# Stemness analysis in hepatocellular carcinoma identifies an extracellular matrix gene–related signature associated with prognosis and therapy response

Lei Chen[1†], Dafang Zhang[1†], Shengmin Zheng[1], Xinyu Li[1] and Pengji Gao[2]*

[1]Department of Hepatobiliary Surgery, Peking University People's Hospital, Beijing, China,
[2]Department of General Surgery, Beijing Jishuitan Hospital, Beijing, China

**Background:** Tumor stemness is the stem-like phenotype of cancer cells, as a hallmark for multiple processes in the development of hepatocellular carcinoma (HCC). However, comprehensive functions of the regulators of tumor cell's stemness in HCC remain unclear.

**Methods:** Gene expression data and clinical information of HCC samples were downloaded from The Cancer Genome Atlas (TCGA) dataset as the training set, and three validation datasets were derived from Gene Expression Omnibus (GEO) and International Cancer Genome Consortium (ICGC). Patients were dichotomized according to median mRNA expression–based stemness index (mRNAsi) scores, and differentially expressed genes were further screened out. Functional enrichment analysis of these DEGs was performed to identify candidate extracellular matrix (ECM)–related genes in key pathways. A prognostic signature was constructed by applying least absolute shrinkage and selection operator (LASSO) to the candidate ECM genes. The Kaplan–Meier curve and receiver operating characteristic (ROC) curve were used to evaluate the prognostic value of the signature. Correlations between signatures and genomic profiles, tumor immune microenvironment, and treatment response were also explored using multiple bioinformatic methods.

**Results:** A prognostic prediction signature was established based on 10 ECM genes, including *TRAPPC4*, *RSU1*, *ILK*, *LAMA1*, *LAMB1*, *FLNC*, *ITGAV*, *AGRN*, *ARHGEF6*, and *LIMS2*, which could effectively distinguish patients with different outcomes in the training and validation sets, showing a good prognostic prediction ability. Across different clinicopathological parameter stratifications, the ECMs signature still retains its robust efficacy in discriminating patient with different outcomes. Based on the risk score, vascular invasion, $\alpha$-fetoprotein (AFP), T stage, and N stage, we further constructed a nomogram (C-index = 0.70; AUCs at 1-, 3-, and 5-year survival = 0.71, 0.75, and 0.78), which is more practical for clinical prognostic risk stratification. The infiltration abundance of macrophages M0,

mast cells, and Treg cells was significantly higher in the high-risk group, which also had upregulated levels of immune checkpoints PD-1 and CTLA-4. More importantly, the ECMs signature was able to distinguish patients with superior responses to immunotherapy, transarterial chemoembolization, and sorafenib.

**Conclusion:** In this study, we constructed an ECM signature, which is an independent prognostic biomarker for HCC patients and has a potential guiding role in treatment selection.

# Introduction

Liver cancer (LC) is one of the most prevalent tumors in the world, ranking as the sixth most common cancer and the third cause of cancer-related deaths worldwide (Sung et al., 2021). The World Health Organization estimated that more than 1 million people will die of LC by 2030 (Villanueva, 2019), which indicates that both the incidence and mortality of LC will increase significantly (Bresnahan et al., 2020). It is a group of heterogeneous malignant tumors with different histological features and outcomes, mainly including hepatocellular carcinoma (HCC), intrahepatic cholangiocarcinoma, and mixed hepatocellular cholangiocarcinoma (Sia et al., 2017; Yang et al., 2019). Patients with HCC are usually diagnosed at advanced stages and are not suitable for the resect surgery because of poor prognosis and a high recurrence rate (Forner et al., 2018). At the same time, the molecular mechanism of HCC is complex, usually involving a variety of genetic abnormalities, such as genomic instability, single nucleotide polymorphisms, somatic mutations, and dysregulation of signaling pathways, which are all related to the occurrence and development of HCC (Tang et al., 2021). Although many advances have been made in the treatments in the past decades, the overall survival of HCC patients is still unsatisfied (Zhang et al., 2018). To improve the management of HCC, novel prognosis biomarkers are urgently needed to evaluate the outcomes and therapy efficacy for patients with HCC.

Cancer stem cells (CSCs) are a small group of undifferentiated cells in tumor tissue and can induce unlimited self-renewal of heterogeneous tumor cells (Kreso and Dick, 2014). A previous study has shown that CSCs are not only crucial for tumor growth and maintenance but also associated with tumor recurrence and metastasis (Peitzsch et al., 2013; Nassar and Blanpain, 2016). Cancer stemness has also been considered to play an important role in HCC (Tsui et al., 2020), supported by the increasing events of genomic, epigenomic, transcriptomic, and proteomic alterations related to it (Eppert et al., 2011). It is well known that the high recurrence rate of HCC is partly due to the presence of CSCs. Furthermore, the sensitivity of conventional radiotherapy and chemotherapy is limited attributed by the biological characteristics of CSCs and protective effect of tumor microenvironment (TME) (Yi et al., 2020). mRNA expression–based stemness index (mRNAsi) is a novel quantitative method to describe the similarities between cancer cells and CSCs, and higher mRNAsi scores are correlated with active biological processes in CSCs and greater tumor dedifferentiation, as reflected by histopathological grades (Pan et al., 2019). It has been identified as an effective marker associated with tumor recurrence and treatment resistance (Malta et al., 2018). Meanwhile, TME not only regulates cancer stemness but also cooperates with CSCs to endure multiple stress condition (Carloni et al., 2014). The extracellular matrix (ECM) as a major structural component of the TME is composed mainly of biochemically distinct components such as fibrous proteins, glycoproteins, proteoglycans, and polysaccharides (Poltavets et al., 2018). It has been revealed that ECM plays a crucial role in normal and CSCs as the niche (Kreso and Dick, 2014). According to previous studies, ECM stiffness had been implicated to promote cancer stemness gene expression to drive the tumor-initiation activity of HCC stem cells (You et al., 2016). The ECM–receptor interaction pathway has been demonstrated as the most significant pathway in cancer (Tian et al., 2021). However, the mechanistic implications of mRNAsi on cancer biology are incompletely understood, and no study has previously attempted to identify the prognostic value of ECMs-related genes and their relationship with mRNAsi in HCC.

In this study, we developed a novel ECM-related signature for predicting the prognosis of HCC using data from the TCGA database. Then we validated its prognostic prediction capacity using data from the GEO database and evaluated its correlation with clinicopathological features, genetic alterations, immune microenvironment, and therapeutic response. This signature is expected to provide potential guidance for personalized treatment of HCC patients.

# Materials and methods

## Study design

A schematic workflow of this study is present in Supplementary Figure S1. In short, we utilized The Cancer

Genome Atlas Liver Hepatocellular Carcinoma (TCGA-LIHC) dataset as the training set, and, three validation datasets were downloaded from Gene Expression Omnibus (GEO) and International Cancer Genome Consortium (ICGC). First, we dichotomized TCGA-LIHC patients according to the median mRNAsi score and compared the overall survival (OS) and disease-free survival (DFS) between the two groups. Subsequently, we compared the gene expression feature of patients from the two groups to screen out differentially expressed genes (DEGs) and then performed a functional enrichment analysis on them to select key pathways and defined the related candidate genes. Third, a least absolute shrinkage and selection operator (LASSO) regression analysis was performed on these candidate genes, and a prognostic model was constructed using TCGA-LIHC data and validated in the GEO cohorts and the ICGC cohort. Then by combination with the prognostic model and clinical characteristics, a nomogram suitable for clinical application was constructed. Finally, the correlation between genomic profile, TME, therapy response, and risk score were analyzed.

## Data source and acquisition of candidate genes

In total, 355 patients with expression data of HCC were obtained from the TCGA database (https://portal.gdc.cancer.gov/). GSE54236 with 161 samples and ICGC-LIRI-JP with 260 samples were retrieved from the GEO (https://www.ncbi.nlm.nih.gov/geo/) or ICGC (https://dcc.icgc.org/) as independent validation datasets. In addition, transcriptome data and calculated tumor volume doubling times of the GSE54236 cohort were used to explore the relationship between the risk score and rapid tumor growth. The tumor volume doubling time was calculated based on imaging data (Villa et al., 2016). The stemness index data for HCC, including mRNAsi and epigenetically regulated mRNAsi (EREG-mRNAsi) were downloaded from a published study (Supplementary Table S1) (Malta et al., 2018). Transcriptome and clinical data from the GSE104580 cohort were used to evaluate the value of the signature to predict the transarterial chemoembolization (TACE) response. By utilization of the R package "IMvigor210CoreBiologies", the association between the signature and the response of atezolizumab (an anti-PD-L1 blockade) was evaluated in the IMvigor210 cohort (an immunotherapy study for bladder cancer). Immunohistochemical (IHC) staining images were downloaded from The Human Protein Atlas database (HPA, https://www.proteinatlas.org/).

## DEG analysis and candidate genes selection

After classifying TCGA HCC patients into two groups based on the median mRNAsi score, the differences in OS and DFS were analyzed by using the "survminer" R package (v.4.0.3). We compared the gene expression feature of the two groups and selected genes with a false discovery rate (FDR) < 0.05 and $|\log_2$ fold change (FC)$| > 1$ as the DEG. Next, we performed a gene set enrichment analysis (GSEA) on the DEGs using GSEA software (version 4.0.1, http://www.broad.mit.edu/gsea/) and selected key pathways, and genes contained in key pathways were selected as candidate ones.

## Construction and validation of an ECM-related prognostic model

First, we used the LASSO analysis to screen variables from high dimensional data to construct an ECM signature (Gui and Li, 2005), and then we calculated the coefficients of this signature using the "glmnet" R package. The risk score formula is as follow: *risk score =Σ Coefficient (ECMs i) * Expression (ECMs i). Coefficient of gene (i)* is the regression coefficient of gene (i) in the LASSO–Cox regression model and *Expression of gene (i)* is the expression value of gene (i) for each patient.

We divided the HCC patients into high- and low-risk groups based on the median risk score and utilized the Kaplan–Meier (K-M) analysis to compare the survival differences between different risk groups; the prognostic capability of ECMs was then measured by the area under the curve (AUC) of a time-dependent receiver operating characteristic (ROC). Last, we utilized datasets from the validation cohorts to confirm this prognostic prediction capacity.

## Correlation between the risk score and clinical characteristics

We analyzed the relationship between the risk score and the clinical features of HCC including AFP, vascular invasion, primary tumor (T) stage, regional lymph nodes (N) stage, and distant metastasis (M) stage. To evaluate the stability of the robustness and accuracy of prognostic models, we compared the survival difference between high- and low-risk groups under clinical characteristic stratification.

## Construction and assessment of a predictive nomogram

We explored the correlation and independence of ECMs signature and clinical characteristics using a multivariate Cox risk regression analysis. In order to be applicable to the clinic, we integrated the risk score with vascular invasion, AFP, T stage, and N stage to construct a nomogram using the "rms" R package. The prognostic predictive ability of nomogram was evaluated by the calibration curve and time-dependent ROC curves. The ROC

curves were plotted by using the "timeROC" R package. The survival net benefits of each variable were estimated by a decision curve analysis (DCA) using the "stdca" R package.

