



# Single-cell analysis revealed a diagnostic model of hepatocellular carcinoma based on cancer stem cell-related gene

Yang Ding<sup>1#</sup>, Qi Liang<sup>1#</sup>, Songbo Ma<sup>1</sup>, Jia Yang<sup>1</sup>, Zengrui Ma<sup>2</sup>, Zhiqi Yang<sup>1</sup>, Baoding Li<sup>1</sup>

<sup>1</sup>Department of Hepatobiliary Surgery, People's Hospital of Ningxia Hui Autonomous Region, Ningxia Medical University, Yinchuan, China;

<sup>2</sup>Department of Anesthesiology, People's Hospital of Ningxia Hui Autonomous Region, Ningxia Medical University, Yinchuan, China

**Contributions:** (I) Conception and design: Y Ding, Q Liang; (II) Administrative support: Y Ding; (III) Provision of study materials or patients: B Li; (IV) Collection and assembly of data: S Ma, J Yang; (V) Data analysis and interpretation: Z Ma, Z Yang, B Li; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

<sup>#</sup>These authors contributed equally to this work as co-first authors.

**Correspondence to:** Baoding Li, MM. Department of Hepatobiliary Surgery, People's Hospital of Ningxia Hui Autonomous Region, Ningxia Medical University, Zheng Yuan North Street 301, Yinchuan 750002, China. Email: 19995350227@163.com.

**Background:** Cancer stem cells (CSCs) play a pivotal role in hepatocellular carcinoma (HCC) pathogenesis, driving tumor initiation, progression, metastasis, and therapeutic resistance. This study aimed to establish a reliable CSC-based signature for HCC through single-cell analysis.

**Methods:** The study integrates single-cell RNA sequencing (scRNA-seq) data with the Dynamic Data Retrieval Tree (DDRTree) algorithm to identify and characterize CSCs in HCC. A prognostic stemness-associated gene signature was constructed using the Cancer Genome Atlas-Liver hepatocellular carcinoma (TCGA-LIHC) cohort as the training set and validated across two independent HCC datasets (GSE14520-GPL571 and GSE14520-GPL3921). Genes with significant prognostic relevance to CSC biology were prioritized for signature inclusion. The efficacy of model was assessed via survival analysis (Kaplan-Meier) and predictive accuracy evaluation [time-dependent receiver operating characteristic (ROC) curves], demonstrating robust stratification of high- versus low-risk HCC patients and strong prognostic discrimination.

**Results:** A total of 17 CSC-related signatures in HCC were identified by copy number variation (CNV) pattern. Then, we constructed 4-CSCs-related gene predictive model by multivariate Cox regression. The model robustly stratified patients into high- and low-risk cohorts, with high-risk individuals exhibiting markedly reduced overall survival (OS). Kaplan-Meier analysis confirmed significant survival disparity between groups ( $P < 0.001$ ), while time-dependent ROC curves validated the model high predictive accuracy. This prognostic signature may guide the development of personalized therapeutic strategies for HCC.

**Conclusions:** The prognostic model consisted of 4-CSC-related genes had a prognostic predictive value, providing a new perspective for precision immuno-oncology studies.

**Keywords:** Cancer stem cells (CSCs); prognosis; hepatocellular carcinoma (HCC); immune

Submitted Mar 18, 2025. Accepted for publication Aug 19, 2025. Published online Oct 22, 2025.

doi: 10.21037/tcr-2025-606

View this article at: <https://dx.doi.org/10.21037/tcr-2025-606>

## Introduction

Hepatocellular carcinoma (HCC), accounting for 75–90% of primary liver cancer cases, is the most prevalent subtype and the second leading cause of cancer-related deaths worldwide (1). HCC shows marked geographic disparities

in incidence, driven by region-specific risks including viral hepatitis, alcohol use, and metabolic disorders (2). Owing to significant heterogeneity of HCC, it is difficult to comprehensively reflect the malignancy degree, invasion and metastasis ability, and sensitivity to treatment. Cancer

stem cells (CSCs)-self-renewing, pluripotent tumor subpopulations-critically contribute to HCC progression, relapse, and treatment resistance (3). CSCs critically underpin tumor progression by orchestrating metastatic dissemination and inducing therapeutic refractoriness, thereby serving as central drivers of disease aggressiveness (4). CSCs not only initiate distant organ colonization but also evade conventional therapies via mechanisms such as enhanced drug efflux (ABC transporters), DNA repair proficiency, and immune evasion, ultimately perpetuating malignant recurrence and poor clinical outcomes (5). Emerging evidence highlights that CSCs drive immune evasion by regulating checkpoint molecules (e.g., PD-L1, CTLA-4), recruiting immunosuppressive cells (e.g., regulatory T cells, myeloid-derived suppressor cells), and secreting cytokines [e.g., interleukin-10 (IL-10), transforming growth factor- $\beta$  (TGF- $\beta$ )] to suppress cytotoxic T-cell function (6). Additionally, CSC-derived exosomes may transfer non-coding RNAs (e.g., miR-146a, lncRNA H19) to reprogram the tumor microenvironment (TME), fostering an immunosuppressive niche conducive to tumor survival and progression (7). CSCs, identified by surface markers such as CD133, CD44, EpCAM, and CD13, exhibit enhanced self-renewal and differentiation capacities driven by dysregulated stemness-related pathways, such as Wnt/ $\beta$ -catenin, Hedgehog, Notch, and STAT3 (8). Many CSC-related markers, including CD13, CD24, EPCAM, CD44,

and CD133 play a critical role in compromising liver regeneration and precipitating liver failure (9). However, the currently discovered markers of CSCs in HCC lack specificity, most of which are also expressed in normal liver cells or other cells, resulting in unreliable early diagnosis and poor treatment. Meanwhile, the mechanism of action of CSC in tumor cells remains unclear. Consequently, there is an urgent demand to identify uniquely expressed CSCs in liver.

