Computational and Structural Biotechnology Journal 23 (2024) 929–941

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

# Prognostic iron-metabolism signature robustly stratifies single-cell characteristics of hepatocellular carcinoma

Zhipeng Zhu [a,b,1], Huang Cao [c,1], Hongyu Yan [c,d], Hanzhi Liu [e], Zaifa Hong [f], Anran Sun [g,h,*,2], Tong Liu [a,b,**,3], Fengbiao Mao [a,b,**,4]

[a] Institute of Medical Innovation and Research, Peking University Third Hospital, Beijing 100191, China
[b] Cancer Center, Peking University Third Hospital, Beijing 100191, China
[c] School of Medicine, Xiamen University, Xiamen, Fujian 361100, China
[d] Key Laboratory of Ministry of Education for Gastrointestinal Cancer, School of Basic Medical Sciences, Fujian Medical University, Fuzhou, Fujian 350122, China
[e] The Third Clinical Medical College, Gannan Medical University, Ganzhou, Jiangxi 341000, China
[f] Department of Hepatobiliary Surgery, The First Affiliated Hospital, School of Medicine, Xiamen University, Xiamen, Fujian 361103, China
[g] Oncology Research Center, Foresea Life Insurance Guangzhou General Hospital, Guangzhou, Guangdong 511300, China
[h] Research Center for Translational Medicine, The First Affiliated Hospital, School of Medicine, Xiamen University, Xiamen, Fujian 361003, China

ARTICLE INFO

*Keywords:*
Hepatocellular carcinoma (HCC)
Iron metabolism-related genes (IRGs)
Immune infiltration
Iron metabolism-related gene prognostic index (IPX)
Personalized medicine

ABSTRACT

Cancer immunotherapy has shown to be a promising method in treating hepatocellular carcinoma (HCC), but suboptimal responses in patients are attributed to cellular and molecular heterogeneity. Iron metabolism-related genes (IRGs) are important in maintaining immune system homeostasis and have the potential to help develop new strategies for HCC treatment. Herein, we constructed and validated the iron-metabolism gene prognostic index (IPX) using univariate Cox proportional hazards regression and LASSO Cox regression analysis, successfully categorizing HCC patients into two groups with distinct survival risks. Then, we performed single-sample gene set enrichment analysis, weighted correlation network analysis, gene ontology enrichment analysis, cellular lineage analysis, and SCENIC analysis to reveal the key determinants underlying the ability of this model based on bulk and single-cell transcriptomic data. We identified several driver transcription factors specifically activated in specific malignant cell sub-populations to contribute to the adverse survival outcomes in the IPX-high subgroup. Within the tumor microenvironment (TME), T cells displayed significant diversity in their cellular characteristics and experienced changes in their developmental paths within distinct clusters identified by IPX. Interestingly, the proportion of Treg cells was increased in the high-risk group compared with the low-risk group. These results suggest that iron-metabolism could be involved in reshaping the TME, thereby disrupting the cell cycle of immune cells. This study utilized IRGs to construct a novel and reliable model, which can be used to assess the prognosis of patients with HCC and further clarify the molecular mechanisms of IRGs in HCC at single-cell resolution.

## 1. Introduction

Liver cancer ranks as the sixth most prevalent tumor, with around 900,000 new cases diagnosed globally each year [1]. However, it ranks third in terms of mortality, with about 830,000 new deaths each year globally [1]. Hepatocellular carcinoma (HCC), constitutes approximately 90% of all liver cancers and stands as the most common type of liver cancer [2]. HCC exhibits high morbidity and mortality since it has

* Corresponding author at: Oncology Research Center, Foresea Life Insurance Guangzhou General Hospital, Guangzhou, Guangdong 511300, China.
** Corresponding authors at: Institute of Medical Innovation and Research, Peking University Third Hospital, Beijing 100191, China.
*E-mail addresses:* 15779722296@163.com (Z. Zhu), maody.cao@foxmail.com (H. Cao), hongyu@fjmu.edu.cn (H. Yan), lhz15870928158@163.com (H. Liu), lantianzaiqianmian@163.com (Z. Hong), sunanran@aliyun.com (A. Sun), tongtongliu@vip.sina.com (T. Liu), fengbiaomao@bjmu.edu.cn (F. Mao).
[1] These authors contributed equally to this work.
[2] 0000–0001-7165–7016;
[3] 0009–0001-4603–2580
[4] 0000–0003-0852–4266.

https://doi.org/10.1016/j.csbj.2024.01.022
Received 26 October 2023; Received in revised form 23 January 2024; Accepted 29 January 2024
Available online 1 February 2024

complex molecular mechanisms due to tumor heterogeneity [3]. Despite the improvements in treatments, HCC remains one of the cancer types with the highest rate of tumor-related fatalities, exhibiting only 21% for 5-year survival rate [4]. For patients in the early stages of HCC, the primary treatment choices include surgical and radiological ablations coupled with localized chemotherapy. However, there are no remedies that can cure advanced HCC, and further therapeutic approaches are required to prevent the progression of HCC. As a crucial factor in the tumor microenvironment (TME), the immunophenotype has played a vital role in influencing the prognosis of patients with HCC. At the same time, the high cellular and molecular heterogeneity results in rather suboptimal response rates in patients with HCC [5]. Although various predictive models can classify individuals into separate risk categories, the ability of these models to differentiate the characteristics at the level of single cells remains unclear. Therefore, it is crucial to explore the TME and analyze the developmental dynamics of tumor cells, which allows us to acquire a profound understanding of the diverse immune responses against heterogeneous HCC tumor cells. The single-cell sequencing methods are state-of-the-art techniques for analyzing the immune landscape and heterogeneity of human cancers [6,7].

The liver is rich in iron, which is crucial for iron metabolism [8,9]. The increase of iron metabolism in liver tissue contributes to the development of HCC via various mechanisms, including providing growth and metabolic needs, generation of ROS, decreasing p53 protein level and its activity, and increasing mesenchymal and metastatic potential [10–15]. The iron metabolism-related genes (IRGs) are significantly modified via transcriptional and translational regulation in HCC. Clarifying potential changes in iron metabolism is necessary for understanding the development and progression of HCC. However, the iron metabolism in TME exhibits different effects in different types of tumors. Tumors can be categorized as 'hot' or 'cold' tumors, based on the level of infiltrated inflammatory cells and the degree of inflammatory response in the TME [16]. Iron often accumulates in hot tumors and supports cancer cell progression via influencing the ability of antigen-presenting cells (APCs) and T/B lymphocytes in TME [17]. As a representative of hot tumors, HCC frequently exhibits elevated levels of inflammation associated with tumor progression. This inflammation has been found to be linked with unfavorable patient survival rates and adverse therapeutic results [18]. Therefore, based on the malignant nature of liver cancer and its sensitive inflammatory response, further investigation is needed to understand how iron metabolism affects tumor-infiltrating lymphocytes, as well as its role in reshaping metabolic pathways within the immune microenvironment of HCC.

In the current study, we developed and validated a stable and reliable prognostic scoring model using genes related to iron metabolism. Additionally, we concentrated on the cellular and molecular heterogeneity of HCC patients with different prognostic scores, and uncovered distinct immune features and molecular mechanisms underlying the capability of the prognostic scoring model. According to our study, the patient prognosis was linked to the makeup of cell composition and molecular heterogeneity. We also discovered that specific subpopulations of malignant cells and Treg cells may account for distinct prognoses based on our scoring model. The study offers innovative perspectives for molecular mechanisms underlying the distinct clinical outcomes at the single-cell level, shedding new light on exploring the molecular and phenotype heterogeneity in HCC.

## 2. Materials and methods

### 2.1. Data collection and process

The 10 sets of IRG were acquired through Gene Set Enrichment Analysis (GSEA) (http://software.broadinstitute.org/gsea/index.jsp) [19] (Supplementary Dataset 1). Both expression and clinical data for the LIHC dataset, encompassing 369 patients with HCC, were obtained from The Cancer Genome Atlas (TCGA, https://cancergenome.nih.gov). The ICGC-LIHC dataset, consisting of 243 patients with HCC, was sourced from the International Cancer Genome Consortium (ICGC, https://dcc.icgc.org/). Genetic alteration information was sourced from UCSC Xena (https://xena.ucsc.edu/). Single-cell transcriptome datasets for infiltrating T cells (GSE98638) [7] and HCC samples (GSE149614) [20] were retrieved from the GEO database (https://www.ncbi.nlm.nih.gov/geo/).

### 2.2. Differential gene expression analysis

Differentially expressed iron metabolism-related genes between patients who survived $< 1$ year and $> 3$ years were identified from the TCGA-LIHC using DEGseq in R software with the standard $|\text{Log2 FC}| >= 0.5$ and adjusted $P < 0.05$ [21].

### 2.3. Classification of HCC clusters

Using the PAM clustering algorithm, the differentially expressed IRGs (DIRGs) were subjected to unsupervised consensus clustering based on TCGA-LIHC data. The clustering analyses were performed by repeating one thousand iterations, each with a random fraction of DIRGs of 0.95. The optimal cluster was selected based on the absence of an appreciable increase.

### 2.4. Prognostic IPX construction

For the construction of the scoring model, the mRNA-seq datasets from 369 HCC patients in TCGA-LIHC were employed. Initially, univariate Cox proportional hazards regression analysis was performed to identify DIRGs with an adjusted *P*-value below 0.05. These selected DIRGs underwent a further 22 stepwise LASSO Cox regression to mitigate overfitting [22]. Subsequently, the iron metabolism-related gene prognostic index (IPX) was utilized to evaluate the prognostic risk for HCC individuals, where a higher score indicates a poorer prognosis. Based on the median risk score, patients with HCC were stratified into high- or low-risk groups. The predictive performance was evaluated using Kaplan-Meier survival analyses, the log-rank test, and the time-ROC analyses [23]. Both univariate and multivariate Cox regression analyses were conducted to ascertain the independence of IPX from other clinical parameters.

### 2.5. Prognostic IPX validation and construction of the quantitative nomogram

A total of 184 TCGA-LIHC patients were selected as the internal testing cohort in a 1:1 ratio by a random manner. Simultaneously, the 243 ICGC-LIHC patients were obtained as the external testing cohort. We computed patient's risk scores in the validation dataset based on IPX, and subsequently divided them into high-risk and low-risk groups using the median risk score from the training cohort. Various analyses, including Kaplan-Meier survival analyses, time-ROC analyses, and both univariate and multivariate Cox regression analyses, were carried out to assess the predictive ability. To improve the predictive accuracy of the model, we established a nomogram by integrating the score model with clinically significant independent prognostic factors.