## Functional and pathway enrichment analysis

To explore biological processes related to the risk score, Reactome pathway (https://reactome.org/) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway (https://www.genome.jp/kegg/) analyses were performed by using the "clusterProfiler" R package.

## Determination of tumor-infiltrating immune cells

CIBERSORT is a tool which can estimate the distribution of 22 common tumor-infiltrating immune cells, including B cells, T cells, NK cells, and macrophage cells, by transcriptome profiles from the TCGA cohort (Newman et al., 2015). Meanwhile, the tumor immune dysfunction and exclusion (TIDE) score was calculated using an online tool (http://tide.dfci.harvard.edu/) (Jiang et al., 2018). Patients' epithelial–mesenchymal transition (EMT) scores were calculated according to the methods described in a previously published study (Qiu et al., 2022).

## Genomic profile analysis

Genetic mutation data were analyzed and visualized by using the "maftools" R package. We compared the differences of somatic variants in most prevalent mutated genes between high- and low-risk groups. The copy number variant (CNV) data were downloaded from the Affymetrix Genome-Wide Human SNP Array 6.0 platform using the "TCGAbiolinks" package, and abnormal chromosomal regions were detected using the "R/Bioconductor GAIA" package (Morganella et al., 2011). Non-synonymous mutation data were obtained from the Ensembl Variation (https://asia.ensembl.org/info/genome/variation/prediction/predicted_data.html) database. Subsequently, the data were read and the TMB values were calculated using the "maftools" R package.

## Prediction of treatment response

In total, four classic drugs were selected from the Genomics of Drug Sensitivity in Cancer (GDSC) database (https://www.cancerrxgene.org/) for analysis. The half-maximal inhibitory concentration ($IC_{50}$) is an important indicator for evaluating drug efficacy or sample response to treatment. The GDSC database helps in predicting each sample's response to targeted therapy or chemotherapy based on the sample's transcriptome. According to the cellular expression profiles in the GDSC database, a ridge regression model was constructed with the "pRRopheticl" R package to calculate the $IC_{50}$ of the drugs in the two risk groups. Transcriptome data and treatment response information from the GSE104580 cohort were used to test the effectiveness of signature in discriminating TACE treatment responders from non-responders. Transcriptome data and treatment response information from the IMvigor210 cohort were used to examine the effectiveness of the signature in identifying complete/partial responses (CR/PR) or stable/progressive disease (SD/PD) with immunotherapy.
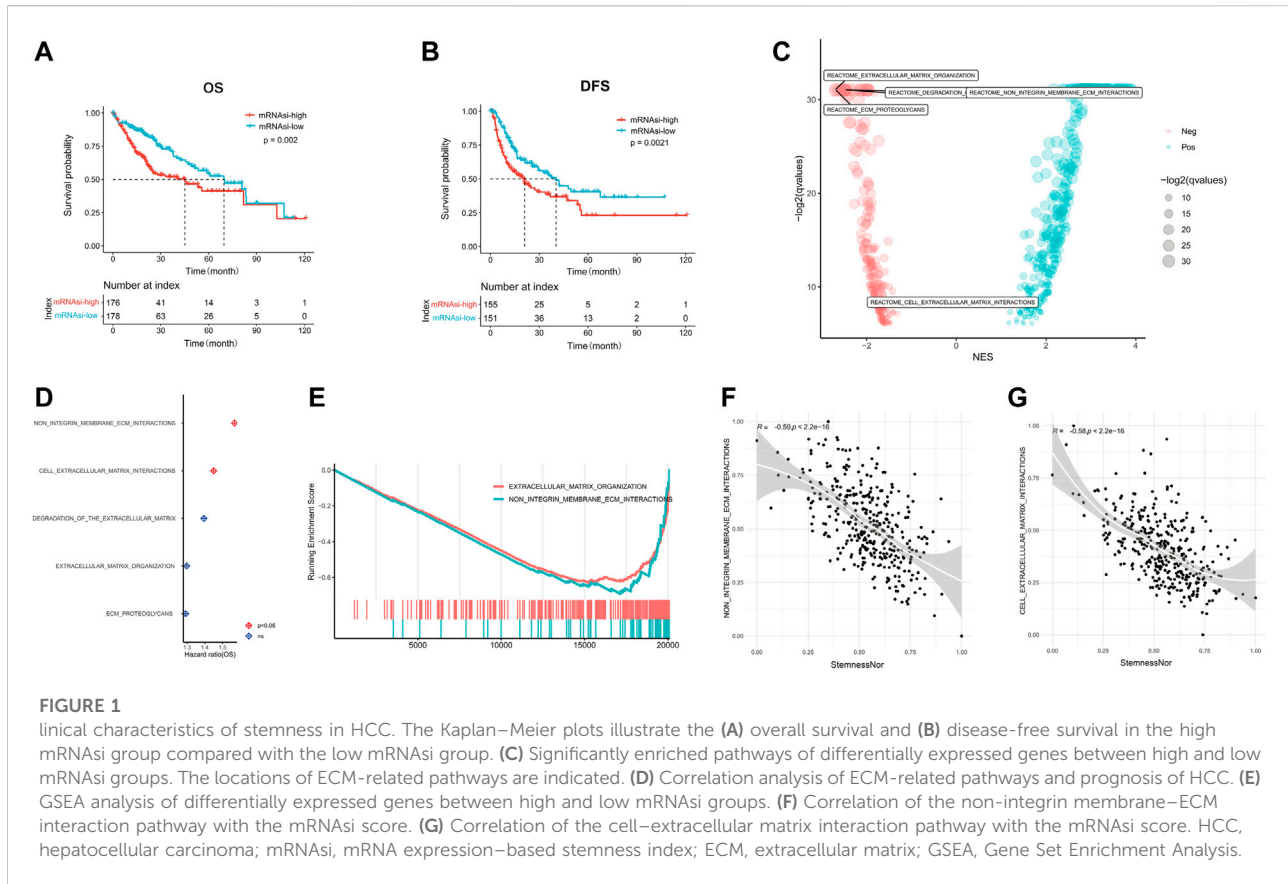
## Statistical analysis

All statistical analyses in this study were based on R software (v.4.0.3). We used Fisher's test to compare the categorical variables and the K-M curve to evaluate the differences in survival between different risk groups. The results of the multivariate Cox regression analysis were visualized using the nomogram. The concordance index, time-dependent ROC, and calibration were also important indicators used to assess the nomogram. All tests were two-sided, and $p < 0.05$ was considered statistically significant, unless otherwise stated.

## Results

### Stemness in HCC

According to the median mRNAsi score, we dichotomized HCC patients in the training set into mRNAsi-high and mRNAsi-low groups and identified that HCC patients with high mRNAsi had both significantly shorter OS (median OS: 45.11 vs. 69.57 months, $p = 0.002$; Figure 1A) and DFS (median DFS: 20.91 vs. 40.41 months, $p = 0.0021$; Figure 1B). The functional enrichment analysis of the DEGs between mRNAsi-high and mRNAsi-low groups revealed that ECM-related pathways were the most significantly enriched (Figure 1C). Among those pathways, non-integrin membrane–ECM interactions and cell–ECM interactions pathways were regarded as the key pathways ($p < 0.05$, Figures 1D,E) with 75 ECM-related genes kept in the two pathways (Supplementary Table S2). The expression levels of these 75 ECM-related genes were mainly negatively correlated with tumor cell stemness ($p < 0.05$, Supplementary Figure S2A). In addition, we also found that non-integrin membrane–ECM interactions and cell–ECM interactions pathways were significantly negatively related with the stemness index (R = −0.59 and −0.58, respectively, $p < 0.001$; Figures 1F,G).

**FIGURE 1**
linical characteristics of stemness in HCC. The Kaplan−Meier plots illustrate the **(A)** overall survival and **(B)** disease-free survival in the high mRNAsi group compared with the low mRNAsi group. **(C)** Significantly enriched pathways of differentially expressed genes between high and low mRNAsi groups. The locations of ECM-related pathways are indicated. **(D)** Correlation analysis of ECM-related pathways and prognosis of HCC. **(E)** GSEA analysis of differentially expressed genes between high and low mRNAsi groups. **(F)** Correlation of the non-integrin membrane−ECM interaction pathway with the mRNAsi score. **(G)** Correlation of the cell−extracellular matrix interaction pathway with the mRNAsi score. HCC, hepatocellular carcinoma; mRNAsi, mRNA expression−based stemness index; ECM, extracellular matrix; GSEA, Gene Set Enrichment Analysis.

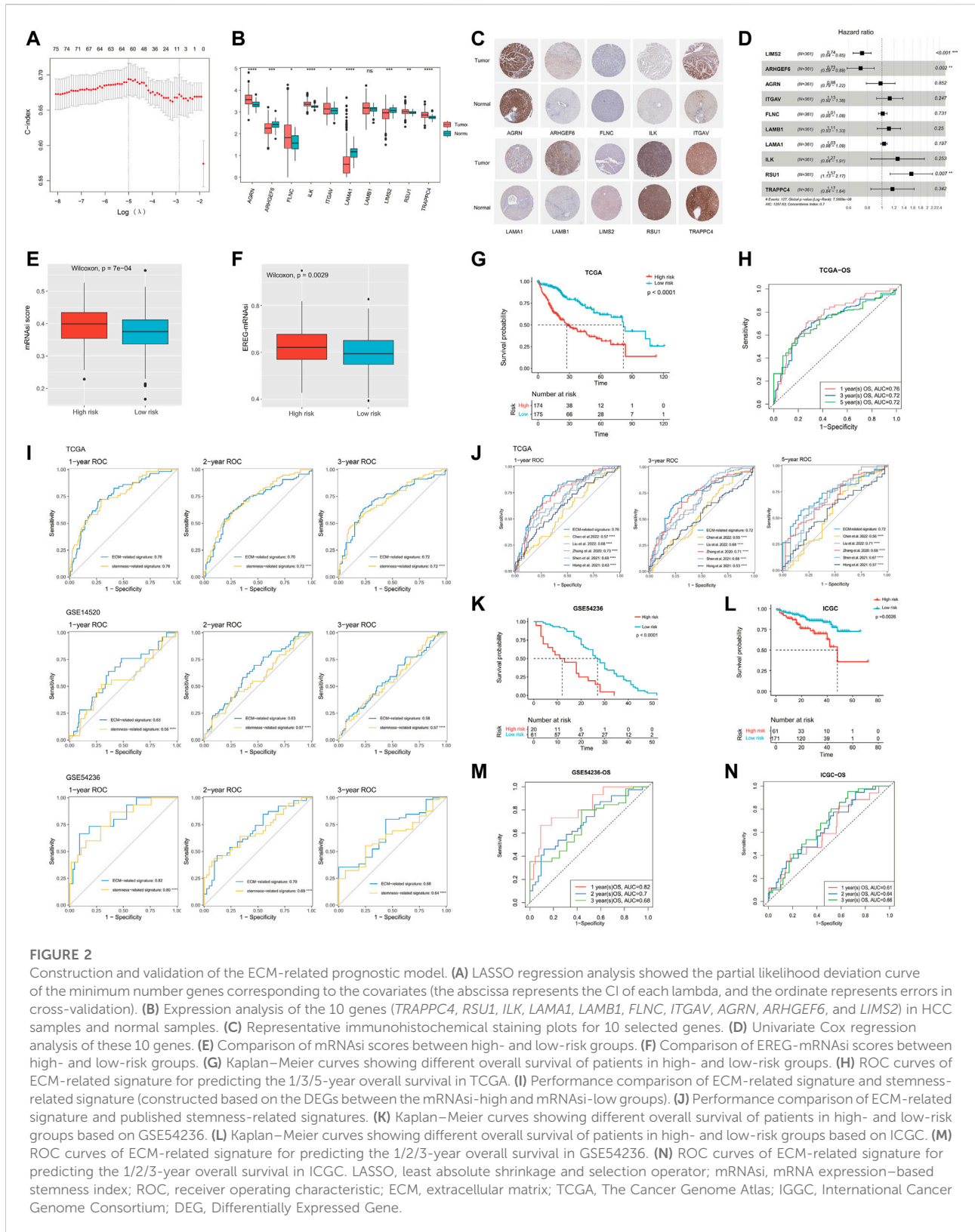## Construction of the ECM-related prognostic model

