



# Identification and validation of a novel innate lymphoid cell-based signature to predict prognosis and immune response in liver cancer by integrated single-cell RNA analysis and bulk RNA sequencing

Meng Pan<sup>1#</sup>, Xiaolong Yuan<sup>2#</sup>, Junlu Peng<sup>3</sup>, Ruiqi Wu<sup>3</sup>, Xiaopeng Chen<sup>3</sup>

<sup>1</sup>Department of Hepatobiliary Surgery, The Second Affiliated Hospital of Wannan Medical College, Wuhu, China; <sup>2</sup>Department of Pharmacy, The Second Affiliated Hospital of Wannan Medical College, Wuhu, China; <sup>3</sup>Department of Hepatobiliary Surgery, The First Affiliated Hospital of Wannan Medical College, Wuhu, China

**Contributions:** (I) Conception and design: All authors; (II) Administrative support: None; (III) Provision of study materials or patients: M Pan, X Yuan, J Peng, R Wu; (IV) Collection and assembly of data: M Pan, X Yuan, J Peng, R Wu; (V) Data analysis and interpretation: M Pan, X Yuan, J Peng, R Wu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

<sup>#</sup>These authors contributed equally to this work.

**Correspondence to:** Xiaopeng Chen, PhD. Department of Hepatobiliary Surgery, The First Affiliated Hospital of Wannan Medical College, 2 Zheshan West Road, Wuhu 241001, China. Email: drchenxp@wnmc.edu.cn.

**Background:** Innate lymphoid cells (ILCs) exert tumor suppressive and tumor promoting effects. However, the prognostic significance of ILC-associated genes remains unclear in hepatocellular carcinoma (HCC). Hence, the aim of this research was to develop an innovative predictive risk classification system using bioinformatics examination.

**Methods:** We explored the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) databases to gather data pertaining to HCC and its clinical details. Significantly different ILC-associated genes were investigated by Seurat analysis. The number of signaling interactions of ILCs with other cells was discovered by CellPhoneDB analysis. ClusterProfiler and Metascape were utilized to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses on ILC genes. In order to identify potential ILC predictors, we utilized univariate Cox regression and least absolute shrinkage and selection operator (LASSO) analyses, subsequently validating these predictors in TCGA and GEO groups. The multi-omics ILC signature model's clinical predictive capabilities, along with drug sensitivity and immune factor relations, were assessed using Cell-type Identification by Estimating Relative Subsets of RNA Transcripts (CIBERSORT) and pRRophetic. We investigated the possible molecular pathways in our predictive ILC signature through the utilization of gene set enrichment analysis (GSEA) and gene set variation analysis (GSVA). Five key genes were screened out by constructing a competing endogenous RNA (ceRNA) network using Cytoscape and their values in clinical indexes were demonstrated. Immunohistochemistry (IHC) in HCC cases confirmed the expression of these genes.

**Results:** ILC cell subsets were identified, and cell-cell communication analysis revealed that the signaling pathways involving ILC cell subsets were mostly abundant in the HCC microenvironment. Subsequently, 270 marker genes of the ILC clusters were subjected to GO and KEGG enrichment analysis. Furthermore, a total of 58 prognostically relevant genes were screened as features for prognostic prediction models. Next, the models were validated and clinically evaluated (P values of Kaplan-Meier survival curves below 0.05). Five key genes (*C2*, *STK4*, *CALM1*, *IL7R*, and *RORA*) were further screened by multi omics analysis of immune cell and factor and drug sensitivity and correlation analysis of tumor regulatory genes in liver cancer. Furthermore, the potential clinical value of the 5 key genes was confirmed in HCC patients. Finally, the IHC results confirmed the expression of *C2*, *STK4*, *CALM1*, *IL7R*, and *RORA* in HCC. Our experimental results provided preliminary evidence supporting the oncogenic roles of *STK4* and *CALM1*, as well as the tumor-

suppressive roles of *C2*, *RORA*, and *IL7R* in HCC.

**Conclusions:** A novel prognostic feature of ILC potentially involved in HCC was discovered. It showed high values in predicting patient overall survival (OS) as well as good differences in immunity and drug sensitivity. Therefore, targeting these ILC signatures may be a potential effective approach in HCC treatment.

**Keywords:** Gene signature; hepatocellular carcinoma (HCC); single cell RNA sequencing (scRNA-seq); innate lymphoid cells (ILCs)

Submitted Apr 29, 2024. Accepted for publication Aug 22, 2024. Published online Oct 28, 2024.

doi: 10.21037/tcr-24-725

View this article at: <https://dx.doi.org/10.21037/tcr-24-725>

## Introduction

Globally, liver cancer ranks fourth in terms of type of cancer. The most common form of liver cancer, hepatocellular carcinoma (HCC), is the third leading cause of death from cancer. Approximately 80% of the 905,677 new liver cancer cases globally in 2020 were diagnosed as HCC (1). Chronic liver injury, inflammation, and fat accumulation in the hepatocytes caused by hepatitis B and C viruses lead to HCC (2). Despite notable progress in treatment, just 40% of cases of HCC are detected early, leading to unfavorable treatment results. Liver cancer deaths are projected to exceed 1 million in 2030 (3).

Enhanced identification and treatment strategies for HCC are urgently required. Despite the advancements in biological techniques like RNA sequencing (RNA-seq)

and single cell RNA sequencing (scRNA-seq) that have enhanced our knowledge of tumors, there is a scarcity of available data concerning research conducted on the individual cell level of HCC. Pioneering scRNA-seq investigated the subpopulations of primary HCC cells versus immune cells in the formation of the shape of the HCC microenvironment (4). This atlas promotes the understanding of the HCC microenvironment.

Innate lymphoid cells (ILCs), a unique category of immune cells, play a key role in immune response regulation during disease states (5). They drive opposite responses such as tissue repair and promotion of tumorigenesis (6). Prior research indicated that ILCs possess anti-cancer properties, while some studies proposed that ILCs contribute to the promotion of tumor growth (7,8). The study examined the role of a specific ILC subgroup, but did not consider the intricate interactions within the ILC system and its controlling cytokines (9). Therefore, more analyses of tumor ILCs are needed to understand their role in the tumor microenvironment.

Patients diagnosed with HCC have a very poor prognosis, as their chances of surviving for 5 years are less than 18% (2). Diagnosis and screening in the early stages of cancer are extremely necessary for a successful treatment, and the genetic screening that is already available in clinical practice is insufficient (10). The prediction of the disease status and stage of patients by screening the risk factors for HCC facilitates its treatment and prolongs the survival time of patients.

Therefore, this study combined scRNA-seq and bulk RNA-seq data to thoroughly examine HCC, aiming to create a novel prognostic risk classification model through bioinformatics analysis. First, the high expressed genes in ILC cells were measured by scRNA-seq cohort. Risk assessment was conducted on these genetic markers, leading

### Highlight box

#### Key findings

- The experimental results provided preliminary evidence supporting the oncogenic roles of *STK4* and *CALM1*, as well as the tumor-suppressive roles of *C2*, *RORA*, and *IL7R* in hepatocellular carcinoma (HCC).

#### What is known and what is new?

- Innate lymphoid cells (ILCs) exert tumor suppressive and tumor promoting effects.
- A novel prognostic feature of ILC potentially involved in HCC was discovered. It showed high values in predicting patient overall survival as well as good differences in immunity and drug sensitivity.

#### What is the implication, and what should change now?

- The potential clinical value of the five key genes was confirmed in HCC patients. Targeting these ILC signatures may be a potential effective approach in HCC treatment.

to the establishment of a robust and dependable model in The Cancer Genome Atlas (TCGA) cohort, which was subsequently verified in cohorts from the Gene Expression Omnibus (GEO) and the International Cancer Genome Consortium (ICGC). Subsequently, the groups categorized as high- and low-risk were assessed for clinical predictive value, immune correlation, and competing endogenous RNA (ceRNA) network construction. Finally, five potential key genes were found as markers of HCC in our prognostic risk model. This research would provide a more thorough insight into HCC. We present this article in accordance with the TRIPOD reporting checklist (available at <https://tcr.amegroups.com/article/view/10.21037/tcr-24-725/rc>).

## Methods

### *Data gathering*

Researchers have access to a range of data in the TCGA database (<https://portal.gdc.cancer.gov/>), such as DNA methylation, copy number variations, single nucleotide polymorphisms (SNPs), microRNA (miRNA), and gene expression profiles. For liver hepatocellular carcinoma (LIHC), messenger RNA (mRNA) expression data were gathered, encompassing both a control group (n=50) and a cancer group (n=374). Gene expression data are stored in the GEO database. Data related to liver cancer from the dataset GSE149614 were obtained from the publicly available GEO database for analysis of correlations at the single cell level. There were a collective of 18 samples, including 10 liver cancer samples and 8 control samples. The data file from the GSE14520 Series Matrix File was retrieved from the publicly accessible GEO database, using the GPL3921 annotation platform. Extracted were 221 comprehensive expression profiles along with survival information of patients diagnosed with LIHC. Furthermore, comprehensive expression profiles and survival data for 202 patients with LIHC were acquired from the ICGC database. There were 460 miRNA sets associated with HCC acquired from the Human microRNA Disease Database (HMDD) database (<http://www.cuilab.cn/hmdd>).

### *Single cell analysis*

Utilizing the Seurat tool (11), expression profiles were analyzed, excluding genes characterized by low expression levels (nFeature\_RNA >50 and percent.mt <5). The information was standardized and homogenized in

sequence, followed by principal component analysis (PCA) to determine the optimal number of principal component (PC) using ElbowPlot. t-Distributed Stochastic Neighbor Embedding (tSNE) analysis was used to group the cells, and the annotation file from the cellDex package was used to determine the spatial relationship between each group. Cell marker genes sourced from the CellMarker website were employed for annotating these groups and infer the potential spatial relationships between cells (12). Cells that play an important role in disease progression are labeled in clusters. Following this, the FindAllMarkers function was utilized to identify marker genes corresponding to each cell subtype within the single-cell profiles, with a logfc.threshold of 1 and a min.pct of 0.1. Genes meeting the criteria of an adjusted P value less than 0.05 and an absolute average log<sub>2</sub> fold change greater than 1 were determined to be distinctive marker genes specific to each cellular subtype.

### *Analysis of the interaction between ligand and receptor*

The database CellPhoneDB contains information on receptors, ligands, and their interactions, as detailed in reference (13). This database, revealing heteromeric complexes in both ligands and receptors, houses information on 978 distinct protein varieties and integrates with other databases like International Union of Pharmacology (IUPHAR), Protein Data Bank (PDB), Ensembl, and Universal Protein (UniProt). It enables the examination of cellular communication and the thorough, methodical study of communication molecules between cells and the exploration of communication among various cell types. We used statistical tools from CellphoneDB to analyze the characteristics of single-cell expression profiles in order to assess the importance of ligand-receptor interactions. Cluster labels for every cell were randomly permuted 1,000 times, and the average expression levels of receptors within clusters and ligands within interacting clusters were calculated. This results in a null distribution, commonly referred to as the Bernoulli distribution, for each receptor-ligand pair in all pairwise comparisons between the two cell types. Ultimately, select ligand-receptor pairs, especially pertinent to our study, were highlighted to demonstrate their connection.

### *Gene function enrichment analysis*

Gene sets pertinent to HCC were systematically annotated

for functional insights using clusterProfiler to delve into their functional significance (14). Functional categories were evaluated utilizing the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses. Pathways from GO and KEGG that had values for both P and Q below 0.05 were deemed significant. Furthermore, the functional annotation of pivotal gene sets was conducted using the Metascape database ([www.metascape.org](http://www.metascape.org)) to thoroughly elucidate their functional significance. GO analysis and KEGG pathway analysis were conducted on specific gene sets. Statistical significance was determined by requiring Min overlap  $\geq 3$  and  $P \leq 0.01$ .