Nowadays, the heterogeneity of HCC cell lines and the dynamic expression of CSCs markers can be dissected by single-cell RNA sequencing (scRNA-Seq) (10). Initial investigations found CSC-related markers including CD24, CD133, and EPCAM that independently predict HCC patient survival (11). High CD133 expression correlates with poor survival and recurrence of HCC patient, and EpCAM cells exhibit enhanced metastatic potential (12). Therefore, identifying reliable CSC markers is critical for prognosis prediction and therapeutic targeting.

This study aimed to identify critical CSC-associated genes driving HCC progression through scRNA-Seq analysis. By integrating clinical survival data, we defined prognostic CSC-related signatures which were then validated across independent cohorts. These signatures accurately stratify HCC patients into high- and low-risk subgroups, offering a tool to avoid overtreatment in low-risk individuals. Our findings further reveal CSC features linked to immune evasion mechanisms. We present this article in accordance with the TRIPOD reporting checklist (available at <https://tcr.amegroups.com/article/view/10.21037/tcr-2025-606/rc>).

### Highlight box

#### Key findings

- This study constructed a 4-gene cancer stem cells (CSCs)-related prognostic model, and time-dependent receiver operating characteristic curves confirmed the model's high predictive accuracy.

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

- CSCs drive hepatocellular carcinoma (HCC) progression, metastasis, and therapy resistance. Prognostic models for HCC exist but often lack specificity in targeting CSCs.
- This study integrated single-cell RNA sequencing analysis to construct a CSC-related prognostic signature, offering a novel perspective on HCC risk stratification and personalized therapy.

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

- The identified CSC-based prognostic model provides a valuable tool for predicting patient outcomes and guiding individualized treatment strategies. Incorporating this model into clinical decision-making could improve risk assessment and precision immuno-oncology approaches in HCC.

## Methods

### Data and acquisition

The study was conducted in accordance with the Declaration of Helsinki and its subsequent amendments. The HCC single-cell data set of GSE222791 was downloaded from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). Additionally, gene expression data set of GSE14520 was downloaded from the GEO database encompassing two cohorts, categorized into GSE14520-GPL3921 and GSE14520-GPL571. Clinical and metadata were accessible from the original studies. Meanwhile, the HCC transcriptome data (TCGA-LIHC) were retrieved from UCSC Xena (<http://xena.ucsc.edu/>).

### *Dimension reduction and clustering analysis*

Based on the Seurat package, the CreateSeuratObject function is used to read and construct single-cell seurat data format. FindIntegrationAnchors and IntegrateData functions are used to remove batch effect. NormalizeData and ScaleData functions are used for data normalization analysis. The top 20 most variable genes were subsequently employed for dimensionality reduction clustering using the FindClusters function (resolution =0.5), and the difference marker between cell clusters was analyzed based on FindMarkers. The cell clusters were further annotated in cellMarker (<http://117.50.127.228/CellMarker/>).

### *InferCNV for single-cell analysis and identification of CSCs in pseudotime analysis*

The InferCNV analysis was employed to identify somatic large-scale chromosomal copy number alterations in each cell. Therefore, Subset function was utilized to extract single-cell data of Hepatocytes and B cells. To quantitatively evaluate copy number variation (CNV) alterations in individual cells, we constructed a CNV scoring system based on reference B cells. First, the average gene expression and standard deviation across all B cells were computed. Thresholds for copy number gain or loss were defined using two standard deviations above and below the mean expression to set the low CNV and high CNV values. Based on the CNV score, a kmeans function was used for cluster analysis, and a fixed random seed [set.seed(123)] was employed to ensure reproducibility. When the number of clusters (k) was between 2 and 6, there was no difference in the distribution of CNV scores among different cluster groups. However, when k was 7, the distribution of CNV scores among different cluster groups showed significant differences. Therefore, the final clustering result consisted of 7 clusters (C1–C7).

Then, Dynamic Data Retrieval Tree (DDRTree) algorithm is implemented through the Monocle2 package and is used to infer the pseudo-temporal trajectory of liver cells based on CNV classification. DDRTree algorithm combines discriminative dimensionality reduction with reverse graph embedding, projecting high-dimensional single-cell data into a low-dimensional space while retaining developmental trajectories. We set “max\_components” to 2, “num\_dim” to 50, and use “reduction\_method = DDRTree”. The sorted genes were selected based on the highly variable genes in HCC classified by CNV clustering.

Furthermore, we ensure robustness by using different random seed repeat analyses.

### *Differential expression analysis*

Differential gene screening was performed by limma package in TCGA-LIHC and GSE14520-GPL3921, respectively, under the conditions  $|\log_2 \text{fold change (FC)}| > 1$  and  $P < 0.05$ . In GSE222791, marker genes were screened according to kmeans clustering results. The VennDiagram package was used to screen the intersection genes of the above genes to identify the CSC-related DEGs.

### *Construction of CSC-related gene prognostic model*

Univariate Cox regression analysis of overlapping genes in TCGA-LIHC was performed using the survival package, and the genes significantly correlated with survival time were screened out (log-rank test,  $P < 0.05$ ). In terms of feature selection, we adopted a stepwise multivariate Cox regression method, conducting both forward and backward selection simultaneously, and used the Aki Information Criterion (AIC) as the optimization criterion to prevent overfitting. Only the genes retained in the final multivariate model are used to calculate the risk score, and the calculation formula is:

$$\text{Risk score} = \sum_{i=1}^n (\beta_i \times \text{Expression}_i) \quad [1]$$

$\beta_i$  represents the regression coefficient of the characteristic genes in the model. The model performance was compared between the high-risk group and the low-risk group using Kaplan-Meier (KM) survival analysis via the survfit function. Meanwhile, a time-dependent receiver operating characteristic (ROC) curve was adopted to calculate the 1- and 3-year survival area under curve (AUC) values. The predictive performance of the model was tested in GSE14520-GPL571 and GSE14520-GPL3921. Before applying the model, we normalized all the gene expression profiles with  $\log_2$  transformation. The risk score in the validation set is calculated using the regression coefficients of prognostic genes derived from the TCGA model. The model was validated by using KM survival analysis and time-dependent ROC curves to evaluate the prognostic predictive ability.