Subsequently, we performed the LASSO regression analysis on the 75 ECM-related genes screened before and finally selected 10 genes (*LIMS2*, *ARHGEF6*, *AGRN*, *ITGAV*, *FLNC*, *LAMB1*, *LAMA1*, *ILK*, *RSU1*, and *TRAPPC4*) to construct a prognostic prediction model (Figure 2A). The 10 genes were found to be expressed in both tumor and normal tissues, among which *AGRN*, *FLNC*, *ILK*, *ITGAV*, *RSU1*, and *TRAPPC4* were significantly upregulated in the tumor tissues, while *ARHGEF6*, *LAMA1*, and *LIMS2* were significantly downregulated in the tumor tissues ($p < 0.05$, Figure 2B). We downloaded representative images of IHC staining of the proteins encoded by these genes in the TCGA-LIHC cohort from the HPA database. The data showed that the expression levels of these proteins in tumor tissues were higher than those in normal tissues, except for the *LIMS2* protein, which was consistent with mRNA expression (Figure 2C). Matrix metalloproteinase (MMP) is a family of zinc-dependent ECM remodeling endopeptidases with the ability to degrade almost all components of the ECM (Cabral-Pacheco et al., 2020). We found that 10 hub ECM-related genes in this study were significantly

correlated with most members of the MMP family at the mRNA expression level ($p < 0.05$, Supplementary Figure S2B). Furthermore, the forest plots showed that the overexpression of *LIMS2* (HR = 0.74, 95% CI = 0.64–0.85, and $p < 0.001$) and *ARHGEF6* (HR = 0.72, 95% CI = 0.59–0.89, and $p = 0.002$) were related to a significantly superior OS; on the contrary, the upregulation of *RSU1* was significantly associated with a worse survival (HR = 1.57, 95% CI = 1.13–2.17, and $p = 0.007$; Figure 2D). The risk scores were calculated for each patient based on the gene expression levels and Cox regression coefficients.

Risk score= $(-0.227 * \mathrm{Exp}_{LIMS2}) + (-0.008 * \mathrm{Exp}_{ARHGEF6}) + (0.002 * \mathrm{Exp}_{AGRN}) + (0.004 * \mathrm{Exp}_{ITGAV}) + (0.005 * \mathrm{Exp}_{FLNC}) + (0.007 * \mathrm{Exp}_{LAMB1}) + (0.015 * \mathrm{Exp}_{LAMA1}) + (0.043 * \mathrm{Exp}_{ILK}) + (0.131 * \mathrm{Exp}_{RSU1}) + (0.132 * \mathrm{Exp}_{TRAPPC4})$.

As the risk score increased, the stemness of HCC got stronger. The median mRNAsi score in the high-risk group was significantly higher than that in the low-risk group ($p < 0.0001$, Figure 2E). In addition, EREG-mRNAsi is also an index to assess tumor stemness (Malta et al., 2018). Likewise, the median EREG-mRNAsi in the high-risk group was significantly higher than that in the low-risk group ($p = 0.0029$, Figure 2F). Meanwhile, HCC patients in the high-risk

**FIGURE 2**

Construction and validation of the ECM-related prognostic model. **(A)** LASSO regression analysis showed the partial likelihood deviation curve of the minimum number genes corresponding to the covariates (the abscissa represents the CI of each lambda, and the ordinate represents errors in cross-validation). **(B)** Expression analysis of the 10 genes (*TRAPPC4*, *RSU1*, *ILK*, *LAMA1*, *LAMB1*, *FLNC*, *ITGAV*, *AGRN*, *ARHGEF6*, and *LIMS2*) in HCC samples and normal samples. **(C)** Representative immunohistochemical staining plots for 10 selected genes. **(D)** Univariate Cox regression analysis of these 10 genes. **(E)** Comparison of mRNAsi scores between high- and low-risk groups. **(F)** Comparison of EREG-mRNAsi scores between high- and low-risk groups. **(G)** Kaplan−Meier curves showing different overall survival of patients in high- and low-risk groups. **(H)** ROC curves of ECM-related signature for predicting the 1/3/5-year overall survival in TCGA. **(I)** Performance comparison of ECM-related signature and stemness-related signature (constructed based on the DEGs between the mRNAsi-high and mRNAsi-low groups). **(J)** Performance comparison of ECM-related signature and published stemness-related signatures. **(K)** Kaplan−Meier curves showing different overall survival of patients in high- and low-risk groups based on GSE54236. **(L)** Kaplan−Meier curves showing different overall survival of patients in high- and low-risk groups based on ICGC. **(M)** ROC curves of ECM-related signature for predicting the 1/2/3-year overall survival in GSE54236. **(N)** ROC curves of ECM-related signature for predicting the 1/2/3-year overall survival in ICGC. LASSO, least absolute shrinkage and selection operator; mRNAsi, mRNA expression−based stemness index; ROC, receiver operating characteristic; ECM, extracellular matrix; TCGA, The Cancer Genome Atlas; IGGC, International Cancer Genome Consortium; DEG, Differentially Expressed Gene.

group had a significantly worse prognosis than that in the low-risk group (median OS: 27.52 vs. 81.73 months, and $p < 0.0001$; Figure 2G). The AUCs of 1-, 3-, and 5-year OS in the TCGA cohort were 0.76, 0.72, and 0.72, respectively (Figure 2H). In addition, we constructed a stemness-related prognostic signature based on the DEGs between the mRNAsi-high and mRNAsi-low groups (Supplementary Figures S3A–C). The results showed that the ECM-related signature was more accurate and robust than the stemness-related signature ($p < 0.0001$, Figure 2I). Also, the performance of the ECM-related signature was also better than other published stemness-related signatures ($p < 0.0001$, Figure 2J) (Zhang et al., 2020; Hong et al., 2021; Shen et al., 2021; Chen et al., 2022; Liu et al., 2022). We further explored the correlation between our prognostic model and ECM collagen or non-collagen genes. The association of 10 ECM-related genes with collagen or non-collagen was consistent. The expressions of *LAMA1, LAMB1, ARHGEF6, ITGAV, FLNC,* and *RSU1* were significantly positively correlated with most collagen and non-collagen genes, while the expressions of *LIMS2, TRAPPC4,* and *ILK* were significantly negatively correlated with most collagen and non-collagen genes ($p < 0.05$, Supplementary Figure S2C).

## Validation of the ECM-related prognostic prediction model

To validate the accuracy of the ECM-related prognostic prediction model, we performed a confirmatory test using three independent validation sets from the GEO and ICGC databases. The K-M curve showed that there was a significant difference in OS between the high- and low-risk groups, and the prognosis of patients in the high-risk group was significantly worse than that in the low-risk group, consistent with the results of the training cohort (GSE54236, median OS 12 vs. 27 months, and $p < 0.0001$; ICGC-LIRI-JP, median OS 48 vs. not reached, and $p = 0.0026$; Figures 2K,l). The ROC curves showed that the AUC of the 1-, 2-, and 3-year survival time of the GSE54236 cohort was 0.82, 0.7, and 0.68, respectively; the AUC of the 1-, 2-, and 3-year survival time of the ICGC cohort was 0.61, 0.64, and 0.66, respectively (Figures 2M,N). Overall, the ECM-related prognostic model could stably and accurately predict the prognosis of HCC patients. Furthermore, the tumor volume doubling time was significantly negatively correlated with the risk score in HCC patients in the GSE54236 cohort ($p = 0.00076$, Supplementary Figure S4).

## Association with the clinical characteristics

As shown in Figure 3A, the risk score was positively correlated with the AFP level (R = 0.28, p = $3e^{-6}$). In addition, the risk score was significantly higher in the HCC