### ***Model construction and prognosis***

Prognosis-associated genes were meticulously chosen, and a least absolute shrinkage and selection operator (LASSO) regression model was constructed using the “glmnet” package to further refine the prognostic model. The integration of the expression of individual genes resulted in the creation of a formula for calculating each patient’s risk score, which was further adjusted by its corresponding LASSO regression coefficient. The patients were stratified into low-risk and high-risk groups by applying this equation, with the median risk score serving as the dividing line. Kaplan-Meier analysis was employed to assess differences in survival across the various groups, with comparisons made using the log-rank test. Additionally, in order to assess the stability of the model, external datasets were employed for validation. Moreover, the predictive accuracy of the model was evaluated by generating a receiver operating characteristic (ROC) curve with the help of the “survivalROC” software.

### ***Drug sensitivity analysis***

Utilizing the R package “pRRophetic”, predictions were made for the sensitivity of specific tumor samples based on the Genomics of Drug Sensitivity in Cancer database (GDSC, <https://www.cancerrxgene.org/>) (15). This approach involved calculating each chemotherapy drug’s half maximal inhibitory concentration (IC<sub>50</sub>) value, followed by validation using 10-fold cross-validation with the GDSC training set to evaluate the accuracy of regression and prediction. Default parameters were employed for all settings, including the application of “combat” to mitigate batch effects and averaging replicate gene expressions.

### ***Immune cell infiltration analysis***

Cell-type Identification by Estimating Relative Subsets of RNA Transcripts (CIBERSORT) stands as one of the most prominently referenced tools for characterizing and inferring cellular composition in accordance with gene expression profiles. In our research, the CIBERSORT algorithm was employed to dissect RNA-seq data from different LIHC patient subgroups, facilitating the estimation of relative proportions for 22 distinct types of immune infiltrating cells (16). Overall, each sample’s estimated score for immune cells was 1. A Spearman correlation analysis was performed to investigate the connection between gene expression and the presence of immune cells, with a significance level of  $P < 0.05$ .

### ***Gene set enrichment analysis (GSEA)***

Differential gene expression analysis was conducted on LIHC expression data to pinpoint genes exhibiting disparate expression levels between high-risk and low-risk groups. Gene sets were filtered to include only those with sizes between 15 and 500 genes. After 1,000 permutations, gene sets were determined to be enriched using a significance threshold of  $P < 0.05$  and a false discovery rate (FDR) value of 0.25. Following that, there were notable enhancements in the GO and KEGG pathways that were revealed.

### ***Gene set variation analysis (GSVA)***

GSVA is a method used to measure the enrichment of transcriptome gene sets without supervision and without relying on parameters (17). GSVA facilitates the conversion of gene-level alterations into pathway-level variations by systematically scoring the gene sets of interest, thereby assessing the biological functionality of the sample. To mitigate the impact of redundant information within pathways, duplicate genes were eliminated from each gene set, along with genes featured in multiple pathways. During this study, a collection of 50 hallmark pathway genes was obtained from the Molecular Signatures Database (MSigDB). Following this, the GSVA algorithm from the “GSVA” package in R software was used to thoroughly evaluate each gene set, allowing for the examination of possible changes in biological function among various samples.

### ***Immunohistochemistry (IHC)***

Participants included four newly diagnosed HCC patients

at varying T stages (T1–T4) from the Second Affiliated Hospital of Wannan Medical College. We acquired four fresh HCC tissue samples from these individuals. Additionally, the expression levels of five crucial ILC genes across different T stages were confirmed using IHC. The IHC was conducted as previously described (18) on HCC clinical samples, utilizing rabbit polyclonal primary antibodies targeting *C2* (1:100, DF14074, Affinity), *STK4* (1:100, DF7691, Affinity), *CALM1* (1:100, AF6353, Affinity), *IL7R* (1:100, DF6362, Affinity), and *RORA* (1:100, DF3161, Affinity). The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was reviewed and approved by the Ethics Committee of The Second Affiliated Hospital of Wannan Medical College (Wuhu, China, WYEFYLS2023139). Informed consent was taken from all participants.

### Statistical analysis

R software (version 4.1) was used to perform statistical analysis. Survival curves were created using Kaplan-Meier analysis and compared by log-rank. Multivariate analysis was conducted using Cox proportional hazards models. A value of  $P < 0.05$  was used to determine statistical significance.

## Results

### Pre-processing of single-cell expression profile data

In this study, a comprehensive analysis was conducted on 63,101 cells from 18 liver samples to identify gene expression patterns. Out of these, only 47,640 cells, which adhered to specific expression criteria (percent.mt  $< 5$ , nFeature\_RNA  $> 50$ ), were selected for further examination. The analysis included these cells to determine the gene expression profile (Figure 1A,1B). The expressions of genes in the samples are shown in Figure 1C and the top 10 genes with the greatest standardized variance are indicated (Figure 1C).

### Analyzing single-cell data for subtype clustering

PCA dimensionality reduction analysis was conducted on 20 genes, revealing varying scores across different dimensions (Figure 1D). Interestingly, this analysis identified no significant variation among the samples (Figure 1E). ElbowPlot analysis determined 17 as the best number of PCs (Figure 1F), while tSNE analysis identified 33 distinct subtypes (Figure 1G).

### Annotation of clustered subtypes

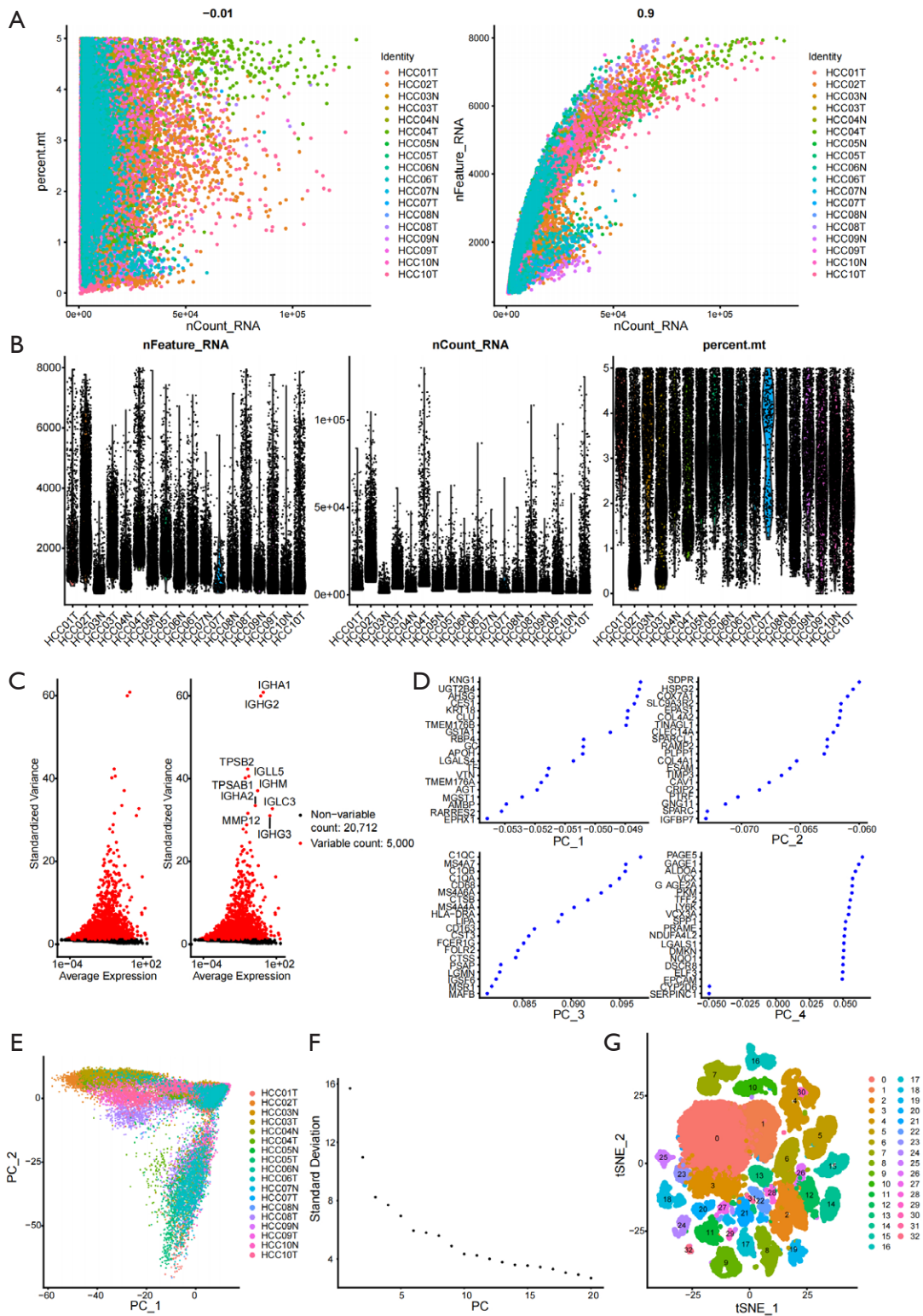
HumanPrimaryCellAtlasData, ImmGenData, and BlueprintEncodeData served as annotation references to categorize each subtype using the R package “SingleR”. A total of 33 clusters were annotated to various cell types, including 11 cell categories such as endothelial cells, B cells, pro, natural killer T cell (NKT), stromal cells, ILC, fibroblasts, epithelial cells, neutrophils, macrophages, hepatocytes, and mast cells (Figure 2A,2B). Following that, FindAllMarkers was utilized to discover marker genes for every cell type by analyzing the single-cell expression profiles.

### Comparative study of ligands and their receptors

The CellphoneDB software was used to analyze the interactions between ligands and receptors in the characteristics of the single-cell expression profile. Subsequently, specific ligand-receptor pairs were chosen and depicted (Figure 3A). Epithelial cells-fibroblasts, ILC-mast cells interacted with CD74\_APP, CD74\_MIF with high interaction scores. Additionally, a significant number of possible ligand-receptor connections were found among mast cells, ILC, stromal cells, and various other cell types (Figure 3B). Following this, the count of ligand-receptor gene pairs linked to each cell cluster was calculated, revealing that ILC subtypes had the most potential interactions with other cell subtypes (Figure 3C).

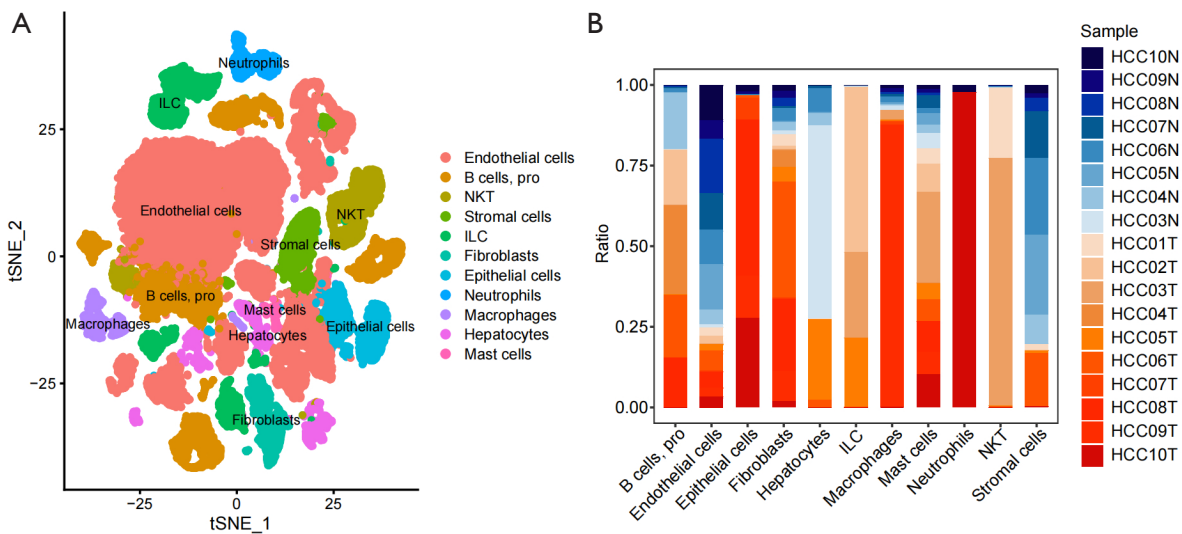
### Analysis of functional markers for significant subcategories

Pathway analysis was performed on 270 marker genes of the ILC subtypes. The analysis of GO enrichment showed that the marker genes were mainly enriched in pathways linked to the humoral immune response, blood particles, and binding to the major histocompatibility complex (MHC) protein complex (Figure 4A). The KEGG enrichment analysis revealed that the marker genes were predominantly enriched in pathways like Graft versus host disease and T helper cell 17 (Th17 cell) differentiation (Figure 4B). Additionally, further pathway analysis of the marker genes conducted using the Metascape database demonstrated enrichment in pathways such as blood microparticles, regulation of cell activation, and collagen-containing extracellular matrix (Figure 4C). Moreover, the Cytoscape software was used to analyze the protein interaction network of the genes in these marker gene sets (Figure 4D).