### *Statistical analysis*

KM survival curves were generated using the survival

package, while ROC curves and boxplots were plotted with the survival ROC and ggpubr packages, respectively. Statistical analyses were conducted in R 4.3.2, with significance defined as  $P < 0.05$ .

## Results

### *The landscape of CSC in HCC*

To assess cellular transcriptomic heterogeneity and explore CSC functions at the single cell level, a total of 4 HCC samples from GSE222791 were analyzed. The expression profile dataset GSE14520 downloaded from the GEO database included two cohorts, namely, cohort 1 (GPL571) including 22 HCC samples and 21 controls, and cohort 2 (GPL3921) including 225 HCC samples and 220 controls, resulting in the identification of 40,805 cells for single-cell analysis. Single-cell analysis of 40,805 cells resolved seven distinct clusters defined by classical markers: T cells, monocytes, macrophages, hepatocytes, fibroblasts, endothelial cells, and B cells (*Figure 1A*). Cellular cluster proportions per sample were quantified using scRNA-seq data and summarized in *Figure 1B*. The expression of marker genes for seven cell subtypes was shown in *Figure 1C-1E*. The results of data quality control and multiple samples integration after batch removal are shown in *Figures S1,S2*. The UMAP plot showing the expression of marker genes for all cell subtypes can be found in *Figure S3*.

### *InferCNV analysis*

InferCNV analysis distinguished malignant cells, and CNV scores for each cluster were calculated to identify HCC malignancy. B cells (CNV-negative per heatmap) served as reference, enabling classification of hepatocytes as malignant (*Figure 2A*). Moreover, the CNV scores of hepatocytes cells were obviously higher than that of B cells ( $P < 0.05$ , *Figure 2B*). Then, hepatocytes were clustered by kmeans. Clusters 2, 3, 4, and 5 with high CNV score, were identified as malignant clusters, while the remaining clusters exhibited significantly lower CNV scores (*Figure 2C*).

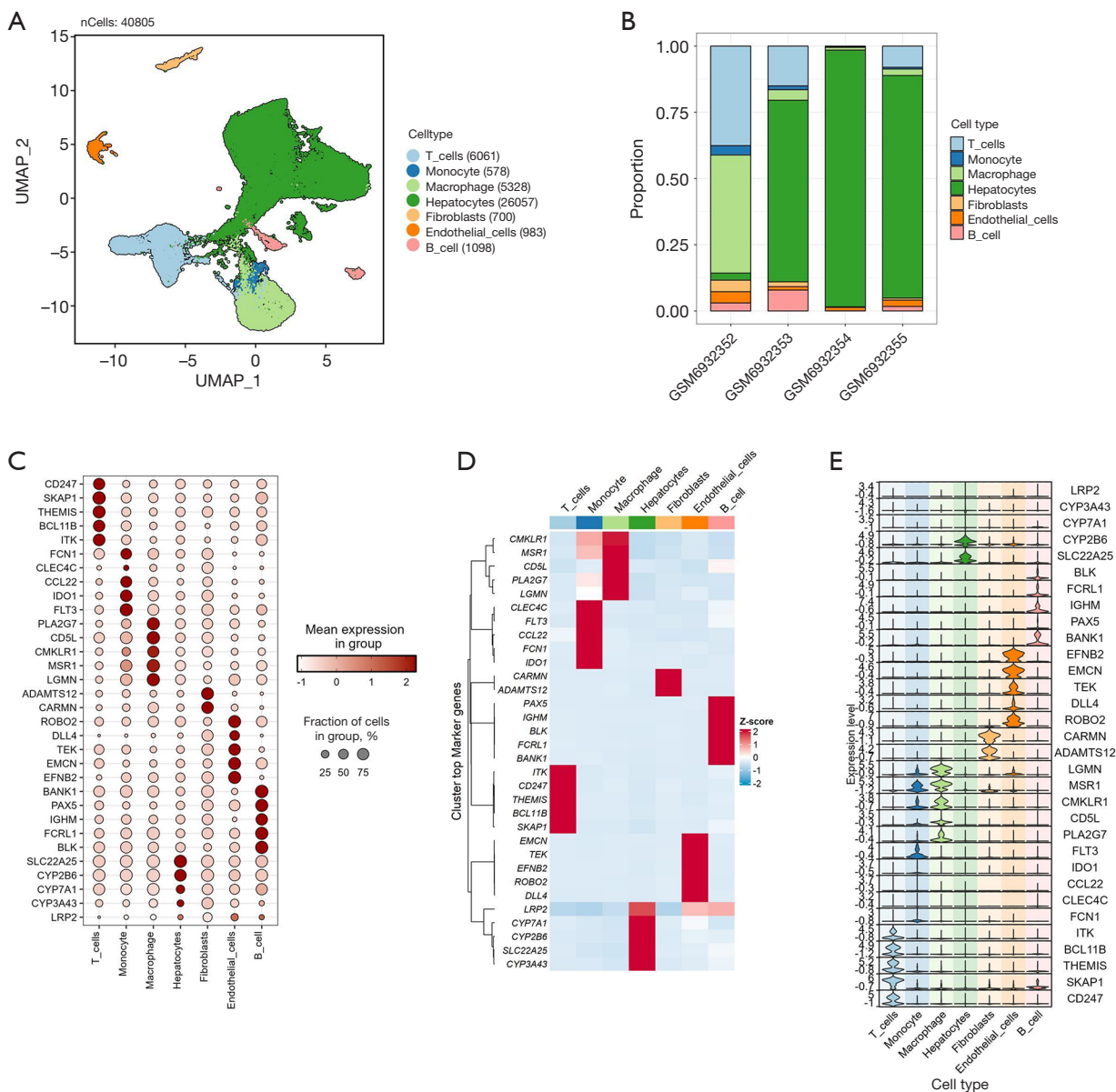
### *Identification of CSC-related differential genes*

