Development and validation of immune-related genomics nomogram for prognostic prediction in left- and right-side colorectal cancer

Mengye Niu
The First Hospital of Hebei Medical University

Chengyang Chen
The First Hospital of Hebei Medical University

Wei Li
The First Hospital of Hebei Medical University

Jianwei Ma
The First Hospital of Hebei Medical University

Xing Zhang
The First Hospital of Hebei Medical University

Chao Zhang
The First Hospital of Hebei Medical University

Xia Jiang
The First Hospital of Hebei Medical University

Zengren Zhao (zzengren@163.com)
The First Hospital of Hebei Medical University

Zhongxin Li
The First Hospital of Hebei Medical University

Research Article

Keywords: Left-sided colon cancer, Right-sided colon cancer, Immune-related gene Biomarkers, nomogram, immune microenvironment, tumor mutation burden, immune checkpoint

Posted Date: October 28th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2203965/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background:

Previous studies have reported that the tumor heterogeneity and immune molecular mechanisms of proximal and distal colorectal cancer (CRC) are divergent. Therefore, our study aims to analyze the difference between left-sided CRC (LCC) and right-sided CRC (RCC), and respectively develop the nomograms based on prognostic immune-related genes for LCC and RCC.

Methods:

We enrolled 443 colon cancer patients (220 LCC patients and 223 patients) from The Cancer Genome Atlas (TCGA) datasets. Firstly, the differential expressed immune-related genes (DE-IRGs), overall survival (OS), and biological functions between LCC and RCC groups were identified. Then, we analyzed the differences between the two groups in the immune microenvironment, immune checkpoint, and tumor mutation burden (TMB). Next, the LCC and RCC data from TCGA dataset are randomly divided into training and internal validation sets at a 7:3 ratio respectively. Additionally, 566 colon cancer patients (342 LCC patients and 224 RCC patients) in the GSE39582 dataset were downloaded from the Gene Expression Omnibus (GEO) database as the external validation set. Then, survival and Lasso Cox regression analyses were applied to identify hub immune-related genes and respectively establish two prognostic gene signatures of LCC and RCC groups. The prognostic signatures were validated by the 10-fold cross-validation, internal validation set, and external validation set. Further, combined with clinical features, we constructed two clinical predictive nomograms and validated them.

Results:

RCC patients have lower survival than LCC. RCC patients have higher proportions of T cells CD8, T cells follicular helper, and lower macrophages M0, T cells CD4 naive. RCC patients have higher ESTIMATE and immune scores and lower tumor purity. The immune checkpoint expression levels and TMB values are higher in RCC patients than in LCC. We respectively selected 10 immune-related genes for LCC and 7 genes for RCC groups to develop and validate the prognostic model and calculate a risk score for each patient. The AUC values of the risk score for OS in LCC were 0.735 in the training set, 0.711 in the internal validation set, and 0.744 in the external validation set, and in RCC were 0.704 in the training set, 0.738 in the internal validation set, and 0.705 in the external validation set. The AUC values of the 10-fold cross-validation range from 0.564 to 0.808 in LCC and from 0.589 to 0.792 in RCC. The nomogram of LCC of RCC includes risk based on prognostic genes, age, pathological T, N, M, stage, and gender. the AUC values of the LCC nomogram were 0.722 in the training set, 0.696 in the internal validation set, and 0.739 in the external validation set, and of the RCC nomogram were 0.774 in the training set, 0.744 in the internal validation set, and 0.737 in the external validation set. We also found that were significantly different
between high- and low-risk patients in the immune score, ESTIMATE score, tumor purity, immune checkpoint expression levels, and TMB values.

Conclusions:

We found significant differences in the multidimensional insight between LCC and RCC patients in clinical features, DE-IRGs, TMB, immune checkpoint expression levels, and immune microenvironment landscape. Our study respectively established two prognostic nomograms based on DE-IRGs in combination with clinical features to provide a basis for personalized and precise treatment of LCC and RCC patients.

1. Introduction

Colorectal cancer (CRC) is the third most common cancer and the second leading cause of cancer death worldwide, posing a serious risk to people's health(1). Given the differences in embryonic origin, vascular and nervous supplies, macrobiotic burden, and main physiological functions of the left and right colons, tumor location is increasingly suggested to dictate tumor behavior affecting pathology, progression, and prognosis(2–4). Although clinical management of CRC generally has not taken into account the primary tumor site, studies have confirmed that left-sided colorectal cancers (LCC) and right-sided colorectal cancers (RCC) have significant differences in epidemiology, clinical and biological characteristics(5). In a systematic review and meta-analysis including 66 studies with more than 1.4 million patients, a significant prognostic impact of tumor site on overall survival was found with a 20% reduced risk of death for LCC(6). A large number of researchers have reported identifying differences between cancers based on location, including gene expression, DNA mutations, microbiota, methylation profiles, tumor microenvironment, and survival rates(2, 3, 6–8). In addition, tumor location also affected the outcome of adjuvant chemotherapy, palliative therapy, or targeted therapy. Therefore, it is of special significance to classify CRC by its location, which may contribute to the determination of clinical diagnosis and treatment.

At present, the main treatment method for CRC is the surgery-based comprehensive treatment, but the therapeutic effect for some patients with unresectable or metastatic CRC is very limited, which is the main reason for the poor prognosis of colorectal cancer(9–12). Immunotherapy utilizes the body's immune system to play a specific and lasting anti-cancer role and has been widely studied and regarded to be an effective treatment for various cancers(13, 14). The most commonly used are immune checkpoint inhibitors, specifically anti-PD-1/PD-L1 interaction, which has made great breakthroughs in the field of tumor immunotherapy in recent years(15–17). This illustrated that immunotherapy has the potential in the treatment of refractory or relapsed cancers, which could extend patients' survival time and improve their quality of life. Therefore, it is greatly significant for precision immunotherapy to explore the molecular characteristics of immunotherapy and search for immune prognostic biomarkers.
Nomograms are widely used for the prognosis of CRC patients. However, few previous studies have separately developed predictive models to predict patients’ prognoses by location. In the present study, we aim to separately develop predictive models of LCC and RCC to identify potential immune prognostic biomarkers. In addition, age, gender, and histological classification are other important factors that affect clinical outcomes and may enhance predictive power. Therefore, our study focused on analyzing the difference between LCC and RCC and respectively constructing a nomogram for each group, containing prognostic immune gene signatures and clinical prognostic factors, which will make more accurate predictions for the prognosis and understand the immune molecular mechanism of CRC, guiding precise personalized diagnosis and treatment.