patients presented with macro-vascular invasion, followed by those with micro-vascular invasion (Figure 3B). Meanwhile, HCC patients with the advanced tumor stage (T) and the positive lymph node stage (N) had significantly higher risk scores (Figures 3C,D). However, as there were limited number of patients with long distant metastasis, there was no significant difference between patients presented with the long distant metastasis (M) stage or not ($p = 0.38$, Figure 3E).
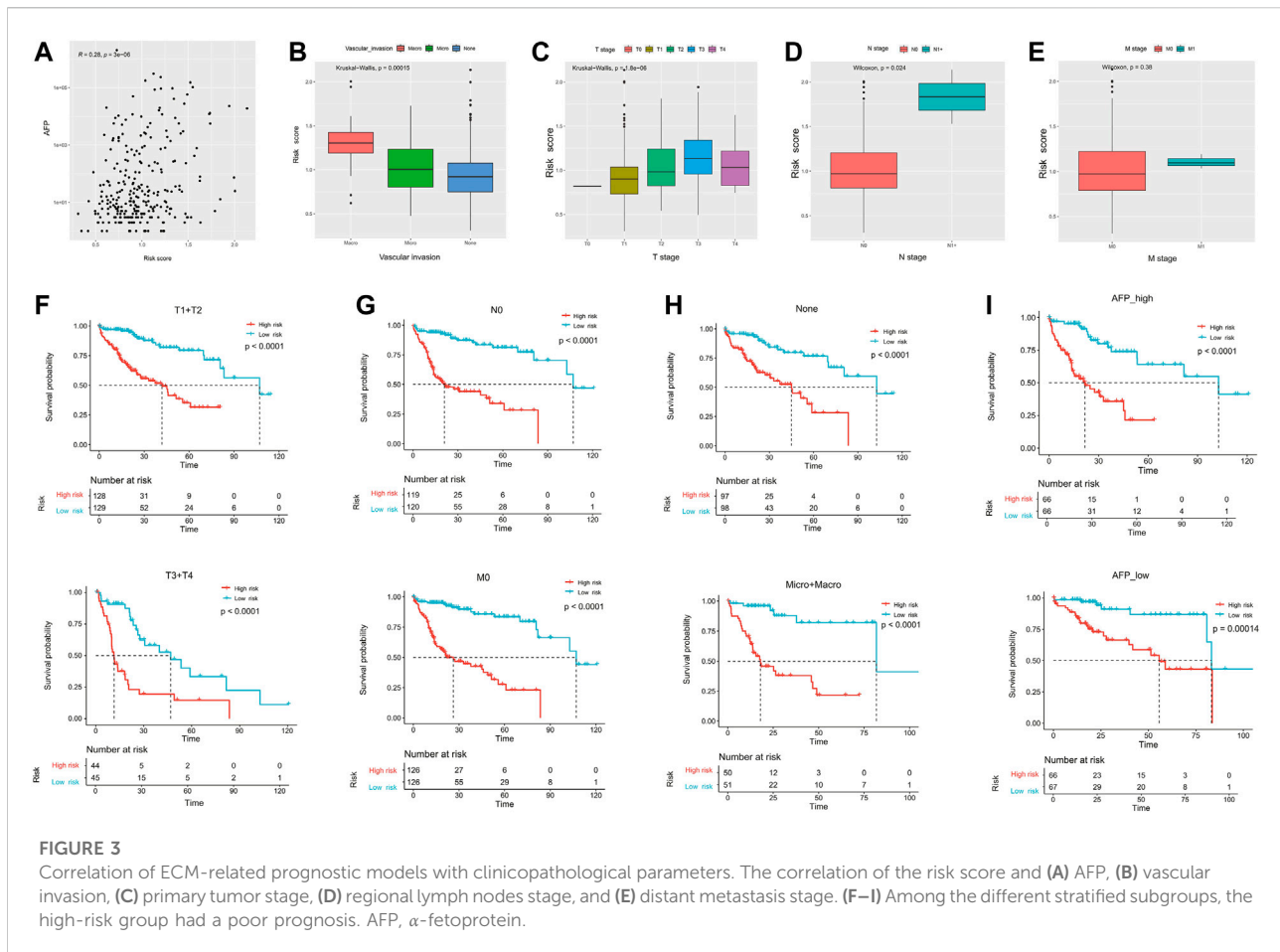
We subsequently divided the clinical features into different subgroups to observe the stability of the ECM prediction efficacy. The results showed that regardless of whether in the advanced tumor stage (T3+T4) or not (T1+T2), patients in the high-risk group had significantly worse OS (Figure 3F). Meanwhile, the trend was similar in patients with N0 (median OS: 21.0 vs. 107.1 months and $p < 0.0001$) and/or M0 (median OS: 26.4 vs. 107.1 months and $p < 0.0001$; Figure 3G). Last, the ECM signature could successfully differentiate HCC patients with worse outcomes regardless of whether present with vascular invasion or not, or whether with high or low AFP levels (Figures 3H,I).

## Establishment of a nomogram integration with independent predictive factors

To better predict the 1-, 3-, and 5-year survival of HCC patients, we constructed a nomogram combining the risk score, vascular invasion, AFP, T stage, and N stage. The C-index of nomogram was 0.70 (95% CI: 0.63–0.77) (Figure 4A). Meanwhile, it could be observed that the calibration curves at 1-, 3-, and 5-year survival showed a good consistency between actual observation and the prediction by nomogram (Figures 4B–D). The results showed the net benefit of nomogram-assisted decisions at a wide range of threshold probabilities, suggesting potential clinical usefulness of the nomograms (Figures 4E–G). The AUCs for the 1-, 3-, and 5-year survival of the constructed nomogram were 0.71, 0.75, and 0.78, respectively (Figure 4H). Finally, the multivariate Cox regression analysis indicated that the risk score and the T stage were the only two independent risk factors (HR = 4.61, 95% CI = 2.36–9.0, and $p < 0.001$; HR = 1.96, 95% CI = 1.17–3.3, and $p = 0.01$; Figure 4I).

## Biological pathways and functional enrichment analysis

We performed a pathway enrichment analysis using DEGs between high- and low-risk groups in the TCGA-LIHC cohort to explore the biological processes associated with the risk scores. A Reactome pathway enrichment results showed that DEGs were significantly enriched in the cell cycle and metabolism-related pathways, such as cell cycle checkpoints, cell cycle mitosis, DNA replication, fatty acid metabolism, and peroxisome lipid

**FIGURE 3**
Correlation of ECM-related prognostic models with clinicopathological parameters. The correlation of the risk score and **(A)** AFP, **(B)** vascular invasion, **(C)** primary tumor stage, **(D)** regional lymph nodes stage, and **(E)** distant metastasis stage. **(F–I)** Among the different stratified subgroups, the high-risk group had a poor prognosis. AFP, $\alpha$-fetoprotein.

metabolism pathways ($p < 0.05$, Figure 5A). The KEGG pathway enrichment results were consistent with the Reactome enrichment, which were also dominated by the cell cycle and metabolism-related pathways, including cell cycle, DNA replication, homologous recombination, pyruvate metabolism, tryptophan metabolism, and starch and sucrose metabolism pathways ($p < 0.05$, Figure 5B).

## Immune landscape, immune checkpoint profile, and immunotherapy response prediction

We evaluated the tumor-infiltrated immune cell contents and immune checkpoint profile and predicted immunotherapy responses in different risk groups. In total, the high-risk group had a significantly higher abundance of macrophage M0, resting mast cells, and Treg cells and lower abundance of activated CD4+ memory T cells, resting NK cells, monocytes, macrophage M1, activated myeloid dendritic cells, and activated mast cells

(Figure 6A). Common immune checkpoints such as PD-1 (Figure 6B) and CTLA-4 (Figure 6C) were significantly upregulated in the high-risk groups and their expression levels were significantly positively correlated with the risk scores. However, the expression level of PD-L1 was not significantly different between the high- and low-risk groups (Figure 6D). In addition, there was no significant difference in the TIDE score between high- and low-risk groups (Figure 6E). In addition, the high-risk group had significantly higher EMT scores than those of the low-risk group ($p < 0.0001$, Figure 6F). We validated the prediction value of the ECMs signature for immunotherapy response of the patients in the IMvigor210 cohort. The K-M survival curve showed that immunotherapy-treated patients in the high-risk group had significantly longer OS (median OS 9.2 vs. 8.0 months, $p = 0.031$; Figure 6G). Moreover, the risk score of immunotherapy patients with objective response (CR/PR status) was significantly higher than those without it (SD/PD, $p = 0.026$; Figure 6H). Also, the proportion of immunotherapy patients with objective response was significantly higher in the high-risk group ($p = 0.03$, Figure 6I).
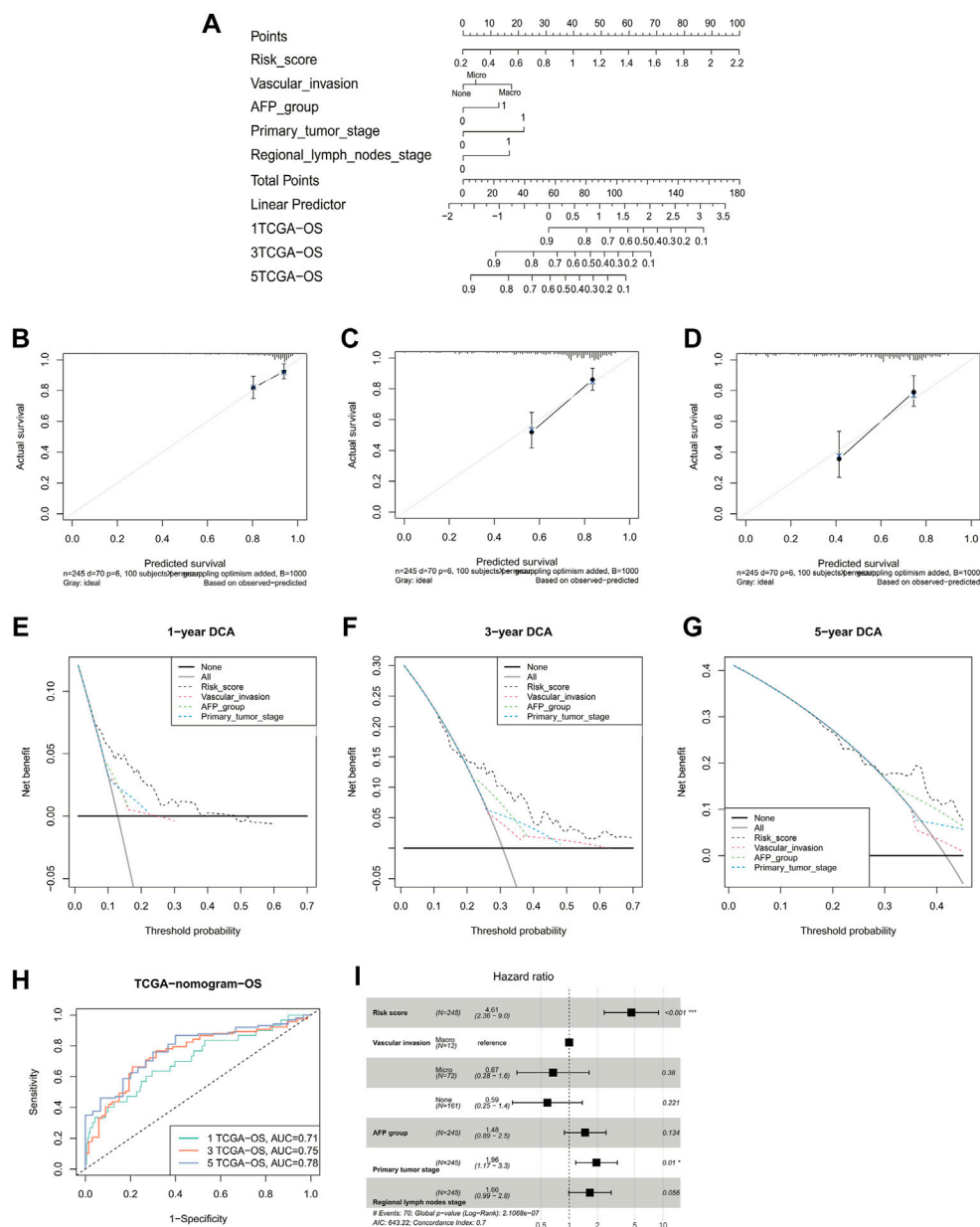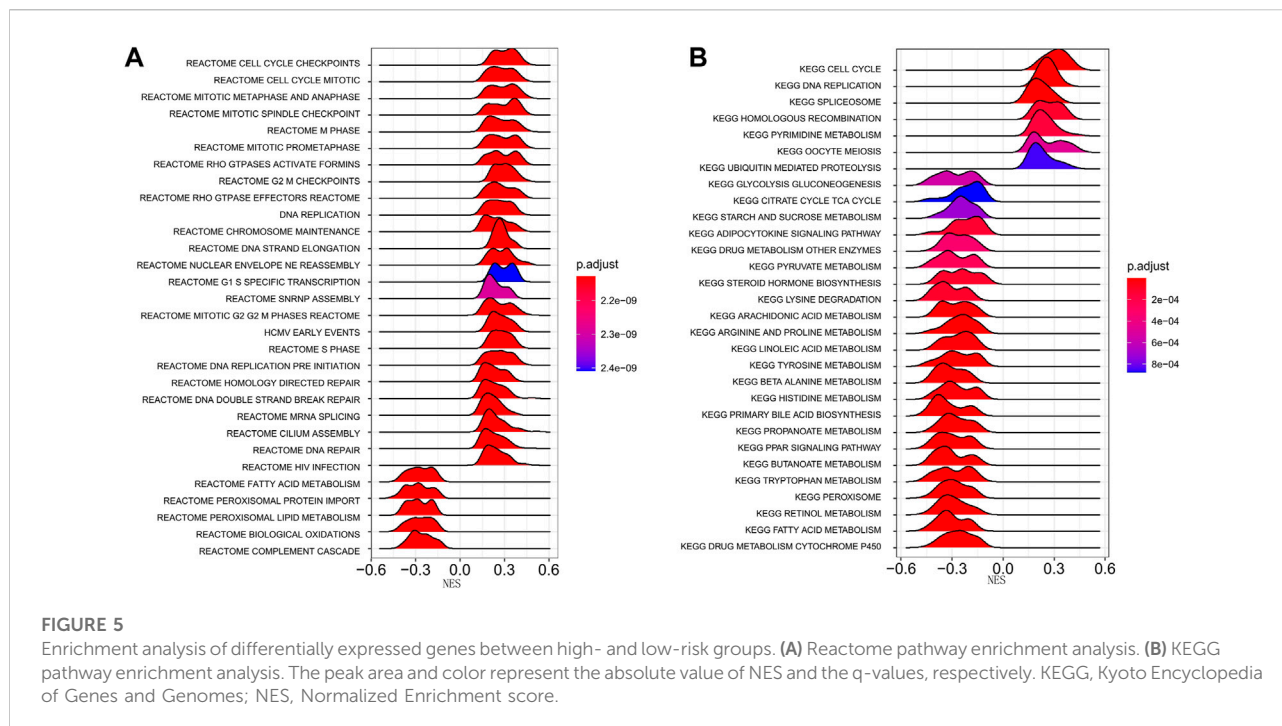
**FIGURE 4**
Validation of a nomogram-integrated independent predictive factors. **(A)** Nomogram with a combination of risk scores and different clinical features. A calibration plot for predicting the accuracy of the nomogram in **(B)** 1-, **(C)** 3-, **(D)** and 5-year. **(E–G)** 1-, 3-, and 5-year DCA analysis of the nomogram. **(H)** ROC curves of the nomogram for predicting the 1/3/5-year overall survival in TCGA. **(I)** Multivariate Cox regression analyses of the risk score and clinicopathological factors for overall survival in TCGA. DCA, decision curve analysis; ROC, receiver operating characteristic; TCGA, The Cancer Genome Atlas.