**Figure 1** Characteristics of liver cancer scRNA-seq. (A) The left figure shows the relationship between cell sequencing depth and mitochondrial content, and the right figure shows the relationship between sequencing depth and gene quantity, and the two are positively correlated. (B) Single-cell quality control showing cell number, gene number, and sequencing depth for each sample. (C) We identified genes

that were significantly different between cells and plotted the characteristic variance. (D,E) Display of PCA and distribution of PC, dots represent cells, and colors represent samples. (F) Variance ranking plot for each PC. (G) We divided cells into 33 clusters by tSNE algorithm according to the important components available in PCA; colors represent cell clusters. PC, principal component; tSNE, t-Distributed Stochastic Neighbor Embedding; scRNA-seq, single cell RNA sequencing; PCA, principal component analysis.



**Figure 2** Annotation of cell subtypes. (A) Cell annotation of 33 clusters, 33 clusters are annotated into 11 cell types, namely endothelial cells, B cells, pro, NKT, stromal cells, ILCs, fibroblasts, epithelial cells, neutrophils, macrophages, hepatocytes, and mast cells. (B) Differences in the proportions of 11 types of cells in the two groups of samples. tSNE, t-Distributed Stochastic Neighbor Embedding; ILCs, innate lymphoid cells; NKT, natural killer T cell.

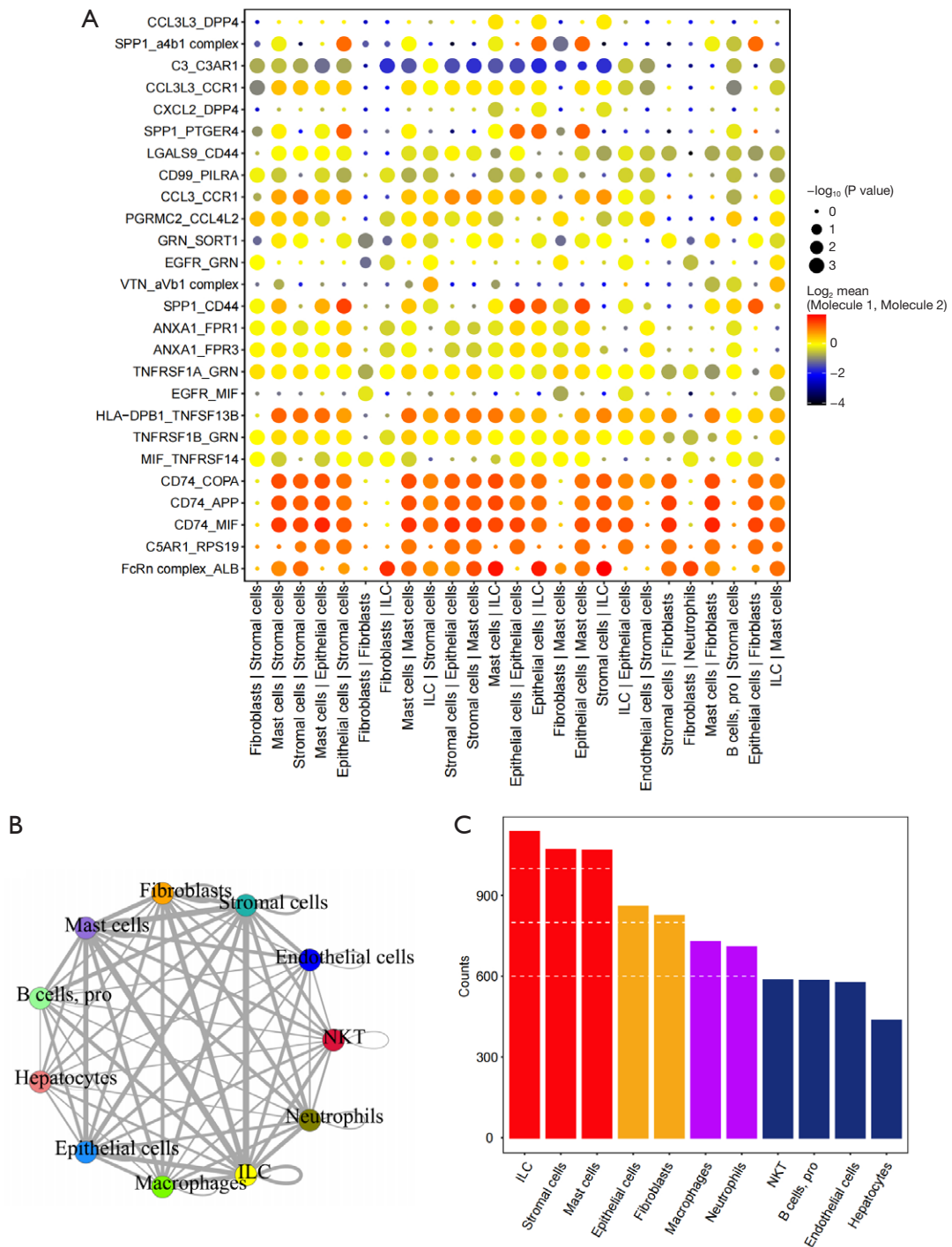
### Prognosis-related genes and prediction model

Clinical data from liver cancer patients were gathered to identify key genes within the modeling candidate gene set. Following this, the Cox univariate regression and LASSO regression algorithms were used to identify specific genes linked to liver cancer (Figure 5A-5C). The findings revealed a collective of 58 prognosis-related genes, which were the following: *SPP1*, *DAB2*, *HSP90AA1*, *RHEB*, *CPS1*, *RBP4*, *TUBA4A*, *CYP3A5*, *SRSF7*, *STK4*, *HSPH1*, *S100A8*, *TREM2*, *KNG1*, *LEPROTL1*, *ADI1*, *UGDH*, *AGXT*, *APOC1*, *KLRB1*, *PDK4*, *ALDOB*, *NPC2*, *SERPING1*, *IER3*, *SULT2A1*, *HMGB2*, *PSAP*, *VTN*, *AMBP*, *FYN*, *CALM1*, *S100A6*, *PRDX6*, *APOC3*, *IL7R*, *C4BPB*, *APOE*, *RARRES2*, *SPINK1*, *GC*, *TTR*, *FTL*, *RORA*, *PCK1*, *PEBP1*, *TSC22D3*, *ALB*, *CD7*, *FCN3*, *LGMN*, *RBCK1*, *C2*, *CFH*, *CTSB*, *STAT4*, *GAMT*, and *CST3*. Patients from the TCGA dataset were partitioned into two cohorts, comprising a training set and an internal validation set, with a randomized split ratio of 4:1. The datasets

GSE14520 and ICGC served as an external validation set. Using LASSO regression analysis, each sample's best risk score was determined ( $0.139578825433095 + PRDX6 \times 0.143823075246877 + SPP1 \times 0.178012356322452$ ), followed by correlation analysis. After stratifying patients based on their risk scores into high-risk and low-risk groups using the median as the threshold, they were evaluated with Kaplan-Meier curves. In both the training and test sets, the high-risk group showed a notably reduced overall survival (OS) in comparison to the low-risk group (Figure 5D, 5E). Furthermore, analysis of the ROC curve showed that the area under the curve (AUC) values for both the training and test groups after 1, 3, and 5 years were all greater than 0.8 (Figure 5F, 5G), demonstrating strong verification performance of the model.

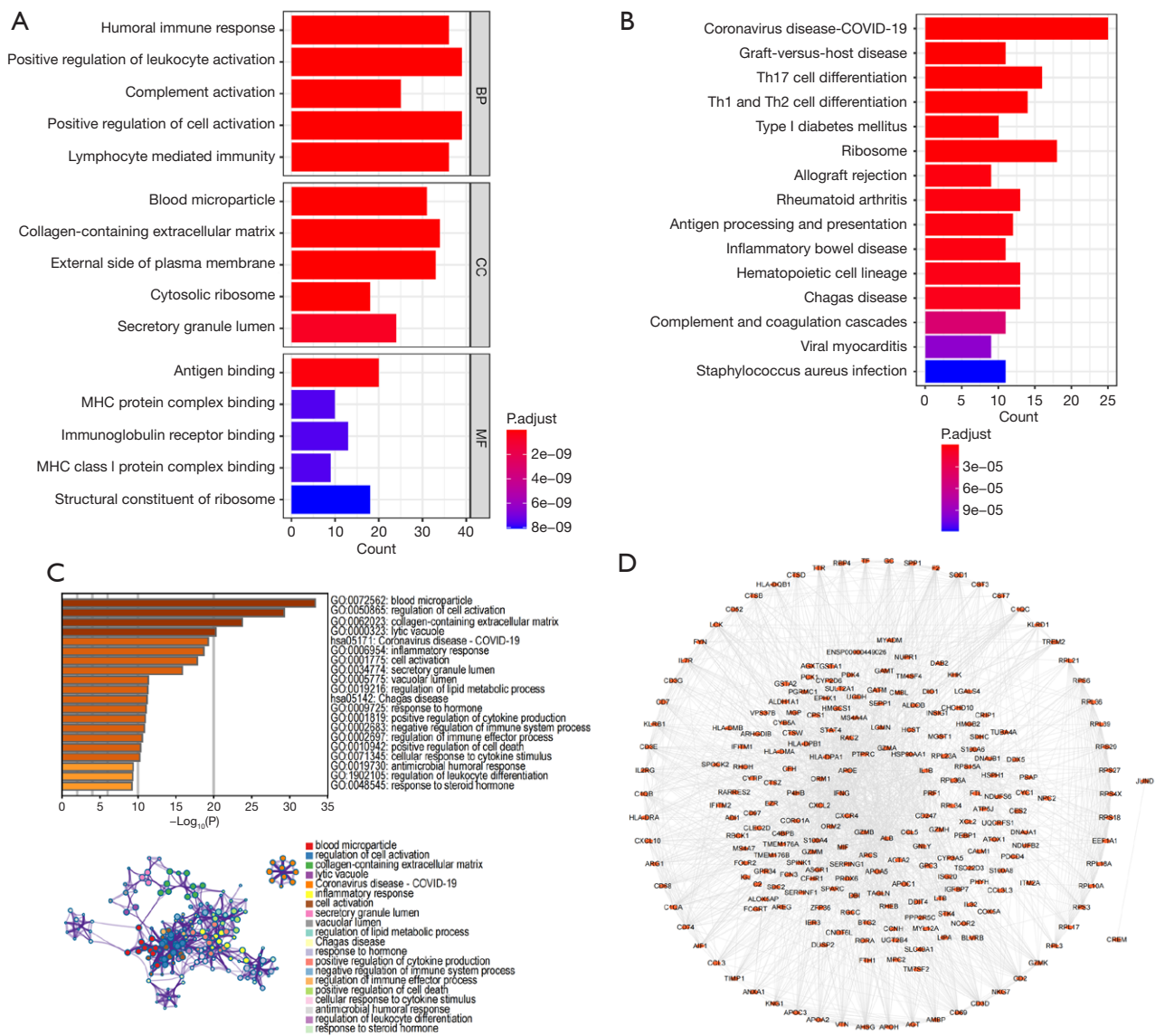
### Validation of the prognostic model's robustness using external datasets

The preprocessed survival data of LIHC patients from public



**Figure 3** CellPhoneDB locates communication between cells. (A) Bubble diagram of receptor and ligand information among 8 types of cells. Red and blue represent the strong and weak interaction of signal axis. (B) Cell interaction network among 11 types of cells, edge width indicates the probability and strength of communication between cells. (C) Comparison of the total number of interactions of the communication network between 11 types of cells, decreasing from left to right, the strongest being the ILC. ILC, innate lymphoid cell; NKT, natural killer T cell.