To distinguish between CSC and non-CSC cells Clusters 2, 3, 4, and 5, we leveraged the monocle2 package to infer cell lineages and pseudotimes from single-cell gene

expression data. Based on the results of the pseudotime developmental trajectory and differentiation states (*Figure 3A*), the C3 was located at the starting point and the first stage of the developmental trajectory, suggesting that this cells cluster may be CSCs. Additionally, the significantly high expression of CD44, a typical CSC marker gene in HCC, in C3 further supported this hypothesis (*Figure S4*). Then, differentially expressed genes (DEGs) in TCGA-LIHC and GSE14520-GPL3921 were identified by limma package with the condition of  $|\log_2FC| > 1$ ,  $P < 0.05$  (*Figure 3B*). By intersecting the marker genes in cluster 3 with the DEGs in TCGA-LIHC and GSE14520-GPL3921 by the Venn diagram, we identified 17 CSC-related genes, namely *HAMP*, *GPC3*, *DNAJC6*, *NT5DC2*, *UBD*, *ATAD2*, *LAMC1*, *GABRE*, *LRRRC1*, *MUC13*, *STK39*, *SDS*, *PPP1R1A*, *TRIM22*, *FGFR2*, *SPINK1*, and *IGF2BP2*, which constituted the CSC signature (*Figure 3C*).

### *Construction of CSC-related gene predictive model*

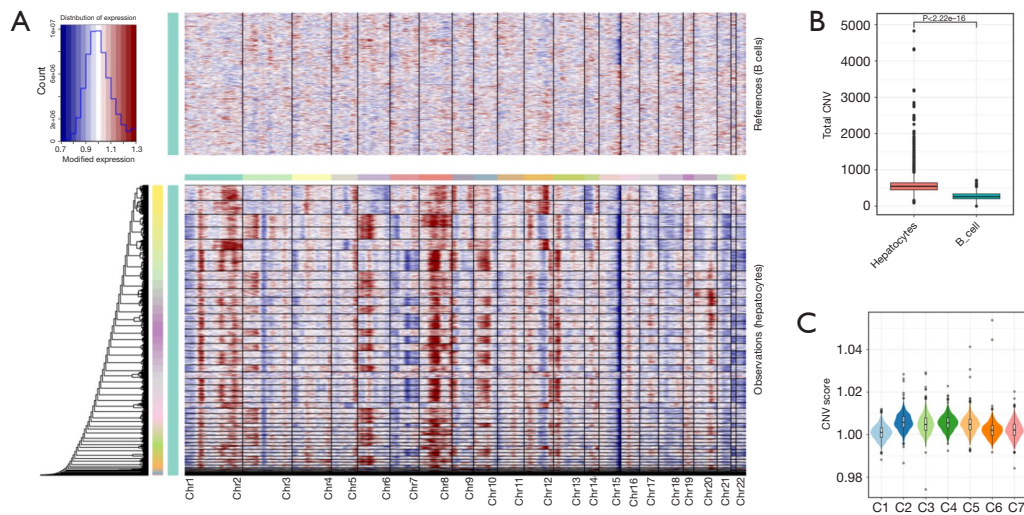
The survival package was used to perform univariate Cox regression for the selected intersected genes in TCGA-LIHC, and the genes significantly correlated with survival time were screened out (log-rank test,  $P < 0.05$ ), and then four genes including *LAMC1*, *TRIM22*, *ATAD2*, and *LRRRC1* were constructed for the prognosis model by multivariate Cox regression. The risk score for patients was calculated according to the formula: Risk score =  $0.187 \times LAMC1 - 0.213 \times TRIM22 + 0.224 \times ATAD2 + 0.102 \times LRRRC1$ . All patients were divided into high-risk and low-risk groups based on the median risk score. The predictive performance of the model was tested in GSE14520-GPL571 and GSE14520-GPL3921. The clinical features (including age, gender, tumor stage and alpha-fetoprotein level) of GSE14520 and TCGA datasets are showed in *Tables 1,2*. Statistical evaluation was conducted using the Chi-squared test (for categorical variables) and the Wilcoxon rank sum test (for continuous variables). This ensures biological comparability between two datasets. Meanwhile, we normalized all the gene expression profiles with  $\log_2$  transformation. The risk score in the validation set is calculated using the regression coefficients of prognostic genes derived from the TCGA model. The KM curves in two validation sets (GSE14520-GPL571 and GSE14520-GPL3921) both showed that the patients in the low risk group exhibited a more significant survival advantage than those in the high risk group (*Figure 4A,4B*,  $P < 0.05$ ). The AUC of the survival time of patients at 1-years in



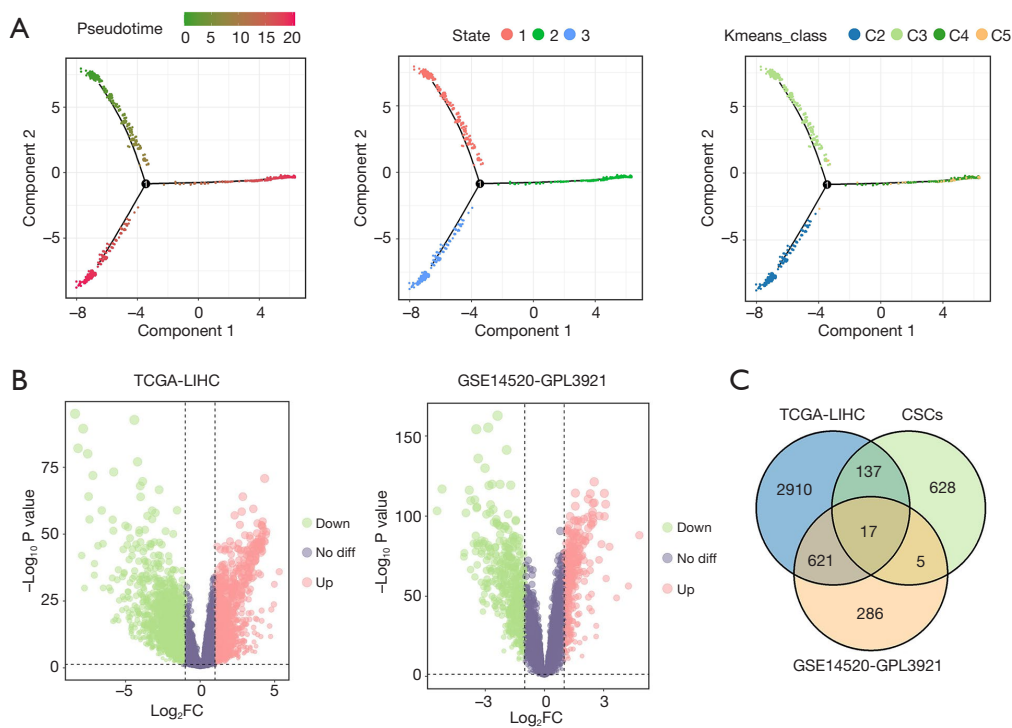
**Figure 1** The landscape of CSCs in hepatocellular carcinoma. (A) Identification of seven distinct clusters in the HCC samples: T cells, monocyte, macrophage, hepatocytes, fibroblasts, endothelial cells, and B cells. (B) The proportion of seven cell clusters in the HCC samples. (C) Dot plot, (D) heat map, and (E) stat plot of marker gene expression in seven cell clusters. CSCs, cancer stem cells; HCC, hepatocellular carcinoma; UMAP, Uniform Manifold Approximation and Projection.