2. Materials And Methods

2.1 Datasets

The transcriptome data and somatic mutation data of CRC patients were downloaded from The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/), which includes 443 CRC patients’ transcriptome data (220 LCC patients and 223 RCC patients) and 305 CRC patients’ somatic mutation data (112 LCC patients and 193 RCC patients). The patients with LCC and RCC in TCGA were divided into training and internal validation sets in the ratio of 7:3, respectively. Additionally, RNA expression profiles of 566 colon cancer patients (342 LCC patients and 224 RCC patients) in the GSE39582 dataset were downloaded from the Gene Expression Omnibus (GEO) database as the external validation sets (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE39582/). An overview of the steps is presented as a flow chart in Fig. 1.

RCC was composed of any histologically confirmed tumor arising from the caecum, ascending colon, hepatic flexure, or the proximal two-thirds of the transverse colon. While LCC was composed of any tumor arising in the splenic flexure, descending colon, sigmoid colon, or rectum.

2.2 Survival Analysis

Kaplan-Meier survival analysis was used to assess the survival differences between different clinicopathological characteristics, between the high- and low-risk groups, and between the LCC and RCC groups in the data sets mentioned above. The ‘survival’ package in R was used to perform a two-sided log-rank test and univariate and multivariate Cox regression analyses(18).

2.3 Differential Gene Analysis and Functional Annotation

The differential expression genes (DEGs) between LCC and RCC, LCC and L_normal, RCC and R_normal were identified by differential expression analysis using the “edgeR” package in R. |log2 FC (fold-change)| > 1 and P < 0.05 were set as the thresholds for screening DEGs. GO enrichment and KEGG pathway analyses were used to explore the potential biological processes, cellular components, and molecular functions of DEGs. A list of immune-related genes (IRGs) was obtained from the ImmPort database (https://immport.niaid.nih.gov), which is a platform for providing accurate and timely immunological
Differentially expressed immune-related genes (DE-IRGs) were obtained from the overlapping of DEGs and IRGs.

2.4 Gene set variation analysis (GSVA)

To reveal pathway enrichment between LCC and RCC patients, we used the “GSVA” package in R to evaluate the t score and assign pathway activity conditions. Moreover, R's “limma” package was also applied to display distinctions in pathway activation between LCC and RCC patients.

2.5 The Proportion of Immune Cell Infiltration and the Calculation of Tumor Purity

We used software CIBERSORT to calculate the relative proportions of 22 immune cells in each cancer sample. To run CIBERSORT, the following packages are required in R software: “e1071”, “parallel”, and “preprocessCore”. A file called “LM22.txt”, which contains a “signature matrix” of 547 genes, in R (obtained under Menu > Download from CIBERSORT web: https://cibersort.stanford.edu/download.php) is also required. Immune, stromal, and ESTIMATE scores and tumor purity were calculated by the ESTIMATE package in R software based on Yoshihara et al. We also compared the expression levels of the human leukocyte antigen (HLA) gene family and immune check-point genes to explore the potential immunotherapy mechanism in LCC and RCC.

2.6 Profiles of tumor mutation burden (TMB) and correlation analysis

The TMB was defined as: TMB = (total count of variants)/ (the whole length of exons). The ‘maftools’ package was used to illustrate the respective mutation profiling of the two group levels by waterfall plot. Afterward, the differential mutation frequencies of mutants were compared using the chi-square test between the two groups. Moreover, TMB was derived for each patient, and calculated with Pearson correlation analysis with estimated P values.

2.7 LASSO Cox Regression Analysis

LASSO Cox regression analysis with R package glmnet was used to further separately identify hub immune genes related to the prognosis of LCC or RCC, and the Risk Score of each sample was calculated using the screened hub genes through the following formula:

\[
\text{Risk Score} = \sum_{i=1}^{N} (\text{Expi} \times \text{Coef})
\]

“N” represents the number of signature genes, “Expi” the gene expression levels, and “Coef” the estimated regression coefficient value from the Cox proportional hazards analysis. Then the optimal cutoff value of the Risk score was determined by R package survival, survminer, and bilateral test, patients were divided into Low Risk and High-Risk groups according to the cutoff values. Moreover, model predictive power was
evaluated by calculating the AUC of 1-year, 3-year, 5-year, 7-year, and all time-dependent ROC curves using the “survivalROC” package.

2.8 Building and validating a predictive nomogram

First, univariate and multivariate Cox regression analyses were performed to identify the proper terms to build the nomogram. A forest plot was used to show the P-value, HR, and 95% CI of each variable through the ‘forest plot’ package in R.

The independent prognostic factors were used to build the nomogram by utilizing the ‘rms’, ‘nomogramEx’, and ‘ggDCA’ packages in R. Next, we estimated whether the predicted survival outcome was close to the actual outcome with calibration curves in the training set, internal validation, external validation, and complete set. Nomogram-predicted survival and the observed outcome were plotted on the x-axis and y-axis, respectively, and the 45° line represents the best prediction. Furthermore, the decision curve analyses (DCA) of the training set, internal validation, external validation, and complete set, which can assess and compare prediction models that incorporate clinical consequences, were used to measure whether our established nomogram was suitable for clinical utility. The x-axis indicates the percentage of threshold probability, and the y-axis represents the net benefit.

2.9 Pan-cancer Analysis

The expression of hub immune genes between various types of cancer and normal tissues were downloaded from the TCGA and GTEx database (https://commonfund.nih.gov/GTEx/). RNA sequencing data for patients with 33 types of cancers, including adrenocortical carcinoma (ACC), bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), cervical squamous cell carcinoma (CESC), cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), lymphoid neoplasm diffuse large B cell lymphoma (DLBC), esophageal carcinoma (ESCA), glioblastoma (GBM), brain lower grade glioma (LGG), head and neck squamous cell carcinoma (HNSC), kidney chromophobe (KICH), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), acute myeloid leukemia (LAML), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), mesothelioma (MESO), ovarian serous cystadenocarcinoma (OV), pancreatic adenocarcinoma (PAAD), pheochromocytoma and paraganglioma (PCPG), prostate adenocarcinoma (PRAD), rectum adenocarcinoma (READ), sarcoma (SARC), skin cutaneous melanoma (SKCM), stomach adenocarcinoma (STAD), testicular germ cell tumors (TGCT), thyroid carcinoma (THCA), thymoma (THYM), uterine corpus endometrial carcinoma (UCEC), uterine carcinosarcoma (UCS), and uveal melanoma (UVM), were obtained from the TCGA database. All expression data were normalized via log2 conversion.

