

A Prognostic Signature Based on Immune-Related Genes Associated with Tumor Microenvironment for Cholangiocarcinoma

Rui Zhang

Xi'an Jiaotong University Medical College First Affiliated Hospital

Chen Chen

Xi'an Jiaotong University Medical College First Affiliated Hospital

Qi Li

Xi'an Jiaotong University Medical College First Affiliated Hospital

Zhechuan Jin

Xi'an Jiaotong University Medical College First Affiliated Hospital

Jingbo Su

Xi'an Jiaotong University Medical College First Affiliated Hospital

Jian Zhang

Xi'an Jiaotong University Medical College First Affiliated Hospital

Jialu Fu

Xi'an Jiaotong University Medical College First Affiliated Hospital

Dong Zhang

Xi'an Jiaotong University Medical College First Affiliated Hospital

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

Xi'an Jiaotong University Medical College First Affiliated Hospital <https://orcid.org/0000-0003-2645-9808>

Primary research

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

Posted Date: October 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-997350/v1>

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

[Read Full License](#)

Abstract

Background

Several studies indicate immune related genes (IRGs) dramatically correlates with tumorigenesis and progress of cancer. However, the association IRGs and the value in predicting the prognosis of cholangiocarcinoma (CCA) patients remains unclear. Thus, we aimed to discover reliable biomarkers based on IRGs signature to predict survival in CCA.

Methods

Gene expression profiles of CCA patients were collected from TCGA and GEO database. Univariate, multivariate and Lasso (Least absolute shrinkage and selection operator) Cox analysis was used to identify IRGs signature. Subsequently, a nomogram that predicts outcomes for CCA patients based on IRGs signature was constructed to visualize prognosis. We also explored the relationships between tumor-infiltrating immune cells (TIICs) and IRGs. GDSC database was used to explore potential targeted drugs. Tissue samples of 23 CCA patients were collected and qRT-PCR were employed to validate the mRNA level of IRGs.

Results

5 IRGs (AVPR1B, CST4, TDGF1, RAET1E and IL9R) were identified as robust indicators for OS. A prognostic model based on IRGs signature can significantly stratify patients into high-risk and low-risk groups. Meanwhile, ROC indicated that the risk model had satisfactory performance. We also observed that the risk score was an independent prognostic factor for OS and built a nomogram. The low-risk group had a significantly higher percentage of infiltrated macrophages M1 than the high-risk group. Through analyzing the GDSC database, the high-risk group demonstrated drug sensitivity to some targeted drugs such as AMG 706 and Thapsigargin. Finally, the results of qRT-PCR were consistent with bioinformatics analysis.

Conclusions

We established a novel IRGs signature to predict prognosis for CCA. The risk model exhibited excellent diagnostic efficiency in predicting OS.

Background

Cholangiocarcinoma (CCA) is a highly aggressive malignancy on biliary tract system with extremely unsatisfactory prognosis, and its incidence is increasing worldwide in recent years [1]. Although surgical resection remains the only possible treatment, the 5-year survival rate for CCA patients underwent radical

resection remains less than 30% due to high frequency of metastasis and recurrence [2]. The benefits from adjuvant therapy appear to be very limited [3]. One reason for the poor prognosis of CCA is that most patients are diagnosed at advanced stage. Immunotherapy has become a new pillar of cancer treatment for more than a decade, and it has offered hope for reducing the morbidity and mortality of this refractory disease [4]. Immunotherapy based on immune checkpoint inhibitors (ICIs) has achieved notable success in clinical treatment of tumor [5, 6]. Many studies have demonstrated that immunotherapy shows a promising safety and efficacy for patients with CCA. Unfortunately, immunotherapy is not effective for all CCA patients [7, 8]. Hence, these problems necessitate the development of an effective risk model that allows to closely monitor progression and shed light on treatment stratification.

Evidence that immune-related genes (IRGs) are closely related to the tumorigenesis and progression of CCA has substantially increased over past years [9, 10]. Several IRGs have been reported to serve as promising biomarkers for predicting survival [11–13]. For example, PD-L1 was significantly correlated with TNM stage and overall survival (OS) for CCA patients after resection [14, 15]. The comprehensive analysis of the association between the IRGs and prognosis is conducive to elucidating the potential prognostic value. However, most of previous investigations mainly focused on a single IRG, which is not sufficiently rigorous to clinical practice. Tumor immune microenvironment (TIME), composed of many immune cells, participated in tumor progression and immunotherapeutic response [16–18]. The proportion and type of tumor-infiltrating immune cells (TIICs) and cytokines in the TIME are in turn determined by IRGs expression. Therefore, it is essential to explore the relevance between the IRGs and immune cells, which will help to uncover the underlying molecular pathway and find a robust immune target serve as prognostic biomarker.

Our study was to investigate the clinical implications of IRGs signature on prognostic stratification for CCA patients. Second, we also explored the association between IRG signature and tumor immune cell infiltration. Finally, a prognostic nomogram that incorporated prognostic IRGs signature and clinical features was established to predict OS.

Materials And Methods

Data acquisition and processing

The RNA-seq data from CCA and matched normal samples were downloaded from The Cancer Genome Atlas cholangiocarcinoma dataset (TCGA-CHOL) (<https://cancergenome.nih.gov/>) and Gene Expression Omnibus (GEO) dataset (GSE107943/GPL18573) (<https://www.ncbi.nlm.nih.gov/gds/>) [19]. The corresponding clinicopathological and survival information were downloaded from open-access resource, which included age, gender, AJCC stage, OS, disease-free survival (DFS), etc.

Meanwhile, we obtained a list of IRGs from the ImmPort database (<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 all selected. Transcription factors (TFs) were obtained from the Cistrome Cancer database (<http://cistrome.org/CistromeCancer>). The list of IRGs and TFs used in our study were listed on Additional file 6: Table S1.

Identification of differentially expressed IRGs and TFs

The normalization of gene expression and screening of differentially expressed genes (DEGs) were determined using “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. To explore regulatory network of IRGs, we draw the IRGs and TFs from all DEGs. Volcano plots and heatmaps were plotted respectively by “gplots” and “pheatmap” package on R software.

GO and KEGG pathway enrichment analysis

To understand biological pathways underlying differentially expressed genes, 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. GO analysis consists of three categories: biological process (BP), cellular component (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 Cox analysis were utilized to assess relationship between IRGs expression and OS. In order to further identify the robust independent prognostic IRGs, we performed least absolute shrinkage and selection operator (Lasso) and multivariate Cox analysis to conduct dimensionality reduction analysis by “survival” and “glmnet” packages. An independent prognostic model for prognostic analysis was designed based on the IRGs signature. We applied a linear model for prognosis risk score prediction:

$$\text{Risk score} = \sum_{i=1}^n (\beta_i * \text{Exprgene}_i)$$

The risk score was calculated based on combination of the expression level of IRGs weighted by the regression coefficient. Exprgene_i refers to the expression of IRGs in the sample, and β_i is the regression coefficient. To prevent overfitting of the model, TCGA-CHOL dataset was utilized as training group. GSE107943 was used as a validation cohort to test reliability and stability of the model. The median risk score was used to divide samples into low-risk and high-risk group. The survival curves were plotted by Kaplan-Meier (K-M) analysis was performed to assess differences in the survival rates between two group. The predictive performance of prognostic model was evaluated by ROC curve. In addition, the risk score distribution and survival status scatter plots for patients in the prognostic model and the heatmap of prognosis-related IRGs were also displayed.

Construction of TF and IRGs network

To uncover the regulatory mechanisms of independent prognostic IRGs, correlations test between TF and IRGs was investigated by Pearson correlation coefficient (r), and the cut-off criteria were set as correlation coefficient >0.4 and $P < 0.05$. The TF-IRGs regulatory network was built on STRING database and further visualized using Cytoscape version 3.7.2.