### 2.6. Comprehensive analysis of molecular and immune characteristics using bulk sequencing data

To decipher the mechanism that underlies the model's ability, we performed weighted correlation network analysis (WGCNA) [24] to investigate modules strongly linked to the high-risk group, The genes in module with high correlation coefficients were considered as suitable candidates for further pathway enrichment analysis using "Metascape" [24,25]. Several immune gene sets were manually collected to demonstrate immune infiltration status among HCC patients [26–28].

Moreover. The CIBERSORT algorithm was applied to elucidate disparities in the distribution of distinct immune cell types within the TME of HCC patients. Immune Cell Abundance Identifier (ImmuneCellAI), the tumor immune dysfunction and exclusion (TIDE), and The Cancer Immunome Atlas (TCIA) were employed to predict the efficacy of immunotherapy [29–31]. The information on genetic alterations was obtained from UCSC Xena (https://xena.ucsc.edu/), and the quantity and quality of gene mutations were analyzed using the Maftools [32]. To assess the transcription factor activity of E2Fs family, we performed a correlation analysis between E2Fs members and all protein-coding genes. The top 50 genes with the strongest correlation with E2Fs members were identified as signature genes to assess the transcription factor activity using ssGSEA [19].

### 2.7. Comprehensive analysis of molecular and immune characteristics using single-cell RNA sequencing data

The single-cell transcriptome dataset of infiltrating T cells comprises 4070 T cells from HCC [7]. The coefficients from the scoring model were used to calculat the risk score for each single cell. We utilized the median score calculated from the scoring model across all individual cells as the threshold to define low-score and high-score subpopulations. The pseudotime developmental cell trajectory was inferred using the Monocle 2 [33]. The DifferentialGeneTest function was performed to detect genes that exhibited differential expression across pseudotime with adjusted *P*-value < 0.01. BEAM function in Monocle 2 was used to identify branch-dependent genes.

The single-cell dataset of HCC samples [20] comprises samples from normal liver tissue(Normal, n = 1788), primary tumor (PT, n = 14202), portal vein tumor thrombus (PVTT, n = 3329) and metastatic lymph node (MLN, n = 1463). The human reference (GRCh38) was obtained from the official website of 10X Genomics. Subsequently, gene expression matrices after quality filtration, were generated using the CellRanger tool (v7.1.0). Next, low-quality cells and doublets were removed according to the following standards: (1) doublet Score > 0.25; (2) unique molecular identifiers (UMIs) < 200 or UMIs > 8000; (3) the number of detected genes < 200 (4) percentage of mitochondrial UMIs > 10%. 71915 cells were retained for downstream analysis. We employed the Harmony algorithm [34] to integrate cells from different patient sources. A consensus list of 2000 highly-variable genes (HVGs) was formed across samples. Principal Component Analysis (PCA) was utilized to calculate 40 principal components (PCs). Nearest neighbor graphs were constructed based on the top 10 PCs using the FindNeighbors function. Subsequently, we identified six major cell types based on canonical cell-type markers, including CD3D (TNK cells), IGHG1 (B cells), ALB (hepatocyte), CD68 (myeloid), VWF (endothelial), ACTA2 (fibroblast). To delineate discrete clusters within hepatocytes, a subsequent round of clustering was conducted individually to identify 15 distinct clusters. The methodology employed in this second round of clustering mirrored that of the initial round, ensuring consistency in the analytical procedures. The risk scoring of hepatocytes, as well as the categorization into high and low-risk groups, aligns with the aforementioned methodology applied to the human infiltrating T cells processing.

Then, we employed the WGCNA [24] to construct the co-expression network. Utilizing default parameters, we established a signed network by integrating genes identified as noise-robust through optimal gene filtering for single-cell data (OGFSC) [35]. To group genes with consistent expression profiles into modules, we utilized average linkage hierarchical clustering, with the topological overlap measure serving as the dissimilarity metric. To evaluate and rank the hubness of modular genes within each gene module, we utilized modular gene centrality, which involves the summation of within-cluster connectivity measures.

Finally, we evaluated the transcription factor regulation strength in the single-cell regulatory network by utilizing pySCENIC with default parameters [36]. The choice of DNA-motif analysis was determined through RcisTarget [37], while the identification of gene networks was achieved using AUCell [36]. Regulon modules were identified utilizing the Connection Specificity Index (CSI), and hierarchical clustering with Euclidean distance was executed on the CSI matrix to discern distinct regulon modules.

### 2.8. Clinical samples collection and RT-qPCR analysis

A total of 12 hepatocellular carcinoma (HCC) specimens, along with corresponding adjacent normal tissues, were procured from the First Affiliated Hospital of Xiamen University in China. Demographic and clinical characteristics of these 12 HCC-diagnosed patients are detailed in Table S1. Institutional review board approval was secured for this study at the First Affiliated Hospital of Xiamen University. Extraction of total RNA from clinical sample tissues adhered to the instructions of RNAiso Plus (Cat#9109, TAKARA, Japan). Reverse transcription involved the use of extracted total RNAs with an RT reagent Kit with gDNA Eraser (Cat# RR047A, TAKARA, Japan), following the provided instructions. The RT-qPCR was executed using the SYBR Green Real-Time PCR Master Mix (Cat#RR820A, TAKARA, Japan) and the LightCycler480II Real-Time PCR system (Roche, CH), following manufacturers' protocols and utilizing appropriate primer pairs, with 18S rRNA employed as a control.

### 2.9. Statistical analysis

All statistical analyses were conducted using R (v4.1.0). Survival analysis employed the Kaplan-Meier method and Log-rank test, with the exclusion of patients lacking completed follow-up data. Two group comparisons utilized the t-test and three or more group comparisons utilized the ANOVA test. Independent prognostic factor analysis involved both univariate and multivariate Cox regression analysis, with the exclusion of patients lacking any characteristics. The prediction accuracy of prognostic indicators was assessed using time-dependent receiver operating characteristic curves. A threshold of adjusted *P-value* < 0.05 was set for statistical significance.

## 3. Result

### 3.1. Establishment and validation of the IPX model

The overall scheme of this study is shown in Fig. 1. Firstly, we collected 481 IRGs from the 10 gene sets [19] and performed differential gene expression (DGE) analysis to compare these genes from the two patient groups with overall survival (OS)< 1 year and OS > 3 years. Thus, we identified 37 DIRGs from the TCGA-LIHC dataset (Fig. 2A). Then, we categorized HCC patients into three clusters based on these DIRGs using unsupervised consensus clustering (Fig. 2B, Supplementary Dataset 2). Clusters A, B, and C exhibited different prognoses (Fig. 2C, *P* < 0.001), indicating that DIRGs possessed the ability to distinguish HCC patients with different survival prognoses. Furthermore, 25 of 37 genes (67.56%) were found to have a correlation with OS (P < 0.05) through the implementation of univariate Cox proportional hazards regression analysis (Fig. 2D). Finally, we performed LASSO Cox regression to screen the ultimate DIRGs for constructing the risk-score models and determined that the expression of 13 DIRGs correlated with OS in HCC patients. The modeling score of the IPX was calculated using these 13 DIRGs, according to the formula: IPX score = -0.00144*CYP3A5+0.00932*CCNB1+0.11536*ABCB6+0.00982*FLVCR1+0.09603*OSBP2+0.00326*G6PD+0.00569*RAP1GAP+0.02172*SLC7A11+0.19662*PPAT-0.00073*CYP2C9+0.01025*PLOD2+0.00098*IGSF3+0.00177*RRM2. In this formula, the gene symbol represents the normalized expression level.

Consequently, we calculated the IPX score for each patient, and classified the patients into high- and low-risk groups based on the