2.10 Statistical Analysis

Statistical analysis was conducted with R software (version 3.6.3; http://www.Rproject.org). The reported statistical significance levels were all two-sided, with statistical significance set at 0.05.
3. Results

3.1 The Differences between LCC and RCC Patients

3.1.1 Differences in Characteristics of Demography between LCC and RCC Patients

Characteristics of demography in LCC and RCC patients of TCGA and GES39582 datasets were summarized in Table 1. The result illustrated statistically significant differences in pN (P = 0.016) and Age (P = 0.003) between the two groups of TCGA set, and in pT (P = 0.042) and Age (P < 0.001) between the two groups of GSE39582 set. Noteworthy, the overall survival of RCC was lower than LCC (Fig. 2A).
Table 1
Clinical features for the LCC and RCC patients in TCGA and GSE39582 datasets.

<table>
<thead>
<tr>
<th></th>
<th>TCGA dataset</th>
<th></th>
<th></th>
<th>GSE39582 dataset</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LCC patients (n = 220)</td>
<td>RCC patients (n = 223)</td>
<td>P value</td>
<td>LCC patients (n = 342)</td>
<td>RCC patients (n = 224)</td>
<td>P value</td>
</tr>
<tr>
<td>Age</td>
<td>0.003</td>
<td></td>
<td></td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 65</td>
<td>103 (46.8%)</td>
<td>73 (32.7%)</td>
<td></td>
<td>150 (43.9%)</td>
<td>60 (26.9%)</td>
<td></td>
</tr>
<tr>
<td>≥ 65</td>
<td>117 (53.2%)</td>
<td>150 (67.3%)</td>
<td></td>
<td>192 (56.1%)</td>
<td>163 (73.1%)</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>0.816</td>
<td></td>
<td></td>
<td>0.285</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>108 (49.1%)</td>
<td>106 (47.5%)</td>
<td></td>
<td>148 (43.3%)</td>
<td>108 (48.2%)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>112 (50.9%)</td>
<td>117 (52.5%)</td>
<td></td>
<td>194 (56.7%)</td>
<td>116 (51.8%)</td>
<td></td>
</tr>
<tr>
<td>pT</td>
<td>0.09</td>
<td></td>
<td></td>
<td>0.042</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>8 (3.64%)</td>
<td>5 (2.24%)</td>
<td></td>
<td>9 (2.76%)</td>
<td>3 (1.38%)</td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>38 (17.3%)</td>
<td>36 (16.1%)</td>
<td></td>
<td>34 (10.4%)</td>
<td>11 (5.07%)</td>
<td></td>
</tr>
<tr>
<td>T3</td>
<td>159 (72.3%)</td>
<td>151 (67.7%)</td>
<td></td>
<td>220 (67.5%)</td>
<td>147 (67.7%)</td>
<td></td>
</tr>
<tr>
<td>T4</td>
<td>15 (6.82%)</td>
<td>31 (13.9%)</td>
<td></td>
<td>63 (19.3%)</td>
<td>56 (25.8%)</td>
<td></td>
</tr>
<tr>
<td>pN</td>
<td>0.031</td>
<td></td>
<td></td>
<td>0.439</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N0</td>
<td>116 (53.0%)</td>
<td>138 (61.9%)</td>
<td></td>
<td>188 (57.3%)</td>
<td>114 (52.3%)</td>
<td></td>
</tr>
<tr>
<td>N1</td>
<td>67 (30.6%)</td>
<td>44 (19.7%)</td>
<td></td>
<td>79 (24.1%)</td>
<td>55 (25.2%)</td>
<td></td>
</tr>
<tr>
<td>N2</td>
<td>36 (16.4%)</td>
<td>41 (18.4%)</td>
<td></td>
<td>61 (18.6%)</td>
<td>49 (22.5%)</td>
<td></td>
</tr>
<tr>
<td>pM</td>
<td>0.156</td>
<td></td>
<td></td>
<td>0.198</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M0</td>
<td>180 (83.3%)</td>
<td>193 (88.5%)</td>
<td></td>
<td>300 (87.7%)</td>
<td>205 (91.5%)</td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td>36 (16.7%)</td>
<td>25 (11.5%)</td>
<td></td>
<td>42 (12.3%)</td>
<td>19 (8.48%)</td>
<td></td>
</tr>
<tr>
<td>pStage</td>
<td>0.17</td>
<td></td>
<td></td>
<td>0.245</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage I</td>
<td>36 (16.7%)</td>
<td>38 (17.4%)</td>
<td></td>
<td>24 (7.06%)</td>
<td>13 (5.83%)</td>
<td></td>
</tr>
<tr>
<td>Stage II</td>
<td>74 (34.3%)</td>
<td>94 (43.1%)</td>
<td></td>
<td>161 (47.4%)</td>
<td>101 (45.3%)</td>
<td></td>
</tr>
<tr>
<td>Stage III</td>
<td>70 (32.4%)</td>
<td>61 (28.0%)</td>
<td></td>
<td>113 (33.2%)</td>
<td>90 (40.4%)</td>
<td></td>
</tr>
<tr>
<td>Stage IV</td>
<td>36 (16.7%)</td>
<td>25 (11.5%)</td>
<td></td>
<td>42 (12.4%)</td>
<td>19 (8.52%)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Dates were displayed in counts (%). LCC: Left-side colon cancer; RCC: Right-side colon cancer; TCGA: The Cancer Genome Atlas.
3.1.2 Differentially expressed immune-related genes (DE-IRGs) and Functional Annotation between LCC and RCC Patients

By comparing the LCC and RCC patients’ transcriptome data, we identified 2653 significant DEGs between LCC and RCC groups, which were visualized in a volcano diagram and heatmap (Fig. 2B-C). In addition, the top 20 different of GSVA in the two groups was shown in Fig. 2D (|log2FC| > 0.2, all p < 0.05).

DE-IRGs were derived from the overlapping of DEGs and IRGs and finally we obtained 23 significant upregulated DE-IRGs in the LCC group and 110 significant upregulated DE-IRGs in the RCC group (Fig. 2E-F). Besides, we analyzed these DE-IRGs using the R software package “cluster Profiler” to identify the functions(20) PMID:22455463. LCC upregulated DE-IRGs were enriched in 73 GO terms and 1 KEGG pathway (FDR < 0.5). RCC upregulated DE-IRGs were enriched in 164 GO terms and 8 KEGG pathways (FDR < 0.5). We chose to show the top terms and pathways in Fig. 2G,2H.

