

Construction and validation of a prognostic model for genes associated with myeloid differentiation in liver hepatocellular carcinoma: bulk RNA sequencing combined with Mendelian randomization and single-cell RNA sequencing

Zhenyu Lin

National Cancer Center/Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College

Lin Xu

National Cancer Center/Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College

Zirong Liu

National Cancer Center/Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College

Xiaozhun Huang

National Cancer Center/Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College

Yihong Ran

National Cancer Center/Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College

Dong Chen

National Cancer Center/Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College

Wei Zhang

National Cancer Center/Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College

Zhangkan Huang

charles846@126.com

National Cancer Center/Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College

Keywords: Hepatocellular Carcinoma, Myeloid cell differentiation, Prognostic genes, Single-cell RNA sequencing

Posted Date: October 20th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9586229/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Construction and validation of a prognostic model for genes associated with myeloid differentiation in liver hepatocellular carcinoma: bulk RNA sequencing combined with Mendelian randomization and single-cell RNA sequencing

Zhenyu Lin^{1†}, Lin Xu^{1†}, Zirong Liu¹, Xiaozhun Huang¹, Yihong Ran¹, Dong Chen¹, Wei Zhang¹, Zhangkan Huang^{1*}

¹ Department of Hepatopancreatobiliary Surgery, National Cancer Center/Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Shenzhen 518116, Guangdong, China

*** Correspondence:**

Zhangkan Huang

CHARLES846@126.COM.

[†] These authors have contributed equally to this work.

Abstract

Background: Myeloid cell differentiation is closely related to the tumour microenvironment and tumour progression, but specific genes and mechanisms have not been investigated in hepatocellular carcinoma (HCC). Therefore, this study developed a prognostic model for myeloid cell differentiation in HCC.

Methods: HCC transcriptomic data and myeloid cell differentiation-related genes (MCDRGs) were sourced from public databases. Candidate genes, derived from the intersection of differentially expressed genes (DEGs) and MCDRGs, underwent Mendelian randomization (MR) to identify MR feature genes causally linked to HCC. Regression and machine learning analyses then identified prognostic genes and established a risk model. A nomogram integrating the risk score with independent prognostic factors was subsequently constructed and validated. Furthermore,

prognostic genes were further investigated through functional enrichment, gene mutation, immune escape correlation, and drug sensitivity analyses. Single-cell analysis revealed their expression on key cell clusters. Finally, expression in clinical samples was verified by reverse transcription-quantitative PCR (RT-qPCR).

Results: Seven genes (FECH, CDKN2B, CAMK4, NBEAL2, RASGRP4, GPR171, EIF6) were identified as prognostic and used to construct a risk model, with high-risk group (HRG) patients showing significantly worse survival. A validated nomogram integrated T stage and risk score. Enrichment analysis revealed differential pathway enrichment (e.g., complement and coagulation cascades, fatty acid metabolism) between risk groups. Moreover, prognostic genes were shown to be correlated with immune escape, and CDKN2B had the highest mutation frequency. Single-cell analysis identified T cells as key cells and the expression of GPR171 showed an upward-downward-upward trend. Finally, RT-qPCR confirmed significant upregulation of CDKN2B, EIF6, NBEAL2, and FECH ($P < 0.05$), alongside downregulation of CAMK4 ($P < 0.0001$) in HCC samples.

Conclusions: This study uncovered 7 myeloid differentiation-related prognostic genes in HCC, offering new insights for its research and treatment.

Keywords: Hepatocellular Carcinoma, Myeloid cell differentiation, Prognostic genes, Single-cell RNA sequencing

Background

Hepatocellular carcinoma (HCC) is the most common primary liver cancer, with 906,000 new cases and 830,000 deaths globally in 2020, ranking sixth in incidence and third in cancer mortality, with a rapid growth trend(1). In China, most patients are diagnosed at intermediate to advanced stages, leaving only 15%-30% eligible for

surgery; the majority are limited to non-surgical treatments(2). Traditional therapies for inoperable intermediate and advanced HCC, such as transarterial chemoembolization and chemoradiotherapy, have shown limited efficacy. Although targeted immunotherapy has achieved notable progress in advanced HCC over the past five years, the five-year survival rate for HCC remains below 20%(3). Therefore, investigating novel prognostic genes at the molecular level is crucial for early HCC diagnosis and treatment.

Originating from bone marrow hematopoietic stem cells, myeloid cells comprise several types, such as granulocytes, mononuclear macrophages, and mast cells(4). Myeloid cell differentiation refers to the process by which hematopoietic stem cells (HSCs) differentiate into various myeloid cell lineages in the bone marrow, regulated by cytokines (e.g., GM-CSF, M-CSF, G-CSF), transcription factors (e.g., PU.1, C/EBP family), and microenvironmental signals(5). Within the tumor microenvironment (TME), abnormal differentiation or dysfunctional activity of myeloid cells significantly influences tumor progression. For instance, immature myeloid cells (such as granulocyte/monocyte precursors) expand in response to tumor-derived factors (e.g., IL-6, VEGF), suppressing T-cell and NK-cell functions and thereby promoting immune evasion(6). Tumor-associated macrophages (TAMs) promote angiogenesis and metastasis by secreting immunosuppressive factors like IL-10 and TGF- β (7). Research indicates the pre-metastatic niche recruits bone marrow-derived cells to facilitate cancer cell colonization(8, 9). For example, Stefanescu et al. demonstrated that myeloid cells suppress CD8⁺ T-cell activity via activation of the TGF- β signaling pathway during lung metastasis, thereby exerting an immunosuppressive effect(10). In addition, there exists a multifaceted and close relationship between myeloid cell differentiation-related genes (MCDRGs) and HCC,

primarily reflected in their influence on the tumor microenvironment, immune regulation, disease progression, and prognosis. For example, a signature constructed based on myeloid cell marker genes exhibits an expression pattern that maps the heterogeneity of the HCC immune microenvironment, is closely associated with myeloid cell infiltration, and demonstrates significant prognostic predictive value(11). Mechanistically, specific MCDRGs, such as myeloid differentiation factor 88 (MyD88), have been shown to promote the growth and metastasis of HCC, indicating their direct role in driving HCC progression(12). Recent studies also suggest that the genetic characteristics of HCC cells can reshape the tumor microenvironment, thereby revealing myeloid cell-centric targetable vulnerabilities, which offer novel insights for immunotherapy(13). However, despite these advances, the mechanisms by which bone marrow cell differentiation-related genes function in HCC remain not fully elucidated.

By leveraging genetic variants as instrumental variables, Mendelian randomization (MR) assesses the effect of an exposure on an outcome while reducing vulnerability to confounders such as environment, socioeconomic conditions, and personal behavior(14). Single-cell RNA sequencing (scRNA-seq) enables analysis at single-cell resolution, revealing the distribution and functional states of diverse cell types. This technology enhances data resolution and precision, allowing observation of dynamic cellular changes during disease and facilitating exploration of individual treatment responses(15).

This study integrated public HCC transcriptomic data with myeloid cell differentiation-related genes for MR, regression, and machine learning analyses to identify prognostic genes and construct a model. Further analyses included enrichment, immune infiltration, and drug sensitivity studies. Prognostic gene

expression was also examined at the single-cell level, and key cell populations were characterized. Finally, reverse transcription-quantitative PCR (RT-qPCR) validated prognostic gene expression in clinical samples, offering new insights for HCC diagnosis and targeted therapy.

Methods

Source of data

HCC transcriptomic, somatic mutation, clinical, and survival data were obtained from The Cancer Genome Atlas (TCGA) (<http://cancergenome.nih.gov/>) on October 25, 2024. The training set (TCGA-HCC) comprised 363 tumor samples with survival data and 50 adjacent tissues. Validation data (196 HCC samples with survival data) came from the International Cancer Genome Consortium (ICGC) database (<https://dcc.icgc.org/>) (ICGC-LIRI). Single-cell RNA sequencing data (GSE149614, GPL24676) containing 10 HCC and 8 control liver samples were sourced from Gene Expression Omnibus (GEO) (www.ncbi.nlm.nih.gov/geo/). Additionally, 423 MCDRGs were obtained from a published paper(16)(**Additional file 1**). What's more, for MR analysis, the ieu-b-4953 dataset which contained 372,184 samples (HCC:control =168:372,016) and 6,304,034 single nucleotide polymorphisms (SNPs) derived from Europeans was extracted from Open genome-wide association study (GWAS) database in Integrative Epidemiology Unit (IEU) (<https://gwas.mrcieu.ac.uk/>).

Selection of candidate genes

DEGs between HCC and control samples in TCGA-HCC were identified using "DESeq2" package(17)(v 1.34.0) (v 1.34.0) ($p < 0.05$, $|\log_2FC| > 0.5$). A volcano plot was generated with "ggplot2" package(18) (v 3.4.1) and "ggrepel" (19) (v 0.9.5), labeling the top 10 up- and down-regulated DEGs. Candidate genes were obtained by

intersecting DEGs with MCDRGs using "VennDiagram" package(20) (v 1.7.1).

Functional analyses and protein-protein interaction (PPI) network of candidate genes

To explore biological functions and pathways of candidate genes, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed using the(21) (v 4.2.2) ($p < 0.05$). GO enrichment included biological processes (BPs), cellular components (CCs), and molecular functions (MFs). The top 5 GO terms in each category and the top 10 KEGG pathways (by ascending p-value) were displayed. On the other hand, with the aim of exploring candidate genes' interactions at protein level, they were uploaded to Search Tool for the Retrieval of Interaction Gene/Proteins (STRING) website (<http://string-db.org/>) with the parameter of interaction score > 0.4 . After removing isolated proteins, "Cytoscape"(22) (v 3.9.1) was applied for the PPI network development.

Two-sample MR analysis

In order to identify genes which had causal relation with HCC from candidate genes, two sample MR analysis was carried out, in which HCC was regarded as the outcome (ieu-b-4953 dataset). Candidate genes were uploaded to Ensembl database (<https://mart.ensembl.org/>) to match their Ensembl ID, and their expression quantitative trait loci (eQTL) data was extracted from the OpenGWAS database for exposure ID acquirement, which were considered as exposure factors.

MR analysis is based on three core premises. IVs (SNPs) were extracted using "TwoSampleMR" (23) (v 0.64) ($p < 5 \times 10^{-8}$), with LD clumping ($R^2 < 0.001$, kb = 10) and outcome-associated IVs removed via PhenoScanner. IV strength was assessed using the F-statistic ($F = R^2 (N - k - 1) / [k (1 - R^2)]$), with $F > 10$ indicating sufficient strength. Data were harmonized, and five MR methods were integrated: inverse

variance weighted (IVW)(24), MR-egger(25), weighted median(26), simple mode(23), and weighted mode(27), with IVW ($p < 0.05$) as primary reference ($OR > 1$ = risk factor; $OR < 1$ = protective factor). Subsequently, results were visualized via scatter plots from "ggplot2" package (v 3.4.1) to demonstrate the correlation of exposure factors as well as outcome. Forest plots derived from "forestplot"(28) (v 3.1.1) were applied to determine the diagnostic efficacy of SNP site for each exposure factor for outcomes, and funnel plots were constructed to assess whether SNPs were randomly distributed.

MR sensitivity analyses

Sensitivity analyses (heterogeneity, pleiotropy, leave-one-out, Steiger) validated MR results. Heterogeneity was assessed via `mr_heterogeneity` (IVW) in "TwoSampleMR" ($p > 0.05$: fixed effects; $p < 0.05$: multiplicative random effects). Pleiotropy was tested; genes with $p < 0.05$ were excluded. Thirdly, in order to test whether each IV caused significant changes to the outcome, the `mr_leaveoneout` function was applied to perform LOO sensitivity test, if the exclusion of a SNP results in a large change in the results, it indicates that the outcome is sensitive to the IV and the result is unreliable. Finally, To confirm the causality direction between exposure and outcome, the Steiger test was performed using the "directionality_test" function. A result of $p < 0.05$ and "correct_causal_direction=TRUE" indicated the MR direction was correct. Thus, candidate genes which selected through MR analysis were recorded as MR feature genes and exported to downstream analyses.

Identification of prognostic genes and risk model construction

Prognostic MR feature genes were identified via PH assumption test ($p > 0.05$), univariate Cox regression ($HR \neq 1$, $p < 0.1$) using "survival" (29) (v 3.5-3) ($HR \neq 1$, $p < 0.1$), and visualized with "forestPlot" (v 3.1.1). Least absolute shrinkage and

selection operator (LASSO) regression with "glmnet"(30) (v 4.1-4) and 10-fold CV selected genes at log(lambda.min). These underwent PH test ($p > 0.05$), multivariate Cox, and stepwise regression ($HR \neq 1$, $p < 0.2$) in "survival". Final genes were visualized with "forestPlot". Expression differences (HCC vs. control) in TCGA-HCC were analyzed with "DESeq2" (v 1.34.0) and "ggplot2" (v 3.4.1) ($p < 0.05$).