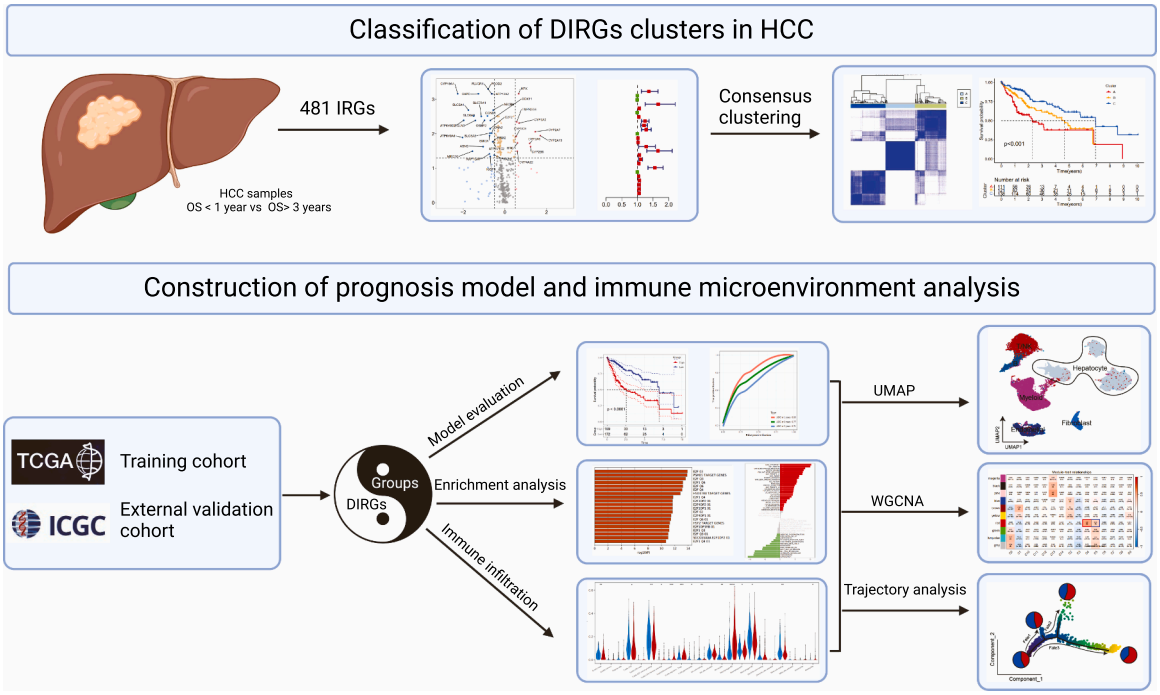


**Fig. 1.** The schematic workflow of the study.

median risk score of the training cohort. The prognosis of patients in the high-risk groups was considerably worse compared to those in the low-risk groups, as indicated by the area under the ROC curve (AUC) value of 0.828, 0.760, and 0.739 for 1-year, 3-year, and 5-year OS, respectively (Fig. 3A, all *P*-value < 0.001). The IPX model demonstrated the ability to predict long-term prognosis in liver cancer patients with relatively high accuracy. In addition, univariate Cox regression and multivariate Cox regression analysis revealed that BMI, TNM stage, and IPX were independent prognostic factors for OS prediction (Table 1). The internal testing cohort (Fig. 3B) and the external testing cohort (Fig. 3C) were further applied to verify the reliability of IPX. The two subgroups exhibited significantly different prognoses and their AUC values for 1-year, 3-year, and 5-year OS were 0.807, 0.817, and 0.774 in the internal testing cohort (Fig. 3B, all *P*-value < 0.001), and 0.734, 0.743, and 0.743 in the external testing cohort (Fig. 3C, all *P*-value < 0.001), respectively. In addition, IPX also showed as an independent prognostic factor for survival prognosis in both validation cohorts (Table 1), demonstrating the reliability and stability of the IPX model. Moreover, we conducted DGE analysis on 13 DIRGs using TCGA and ICGC datasets, and found significant statistical differences in six genes. In cancer tissue, the mRNA levels of ABCB6, CCNB1, RRM2, FLVCR1, and G6PD exhibited a notable increase compared to normal tissues (Fig. 4A-E, G-K) while CYP2C9 showed the opposite trend (Fig. 4F, L). Further, we used The Human Protein Atlas (https://www.proteinatlas.org/) to confirm the protein expression level, and found five of these genes showed a consistent trend except gene ABCB6 which was lack of immunohistochemical result (Fig. 4M-Q). Meanwhile, we employed RT-qPCR analysis to compare differences in gene expression levels between 12 tumors and adjacent tissues from HCC samples. The results confirmed the changes of gene expression in six genes described above (Fig. 4R-W).

### 3.2. Building and validating a nomogram to predict prognosis

To offer an adaptable clinical approach for evaluating the risk of patients, we created a nomogram (combined model) using independent prognostic factors including TNM stage and IPX, for predicting the 1-year, 3-year, and 5-year OS based on the TCGA dataset (Supplementary Fig. 1A). The calibration curve demonstrated a strong correlation between the predicted survival probabilities for 1, 3, and 5 years and the actual survival probability (Supplementary Fig. 1B). Additionally, the nomogram had a C-index value of 0.746 (0.709–0.784), which was higher than that from only IPX (0.727, [0.681–0.773]) or only TNM stage (0.664, [0.589–0.739]) (Table 2). The prognosis of patients in the high-risk subgroup was considerably worse compared with those in the low-risk subgroup, as indicated by the AUCs for 1-year, 3-year, and 5-year OS rates, which were 0.83, 0.73, and 0.71, respectively (Fig. 3D). In addition, the two subgroups exhibited significantly different prognosis and the AUCs for 1-year, 3-year, and 5-year OS were 0.82, 0.86, and 0.78 in the internal testing cohort (Fig. 3E, all *P*-value < 0.001), and 0.85, 0.78, and 0.78 in the external testing cohort (Fig. 3F, all *P*-value < 0.001), respectively. Therefore, the AUCs of the nomogram indicated consistent and robust discriminative power, and the Kaplan-Meier analysis revealed that patients in the high-risk subgroup had a considerably worse prognosis compared with those in the low-risk subgroup.

To assess the performance of the combined model in comparison to other gene signatures, we systematically reviewed 37 previously published gene signatures and conducted a thorough comparative analysis with IPX. Notably, IPX exhibited superior performance, demonstrating the highest efficacy among these published gene signatures (Supplementary Fig. 2). It ranked as the top one in the 1-year AUC and displayed relatively higher AUC values at 3 years compared with other methods. Concordantly, the C index of IPX-based nomogram ranked highest among recently published nomograms. Comprehensive details regarding previously published gene models for HCC prognosis are summarized in Supplementary Dataset 3.

### 3.3. Heterogeneity of genetic phenotype of IPX-defined subgroups based on bulk transcriptomic data

To clarify the molecular mechanisms underlying the predictive scoring model's ability, we applied WGCNA to identify modules associated with IPX and observed nine IPX-related modules with excluded gray module (Supplementary Fig. 3A). Genes in the blue module displayed a robust positive correlation with the high-risk group and showed enrichment in the cell cycle signaling pathway (Supplementary Fig. 3B) and the E2Fs family (Supplementary Fig. 3C). To validate our

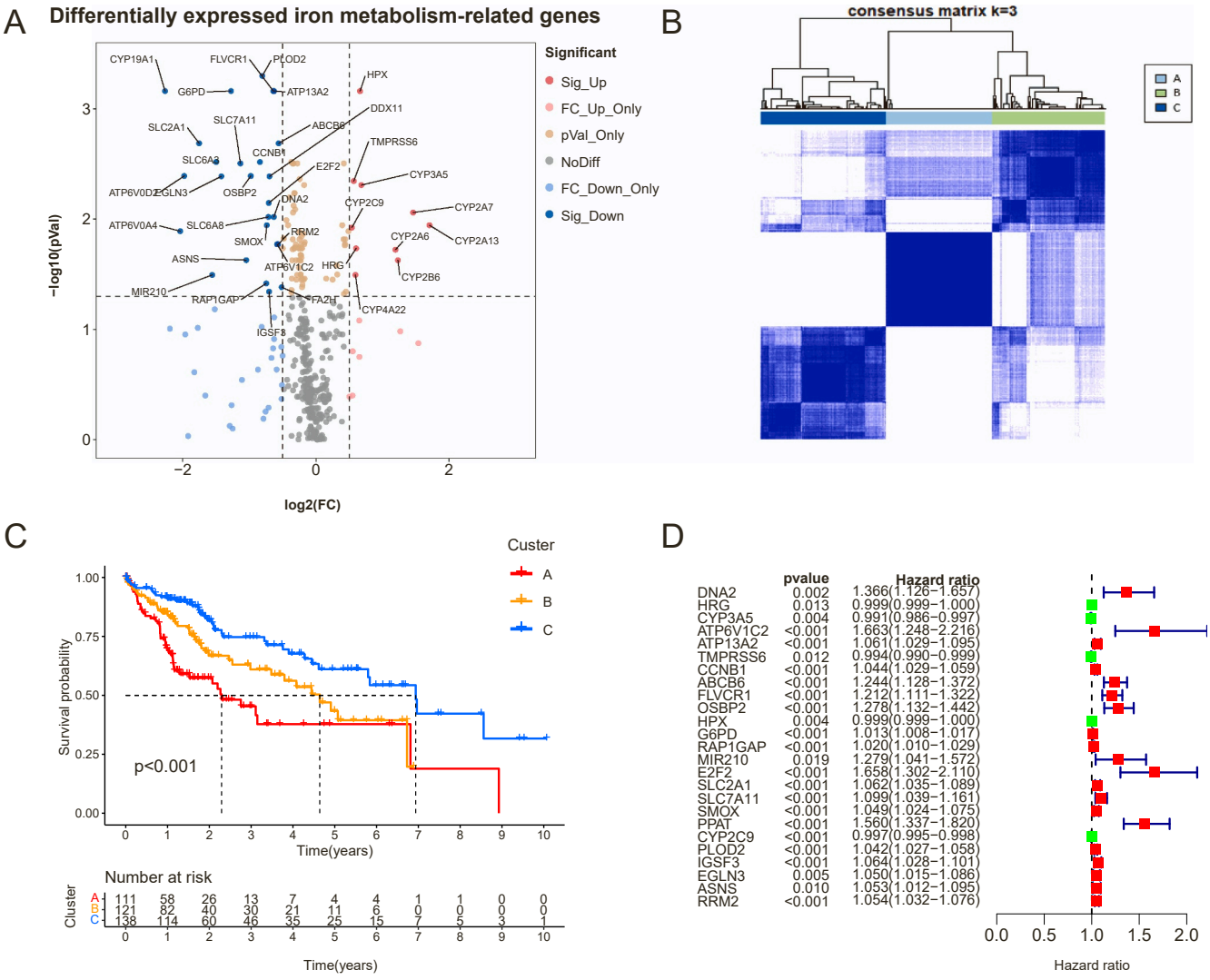


**Fig. 2.** Prognostic DIRGs could classify patients with HCC into three clusters with distinct prognostic outcomes. (A) Volcano plot showing DIRGs based on TCGA training cohort. (B) Heatmap showing similarity matrix of patients in TCGA training cohort derived from unsupervised consensus clustering. (C) Kaplan-Meier plot showing survival difference within three clusters. (D) Forest plot showing prognostic DIRGs associated with OS.

hypothesis, we conducted additional enrichment analysis using GSVA with the Hallmark database as the reference (https://www.gsea-msigdb.org/gsea/msigdb/collections.jsp). Consistently, we identified the cell cycle pathway and E2Fs pathway as two of the top five enriched pathways (Supplementary Fig. 3D). Besides, the molecular mechanisms were validated in the ICGA dataset (Supplementary Fig. 4). Moreover, the E2Fs activities were notably elevated in the IPX-high group compared with those in the IPX-low group, as demonstrated in both the TCGA dataset (Supplementary Fig. 5A) and the ICGC dataset (Supplementary Fig. 5B). Therefore, the poor survival rate of the IPX-high group may be attributed to the activation of cell cycle and E2Fs pathways in malignant cells. Ferritin proteins have been reported to be involved in maintaining the cell cycle and genomic stability since imbalances in iron metabolism can lead to genomic instability and generation of DNA repair defects, which favors carcinogenesis [13,15]. Moreover. TP53 mutation also mediates metabolic alteration to promote tumor progression [11]. We conducted a more in-depth examination of genomic alterations to pinpoint genetic regulators and displayed the top 20 genes having the highest mutation frequencies in individuals diagnosed with HCC in Supplementary Fig. 6A. Notably, TP53 mutations were considerably more common in IPX-high group (40%) than IPX-low group (15%, Supplementary Fig. 6B).