Construction and validation of the nomogram

A nomogram was visualized by means of “rms” package based on risk scores and other clinicopathological parameters (age, gender and AJCC stage). The nomogram was subjected to 1000 bootstrap resamples for internal validation. Calibration of the nomogram for 1-, 2- and 3- year OS was performed by comparing the actual survival with the predicted survival. Further, the discriminatory ability of nomogram was evaluated by calculating the concordance index (C-index).

Estimation of tumor-infiltrating immune cells

We used CIBERSORT algorithm to quantify the proportion of immune cells [20]. The RNA-seq data of CCA patients from TCGA-CHOL and GSE107943 were integrated. Normalized gene expression data were analyzed using the CIBERSORT algorithm, running with 1,000 permutations. The correlation matrix for all 22 immune cell proportions were displayed by the “corheatmap” package. The difference of abundance of 22 TIICs between high- and low- risk groups was compared by Wilcoxon test. Furthermore, the relationship between TIICs and prognostic IRGs were analyzed using Spearman's rank correlation. The distribution of gene expression and immune cell infiltration in each sample were also displayed by “ggscatterstats” package.

Chemotherapeutic response prediction

According to the immune related risk score, we divided the CCA patients into high risk group and low risk group. Based on the Genomics of Drug Sensitivity in Cancer (GDSC, <https://www.cancerrxgene.org/>)[21], the largest publicly available pharmacogenomics database, we explored the chemotherapeutic response of targeted drugs for the patients from different risk groups. Two commonly used chemo drugs, cisplatin and gemcitabine, were also selected. The prediction process was implemented by R package “pRRophetic”, which estimated the half-maximal inhibitory concentration (IC50) by ridge regression and the prediction accuracy was evaluated by 10-fold cross-validation based on the GDSC training set.

Quantitative Real-Time -PCR

Twenty-three matched tumorous and non-tumorous tissue specimens of cholangiocarcinoma are collected from the Biobank of First Affiliated Hospital of Xi'an Jiaotong University. Total RNA was extracted using TRIzol reagent (Invitrogen) and the total RNA integrity were checked by RNA 6000 Nano kit. Before reverse transcription to cDNA, RevertAid First Strand cDNA Synthesis Kit (Thermoscientific K1622) was used and reacted at 42°C for 2 min to remove the residual genomic DNA from total RNA.

PrimeScript[®], RT reagent kit was used to synthesize the complementary RNA. The FastStart Essential DNA Green Master (Roche 53510700) was utilized to perform real-time quantification. The relative expression levels of target genes were normalized by GAPDH and estimated using the $2^{-\Delta\Delta C_t}$ method. The PCR primers are listed in Additional file 7: Table S2.

Statistical analysis

Statistical analyses of all data utilized in our study were completed by R software (version 4.0.3). 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. Univariate and multivariate Cox proportional hazards regression model was used to identify independent predictors of OS. All statistical tests were two-sided and $P < 0.05$ was considered to be statistically significant.

Results

Differential expressions of IRGs and TFs in CCA

The main design and process of this study are shown in the following figure (Fig. 1). We used RNA-seq data of 45 samples, including 36 CCA and 9 non-tumor tissues, collected from TCGA-CHOL. Totally, 1755 DEGs were identified, which consist of 743 significantly down-regulated genes and 1012 significantly up-regulated genes, for subsequent bioinformatics analysis (Fig. 2A and 2B).

The mRNA levels of 2498 IRGs (ImmPort database) and 318 TFs (Cistrome Cancer database) were examined. This analysis eventually revealed 156 IRGs (73 upregulated and 83 downregulated) and 18 TFs (11 upregulated and 7 downregulated) according to the screening criteria. Then, the heatmap and volcano plot were visualized to show the expression pattern of differentially expressed IRGs (Fig. 2C and 2D) and TFs (Additional file 1: Figure S1A and B) between CCA and non-tumor tissues. The names of DEGs were listed on Additional file 8: Table S3.

Construction of TF regulatory network

To investigate the biological roles and pathways of the DEGs, we found that top enriched GO terms were associated with regulation of signaling in immune and inflammatory response, such as "immune response", "inflammatory response", and "cytokine activity" (Fig. 2E). The significant pathways identified by KEGG pathway analysis included "cytokine- cytokine receptor interaction", "natural killer cell mediated cytotoxicity" and "chemokine signaling pathway".

Furthermore, GO enrichment analyses of differentially expressed IRGs demonstrated that the differentially expressed IRGs were mainly primarily participated in the pathway about tumor immunity, such as "regulation of immune effector process" and "regulation of cell killing" (Fig. 2F).

Identification of prognostic immune-related genes signature

For observing the association between IRGs and OS, we further conducted a Cox regression analysis to screen the IRGs. We found 11 prognostic IRGs (RAET1E, CST4, CCL24, CCK, CGB, GUCA2A, TDGF1, TDGF3, THPO, AVPR1B and IL9R) that were significantly related to survival of CCA patients in the training group (Fig. 3A). To further explore potential regulatory mechanisms behind deregulation of 11 prognostic IRGs in CCA, we analyzed the correlation between TFs and IRGs expression, as showed on Additional file 1: Figure S1C.

To investigate significance of risk genes in estimating prognosis, the 11 survival-related IRGs were further submitted to Lasso method and multivariate Cox regression analysis (Fig. 3B-C; Table 1). Finally, only 5 candidates prognostic IRGs (RAET1E, CST4, TDGF1, AVPR1B and IL9R) which may serve as significant predictors were obtained for 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). In the TCGA training cohort, the expression distribution of five IRGs and K-M curves for OS were shown in Additional file 2: Figure S2.

A prognostic index was established based on the expression levels and estimated regression coefficients of the risk genes. The formula of prognostic model containing five IRGs 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).

Analysis of the immune-related signature in the training set

The survival status of each patient in the training cohort 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. 4A and 4B). Among high-risk patients, three risk genes (RAET1E, CST4 and TDGF1) were upregulated, while two protective genes (AVPR1B and IL9R) were downregulated (Fig. 4C). In low-risk patients, these genes displayed an opposite expression pattern. The survival risk curve showed that patients with high risk had a significantly shorter OS time than those with low risk ($P < 0.001$, Fig. 4D). Interestingly, we also explored the difference of DFS between two groups, and found the similar results (Fig. 4E). Then ROC analysis revealed reliable performance in the survival prediction of the model, with the AUC value of ROC at 1-, 3-, 5-year were 0.920, 0.966 and 0.940, respectively (Fig. 4F).

Analysis of the immune-related signature in the validation set

The GSE107943 cohort including 30 CCA samples was used for validation of prognostic model. The risk score distribution, survival status and risk gene expression in the validation set were displayed in Fig. 5A-C. The K-M curves demonstrated that the OS ($P < 0.05$; Fig. 5D) and DFS ($P < 0.01$; Fig. 5E) was significantly poorer in high-risk group than low-risk group. The AUC at 1-, 3-, 5-year in the validation set were 0.689, 0.713 and 0.710, respectively (Fig. 5F). Taking together, these results indicated the risk score could well assess patients' prognosis. K-M analysis of OS among patients with different expression of each IRGs were displayed in Additional file 3: Figure S3.

Independent prognostic value of the risk model

Next, combined with other clinicopathological factors, we performed Cox regression analysis to assess performance of our model, such as age, gender and AJCC stage. The univariate analysis and multivariate analysis revealed that risk score was an independent risk factor in the TCGA cohort (HR=16.459, 95%CI, 3.923-69.046, $P < 0.001$) (Table 2).

Nomogram 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 and AJCC stage) (Fig. 6A). The nomogram could accurately predict OS, and demonstrated that risk score contributed much more risk points than other variables. C-index for OS was 0.814, and calibration plot of internal verification for probability of survival at 1-, 2- and 3-year showed better conformity between the predicted and realistic observation results (Fig. 6B-D).