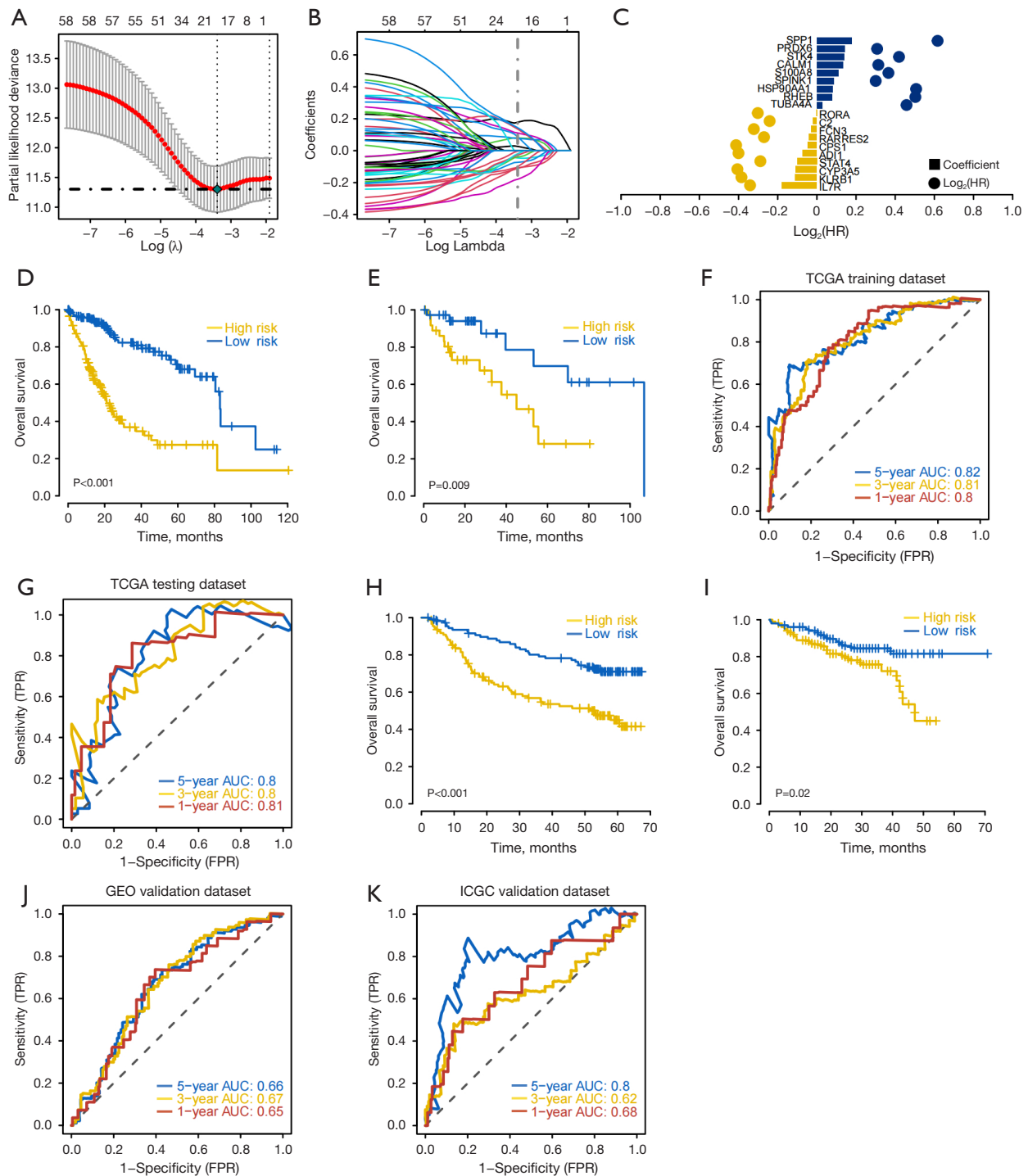




**Figure 4** Enrichment analysis of prognosis-related genes. (A,B) GO-KEGG enrichment analysis based on ClusterProfiler. (C) Metascapable GO-KEGG enrichment analysis, the figure above is a cluster network composed of enrichment pathways, where nodes sharing the same cluster are usually close to each other. (D) Protein interaction network, based on String database, visualized by Cytoscape software. MHC, major histocompatibility complex; BP, biological process; CC, cellular component; MF, molecular function; COVID-19, coronavirus disease 2019; Th17 cell, T helper cell 17; Th1 cell, T helper cell 1; Th2 cell, T helper cell 2; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

databases (GSE14520 and ICGC) were obtained. The model was utilized to forecast the clinical categorization of LIHC patients within the GEO database. Kaplan-Meier analysis was employed to assess survival disparities between the predicted groups, and the prediction model was examined for stability. The findings indicated a significant reduction in OS

within the high-risk group as opposed to the low-risk group in both external validation datasets from the GEO and ICGC (Figure 5H,5I). Furthermore, the accuracy of the model in predicting patient outcomes was validated by analyzing ROC curves with external datasets, demonstrating its significant prognostic utility (Figure 5J,5K).



**Figure 5** Prognosis-related genes and prediction model. (A) Ten-fold cross-validation of tuning parameter selection in the LASSO model to determine the minimum lambda value. (B) Distribution of LASSO coefficients of prognostic genes and the gene combination at the minimum lambda value. (C) Coefficients of LASSO genes. (D,E) Survival curves of TCGA training set and test set models. (F,G) ROC curves (1-, 3- and 5-year OS) of the TCGA training set and test set models. (H,I) Survival curves of GEO and ICGC models. (J,K) ROC curves (1-, 3- and 5-year OS) of GEO and ICGC models. HR, hazard ratio; TCGA, The Cancer Genome Atlas; AUC, area under the curve; TPR, true positive rate; FPR, false positive rate; LASSO, least absolute shrinkage and selection operator; GEO, Gene Expression Omnibus; ICGC, International Cancer Genome Consortium; ROC, receiver operating characteristic.

### *An exploration of the clinical predictive value of the model using multi-omics*

The tumor microenvironment consists of immune cells, tumor-associated fibroblasts, inflammatory factors, growth factors, cancer cells, and extracellular matrix, all of which play a vital role in the detection of tumors, patient prognosis, and treatment response. In order to explore the molecular mechanisms driving the progression of HCC, our analysis focused on the correlation between the risk score and the infiltration of immune cells within tumor tissues. The findings indicated that there was variability in the distribution of immune levels of various immune factors within the samples (Figure 6A). Additionally, significant correlations were observed among multiple immune factors (Figure 6B). It is worth mentioning that samples in the low-risk category showed increased levels of naive B cells and CD8 T cells, while having decreased levels of macrophages M0 and macrophages M2 (Figure 6C). The risk score demonstrated a significant positive association with macrophages M2, dendritic cells activated, T cells CD4 memory activated, neutrophils, eosinophils, and macrophages M0, while displaying a negative association with macrophages M1, plasma cells, T cells CD8, B cells naive and T cells CD4 memory resting (Figure 6D). Furthermore, combining surgery with chemotherapy for early liver cancer showed evident efficacy. We utilized the R package “pRRophetic” to forecast the sensitivity of chemotherapy for individual tumor samples based on GDSC database drug sensitivity data, revealing significant correlations between risk score levels and drug sensitivity to bexarotene, bleomycin, camptothecin, cisplatin, doxorubicin, and gemcitabine (Figure 6E). Furthermore, notable variances in microsatellite instability (MSI) levels were identified between the high-risk and low-risk groups, whereas no statistically significant distinctions in Neoantigen and tumor mutation burden (TMB) were evident between the aforementioned groups (Figure 6F-6H). Exploring the correlation between key genes and immune factors from the TISIDB database further elucidated immune factor distinctions between high- and low-risk groups (Figure 7A-7E).

### *Specific signaling mechanism related to the prognostic model*

The research examined particular communication pathways linked to the high-low risk relationship model

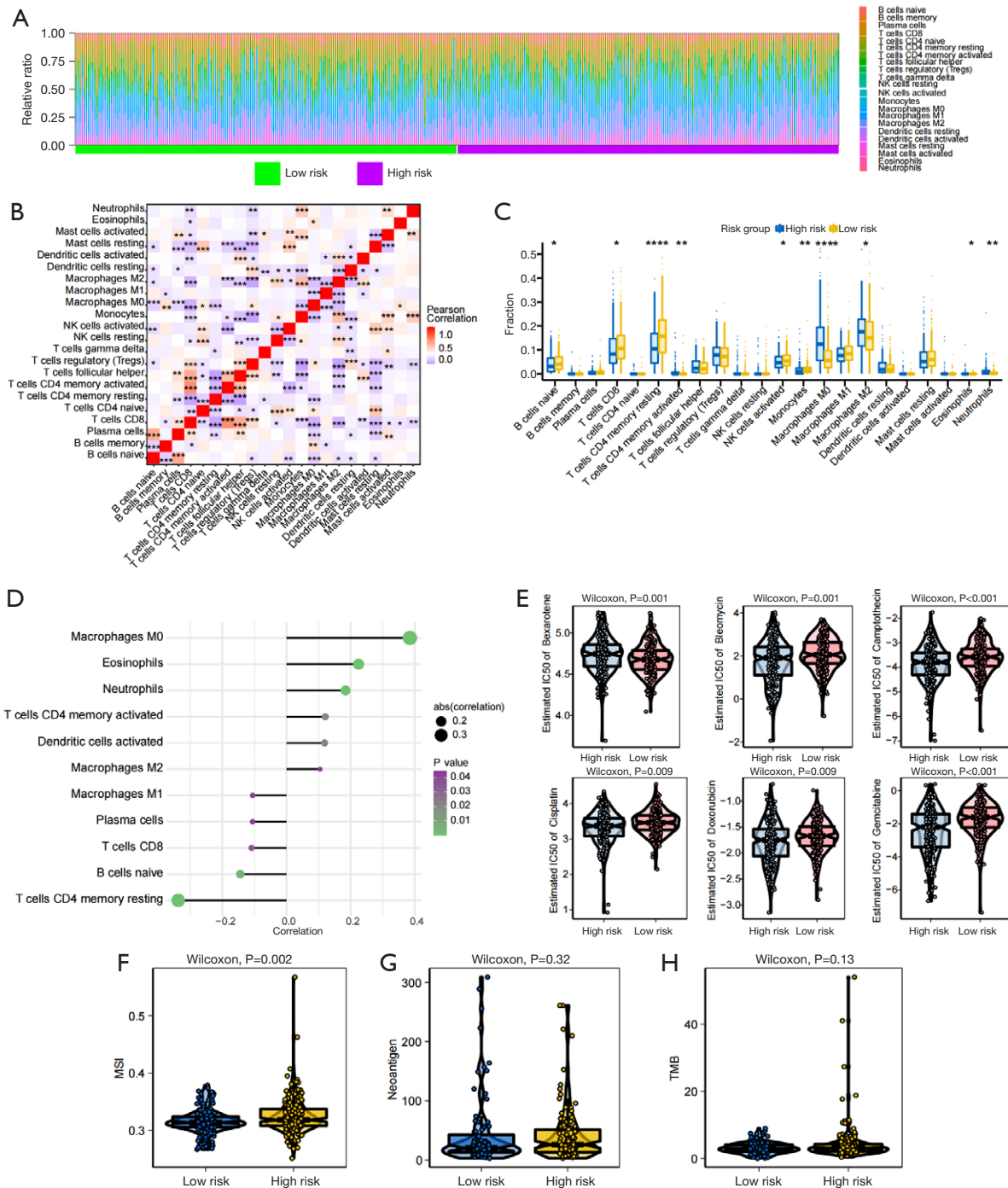
in order to reveal the molecular processes that explain how risk scores affect tumor advancement. GSEA analysis revealed significant enrichment of various pathways. GO analysis revealed pathways including CELLULAR\_AMINO\_ACID\_CATABOLIC\_PROCESS and POSITIVE\_REGULATION\_OF\_CELL\_CYCLE\_PHASE\_TRANSITION, whereas KEGG analysis pinpointed pathways such as COMPLEMENT\_AND\_COAGULATION\_CASCADES and FATTY\_ACID\_METABOLISM. Notably, certain highly significant pathways were separately clustered (Figure 8A, 8B). Furthermore, GSVA analysis revealed that the two patient groups exhibited differential pathways that were predominantly enriched in UNFOLDED\_PROTEIN\_RESPONSE, MYC\_TARGETS\_V1, MTORC1\_SIGNALING, BILE\_ACID\_METABOLISM, and COAGULATION (Figure 8C), suggesting that alterations in these signaling pathways among patients in high- and low-risk groups significantly impacted the prognosis of individuals with liver cancer.