GSE14520-GPL3921, GSE14520-GPL571, and TCGA-LIHC cohort is 0.707, 0.73, and 0.596, respectively (Figure 4A-4C). Moreover, The AUC of the survival time of patients at 3 years in GSE14520-GPL3921, GSE14520-GPL571, and TCGA-LIHC cohort is 0.681, 0.7, and 0.561, respectively (Figure 4A-4C). This indicated that the model has an excellent predictive effect. Meanwhile,

under the different clinical level like the older, the younger, the patients at the stage I-II, the patients at the stage III-IV, the male, and the female, the survival rate of patients in the high-risk group is all significantly lower than that in the low-risk group ( $P < 0.05$ , Figure 5A-5F). This model can clearly demonstrate the close connection between this prognostic model and clinical features, proving that it



**Figure 2** The inferCNV analysis was performed to discrete malignant cells. (A) Hierarchical heatmap illustrating large-scale copy number variations in cancer cells and spiked-in controls (B cells). The green bar to the far left represents all cells used as the observation cells, encompassing B cells and hepatocytes cells within the dataset. The different colors in the top row correspond to the 32 pairs of chromosomes on the x-axis. (B) Boxplot of the copy number variation scores of hepatocytes cells and B cells. (C) Stat plot of kmeans clustering of hepatocytes. CNV, copy number variation.



**Figure 3** Identification of CSCs-related differential genes in liver cancer. (A) Results of quasi-temporal analysis. (B) Volcano of DEGs in TCGA-LIHC and GSE14520-GPL3921. (C) Venn diagram of intersection genes of CSCs, DEGs in TCGA-LIHC, and GSE14520-GPL3921. CSCs, cancer stem cells; DEGs, differentially expressed genes; FC, fold change; TCGA-LIHC, The Cancer Genome Atlas-Liver Hepatocellular Carcinoma.

**Table 1** Clinical pathological characteristics of patients in the GSE14520 database

Variables	High risk (n=121)	Low risk (n=121)	P
Gender			0.70
Female	14 (11.6)	17 (14.0)	
Male	107 (88.4)	104 (86.0)	
Age (years)	48.9±11.0	52.8±10.5	0.006
Stage			<0.001
I	34 (28.1)	51 (42.1)	
II	31 (25.6)	37 (30.6)	
III	42 (34.7)	24 (19.8)	
IV	14 (11.6)	0 (0.0)	
Unknown	0 (0.0)	9 (7.4)	
AFP			0.003
High (>300 ng/mL)	69 (57.0)	45 (37.2)	
Low (≤300 ng/mL)	52 (43.0)	76 (62.8)	

Data are presented as mean ± standard deviation or n (%). AFP, alpha-fetoprotein.

**Table 2** Clinical pathological characteristics of patients in the TCGA database

Variables	High risk (n=211)	Low risk (n=211)	P
Age (years)	57.7±13.3	61.8±14.0	0.002
Gender			0.837
Female	70 (33.2)	73 (34.6)	
Male	141 (66.8)	138 (65.4)	
Stage			<0.001
I	17 (8.1)	43 (20.4)	
II	88 (41.7)	117 (55.5)	
III	93 (44.1)	44 (20.9)	
IV	10 (4.7)	2 (0.9)	
Unknown	3 (1.4)	5 (2.4)	
AFP			<0.001
High (>300 ng/mL)	56 (26.5)	16 (7.6)	
Low (≤300 ng/mL)	102 (48.3)	138 (65.4)	
Unknown	53 (25.1)	57 (27.0)	

Data are presented as mean ± standard deviation or n (%). AFP, alpha-fetoprotein; TCGA, The Cancer Genome Atlas.