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

To determine whether our IRGs signature could reflect the proportion of TIICs on tumor immune microenvironment, we investigated the correlation between risk score and 22 immune cells in CCA patients. The difference on proportion of 22 TIICs types between high- and low- risk group were investigated (Fig. 7A and Additional file 4: Figure S4). As shown on the violin plot, we found the low- risk group had more infiltrated macrophages M1 compared to high- risk group. In addition, we also analyzed the association between immune cells subpopulations, which indicated that memory resting CD4+ T cell is negatively related to CD8+ T cell, while active NK cell is negatively related to resting NK cell in CCA patients (Fig. 7B). To analyze whether prognostic IRGs might contribute to TIICs, we analyzed correlation between IRGs and tumor-immune cell proportions (Additional file 5: Figure S5). Interestingly, we found the expression of RAET1E is positively correlated to the proportion of B cells naïve; TDGF1 is positively correlated to memory active CD4+ T cells. IL9R is positively correlated to follicular helper T cells, and negatively correlated to M2 Macrophages (Fig. 7C-F).

Drug Sensitivity prediction

We used GDSC database to predict the likelihood of response to several common chemotherapy and targeted drugs. We estimated the IC50 of each sample and observed a significant difference of IC50 between high-risk and low-risk groups among anticancer drugs listed on the GDSC database. Patients in the high-risk group were more sensitive to commonly administered chemodrugs ($P = 0.037$ for AMG 706, $P = 0.019$ for Thapsigargin) (Fig. 8A-D). In contrast, the chemotherapeutic response of cisplatin and gemcitabine was not significantly different between both groups.

External cohort validation for 5- IRGs signature

Finally, the 5-gene immune related signature was also validated in our cohort. Compared with adjacent normal tissues, the expressions of RAET1E, CST4, AVPR1B and IL9R were significantly higher (Fig. 8E),

while TDGF1 were significantly lower in CCA samples.

Discussion

Activation of immune system has been widely proven to be a pivotal factor on tumorigenesis and metastasis [22]. With development of tumor immunotherapy in recent years, the function of immune system on cancer growth and metastasis has been addressed, in which immune cells participate in tumor biological activity by dysregulating expression of IRGs [23, 24]. Hence, IRGs may be an important predictor for prognosis of CCA patients. Although some considerable efforts had been made to identify prognostic model for CCA, the reports on prognostic role of IRGs signature is lacking [25]. In our study, we identified a new prognostic IRGs signature and built a dependable model to predict survival for CCA patients.

In order to obtain a reliable prognostic model, we used LASSO and multivariate Cox regression analyses. Five IRGs (AVPR1B, CST4, TDGF1, RAET1E and IL9R) were identified as robust prognostic gene signature on CCA patients. The reliability of our model was further validated by GSE107943 cohort. The results demonstrated that the model had better discriminating power for risk stratification, and showed prominent performance for prediction survival. After adjusting to other clinical factors, risk score generated by our model was demonstrated to be an independent prognostic factor. Finally, a nomogram model based on prognostic IRGs signature and clinical features was constructed to predict the survival rates of CCA patients. The immune risk score has the greatest effect on prognosis. Our results indicated that the risk model could act as an effective biomarker for survival prediction of CCA patients.

The functional enrichment analysis suggests that the IRGs are widely involved in the tumor immunological process, such as “regulation of immune effector process” and “regulation of cell killing”. Retinoic Acid Early Transcript 1E (RAET1E) is one important component of the RAET1 family, which consists of major histocompatibility complex (MHC) class I-related genes. RAET1E binds and activates killer cell lectin like receptor K1 (KLRK1), mediating natural killer cell cytotoxicity. Eisele et al found that RAET1E participated in glioma immune escape in glioblastoma [26]. Interleukin-9 (IL-9) is a T cell cytokine that acts through a γ C-family receptor on target cells and is associated with inflammation and allergy. One study demonstrated that IL9R showed suppressed tumor growth on melanoma mice, and administration of recombinant IL-9 (rIL-9) inhibited melanoma as well as lung carcinoma growth[27]. In our study, we also found that IL9R was a protective prognostic factor for CCA patients, which is consistent with previous study. As for Cystatin 4 (CST4), some studies demonstrated that CST4 enhanced gastric cancer (GC) and colorectal cancer (CRC) aggressiveness, and found that blood biomarkers of CST4 have enormous potential in terms of clinical diagnosis in GC and CRC [28, 29]. Teratocarcinoma-Derived Growth Factor 1 (TDGF1) is a member of the epidermal growth factor family, which had been regarded as a predictive marker for metachronous metastasis in CRC patients [30].

Recently, some studies have demonstrated that immune infiltration was a crucial determinant of responsiveness to immunotherapy and prognosis prediction of CCA [31, 32]. The immune milieu on tumor

microenvironment may account to the difference of survival outcome between low- and high- risk group. We found that M1 macrophages were the most infiltrated immune cells in low- risk compared with high-risk group, but the populations of M2 macrophage between two groups had no significant difference. Tumor-associated macrophages (TAMs) are one of the main components of the tumor immune contexture and have a major role in cancer progression and remodeling of the microenvironment [33]. Macrophages have different subtypes, which are differentiated under different cytokines. According to their activation, macrophages have three subtypes (M0, M1 and M2), and each subtype serves different immune functions. M1 macrophages secrete IL-12, IL-16, INF- γ and other pro-inflammatory cytokines to activate the inflammatory response and also participate in the host innate immunity, killing tumor cells in the TME. Our results indicated that M1 tumor infiltrated macrophages may exhibit anti-tumor roles on CCA.

GDSC database is a useful tool for predicting the chemotherapeutic response of targeted drugs for the patients from different risk groups. To determine small molecule drugs with potential roles to treat CCA, we observed that AMG 706 and Thapsigargin (TG) may have potential therapeutic efficacy against CCA with high risk. In contrast, cisplatin and gemcitabine haven't been showed significant difference between both groups. AMG 706 is an orally administered, small-molecule angiogenesis inhibitor that acts as a multi-targeted tyrosine kinase inhibitor of VEGFR-1, -2 and -3, PDGF receptor (PDGFR), and stem cell factor receptor[34]. AMG 706 has also exhibited anticancer activity on clinical trial in solid tumors including low-grade neuroendocrine tumors and NSCLC[35, 36]. Of which, TG is a non-competitive inhibitor of endoplasmic-reticulum Ca ($^{2+}$)-ATPase pump [37]. Moreover, TG can directly induce Estrogen receptor (ER) stress and decline the cell stemness[38]. The treatment of TG could inhibit cell viability via mediating NF-kappaB translocation and PI3K pathway[39]. Therefore, we concluded that these identified chemotherapy drugs might have potential to treat CCA.

There have been multiple studies for the value of IRG models to predict prognosis in various cancers, such as clear cell renal cell carcinoma [40], hepatocellular carcinoma [41] and cervical cancer [42]. 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. But some limitations should be addressed. The samples included in our study has limited sample size, and the results of our study were only validated in GEO database. The model requires more data support from clinical patients. Next, the potential molecular mechanisms of IRGs impact on CCA are not fully elucidated, which require further in vivo and in vitro experiments to investigate.

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. Thus, our study demonstrated the potential clinical application for survival prediction of CCA patients.

Declarations

Ethics approval and consent to participate

The research was approved by the Institutional Research Ethics Committees of the First Affiliated Hospital of Xi'an Jiaotong University. Informed consent for publication was obtained from all patients for collection of tissue samples prior to the surgery.

Consent for publication

Not applicable.

Availability of data and materials

All data used in this study were acquired from The Cancer Genome Atlas (<https://portal.gdc.cancer.gov/>) (up to June 10, 2021) and Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/gds>) (up to June 10, 2021).

Competing interests

All authors declare no conflict of interests.