### *Incidence risk and independent prognosis analysis*

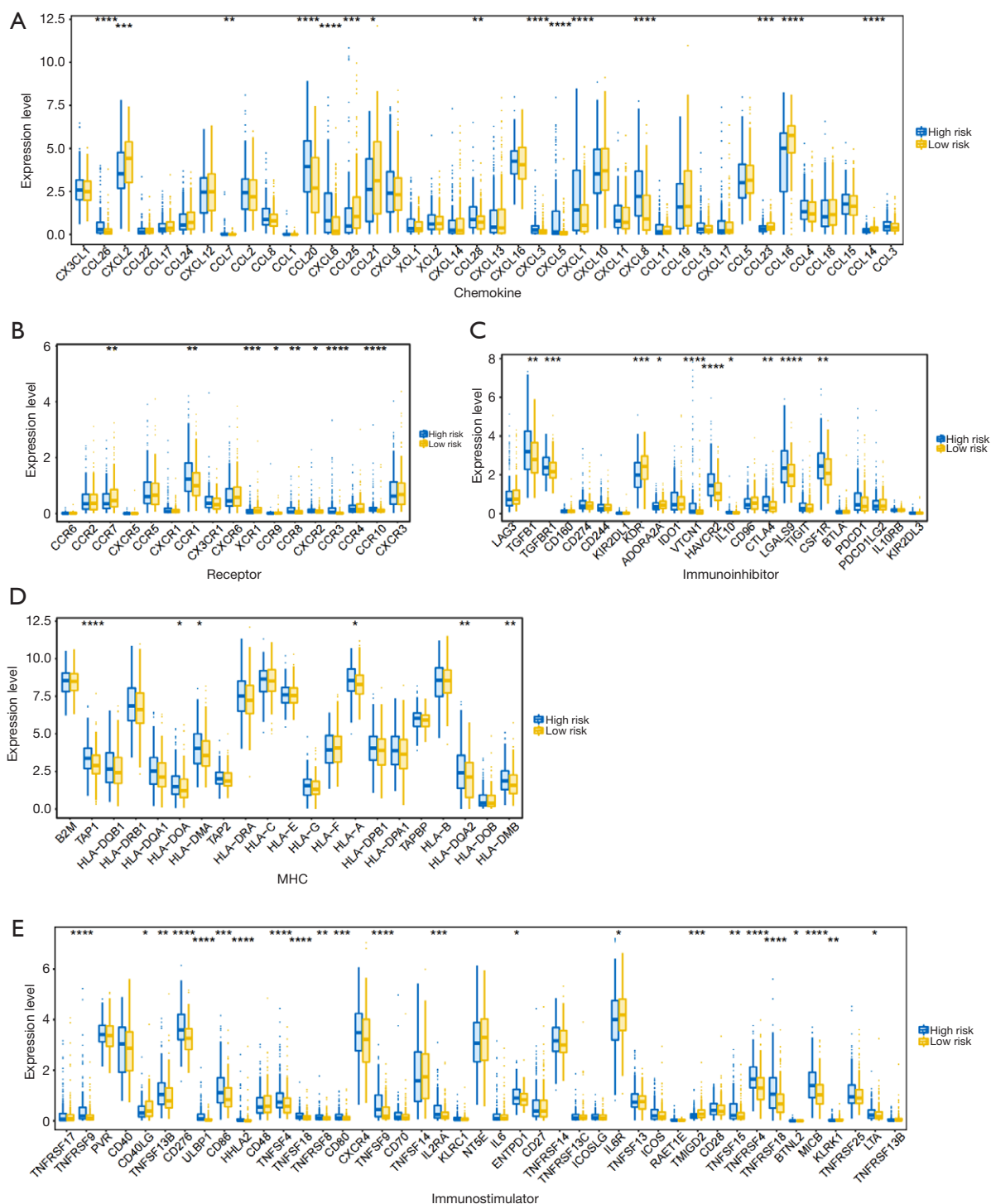
Patient clinical information and risk assessments from high- and low-risk groups were combined and used to create a nomogram illustrating the results of regression analysis. The results from logistic regression analysis indicated that different clinical markers of liver cancer in all samples, as well as the range of risk score values, played a role in several scoring procedures (Figure 9A). Additionally, predictive analysis on the 3- and 5-year OS of liver cancer was conducted, demonstrating a closer alignment between predicted and observed OS, underscoring the model's robust predictive performance (Figure 9B).

### *Evaluation of model gene ceRNA networks*

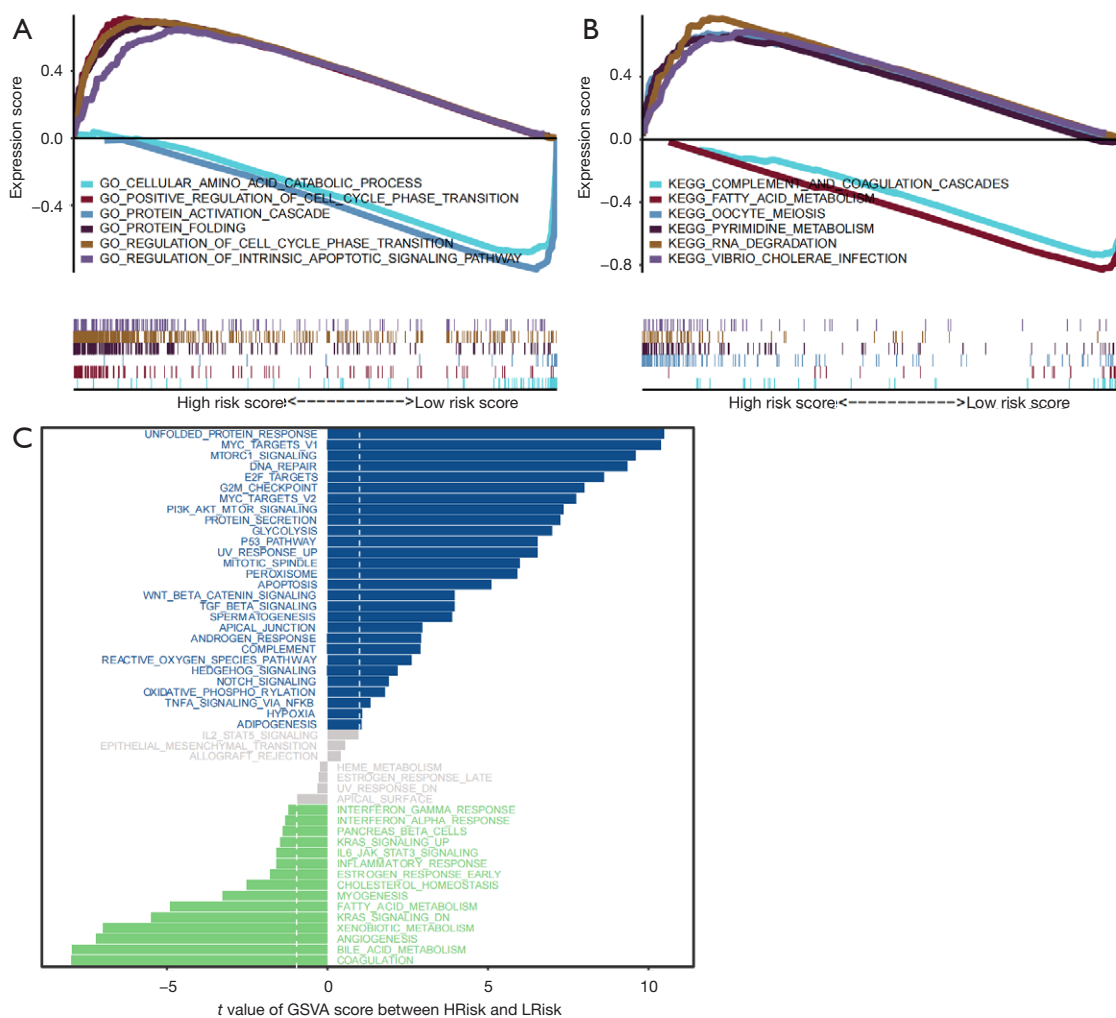
A collective of 460 microRNAs associated with liver cancer was identified through the HMDD database. By harnessing the miRWalk and The Encyclopedia of RNA Interactomes (ENCORI) databases, a comprehensive roster was compiled, encompassing potential miRNAs and long non-coding RNAs (lncRNAs) pertinent to 19 genes featured in the model. Initially, the mRNA-miRNA relationship pairs related to disease miRNAs were only retained (including 6 mRNAs and 17 miRNAs) using the miRWalk database to extract the targetScan or miRDB database detectable modeling gene-related



**Figure 6** Multi-omics research explores the clinical predictive value of the model. (A) Relative percentages of 22 immune cell subsets in patients in high-risk and low-risk groups. The green bar represents the low-risk group, while the purple bar signifies the high-risk group. (B) Pearson correlation between 22 types of immune cells, red indicates positive correlation and purple indicates negative correlation. (C) Differences in immune cell content between high-risk and low-risk patients, blue indicates high-risk patients and yellow indicates low-risk patients. (D) Correlation of risk score with immune cells. (E) Sensitivity analysis of risk score and common chemotherapeutic drugs. (F-H) Differences of TMB, MSI, neoantigen in high- and low-risk groups. Statistical significance was determined at the following thresholds: \*, P<0.05; \*\*, P<0.01; \*\*\*, P<0.001; and \*\*\*\*, P<0.0001. LRisk, low-risk group; HRisk, high-risk group; NK cells, natural killer cells; abs, absolute value; MSI, microsatellite instability; TMB, tumor mutation burden; IC50, half maximal inhibitory concentration.



**Figure 7** The relationship between the prognostic model and immune factors. Expression differences of chemokines (A), receptors (B), immunoinhibitors (C), MHC (D), and immunostimulators (E) between high and low risk patients, blue indicates high risk patients, yellow indicates low risk patients. Statistical significance was determined at the following thresholds: \*, P<0.05; \*\*, P<0.01; \*\*\*, P<0.001; and \*\*\*\*, P<0.0001. LRisk, low-risk group; HRisk, high-risk group; MHC, major histocompatibility complex.



**Figure 8** Model-dependent enrichment analysis. (A,B) GO-KEGG pathways enriched in high- and low-risk groups. (C) Pathway activity was scored by GSVA between high-risk and low-risk groups to explore differences in pathways between high-risk and low-risk groups, and the background gene set was Hallmark. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; HRisk, high-risk group; LRisk, low-risk group; GSVA, gene set variation analysis.

mRNA-miRNA relationship pairs (that acquired a total of 470 miRNAs) (Figure 9C). Based on these miRNAs to predict the interaction of lncRNAs, a total of 3,477 pairs of interactions (including 10 miRNAs and 2,202 lncRNAs) were predicted. Finally, utilizing Cytoscape (v3.7), a ceRNA network was assembled, integrating 5 mRNAs, 10 miRNAs, and 2,202 lncRNAs (Figure 9D).