Risk model development and verification

According to the coefficients of prognostic genes acquired from stepwise regression, a risk model could be developed according to the formula:

$$\text{Riskscore} = \sum_{i=1}^n \text{coef}(\text{gene}_i) * \text{expr}(\text{gene}_i)$$

Where "coef" denoted the risk coefficients for each prognostic gene derived from stepwise regression, "expr" represented the expression of them.

Risk scores were calculated for each tumor sample in TCGA-HCC and ICGC-LIRI using the risk formula. Patients were stratified into high-risk (HRG) and low-risk (LRG) groups based on optimal cutoffs. Model validation included risk curves and survival diagrams generated with "survminer"(29) (v 0.4.9). KM curves assessed survival differences between HRG and LRG (log-rank test, $p < 0.05$). Prediction reliability was evaluated using 2-, 3-, and 5-year ROC curves with "survivalROC" package(31) (v 1.0.3); $AUC > 0.6$ indicated accurate prediction.

Independent prognostic analysis and nomogram development

Independent prognostic factors in TCGA-HCC were identified by applying PH test ($p > 0.05$), univariate and multivariate Cox regression ($p < 0.05$, $HR \neq 1$) to clinical traits and risk score using "survival" (v 3.5-3). Results were visualized with "forestPlot" (v 3.1.1). Significant factors were used to build a nomogram for 2-, 3-, and 5-year OS via "rms" (v 6.5-1)(29). Nomogram performance was assessed via calibration curves

(slope near 1) and time-dependent ROC curves ($AUC > 0.6$) using "rms" and "survivalROC" (v 1.0.3).

Gene set enrichment analysis (GSEA)

GSEA on TCGA-HCC explored biological functions between risk groups. Differential expression (HRG vs. LRG) used "DESeq2" (v 1.34.0) with reference set "c2.cp.kegg.v7.5.1.symbols.gmt" from MSigDB. Resulting DEGs were ranked by descending log2FC. GSEA was then carried out with the "clusterProfiler" package (v 4.2.2) ($\text{adj.p} < 0.05$, false discovery rate (FDR) < 0.25 , $|\text{normalized enrichment score (NES)}| > 1$), and the top 10 enriched pathways were displayed.

Prognostic gene mutation analysis and immune escape analysis

To assess the impact of mutations on HCC, genomic alterations of prognostic genes were analyzed in the cBioPortal database (<https://www.cbioportal.org/>). TCGA-HCC samples were divided into mutant and unmutant groups based on these mutations. Survival differences were evaluated using KM curves generated with the "survminer" package (v 0.4.9) and the log-rank test ($p < 0.05$). On the other hand, immune escape related genes made tumor cells resistant to immune system and thus promoted the proliferation of cancer cells. Thus, exploring the association between prognostic genes and immune escape genes can reveal whether prognostic genes are involved in immune escape. Here, we employed 2 immune checkpoints (ICs) CTLA4(32) and PDCD1(33, 34), along with gene TP53(35, 36), exploring the correlations between these 3 genes and prognostic genes. The single sample GSEA (ssGSEA) scores of prognostic genes were computed in TCGA-HCC by "GSVA" package(37) (v 1.46.0), then "psych" package(38) (v 2.1.6) was implemented to analyze the correlations between ssGSEA scores of prognostic genes and 3 immune escape related genes, $p < 0.05$ and $|\text{correlation (cor)}| > 0.3$ indicated significant correlation.

Drug sensitivity analysis

Drug sensitivity between HRG and LRG in TCGA-HCC was assessed using IC₅₀ values. IC₅₀s for 198 drugs from the Genomics of GDSC (<https://www.cancerrxgene.org/>) were calculated with the "oncoPredict" package(39) (v 1.2). Correlations between drug IC₅₀ and risk score were analyzed using the "psych" package (v 2.1.6) ($p < 0.05$) to identify sensitive drugs. Wilcoxon tests compared IC₅₀ differences between groups ($p < 0.05$), and the top 5 drugs were visualized with "ggplot2" (v 3.4.1).

Single-cell data processing and cell annotation

Single-cell analysis of the GSE149614 dataset was performed using the "Seurat" package(40) (v 5.0.1). Quality control (QC) thresholds were set at $200 \leq \text{nFeature-RNA} \leq 2,500$ and $\text{percent.mt} < 5\%$. Post-QC metrics were visualized with violin plots. Data were normalized, and the top 2,000 highly variable genes (HVGs) were identified (FindVariableFeatures, vst method) and displayed (LabelPoints). Scaling (ScaleData) and principal component analysis (PCA) (RunPCA) were applied to HVGs, with PCs filtered via JackStrawPlot and ElbowPlot ($p < 0.05$). Selected PCs were used for unsupervised clustering (FindNeighbors, FindClusters) and uniform manifold approximation and projection (UMAP) (RunUMAP, resolution = 0.1) to identify cell clusters.

To annotate cell clusters, the "singleR" (41) (v 2.0.0) was applied (min.pct=0.2, only.pos=TRUE) using three reference datasets (HumanPrimaryCellAtlasData, BlueprintEncodeData, ImmuneCellExpressionData) and marker genes from the CellMarker database (<http://xteam.xbio.top/CellMarker/>). Results were visualized with dot plots and UMAP.

Selection of differential cells, functional and cellular communication of

differential cells

To identify differential cells between control and HCC samples, the Benjamini-Hochberg (BH) test identified cell clusters with significantly different abundance ($p < 0.05$). Gene set variation analysis (GSVA) was then performed on these populations using the "ReactomeGSA" (42) (v 1.16.1), with the top 10 pathways by ssGSEA score displayed. Cellular communication in GSE149614 was analyzed using "CellChat" (43) (v 1.6.1), examining communication number, strength, and ligand-receptor interactions via the netVisual_bubble function.

Identification of key cells and pseudotime analysis

To identify key cells, a Wilcoxon test compared prognostic gene expression in differential cells between HCC and control samples ($p < 0.05$). Cells with distinct expression were designated key cells. Their differentiation process was explored by subjecting them to standardization, normalization, dimensionality reduction, clustering, and annotation, following the steps in section 2.11, to identify subgroups. Then subgroups in key cells were exported to pseudotime analysis implementing "Monocle2" package(44) (v 2.26.0), cell differentiation trajectories were analyzed through DDRTree function. Meanwhile, the plot_genes_in_pseudotime function visualized expression trends of prognostic genes during cell differentiation.

Immune histochemistry (IHC) staining and experimental verification

To validate prognostic gene expression in HCC, IHC staining images for their corresponding proteins were obtained from the Human Protein Atlas (HPA, www.proteinatlas.org). Expression levels between control and HCC tissues were assessed based on staining intensity.

Following ethics approval, 5 paired HCC and paracancerous samples from Shenzhen Hospital were used for RT-qPCR validation. RNA was extracted with TRIzol

(Ambion), quantified (NanoPhotometer N50), and reverse-transcribed (Hifair® III or SureScript kits) on a S1000TM Thermal Cycler. qPCR on a CFX Connect system (95°C 60s; 40 cycles of 95°C 20s, 55°C 20s, 72°C 30s) used the $2^{-\Delta\Delta CT}$ method. Data were analyzed with GraphPad Prism 10 (v 5.0)(45) ($P < 0.05$). Primer sequences are in **Additional file 2**.

Statistical analysis

Bioinformatics analyses were performed with R software (v 4.2.2). Group comparisons used the Wilcoxon test with BH correction; RT-qPCR results were analyzed by t-test. Significance was set at $p < 0.05$.

Results

Functional enrichment results and PPI network of 173 candidate genes

In differential expression analysis, 8,182 DEGs were identified in TCGA-HCC with 5,657 up-regulated (such as CPLX2, PGC and MAGEA1) and 2,525 down regulated DEGs (such as STAB2, BMPER and CRHBP) (**Figure 1a**), and 173 candidate genes were caught by overlapping 423 MCDRGs with 8,182 DEGs (**Figure 1b**). GO and KEGG enrichment analyses of the 173 candidate genes yielded 1,707 GO terms, comprising 1,639 BP, 24 CC, and 44 MF (**Additional file 3**). Among the enriched GO terms, BP included myeloid cell differentiation, CC included the RNA polymerase II transcription regulator complex, and MF included DNA-binding transcription factor binding (**Figure 1c**), suggesting roles in myeloid differentiation and gene expression regulation. KEGG analysis further enriched 122 pathways (**Additional file 4**), such as osteoclast differentiation, hepatitis B, and cytokine-cytokine receptor interaction (**Figure 1d**). Moreover, the PPI network demonstrated that 158 candidate genes exhibited interactions at the protein level (**Figure S1**), in which TNF, STAT3 and MYC showed the strongest interactions with other genes.

A total of 40 MR feature genes were obtained

Next, the 173 candidate genes were analyzed by two-sample MR, yielding 40 genes with significant IVW results ($p < 0.05$), including GATA3, HSF1, and FOS. Among these, 16 were identified as risk factors ($OR > 1$) and 24 as protective factors ($OR < 1$) (**Additional file 5**). The forest plots of each gene indicated that the diagnostic efficacy of SNPs and exposure factors (genes) for resolution was consistent (**Figure S2**). Scatter plots confirmed MR results: an IVW slope > 0 indicated a risk gene, while < 0 indicated a protective gene. Intercepts near 0 suggested minimal confounding (**Figure S3**). Funnel plots showed random SNP distribution, aligning with Mendel's second law (**Figure S4**). Heterogeneity testing further validated the MR results: 35 genes passed (Q test $p > 0.05$), analyzed with a fixed effects model, while 5 genes (PARP1, CCR1, FECH, STAT3, ZNF675) showed heterogeneity and required a multiplicative random effects model (**Additional file 5, Additional file 6**). Then the pleiotropy test results indicated that no horizontal pleiotropy was existed in these genes (**Additional file 7**), suggesting the MR results of these 40 genes were reliable. Furthermore, the LOO test results of these genes indicated that the outcome was not sensitive to each IV (**Figure S5**). Finally, when the correct_causal_direction set to TRUE in the steiger test, p values of all 40 genes were < 0.05 , indicating a valid unidirectional causal association between these genes and HCC (**Additional file 8**). Thus, a total of 40 MR feature genes were obtained.

A risk model was established and verified with 7 prognostic genes

Firstly, conducting PH assumption test, 30 genes such as UFL1, GAB2 and FECH were identified ($p > 0.05$) (**Additional file 9**). Subsequently, 9 genes were selected by the univariate Cox regression ($p < 0.1$), in which CAMK4 and GPR171 were protective genes ($HR < 1$), HSPA1B, NBEAL2, EIF6, FECH, CDKN2B, ZNF675 and

RASGRP4 were risk genes ($HR > 1$) (**Figure 2a**). Moreover, in the LASSO regression analysis, all 9 genes passed the LASSO regression when $\log(\lambda_{\min}) = -4.978998$ (**Figure 2b**), leading to the utilization of multivariate Cox and stepwise regression model. Finally, according to the multivariate Cox and stepwise regression model, FECH, CDKN2B, CAMK4, NBEAL2, RASGRP4, GPR171 and EIF6 were identified as prognostic genes ($p < 0.2$), in which CAMK4 and GPR171 were protective genes ($HR < 1$) while other 5 genes were risk genes ($HR > 1$) (**Figure 2c**). Therefore, FECH, CDKN2B, CAMK4, NBEAL2, RASGRP4, GPR171 and EIF6 were utilized to construct a risk model with coefficients obtained from stepwise regression

$$\text{Risk score} = \text{FECH} \times (0.295) + \text{CDKN2B} \times (0.276) + \text{CAMK4} \times (-1.811) + \text{NBEAL2} \times (0.397) + \text{RASGRP4} \times (1.230) + \text{GPR171} \times (-0.477) + \text{EIF6} \times (0.308).$$

Risk scores for each HCC patient were calculated in TCGA-HCC and ICGC-LIRI. Using optimal cutoffs (0.2231887 in TCGA-HCC; 34.66562 in ICGC-JP), patients were stratified into HRG and LRG (**Figure 2d**). Higher risk scores correlated with increased mortality (**Figure 2e**). KM curves showed significantly lower survival in HRG across both datasets ($p < 0.05$) (**Figure 2f**). Predictive reliability was confirmed by 2-, 3-, and 5-year ROC curves ($AUC > 0.6$) in both cohorts (**Figure 2g**). In TCGA-HCC, CDKN2B, EIF6, FECH, GPR171, and NBEAL2 were upregulated in HCC, while CAMK4 and RASGRP4 were downregulated ($p < 0.05$) (**Figure 2h**).