### *3.4. Application of IPX model in malignant cells based on single-cell transcriptomic data*

To uncover the underlying mechanisms of malignant cells behind the IPX model, we performed single-cell integration on hepatocellular carcinoma tissues obtained from distinct anatomical locations of 10 patients, including normal liver tissue, primary tumor (PT), portal vein tumor thrombus (PVTT) patients, metastatic lymph node (MLN). We then identified six major cell types (Fig. 5B) with canonical markers (Fig. 5A), including TNK cells, B cells, hepatocytes, myeloid, endothelial, and fibroblast. Then, the hepatocytes were divided into 15 clusters (Fig. 5C). Interestingly, we observed a notable enrichment of C4 and C5 in cancer tissues, while they were absent in normal tissues (Fig. 5D). We identified 20782 hepatocytes and applied the IPX model to assign a risk score to each hepatic cell. We found higher risk scores were positively correlated with HCC stages (Fig. 5E) and HCC metastasis progression (Fig. 5F). Besides, we found C4 and C5 exhibited the highest risk scores among hepatocyte clusters (Fig. 5G). The result suggested that C4 and C5 may constitute pivotal malignant cell subpopulations underlying the capability of the IPX model. To investigate the potential genes network of C4 and C5, we conducted a WGCNA analysis. Surprisingly, the red module was positively correlated with C4 and C5 (Figure 5H). The genes

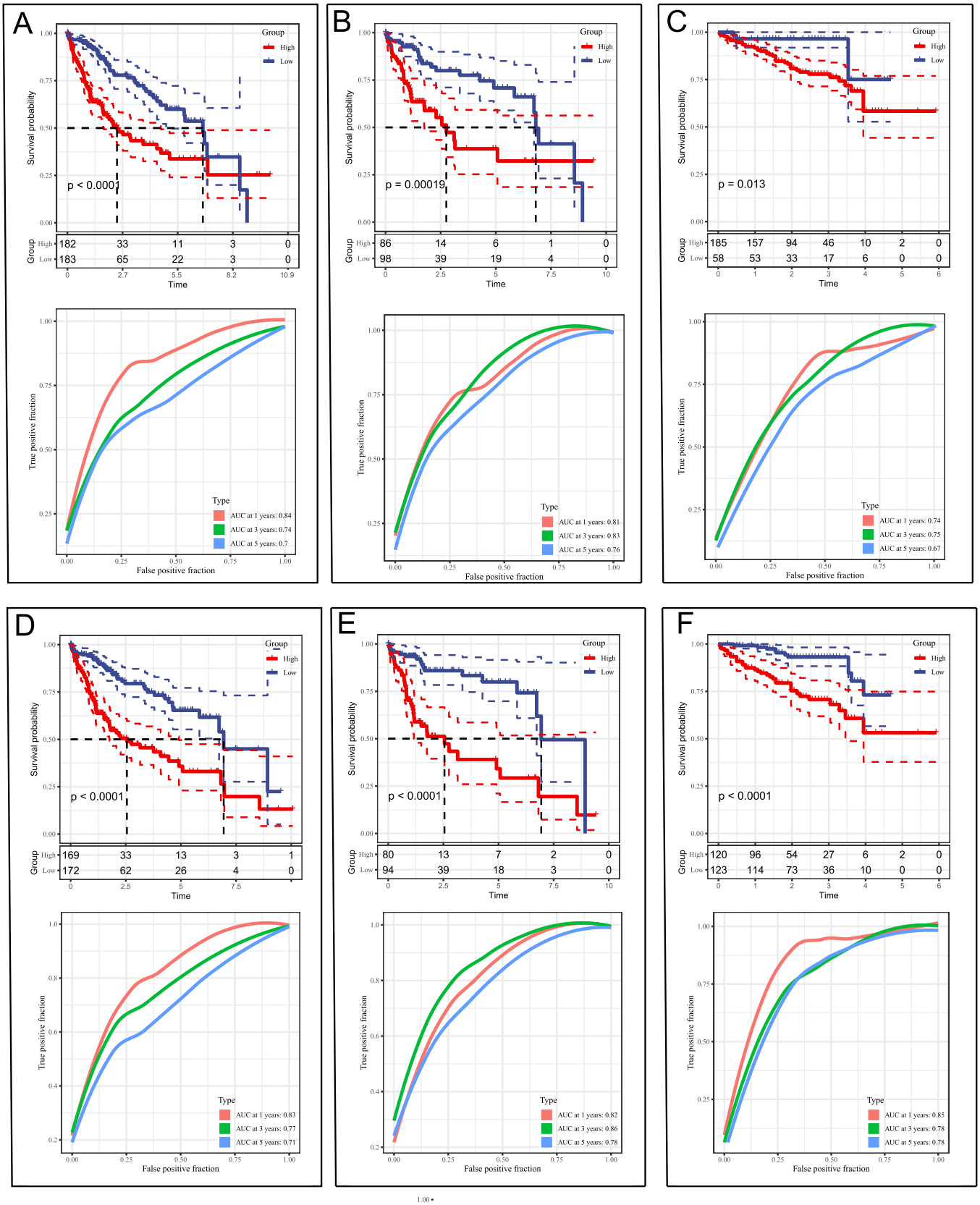


**Fig. 3.** Performance of the predictive score model (A-C) and the nomogram (D-F). (A, D) Kaplan-Meier survival curves for comparison of the overall survival rates between patients in the low-risk group and the high-risk group for the training cohort (upper), The 1,3,5-year ROC curve of predictive score model for the training cohort (lower, all *P*-value < 0.001). (B, E) Comparison of the overall survival rates between patients in the low-risk group and the high-risk group for internal testing cohort (upper), The 1,3,5-year ROC curve of predictive score model for internal testing cohort (lower, all *P*-value < 0.001). (C, F) Comparison of the overall survival rates between patients in the low-risk group and the high-risk group for external testing cohort (upper), The 1,3,5-year ROC curve of predictive score model the external testing cohort (lower, all *P*-value < 0.001).

**Table 1**
Univariate and multivariate analysis the association between clinicopathological variables and overall survival.

| Characters | Training cohort | | | | Internal testing cohort | | | | External testing cohort | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Univariate analysis | | Multivariate analysis | | Univariate analysis | | Multivariate analysis | | Univariate analysis | | Multivariate analysis | |
| | HR 95%CI | *P* | HR 95%CI | *P* | HR 95%CI | *P* | HR 95%CI | *P* | HR 95%CI | *P* | HR 95%CI | *P* |
| Age | 1.012 (0.989-1.035) | 0.322 | | | 1.003 (0.971-1.036) | 0.848 | 1.897 (1.045-3.443) | 0.035 | 1.001 (0.971-1.032) | 0.027 | 0.387 (0.203-0.737) | 0.004 |
| Gender | 0.624 (0.356-1.093) | 0.099 | | | 0.531 (0.232-1.215) | 0.134 | | | 0.497 (0.267-0.923) | 0.929 | | |
| Grade | 1.232 (0.836-1.817) | 0.292 | | | 1.166 (0.675-2.014) | 0.583 | | | | | | |
| TNM stage | 1.865 (1.293-2.690) | < 0.001 | 1.547 (1.028-2.328) | 0.036 | 2.021 (1.225-3.335) | 0.006 | | | 2.026 (1.407-2.917) | < 0.001 | | |
| Vascular invision | 1.377 (0.773-2.451) | 0.277 | | | 1.745 (0.775-3.927) | 0.179 | | | | | | |
| BMI | 1.045 (1.012-1.078) | 0.006 | 1.044 (1.011-1.078) | 0.008 | 1.056 (1.016-1.097) | 0.006 | 1.897 (1.045-3.443) | 0.035 | | | | |
| Prior malignancy | | | | | | | | | 1.754 (0.775-3.971) | 0.178 | 1.883 (1.302-2.722) | < 0.001 |
| IPX | 3.332 (2.166-5.125) | < 0.001 | 3.125 (1.995-4.896) | < 0.001 | 4.620 (2.220-9.613) | < 0.001 | 1.897 (1.045-3.443) | 0.035 | 6.086 (3.015-12.281) | < 0.001 | 4.754 (2.328-9.706) | < 0.001 |