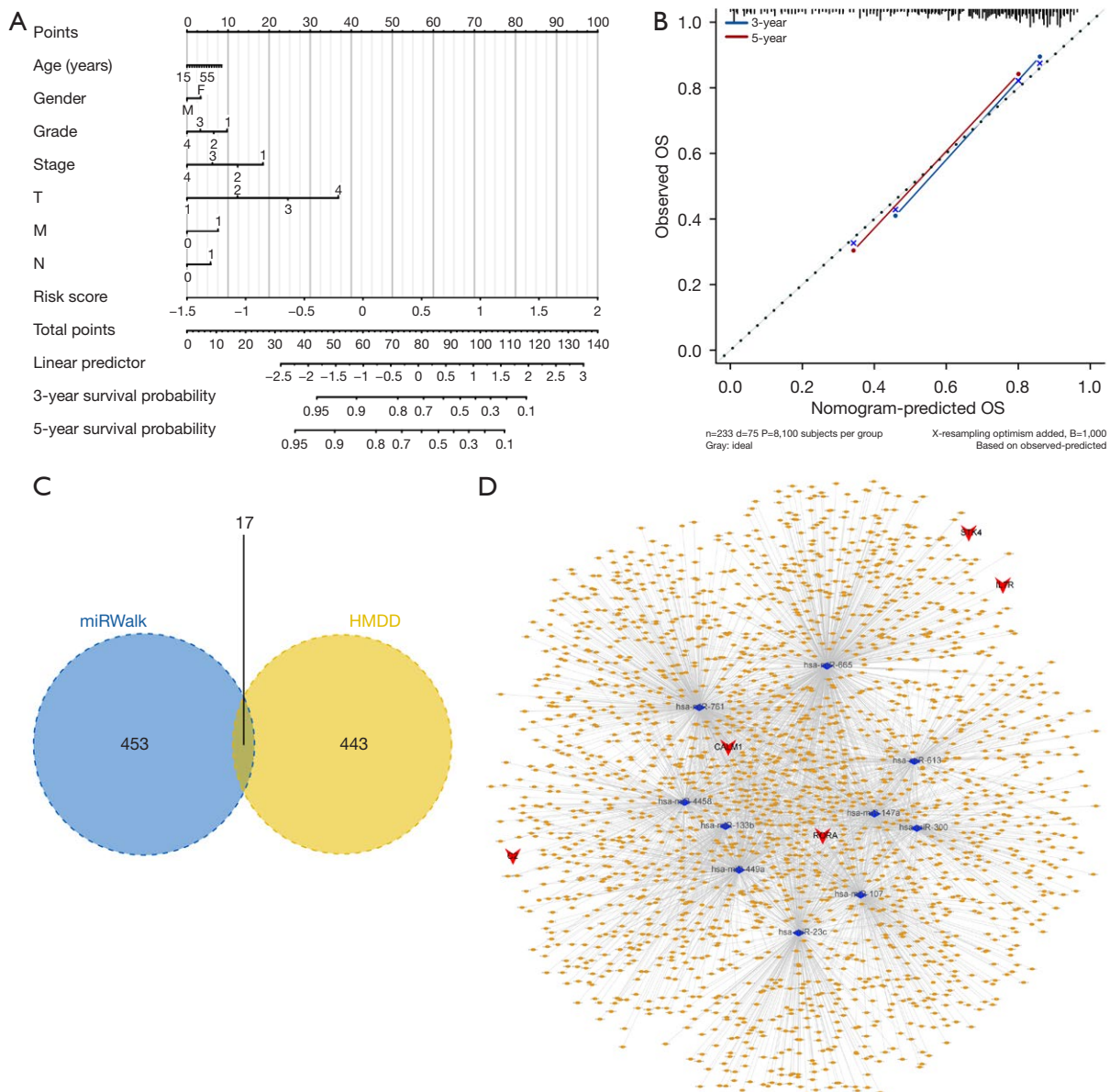
### CeRNA-network gene modeling

The GeneCards database (<https://www.genecards.org/>) was used to identify a thorough collection of 8,748 genes associated with liver cancer. Further

investigations concentrated on analyzing the gene expression patterns of five important genes in the ceRNA network in addition to the top 20 genes ranked by their relevance score. The investigation revealed an important finding: a strong connection was found between the expression levels of these specific model genes and several genes linked to liver cancer (Figure 10A,10B).

### Clinical experimental validation

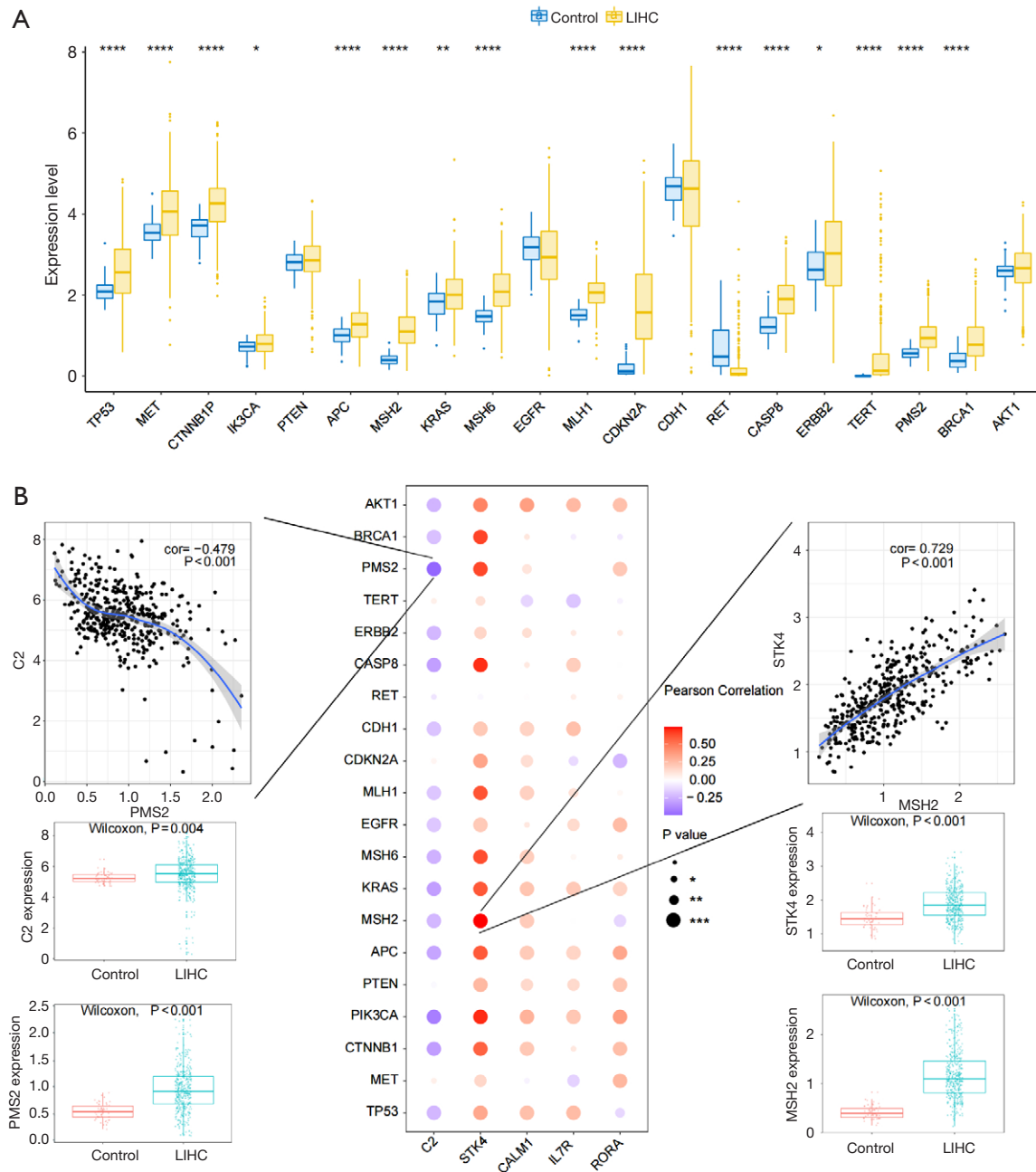
The analysis in our findings investigates the relationship between the levels of expression of five particular genes and clinical indicators (Figure 11A-11G). This



**Figure 9** Construction and performance evaluation of nomogram model related to risk score. (A) Nomogram associated with the risk model. (B) Model-dependent correction curves. (C) Venn diagram identifying the intersection miRNAs of the miRWalk and HMDD databases. (D) ceRNA network of model genes, red indicates mRNA, blue indicates miRNA, and orange indicates lncRNA. OS, overall survival; HMDD, Human microRNA Disease Database; miRNA, microRNA; ceRNA, competing endogenous RNA; mRNA, messenger RNA; lncRNA, long non-coding RNA.

examination unveiled statistically significant variations in the expression of *STK4* and *RORA*, alongside other genes, across distinct grade and stage subgroups. Furthermore, the expression of the five key ILC genes was verified by IHC. Representative IHC images of *STK4* and *CALM1* (Figure 12A), as well as *C2*, *RORA*, and *IL7R* (Figure 12B)

in different T stages of HCC were showed. The IHC findings suggested a discernible trend: *STK4* and *CALM1* displayed a positive correlation with malignancy potential in HCC, contrasting with *C2*, *RORA*, and *IL7R*, which exhibited an inverse association with the disease’s malignancy potential.



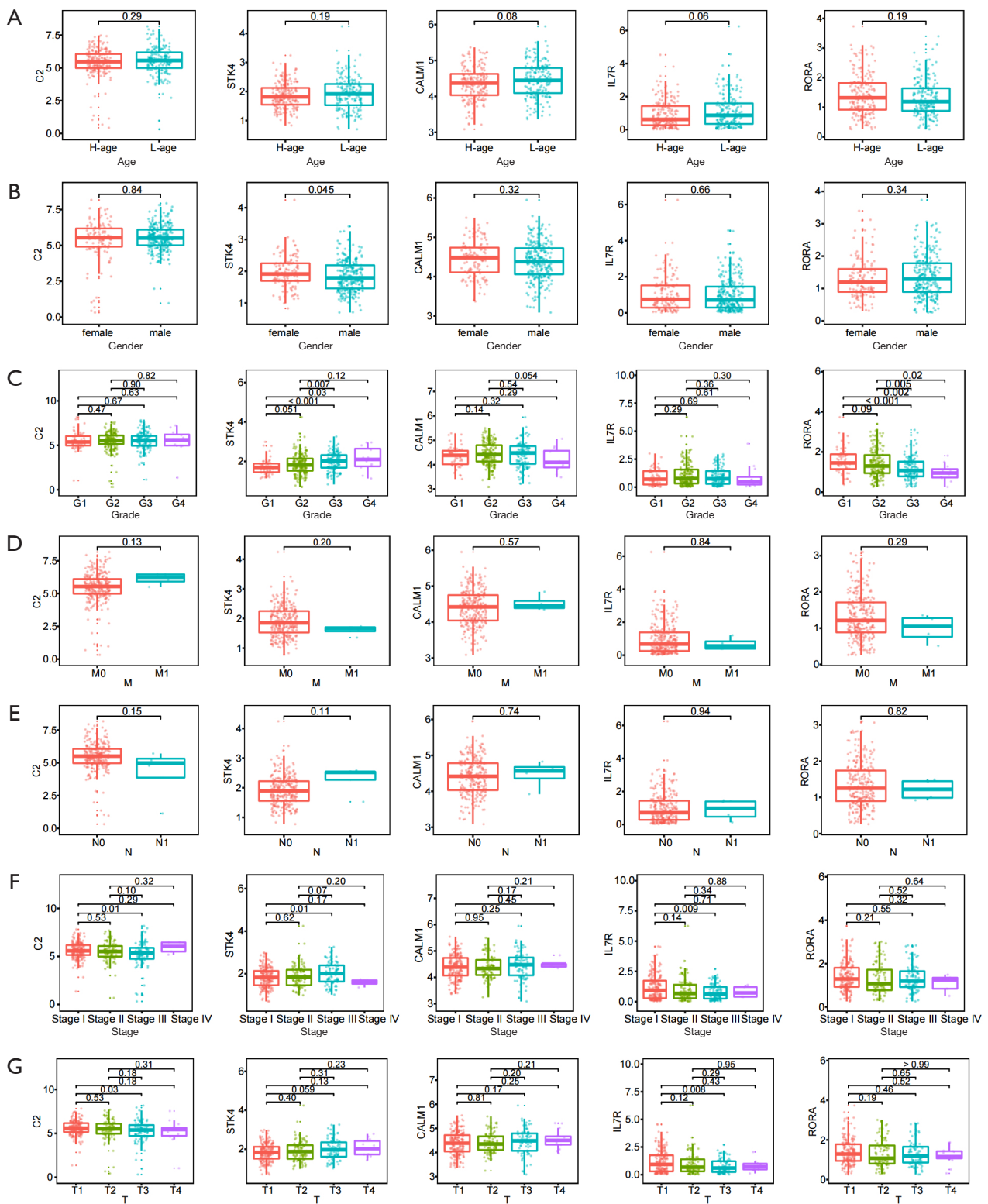
**Figure 10** Analysis of tumor regulatory genes in liver cancer. (A) Differences in expression of tumor-regulated genes, blue indicates control patients, yellow indicates tumor patients. (B) Pearson correlation analysis of model genes and tumor genes, blue indicates negative correlation, red indicates positive correlation. Statistical significance was determined at the following thresholds: \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ ; and \*\*\*\*,  $P < 0.0001$ . LIHC, liver hepatocellular carcinoma; cor, correlation.