A nomogram was developed and verified with T stage and risk score

Based on PH assumption test results, 6 clinical features and risk score (all $p > 0.05$) (**Additional file 10**) were included in downstream analyses. Univariate Cox analysis identified stage, T stage, M stage, and risk score as significant ($p < 0.05$, $HR \neq 1$) (**Figure 3a**). Multivariate analysis confirmed T stage and risk score as independent

prognostic factors ($p < 0.05$, $HR \neq 1$) (**Figure 3b**), which were integrated into a nomogram (**Figure 3c**). Higher total points correlated with lower OS probability. Subsequently, the calibration curve demonstrated the slopes of 2-, 3-, 5-year were close to 1, confirming the nomogram's excellent predictive accuracy (**Figure 3d**). Additionally, the nomogram showed excellent predictive capability ($AUC > 0.6$ at 2, 3, and 5 years) (**Figure 3e**), supporting its utility for HCC risk assessment.

Differential pathways and drugs between HRG and LRG were uncovered

After conducting GSEA between HRG and LRG in TCGA-HCC, 38 significantly enriched pathways were identified (**Additional file 11**), the top 10 pathways according to |enrichment score| included complement and coagulation cascades, fatty acid metabolism, primary bile acid biosynthesis and so on (**Figure 4a**), all of them showed down-regulated trends (HRG vs. LRG). Meanwhile, after assessing the IC_{50} values of 198 common chemotherapy drugs, HRG patients were more sensitive to 71 drugs such as afatinib and GDC0810, while 33 drugs were more sensitive to LRG patients such as IAP and doramapimod (**Figure 4b**). Furthermore, the top 5 drugs with remarkably different IC_{50} values were MK.1775, ML323, SB505124, ulixertinib, and Wee1.inhibitor, all showed higher IC_{50} values in LRG except for SB505124 (**Figure 4c**).

Mutation status and immune escape correlations of prognostic genes were revealed

In mutation analysis, 7 prognostic genes mainly included missense mutation, amplification and deep deletion (**Figure 5a**), gene CDKN2B showed the highest mutation frequency (3%) with deep deletion, while CAMK4, NBEAL2, RASGRP4, GPR171 and EIF6 mainly comprised amplification. According to the mutation status, patients with mutation data in TCGA-HCC were sorted into mutant ($n=34$) and

wild-type groups (n=313), and patients which classified into mutant group had obviously lower survival rate ($p < 0.05$) (**Figure 5b**). Besides, we also analyzed the correlations between prognostic genes' ssGSEA scores with CTLA4, PDCD1 and TP53, and it was found that prognostic genes were positively correlated with CTLA4 and PDCD1 ($\text{cor} > 0.3$, $p < 0.05$) (**Figure 5c**), indicating that prognostic genes might be related to tumour immune escape.

Single-cell data procession and identification of 7 cell types

In single-cell analysis, 38,920 cells and 25,712 genes were selected after rigorous QC (**Figure S6**). Next, top 2,000 HVGs were identified such as IGHA1, IGHG2 and IGLC3 (**Figure 6a**). Based on JackStraw and Elbow plots (decreased significance after PC30, plateau at PC30) (**Figure 6b, c**), the first 30 PCs were selected. UMAP clustering classified cells into 21 distinct clusters (**Figure 6d**). Using marker genes and "singleR" (v 2.0.0), clusters were annotated as T cells, NK cells, myeloid cells, Endo, B cells, Hepato, and Fibro. Unidentified clusters were labeled as undefined (**Figure 6e, f**).

Differential cell types along with their functional difference and cellular communication were uncovered

After cell type annotation, the abundance of all seven cell types differed significantly between HCC and control samples ($p < 0.05$). T cells, NK cells, endothelial cells (Endo), and B cells were more abundant in controls (**Figure 7a**). GSVA of differential cells revealed enriched pathways, including sterol 12-hydroxylation by CYP8B1, NEIL3-mediated ICL resolution, and serotonin metabolism (**Figure 7b**), highlighting their diverse functions. According to the cellular communication network, T cells showed the highest number as well as strongest interactions in control group (**Figure 7c,d**), while myeloid cells and Hepato increased their interaction numbers and

strength in HCC group (**Figure 7e,f**). Furthermore, in both HCC and control groups, the ligand-receptor pair MIF-(CD74+CXCR4) showed the highest expression level (**Figure 7g,h**), indicating that cells in HCC microenvironment mainly interacted via the MIF-(CD74+CXCR4).

T cells were identified as key cells and results of pseudotime analysis

Among the seven differential cell types, prognostic gene expression was examined. In T cells, all prognostic genes except RASGRP4 and EIF6 showed differential expression between HCC and control groups (**Figure S7**). T cells, where prognostic genes were detectable in both groups, were selected as key cells. Dimensionality reduction clustering and annotation divided T cells into 10 clusters, identifying CD8⁺ and CD4⁺ T cells (**Figure 8a,b**). After that, Pseudotime analysis mapped the T cell differentiation trajectory, with CD4⁺ T cells at the initial stage and CD8⁺ T cells at the terminal stage (**Figure 8c,d**). Most prognostic genes showed low, stable expression, except GPR171, which exhibited an upward-downward-upward trend during differentiation (**Figure 8e**).

IHC staining and RT-qPCR results

IHC staining images between HCC and control tissues (**Figure 9a-d**) showed higher staining for FECH, CAMK4, EIF6, and CDKN2B in HCC samples. No staining images were available for NBEAL2, RASGRP4, or GPR171.

RT-qPCR validation in clinical samples confirmed significant upregulation of CDKN2B ($P = 0.0244$), EIF6 ($P = 0.0012$), NBEAL2 ($P = 0.0165$), and FECH ($P = 0.0036$) in HCC versus controls, alongside downregulation of CAMK4 ($P < 0.0001$) (**Figure 9e-i**). These findings underscore their relevance in HCC and potential as prognostic biomarkers.

Discussion

As the predominant form of primary liver cancer, HCC exhibits high malignancy and is associated with an unfavorable prognosis(46). Research has demonstrated that MCDRGs play a key role in remodeling the immune landscape of HCC and contribute to its resistance to therapy. For instance, myeloid cells—particularly MDSCs—can enhance the local immunosuppressive state within tumors by secreting IL-8 and activating the CXCR2 signaling pathway, thereby promoting resistance to immunotherapy in HCC(47). Seven prognostic genes—FECH, CDKN2B, CAMK4, NBEAL2, RASGRP4, GPR171, and EIF6—were identified in this study through bioinformatics analyses of HCC transcriptome data from TCGA and GEO, integrated with myeloid cell differentiation-related genes. Based on these, a prognostic risk model was constructed. Mechanistic exploration linked these genes to tumor immune evasion, with CDKN2B showing the highest mutation frequency. Single-cell analysis identified T cells as the key population driving their differential expression, and GPR171 expression varied dynamically during T-cell differentiation. Finally, RT-qPCR validation confirmed significant upregulation of CDKN2B, EIF6, NBEAL2, and FECH ($P < 0.05$) and downregulation of CAMK4 ($P < 0.0001$) in HCC tissues, consistent with bioinformatic predictions. This study elucidates the mechanisms and prognostic value of myeloid differentiation-related genes in HCC, offering new directions for targeted therapy and diagnosis.

Eukaryotic translation initiation factor 6 (EIF6) primarily binds to the 60S ribosomal subunit, preventing its association with the 40S subunit and thus inhibiting formation of the cytoplasmic 80S initiation complex(48). In HCC-related studies, this gene has been shown to promote malignant progression predominantly through the mTOR signaling pathway(49). Both in vitro and in vivo experiments confirm that suppressing

EIF6 expression reduces HCC growth(50). Notably, liver-specific overexpression of EIF6 in transgenic mouse models was sufficient to spontaneously induce HCC, constituting one of the most robust pieces of evidence supporting its oncogenic properties(51). RT-qPCR validation in clinical samples confirmed significant EIF6 upregulation in HCC tissues ($P < 0.05$), supporting its clinical relevance as a prognostic gene. Calcium/calmodulin-dependent protein kinase IV (CAMK4) primarily functions within the CaMKK–CaMK4 signaling cascade. In neurons and immune cells, it regulates gene expression by phosphorylating transcription factors such as CREB, thereby influencing cell proliferation, differentiation, and survival(52). Previous studies have confirmed that CAMK4 functions as a promoter of HCC progression and can be directly targeted and suppressed by the tumor-suppressive microRNA miR-129-5p, thereby inhibiting HCC growth both in vitro and in vivo. Consistent with the present findings, miR-129-5p is expressed at low levels in HCC tissues and cell lines(53). RT-qPCR validation in this study revealed significant downregulation of CAMK4 in HCC tissues ($P < 0.0001$), suggesting a potential tumor suppressor role in HCC. Despite this, the tumor type specificity of CAMK4 functions presents challenges for its development as a therapeutic target.

This study confirms that cyclin-dependent kinase inhibitor 2B (CDKN2B), a key cell cycle regulator, is significantly upregulated in HCC tissues. This gene serves as a critical determinant of HCC sensitivity to sorafenib(54). Its antisense lncRNA, CDKN2B-AS1, promotes HCC progression via the E2F1/GNAZ axis and is associated with disease risk and clinical progression(55, 56). This evidence suggests CDKN2B and its non-coding RNA network are important in HCC malignancy.

Neurobeachin-like protein 2 (NBEAL2), a key regulator of cellular secretion and vesicular trafficking, was significantly upregulated in HCC tissues, suggesting its

potential pro-tumorigenic role(57). Concurrently, NBEAL2 is crucial for the function of innate immune cells such as neutrophils and natural killer (NK) cells(58), which are central players in myeloid differentiation and tumor immune surveillance(59). By modulating protein secretion and cytotoxic functions in these cells, NBEAL2 directly influences the intensity of the host's anti-tumor immune response.

Ferrochelatase (FECH) catalyzes the insertion of ferrous iron into protoporphyrin IX, serving as the terminal and rate-limiting enzyme in heme biosynthesis(60). Relevant literature indicates that this gene is highly expressed in senescent cells and may be associated with the anti-apoptotic properties of tumor cells(61). Inhibition of FECH has been demonstrated as an effective strategy for inducing ferroptosis in cancer cells. For example, the flavonoid 4,4'-dimethoxychalcone (DMC) exerts pro-ferroptotic and anti-tumor effects by inhibiting FECH(62). Therefore, targeting the expression of this gene may represent a promising therapeutic strategy for HCC.

GSEA revealed significant enrichment of the complement and coagulation cascades, fatty acid metabolism, and primary bile acid biosynthesis pathways, all of which were consistently down-regulated in both risk groups. Physiologically, the complement system contributes to anti-tumor immune surveillance as part of innate immunity. However, during progression from chronic liver diseases to HCC, persistent aberrant activation in the inflammatory tumor microenvironment promotes tumor growth, immune evasion, and metastasis(63). Moreover, the complement system, particularly the C5a/C5aR1 axis, is a key regulator shaping the immunosuppressive microenvironment and influencing immune checkpoint inhibitor efficacy(64). Additionally, recent studies have further confirmed that complement-related gene signatures can effectively stratify HCC patients into prognostic subtypes, with high-risk subtypes closely linked to an immunosuppressive microenvironment rich in

myeloid-derived suppressor cells(65). This suggests aberrant complement system activation is a critical driver of the immunosuppressive landscape in HCC.

The fatty acid metabolic pathway encompasses fatty acid uptake, de novo synthesis, and conversion into complex lipids, with the liver serving as the central organ for this metabolism(66). In HCC, ACSL4 reprograms metabolism via the c-Myc/SREBP1 pathway, promoting polyunsaturated fatty acid synthesis and driving tumor malignant progression(67). Concurrently, fatty acids can stabilize HIF-1 α through FABP5, establishing a positive feedback loop that upregulates the expression of fatty acid synthesis genes and enhances cell proliferation(68). This intrinsic metabolic reprogramming within tumor cells may alter their lipid secretion profile. Consequently, we hypothesize that the resulting specific lipid signals (potentially mediated via the FABP5/HIF-1 α axis) might partially shape the aberrant differentiation and function of myeloid cells within the hepatocellular carcinoma microenvironment.

Our single-cell analysis identified T cells, primarily CD4⁺ and CD8⁺ subsets, as the key population most correlated with the expression of MCDRGs. T cells play a dual role in HCC: CD8⁺ T cells possess tumor-killing capabilities, while tumor-infiltrating T cells often exhibit an exhausted phenotype and contribute to immune evasion(69). There was regulatory crosstalk between T cells and myeloid cells in terms of their development and function(70). We observed differential expression of MCDRGs in T cells, with dynamic changes along T cell differentiation trajectories, suggesting that these genes might indirectly influence the HCC immune microenvironment by modulating T cell states. Myeloid cells (e.g., tumor-associated macrophages) can directly regulate T cell function by secreting cytokines or expressing immune checkpoint molecules(71), and some MCDRGs may be involved in this process.

Therefore, T cells, acting as both responders to and executors of "myeloid-related signals," may serve as a critical bridge linking myeloid differentiation status with anti-tumor immunity.