3.1.3 Differential Immune microenvironment between LCC and RCC Patients

Comparing the immune microenvironment of LCC and RCC by two different approaches, we observed that there were significant differences in the components of immune infiltration between the two groups. In the LCC group, the proportions of B cells memory, T cells CD4 naïve, and macrophages M0 were significantly higher than in the RCC group (Wilcoxon test, all p < 0.05). Similarly, T cells CD8, and T cells follicular helper were significantly infiltrations in the RCC group (Wilcoxon test, all p < 0.05; Fig. 3A).

Comparing the stromal score, ESTIMATE score, immune score, and tumor purity of the two groups, we found that patients in the RCC group had a lower tumor purity than the LCC group (Wilcoxon test, p < 0.05). While the LCC group had a lower ESTIMATE score and immune score (Wilcoxon test, p < 0.05) than patients in the RCC group (Fig. 3B).

We also analyzed the expression level of immune checkpoint genes (PD-1, PD-L1, CTLA4, CD86, LAG3, HAVCR2, TIGIT), which are considered biomarkers for predicting the efficacy of immunotherapy, between the two groups, and found that the expression levels of these genes were higher in the RCC group than the LCC group patients. (Wilcoxon test, all p < 0.05; Fig. 3C).

<table>
<thead>
<tr>
<th></th>
<th>TCGA dataset</th>
<th>GSE39582 dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survival</td>
<td>0.005</td>
<td>0.961</td>
</tr>
<tr>
<td>Alive</td>
<td>189 (85.9%)</td>
<td>167 (74.9%)</td>
</tr>
<tr>
<td>Dead</td>
<td>31 (14.1%)</td>
<td>56 (25.1%)</td>
</tr>
</tbody>
</table>

Note: Dates were displayed in counts (%). LCC: Left-side colon cancer; RCC: Right-side colon cancer; TCGA: The Cancer Genome Atlas.
The expression of HLA-related genes has been described between LCC and RCC, we found that 16 genes, including HLA-F, HLA-E, HLA-DRB6, HLA-DRB5, HLA-DRB1, HLA-DRA, HLA-DQB2, HLA-DQB1, HLA-DQA2, HLA-DQA1, HLA-DPB2, HLA-DPB1, HLA-DPA1, HLA-DOA, HLA-DMB, HLA-C, were significantly different between two groups, the expression of which were higher in RCC group than LCC group (Wilcoxon test, all \( p < 0.05 \); Fig. 3D).

### 3.1.4 Differential TMB landscape between LCC and RCC Patients

As shown in Fig. 3E, mutation prevalence varies dramatically within CRC in different locations, which highlights the mutation frequency of RCC was relatively higher than LCC. Moreover, LCC and RCC groups contain different mutant genes. Waterfall plots (Fig. 3F, 3G) show the first 30 gene mutation rates in each location. This major discrepancy can be seen that the TP53 presents a higher mutation rate in LCC (LCC: 71%, RCC: 47%), while the PIK3CA (LCC: 21%, RCC: 36%), KRAS (LCC: 33%, RCC: 50%) showed a higher yield mutation rate in RCC. The result illustrates that CRC is not one type of disease, rather it acts as two different diseases in the same organ.

### 3.2 Identifying DE-IRGs and Functional Annotation in tumor and normal

By comparing the LCC and L_normal patients’ transcriptome data, we identified upregulated DEGs and downregulated DEGs shown in Fig. 4A, 4C, and further obtained 658 DE-IRGs and visualized them in a Venn diagram (Fig. 4E). We analyzed these DEGs to identify the functions. This evaluation revealed the enrichment of 2086 GO terms and 90 KEGG pathways in DE-IRGs (FDR < 0.05). We chose to show the top terms and pathways in Fig. 4J. Likewise, we obtained DEGs between RCC and R_normal (Fig. 4B, 4D) and collected 644 DE-IRGs (Fig. 4F) The functional annotations show that 1871 GO terms and 58 KEGG pathways (FDR < 0.05) in DE-IRGs have been enriched and we chose to show the top terms and pathways in Fig. 4H).

Compared with normal tissues, the LCC and RCC have different DE-IRGs and functional annotations.

### 3.3 Construction of prognostic immune genes model

To identify prognosis-related genes, we used the Kaplan-Meier method in DE-IRGs, respectively in LCC and RCC groups, with \( P < 0.01 \), to screen survival-related DE-IRGs. Then to avoid model overfitting, we performed multivariate Cox regression analyses with the LASSO penalty algorithm to solve the multicollinearity problem (Fig. 5A, 5B, 6A, 6B). Finally, we obtained 10 immune genes associated with LCC patients’ prognosis and 7 immune genes associated with RCC patients’ prognosis. These genes have a significant impact on the survival of CRC patients (Supplementary Fig. 1). In other types of cancer, These genes are also differences in the expression between cancer tissues and adjacent tissues(Supplementary Fig. 2).
The LCC patients' prognosis features and risk scores were calculated as: -CCL11*0.268- CXCL13*0.349 + FGF7*0.193- FGFR4*0.206 + FOS*0.350 + INHBB*0.237- MTNR1A*0.229 + NOS1*0.184 + RBP7*0.183- TNFSF11*0.229. Overall survival analysis was performed based on risk score (Fig. 5C, 5D). The AUC values of the risk score in the training set for 1-year, 3-year, 5-year, and 7-year and all-year OS were 0.527, 0.583, 0.571, 0.580 and 0.735 (Fig. 5E).

The RCC prognosis features and risk scores were calculated as: FGF2*0.274- IGHV1-69*0.239 + IL1RL2*0.170 + INHBB*0.240- PTGS2*0.277 + SLPI*0.317- TNFRSF10A*0.183. Overall survival analysis was performed based on risk score (Fig. 5C, 5D). The AUC values of the risk score in the training set for 1-year, 3-year, 5-year, and 7-year and all-year OS were 0.540, 0.560, 0.619, 0.570, and 0.704 (Fig. 6E).

Internal validation of the prognosis immune genes model and stratified analysis by Clinical Factors among LCC patients, the area under curves (AUC) values of risk score predicted in the internal validation set of TCGA for 1-year, 3-year, 5-year, and 7-year and all-year OS were 0.534, 0.597, 0.549, 0.574, and 0.711, respectively (Fig. 5F). The AUC value of the risk score predicted in all TCGA set for all-year OS was 0.714 (Fig. 5G). The AUC values of 10-fold cross-validation were 0.670, 0.781, 0.713, 0.808, 0.720, 0.672, 0.652, 0.564, 0.646, 0.774. The results of the 10-fold cross-validation were high and similar, indicating the model had good predictability and stability.