## Genomic profile related to the ECM prognostic signature

There are two figures presenting the top 20 frequently mutated genes in the high-risk and low-risk groups (Figures 7A,B). The prevalence of *TP53*, *MYO18B*, *JARID2*, and *HUWE1* alterations in the high-risk group was significantly higher than those in the low-risk group; on the contrary, *HTT*, *PIK3CA*, and *LRRC7* mutations were more enriched in the low-risk group (*p* < 0.01, Figure 7C). As shown in Figure 7D, *TP53* alterations were mainly located in the P53 domain, and the high-risk group had more mutations than the low-risk group in this domain. In addition, the low-risk group had significantly high frequencies of amplifications and deletions in chromosome 11 and 13,

**FIGURE 5**
Enrichment analysis of differentially expressed genes between high- and low-risk groups. **(A)** Reactome pathway enrichment analysis. **(B)** KEGG pathway enrichment analysis. The peak area and color represent the absolute value of NES and the q-values, respectively. KEGG, Kyoto Encyclopedia of Genes and Genomes; NES, Normalized Enrichment score.

respectively (FDR<0.01, Figures 7E,F). There was no significant difference in the TMB value between the two risk groups ($p = 1$, Figure 7G).
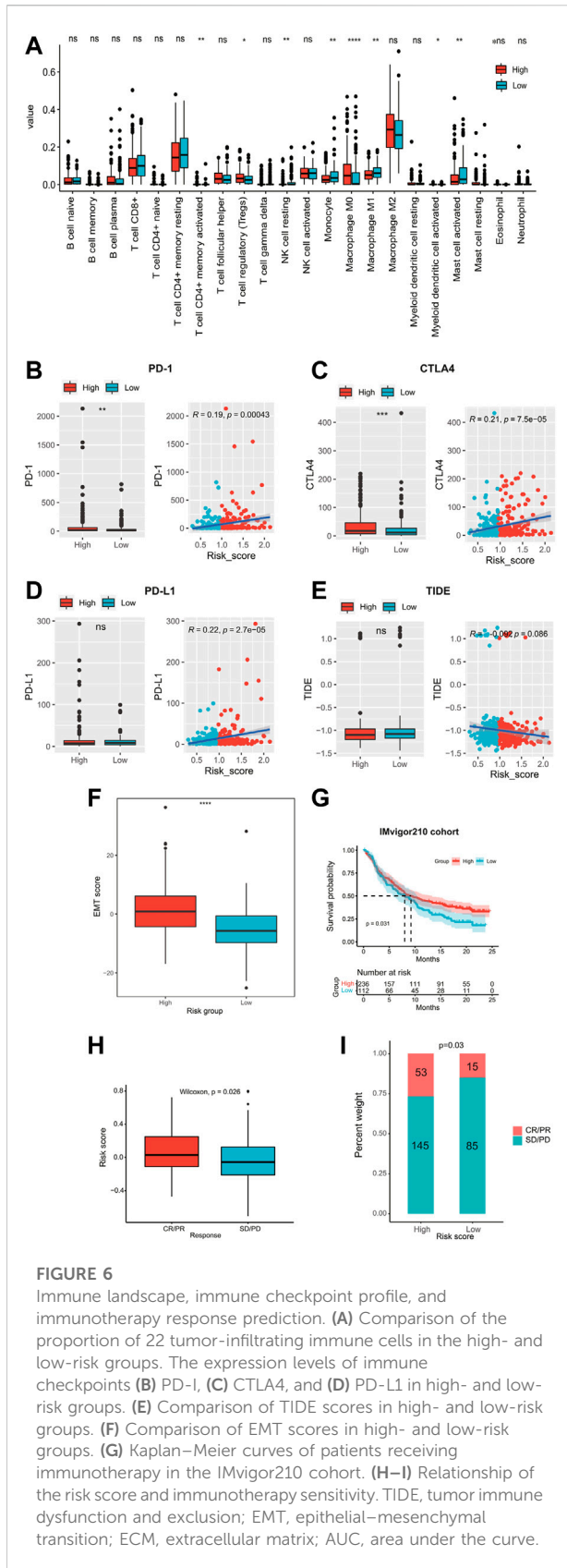
## Chemotherapy response prediction

The results of drug sensitivity analysis indicated that the patients in the high-risk group were more sensitive to commonly used drugs such as sorafenib and cisplatin ($p < 0.0001$, Figures 8A,B), but more resistant to doxorubicin ($p < 0.01$, Figure 8C). Meanwhile, there was no significant difference in sensitivity to gemcitabine between the two risk groups (Figure 8D). In addition, we assessed whether the ECM signature could predict the TACE benefit for HCC patients. The boxplots showed that non-responders of TACE had significantly higher risk scores than responders (Figure 8E). We also assessed the distribution of non-responders and responders in different risk groups (Figure 8F) and found that patients in the low-risk group had higher sensitivity to TACE than those in the high-risk group. The AUC of the ECMs signature to predict response to TACE in HCC patients was 0.72 (Figure 8G).

## Discussion

The resistance of HCC to traditional treatments and high tumor recurrence after therapy are pivotal causes of cancer-related deaths (Liu et al., 2020); among this process, CSCs play