This study validated prognostic gene expression in clinical samples by integrating IHC and PCR results from specimens with database expression data. Comparative IHC staining images of prognostic genes in HCC versus control tissues showed stronger staining intensity for FECH, CAMK4, EIF6, and CDKN2B in HCC samples, while NBEAL2, RASGRP4, and GPR171 exhibited no detectable staining. RT-qPCR revealed significantly elevated mRNA levels of CDKN2B, EIF6, NBEAL2, and FECH in HCC tissues, while CAMK4 expression was markedly reduced. Overall, these findings are largely consistent with previous reports. Furthermore, we observed discordance between protein-level and gene-level expression for CAMK4 and NBEAL2, a phenomenon that was common yet critical in cancer research, often reflecting deeper biological regulatory mechanisms. The discrepancy between mRNA and protein expression of CAMK4 likely stems from multilayered post-transcriptional regulation between mRNA translation and protein synthesis/degradation, as well as cellular heterogeneity within tumor tissues(72). For example, if CAMK4 is highly expressed in tumor-infiltrating immune cells (e.g., T cells), RT-qPCR may detect an overall increase in mRNA levels, while IHC could reveal predominant localization of the protein in stromal lymphocytes rather than in HCC cells(73). This discrepancy suggests that CAMK4 may exhibit cell-type-specific functions within the HCC microenvironment, linking tumor malignancy to immune regulation, which warrants further investigation.

This study integrated multi-omics data to construct and validate an HCC prognostic model based on myeloid cell differentiation-related genes. Seven core genes were

associated with patient prognosis, immune microenvironment, and metabolic reprogramming, with T cells identified as the key cellular population for their expression. However, several limitations remained in this work. First, the conclusions are derived from retrospective analyses of public databases, with a limited scale of clinical sample validation. Second, the GWAS data for MR analysis originated primarily from European populations, necessitating further validation of the findings' generalizability to other ethnic groups. Additionally, the single-cell analysis was based on a single dataset, and the specific mechanisms through which these genes influence immune cell function awaited experimental confirmation. Future research should validate this model in larger prospective cohorts and elucidate key gene mechanisms through functional experiments.

Conclusions

In summary, this study constructed and validated a novel prognostic model for HCC based on seven myeloid differentiation-related genes (FECH, CDKN2B, CAMK4, NBEAL2, RASGRP4, GPR171, EIF6). The model effectively stratified patients and was associated with the immune microenvironment, metabolic reprogramming, and patient survival. T cells were identified as the key cellular hub linking gene expression to the immunosuppressive landscape. Experimental validation confirmed dysregulated expression of core genes in clinical samples. Our findings reveal the prognostic and immunological significance of myeloid-related genes in HCC, offering insights for risk stratification and potential immunotherapeutic strategies.

List of abbreviations

HCC hepatocellular carcinoma

MCDRGs myeloid cell differentiation related genes;

DEGs differentially expressed genes

MR Mendelian randomization

RT-qPCR reverse transcription-quantitative PCR

TCGA The Cancer Genome Atlas

ICGC International Cancer Genome Consortium

GEO Gene Expression Omnibus

HRG high-risk group

HSCs hematopoietic stem cells

TME tumor microenvironment

BP biological process

CC cellular components

MF molecular function

GO Gene Ontology

KEGG Kyoto Encyclopedia of Genes and Genomes

TAMs Tumor-associated macrophages

MyD88 myeloid differentiation factor 88

SNPs single nucleotide polymorphisms

GWAS genome-wide association study

eQTL expression quantitative trait loci

IVs instrumental variables

LD linkage disequilibrium

IVW inverse variance weighted

OR odds ratio

LOO leave-one-out

PH proportional hazards

LASSO least absolute shrinkage and selection operator

KM Kaplan Meier

ROC receiver operating characteristic

GSEA Gene set enrichment analysis

FDR false discovery rate

NES normalized enrichment score

GDSC Genomics of Drug Sensitivity in Cancer

QC quality control

HVGs highly variable genes.

Declarations

Ethics approval and consent to participate

The study process was conducted in accordance with the Declaration of Helsinki and approved by the ethics committee of National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, and all participants have signed informed consent.

Consent for publication

Not applicable.

Availability of data and materials

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

Competing Interests

The authors declare no conflict of interest.

Funding

This research supported by Shenzhen High-level Hospital Construction Fund and Research project of National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital & Shenzhen Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College (E010322004; SZ2020MS011).

Authors' Contributions

Z.L. (Zhenyu Lin), L.X. (Lin Xu) and Z.H.(Zhangkan Huang) conceived and designed the study. Z.L. (Zhenyu Lin), X.H.(Xiaozhun Huang), Y.R.(Yihong Ran), ZL(Zirong Liu) collected the data. Z.L. (Zhenyu Lin), W.Z. (Wei Zhang) and Z.H.(Zhangkan Huang) performed the analyses. Z.L. (Zhenyu Lin) and Z.H.(Zhangkan Huang) wrote the manuscript. L.X. (Lin Xu) and Z.H.(Zhangkan Huang) supervised the study. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

Not applicable.

Authors' information

Correspondence: Zhangkan Huang:charles846@126.com.

Zhenyu Lin and Lin Xu have contributed equally to this work.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, 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(3):209-49.
2. Foglia B, Turato C, Cannito S. Hepatocellular Carcinoma: Latest Research in

- Pathogenesis, Detection and Treatment. *Int J Mol Sci.* 2023;24(15).
3. Gabbia D, De Martin S. Insights into Hepatocellular Carcinoma: From Pathophysiology to Novel Therapies. *Int J Mol Sci.* 2024;25(8).
 4. Gao X, Xu C, Asada N, Frenette PS. The hematopoietic stem cell niche: from embryo to adult. *Development.* 2018;145(2).
 5. Kusmartsev S, Gabrilovich DI. Role of immature myeloid cells in mechanisms of immune evasion in cancer. *Cancer Immunol Immunother.* 2006;55(3):237-45.
 6. Weiskopf K, Schnorr PJ, Pang WW, Chao MP, Chhabra A, Seita J, et al. Myeloid Cell Origins, Differentiation, and Clinical Implications. *Microbiol Spectr.* 2016;4(5).
 7. Greene JT, Brian BFt, Senevirathne SE, Freedman TS. Regulation of myeloid-cell activation. *Curr Opin Immunol.* 2021;73:34-42.
 8. Gabrilovich DI, Ostrand-Rosenberg S, Bronte V. Coordinated regulation of myeloid cells by tumours. *Nat Rev Immunol.* 2012;12(4):253-68.
 9. Hirai H, Fujishita T, Kurimoto K, Miyachi H, Kitano S, Inamoto S, et al. CCR1-mediated accumulation of myeloid cells in the liver microenvironment promoting mouse colon cancer metastasis. *Clin Exp Metastasis.* 2014;31(8):977-89.
 10. Stefanescu C, Van Gogh M, Roblek M, Heikenwalder M, Borsig L. TGF β Signaling in Myeloid Cells Promotes Lung and Liver Metastasis Through Different Mechanisms. *Front Oncol.* 2021;11:765151.
 11. Zheng SS, Wu YF, Zhang BH, Huang C, Xue TC. A novel myeloid cell marker genes related signature can indicate immune infiltration and predict prognosis of hepatocellular carcinoma: Integrated analysis of bulk and single-cell RNA sequencing. *Front Mol Biosci.* 2023;10:1118377.
 12. Liang B, Chen R, Wang T, Cao L, Liu Y, Yin F, et al. Myeloid differentiation factor 88 promotes growth and metastasis of human hepatocellular carcinoma. *Clin*

Cancer Res. 2013;19(11):2905-16.

13. Ramirez CFA, Taranto D, Ando-Kuri M, de Groot MHP, Tsouri E, Huang Z, et al. Cancer cell genetics shaping of the tumor microenvironment reveals myeloid cell-centric exploitable vulnerabilities in hepatocellular carcinoma. *Nat Commun.* 2024;15(1):2581.

14. Fang Z, Song M, Lee DH, Giovannucci EL. The Role of Mendelian Randomization Studies in Deciphering the Effect of Obesity on Cancer. *J Natl Cancer Inst.* 2022;114(3):361-71.

15. Zhang J, Li Z, Chen Z, Shi W, Xu Y, Huang Z, et al. Comprehensive analysis of macrophage-related genes in prostate cancer by integrated analysis of single-cell and bulk RNA sequencing. *Aging (Albany NY).* 2024;16(8):6809-38.

16. Wu D, Liu Y, Liu J, Ma L, Tong X. Myeloid cell differentiation-related gene signature for predicting clinical outcome, immune microenvironment, and treatment response in lung adenocarcinoma. *Sci Rep.* 2024;14(1):17460.

17. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.

18. Gustavsson EK, Zhang D, Reynolds RH, Garcia-Ruiz S, Ryten M. ggtranscript: an R package for the visualization and interpretation of transcript isoforms using ggplot2. *Bioinformatics.* 2022;38(15):3844-6.

19. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics.* 2011;12:35.

20. Coleman A, Bose A, Mitra S. Metagenomics Data Visualization Using R. *Methods Mol Biol.* 2023;2649:359-92.

21. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing

- biological themes among gene clusters. *Omics*. 2012;16(5):284-7.
22. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498-504.
 23. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7.
 24. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *Eur J Epidemiol*. 2015;30(7):543-52.
 25. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015;44(2):512-25.
 26. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol*. 2016;40(4):304-14.
 27. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol*. 2017;46(6):1985-98.
 28. Jiang H, Li Y, Wang T. Comparison of Billroth I, Billroth II, and Roux-en-Y reconstructions following distal gastrectomy: A systematic review and network meta-analysis. *Cir Esp (Engl Ed)*. 2021;99(6):412-20.
 29. Liu TT, Li R, Huo C, Li JP, Yao J, Ji XL, et al. Identification of CDK2-Related Immune Forecast Model and ceRNA in Lung Adenocarcinoma, a Pan-Cancer Analysis. *Front Cell Dev Biol*. 2021;9:682002.

30. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw.* 2010;33(1):1-22.
31. Zhang S, Sun L, Cai D, Liu G, Jiang D, Yin J, et al. Development and Validation of PET/CT-Based Nomogram for Preoperative Prediction of Lymph Node Status in Esophageal Squamous Cell Carcinoma. *Ann Surg Oncol.* 2023;30(12):7452-60.
32. Rowshanravan B, Halliday N, Sansom DM. CTLA-4: a moving target in immunotherapy. *Blood.* 2018;131(1):58-67.
33. Wu ZH, Yang DL, Wang L, Liu J. Epigenetic and Immune-Cell Infiltration Changes in the Tumor Microenvironment in Hepatocellular Carcinoma. *Front Immunol.* 2021;12:793343.
34. Wei J, Liao Z, Tao Y, Liu S. Evaluation of the possible association of PDCD-1 and LAG3 gene polymorphisms with hepatocellular carcinoma risk. *BMC Med Genomics.* 2023;16(1):92.
35. Long J, Wang A, Bai Y, Lin J, Yang X, Wang D, et al. Development and validation of a TP53-associated immune prognostic model for hepatocellular carcinoma. *EBioMedicine.* 2019;42:363-74.
36. Yu J, Ling S, Hong J, Zhang L, Zhou W, Yin L, et al. TP53/mTORC1-mediated bidirectional regulation of PD-L1 modulates immune evasion in hepatocellular carcinoma. *J Immunother Cancer.* 2023;11(11).
37. Zhao P, Zhen H, Zhao H, Huang Y, Cao B. Identification of hub genes and potential molecular mechanisms related to radiotherapy sensitivity in rectal cancer based on multiple datasets. *J Transl Med.* 2023;21(1):176.
38. Kasyanov ED, Yakovleva YV, Mudrakova TA, Kasyanova AA, Mazo GE. [Comorbidity patterns and structure of depressive episodes in patients with bipolar disorder and major depressive disorder]. *Zh Nevrol Psikhiatr Im S S Korsakova.*

2023;123(11. Vyp. 2):108-14.

39. Maeser D, Gruener RF, Huang RS. oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. *Brief Bioinform.* 2021;22(6).

40. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47.

41. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol.* 2019;20(2):163-72.

42. Griss J, Viteri G, Sidiropoulos K, Nguyen V, Fabregat A, Hermjakob H. ReactomeGSA - Efficient Multi-Omics Comparative Pathway Analysis. *Mol Cell Proteomics.* 2020;19(12):2115-25.

43. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, et al. Inference and analysis of cell-cell communication using CellChat. *Nat Commun.* 2021;12(1):1088.

44. Hong B, Li Y, Yang R, Dai S, Zhan Y, Zhang WB, et al. Single-cell transcriptional profiling reveals heterogeneity and developmental trajectories of Ewing sarcoma. *J Cancer Res Clin Oncol.* 2022;148(12):3267-80.