Likewise, in RCC patients, the AUC values of risk score predicted in the testing set of TCGA for 1-year, 3-year, 5-year, and 7-year and all-year OS were 0.674, 0.555, 0.600, 0.625, and 0.738 (Fig. 6F). The AUC value of the risk score predicted in all TCGA set for all-year OS was 0.734 (Fig. 6G). The AUC values of 10-fold cross-validation were 0.635, 0.792, 0.640, 0.625, 0.654, 0.673, 0.755, 0.589, 0.608 and 0.747 (Fig. 6J). The results of the 10-fold cross-validation were high and stable, indicating the model had good predictability and stability.

### 3.4 External validation of the prognosis signature and stratified analysis by Clinical Factors

To further evaluate the performance of the prognostic feature model, the GSE39582 dataset was used as an external validation set. Among LCC patients, the AUC values of risk score predicted in GSE39582 and complete datasets (TCGA + GEO) for all-year OS were 0.744 and 0.726 (Fig. 5H, 5I). Based on the obtained sample clinical characteristics, complete datasets patients were stratified into ages < 65 years and age ≥ 65 years subgroups (Fig. 5K, 5L), female and male subgroups (Fig. 5M, 5N), pathological tumor Stage I/II, and Stage III/IV subgroups (Fig. 5O, 5P). Overall survival analysis was performed in each subgroup, a significant survival disparity persists between these subgroups (Log-rank, all p < 0.05).

While in RCC patients, the AUC values of risk score predicted in GSE39582 and complete datasets (TCGA + GEO) for all-year OS were 0.705 and 0.713 (Fig. 6H, 6I). Complete datasets patients were also stratified into ages < 65 years and age ≥ 65 years subgroups (Fig. 6K, 6L), female and male subgroups (Fig. 6M, 7N), pathological tumor Stage I/II, and Stage III/IV subgroups (Fig. 6O, 6P). Overall survival analysis was
performed in each subgroup, a significant survival disparity persists between these subgroups (Log-rank, all p < 0.05).

3.5 Incorporating clinical factors to develop individualized nomograms and verify prediction performance

Clinical characteristics, including Age, Gender, pT, PN, pM, pStage, and risk (the level of risk score), were enrolled to perform univariate analysis in the training set of LCC and RCC respectively (Fig. 7A, 8A). After statistical adjustment for other variables, multivariate Cox regression models were developed to predict the overall survival rates of LCC and RCC groups (Fig. 7B, 8B). Multivariate Cox analysis further identified that the risk score was independently prognostic of OS in LCC (P < 0.001), the same goes for RCC (P < 0.001). Nomograms of LCC and RCC were developed by the above prognostic features, with the total points calculated by adding the points of prognostic features (Fig. 7C, 8C).

Among LCC patients, the calibration curve for predicting median survival time OS in the training set, internal validation set, all TCGA set, external validation set (GSE39582), and complete dataset indicated that the nomogram-predicted survival similarly corresponded with actual survival outcomes (Fig. 7D, 7E). The nomogram’s AUC value was 0.722 in the training set, 0.696 in the internal validation set, 0.703 in all TCGA set, 0.739 in the external validation set (GSE39582), and 0.713 in the complete dataset (Fig. 7H-L). The decision curve analysis for the nomograms of LCC patients was presented in Fig. 7F, 7G.

Among RCC patients, the calibration curve for predicting median survival time OS in the training set, internal validation set, all TCGA set, external validation set (GSE39582), and complete dataset indicated that the nomogram-predicted survival similarly corresponded with actual survival outcomes (Fig. 8D, 8E). The nomogram’s AUC value was 0.774 in the training set, 0.744 in the internal validation set, 0.799 in all TCGA sets, 0.737 in the external validation set (GSE39582), and 0.765 in the complete dataset (Fig. 8H-L). The decision curve analysis for the nomograms of LCC patients was presented in Fig. 8F, 8G.

3.6 Differences in the immune microenvironment, immune checkpoint levels, TMB Landscape, and HLA-related gene level Between High- and Low-Risk Patients

Considering the robust correlation between prognostic genes and immune infiltration, we further explored the immune microenvironment, and TMB between high- and low-risk patients based on prognostic genes models. The results were that, in the LCC group, high-risk patients were lower proportions of T cells follicular helper and Dendritic cells resting, higher tumor purity, lower immune, and ESTIMATE scores than low-risk patients, which had a poor prognosis (Wilcoxon test, p < 0.05; Fig. 9A, 9C). In the RCC group, high-risk patients were lower proportions of T cells CD4 memory activated and T cells follicular helper and Macrophages M1 than low-risk patients, which had a poor prognosis (Wilcoxon test, p < 0.05; Fig. 9B, 9D).

We also found that the expression levels of immune checkpoint genes were partly different between high- and low-risk patients in LCC and RCC groups. The expression levels of PD-L1, CTLA4, CD86, and TIGIT were lower in high-risk patients of the LCC group, while the expression levels of LAG3 and TIGIT were
lower in high-risk patients of the RCC group (Wilcoxon test, p < 0.05; Fig. 9E, 9F). In addition, the expression levels of HLA-related genes between high- and low-risk patients in LCC and RCC groups were also analyzed. Both in the LCC and RCC model high-risk patients had lower expression of levels of HLA-related genes than low-risk (Fig. 9G, 9H).

Patients with high-risk score have higher immune related scores and immune checkpoint gene expression, which indicates that patients with high risk may benefit more from immunotherapy.

3.7 Correlation of Hub Gene and Risk Score with Immune-Related Score and Genes

As illustrated in Fig. 10, the correlation analyses between 22 types of immune cells, immune checkpoint genes, and prognostic genes of LCC and RCC groups have been performed respectively. As we can see, both the left risk score and hub genes and the right risk score and hub genes have a strong correlation with immune related scores and genes. This shows that our model may predict the immunotherapy of patients.

4. Discussion

Colorectal cancer (CRC) is the third most common cancer in the world with increasing morbidity and mortality, which has been highly concerned(1). The traditional comprehensive treatment based on surgery were not effective in recurrent and refractory CRC patients(10, 28). Recent studies shown that immunotherapy has been regarded to be an effective treatment for various cancers, playing a specific and lasting anti-cancer role(13). However, immunotherapy only works for long-term durable remission in a subset of patients, possibly because researches of immune-related molecular mechanisms remain unclear and biomarkers for guiding patient stratification are still lacking(29). Furthermore, due to growth nonuniformity and sophisticated oncogenic mechanisms in CRC, developing personalized treatment methods and accurately predicting patient prognosis by growth location is extraordinarily imperative. Therefore, we take the lead to severally establish an immune prophetic model of LCC and RCC that can effectively stratify CRC patients and predict patients’ survival.