key roles through genetic mutations, dysregulation of signaling pathways, epigenetic disruption, and/or TME regulation (Liu et al., 2020). In this study, by analyzing the characteristics of CSCs in HCC, we found that ECM is closely related to the occurrence and development of HCC. The ECM is the major structural component of the TME and is a highly dynamic structure (Nallanthighal et al., 2019). ECM is both a structural scaffold and a regulator of cell signal transduction in tissues (Dalton and Lemmon, 2021). Research in recent years has mainly focused on uncovering the cellular signal transduction mechanisms involved in the development of HCC. Gene modification plays an important role in tumorigenesis, and anomalies of many related signal transduction pathways, such as MAPK pathway, PI3K/AKT/mTOR pathway, WNT/β-catenin pathway, and JAK/STAT pathway, are also related to the progression of HCC (Mir et al., 2021). A previous study has supported that ECM-related proteins establishes a physical and biochemical niche for CSCs, providing structural and biochemical support for regulating CSCs proliferation, self-renewal, and differentiation (Nallanthighal et al., 2019). A total of 10 key ECM-related genes were obtained from ECM-related pathways to construct a novel prognostic prediction signature, including *TRAPPC4*, *RSU1*, *ILK*, *LAMA1*, *LAMB1*, *FLNC*, *ITGAV*, *AGRN*, *ARHGEF6*, and *LIMS2*. Previous studies have demonstrated that these 10 ECM-related genes are associated with the prognosis of various types of cancer, especially HCC. In published studies, *AGRN* (Zou et al., 2019), *ITGAV* (Zhang et al., 2019; Weiler et al., 2020), *FLNC* (Qi et al., 2016; Yang et al., 2017), *ILK* (Chan et al., 2011), and
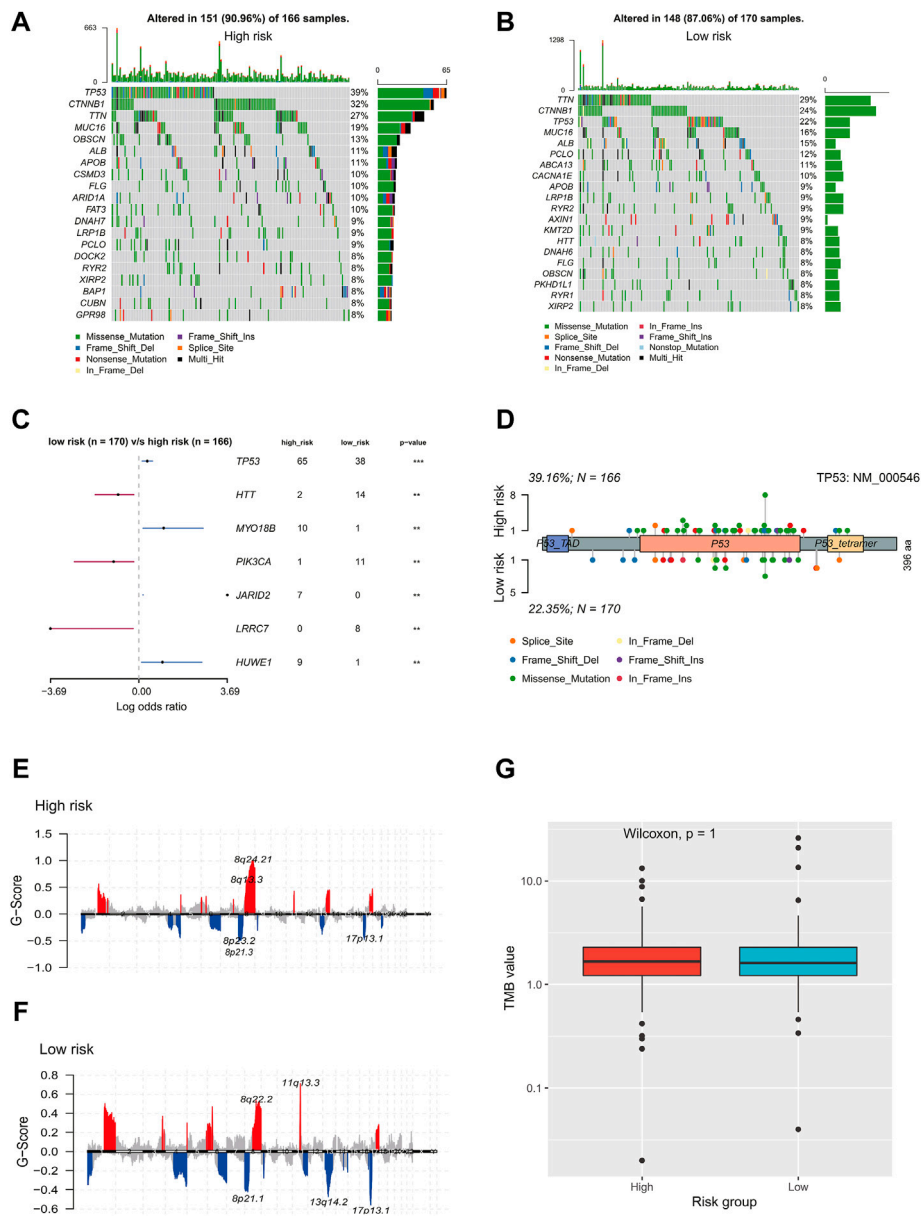
**FIGURE 6**
Immune landscape, immune checkpoint profile, and immunotherapy response prediction. **(A)** Comparison of the proportion of 22 tumor-infiltrating immune cells in the high- and low-risk groups. The expression levels of immune checkpoints **(B)** PD-I, **(C)** CTLA4, and **(D)** PD-L1 in high- and low-risk groups. **(E)** Comparison of TIDE scores in high- and low-risk groups. **(F)** Comparison of EMT scores in high- and low-risk groups. **(G)** Kaplan−Meier curves of patients receiving immunotherapy in the IMvigor210 cohort. **(H−I)** Relationship of the risk score and immunotherapy sensitivity. TIDE, tumor immune dysfunction and exclusion; EMT, epithelial−mesenchymal transition; ECM, extracellular matrix; AUC, area under the curve.

*RSU1* (Gkretsi and Bogdanos, 2015) were overexpressed in HCC tumor tissues, thereby promoting tumor development, metastasis, and even leading to poor prognosis, which is consistent with our findings. These genes were more upregulated in the high-risk group, which was responsible for causing the inferior outcomes of patients with the high-risk scores. When these genes were knocked out or downregulated, the proliferation and metastasis of HCC cells could be successfully inhibited, which indicates that they also have the potential to be developed as new therapeutic targets (Yang et al., 2017; Zou et al., 2019). Contrary to our findings, studies have shown that *TRAPPC4* is expressed at low levels in HCC tissues, and HCC patients with low *TRAPPC4* expression have a shorter survival time than those with high expression (Shen et al., 2018). This difference may be due to the tumor heterogeneity of HCC. In addition, *LIMS2* and *ARHGEF6* were expressed at lower levels in tumors than in normal tissues in our study and associated with better clinical outcomes. The dysregulation of *LIMS2*, also known as *PINCH2*, caused liver enlargement and tumorigenesis (Donthamsetty et al., 2013).

Subsequently, we explored the probability of the ECM signature in clinical application. By evaluating the relationship between clinical features and the risk score, we found that ECM signature was closely associated with AFP, vascular invasion, T stage, and N stage. AFP is the most prevalent clinically applied biomarker for the detection and treatment monitor of HCC, associated with HCC differentiation (Tangkijvanich et al., 2000). Previous studies found that the components of ECM, such as fibroblasts, would stimulate the paracrine and upregulate the AFP expression level (Ogunwobi et al., 2019). Meanwhile, vascular invasion has been reported not only to be present in about 50% of HCC but also to be a major risk factor for disease recurrence, associated with shorter survival in HCC, which is consistent with our results (Krishnan et al., 2021). The positive correlation between ECM and vascular invasion may be mainly attributed to the reason that ECM provides an essential scaffold supporting the vascular endothelium (Davis and Senger, 2005). Meanwhile, ECM induces tumor cell to transfer to an endothelial-like phenotype, imitating the vasculature that connect to blood vessels; and on the other hand, hypoxic tumor microenvironment also facilitates the ECM to release VEGFR and further angiogenic events (Winkler et al., 2020).

In recent years, immune-based therapies have revolutionized the systematic management of advanced cancers (Whiteside et al., 2016). The application of immune checkpoint inhibitor (ICI) therapy targeting PD-1, PD-L1, or CTLA-4 represents a major breakthrough for many types of cancers, including HCC (Hoos, 2016). However, the objective response rate of these agents as monotherapy for HCC is only 15%–20% (Cheng et al., 2020). Previously identified biomarkers such as PD-L1 expression and tumor mutation burden cannot reliably predict the benefit of ICI therapy (Mushtaq et al., 2018). Therefore,

**FIGURE 7**
Genomic landscape of HCC patients in the high- and low-risk group. Oncoprints showed the most prevalent alter genes in **(A)** high-risk and **(B)** low-risk. **(C)** Significantly different mutated genes between low- and high-risk groups. **(D)** Distribution of TP53 alterations in low- and high-risk groups. **(E,F)** Detection and comparison of significant amplifications and deletions of copy number between high- and low-risk groups. **(G)** Difference in the TMB values between low- and high-risk groups. HCC, hepatocellular carcinoma; TMB: tumor mutation burden.

strategies to improve their efficacy through patient stratification, or selection of potential combinations are still urgently needed. The composition of the tumor microenvironment has been shown to influence ICI response (Petitprez et al., 2020). In this study, we found that HCC patients with high-risk scores have higher sensitivity to ICIs. However, there was no significant difference in tumor mutation counts or in T cell function (revealed by the TIDE algorism) between high- and low-risk

groups. Then this correlation between the risk score and ICI response may be interpreted as the positive correlation between the risk score and the expression of immune checkpoint genes, including PD-1 and CTLA4. Previous studies have shown that tumor-infiltrating lymphocytes, such as Treg cells and tumor-associated macrophages, play an important role in driving immune evasion, which in turn drives HCC progression and affects patient treatments (Lurje et al., 2021). In our study, there
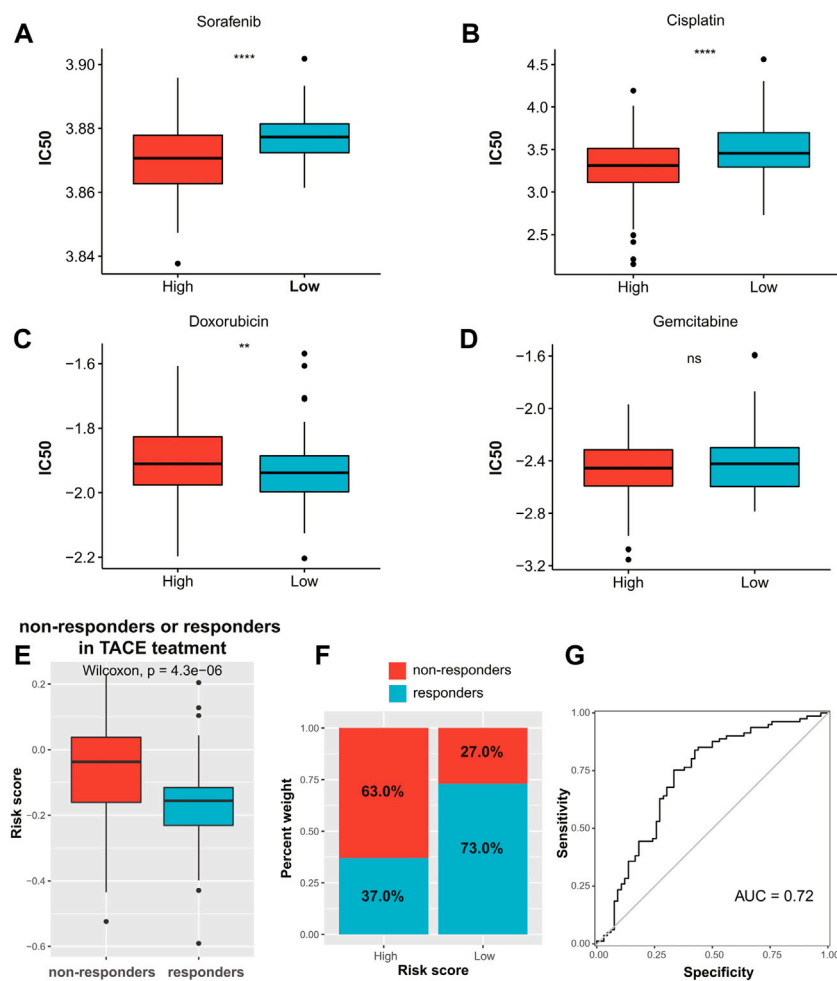
**FIGURE 8**
Chemotherapy response prediction. The boxplots of the estimated $IC_{50}$ for **(A)** sorafenib, **(B)** cisplatin, **(C)** doxorubicin, and **(D)** gemcitabine in the high- and low-risk groups. **(E)** Distribution of risk scores among TACE treatment responders and non-responders in the GSE104580 cohort. **(F)** Proportion of TACE treatment responders or non-responders in the high- and low-risk groups in the GSE104580 cohort. **(G)** ROC curve of ECM-related signature in predicting TACE response in the GSE104580 cohort. $IC_{50}$, half-maximal inhibitory concentration; TACE, transarterial chemoembolization; ROC, receiver operating characteristic; ECM, extracellular matrix.

were significant differences in the abundance of immune cell infiltration between the two groups of patients, including $CD4^+$ T cells, Treg cells, macrophages M0, and myeloid dendritic cells. It has been reported that the number of Tregs in tumor tissue or peripheral blood of HCC patients is increased compared with healthy individuals (Oura et al., 2021), and Tregs are associated with poorer median OS (Lim et al., 2019). Both macrophages and myeloid dendritic cells belong to myeloid cells, which are important components of tumor tissues and key regulators of the immune environment (Wu et al., 2020). They promote tumor development and are associated with prognosis in HCC patients (Wu et al., 2020). These findings explain the poor prognosis of the high-risk group in our study from the perspective of the immune environment. This shows the potential of our signature

in predicting the HCC tumor immune microenvironment, which may be beneficial for immunotherapy of this malignancy. Different numbers, phenotypes, and localization of tumor-infiltrating lymphocytes are not only key regulators of disease progression but also potential biomarkers for predicting immunotherapy response (Maibach et al., 2020). Previous studies have shown that higher levels of immune cell infiltration are positively associated with immunotherapy response in multiple tumor types (Karn et al., 2017; Kümpers et al., 2019). Interestingly, the high-risk group not only had higher levels of immune cell infiltration in our study but also showed a better response to immunotherapy. Therefore, ECMs signature had the potency for assisting oncologists to decide which patients are likely to respond to ICI in order to take the best course of treatment.