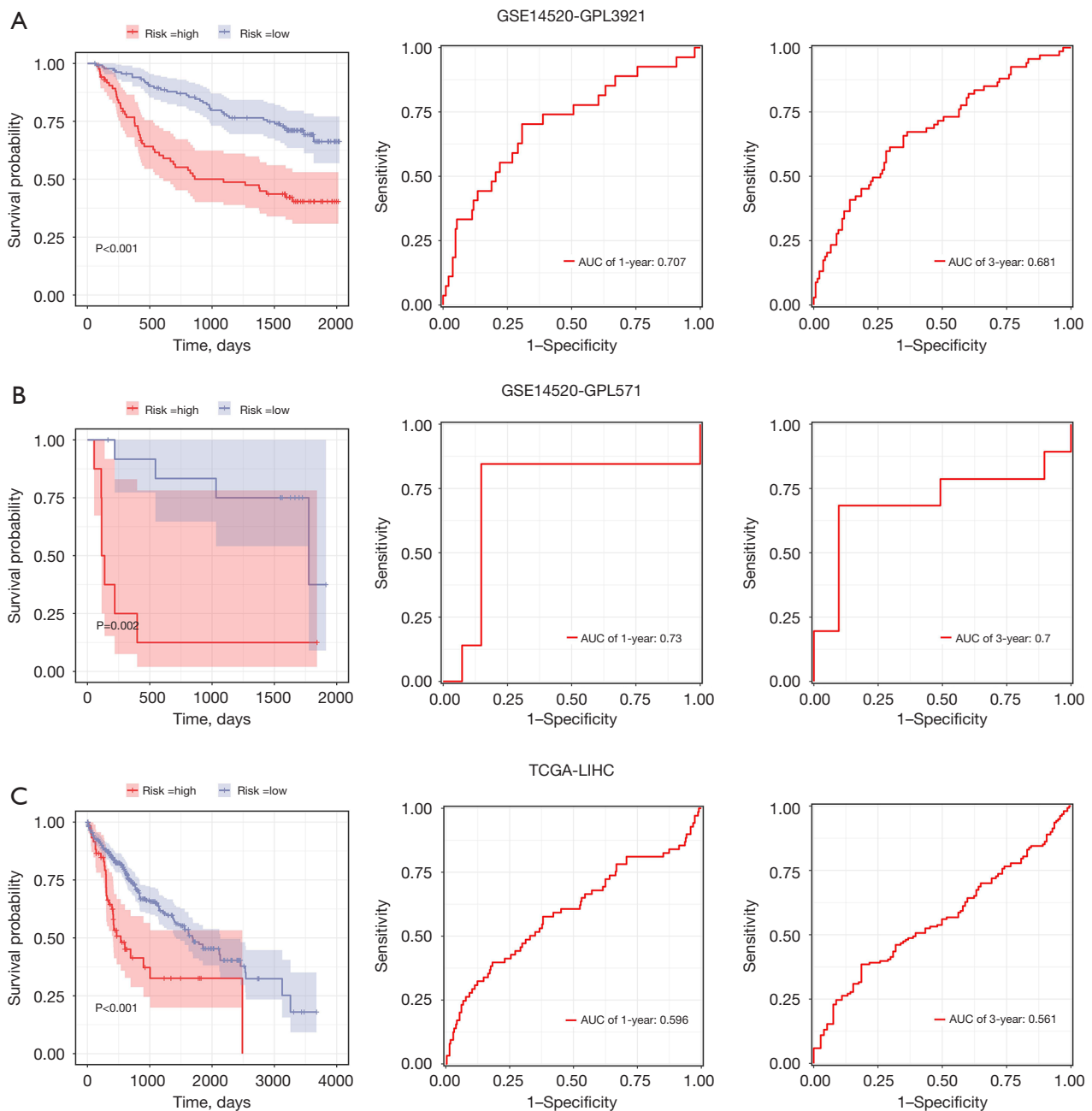
not only has good predictive efficacy but also can provide valuable information for clinical treatment.

## Discussion

HCC is a notoriously aggressive cancer marked by poor prognosis and suboptimal treatment efficacy. CSCs are pivotal drivers of HCC progression, orchestrating tumor metastasis, therapy resistance, and immunosuppressive TME remodeling via cytokine crosstalk and stemness-related pathways (13). CSCs have the capacity to dynamically remodel the immunosuppressive TME through cytokine signaling (e.g., IL-6, TGF-β) and metabolic reprogramming. Despite their clinical relevance, the spatial architecture of CSCs within HCC ecosystems and their interactions with stromal components remain underexplored. Here, we developed a reliable CSC signature from four HCC tissue samples using single-cell data, and established a robust 4-gene signature with cross-platform compatibility based on a panel of CSC-related genes. The proposed model exhibited better predictive performance for overall survival (OS).

Autophagy-, EMT-, and immune-related gene signatures of cancers have been extensively reported (14). While CSCs critically drive HCC therapy resistance, metastasis, and poor outcomes, CSC-based prognostic models remain underexplored (15). Existing models derived from single TCGA cohorts show limited reproducibility due to platform heterogeneity and analytical variability. To address this, we developed a cross-validated signature using consensus genes from three independent cohorts, ensuring enhanced precision and generalizability compared to single-dataset approaches.

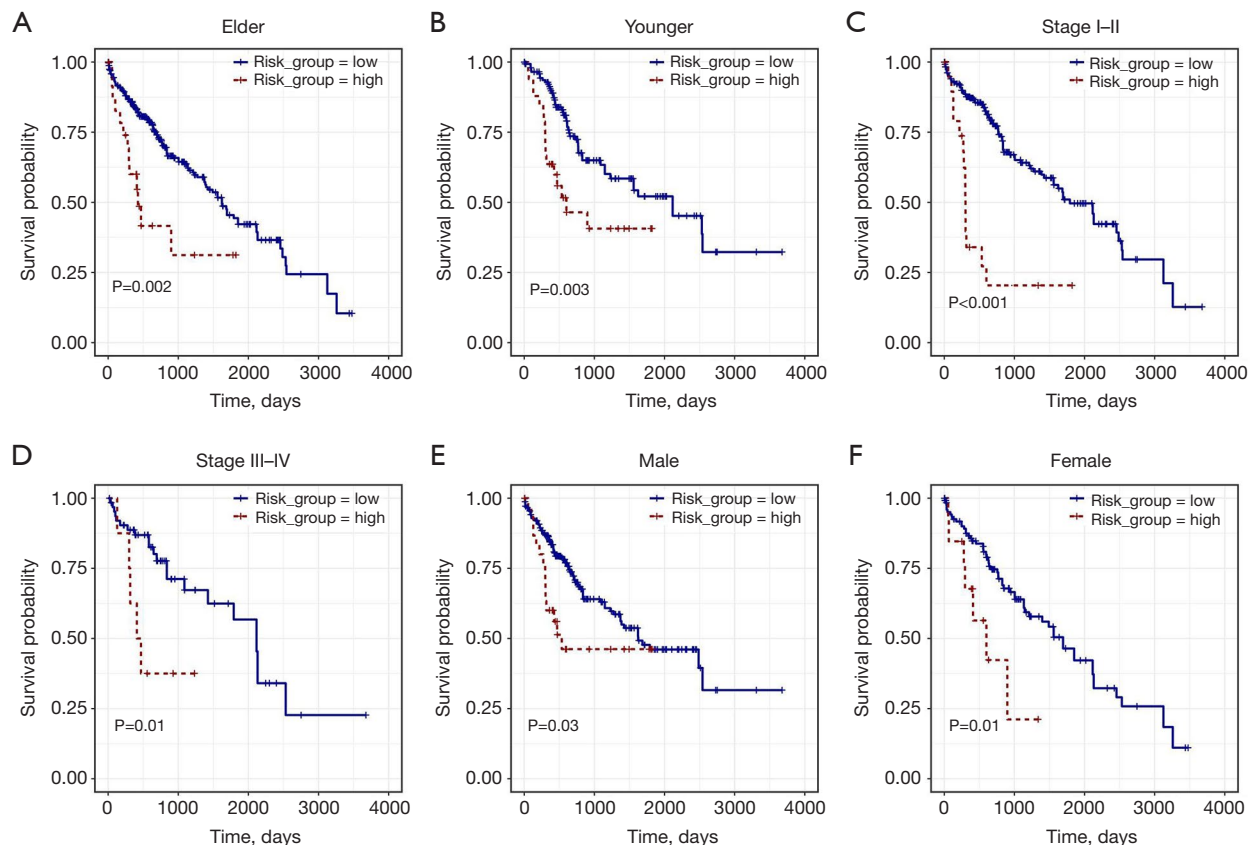
During the construction of the gene signature, we initially identified 17 genes (*HAMP*, *GPC3*, *DNAJC6*, *NT5DC2*, *UBD*, *ATAD2*, *LAMC1*, *GABRE*, *LRRC1*, *MUC13*, *STK39*, *SDS*, *PPP1R1A*, *TRIM22*, *FGFR2*, *SPINK1*, and *IGF2BP2*) linked to OS in HCC were prioritized via univariate Cox regression and visualized in a Venn diagram. Survival analyses confirmed the signature's ability to stratify patients into distinct OS subgroups, while ROC curves demonstrated high AUC values, validating its superior prognostic accuracy. Among those 17 genes, *GPC3* (Glypican-3), cell membrane protein, regulates the Wnt/β-catenin and Hedgehog signaling pathways. It is highly expressed in liver cancer and is a commonly used diagnostic marker (serum or tissue detection) in clinical practice (16). Moreover, highly expression of *SPINK* and *FGFR2* is