In this study, we tend to develop 2 nomograms for CRC classified by each neoplasm aspect and placement containing prognostic immune signatures and clinical prognostic factors, which might accurately distinguish high-and low-risk patients. Besides, the immune signature may act as a predictor of immunotherapy response. The LCC nomogram includes the risk score calculated by prognostic genes (CCL11, CXCL13, FGF7, FGFR4, FOS, INHBB, MTNR1A, NOS1, RBP7, TNFSF11), age, gender, pathological T, pathological N, pathological M and pathological stage that would predict the survival rate. RCC nomogram contains the risk score calculated by prognostic genes (FGF2, IGHV1-69, IL1RL2, INHBB, PTGS2, SLPI, TNFRSF10A), age, gender, pathological T, pathological N, pathological M and pathological stage that may predict the survival rate. Our study extends these findings to demonstrate the presence of great variations in CRC patients’ prognosis supported tumor location indicating that the tumor cell population harbored the foremost essential transcriptomic variations between LCC and RCC.
The colon is approximately 150 cm in length and extends from the ileocecal valve up to the anus. It comprises seven parts: the cecum, ascending colon, transverse colon, descending colon, sigmoid colon, rectum, and anus. As antecedently mentioned, multiple studies have incontestable that RCC and LCC patients are distinct entities based on their embryological origins. The cecum, appendix, ascending colon, hepatic flexure, and proximal two-thirds of the transverse colon have originated from the midgut, whereas distal one-third of the transverse colon, splenic flexure, sigmoid colon, and descending colon have originated from the hindgut. This study confirmed that there were significant differences in the multidimensional insight between LCC and RCC patients. RCC patients have a worse prognosis than LCC patients. We believe that the progression distinction between RCC and LCC patients might be due to higher frequency mutations in key pathways to change in the tumor microenvironment associated with tumor purity.

Recent studies have shown that mutation prevalence differed by side and location. Within right- and left-sided CRC patients, RAS mutations decreased from 70% for cecal to 43% for hepatic flexure location, while BRAFV600 mutations increased from 10–22% between the same locations. The sigmoid and rectal regions had more TP53 mutations and fewer PIK3CA, BRAF, or CTNNB1 mutations within left-sided tumors which is consistent with our results\(^{(3)}\). Mutation patterns noted in our study align well with a recent report by Marshall et al. demonstrate significant differences between LCC and RCC.

The tumor microenvironment (TME) is generally defined as the environment surrounding the tumor, including the extracellular matrix, blood vessels, immune cells, neurons, and other cellular functions, all of which are closely related to tumor progression and therapeutic effects\(^{(30, 31)}\). We also confirmed that the prognosis of CRC patients was closely related to the immune microenvironment. Immune responses in TME are also considered important determinants of tumor invasiveness and progression. Immune score and tumor purity refer to the percentage of tumor cells in the tumor immune microenvironment, which can promote the quantification of immune components, such as immune cells, in tumors and be considered to be an important factor affecting the prognosis of cancer patients. In several studies, it has been confirmed that low immune scores and high tumor purity were associated with a better prognosis\(^{(32, 33)}\). On this basis, this study explored the difference in tumor immune microenvironment between LCC and RCC patients. In our study, RCC patients not only had a poor prognosis but also had a low ESTIMATE score, low immune score, and high tumor purity. Thus, we further analyzed the effect of high- and low-risk on immune infiltration in patients in both LCC and RCC models in our study. We found that both in the LCC and RCC model high-risk patients had lower immune scores, ESTIMATE scores, and higher tumor purity than low-risk with statistical difference. Prior research result was in line with our study that poor prognosis was closely related to low tumor purity in glioma and CRC\(^{(32)}\). Besides, LCC patients were significantly different from RCC patients in terms of immune infiltrating cell types, for example, the proportions of T cells CD8 and T cells follicular helper were significantly higher in the RCC group than LCC group, while macrophages M0 had higher infiltrations in the LCC group. Furthermore, we found in our LCC model that Macrophages M0 and Mast cells activated were higher in high-risk patients, Macrophages M2, Mast cells activated and Neutrophils were higher in low-risk. In summary, all kinds of
cells, cytokines, and chemokines that interact with tumor cells in the tumor microenvironment, especially immune cells, are increasingly recognized as important roles in the body against tumors.

Since patients with refractory malignant tumors, including colorectal cancers, benefit significantly from immune checkpoint inhibitors, immunotherapy is becoming a new therapeutic option for cancer patients. TMB, TME, and immune checkpoint levels (PD-1, PD-L1, CTLA4, CD86, LAG3, HAVCR2, TIGIT), and HLA gene family are considered biomarkers for predicting the efficacy of immunotherapy are considered biomarkers for predicting the efficacy of immunotherapy(8, 34–36). CTLA4 is a member of the immunoglobulin superfamily and encodes a protein that transmits an inhibitory signal to T cells in humans, closely related to human immune system function, and it also is an important genetic genome of the human immune system. PD-1 is an immune-inhibitory receptor expressed in activated T cells which can involve in the regulation of T-cell functions, including those of effector CD8 + T cells. In addition, this protein can also promote the differentiation of CD4 + T cells into T regulatory cells, however, it can also contribute to the inhibition of effective anti-tumor and anti-microbial immunity. PD-L1 is a type I transmembrane protein that Interaction of this ligand with its receptor inhibits T-cell activation and cytokine production. In tumor microenvironments, this interaction provides an immune escape for tumor cells through cytotoxic T-cell inactivation. CD86, one of the co-stimulatory molecules on antigen-presenting cells, plays an important role not only in autoimmunity and transplantation but also in tumor immunity(37, 38). LAG3 is a significant immune checkpoint with relevance in cancer, infectious disease, and autoimmunity(39). HAVCR2 encodes T-cell immunoglobulin mucin 3 protein, a critical checkpoint molecule found to be expressed on the surface of Th1 cells, acting as a negative regulator and binding to the ligand galectin-9 to mediate Th1 cell the apoptosis and regulating inflammatory responses(40). TIGIT is an inhibitory receptor expressed on lymphocytes that are recently propelled under the spotlight, which interacts with CD155 expressed on antigen-presenting cells or tumor cells to down-regulate T cell and natural killer (NK) cell functions(41). The previous study demonstrates that the prognostic impact of PD-L1 and PD-1 expression differs according to the primary tumor site in CRC patients. Moreover, PD-L1 is an independent favorable prognostic factor in right-side tumors(42). Previous studies have shown that the high expression of HLA class I genes and PD-L1 are both closely associated with RCC(8). This finding was in line with our study, which demonstrated that there were significant differences in immune checkpoint genes (PD-1, PD-L1, CTLA4, CD86, LAG3, HAVCR2, and TIGIT) and HLA gene family (HLA-F, HLA-E, HLA-DRB6, HLA-DRB5, HLA-DRB1, HLA-DRA, HLA-DQB2, HLA-DQB1, HLA-DQA2, HLA-DQA1, HLA-DPB2, HLA-DPB1, HLA-DPA1, HLA-DOA, HLA-DMB, HLA-C) expression between LCC and RCC patients. All of these genes were a higher expression in RCC than in the LCC group.