Funding

This work was supported by the National Natural Science Foundation of China, No. 62076194, the Key Research and Development Plan of Shaanxi Province (No. 2017KW-060), the General Project of Natural Science Basic Research Plan of Shaanxi Province (No. 2020SF-070), and Clinical Research Fund of the First Affiliated Hospital of Xi'an Jiaotong University (XJTU1AF-CRF-2018-022). These funds were used for design of the study, collection and interpretation of data, bioinformatic analysis, and writing of the manuscript. We also thank the Biobank of First Affiliated Hospital of Xi'an Jiaotong University for providing clinical samples.

Authors' contributions

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

Acknowledgements

RZ and CC contributed equally to this work and should be considered co-first authors.

References

1. 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(5):261-280.
2. 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(1):235-243.
3. Lamarca A, Hubner RA, David Ryder W, Valle JW: Second-line chemotherapy in advanced biliary cancer: a systematic review. *Ann Oncol* 2014, 25(12):2328-2338.
4. Fritz JM, Lenardo MJ: Development of immune checkpoint therapy for cancer. *J Exp Med* 2019, 216(6):1244-1254.
5. Camidge DR, Doebele RC, Kerr KM: Comparing and contrasting predictive biomarkers for immunotherapy and targeted therapy of NSCLC. *Nat Rev Clin Oncol* 2019, 16(6):341-355.
6. Havel JJ, Chowell D, Chan TA: The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat Rev Cancer* 2019, 19(3):133-150.
7. 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). *The oncologist* 2018, 23(12):1407-e1136.
8. 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.
9. 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(1):250-256.
10. 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* 2020,111(2):323-333.
11. 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.* 2020, 65(2):480-488
12. 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(1):77.
13. 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(10):3197-3203.

14. 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(16):4678-4687.
15. 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.
16. Fridman WH, Zitvogel L, Sautes-Fridman C, Kroemer G: The immune contexture in cancer prognosis and treatment. *Nat Rev Clin Oncol* 2017, 14(12):717-734.
17. 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.
18. 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(2):227-240.
19. Ahn KS, O'Brien D, Kang YN, Mounajjed T, Kim YH, Kim TS, Kocher JA, Allotey LK, Borad MJ, Roberts LR et al: Prognostic subclass of intrahepatic cholangiocarcinoma by integrative molecular-clinical analysis and potential targeted approach. *Hepatol Int* 2019, 13(4):490-500.
20. 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(5):453-457.
21. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, Bindal N, Beare D, Smith JA, Thompson IR et al: Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* 2013, 41(Database issue):D955-961.
22. Upadhyay S, Sharma N, Gupta KB, Dhiman M: Role of immune system in tumor progression and carcinogenesis. *J Cell Biochem* 2018, 119(7):5028-5042.
23. Kalbasi A, Ribas A: Tumour-intrinsic resistance to immune checkpoint blockade. *Nat Rev Immunol* 2020, 20(1):25-39.
24. 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(1):155.
25. 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(5):869-881.
26. Eisele G, Wischhusen J, Mittelbronn M, Meyermann R, Waldhauer I, Steinle A, Weller M, Friese MJBajon: TGF-beta and metalloproteinases differentially suppress NKG2D ligand surface expression on malignant glioma cells. *Brain* 2006, 129:2416-2425.
27. Purwar R, Schlapbach C, Xiao S, Kang HS, Elyaman W, Jiang X, Jetten AM, Houry SJ, Fuhlbrigge RC, Kuchroo VK et al: Robust tumor immunity to melanoma mediated by interleukin-9-producing T cells. *Nature medicine* 2012, 18(8):1248-1253.
28. Zhang Y, Zhang J, Song H, Li DJA: Overexpression of CST4 promotes gastric cancer aggressiveness by activating the ELFN2 signaling pathway. *Am J Cancer Res* 2017, 7(11):2290-

2304.

29. Dou Y, Lv Y, Zhou X, He L, Liu L, Li P, Sun Y, Wang M, Gao M, Wang CJO et al: Antibody-sandwich ELISA analysis of a novel blood biomarker of CST4 in gastrointestinal cancers. *Onco Targets Ther* 2018, 11:1743-1756.
30. Miyoshi N, Ishii H, Mimori K, Sekimoto M, Doki Y, Mori MJ, Ijoo T: TDGF1 is a novel predictive marker for metachronous metastasis of colorectal cancer. *Int J Oncol* 2010, 36(3):563-568.
31. 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(2):470-478.
32. 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 management and research, Cancer Manag Res* 2018, 10:4113-4123.
33. Paillet J, Kroemer G, Pol JJ, Coig: Immune contexture of cholangiocarcinoma. *Curr Opin Gastroenterol* 2020, 36(2):70-76.
34. Polverino A, Coxon A, Starnes C, Diaz Z, DeMelfi T, Wang L, Bready J, Estrada J, Cattley R, Kaufman S et al: AMG 706, an oral, multikinase inhibitor that selectively targets vascular endothelial growth factor, platelet-derived growth factor, and kit receptors, potently inhibits angiogenesis and induces regression in tumor xenografts. *Cancer research* 2006, 66(17):8715-8721.
35. Lubner S, Feng Y, Mulcahy M, O'Dwyer P, Giang GY, Hinshaw JL, Deming D, Klein L, Teitelbaum U, Payne J et al: E4206: AMG 706 and Octreotide in Patients with Low-Grade Neuroendocrine Tumors. *The oncologist* 2018, 23(9):1006-e1104.
36. Coxon A, Ziegler B, Kaufman S, Xu M, Wang H, Weishuhn D, Schmidt J, Sweet H, Starnes C, Saffran D et al: Antitumor activity of motesanib alone and in combination with cisplatin or docetaxel in multiple human non-small-cell lung cancer xenograft models. *Mol Cancer* 2012, 11:70.
37. Arita Y, Ito T, Oono T, Kawabe K, Hisano T, Takayanagi R: Lysophosphatidic acid induced nuclear translocation of nuclear factor-kappaB in Panc-1 cells by mobilizing cytosolic free calcium. *World J Gastroenterol* 2008, 14(28):4473-4479.
38. Tadros S, Shukla SK, King RJ, Gunda V, Vernucci E, Abrego J, Chaika NV, Yu F, Lazenby AJ, Berim L et al: De Novo Lipid Synthesis Facilitates Gemcitabine Resistance through Endoplasmic Reticulum Stress in Pancreatic Cancer. *Cancer research* 2017, 77(20):5503-5517.
39. Jie Y, Peng W, Li YY: Identification of novel candidate biomarkers for pancreatic adenocarcinoma based on TCGA cohort. *Aging* 2021, 13(4):5698-5717.
40. 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(1):101.
41. 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(3):277-294.

42. 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(12):e1659094.

Tables

Table 1 The multivariate Cox regression analysis of key immune-related genes

IRGs	Coef	HR	95% CI	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

Coef regression coefficient generated from multivariate Cox analysis, *HR* hazard ratio, *CI* confidence interval

Table 2 Cox proportional hazards model analysis of prognostic factors

Variables	Univariate analysis			Multivariate analysis		
	HR	95% CI	P value	HR	95% CI	P value
Age (<65 vs. ≥65)	1.268	0.499-3.221	0.617	0.742	0.249-2.209	0.592
Gender (male vs. female)	1.387	0.544-3.534	0.494	1.010	0.364-2.798	0.985
AJCC stage (III + IV vs. I + II)	1.475	0.517-4.207	0.467	3.488	0.954-12.751	0.059
Risk score	10.929	3.073-38.865	2.20E-04	16.459	3.923-69.046	1.29E-04