In conclusion, this study developed a prognostic signature composed of 10 ECM-related genes that can accurately and robustly predict the prognosis of HCC. These genes have the potential to be developed as therapeutic targets for HCC. Also, this ECMs signature has an important value for HCC in the selection of treatment.

## Data availability statement

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

## Author contributions

LC: data collection and drafting the article, DZ: drafting and revision, SZ: data analysis, XL: data collection, PG: designing this work and revision. All authors reviewed and approved the final manuscript.

## Acknowledgments

## Conflict of interest

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

## Publisher's note

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.959834/full#supplementary-material

**SUPPLEMENTARY FIGURE S1**
Schematic illustration of study's design. This study is mainly divided into three parts: first, identifying ECM-related genes; second, constructing a prognostic signature based on ECM-related genes; and finally analyzing the multi-omics features underlying the signature. ECM, extracellular matrix.

**SUPPLEMENTARY FIGURE S2**
(A) Correlation analysis of 75 ECM-related genes with stemness. (B) Correlation analysis of 10 selected ECM-related genes and MMP family. (C) Correlation analysis of 10 selected ECM-related genes and collagen or non-collagen. ECM, extracellular matrix; MMP, matrix metalloproteinase.

**SUPPLEMENTARY FIGURE S3**
Performance comparison of ECM-related prognostic prediction signature and stemness-related prognostic prediction signatures. (A) LASSO regression analysis showed the partial likelihood deviation curve of the minimum number genes corresponding to the covariates. (B) Kaplan–Meier curves showing different overall survival of patients in high- and low-risk groups. (C) ROC curves of stemness-related signature for predicting the 1/3/5-year overall survival. ECM, extracellular matrix; LASSO, least absolute shrinkage and selection operator; ROC, receiver operating characteristic.

**SUPPLEMENTARY FIGURE S4**
Correlation analysis of the tumor doubling time and the risk score.

## References

Bresnahan, E., Ramadori, P., Heikenwalder, M., Zender, L., and Lujambio, A. (2020). Novel patient-derived preclinical models of liver cancer. *J. Hepatol.* 72 (2), 239–249. doi:10.1016/j.jhep.2019.09.028

Cabral-Pacheco, G. A., Garza-Veloz, I., Castruita-De la Rosa, C., Ramirez-Acuña, J. M., Perez-Romero, B. A., Guerrero-Rodriguez, J. F., et al. (2020). The roles of matrix metalloproteinases and their inhibitors in human diseases. *Int. J. Mol. Sci.* 21 (24), E9739. doi:10.3390/ijms21249739

Carloni, V., Luong, T., and Rombouts, K. (2014). Hepatic stellate cells and extracellular matrix in hepatocellular carcinoma: More complicated than ever. *Liver Int.* 34 (6), 834–843. doi:10.1111/liv.12465

Chan, J., Ko, F. C., Yeung, Y. S., Ng, I. O., and Yam, J. W. (2011). Integrin-linked kinase overexpression and its oncogenic role in promoting tumorigenicity of hepatocellular carcinoma. *PloS one* 6 (2), e16984. doi:10.1371/journal.pone.0016984

Chen, D., Liu, J., Zang, L., Xiao, T., Zhang, X., Li, Z., et al. (2022). Integrated machine learning and bioinformatic analyses constructed a novel stemness-related classifier to predict prognosis and immunotherapy responses for hepatocellular carcinoma patients. *Int. J. Biol. Sci.* 18 (1), 360–373. doi:10.7150/ijbs.66913

Cheng, A. L., Hsu, C., Chan, S. L., Choo, S. P., and Kudo, M. (2020). Challenges of combination therapy with immune checkpoint inhibitors for hepatocellular carcinoma. *J. Hepatol.* 72 (2), 307–319. doi:10.1016/j.jhep.2019.09.025

Dalton, C. J., and Lemmon, C. A. (2021). Fibronectin: Molecular structure, fibrillar structure and mechanochemical signaling. *Cells* 10 (9), 2443. doi:10.3390/cells10092443

Davis, G., and Senger, D. (2005). Endothelial extracellular matrix: Biosynthesis, remodeling, and functions during vascular morphogenesis and neovessel stabilization. *Circ. Res.* 97 (11), 1093–1107. doi:10.1161/01.RES.0000191547.64391.e3

Donthamsetty, S., Bhave, V. S., Mars, W. M., Bowen, W. C., Orr, A., Haynes, M. M., et al. (2013). Role of PINCH and its partner tumor suppressor Rsu-1 in regulating liver size and tumorigenesis. *PloS one* 8 (9), e74625. doi:10.1371/journal.pone.0074625

Eppert, K., Takenaka, K., Lechman, E. R., Waldron, L., Nilsson, B., van Galen, P., et al. (2011). Stem cell gene expression programs influence clinical outcome in human leukemia. *Nat. Med.* 17 (9), 1086–1093. doi:10.1038/nm.2415

Forner, A., Reig, M., and Bruix, J. (2018). Hepatocellular carcinoma. *Lancet* 391 (10127), 1301–1314. doi:10.1016/S0140-6736(18)30010-2

Gkretsi, V., and Bogdanos, D. P. (2015). Elimination of Ras Suppressor-1 from hepatocellular carcinoma cells hinders their *in vitro* metastatic properties. *Anticancer Res.* 35 (3), 1509–1512.

Gui, J., and Li, H. (2005). Penalized Cox regression analysis in the high-dimensional and low-sample size settings, with applications to microarray gene

expression data. *Bioinforma. Oxf. Engl.* 21 (13), 3001–3008. doi:10.1093/bioinformatics/bti422

Hong, L., Zhou, Y., Xie, X., Wu, W., Shi, C., Lin, H., et al. (2021). A stemness-based eleven-gene signature correlates with the clinical outcome of hepatocellular carcinoma. *BMC cancer* 21 (1), 716. doi:10.1186/s12885-021-08351-0

Hoos, A. (2016). Development of immuno-oncology drugs - from CTLA4 to PD1 to the next generations. *Nat. Rev. Drug Discov.* 15 (4), 235–247. doi:10.1038/nrd.2015.35

Jiang, P., Gu, S., Pan, D., Fu, J., Sahu, A., Hu, X., et al. (2018). Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* 24 (10), 1550–1558. doi:10.1038/s41591-018-0136-1

Karn, T., Jiang, T., Hatzis, C., Sänger, N., El-Balat, A., Rody, A., et al. (2017). Association between genomic metrics and immune infiltration in triple-negative breast cancer. *JAMA Oncol.* 3 (12), 1707–1711. doi:10.1001/jamaoncol.2017.2140

Kreso, A., and Dick, J. E. (2014). Evolution of the cancer stem cell model. *Cell Stem Cell* 14 (3), 275–291. doi:10.1016/j.stem.2014.02.006

Krishnan, M., Rajan, Kd A., Park, J., Arjunan, V., Garcia Marques, F., Bermudez, A., et al. (2021). Genomic analysis of vascular invasion in HCC reveals molecular drivers and predictive biomarkers. *Hepatol. Baltim. Md)* 73 (6), 2342–2360. doi:10.1002/hep.31614

Kümpers, C., Jokic, M., Haase, O., Offermann, A., Vogel, W., Grätz, V., et al. (2019). Immune cell infiltration of the primary tumor, not PD-L1 status, is associated with improved response to checkpoint inhibition in metastatic melanoma. *Front. Med.* 6, 27. doi:10.3389/fmed.2019.00027

Lim, C. J., Lee, Y. H., Pan, L., Lai, L., Chua, C., Wasser, M., et al. (2019). Multidimensional analyses reveal distinct immune microenvironment in Hepatitis B virus-related hepatocellular carcinoma. *Gut* 68 (5), 916–927. doi:10.1136/gutjnl-2018-316510

Liu, P., Zhou, Q., and Li, J. (2022). Integrated multi-omics data analysis reveals associations between glycosylation and stemness in hepatocellular carcinoma. *Front. Oncol.* 12, 913432. doi:10.3389/fonc.2022.913432

Liu, Y. C., Yeh, C. T., and Lin, K. H. (2020). Cancer stem cell functions in hepatocellular carcinoma and comprehensive therapeutic strategies. *Cells* 9 (6), E1331. doi:10.3390/cells9061331

Lurje, I., Werner, W., Mohr, R., Roderburg, C., Tacke, F., and Hammerich, L. (2021). *In situ* vaccination as a strategy to modulate the immune microenvironment of hepatocellular carcinoma. *Front. Immunol.* 12, 650486. doi:10.3389/fimmu.2021.650486

Maibach, F., Sadozai, H., Seyed Jafari, S. M., Hunger, R. E., and Schenk, M. (2020). Tumor-infiltrating lymphocytes and their prognostic value in cutaneous melanoma. *Front. Immunol.* 11, 2105. doi:10.3389/fimmu.2020.02105

Malta, T., Sokolov, A., Gentles, A., Burzykowski, T., Poisson, L., Weinstein, J., et al. (2018). Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell* 173 (2), 338–354.e15. e315. doi:10.1016/j.cell.2018.03.034

Mir, I. H., Guha, S., Behera, J., and Thirunavukkarasu, C. (2021). Targeting molecular signal transduction pathways in hepatocellular carcinoma and its implications for cancer therapy. *Cell Biol. Int.* 45 (11), 2161–2177. doi:10.1002/cbin.11670

Morganella, S., Pagnotta, S. M., and Ceccarelli, M. (2011). Finding recurrent copy number alterations preserving within-sample homogeneity. *Bioinforma. Oxf. Engl.* 27 (21), 2949–2956. doi:10.1093/bioinformatics/btr488