45. Chang J, Wu H, Wu J, Liu M, Zhang W, Hu Y, et al. Constructing a novel mitochondrial-related gene signature for evaluating the tumor immune microenvironment and predicting survival in stomach adenocarcinoma. *J Transl Med.* 2023;21(1):191.

46. Sankar K, Gong J, Osipov A, Miles SA, Kosari K, Nissen NN, et al. Recent advances in the management of hepatocellular carcinoma. *Clin Mol Hepatol.*

2024;30(1):1-15.

47. Kwong TT, Xiong Z, Zhang Y, Wu H, Cao J, Pak-Chun Wong P, et al. Overcoming immunotherapy resistance in hepatocellular carcinoma by targeting myeloid IL-8/CXCR2 signaling. *Mol Ther*. 2025;33(4):1659-73.

48. Gartmann M, Blau M, Armache JP, Mielke T, Topf M, Beckmann R. Mechanism of eIF6-mediated inhibition of ribosomal subunit joining. *J Biol Chem*. 2010;285(20):14848-51.

49. Sun L, Liu S, Wang X, Zheng X, Chen Y, Shen H. eIF6 promotes the malignant progression of human hepatocellular carcinoma via the mTOR signaling pathway. *J Transl Med*. 2021;19(1):216.

50. Scagliola A, Miluzio A, Ventura G, Oliveto S, Cordiglieri C, Manfrini N, et al. Targeting of eIF6-driven translation induces a metabolic rewiring that reduces NAFLD and the consequent evolution to hepatocellular carcinoma. *Nat Commun*. 2021;12(1):4878.

51. Scagliola A, Miluzio A, Mori G, Ricciardi S, Oliveto S, Manfrini N, et al. Inhibition of eIF6 Activity Reduces Hepatocellular Carcinoma Growth: An In Vivo and In Vitro Study. *Int J Mol Sci*. 2022;23(14).

52. Naz H, Islam A, Ahmad F, Hassan MI. Calcium/calmodulin-dependent protein kinase IV: A multifunctional enzyme and potential therapeutic target. *Prog Biophys Mol Biol*. 2016;121(1):54-65.

53. Li Z, Lu J, Zeng G, Pang J, Zheng X, Feng J, et al. MiR-129-5p inhibits liver cancer growth by targeting calcium calmodulin-dependent protein kinase IV (CAMK4). *Cell Death Dis*. 2019;10(11):789.

54. Weng X, Zeng L, Yan F, He M, Wu X, Zheng D. Cyclin-dependent kinase inhibitor 2B gene is associated with the sensitivity of hepatoma cells to Sorafenib.

Onco Targets Ther. 2019;12:5025-36.

55. Tao ZG, Yuan YX, Wang GW. Long non-coding RNA CDKN2B-AS1 promotes hepatocellular carcinoma progression via E2F transcription factor 1/G protein subunit alpha Z axis. *World J Gastrointest Oncol.* 2023;15(11):1974-87.

56. Hsueh KC, Lee HL, Ho KH, Chang LC, Yang SF, Chien MH. Disease-Associated Risk Variants and Expression Levels of the lncRNA, CDKN2B-AS1, Are Associated With the Progression of HCC. *J Cell Mol Med.* 2025;29(6):e70496.

57. Wen Y, Ren F, Wang W. Prognostic significance and regulatory mechanisms of NBEAL2 in HCC. *Asian J Surg.* 2024;47(6):2649-51.

58. Sowerby JM, Thomas DC, Clare S, Espéli M, Guerrero JA, Hoenderdos K, et al. NBEAL2 is required for neutrophil and NK cell function and pathogen defense. *J Clin Invest.* 2017;127(9):3521-6.

59. Raskov H, Orhan A, Salanti A, Gaggar S, Gögenur I. Natural Killer Cells in Cancer and Cancer Immunotherapy. *Cancer Lett.* 2021;520:233-42.

60. Obi CD, Bhuiyan T, Dailey HA, Medlock AE. Ferrochelatase: Mapping the Intersection of Iron and Porphyrin Metabolism in the Mitochondria. *Front Cell Dev Biol.* 2022;10:894591.

61. Wang T, Yang C, Li Z, Li T, Zhang R, Zhao Y, et al. Flavonoid 4,4'-dimethoxychalcone selectively eliminates senescent cells via activating ferritinophagy. *Redox Biol.* 2024;69:103017.

62. Yang C, Wang T, Zhao Y, Meng X, Ding W, Wang Q, et al. Flavonoid 4,4'-dimethoxychalcone induced ferroptosis in cancer cells by synergistically activating Keap1/Nrf2/HMOX1 pathway and inhibiting FECH. *Free Radic Biol Med.* 2022;188:14-23.

63. Malik A, Thanekar U, Amarachintha S, Mourya R, Nalluri S, Bondoc A, et al.

"Complimenting the Complement": Mechanistic Insights and Opportunities for Therapeutics in Hepatocellular Carcinoma. *Front Oncol.* 2020;10:627701.

64. Xiao Z, Yeung CLS, Yam JWP, Mao X. An update on the role of complement in hepatocellular carcinoma. *Front Immunol.* 2022;13:1007382.

65. Huang R, Zhu G, Fu X, Liu W, Tao C, Gao J, et al. Comprehensive analysis of complement-associated molecular features in hepatocellular carcinoma. *Acta Biochim Biophys Sin (Shanghai).* 2022;54(11):1694-707.

66. Heeren J, Scheja L. Metabolic-associated fatty liver disease and lipoprotein metabolism. *Mol Metab.* 2021;50:101238.

67. Chen J, Ding C, Chen Y, Hu W, Yu C, Peng C, et al. ACSL4 reprograms fatty acid metabolism in hepatocellular carcinoma via c-Myc/SREBP1 pathway. *Cancer Lett.* 2021;502:154-65.

68. Seo J, Jeong DW, Park JW, Lee KW, Fukuda J, Chun YS. Fatty-acid-induced FABP5/HIF-1 reprograms lipid metabolism and enhances the proliferation of liver cancer cells. *Commun Biol.* 2020;3(1):638.

69. Chen X, Liu X, Du S. Unveiling the Role of Tumor-Infiltrating T Cells and Immunotherapy in Hepatocellular Carcinoma: A Comprehensive Review. *Cancers (Basel).* 2023;15(20).

70. Godini R, Yan J, Chopin M. Transcriptional regulatory logic orchestrating lymphoid and myeloid cell fate decisions. *Front Immunol.* 2025;16:1544483.

71. Cheng X, Wang H, Wang Z, Zhu B, Long H. Tumor-associated myeloid cells in cancer immunotherapy. *J Hematol Oncol.* 2023;16(1):71.

72. Filipowicz W, Bhattacharyya SN, Sonenberg N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet.* 2008;9(2):102-14.

73. Yang Z, Qiu L, Jia G, Sun Z, Gong Y, Chen Y, et al. Multi-omics analysis reveals the role of tumor-infiltrating CD4(+)CCR7(+) T cells in EGFR antibody resistance and prognosis of hepatocellular carcinoma. BMC Cancer. 2025;26(1):17.

Figure legends

Figure 1. Identification and functional enrichment analysis of myeloid cell differentiation-related genes (MCDRGs) in hepatocellular carcinoma. (a) Volcano plot of differentially expressed genes (DEGs) in TCGA-HCC. (b) Venn diagram showing the overlap between DEGs and MCDRGs for candidate gene identification. (c) Top enriched Gene Ontology (GO) terms of the candidate genes. (d) Top enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways of the candidate genes.

Figure 2. Construction and validation of a myeloid cell differentiation-related prognostic signature in HCC. (a) Forest plot of univariate Cox regression analysis of the candidate MR feature genes. (b) LASSO coefficient profiles of the candidate genes. (c) Forest plot of multivariate Cox regression analysis identifying seven prognostic genes. (d) Distribution of risk scores and patient grouping (high-risk group [HRG] vs. low-risk group [LRG]) in the TCGA-HCC and ICGC-LIRI cohorts. (e) Distribution of survival status according to risk score. (f) Kaplan-Meier survival curves comparing the HRG and LRG. (g) Time-dependent ROC curves evaluating the predictive performance at 2, 3, and 5 years. (h) Expression levels of the seven prognostic genes in TCGA-HCC tumor and adjacent normal tissues.

Figure 3. Development and evaluation of a clinical prognostic nomogram integrating the risk score. (a) Univariate Cox regression analysis of clinical features and the risk score. (b) Multivariate Cox regression analysis identifying independent

prognostic factors. (c) Nomogram constructed based on the independent prognostic factors (T stage and risk score) for predicting 2-, 3-, and 5-year overall survival (OS). (d) Calibration curves assessing the agreement between nomogram-predicted and observed OS at 2, 3, and 5 years. (e) Time-dependent ROC curves evaluating the predictive performance of the nomogram at 2, 3, and 5 years.

Figure 4. Pathway enrichment and drug sensitivity analyses between the high- and low-risk groups. (a) Top 10 significantly enriched pathways identified by GSEA between the high-risk group (HRG) and low-risk group (LRG), ranked by absolute normalized enrichment score ($|NES|$). (b) Heatmap showing predicted drug sensitivity differences, based on IC50 values, between the HRG and LRG among the tested drugs. (c) Comparison of the top 5 drugs with the most significant differences in IC50 values between the HRG and LRG.

Figure 5. Mutation landscape, prognostic impact, and immune escape-related correlation of the seven prognostic genes. (a) Mutation profiles of the seven prognostic genes, showing mutation types and frequencies. (b) Kaplan-Meier survival curves comparing overall survival between patients with mutations in the prognostic genes and those without mutations. (c) Correlation heatmap showing the associations between the ssGSEA scores of the seven prognostic genes and immune escape-related genes (CTLA4, PDCD1, and TP53).

Figure 6. Single-cell RNA sequencing analysis of the HCC tumor microenvironment. (a) Selection of the top 2,000 highly variable genes (HVGs). (b, c) Determination of the number of principal components (PCs) for downstream analysis using JackStraw and Elbow plots, respectively. (d) UMAP visualization of 21 transcriptionally distinct cell clusters. (e, f) Cell type annotation of the clusters using canonical marker genes and the "singleR" package, identifying major cell types (T

cells, NK cells, myeloid cells, endothelial cells, B cells, hepatocytes, fibroblasts) and undefined clusters.

Figure 7. Cellular composition and intercellular communication analysis in the HCC single-cell landscape. (a) Comparison of the abundance of major annotated cell types between HCC and control samples. (b) GSVA pathway enrichment analysis of differentially abundant cell types. (c, d) Number and strength of cell-cell interactions in the control group. (e, f) Number and strength of cell-cell interactions in the HCC group. (g, h) Ligand-receptor interaction analysis showing the dominant expression of the MIF-(CD74+CXCR4) pair in both groups.

Figure 8. Expression and dynamics of prognostic genes in T cell subpopulations. (a,b) Sub-clustering and annotation of T cells, identifying CD8⁺ and CD4⁺ T cell clusters. (c,d) Pseudotime trajectory analysis showing the inferred differentiation path of T cells, with CD4⁺ T cells at the start and CD8⁺ T cells at the end. (e) Expression patterns of the prognostic genes along the pseudotime trajectory, with GPR171 showing a distinct fluctuation.

Figure 9. Experimental validation of prognostic gene expression in HCC clinical samples. (a-d) Representative immunohistochemistry (IHC) staining images showing the protein expression of FECH, CAMK4, EIF6, and CDKN2B in HCC and control tissues. (e-i) RT-qPCR validation of the mRNA expression levels of CDKN2B, EIF6, NBEAL2, FECH, and CAMK4 in clinical samples.

Additional files

Figure S1. The PPI network demonstrated that 158 candidate genes exhibited interactions at the protein level.

Figure S2. Forest plots of MR results.

Figure S3. Scatter plots of MR results.

Figure S4. Funnel plots of MR results.

Figure S5. LOO test results of MR results.

Figure S6. Single-cell data procession of 7 cell types.

Figure S7. Prognostic gene expression of 7 cell types

Additional file 1. The list of 423 macrophage-related cell death regulator genes (MCDRGs) used in this study.

Additional file 2.Primer sequences used for reverse transcription quantitative polymerase chain reaction (RT-qPCR) validation.

Additional file 3.Results of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses for the 173 candidate genes.

Additional file 4.Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis for the candidate genes.

Additional file 5.Results of MR analysis (IVW method)

Additional file 6.Heterogeneity testing and effect model selection for the Mendelian randomization (MR) analysis of the 40 significant candidate genes.

Additional file 7.Assessment of horizontal pleiotropy for the Mendelian randomization (MR) analysis of the 40 significant candidate genes.

Additional file 8.Results of the Steiger directionality test for the Mendelian randomization (MR) analysis of the 40 significant candidate genes.

Additional file 9.Results of the proportional hazards (PH) assumption test for the candidate genes in the survival analysis.

Additional file 10.Results of the proportional hazards (PH) assumption test for clinical features and the risk score in the survival model.