**P* value <0.05 was considered statistically significant, *HR* hazard ratio, *CI* confidence interval

Figures

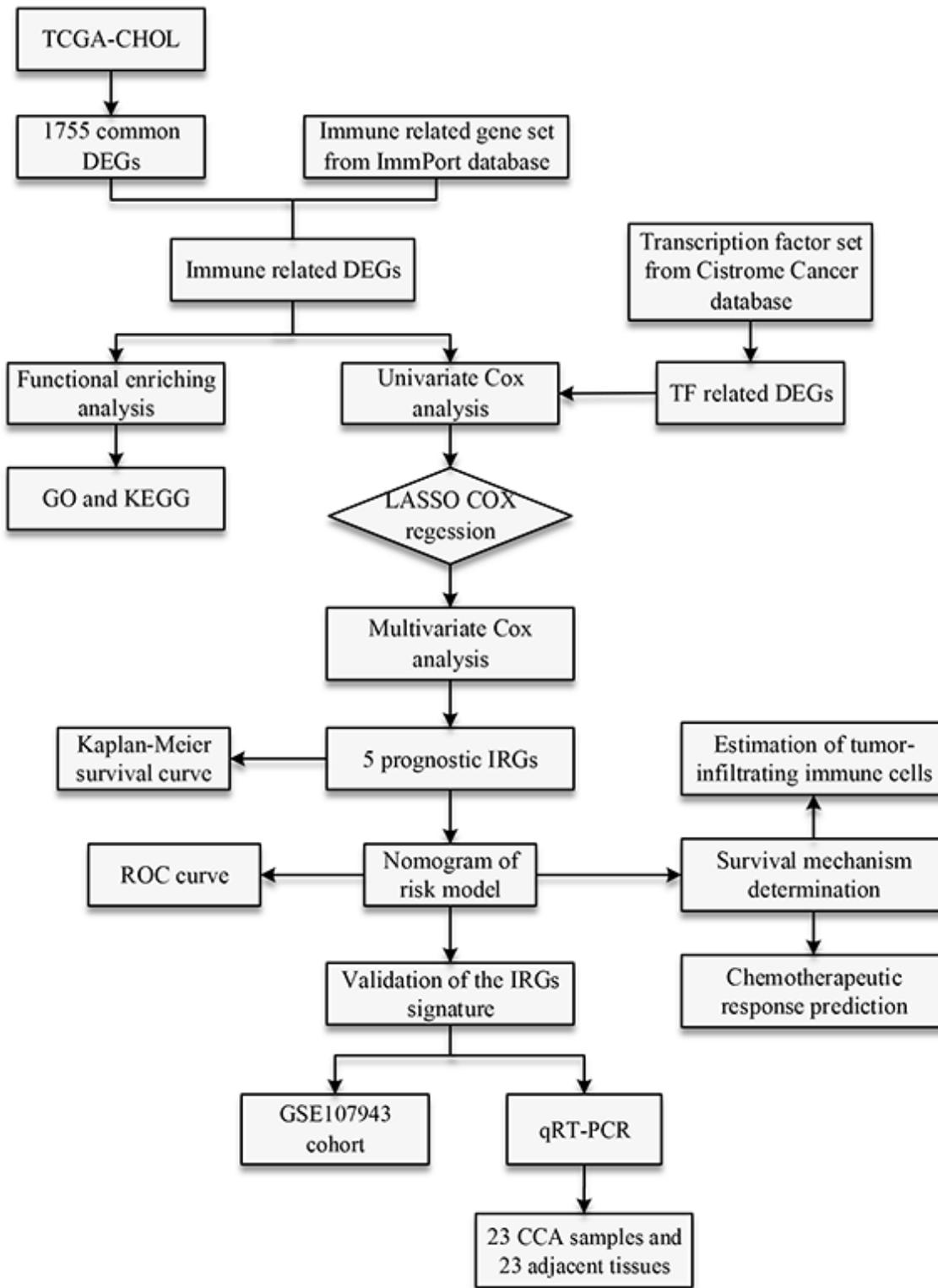


Figure 1

Flow chart of the main procedures of this study.

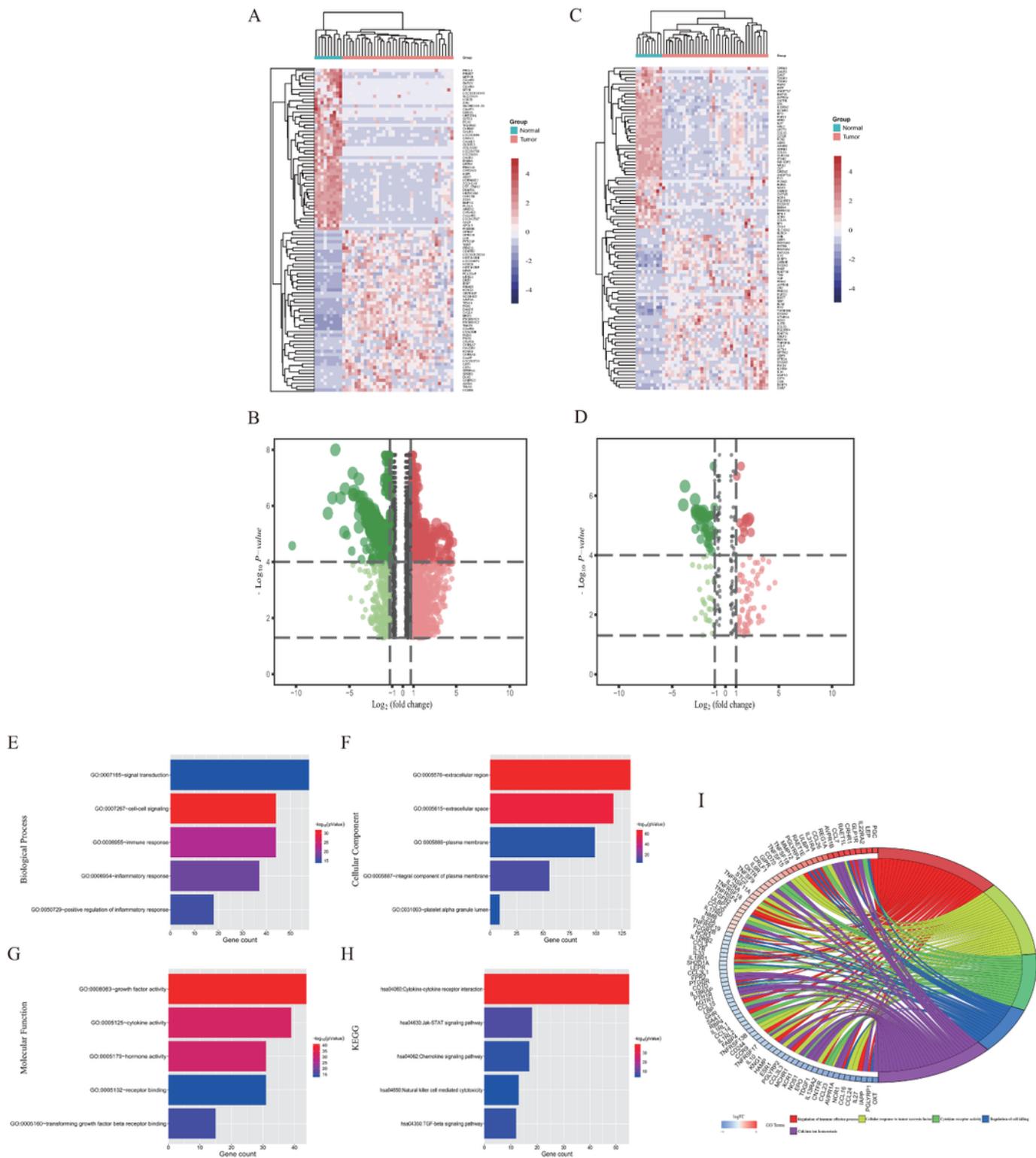


Figure 2

Differential expression of IRGs and functional enrichment analysis of DEGs in CCA. Heatmaps and volcano plots of the differential expressed genes (DEGs) between CCA and non-tumor samples, including all differential expressed genes (A, B) and IRGs (C, D). 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. (E) Gene ontology analysis of all DEGs: BP, CC, MF and KEGG. (F) GO enrichment analyses of differentially expressed IRGs.