Mushtaq, M. U., Papadas, A., Pagenkopf, A., Flietner, E., Morrow, Z., Chaudhary, S. G., et al. (2018). Tumor matrix remodeling and novel immunotherapies: The promise of matrix-derived immune biomarkers. *J. Immunother. Cancer* 6 (1), 65. doi:10.1186/s40425-018-0376-0

Nallanthighal, S., Heiserman, J. P., and Cheon, D. J. (2019). The role of the extracellular matrix in cancer stemness. *Front. Cell Dev. Biol.* 7, 86. doi:10.3389/fcell.2019.00086

Nassar, D., and Blanpain, C. (2016). Cancer stem cells: Basic concepts and therapeutic implications. *Annu. Rev. Pathol.* 11, 47–76. doi:10.1146/annurev-pathol-012615-044438

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12 (5), 453–457. doi:10.1038/nmeth.3337

Ogunwobi, O. O., Harricharran, T., Huaman, J., Galuza, A., Odumuwagun, O., Tan, Y., et al. (2019). Mechanisms of hepatocellular carcinoma progression. *World J. Gastroenterol.* 25 (19), 2279–2293. doi:10.3748/wjg.v25.i19.2279

Oura, K., Morishita, A., Tani, J., and Masaki, T. (2021). Tumor immune microenvironment and immunosuppressive therapy in hepatocellular carcinoma: A review. *Int. J. Mol. Sci.* 22 (11), 5801. doi:10.3390/ijms22115801

Pan, S., Zhan, Y., Chen, X., Wu, B., and Liu, B. (2019). Identification of biomarkers for controlling cancer stem cell characteristics in bladder cancer by network analysis of transcriptome data stemness indices. *Front. Oncol.* 9, 613. doi:10.3389/fonc.2019.00613

Peitzsch, C., Kurth, I., Kunz-Schughart, L., Baumann, M., and Dubrovska, A. (2013). Discovery of the cancer stem cell related determinants of radioresistance. *Radiother. Oncol.* 108 (3), 378–387. doi:10.1016/j.radonc.2013.06.003

Petitprez, F., Meylan, M., de Reyniès, A., Sautès-Fridman, C., and Fridman, W. H. (2020). The tumor microenvironment in the response to immune checkpoint blockade therapies. *Front. Immunol.* 11, 784. doi:10.3389/fimmu.2020.00784

Poltavets, V., Kochetkova, M., Pitson, S., and Samuel, M. (2018). The role of the extracellular matrix and its molecular and cellular regulators in cancer cell plasticity. *Front. Oncol.* 8, 431. doi:10.3389/fonc.2018.00431

Qi, Y., Xu, F., Chen, L., Li, Y., Xu, Z., Zhang, Y., et al. (2016). Quantitative proteomics reveals FLNC as a potential progression marker for the development of hepatocellular carcinoma. *Oncotarget* 7 (42), 68242–68252. doi:10.18632/oncotarget.11921

Qiu, Z. Q., Wang, X., Ji, X. W., Jiang, F. J., Han, X. Y., Zhang, W. L., et al. (2022). The clinical relevance of epithelial-mesenchymal transition and its correlations with tumorigenic immune infiltrates in hepatocellular carcinoma. *Immunology* 166, 185–196. doi:10.1111/imm.13465

Shen, J. D., Cai, X. F., Wei, J., Gu, C. Y., Ju, L. L., Wang, Y. F., et al. (2018). Low expression of trafficking protein particle complex 4 predicts poor prognosis in hepatocellular carcinoma. *Int. J. Clin. Exp. Pathol.* 11 (5), 2691–2698.

Shen, S., Wang, R., Qiu, H., Li, C., Wang, J., Xue, J., et al. (2021). Development of an autophagy-based and stemness-correlated prognostic model for hepatocellular carcinoma using bulk and single-cell RNA-sequencing. *Front. Cell Dev. Biol.* 9, 743910. doi:10.3389/fcell.2021.743910

Sia, D., Villanueva, A., Friedman, S., and Llovet, J. (2017). Liver cancer cell of origin, molecular class, and effects on patient prognosis. *Gastroenterology* 152 (4), 745–761. doi:10.1053/j.gastro.2016.11.048

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca. Cancer J. Clin.* 71 (3), 209–249. doi:10.3322/caac.21660

Tang, B., Zhu, J., Zhao, Z., Lu, C., Liu, S., Fang, S., et al. (2021). Based on tumor mutation burden to construct diagnosis and prognosis models and provide guidance for the treatment in hepatocellular carcinoma patients. *J. Adv. Res.*

Tangkijvanich, P., Anukulkarnkusol, N., Suwangool, P., Lertmaharit, S., Hanvivatvong, O., Kullavanijaya, P., et al. (2000). Clinical characteristics and prognosis of hepatocellular carcinoma: Analysis based on serum alpha-fetoprotein levels. *J. Clin. Gastroenterol.* 31 (4), 302–308. doi:10.1097/00004836-200012000-00007

Tian, R., Li, Y., Liu, Q., and Shu, M. (2021). Identification and validation of an immune-associated RNA-binding proteins signature to predict clinical outcomes and therapeutic responses in glioma patients. *Cancers (Basel)* 13 (7), 1730. doi:10.3390/cancers13071730

Tsui, Y. M., Chan, L. K., and Ng, I. O. (2020). Cancer stemness in hepatocellular carcinoma: Mechanisms and translational potential. *Br. J. Cancer* 122 (10), 1428–1440. doi:10.1038/s41416-020-0823-9

Villa, E., Critelli, R., Lei, B., Marzocchi, G., Cammà, C., Giannelli, G., et al. (2016). Neoangiogenesis-related genes are hallmarks of fast-growing hepatocellular carcinomas and worst survival. Results from a prospective study. *Gut* 65 (5), 861–869. doi:10.1136/gutjnl-2014-308483

Villanueva, A. (2019). Hepatocellular carcinoma. *N. Engl. J. Med.* 380 (15), 1450–1462. doi:10.1056/NEJMra1713263

Weiler, S. M. E., Lutz, T., Bissinger, M., Sticht, C., Knaub, M., Gretz, N., et al. (2020). TAZ target gene ITGAV regulates invasion and feeds back positively on YAP and TAZ in liver cancer cells. *Cancer Lett.* 473, 164–175. doi:10.1016/j.canlet.2019.12.044

Whiteside, T. L., Demaria, S., Rodriguez-Ruiz, M. E., Zarour, H. M., and Melero, I. (2016). Emerging opportunities and challenges in cancer immunotherapy. *Clin. Cancer Res.* 22 (8), 1845–1855. doi:10.1158/1078-0432.CCR-16-0049

Winkler, J., Abisoye-Ogunniyan, A., Metcalf, K. J., and Werb, Z. (2020). Concepts of extracellular matrix remodelling in tumour progression and metastasis. *Nat. Commun.* 11, 5120. doi:10.1038/s41467-020-18794-x

Wu, C., Lin, J., Weng, Y., Zeng, D. N., Xu, J., Luo, S., et al. (2020). Myeloid signature reveals immune contexture and predicts the prognosis of hepatocellular carcinoma. *J. Clin. Invest.* 130 (9), 4679–4693. doi:10.1172/JCI135048

Yang, B., Liu, Y., Zhao, J., Hei, K., Zhuang, H., Li, Q., et al. (2017). Ectopic overexpression of filamin C scaffolds MEK1/2 and ERK1/2 to promote the progression of human hepatocellular carcinoma. *Cancer Lett.* 388, 167–176. doi:10.1016/j.canlet.2016.11.037

Yang, J. D., Hainaut, P., Gores, G. J., Amadou, A., Plymoth, A., and Roberts, L. R. (2019). A global view of hepatocellular carcinoma: Trends, risk, prevention and management. *Nat. Rev. Gastroenterol. Hepatol.* 16 (10), 589–604. doi:10.1038/s41575-019-0186-y

Yi, L., Huang, P., Zou, X., Guo, L., Gu, Y., Wen, C., et al. (2020). Integrative stemness characteristics associated with prognosis and the immune microenvironment in esophageal cancer. *Pharmacol. Res.* 161, 105144. doi:10.1016/j.phrs.2020.105144

You, Y., Zheng, Q., Dong, Y., Xie, X., Wang, Y., Wu, S., et al. (2016). Matrix stiffness-mediated effects on stemness characteristics occurring in HCC cells. *Oncotarget* 7 (22), 32221–32231. doi:10.18632/oncotarget.8515

Zhang, L., Huang, Y., Ling, J., Zhuo, W., Yu, Z., Luo, Y., et al. (2019). Is integrin subunit alpha 2 expression a prognostic factor for liver carcinoma? A validation experiment based on bioinformatics analysis. *Pathol. Oncol. Res.* 25 (4), 1545–1552. doi:10.1007/s12253-018-0551-0

Zhang, Q., Lou, Y., Bai, X., and Liang, T. (2018). Immunometabolism: A novel perspective of liver cancer microenvironment and its influence on tumor progression. *World J. Gastroenterol.* 24 (31), 3500–3512. doi:10.3748/wjg.v24.i31.3500

Zhang, Q., Wang, J., Liu, M., Zhu, Q., Li, Q., Xie, C., et al. (2020). Weighted correlation gene network analysis reveals a new stemness index-related survival model for prognostic prediction in hepatocellular carcinoma. *Aging* 12 (13), 13502–13517. doi:10.18632/aging.103454

Zou, Q., Xiao, Z., Huang, R., Wang, X., Wang, X., Zhao, H., et al. (2019). Survey of the translation shifts in hepatocellular carcinoma with ribosome profiling. *Theranostics* 9 (14), 4141–4155. doi:10.7150/thno.35033

# Glossary

**HCC** hepatocellular carcinoma

**TCGA** The Cancer Genome Atlas

**GEO** Gene Expression Omnibus

**ICGC** International Cancer Genome Consortium

**mRNAsi** mRNA expression–based stemness index

**DEGs** differentially expressed genes

**ECM** extracellular matrix

**LASSO** least absolute shrinkage and selection operator

**ROC** receiver operating characteristic

**TACE** transarterial chemoembolization

**LC** liver cancer

**CSCs** cancer stem cells

**TME** tumor microenvironment

**TCGA-LIHC** The Cancer Genome Atlas Liver Hepatocellular Carcinoma

**OS** overall survival

**DFS** disease-free survival

**OCLR** one-class logistic regression

**GSEA** gene set enrichment analysis

**K-M** Kaplan–Meier

**AUC** area under the curve

**T stage** primary tumor stage

**N stage** regional lymph nodes stage

**M stage** distant metastasis stage

**KEGG** Kyoto Encyclopedia of Genes and Genomes

**TIDE** tumor immune dysfunction and exclusion

**EMT** Epithelial–mesenchymal transition

**CNV** copy number variant

**GDSC** Genomics of Drug Sensitivity in Cancer

**CR/PR** complete/partial responses

**SD/PD** stable/progressive disease

**DCA** decision curve analysis.