Additional file 11.Results of the Gene Set Enrichment Analysis (GSEA) between the high-risk group (HRG) and low-risk group (LRG) in the TCGA-HCC cohort.

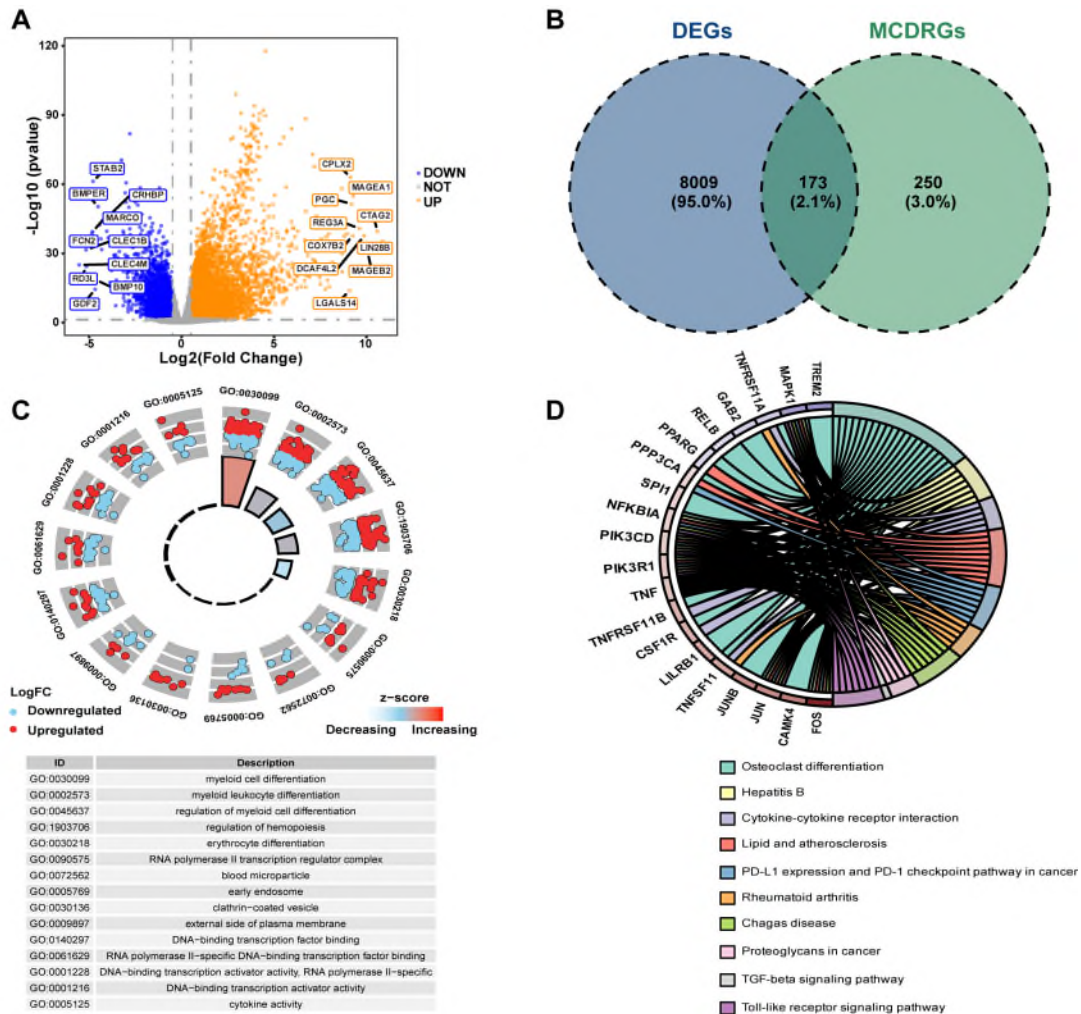
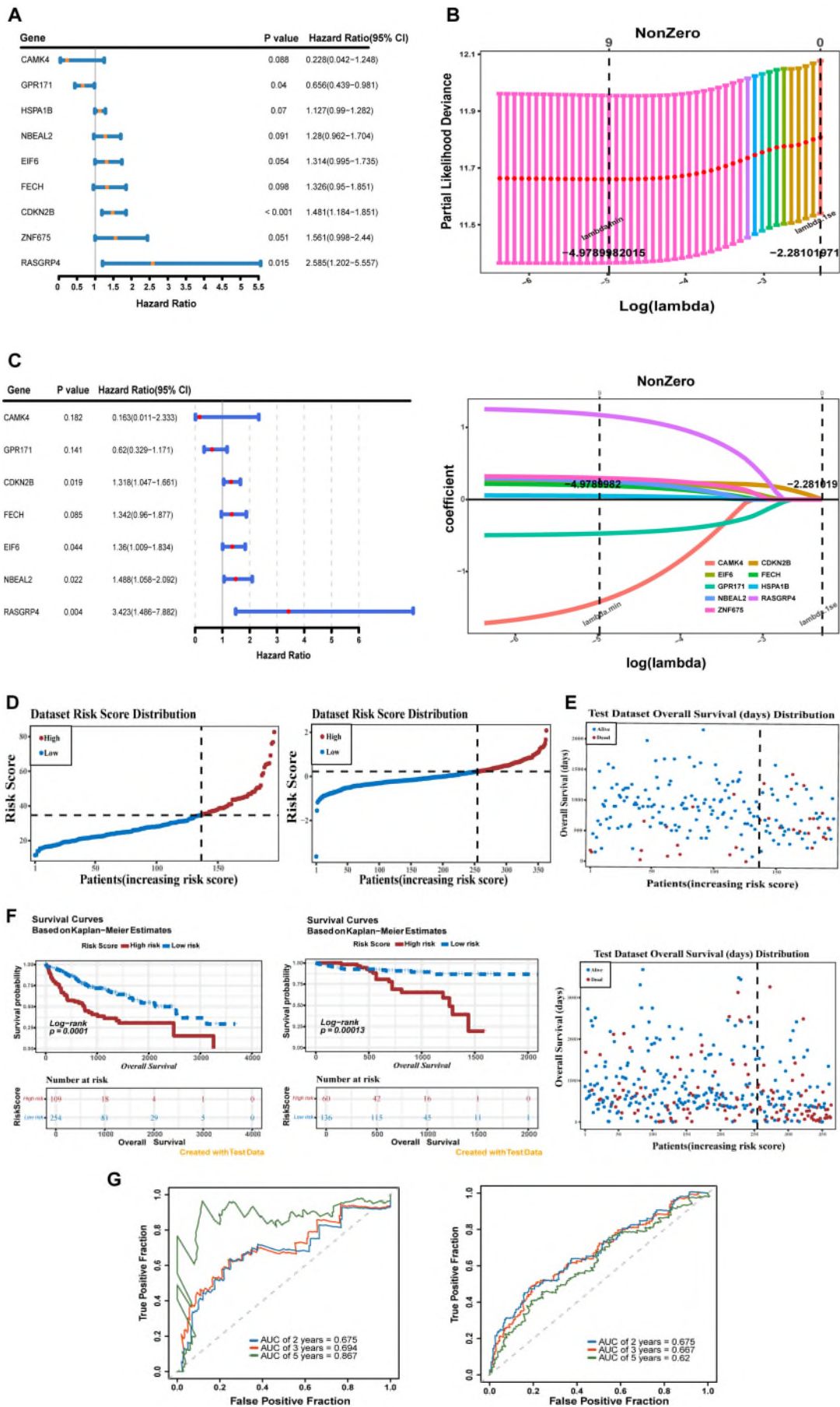


Figure 1. Identification and functional enrichment analysis of myeloid cell differentiation-related genes (MCDRGs) in hepatocellular carcinoma. (a) Volcano plot of differentially expressed genes (DEGs) in TCGA-HCC. (b) Venn diagram showing the overlap between DEGs and MCDRGs for candidate gene identification. (c) Top enriched Gene Ontology (GO) terms of the candidate genes. (d) Top enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways of the candidate genes.



H

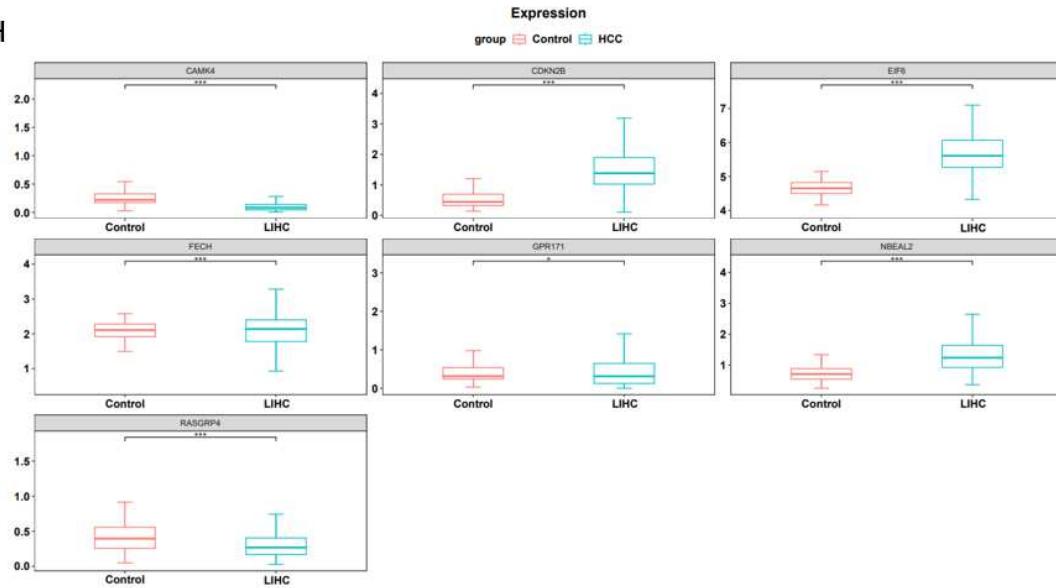


Figure 2. Construction and validation of a myeloid cell differentiation-related prognostic signature in HCC. (a) Forest plot of univariate Cox regression analysis of the candidate MR feature genes. (b) LASSO coefficient profiles of the candidate genes. (c) Forest plot of multivariate Cox regression analysis identifying seven prognostic genes. (d) Distribution of risk scores and patient grouping (high-risk group [HRG] vs. low-risk group [LRG]) in the TCGA-HCC and ICGC-LIRI cohorts. (e) Distribution of survival status according to risk score. (f) Kaplan-Meier survival curves comparing the HRG and LRG. (g) Time-dependent ROC curves evaluating the predictive performance at 2, 3, and 5 years. (h) Expression levels of the seven prognostic genes in TCGA-HCC tumor and adjacent normal tissues.

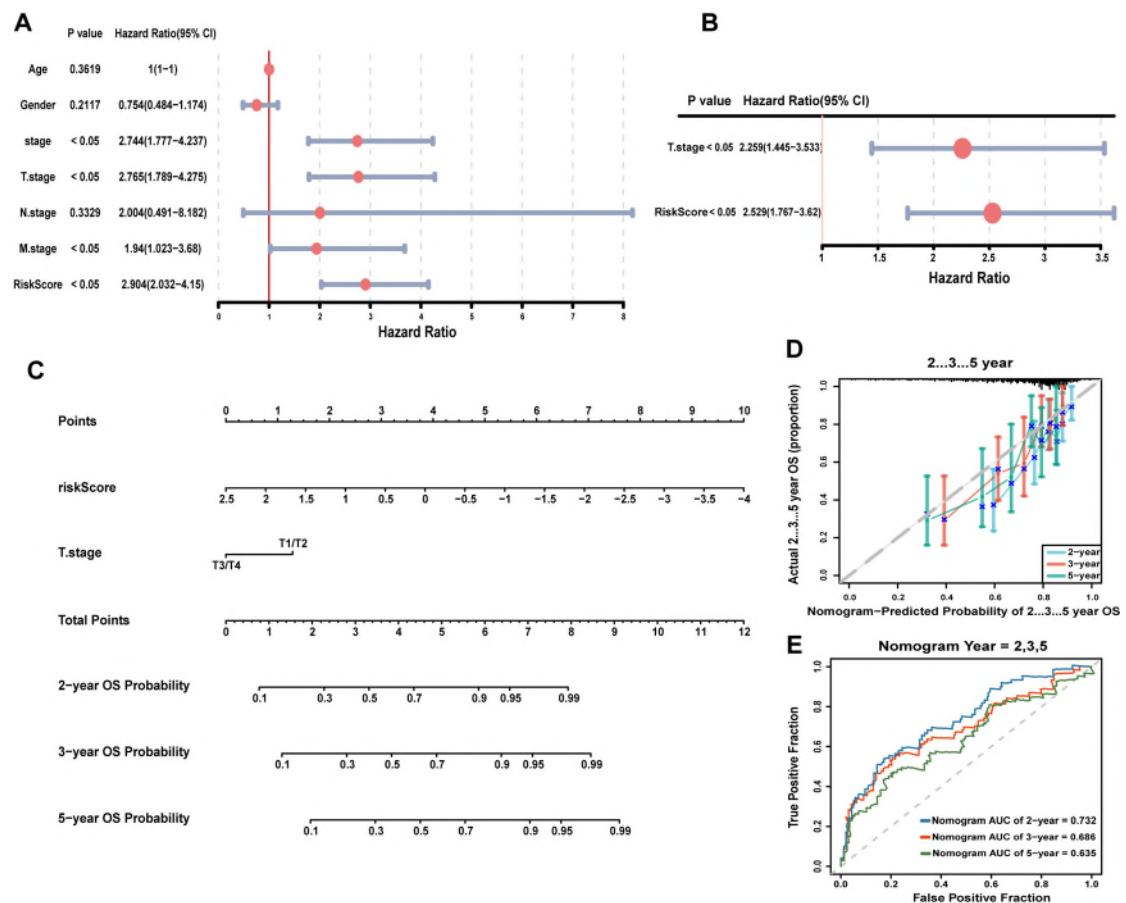


Figure 3. Development and evaluation of a clinical prognostic nomogram integrating the risk score. (a) Univariate Cox regression analysis of clinical features and the risk score. (b) Multivariate Cox regression analysis identifying independent prognostic factors. (c) Nomogram constructed based on the independent prognostic factors (T stage and risk score) for predicting 2-, 3-, and 5-year overall survival (OS). (d) Calibration curves assessing the agreement between nomogram-predicted and observed OS at 2, 3, and 5 years. (e) Time-dependent ROC curves evaluating the predictive performance of the nomogram at 2, 3, and 5 years.