**Figure 4** Construction and validation of CSCs-related gene predictive model. Kaplan-Meier curve compares the overall liver cancer patients and time-dependent receiver operator curves analysis in the GSE14520-GPL3921 (A), GSE14520-GPL571 (B), and TCGA-LIHC (C) cohorts. AUC, area under curve; CSCs, cancer stem cells; TCGA-LIHC, The Cancer Genome Atlas-Liver Hepatocellular Carcinoma.

associated with HCC and promotes tumor cell invasion and metastasis (17). The survival rate of *SPINK1* and *FGFR2*-overexpressed patients is significantly reduced. *FGFR2* overexpression promotes HCC progression by activating the MAPK/ERK pathway. Gene fusion events, such as *FGFR2*-

*TACC3*, drive tumorigenesis in some liver cancers (18). *IGF2BP2* is RNA binding proteins that regulate mRNA stability and translation (e.g., *MYC*, *CCND1*). *IGF2BP2* upregulation correlates with liver CSCs, leading to self-renewal and drug resistance. Additionally, tumor growth



**Figure 5** The hepatocellular carcinoma survival probability of the elder (A), the younger (B), the patients at the stage I–II (C), the patients at the stage III–IV (D), the male (E), and the female (F), in the high-risk group and in the low-risk group, respectively.

is promoted through the Wnt/ $\beta$ -catenin pathway and is associated with postoperative recurrence (19). Four genes including *LAMC1*, *TRIM22*, *ATAD2*, and *LRRC1* were further constructed for the prognosis model by multivariate Cox regression. *ATAD2* protein is an emerging oncogene that has strongly been linked to the HCC progression driven by epigenetic mechanisms such as histone acetylation (20). *ATAD2* is involved in multiple pathways including steroid hormone signaling pathway, p53 and p38-MAPK-mediated apoptosis pathway, Akt pathway, Hedgehog signaling pathway, HIF 1 $\alpha$  signaling pathway, and epithelial-mesenchymal transition (EMT) pathway, etc. (21). The self-renewal, differentiation and migration characteristics of CSC depend on the regulation of these signaling pathways, indicating that *ATAD2* may participate in the regulation of characteristics such as CSC self-renewal by influencing this signaling pathway. Meanwhile, multiple omics data show that *LAMC1* is significantly overexpressed in HCC tissues and is positively correlated with tumor stage, vascular

invasion, and distant metastasis. It is inferred that *LAMC1* activates the Wnt/ $\beta$ -catenin pathway through integrin- $\beta$ 1 mediation, up-regulating the stemness markers (22). The expression of *TRIM22* is significantly decreased in HCC, and *TRIM22* regulates cancer cell senescence by modulating the proteasome degradation of *PHLPP2* in HCC cells (23). *LRRC1* is aberrantly overexpressed in HCC tissues and it can promote HCC progression by regulating amino acid and carbohydrate metabolism (24). CSCs rely on a complex network of amino acid metabolism to sustain their self-renewal and differentiation abilities (25). It is inferred that *LRRC1* might regulate the function of CSCs by regulating amino acid metabolism. Other potentially associated gene like *HAMP* is further verification required in the role of HCC progression. This study identified 17 CSC-related genes via single-cell RNA analysis and established a 4-gene prognostic model using multivariate Cox regression. The signature demonstrated robust predictive power for HCC OS, outperforming previous models. Key strengths include

the first comprehensive characterization of CSC spatial organization in HCC and validation of a CSC-specific signature for clinical targeting. However, future work should validate the genes' immune-modulatory roles, test the model in larger cohorts, and elucidate mechanisms linking the signature to poor prognosis.

## Conclusions

In summary, HCC patient stratification via CSC-related gene clusters demonstrated prognostic significance. The 4-gene risk model, derived from CSC-associated markers, offers clinical potential for prognosis assessment and immunotherapy response prediction in HCC, advancing personalized therapeutic strategies.