## Discussion

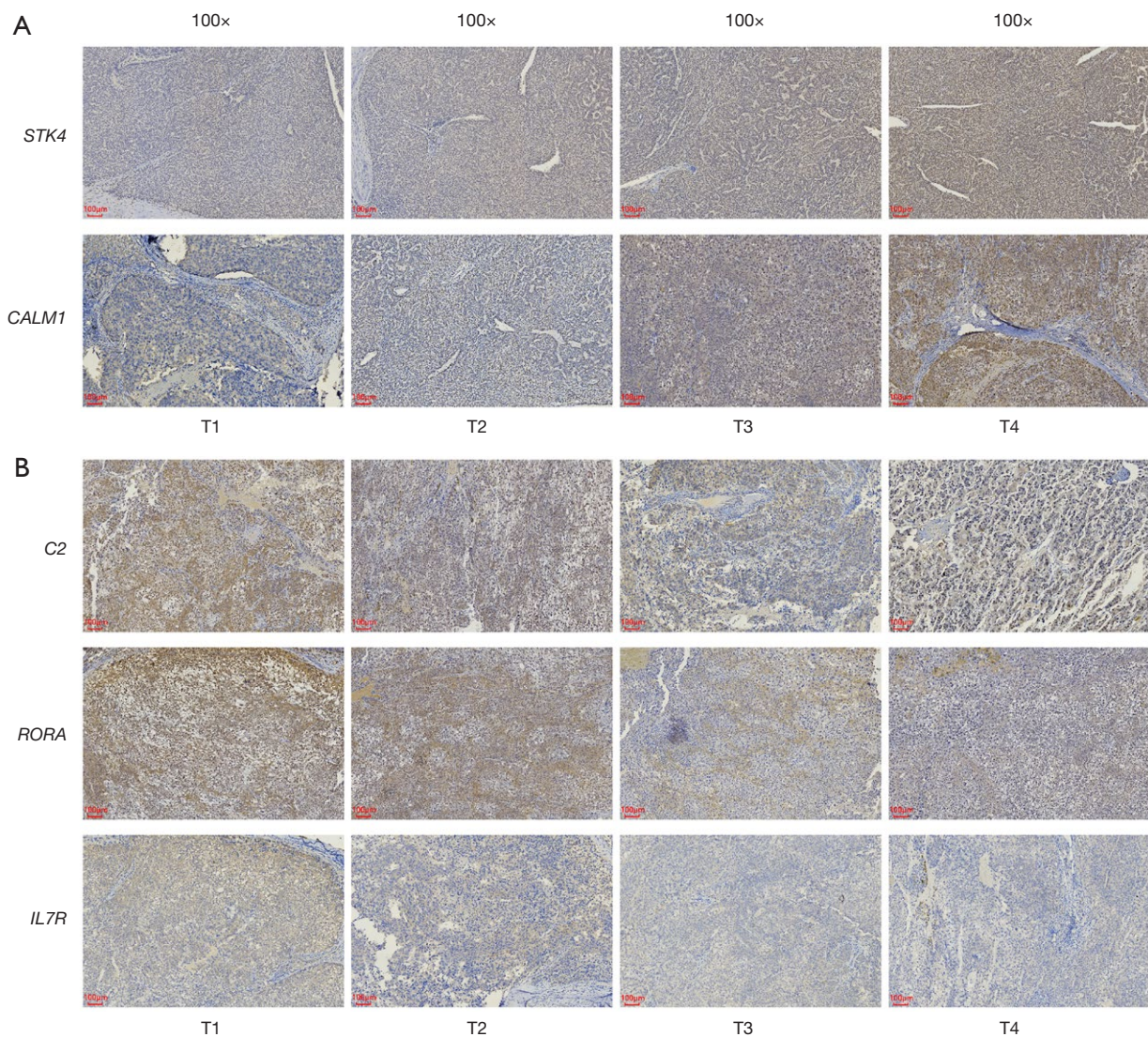
The development of scRNA-seq technology has helped progress our knowledge of tumorigenesis and the microenvironment of tumors. Previous study (19) has

hinted at the involvement of a subset of inducible T cell co-stimulators, ILC2 cells, in HCC's poor prognosis. On the other hand, the presence of ILC3, a subset of helper ILCs, in the tumor tissues of patients with non-small





**Figure 11** Risk score and clinical relevance. The difference between risk score and age (A), gender (B), grade (C), M (D), N (E), stage (F), T (G),  $P < 0.05$  is considered to be statistically significant. H-age stands for the higher-age group, while L-age denotes the lower-age group. We define the higher-age group as individuals who are above 60 years, whereas the lower-age group consists of those who are 60 years or younger.



**Figure 12** The expression of the five key ILC genes verified by IHC. Representative 100× IHC images of (A) *STK4*, *CALM1*, (B) *C2*, *RORA*, and *IL7R* in different T stages of HCC. ILC, innate lymphoid cell; IHC, immunohistochemistry; HCC, hepatocellular carcinoma.

cell lung cancer is linked to a better prognosis, possibly because it aids in the formation and advancement of tertiary lymphoid structures (20). However, insights into ILCs in HCC patients remain limited. Recently, Song *et al.* developed a signature based on natural killer cell marker genes to assess its prognosis and immune response in lung adenocarcinoma (21), inspiring an exploration of ILC marker genes in HCC using scRNA analysis.

A prognostic model was created in our research utilizing the ILC signature in the TCGA group, which was then validated with external datasets from the GEO and ICGC groups. The ILC low-risk score was associated with

substantial immune cell infiltration, MSI, and diversity levels. Moreover, individuals diagnosed with HCC in the high-risk category showed notably reduced IC50 levels for drug effectiveness compared to those in the low-risk category, indicating that drug treatment could be more advantageous for the high-risk HCC demographic.

The ILC prognostic model primarily comprised 58 genes, with 5 key genes—*C2*, *STK4*, *CALM1*, *IL7R*, and *RORA*—playing pivotal roles. Among these, *C2*, a complement protein, has been linked to age-related macular degeneration and type 2 diabetes (22,23). *STK4* modulation has been associated with inhibiting pro-

inflammatory cytokine secretion, potentially offering protective effects against chronic inflammation associated with HCC (24). The lncRNA *GAS5* suppresses the epithelial-to-mesenchymal transition (EMT) and enhances the sensitivity to paclitaxel in prostate cancer through the regulation of the *miR-18a-5p/STK4* axis, as evidenced in cancer stem cell study (25). *CALM1* overexpression is observed in multiple cancers and is significantly upregulated in gastric cancer, correlating with poor OS (26). Additionally, it aids in the oncogenic advancement of esophageal squamous cell carcinoma by promoting EMT and reducing the sensitivity to EGFR inhibitors (27). *IL7R* is downregulated in primary cutaneous T-cell lymphoma, serving as a prognostic biomarker for disease progression (28). Additionally, it restrains tumor growth by modulating the composition of immune infiltrating cells within the tumor microenvironment in lung adenocarcinoma (29). Its expression level is inversely associated with the pathological stage of skin cutaneous melanoma (30). *RORA* transcription factor expression inhibits T cell maturation in early embryonic stages while promoting *ILC2* development in the thymus (31). Moreover, Ma *et al.* reported that *RORA* overexpression effectively impedes the progression of gastric cancer by disrupting the interaction between *circGSK3B*'s partner regulator, *EZH2*, and the *RORA* promoter (32). Our findings indicated that the ILC signature might offer potential targets for laboratory and clinical therapies to elucidate HCC's underlying molecular mechanisms, consistent with existing reports. Furthermore, our model demonstrated robust predictive power for patient outcomes across different cohorts, reflecting the underlying mechanisms of ILCs. Variances in immune cell infiltration and levels of inflammation were noted among high- and low-risk groups, which could impact the prognosis and response to treatment of HCC.

Initially, our developed model demonstrated strong predictive efficacy for patient outcomes within both the training set of the TCGA cohort and the external validation dataset from the GEO and ICGC cohort. The robust predictive capability of the ILC signatures spurred our exploration into their underlying mechanisms. CIBERSORT was employed to assess the extent of immune cell infiltration in cohorts categorized as high- and low-risk. The group at high risk displayed increased presence of macrophages M0 and M2, whereas the group at low risk demonstrated higher levels of CD8 T cells, CD4 memory T cells, and activated natural killer cells. The results indicate a possible link between the make-up of immune cells and the

prognosis, specifically emphasizing the connection between certain immune cell categories and the poorer prognosis seen in high-risk HCC individuals.

Further investigation into the expression of immunoregulatory factors was conducted to elucidate the association between our model and immune and inflammatory processes. Increased concentrations of the chemokine *CXCL8* and its corresponding receptor *CXCR2* were detected in the high-risk cohort. Prior research has emphasized the upregulation of *CXCL8* in pancreatic cancer patients, who show elevated levels of *CXCR2* on CD68 macrophages in both the periphery and tumor site, which is associated with later stage tumors and poor prognosis (33). Furthermore, a variety of immunosuppressants such as *TGFB1*, *VTCN1*, *HAVCR2*, *IL10*, *CTLA4*, *LGALS9*, and *CSF1R* showed significant increases in expression in the high-risk cohort. Targeting immunosuppressive checkpoints on tumor-associated macrophages and myeloid cells may enhance immune cell activity within the microenvironment (34). Our findings align with previous research and potentially contribute to the adverse prognosis observed in HCC patients with high-risk scores.

Additionally, the high-risk group showed enrichment of pathways like apoptosis, mTORC1 signaling, DNA repair, and the p53 pathway according to KEGG and GO analyses, highlighting important signaling pathways in the development and advancement of tumors. On the other hand, the group with low risk showed an increase in pathways related to interferon-gamma, interferon-alpha, and inflammatory reactions. The good prognosis seen in the low-risk category could be due to increased immune cell presence. As a result, the ILC signature model highlights a clear difference in immune landscapes between high- and low-risk groups in HCC, showing that the high-risk group has notably weaker immune profiles than the low-risk group.

Moreover, variations in immune infiltration and inflammatory activity across risk groups motivated our investigation into the potential of an ILC signature to predict drug sensitivity during patient treatment. Examining the response to chemotherapy in high- and low-risk categories based on GDSC data revealed increased IC50 levels for bleomycin, camptothecin, cisplatin, doxorubicin, and gemcitabine in the low-risk group, whereas bexarotene displayed higher IC50 values in the high-risk group. These medications are frequently employed in the treatment of different types of solid tumors, such as colorectal cancer, breast cancer, squamous cell carcinoma, malignant lymphoma, bladder cancer, and liver cancer (35-40). Except

for bexarotene, the group at high risk showed reduced sensitivity to these medications at the IC50 levels, indicating a possible improvement in chemotherapy effectiveness. Following this, analysis of MSI, neoantigen, and TMB findings among various risk categories revealed that the low-risk category exhibited decreased MSI levels, with no notable variances in neoantigen and TMB levels, although there was a trend towards lower TMB levels in the low-risk category. The results indicate that low-risk HCC shows reduced immunogenicity, despite a significant rise in CD8 T cell infiltration in low-risk tumors compared to high-risk tumors.