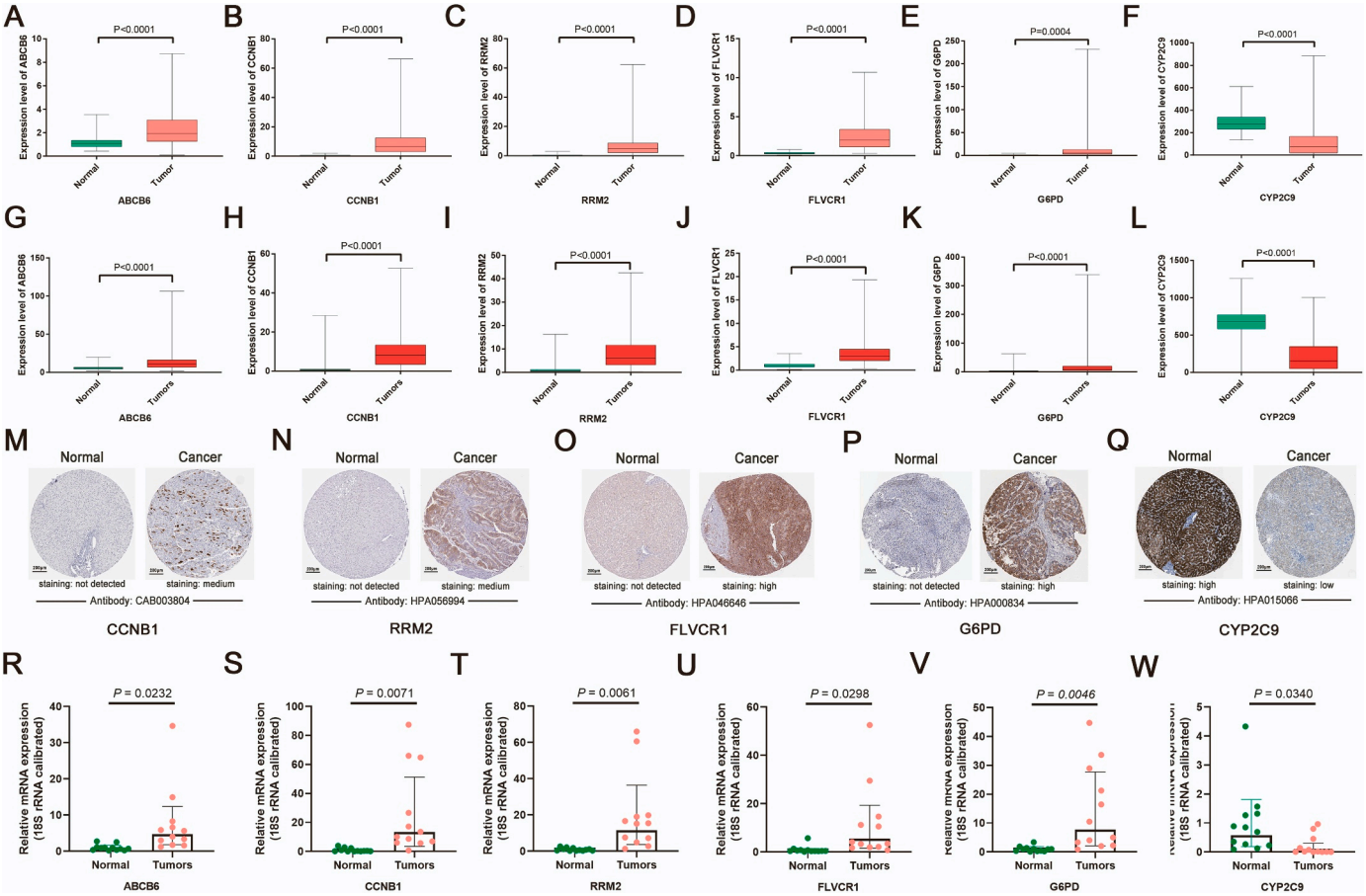


**Fig. 4.** mRNA and protein expression levels of individual genes in DIRGs in different datasets. The expression differences of 6 of the 13 DIRGs were statistically significant, and the expression trends were consistent in the TCGA dataset (A-F), ICGC dataset (G-L), The Human Protein Atlas (M-Q), and 12 HCC clinical samples (using RT-qPCR) (R-W).

**Table 2**
The C-index of Signature, Stage and Combined model.

| Biomarker | C-index | 95%CI Lower | 95%CI Higher | P-value |
|---|---|---|---|---|
| Signature | 0.726819 | 0.680894 | 0.772744 | 3.67E-22 |
| Stage | 0.664469 | 0.589597 | 0.73934 | 1.67E-05 |
| Combined model | 0.746342 | 0.708523 | 0.784161 | 4.65E-12 |

belonging to the red module was enriched in chromosome segregation, nuclear division, and mitotic nuclear division (Supplementary Figure 7). We then analyzed the interaction strength of genes in the red module and found the interaction strength of the NUF2, CKS1B, HMGB2, CENPF, and CDC20 ranked as the top five genes (Fig. 5I). We further explored the association between the genes in the red module with C4 and C5, and genes such as NUF2, and CKS1B were significantly associated with C4 and C5, respectively (Fig. 5J, Fig. 5K).

To explore the potential driver transcription factors for C4 and C5, we conducted a thorough comparison of the atlas-wide concordance in regulon specificity score (RSS) scores for each regulon pair using the Connection Specificity Index (CSI) using SCENIC [38]. Notably, these TFs are structured into 7 primary modules (Fig. 5L). Interestingly, the module7 occupies C4 and C5 region with several representative regulators, such as FOXM1, E2F1, E2F7 (Supplementary Figure 8, Fig. 5L), but other modules have minimal impact on C4 and C5 (Supplementary

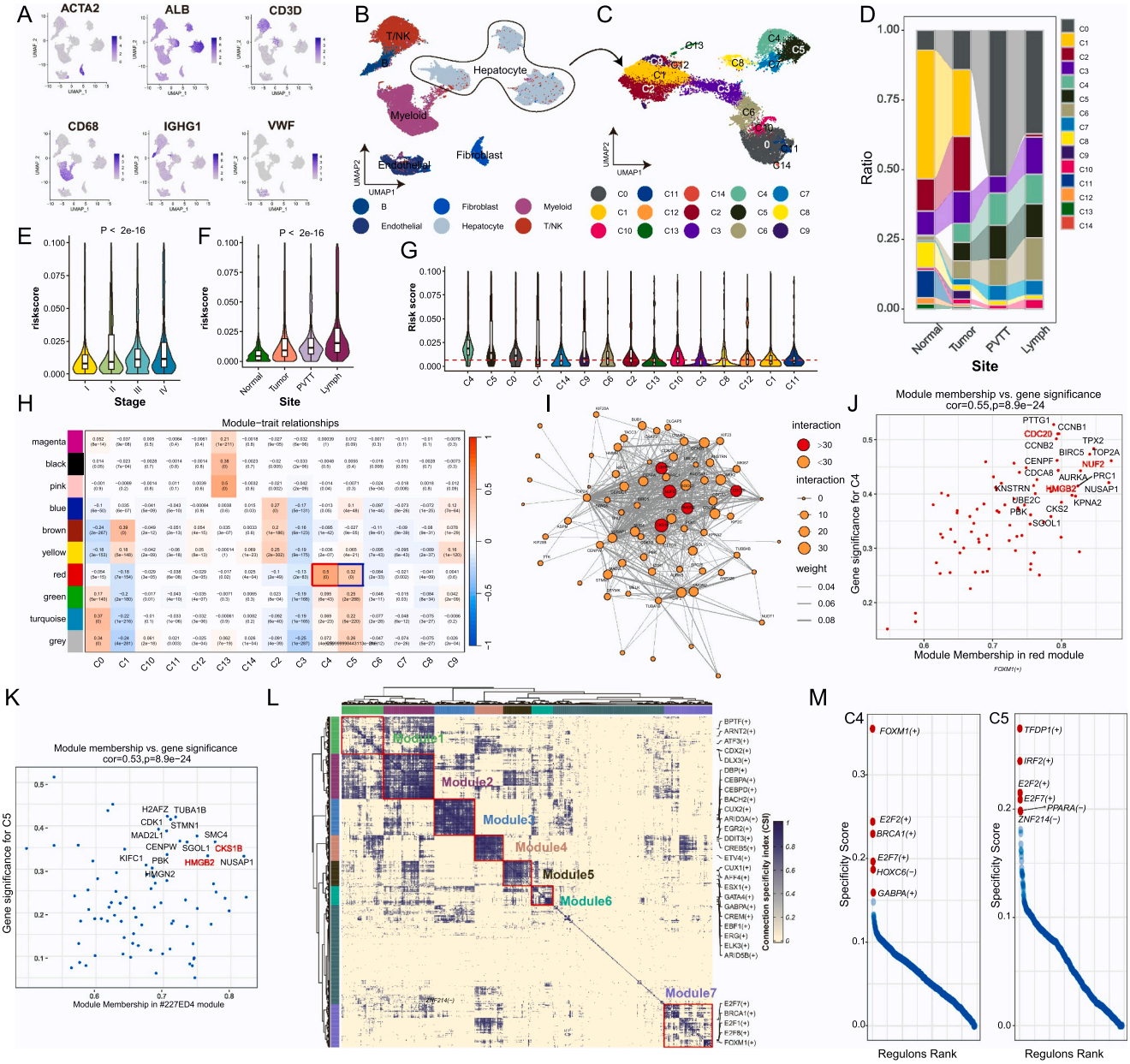


**Fig. 5.** Detailed characterization of malignant cells based on single-cell RNA sequencing data. (A) UMAP visualization of signature genes that were used to annotat major cell types. (B) UMAP visualization of integrated 6 major cell types. (C) UMAP visualization of hepatocytes that were clustered into 15 sub-clusters. (D) Stacked barplot showing the proportion of hepatocytes cell types in each site. (E) Violin plot showing the risk scores of hepatocytes across different HCC stages. (F) Violin plot showing the risk scores of hepatocytes across different sites. (G) Violin plot showing the risk scores of hepatocytes across different hepatocyte sub-clusters. (H) Module–trait relationships in all hepatocyte sub-clusters. (I) Relatedness network of genes in the red module. (J) The red module membership and gene significance for the C4 sub-cluster. (K) The red module membership and gene significance for the C5 sub-cluster. (L) Heatmap showing the regulon modules using regulon connection specificity index (CSI) matrix with representative regulators. (M) Rank plot showing the transcription factors of C4 (Left) and C5 (Right) sub-cluster based on regulon specificity score (RSS).