Given the above, our study independently assessed the effect of tumor microenvironment in LCC and RCC from two aspects (TMB and immune microenvironment) and speculated that RCC patients may benefit more from immunotherapy than LCC. It is worth noting that although immunotherapy can bring good benefits to some lung cancer patients, some patients still do not show the desired results after using immune checkpoint inhibitors(43, 44). These results need validation but may be clinically significant, as they indicate that tumor location might be an important factor to take into consideration in therapeutic decisions, including eligibility for immunotherapy. This study revealed the sensitivity of RCC patients may
have to immunotherapy, which would provide a visual field for researchers to develop drugs with a high therapeutic index or high efficacy.

The prognostic genes in the signature have previously been shown to be potential biomarkers. CCL11, also known as eotaxin-1, is regarded as an eosinophil chemoattractant, which has been implicated in allergic and Th2 inflammatory diseases. Researches showed CCL11 is significantly increased in the serum of inflammatory bowel disease (IBD) patients and involved in the pathogenesis of colon tumorigenesis(45, 46). CXCL13, also known as B-cell attracting chemokine 1 (BCA-1), is originally identified as a homeostatic chemokine to attract B cells, a minority of T cells, and macrophages. Several studies have shown that CXCL13 is involved to mediate 5-Fu resistance mechanism and considered to be a biomarker for predicting the prognosis of colorectal cancer, which is consistent with our findings(47, 48). FGF2, a member of the fibroblast growth factor (FGF) family whose signaling network has been implicated in several pathways, such as normal cell growth, differentiation, angiogenesis and tumor development, significantly promotes tumor cell differentiation and proliferation(49, 50). FGF7, also called keratinocyte growth factor (KGF), is involved in inducing vascular endothelial growth factor (VEGF)-A expression, which plays important roles in the angiogenesis and metastasis of various cancer cells including colorectal cancer(51, 52). FGFR4, a receptor tyrosine kinase, plays a key role in cancer development and prognosis via the activation of its downstream oncogenic signaling pathways(53). Studies have shown that Patients with high FGFR4 expression exhibited a lower 5-year survival rate compared with patients with low FGFR4 expression (64 vs. 74%). Besides, the combination of 5-FU and FGFR4-selective inhibitor has a synergistic effect in reducing colorectal cancer cell proliferation and preventing the cell cycle(54, 55). FOS, located on human chromosome 14q21-31, encodes the nuclear oncprotein c-Fos. Prior studies revealed that overexpression of FOS can promote the incidence and development of colorectal cancer(56). INHBB, a protein-coding gene that participated in the synthesis of the transforming growth factor-β (TGF-β) family members, is increased in CRC tissue and associated with poor overall survival (OS) and disease-free survival (DFS)(57, 58). MTNR1A, called melatonin receptors 1a, is increasingly recognized as an anti-tumor molecule in a variety of malignancies(59, 60). Studies showed common genetic variation in the MTNR1A genes might contribute to breast cancer susceptibility(61). However, no evidence of correlation has been investigated between MTNR1A and CRC. NOS1, called nitric oxide synthase 1, can synthesize nitric oxide that is closely related to the carcinogenesis and progression of colon cancer(62). RBP7, a member of the cellular retinol-binding protein (CRBP) family, is involved in the malignant transformation of colon cancer cells. Prior research found that high expression of RBP7 promoted migration and invasion of colon cancer cells(63). PTGS2, also known as cyclooxygenase-2 (COX-2), is highly increased in tumor cells and contributes to tumor growth, recurrence and metastasis(64). Prior research showed that PTGS2-driven inflammatory responses can induce tumor expression of microRNA-21, which contributes to colorectal cancer development(65). SLPI, called secretory leukocyte protease inhibitor, plays important roles in colorectal cancer cell growth, migration and invasion via upregulation of AKT signaling, associated with poor prognosis of CRC(66, 67).
Currently, there is no reported experimental research between MTNR1A, TNFSF11, TNFRSF10A, IGHV1-69, IL1RL2 and colorectal cancer. Moreover, the GSE39582 dataset was used as the external validation set to analyze the overall survival of each prognostic immune gene and the accuracy of our prognostic genetic models of LCC and RCC groups. In the end, we analyzed the differential expression of prognostic immune genes in 33 types of cancer and found that most of them were expressed differently in cancer and normal tissue. The results further confirmed that these key genes were related to the occurrence and development of CRC, and enable to be potential prognostic biomarkers for other cancers.

There are some limitations to the current study. First, although the signature and nomogram constructed in this study using massive data from TCGA and verified in GEO have reliable robustness, the nature of retrospective analysis still exists. Second, we tend to explore the TMB and immune microenvironment landscape between right- and left-sided CRC patients and between different risks patients, however, the study still lacked experimental verification. Third, this is often a serious downside that acquisition of risk score required knowledge of the expression of multiple genes in tumor tissues, which undoubtedly increased the burden of nomogram application. This seems to be a common problem for most molecular diagnostic or prognostic models. The way to change and simplify the clinical application of prognosticative models may be a question for researchers and clinicians to contemplate. We believe that the development of molecular detection technology in the future is bound to improve this dilemma. The nomogram may be used routinely in the future.