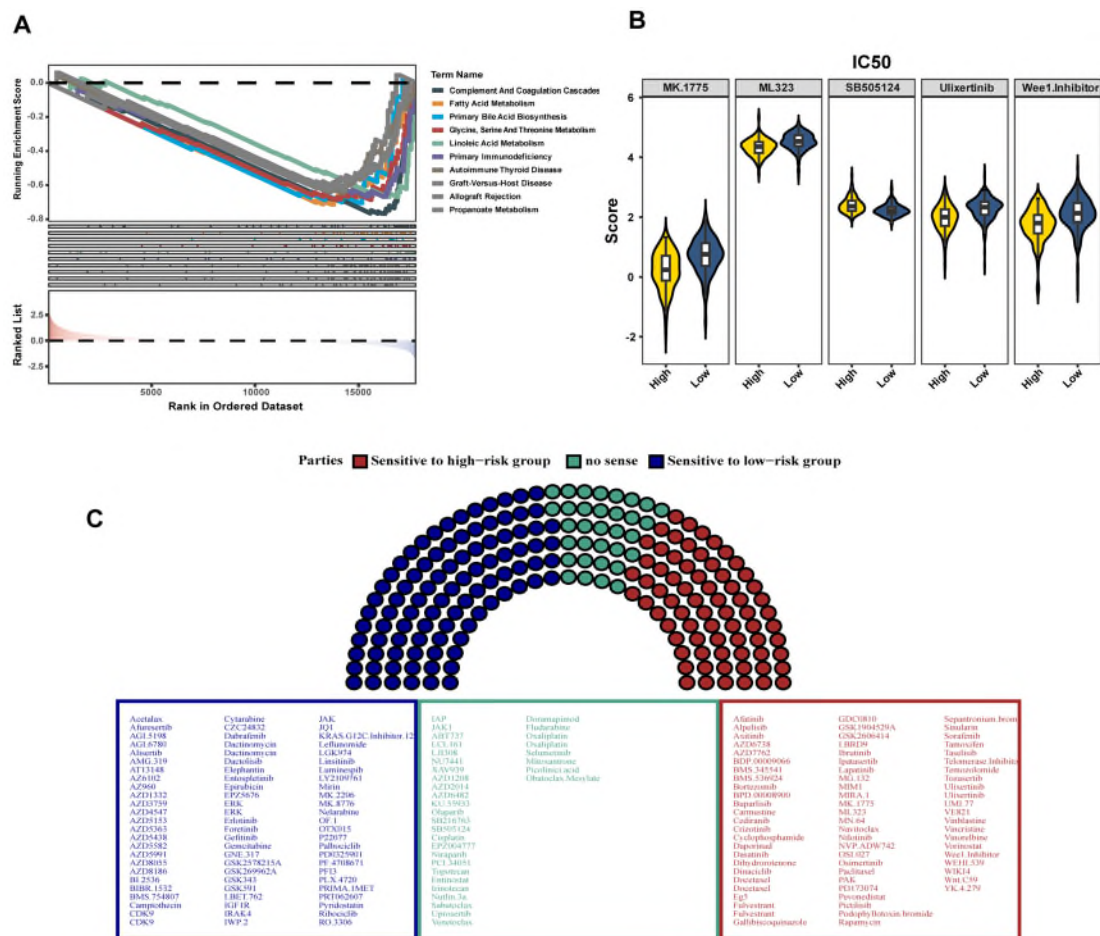


Figure 4. Pathway enrichment and drug sensitivity analyses between the high- and low-risk groups. (a) Top 10 significantly enriched pathways identified by GSEA between the high-risk group (HRG) and low-risk group (LRG), ranked by absolute normalized enrichment score ($|NES|$). (b) Heatmap showing predicted drug sensitivity differences, based on IC50 values, between the HRG and LRG among the tested drugs. (c) Comparison of the top 5 drugs with the most significant differences in IC50 values between the HRG and LRG.

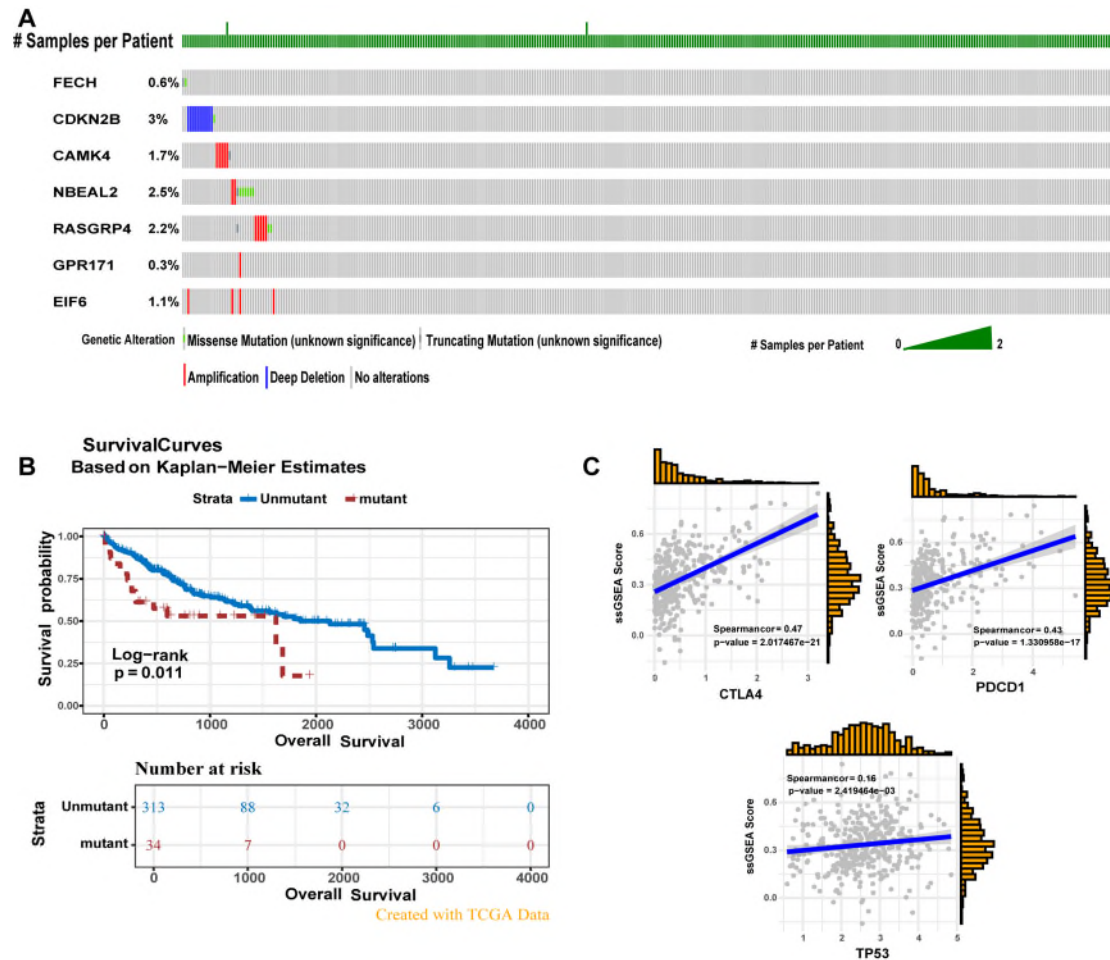


Figure 5. Mutation landscape, prognostic impact, and immune escape-related correlation of the seven prognostic genes. (a) Mutation profiles of the seven prognostic genes, showing mutation types and frequencies. (b) Kaplan-Meier survival curves comparing overall survival between patients with mutations in the prognostic genes and those without mutations. (c) Correlation heatmap showing the associations between the ssGSEA scores of the seven prognostic genes and immune escape-related genes (CTLA4, PDCD1, and TP53).

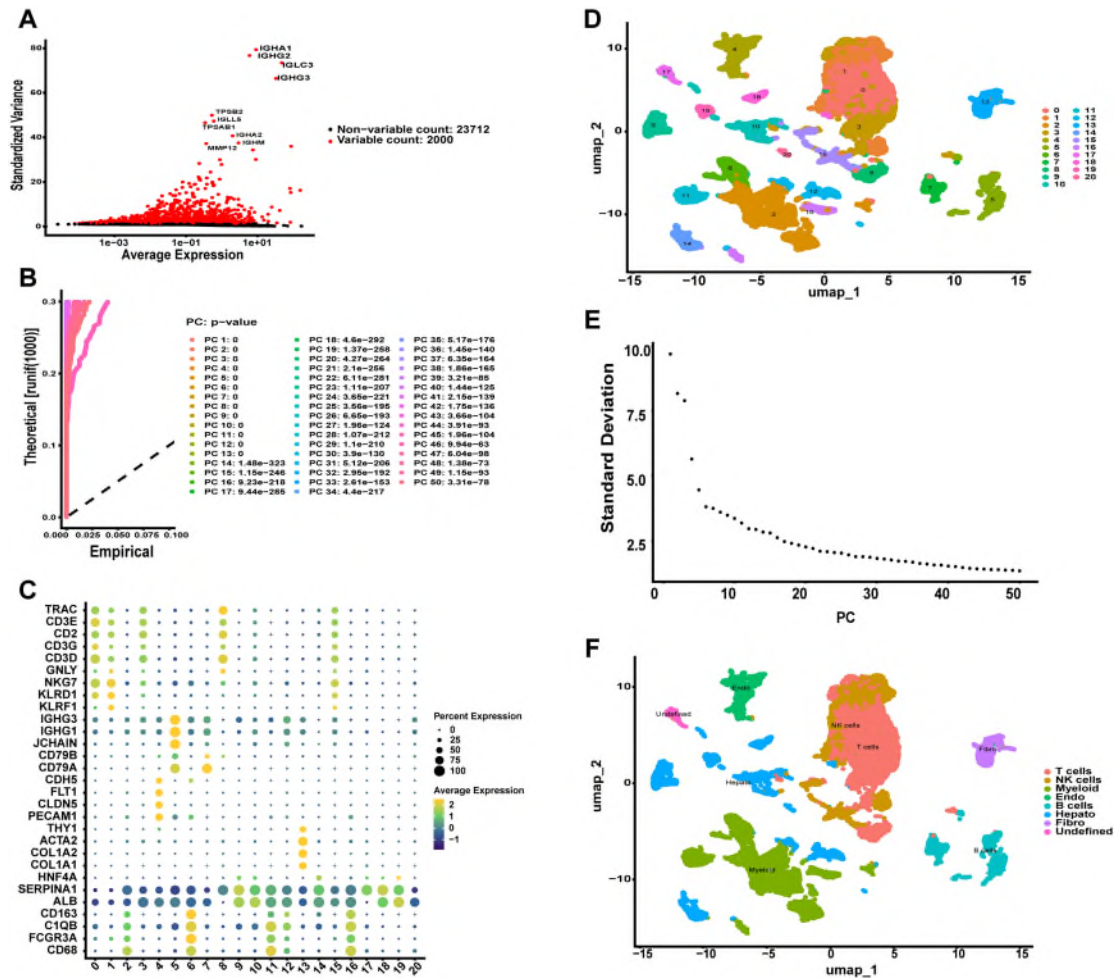


Figure 6. Single-cell RNA sequencing analysis of the HCC tumor microenvironment. (a) Selection of the top 2,000 highly variable genes (HVGs). (b, c) Determination of the number of principal components (PCs) for downstream analysis using JackStraw and Elbow plots, respectively. (d) UMAP visualization of 21 transcriptionally distinct cell clusters. (e, f) Cell type annotation of the clusters using canonical marker genes and the "singleR" package, identifying major cell types (T cells, NK cells, myeloid cells, endothelial cells, B cells, hepatocytes, fibroblasts) and undefined clusters.