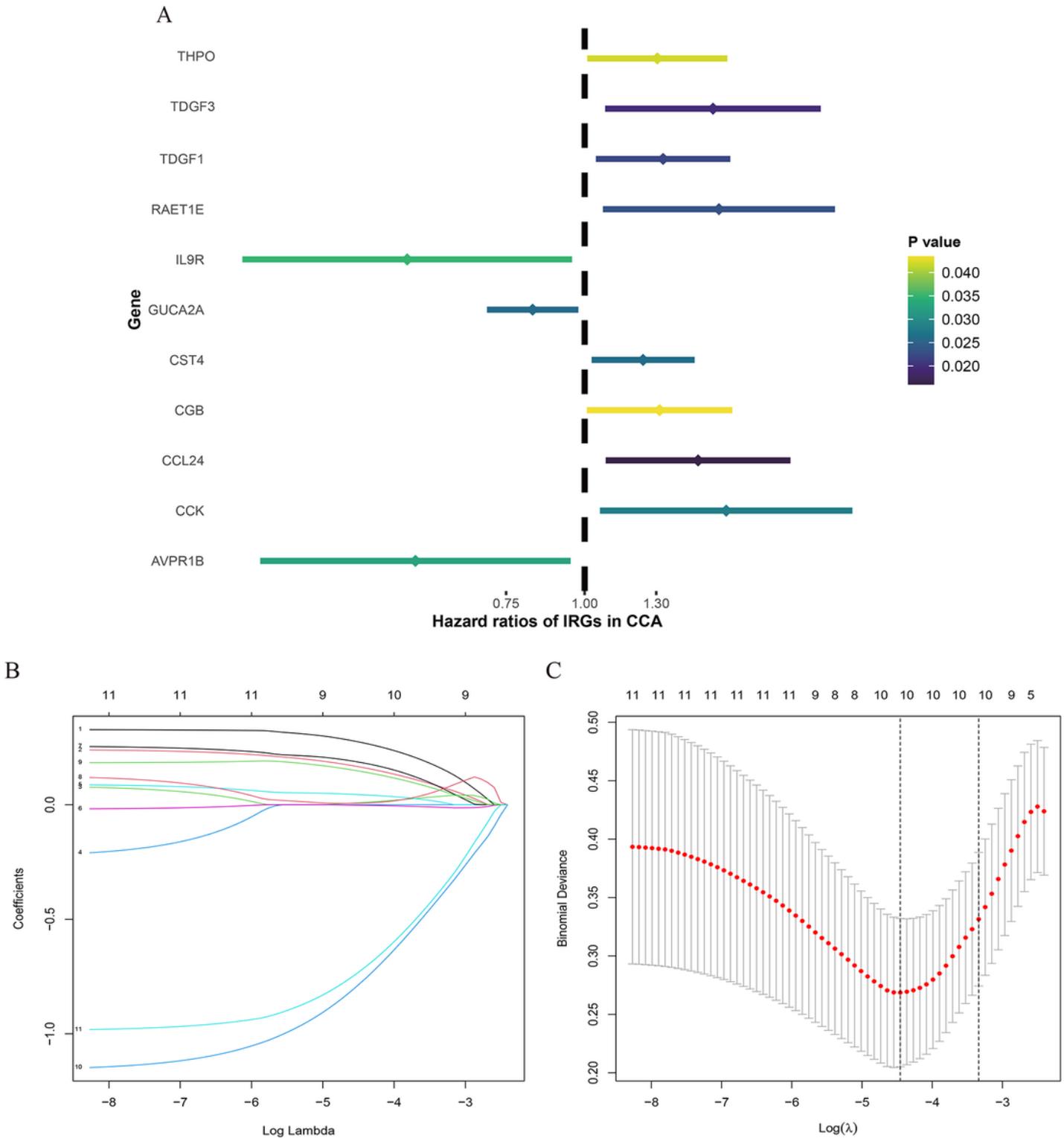


Figure 3

Univariate Cox and Lasso regression was used to screen the IRG signature for the predictive model. (A) The forest plot of 11 prognostic IRGs screened out by univariate Cox proportional hazards regression. (B)

Lasso coefficient profiles of IRGs. (C) A coefficient profile plot was generated to find the optimal parameter (lambda).

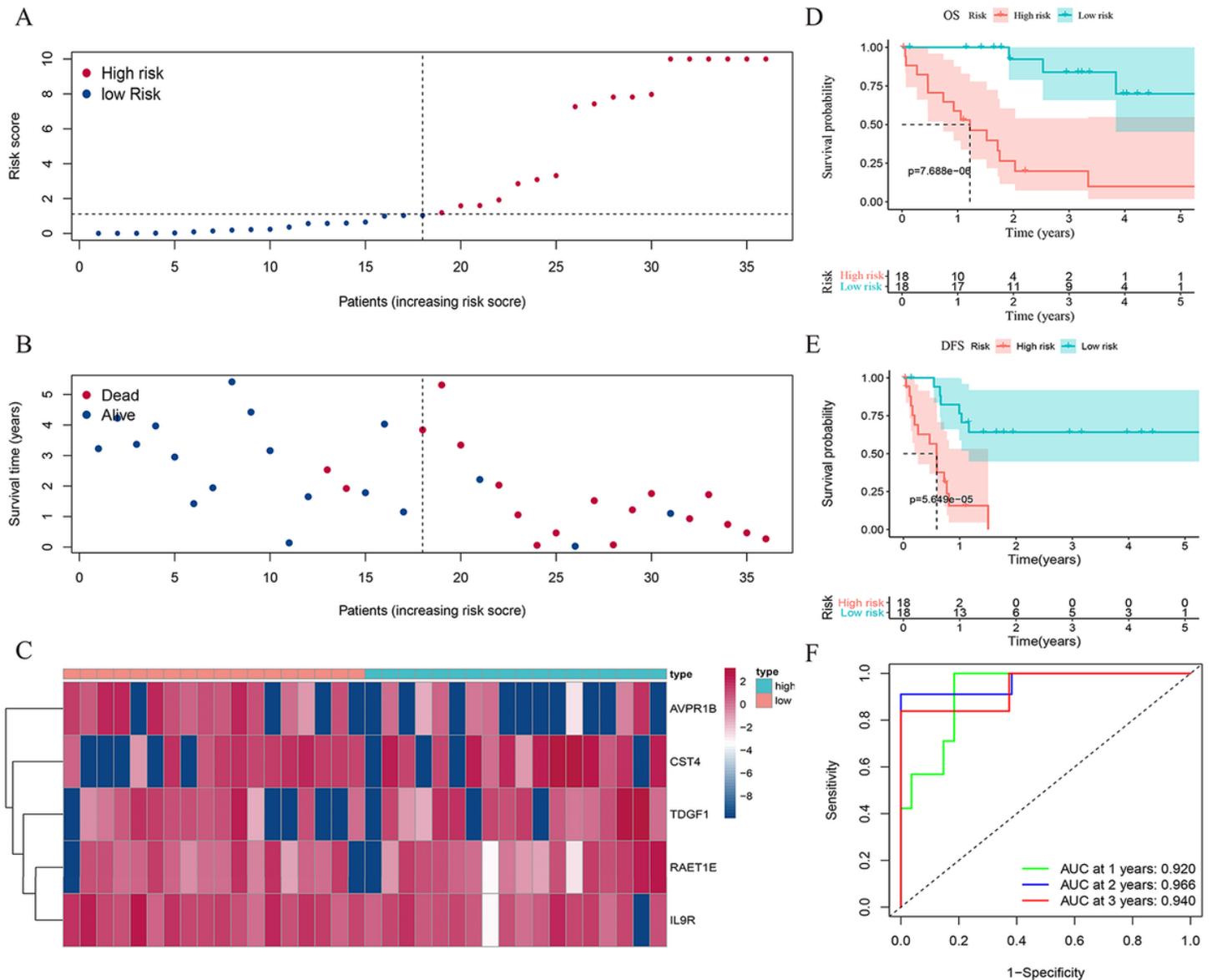


Figure 4

Construction of the prognostic model in the training group. Risk score analysis, survival and ROC curves for the IRGs signature in training group. (A) Scatterplots of patients with different survival status. (B) Risk score distribution of patients with different risks (low, green; high, red) (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, E) Kaplan-Meier curve showing OS and DFS according to high- and low-risk group. (F) ROC analysis of the model consisted of IRGs signature. ROC, Receiver operating characteristic; AUC, Area under curve.