**Conclusion**

We found significant differences between left- and right-sided CRC patients in clinical features, transcriptome, TMB, and immune microenvironment landscape in the multidimensional insight, suggesting that CRC can be classified and analyzed into different clinical types by differences in anatomical location and gene expression to aid in early diagnosis and prognosis. In our study, we respectively developed and validated two prognostic nomograms based on DE-IRGs in combination with clinical features that not only provide a basis for personalized and precise treatment of CRC patients but also reflect the immune status of LCC and RCC. These prognostic immune genes may become promising biomarkers for the diagnosis, treatment, and prognosis of CRC. Moreover, these findings further support previous evidence suggesting that proximal and distal CRC may represent different epidemiological, pathological, genetic, immune and clinical entities(68, 69). Compared with LCC, RCC patients are generally poorer differentiated, diagnosed in more advanced tumor stages, display different immune molecular patterns, and carry a poorer prognosis, but might benefit more from immunotherapy.

**Declarations**

**Data Availability Statement**

The original contributions presented in the study are included in the article/supplementary materials, further inquiries can be directed to the corresponding author.
Ethics Statement

Data from public database

Author Contributions

MYN and CYC conceptualized and designed the study, collected data and performed the data analysis. MYN, CYC and ZRZ designed and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by Medical Scientific Research Foundation of Hebei Province, China (H2020206374, H2021206306, 8220160183), and the Talent Project of Hebei, China (LS202001).

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References


**Supplementary Figure 2**

**Figures**
Figure 1

Flowchart presenting the process of establishing the immune signature and prognostic nomogram of LCC and RCC in this study.
Figure 2

The difference between LCC and RCC groups.

(A) Comparison of survival rates of the two groups. (B) Volcano plot for DEGs of the two groups. (C) Heatmap of top 40 DEGs of the two groups. (D) Heatmap demonstrated the top 20 different GSVA pathways of the two groups. Venn diagram for differentially expressed immune-related genes (DE-IRGs)
of LCC (E) and RCC (F) groups. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway functional enrichment analyses of the up-regulated DE-IRGs in LCC (G) and RCC (H) groups.

Figure 3

The difference between LCC and RCC groups.
Comparison of 22 immune cells between the LCC and RCC groups (A). The Immune Score, ESTIMATES core, and Tumor Purity level between the two groups (B). The expressed level of PD-1, PD-L1, CTLA4, CD86, LAG3, HAVCR2, and TIGIT between the two groups (C). The expressed level of HLA-related genes between the two groups (D). The tumor mutation burden of samples between the two groups (E). The waterfall plot demonstrated the top 30 frequently mutated genes in LCC (F) and RCC (G) groups. (ns P>0.05, * P < 0.05, ** P < 0.01, *** P < 0.001).
Figure 4

Volcano plot for DEGs between LCC and L_normal groups (A) and RCC and R_normal groups (B). Heatmap of the top 40 DEGs between LCC and L_normal groups (C) and RCC and R_normal groups (D). Venn diagram for DE-IRGs between LCC and L_normal groups (E) and RCC and R_normal groups (F). Top GO and KEGG functional enrichment analysis of the DE-IRGs between LCC and L_normal groups (G) and RCC and R_normal groups (H).
Figure 5

Construction and validation of the prognostic model in the LCC group.

(A-B) Optimal parameters for Lasso Regression Analysis. Differences in overall survival between high-risk and low-risk groups based on the risk scores in TCGA dataset (C) and GSE39582 dataset (D). Time-dependent ROC curves at 1-year, 2-year, 3-year, 5-year, 7-year and all-year in the training set (E), in internal validation set (F), in all TCGA set (G), in GSE39582 set (H), and in complete dataset (I). The ROC curves of 10-fold cross-validation. (J) Comparison of survival rates of high-risk and low-risk groups in different clinical subtypes. Survival analysis of different clinical characteristics including (K) Age < 65, (L) Age ≥ 65, (M) Female, (N) Male, (O) Stage I-II, (P) Stage III-IV.
Figure 6

Construction and validation of the prognostic model in the RCC group.

(A-B) Optimal parameters for Lasso Regression Analysis. Differences in overall survival between high-risk and low-risk groups based on the risk scores in TCGA dataset (C) and GSE39582 dataset (D). Time-dependent ROC curves at 1-year, 2-year, 3-year, 5-year, 7-year and all-year in the training set (E), in internal
validation set (F), in all TCGA set (G), in GSE39582 set (H), and in complete dataset (I). The ROC curves of 10-fold cross-validation. (J) Comparison of survival rates of high-risk and low-risk groups in different clinical subtypes. Survival analysis of different clinical characteristics including (K) Age < 65, (L) Age ≥ 65, (M) Female, (N) Male, (O) Stage I-II, (P) Stage III-IV.

### Figure 7

Independent prognostic analysis of prognostic immune signatures in LCC.

Univariate (A) and multivariate (B) Cox regression analysis of prognostic signatures and clinical characteristics. Developed Incorporating clinical factors nomogram of LCC (C). Calibration curve of the
nomogram in the training set, internal validation set, in all TCGA set, in GSE39582 set and in complete dataset (D-E). Decision curve analysis of the nomogram in the training set, internal validation set, in all TCGA set, in GSE39582 set and in complete dataset (F-G). Time-dependent ROC curves at 1-year, 2-year, 3-year, 5-year, 7-year and all-year in the training set (H), in internal validation set (I), in all TCGA set (J), in GSE39582 set (K), and in complete dataset (L).

**Figure 8**

Independent prognostic analysis of prognostic immune signatures in RCC.
Univariate(A) and multivariate(B) Cox regression analysis of prognostic signatures and clinical characteristics. Developed Incorporating clinical factors nomogram of RCC (C). Calibration curve of the nomogram in the training set, internal validation set, in all TCGA set, in GSE39582 set and in complete dataset (D-E). Decision curve analysis of the nomogram in the training set, internal validation set, in all TCGA set, in GSE39582 set and in complete dataset (F-G). Time-dependent ROC curves at 1-year, 2-year, 3-year, 5-year, 7-year and all-year in the training set (H), in internal validation set (I), in all TCGA set (J), in GSE39582 set (K), and in complete dataset (L).
Figure 9

Comparison of immune cells between high-risk and low-risk groups in LCC (A) and RCC. (B) The Immune Score, ESTIMATE Score, and Tumor Purity level between high- and low-risk groups in LCC (C) and RCC (D). The expression level of PD-1, PD-L1, CTLA4, CD86, LAG3, HAVCR2 and TIGIT between high-risk and low-risk groups in LCC (E) and RCC (F). The HLA-related genes between high-risk and low-risk groups in LCC (G) and RCC (H). (ns P>0.05, * P < 0.05, ** P < 0.01, *** P < 0.001).
Figure 10

Correlation matrix between 22 immune cell proportions, immune check point genes and 10 prognostic immune genes in LCC (A) and 7 prognostic immune genes in RCC (B).

Supplementary Files

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

- S1.tif