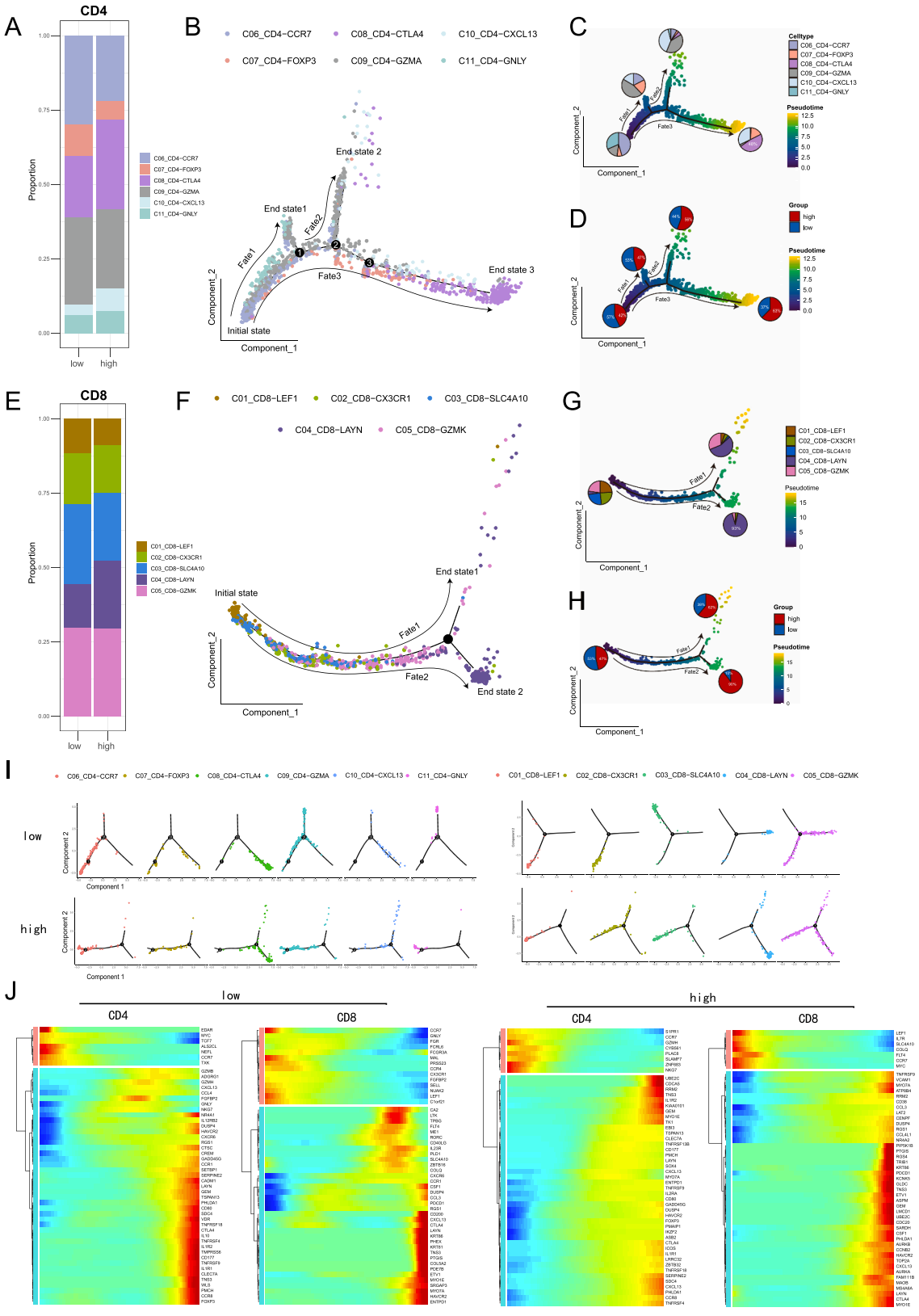


**Fig. 6.** Discriminating T-cell populations based on single-cell RNA sequencing data. (A) Stacked barplot showing the proportion of $CD4^+$ T cells between the high-score and low-score populations. (B) The trajectory of $CD8^+$ T cells inferred by Monocle2. (C) Pseudotime analysis showed six $CD8^+$ T subpopulations on each branch. (D) Pseudotime analysis showed high-score and low-score $CD8^+$ T cells on each branch. (E) Stacked barplot showing the proportion of $CD8^+$ T cells between the high-score and low-score populations. (F) The trajectory of $CD4^+$ T cells inferred by Monocle2. (G) Pseudotime analysis showed six $CD4^+$ T subpopulations on each branch. (H) Pseudotime analysis showed high-score and low-score $CD4^+$ T cells on each branch. (I) Pseudotime plot showing distinct developmental trajectories of $CD4^+$ T cell and $CD8^+$ T cell for high-score and low-score populations. (J) Heatmap for the dynamic top genes along pseudotime for high-score and low-score populations for $CD4^+$ and $CD8^+$ T cells.

Figure 8). We further identified cell cluster-specific transcription factors across all 15 clusters (Supplementary Figure 9). We identified E2F2, E2F7, FOXM1, TFDP1 were the most specific regulons for C4 and C5 (Fig. 5M). FOXM1, E2F7, E2F2, and TFDP1 genes are known for their involvement in cell cycle control and are crucial for normal cell division [39–41]. Additionally, they have been implicated in cancer progression [39,40]. These results suggested that FOXM1, E2F7, E2F2, and TFDP1 genes were the driver transcription factors in C4 and C5, thereby contributing to the adverse survival outcomes observed in the IPX-high group.

### 3.5. The TME heterogeneity of IPX-defined subgroups based on bulk transcriptomic data

To explore variances of immune microenvironments between the IPX-high and IPX-low groups, we utilized the CIBERSORT algorithm to estimate the proportions of immune cells using RNA expression data from both the TCGA and ICGC datasets. The proportion of Treg cells exhibited significantly higher in the IPX-high group compared with that of the IPX-low group in both the TCGA cohort (Supplementary Figure 10 A) and ICGC cohort (Supplementary Figure 10B). To confirm our hypothesis, we employed ssGSEA to compute the scores for various immune cell types and observed a higher abundance of Tregs in the IPX-high group compared with the IPX-low group in both the TCGA (Supplementary Figure 10 C) and ICGC cohorts (Supplementary Figure 10D). The abundance of Tregs reflects the immunosuppressive status of the TME [42–44]. We further evaluated the immunosuppressive status of patients with HCC using several well-known signatures [26–28]. Indeed, we found that the IPX signature was significantly positively associated with TGFβ extracellular matrix pathway, T cell inhibitors, and Treg transcriptional signature (Supplementary Figure 11). Therefore, these findings suggested a possible mechanistic connection between the diverse immune infiltration and IPX scores observed in HCC patients. Specifically, the IPX-high group appeared to display a more pronounced immune escape phenotype, possibly due to the elevation of activated Tregs.

In addition, we utilized ImmuneCellAI to forecast the response to ICB therapy and revealed a notable enrichment of 'no response' in the IPX-high group (Supplementary Figure 12 A). This observation was consistently corroborated by the TIDE algorithm [30] (Supplementary Figure 12B). On the other hand, a patient's Immune Prognostic Score (IPS) can be derived in an unbiased manner using machine learning by considering the four major categories of genes that determine immunogenicity including effector cells, immunosuppressive cells, MHC molecules, and immunomodulators. A higher IPS score indicates higher level of immunogenicity [45]. The IPS scores of HCC patients in the high IPX group were significantly lower for ctla4_neg_pd1_neg, ctla4_neg_pd1_pos, ctla4_pos_pd1_neg, and ctla4_pos_pd1_pos categories (Supplementary Figure 12 C). The high IPX group was associated with a reduced likelihood of benefiting from ICI therapy. Therefore, individuals with higher IPX scores could potentially benefit less from immunotherapy when compared with those belonging to the IPX-low group.

### 3.6. Application of IPX model in infiltrating T cells based on single-cell transcriptomic data

To confirm our observation of TME dynamics from the deconvolution of bulk RNA-seq data, we applied the IPX scoring model to single-cell transcriptomic data of infiltrating T cells in HCC patients from the previous study (GSE98638) [7]. The 4070 T cells from the single-cell dataset were classified into high-score (n = 2103) and low-score (n = 1967) subgroups based on our scoring model. We observed obvious differences in the composition of $CD4^+$ T cells and $CD8^+$ T cells between the two subgroups (Figs. 6A, 6E). The low-score subgroup exhibited higher proportions of naive $CD4^+$ (C06_CD4 −CCR7) and $CD8^+$ T cells (C01_CD8 −LEF1) compared with the high-score subgroup. In contrast, the high-score subgroup displayed elevated proportions of exhausted $CD8^+$ T cells (CC04_CD8 −LAYN), exhausted $CD4^+$ T cells (C10_CD4 −CXCL13), and tumor-infiltrating Tregs (C08_CD4 −CTLA4) compared to the low-score subgroup. In addition, we noticed a significant increase in the expression of exhaustion marker genes in the high-score subgroup compared with those in the low-score subgroup (Supplementary Figure 13).

Then, we performed a pseudotime analysis to examine the potential transitions and differentiation trajectories among $CD4^+$ T cells and $CD8^+$ T cells. $CD4^+$ T cells exhibited three major fate branches (Fig. 6B). We found that tumor-infiltrating Tregs (C08_CD4 −CTLA4) exhibited depletion in the initial state and constituted most of the fate 3 branch, indicating that the fate 3 branch represented the differentiation trajectory of Treg phenotype (Fig. 6C). Interestingly, the proportion of high IPX-score cells increased from an initial 42% in the initial state to 63% in the end fate 3 branch (Fig. 6D). On the other hand, $CD8^+$ T cells exhibited two major fate branches (Fig. 6F) and the exhaustion-associated $CD8^+$ T cells (CC04_CD8 −LAYN) constituted the predominant portion (accounting for 93%) of the fate 2 branch (Fig. 6G), suggesting that the fate 2 branch represented the differentiation trajectory of exhausted phenotype. Interestingly, the fate 2 branch was dominated by high IPX-score cells, constituting up to 90% of the sub-population (Fig. 6H). These results suggested that Treg and exhausted $CD8^+$ T cells might contribute to the unfavorable survival outcomes identified in the IPX-high group. We also conducted pseudotime analyses using Monocle for a deeper investigation into the developmental trajectories of low-score and high-score cells. This analysis enabled us to elucidate the dynamic evolution of cell states and immune cell transitions in both subgroups of $CD4^+$ and $CD8^+$ T cells [46]. Interestingly, high-score and low-score clusters displayed distinct developmental paths and distributions of cell types in both $CD4^+$ and $CD8^+$ T cells (Fig. 6I). In the low-score subgroup, the naïve $CD4^+$ T cells (C06_CD4-CCR7) were predominantly found at the initial stage of the pseudotime trajectory while the cytotoxic $CD4^+$ T cells (C11_CD4-GNLY) mostly localized at a distal branch. In contrast, these two cell types showed substantial overlap and primarily occupied intermediate states within the high-score subgroup. In low-score subgroup, the effector $CD8^+$ T cells (C02_CD8-CX3CR1) mostly resided along one trajectory path and overlapped extensively with naive $CD8^+$ T cells (C01_CD8 −LEF1 cluster), but the exhausted $CD8^+$ T cells (C04_CD8-LAYN) mainly localized at another distal branch. Unlike the low-score subgroup, the effector $CD8^+$ T cells (C02_CD8-CX3CR1) and exhausted $CD8^+$ T cells (C04_CD8-LAYN) bifurcated into two branches in the high-score subgroup.