The tumor immune response is influenced by numerous factors, and the ceRNA network, comprising lncRNA-miRNA-mRNA interactions, plays a critical role in regulating tumor behavior and drug resistance. For instance, in bladder cancer, a ceRNA network orchestrated by lncRNAs governs cancer cell proliferation, invasion, and apoptosis (41). Additionally, the *lncRNA-bas-mir-222-3p-XBP1* axis has been implicated in regulating benign prostatic hyperplasia through autophagy (42). Leveraging miRWalk and HMDD databases, we investigated the interconnections among ILC signature genes within a ceRNA network. Subsequently, we identified 5 hub nodes within this network as pivotal genes of our model (*C2*, *STK4*, *CALM1*, *IL7R*, and *RORA*). Analyzing the expression landscape of tumor-related genes, specifically the top 20 genes associated with HCC sourced from GeneCards, we found significant associations with our 5 key genes. Notably, a positive correlation was observed between *STK4* and *MSH2* (cor = 0.729, P < 0.001). *MSH2*, a key gene in the mismatch repair system, is crucial for HCC prognosis, with its mutations impacting patient outcomes (43). Conversely, a negative correlation was noted between *C2* and *PMS2* (cor = -0.479, P < 0.001). *PMS2* harbors mutations leading to systemic mismatch repair system deficiency syndrome and pediatric intestinal cancer (44). These findings from our model suggest a potential linkage between ILC signatures and the expression of HCC-related genes, which could partly underlie the unfavorable prognosis in HCC patients. Furthermore, we evaluated the association of the five key genes with various prognostic indicators in an HCC patient cohort. Notably, *STK4* exhibited significant correlations with gender, grade, and stage of HCC patients. Additionally, our experimental results provided preliminary evidence supporting the oncogenic roles of *STK4* and *CALM1*, as well as the tumor-suppressive roles of *C2*, *RORA*, and *IL7R* in HCC.

While our study has yielded valuable insights, several limitations should be acknowledged. Additional research

is required to evaluate the gene expression and prognostic importance at different molecular levels, including protein expression and DNA methylation patterns. Secondly, our model primarily focused on ILC cell marker genes, neglecting the complexity of the heterogeneous tumor microenvironment, which may limit the prognostic efficacy of the model. Lastly, our analysis was conducted at a single level, highlighting the necessity for additional research to elucidate the underlying mechanisms linking the expression of ILC signatures and the prognosis of HCC.

## Conclusions

Our study presents a validated 5-key gene signature based on ILC cell signature genes, offering a powerful tool for predicting HCC prognosis. This signature shows potential as a predictive biomarker for making clinical decisions, potentially leading to personalized treatments for better patient results.

## Acknowledgments

We would like to thank the databases TCGA, ICGC and GEO that give free access to many useful data. We appreciate the contributors of these public databases used in this study.

**Funding:** This work was financially supported by the Natural Science Foundation of Anhui Province (No. 2108085MH283); Key Programs of Wannan Medical College (No. WK2022ZF26); and Climbing Scientific Peak Project for Talents, The Second Affiliated Hospital of Wannan Medical College (No. DFJH2022012).

## Footnote

**Reporting Checklist:** The authors have completed the TRIPOD reporting checklist. Available at <https://tcr.amegroups.com/article/view/10.21037/tcr-24-725/rc>

**Data Sharing Statement:** Available at <https://tcr.amegroups.com/article/view/10.21037/tcr-24-725/dss>

**Peer Review File:** Available at <https://tcr.amegroups.com/article/view/10.21037/tcr-24-725/prf>

**Conflicts of Interest:** All authors have completed the ICMJE uniform disclosure form (available at <https://tcr.amegroups.com/article/view/10.21037/tcr-24-725/coif>). The authors

have no conflicts of interest to declare.

**Ethical Statement:** The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was reviewed and approved by the Ethics Committee of The Second Affiliated Hospital of Wannan Medical College (Wuhu, China, WYEFYLS2023139). Informed consent was taken from all participants.

**Open Access Statement:** This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

## References

- Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021;71:209-49.
- Forner A, Reig M, Bruix J. Hepatocellular carcinoma. *Lancet* 2018;391:1301-14.
- Howell J, Pedrana A, Schroeder SE, et al. A global investment framework for the elimination of hepatitis B. *J Hepatol* 2021;74:535-49.
- Lu Y, Yang A, Quan C, et al. A single-cell atlas of the multicellular ecosystem of primary and metastatic hepatocellular carcinoma. *Nat Commun* 2022;13:4594.
- Warner K, Ohashi PS. ILC regulation of T cell responses in inflammatory diseases and cancer. *Semin Immunol* 2019;41:101284.
- Jacquelot N, Seillet C, Vivier E, et al. Innate lymphoid cells and cancer. *Nat Immunol* 2022;23:371-9.
- Dadi S, Chhangawala S, Whitlock BM, et al. Cancer Immunosurveillance by Tissue-Resident Innate Lymphoid Cells and Innate-like T Cells. *Cell* 2016;164:365-77.
- Trabanelli S, Chevalier MF, Derré L, et al. The pro- and anti-tumor role of ILC2s. *Semin Immunol* 2019;41:101276.
- Heinrich B, Gertz EM, Schäffer AA, et al. The tumour microenvironment shapes innate lymphoid cells in patients with hepatocellular carcinoma. *Gut* 2022;71:1161-75.
- Matsuoka T, Yashiro M. Biomarkers of gastric cancer: Current topics and future perspective. *World J Gastroenterol* 2018;24:2818-32.
- Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. *Cell* 2021;184:3573-3587.e29.
- Aran D, Looney AP, Liu L, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* 2019;20:163-72.
- Efremova M, Vento-Tormo M, Teichmann SA, et al. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat Protoc* 2020;15:1484-506.
- Wu T, Hu E, Xu S, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* 2021;2:100141.
- Geeleher P, Cox N, Huang RS. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS One* 2014;9:e107468.
- Chen B, Khodadoust MS, Liu CL, et al. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol Biol* 2018;1711:243-59.
- Ferreira MR, Santos GA, Biagi CA, et al. GSEA score reveals molecular signatures from transcriptomes for biomaterials comparison. *J Biomed Mater Res A* 2021;109:1004-14.
- Hussaini HM, Seo B, Rich AM. Immunohistochemistry and Immunofluorescence. *Methods Mol Biol* 2023;2588:439-50.
- He Y, Luo J, Zhang G, et al. Single-cell profiling of human CD127(+) innate lymphoid cells reveals diverse immune phenotypes in hepatocellular carcinoma. *Hepatology* 2022;76:1013-29.
- Sivori S, Pende D, Quatrini L, et al. NK cells and ILCs in tumor immunotherapy. *Mol Aspects Med* 2021;80:100870.
- Song P, Li W, Guo L, et al. Identification and Validation of a Novel Signature Based on NK Cell Marker Genes to Predict Prognosis and Immunotherapy Response in Lung Adenocarcinoma by Integrated Analysis of Single-Cell and Bulk RNA-Sequencing. *Front Immunol* 2022;13:850745.
- Gold B, Merriam JE, Zernant J, et al. Variation in factor B (BF) and complement component 2 (C2) genes is associated with age-related macular degeneration. *Nat Genet* 2006;38:458-62.
- Steffen BT, Tang W, Lutsey PL, et al. Proteomic analysis of diabetes genetic risk scores identifies complement C2 and neuropilin-2 as predictors of type 2 diabetes:

- the Atherosclerosis Risk in Communities (ARIC) Study. *Diabetologia* 2023;66:105-15.
24. Li W, Xiao J, Zhou X, et al. STK4 regulates TLR pathways and protects against chronic inflammation-related hepatocellular carcinoma. *J Clin Invest* 2015;125:4239-54.
  25. Lu TT, Tao X, Li HL, et al. LncRNA GAS5 enhances tumor stem cell-like mediated sensitivity of paclitaxel and inhibits epithelial-to-mesenchymal transition by targeting the miR-18a-5p/STK4 pathway in prostate cancer. *Asian J Androl* 2022;24:643-52.
  26. Liu Y, Xu Y, Xiao F, et al. Comprehensive Analysis of a circRNA-miRNA-mRNA Network to Reveal Potential Inflammation-Related Targets for Gastric Adenocarcinoma. *Mediators Inflamm* 2020;2020:9435608.
  27. Liu T, Han X, Zheng S, et al. CALM1 promotes progression and dampens chemosensitivity to EGFR inhibitor in esophageal squamous cell carcinoma. *Cancer Cell Int* 2021;21:121.
  28. Rindler K, Jonak C, Alkon N, et al. Single-cell RNA sequencing reveals markers of disease progression in primary cutaneous T-cell lymphoma. *Mol Cancer* 2021;20:124.
  29. Wang X, Chang S, Wang T, et al. IL7R Is Correlated With Immune Cell Infiltration in the Tumor Microenvironment of Lung Adenocarcinoma. *Front Pharmacol* 2022;13:857289.
  30. Li G, Zhang J, Liu Y, et al. Analyzing Prognostic Hub Genes in the Microenvironment of Cutaneous Melanoma by Computer Integrated Bioinformatics. *Comput Intell Neurosci* 2022;2022:4493347.
  31. Ferreira ACF, Szeto ACH, Heycock MWD, et al. ROR $\alpha$  is a critical checkpoint for T cell and ILC2 commitment in the embryonic thymus. *Nat Immunol* 2021;22:166-78.
  32. Ma X, Chen H, Li L, et al. CircGSK3B promotes RORA expression and suppresses gastric cancer progression through the prevention of EZH2 trans-inhibition. *J Exp Clin Cancer Res* 2021;40:330.
  33. Zhang M, Huang L, Ding G, et al. Interferon gamma inhibits CXCL8-CXCR2 axis mediated tumor-associated macrophages tumor trafficking and enhances anti-PD1 efficacy in pancreatic cancer. *J Immunother Cancer* 2020;8:e000308.
  34. Chen HM, van der Touw W, Wang YS, et al. Blocking immunoinhibitory receptor LILRB2 reprograms tumor-associated myeloid cells and promotes antitumor immunity. *J Clin Invest* 2018;128:5647-62.
  35. Ghosh S. Cisplatin: The first metal based anticancer drug. *Bioorg Chem* 2019;88:102925.
  36. Mollica A, Stefanucci A, Feliciani F, et al. Delivery methods of camptothecin and its hydrosoluble analogue irinotecan for treatment of colorectal cancer. *Curr Drug Deliv* 2012;9:122-31.
  37. Fu J, Wang Y, Zhang J, et al. The safety and efficacy of transarterial chemoembolisation with bleomycin for hepatocellular carcinoma unresponsive to doxorubicin: a prospective single-centre study. *Clin Radiol* 2021;76:864.e7-864.e12.
  38. Monteran L, Ershaid N, Doron H, et al. Chemotherapy-induced complement signaling modulates immunosuppression and metastatic relapse in breast cancer. *Nat Commun* 2022;13:5797.
  39. Oh DY, Lee KH, Lee DW, et al. Gemcitabine and cisplatin plus durvalumab with or without tremelimumab in chemotherapy-naive patients with advanced biliary tract cancer: an open-label, single-centre, phase 2 study. *Lancet Gastroenterol Hepatol* 2022;7:522-32.
  40. Rocuzzo G, Fava P, Avallone G, et al. Time to next treatment and safety assessment in cutaneous-T-cell lymphomas: a retrospective analysis on patients treated with bexarotene and acitretin. *Br J Dermatol* 2022;187:1019-21.
  41. Lyu L, Xiang W, Zhu JY, et al. Integrative analysis of the lncRNA-associated ceRNA network reveals lncRNAs as potential prognostic biomarkers in human muscle-invasive bladder cancer. *Cancer Manag Res* 2019;11:6061-77.
  42. Zhou L, Li Y, Li J, et al. Decoding ceRNA regulatory network and autophagy-related genes in benign prostatic hyperplasia. *Int J Biol Macromol* 2023;225:997-1009.
  43. Hong S, Zhang J, Liu S, et al. Protein profiles reveal MSH6/MSH2 as a potential biomarker for hepatocellular carcinoma with microvascular invasion. *Hepatol Res* 2024;54:189-200.
  44. Herkert JC, Niessen RC, Olderode-Berends MJ, et al. Paediatric intestinal cancer and polyposis due to bi-allelic PMS2 mutations: case series, review and follow-up guidelines. *Eur J Cancer* 2011;47:965-82.

**Cite this article as:** Pan M, Yuan X, Peng J, Wu R, Chen X. Identification and validation of a novel innate lymphoid cell-based signature to predict prognosis and immune response in liver cancer by integrated single-cell RNA analysis and bulk RNA sequencing. *Transl Cancer Res* 2024;13(10):5395-5416. doi: 10.21037/tcr-24-725