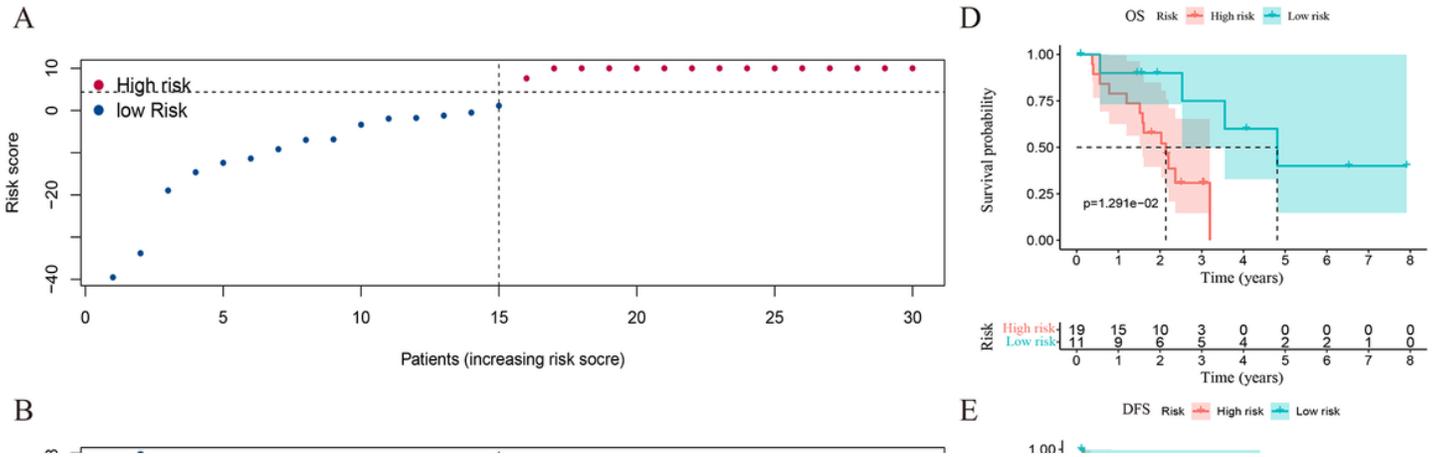


Figure 5

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) Scatterplots of patients with different survival status. (B) Risk score distribution of patients with different risks. (C) Heatmap of five IRGs expression. (D, E) Kaplan-Meier curve of OS and DFS according to high- and low-risk group. (F) ROC analysis of the model.

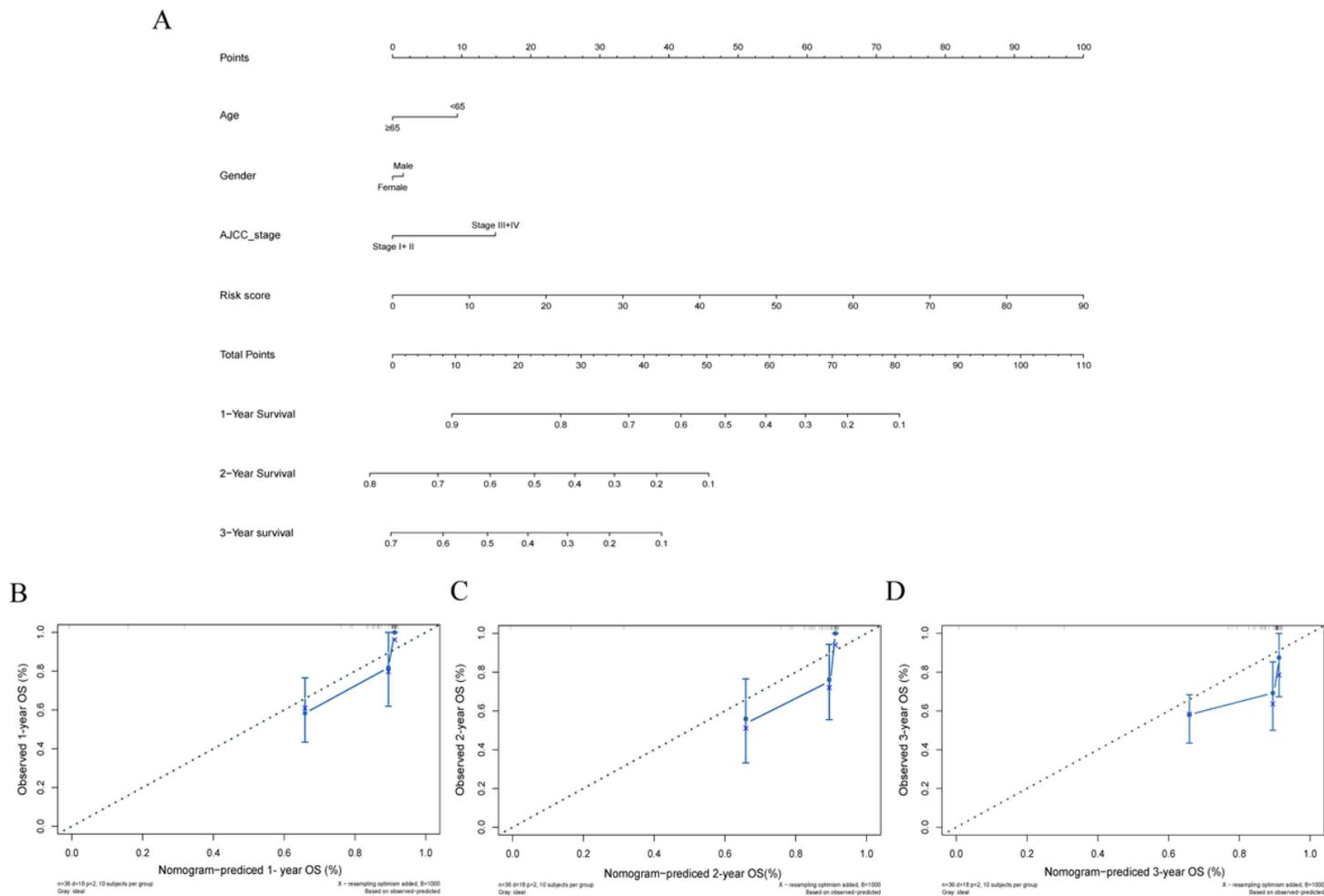


Figure 6

Independent prognostic value of the risk model, nomogram for the prediction of prognosis in the TCGA cohort. (A) The nomogram based on the IRGs signature and clinical information to predict the 1-, 2- and 3-year survival of patients. (B, C, D) Calibration curves evaluating the predictive accuracy of nomogram at 1-, 2- and 3-year survival.