Further, we delved into the transcriptional alterations across the pseudotime trajectory and pinpointed genes that underwent significant changes along the differentiation processes. We found the top 50 differentially expressed genes (DEGs) were associated with cell cycle and immune inflammation (Fig. 6J). We performed branch-dependent DGE analysis and found that distinct DEGs were associated with cell cycle and immune inflammation in both $CD4^+$ T cells and $CD8^+$ T cells (Supplementary Figure 14). T cells in the high-score subgroup highly expressed genes enriched in pathways related to the cell cycle, including transitions in the mitotic cell cycle phases, nuclear division, and mitotic nuclear division (Fig. 7A). In contrast, the top up-regulated genes in the low-score subgroup were notably enriched in several critical biological processes, including the regulation of T cell activation, cell-cell adhesion involving leukocytes, and immune responses mediated by leukocytes (Fig. 7B). We further performed ssGSEA to investigate the different pathway activities between two cell subpopulations. We found that the gene set scores were notably elevated in the high-score subgroup compared to the low-score cluster for categories such as "Cell cycle," "Treg," and "T exhaustion" (Fig. 7C). Conversely, the gene set score for "Positive regulation of cell activation" was higher in the low-score subgroup compared with that in the high-score subgroup (Fig. 7C). Therefore, our results indicated that cell-cycle perturbation and TME

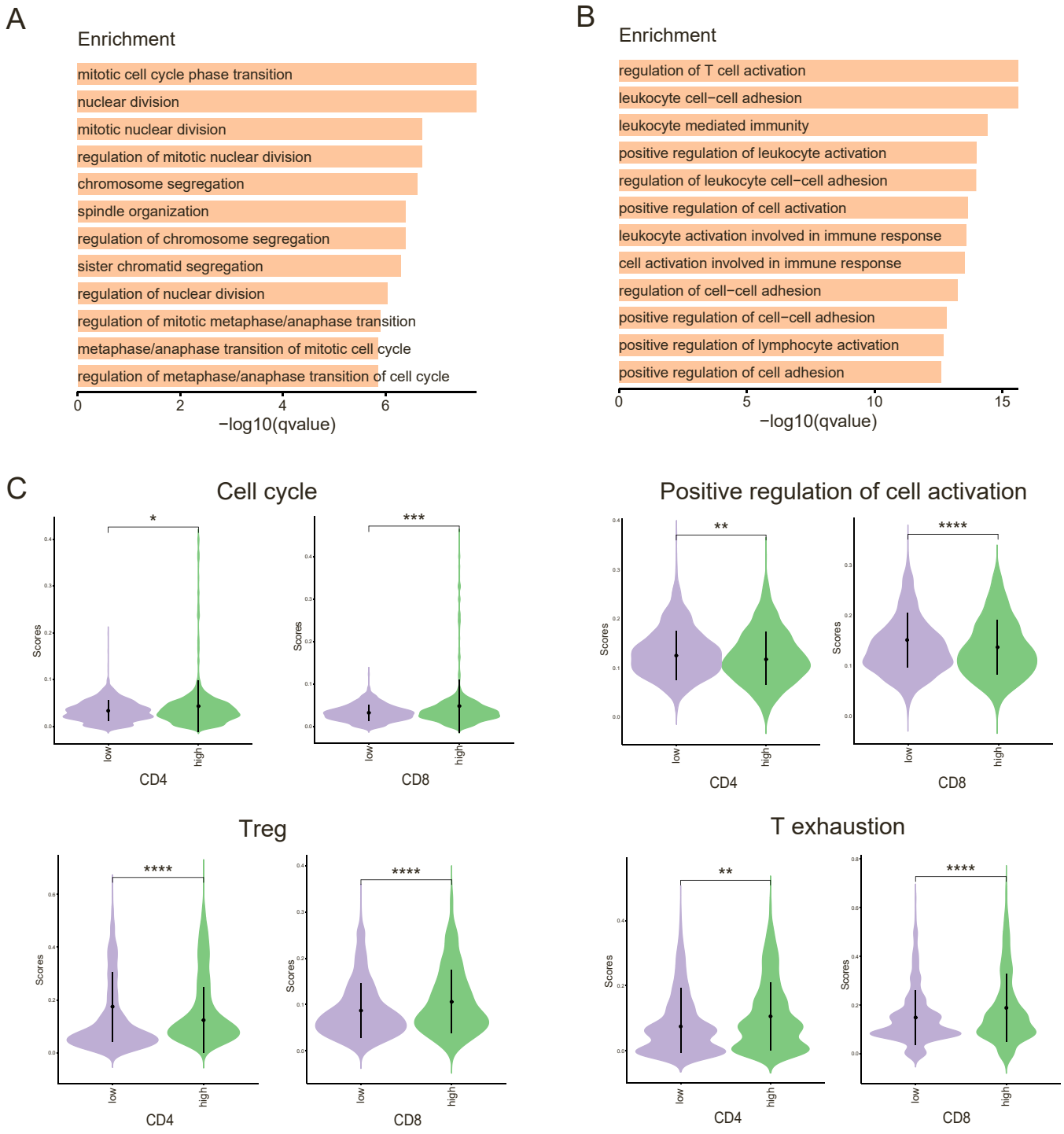


**Fig. 7.** Enrichment analysis showed distinct biological processes between high-score and low-score clusters. (A) Biological pathways that are associated with significantly branch-dependent genes for high-score. (B) Biological pathways that are associated with significantly branch-dependent genes for low-score. (C) Single sample gene set enrichment analysis results for comparing the expression of the dynamic genes during branch evolution for the "Cell cycle" term, the "Positive regulation of cell activation" term, the "Treg" term, and the "T exhaustion" term. * *P*-value $< 0.05$, ** *P*-value $< 0.01$, *** *P*-value $< 0.001$ and **** *P*-value $< 0.0001$.

reprogramming appeared to influence immunosuppressive features exhibited by T cells in the high-score subpopulation. Consistently, we investigated the iron metabolic score across human cancers using the ssGSEA algorithm and found that iron metabolism scores were significantly elevated in cancer than those in the normal group among various tumor types (n = 13), including BLCA, UCEC, HNSC, PRAD, COAD, LUSC, LIHC, BRCA, KICH, THCA, LUAD, CHOL and ESCA (Supplementary Figure 15), expanding the broad application value of the score model.

## 4. Discussion

Most HCC staging systems are based on tumor mutation burden and disease staging, with stratification based on prognosis [47]. However,

due to the complexity of liver cancer's response to immunotherapy, it is difficult to explain the adverse biological characteristics that affect treatment and survival responses. The iron metabolism in the tumor site is different from the adjacent site in HCC, and IRGs contribute significantly to the progression of HCC [48]. For the proliferation and replication of DNA, HCC necessitates a substantial concentration of iron [49]. On the other hand, high levels of cellular iron would alter the T cell infiltrating microenvironment, then promote the proliferation, invasion, and migration of tumors [50]. Thus, IRGs could potentially serve as a promising target for the treatment of HCC.

In the current study, we developed and verified a prognostic scoring model by utilizing IRGs that have a crucial function in malignant cells and the TME. The scoring model effectively categorized HCC patients into distinct risk groups since the prognosis for patients in the high-risk group was significantly poorer in comparison to those in the low-risk group. Then, a nomogram was constructed by combining the scoring model and clinicopathological risk feature, offering a comprehensive model suitable for clinical implementation. Several previous studies have built prognostic signatures using IRGs [51,52], but there has been limited exploration of the impact of cellular heterogeneity and microenvironment variations on the prognostic gene models in HCC. Our study analyzed differences in cell populations at different tumor stages and different tumor sites, and found that cellular and molecular heterogeneity can influence patient prognosis. We found several representative regulators were specifically activated in specific malignant cell subpopulations to contribute to the adverse survival outcomes in the IPX-high subgroup. In addition, we observed that cell-cycle perturbation and TME reprogramming appeared to influence the immunosuppressive features of T cells in the IPX-high subgroup. Previous predictive models for metabolic reprogramming of HCC have identified genes that are consistent with the predictions of this study, such as FLVCR1 [51,52]. And it has been validated that knockout of FLVCR1 could modulate the proliferation, migration, and invasion of HCC [53]. The further functional study is needed to investigate whether IRGs can influence tumor progression and immune cell infiltration in mouse models. It has been confirmed that ferroptosis has opened up an emerging strategy for anti-tumor treatment in systemic tumor therapy, radiation therapy, and immunotherapy [20]. Multiple studies have confirmed that targeting SLC7A11 can reverse drug resistance during the treatment of malignant tumors. Data from PharmSnap shows that one anti-xCT antibody drug (Agilvax) targeting SLC7A11 is currently in preclinical use for cancer treatment. At present, SLC7A11 has been widely endowed with chemotherapy resistance for various types of cancer [54]. Therefore, these modeling genes are expected to become special therapeutic targets in HCC treatments.

We further attempted to uncover key determinants behind the ability of this model to identify high-risk populations. The single-cell RNA-seq analysis showed that several driver transcription factors activated in special malignant subsets belonging to E2Fs family, such as E2F7, E2F2, and TFDP1. It is reasonable that the E2F family was identified in a high–risk subgroup since E2Fs are a prevalent and crucial group of transcriptional regulators to control cell cycle and genomic integrity, and cope with replication pressure and DNA damage. Moreover, the dysregulated functions of E2Fs were associated with unfavorable prognoses in human cancers [55]. The bulk RNA-seq analysis revealed that poor prognosis may be attributed to dysregulation of the cell cycle and E2Fs pathway that was induced by TP53 mutations. Besides, iron overload in HCC cells was likely to generate ROS that promotes genomic instability and generation of DNA repair defects, especially the TP53 mutation and p53 acetylation [56,57], resulting in the abnormal expression of E2Fs that induce malignant transformation [58].

On the other hand, the TME consists of various cellular and non-cellular elements that play vital roles in fostering or preventing tumor growth. Intertumoral immune cell heterogeneity may also be a key determinant for the distinct clinical outcomes. In the HCC microenvironment, iron is usually enriched and released from macrophages and neutrophils, which can support tumor progression via multiple iron-dependent pathways [59]. The cellular function of iron metabolism is complex and even conflicts in the microenvironment. For instance, Agoro et. al. stated that an iron-rich diet in mouse models promotes the expression of M2 markers and inhibits the M1 phenotype in the liver [60]. In contrast, Handa et. al. reported an opposite conclusion that cell iron loading triggers the expression of M1 markers but reduces M2 polarization in macrophages [61]. To unveil the key determinants in the microenvironment of our scoring model, we deconvoluted bulk transcriptomes using CIBERSORT and found that the Treg abundance of the high-risk subgroup increased notably than that of the low-risk subgroup in both the TCGA and ICGC datasets. We extended the scoring model to single-cell transcriptomes and confirmed that Treg cells exhibited a higher ratio in the high-score cluster than in the low-score cluster. And cell cycle showed higher enrichment in high-score cluster, providing an explanation that IRGs may induce cell-cycle perturbation to influence the function and composition of immune cells in TME. The construction and calibration of this predictive model were based on multiple sets of different datasets, providing a relatively reliable predictability. Unfortunately, this predictive model lacks experimental validation and evidence from multi-omics studies on the same set of samples. Thus, more functional experiments and multi-omics studies from the same group of individuals should be carried out in the future.

## Funding

Beijing Nova Program [Z211100002121039]; National Natural Science Foundation of China [32170656, 82073057]; Research start-up funding, Peking University Third Hospital [BYSYYZD2021001].