## Acknowledgments

None

## Footnote

*Reporting Checklist:* The authors have completed the TRIPOD reporting checklist. Available at <https://tcr.amegroups.com/article/view/10.21037/tcr-2025-606/rc>

*Peer Review File:* Available at <https://tcr.amegroups.com/article/view/10.21037/tcr-2025-606/prf>

*Funding:* This work was supported by the Natural Science Foundation of Ningxia (No. 2024AAC03466).

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at <https://tcr.amegroups.com/article/view/10.21037/tcr-2025-606/coif>). All authors report that this work was supported by the Natural Science Foundation of Ningxia (No. 2024AAC03466). The authors have no other conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki and its subsequent amendments.

*Open Access Statement:* This is an Open Access article distributed in accordance with the Creative Commons

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

## References

- Toh MR, Wong EYT, Wong SH, et al. Global Epidemiology and Genetics of Hepatocellular Carcinoma. *Gastroenterology* 2023;164:766-82.
- Singal AG, Lampertico P, Nahon P. Epidemiology and surveillance for hepatocellular carcinoma: New trends. *J Hepatol* 2020;72:250-61.
- Liu YC, Yeh CT, Lin KH. Cancer Stem Cell Functions in Hepatocellular Carcinoma and Comprehensive Therapeutic Strategies. *Cells* 2020;9:1331.
- Toh TB, Lim JJ, Chow EK. Epigenetics in cancer stem cells. *Mol Cancer* 2017;16:29.
- Baumann M, Krause M, Hill R. Exploring the role of cancer stem cells in radioresistance. *Nat Rev Cancer* 2008;8:545-54.
- Zhou H, Tan L, Liu B, et al. Cancer stem cells: Recent insights and therapies. *Biochem Pharmacol* 2023;209:115441.
- Paskeh MDA, Entezari M, Mirzaei S, et al. Emerging role of exosomes in cancer progression and tumor microenvironment remodeling. *J Hematol Oncol* 2022;15:83.
- Zeng Z, Fu M, Hu Y, et al. Regulation and signaling pathways in cancer stem cells: implications for targeted therapy for cancer. *Mol Cancer* 2023;22:172.
- Marquardt S, Solanki M, Spitschak A, et al. Emerging functional markers for cancer stem cell-based therapies: Understanding signaling networks for targeting metastasis. *Semin Cancer Biol* 2018;53:90-109.
- Zhang Y, Wang D, Peng M, et al. Single-cell RNA sequencing in cancer research. *J Exp Clin Cancer Res* 2021;40:81.
- Zheng H, Pomyen Y, Hernandez MO, et al. Single-cell analysis reveals cancer stem cell heterogeneity in hepatocellular carcinoma. *Hepatology* 2018;68:127-40.
- Khosla R, Hemati H, Rastogi A, et al. miR-26b-5p helps in EpCAM+cancer stem cells maintenance via HSC71/HSPA8 and augments malignant features in HCC. *Liver Int* 2019;39:1692-703.

13. Lee TK, Guan XY, Ma S. Cancer stem cells in hepatocellular carcinoma - from origin to clinical implications. *Nat Rev Gastroenterol Hepatol* 2022;19:26-44.
14. Zeng F, Liu X, Wang K, et al. Transcriptomic Profiling Identifies a DNA Repair-Related Signature as a Novel Prognostic Marker in Lower Grade Gliomas. *Cancer Epidemiol Biomarkers Prev* 2019;28:2079-86.
15. Zhou HM, Zhang JG, Zhang X, et al. Targeting cancer stem cells for reversing therapy resistance: mechanism, signaling, and prospective agents. *Signal Transduct Target Ther* 2021;6:62.
16. Piñero F, Dirchwolf M, Pessôa MG. Biomarkers in Hepatocellular Carcinoma: Diagnosis, Prognosis and Treatment Response Assessment. *Cells* 2020;9:1370.
17. Yang C, Guo L, Du J, et al. SPINK1 Overexpression Correlates with Hepatocellular Carcinoma Treatment Resistance Revealed by Single Cell RNA-Sequencing and Spatial Transcriptomics. *Biomolecules* 2024;14:265.
18. Shek FH, Luo R, Lam BYH, et al. Serine peptidase inhibitor Kazal type 1 (SPINK1) as novel downstream effector of the cadherin-17/ $\beta$ -catenin axis in hepatocellular carcinoma. *Cell Oncol (Dordr)* 2017;40:443-56.
19. Yang X, Yang C, Zhang S, et al. Precision treatment in advanced hepatocellular carcinoma. *Cancer Cell* 2024;42:180-97.
20. Hussain M, Zhou Y, Song Y, et al. ATAD2 in cancer: a pharmacologically challenging but tractable target. *Expert Opin Ther Targets* 2018;22:85-96.
21. Nayak A, Dutta M, Roychowdhury A. Emerging oncogene ATAD2: Signaling cascades and therapeutic initiatives. *Life Sci* 2021;276:119322.
22. Zhang Y, Xi S, Chen J, et al. Overexpression of LAMC1 predicts poor prognosis and enhances tumor cell invasion and migration in hepatocellular carcinoma. *J Cancer* 2017;8:2992-3000.
23. Kang D, Hwang HJ, Baek Y, et al. TRIM22 induces cellular senescence by targeting PHLPP2 in hepatocellular carcinoma. *Cell Death Dis* 2024;15:26.
24. Cai Q, Wu D, Shen Y, et al. Prognostic significance of LRRC1 in hepatocellular carcinoma and construction of relevant prognostic model. *Medicine (Baltimore)* 2023;102:e34365.
25. Gong Y, Wang X, Chen W, et al. Cancer stem cells amino acid metabolism: Roles, mechanisms, and intervention strategies. *Cell Signal* 2025;134:111903.

**Cite this article as:** Ding Y, Liang Q, Ma S, Yang J, Ma Z, Yang Z, Li B. Single-cell analysis revealed a diagnostic model of hepatocellular carcinoma based on cancer stem cell-related gene. *Transl Cancer Res* 2025;14(10):6803-6813. doi: 10.21037/tcr-2025-606