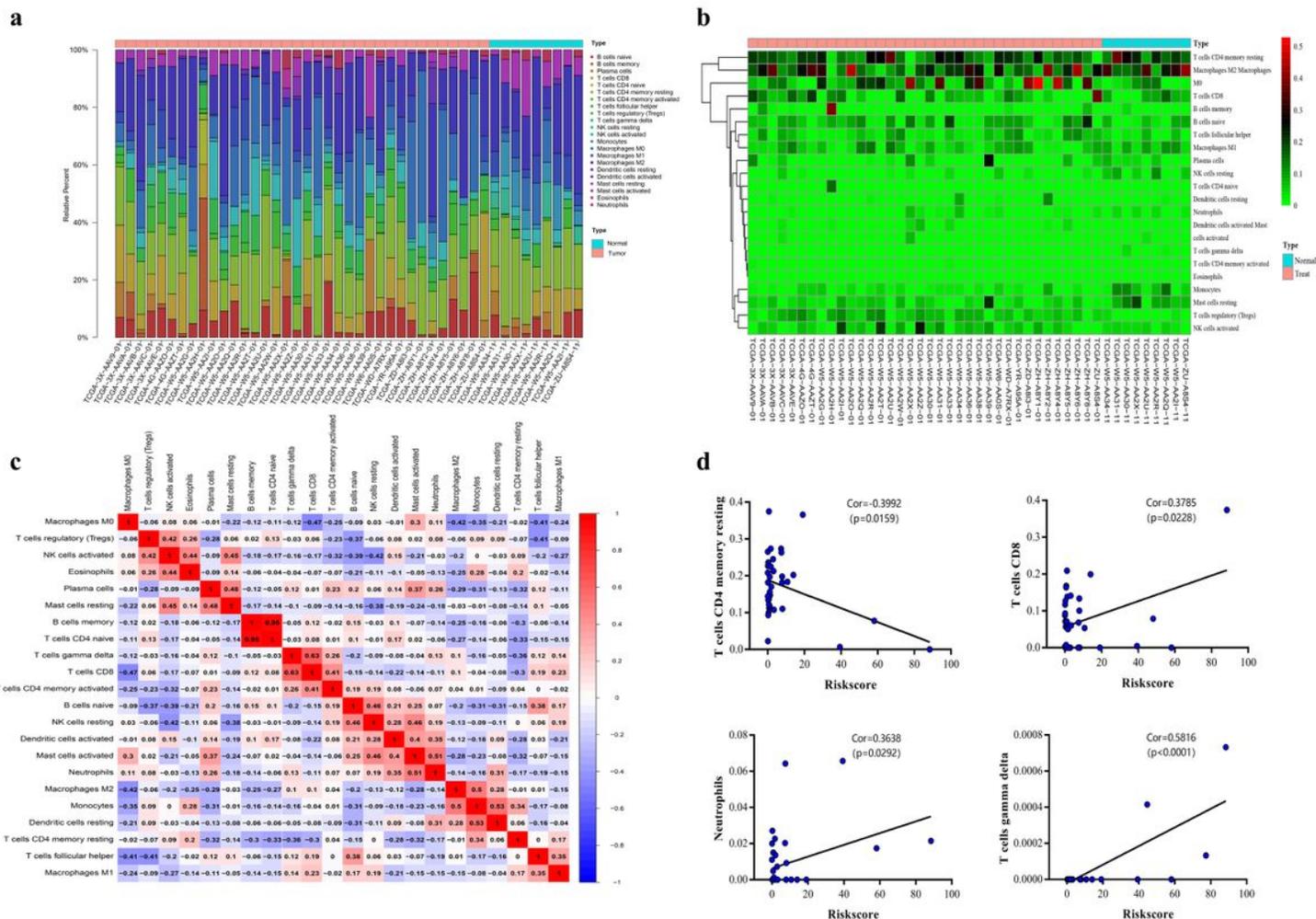


Figure 6

The correlation between the risk score and tumor-infiltrating immune cells in the TCGA cohort. The landscape of tumor-infiltrating immune cells in CCA and normal samples. a Stacked bar chart represents deviations in immune infiltration in each sample. Each column represents a sample, and each column with a different color and height indicates the abundance ratios of immune cells in this sample. b Hierarchical clustering of normal and tumor samples based on 22 immune cell proportions. The horizontal axis represents samples that were divided into two clusters. Red, black and green color indicates the high, moderate and low proportion of immune cells, respectively. c Correlation matrix of immune cell proportions and risk score. The value represents the correlation value. Red color represents positive correlation and blue color represents negative correlation. d The significant correlation the risk score and tumor-infiltrating immune cells. CD4+ memory resting T cells; CD8+ T cells; Neutrophils; T gamma delta cells.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFig.1.tif](#)
- [SupplementaryFig.2.tif](#)
- [SupplementaryFig.3.tif](#)