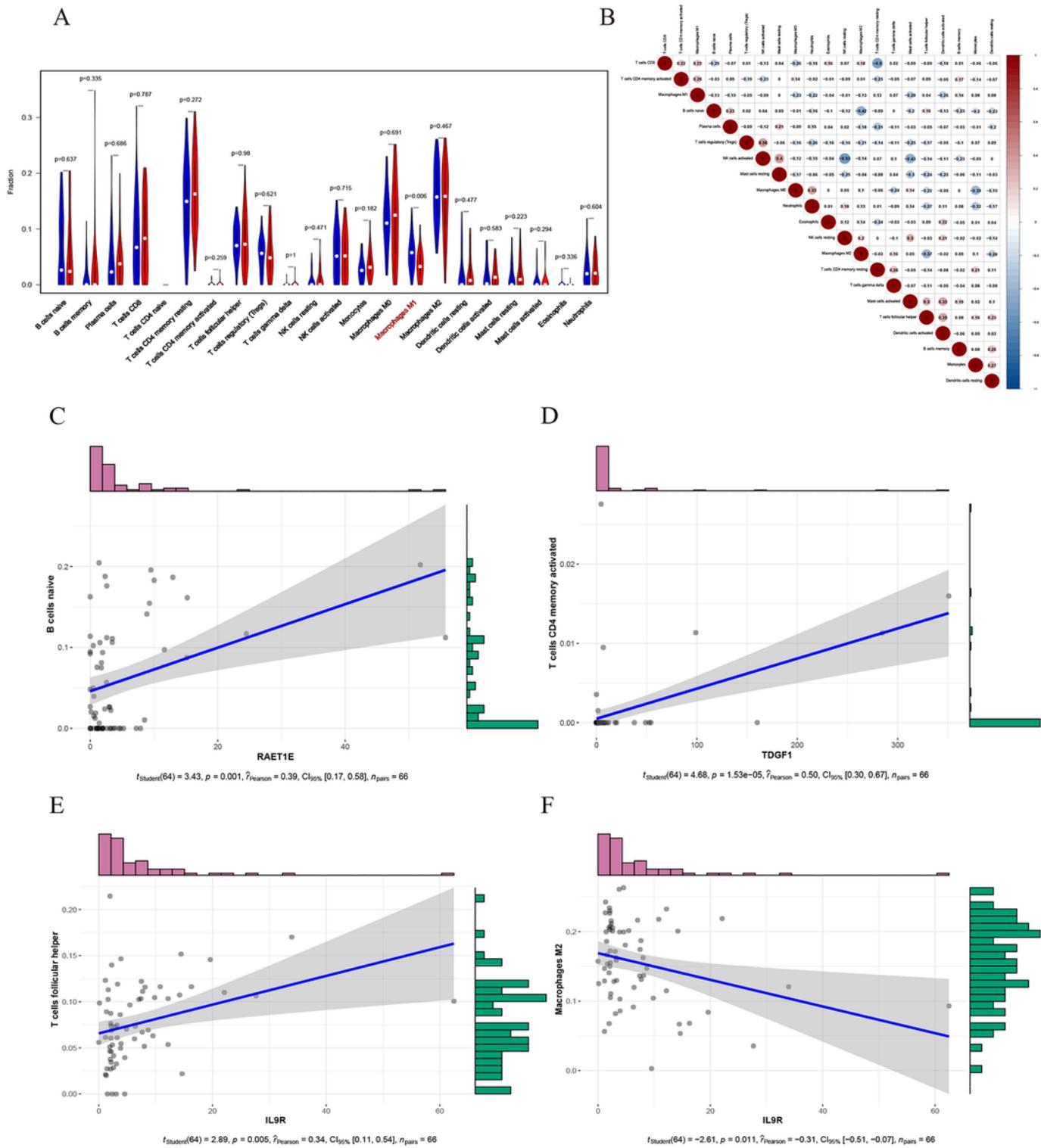


Figure 7

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) Difference of TIICs between high- and low- risk group. (B) The correlation matrix of immune cells subpopulations. (D-F) The significant correlation between the IRGs and tumor-infiltrating immune cells.

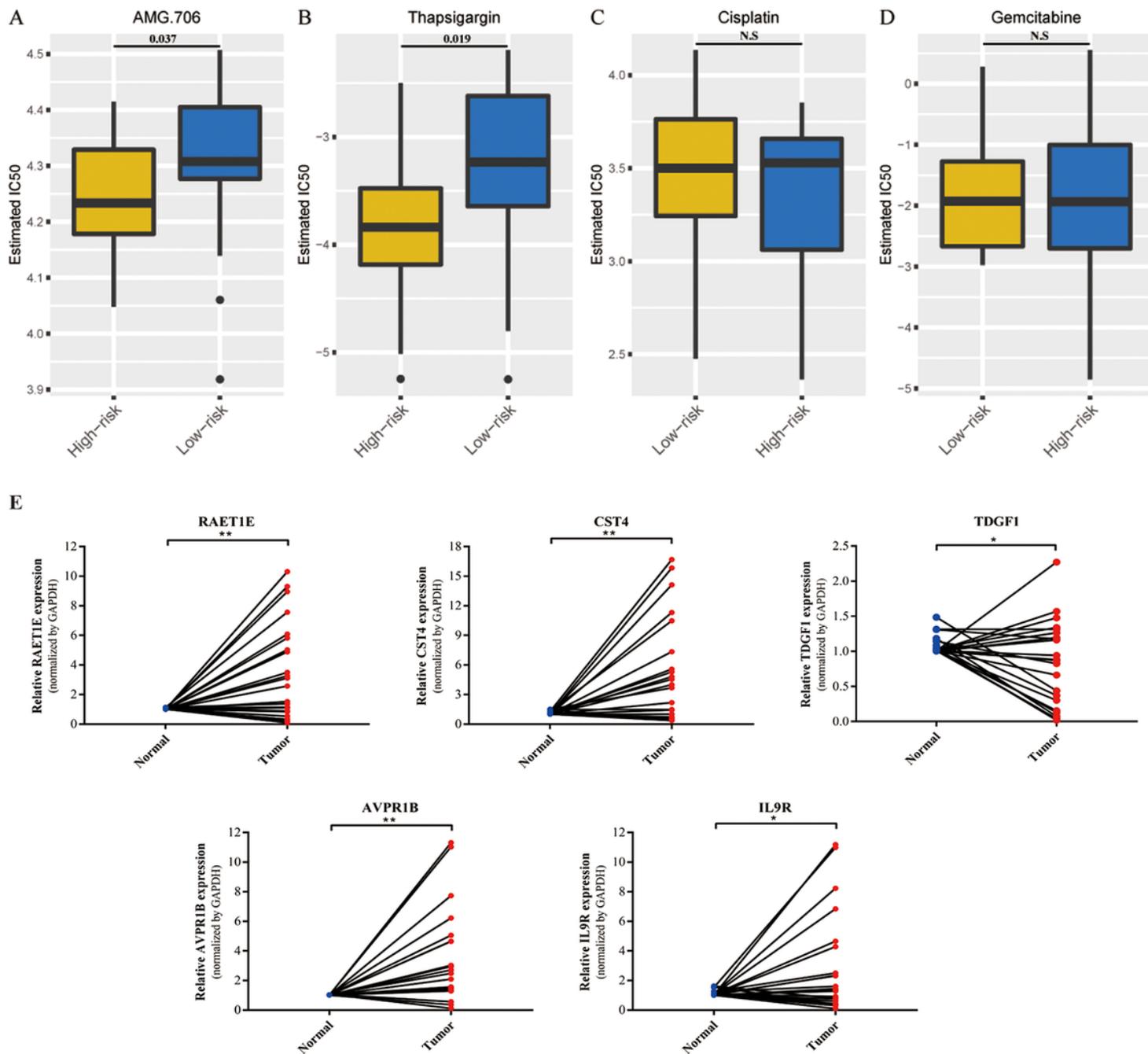


Figure 8

Chemotherapeutic response prediction and external cohort validation for 5-IRGs signature. (A-D) Chemotherapeutic response in the high-risk and low-risk groups. (E) Validation of five-gene prognostic signature by qRT-PCR.

Supplementary Files

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

- [FigureS1.tif](#)

- [FigureS2.tif](#)
- [FigureS3.tif](#)
- [FigureS4.tif](#)
- [FigureS5.tif](#)
- [SupplementaryTable1.xlsx](#)
- [SupplementaryTable2.xlsx](#)
- [SupplementaryTable3.xlsx](#)