## CRediT authorship contribution statement

**Zhipeng Zhu:** Conceptualization, Methodology, Formal analysis, Writing – original draft. **Huang Cao:** Validation, Project administration. **Hongyu Yan**: Data curation. **Hanzhi Liu**: Resources. **Zaifa Hong**: Data curation. **Anran Sun**: Visualization, Supervision, Writing - Original Draft, Writing - Review & Editing. **Tong Liu:** Writing – review & editing. **Fengbiao Mao**: Writing – review & editing.

## Declaration of Competing Interest

The authors report no conflict of interest.

## Acknowledgements

Not applicable.

## Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2024.01.022.

## References

[1] 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.
[2] Donne R, Lujambio A. The liver cancer immune microenvironment: therapeutic implications for hepatocellular carcinoma. Hepatology 2023;77:1773–96.
[3] Yang F, Hilakivi-Clarke L, Shaha A, et al. Metabolic reprogramming and its clinical implication for liver cancer. Hepatology 2023;78:1602–24.
[4] Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA Cancer J Clin 2023;73:17–48.
[5] Wang L, Chen Y, Chen R, Mao F, Sun Z, Liu X. Risk modeling of single-cell transcriptomes reveals the heterogeneity of immune infiltration in hepatocellular carcinoma. J Biol Chem 2023;299:102948.
[6] Zhang Q, He Y, Luo N, et al. Landscape and dynamics of single immune cells in hepatocellular carcinoma. Cell 2019;179:829–45. e820.

[7] Zheng C, Zheng L, Yoo JK, et al. Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing. Cell 2017;169:1342–56. e1316.
[8] Toyokuni S. Role of iron in carcinogenesis: cancer as a ferrotoxic disease. Cancer Sci 2009;100:9–16.
[9] Andrews NC. Disorders of iron metabolism. N Engl J Med 1999;341:1986–95.
[10] Hsu MY, Mina E, Roetto A, Porporato PE. Iron: an essential element of cancer metabolism. Cells 2020;9.
[11] Luo K, Qian Z, Jiang Y, et al. Characterization of the metabolic alteration-modulated tumor microenvironment mediated by TP53 mutation and hypoxia. Comput Biol Med 2023;163:107078.
[12] Mehta KJ, Sharp PA. Iron elevates mesenchymal and metastatic biomarkers in HepG2 cells. Sci Rep 2020;10:21926.
[13] Paul VD, Lill R. Biogenesis of cytosolic and nuclear iron-sulfur proteins and their role in genome stability. Biochim Biophys Acta 2015;1853:1528–39.
[14] Shen J, Sheng X, Chang Z, et al. Iron metabolism regulates p53 signaling through direct heme-p53 interaction and modulation of p53 localization, stability, and function. Cell Rep 2014;7:180–93.
[15] Zhang C. Essential functions of iron-requiring proteins in DNA replication, repair and cell cycle control. Protein Cell 2014;5:750–60.
[16] Duan Q, Zhang H, Zheng J, Zhang L. Turning cold into hot: firing up the tumor microenvironment. Trends Cancer 2020;6:605–18.
[17] Brown RAM, Richardson KL, Kabir TD, Trinder D, Ganss R, Leedman PJ. Altered iron metabolism and impact in cancer biology, metastasis, and immunology. Front Oncol 2020;10:476.
[18] Llovet JM, Kelley RK, Villanueva A, et al. Hepatocellular carcinoma. Nat Rev Dis Prim 2021;7:6.
[19] Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 2005;102:15545–50.
[20] Li S, Lu Z, Sun R, et al. The role of SLC7A11 in cancer: friend or foe? Cancers (Basel) 2022;14.
[21] Wang L, Feng Z, Wang X, Wang X, Zhang X. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics 2010;26:136–8.
[22] Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med 1997;16:385–95.
[23] Efron B. Estimation and Accuracy after Model Selection. J Am Stat Assoc 2014;109:991–1007.
[24] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinforma 2008;9:559.
[25] Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun 2019;10:1523.
[26] Chakravarthy A, Khan L, Bensler NP, Bose P, De Carvalho DD. TGF-beta-associated extracellular matrix genes link cancer-associated fibroblasts to immune evasion and immunotherapy failure. Nat Commun 2018;9:4692.
[27] Lachenmayer A, Alsinet C, Savic R, et al. Wnt-pathway activation in two molecular classes of hepatocellular carcinoma and experimental modulation by sorafenib. Clin Cancer Res 2012;18:4997–5007.
[28] Magnuson AM, Kiner E, Ergun A, et al. Identification and validation of a tumor-infiltrating Treg transcriptional signature conserved across species and tumor types. Proc Natl Acad Sci USA 2018;115:E10672–81.
[29] Charoentong P, Finotello F, Angelova M, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. Cell Rep 2017;18:248–62.
[30] Jiang P, Gu S, Pan D, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat Med 2018;24:1550–8.
[31] Miao YR, Zhang Q, Lei Q, et al. ImmuCellAI: a unique method for comprehensive t-cell subsets abundance prediction and its application in cancer immunotherapy. Adv Sci (Weinh) 2020;7:1902880.
[32] Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res 2018;28:1747–56.
[33] Qiu X, Mao Q, Tang Y, et al. Reversed graph embedding resolves complex single-cell trajectories. Nat Methods 2017;14:979–82.
[34] Korsunsky I, Millard N, Fan J, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods 2019;16:1289–96.
[35] Hao J, Cao W, Huang J, Zou X, Han ZG. Optimal Gene Filtering for Single-Cell data (OGFSC)-a gene filtering algorithm for single-cell RNA-seq data. Bioinformatics 2019;35:2602–9.
[36] Aibar S, Gonzalez-Blas CB, Moerman T, et al. SCENIC: single-cell regulatory network inference and clustering. Nat Methods 2017;14:1083–6.
[37] Aerts S, Quan XJ, Claeys A, et al. Robust target gene discovery through transcriptome perturbations and genome-wide enhancer predictions in Drosophila uncovers a regulatory basis for sensory specification. PLoS Biol 2010;8:e1000435.
[38] Fuxman Bass JI, Diallo A, Nelson J, Soto JM, Myers CL, Walhout AJ. Using networks to measure similarity between genes: association index selection. Nat Methods 2013;10:1169–76.
[39] Kent LN, Leone G. The broken cycle: E2F dysfunction in cancer. Nat Rev Cancer 2019;19:326–38.
[40] Raychaudhuri P, Park HJ. FoxM1: a master regulator of tumor metastasis. Cancer Res 2011;71:4329–33.
[41] Adams PD, Kaelin Jr WG. Transcriptional control by E2F. Semin Cancer Biol 1995;6:99–108.
[42] Berraondo P, Sanmamed MF, Ochoa MC, et al. Cytokines in clinical cancer immunotherapy. Br J Cancer 2019;120:6–15.
[43] Furukawa A, Wisel SA, Tang Q. Impact of immune-modulatory drugs on regulatory T cell. Transplantation 2016;100:2288–300.
[44] Groth C, Hu X, Weber R, et al. Immunosuppression mediated by myeloid-derived suppressor cells (MDSCs) during tumour progression. Br J Cancer 2019;120:16–25.
[45] Givechian KB, Wnuk K, Garner C, et al. Identification of an immune gene expression signature associated with favorable clinical features in Treg-enriched patient tumor samples. NPJ Genom Med 2018;3:14.
[46] Qiu X, Hill A, Packer J, Lin D, Ma YA, Trapnell C. Single-cell mRNA quantification and differential analysis with Census. Nat Methods 2017;14:309–15.
[47] Cunha GM, Hosseini M, Furlan A, Fowler KJ. Hepatocellular carcinoma staging: differences between radiologic and pathologic systems and relevance to patient selection and outcomes in liver transplantation. AJR Am J Roentgenol 2022;218:77–86.
[48] Rah B, Farhat NM, Hamad M, Muhammad JS. JAK/STAT signaling and cellular iron metabolism in hepatocellular carcinoma: therapeutic implications. Clin Exp Med 2023.
[49] Kew MC. Hepatic iron overload and hepatocellular carcinoma. Liver Cancer 2014;3:31–40.
[50] Sacco A, Battaglia AM, Botta C, et al. Iron metabolism in the tumor microenvironment-implications for anti-cancer immune response. Cells 2021;10.
[51] Tang B, Zhu J, Li J, et al. The ferroptosis and iron-metabolism signature robustly predicts clinical diagnosis, prognosis and immune microenvironment for hepatocellular carcinoma. Cell Commun Signal 2020;18:174.
[52] Wang Z, Embaye KS, Yang Q, et al. A novel metabolism-related signature as a candidate prognostic biomarker for hepatocellular carcinoma. J Hepatocell Carcinoma 2021;8:119–32.
[53] Zhang K, Zhao Z, Yu J, Chen W, Xu Q, Chen L. LncRNA FLVCR1-AS1 acts as miR-513c sponge to modulate cancer cell proliferation, migration, and invasion in hepatocellular carcinoma. J Cell Biochem 2018;119:6045–56.
[54] Lin Y, Dong Y, Liu W, Fan X, Sun Y. Pan-cancer analyses confirmed the ferroptosis-related gene SLC7A11 as a prognostic biomarker for cancer. Int J Gen Med 2022;15:2501–13.
[55] Kent LN, Bae S, Tsai SY, et al. Dosage-dependent copy number gains in E2f1 and E2f3 drive hepatocellular carcinoma. J Clin Invest 2017;127:830–42.
[56] Engeland K. Cell cycle regulation: p53-p21-RB signaling. Cell Death Differ 2022;29:946–60.
[57] Zhou Y, Que KT, Zhang Z, et al. Iron overloaded polarizes macrophage to proinflammation phenotype through ROS/acetyl-p53 pathway. Cancer Med 2018;7:4012–22.
[58] Sherr CJ, McCormick F. The RB and p53 pathways in cancer. Cancer Cell 2002;2:103–12.
[59] Liang W, Ferrara N. Iron metabolism in the tumor microenvironment: contributions of innate immune cells. Front Immunol 2020;11:626812.
[60] Agoro R, Taleb M, Quesniaux VFJ, Mura C. Cell iron status influences macrophage polarization. PLoS One 2018;13:e0196921.
[61] Handa P, Thomas S, Morgan-Stevenson V, et al. Iron alters macrophage polarization status and leads to steatohepatitis and fibrogenesis. J Leukoc Biol 2019;105:1015–26.