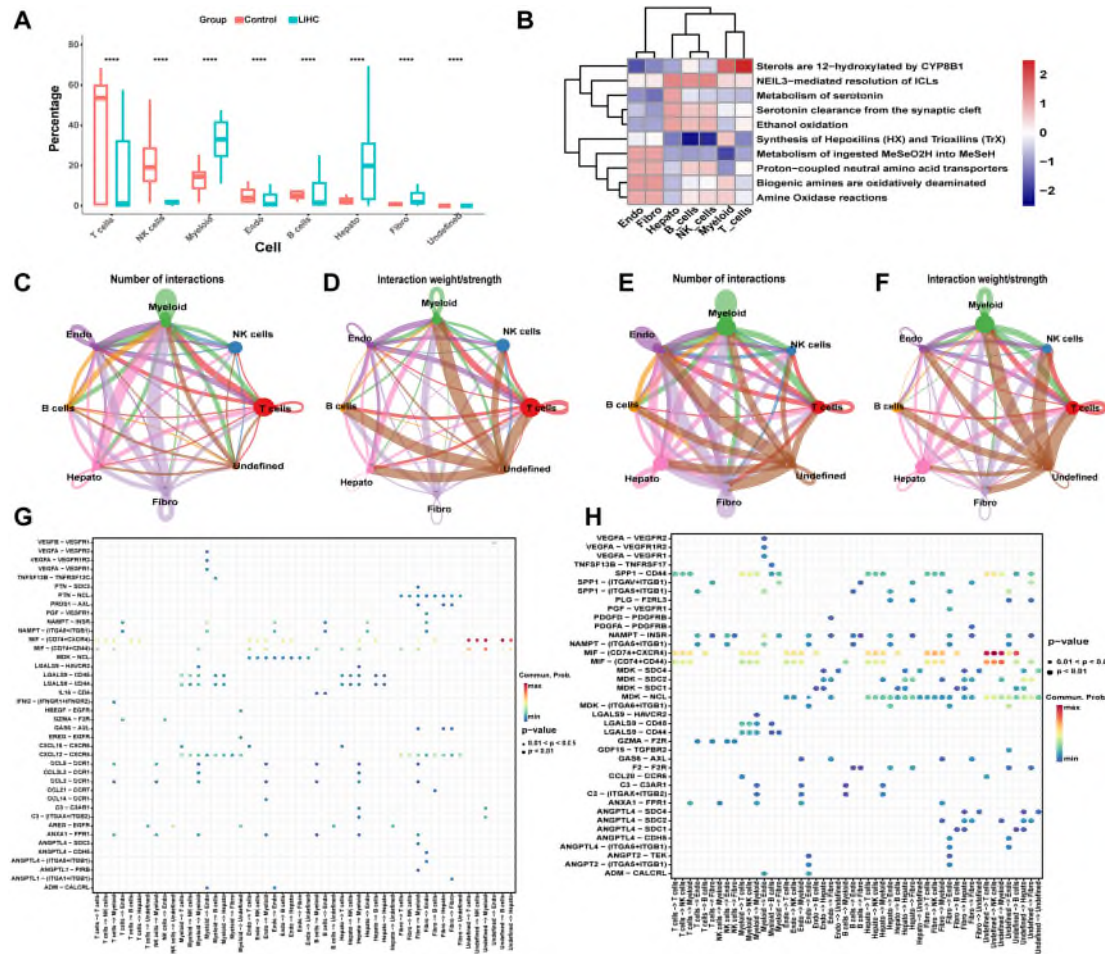


Figure 7. Cellular composition and intercellular communication analysis in the HCC single-cell landscape. (a) Comparison of the abundance of major annotated cell types between HCC and control samples. (b) GSVA pathway enrichment analysis of differentially abundant cell types. (c, d) Number and strength of cell-cell interactions in the control group. (e, f) Number and strength of cell-cell interactions in the HCC group. (g, h) Ligand-receptor interaction analysis showing the dominant expression of the MIF-(CD74+CXCR4) pair in both groups.

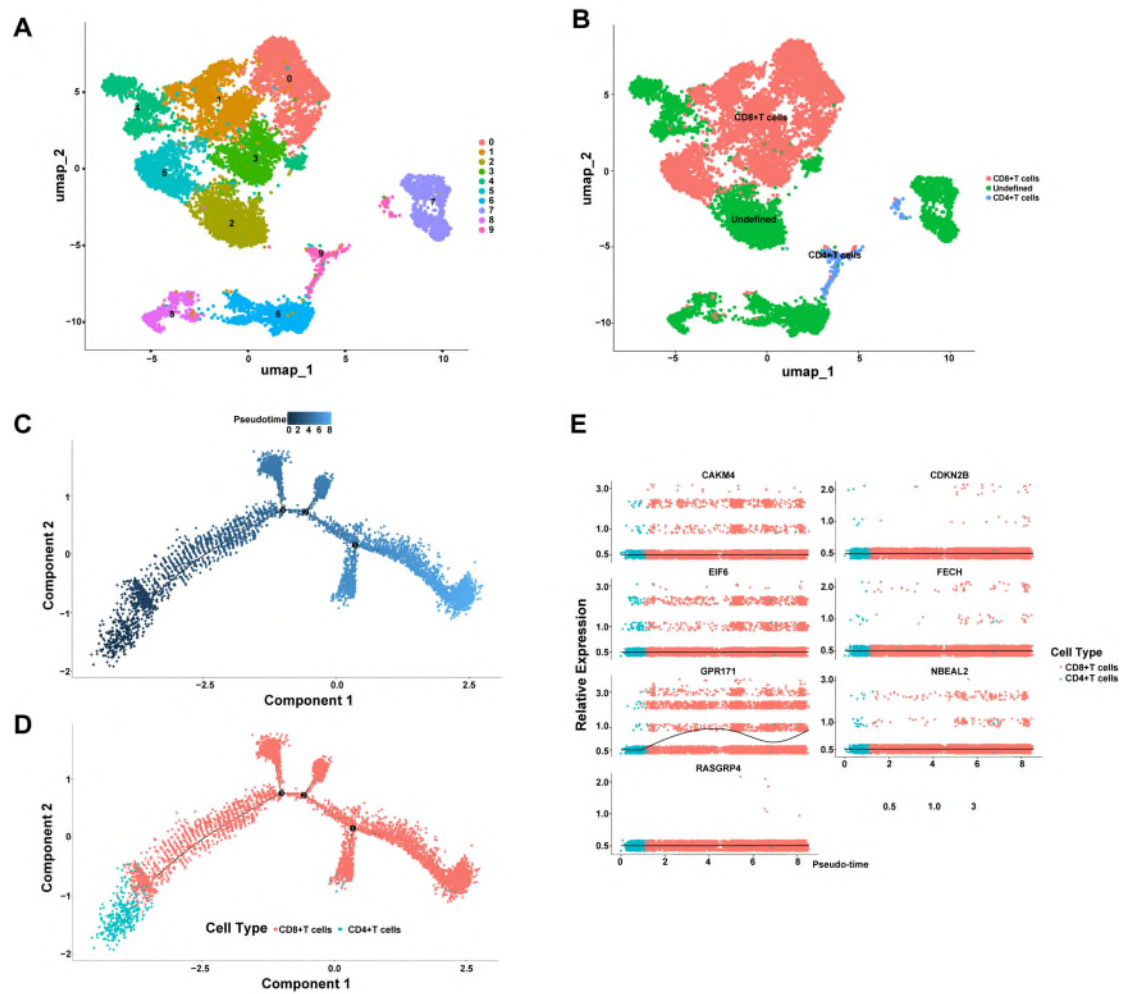


Figure 8. Expression and dynamics of prognostic genes in T cell subpopulations. (A,B) Sub-clustering and annotation of T cells, identifying CD8⁺ and CD4⁺ T cell clusters. (C,D) Pseudotime trajectory analysis showing the inferred differentiation path of T cells, with CD4⁺ T cells at the start and CD8⁺ T cells at the end. (E) Expression patterns of the prognostic genes along the pseudotime trajectory, with GPR171 showing a distinct fluctuation.

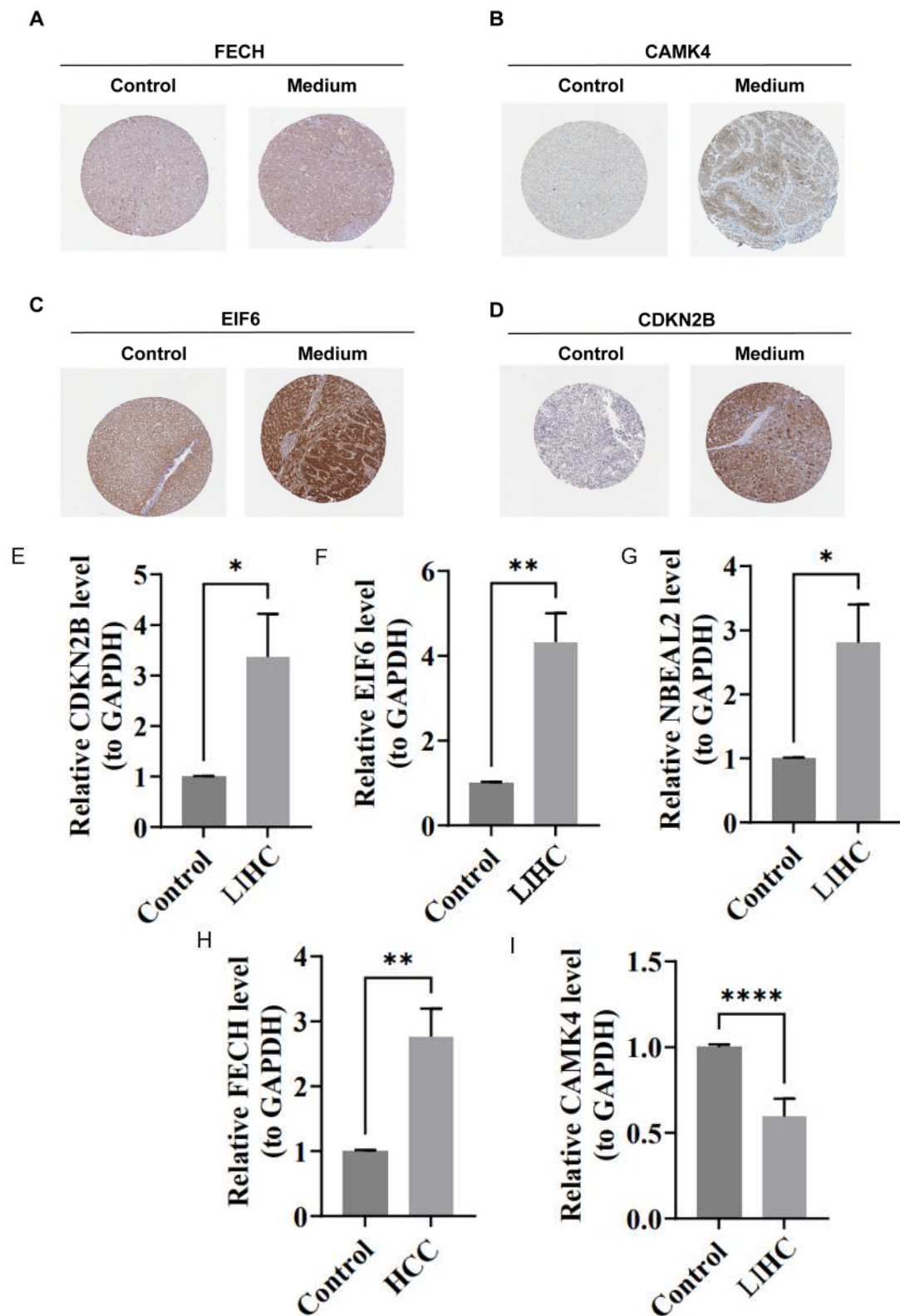


Figure 9. Experimental validation of prognostic gene expression in HCC clinical samples. (a-d)

Representative immunohistochemistry (IHC) staining images showing the protein expression of FECH,

CAMK4, EIF6, and CDKN2B in HCC and control tissues. (e-i) RT-qPCR validation of the mRNA expression levels of CDKN2B, EIF6, NBEAL2, FECH, and CAMK4 in clinical samples.

Supplementary Files

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

- [Additionalfile.zip](#)