



# Extracellular Matrix-Based Gene Expression Signature Defines Two Prognostic Subtypes of Hepatocellular Carcinoma With Different Immune Microenvironment Characteristics

Hui Tang, Tingting You, Zhao Sun, Chunmei Bai and Yingyi Wang\*

Department of Medical Oncology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

## OPEN ACCESS

### Edited by:

Monika Vishnoi,  
Houston Methodist Research Institute,  
United States

### Reviewed by:

Jabed Iqbal,  
Singapore General Hospital,  
Singapore  
Wenzhi Guo,  
The First Affiliated Hospital of  
Zhengzhou University, China

### \*Correspondence:

Yingyi Wang  
wangyingyi@pumch.cn

### Specialty section:

This article was submitted to  
Molecular Diagnostics and  
Therapeutics,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 20 December 2021

**Accepted:** 31 January 2022

**Published:** 25 March 2022

### Citation:

Tang H, You T, Sun Z, Bai C and  
Wang Y (2022) Extracellular Matrix-  
Based Gene Expression Signature  
Defines Two Prognostic Subtypes of  
Hepatocellular Carcinoma With  
Different Immune  
Microenvironment Characteristics.  
*Front. Mol. Biosci.* 9:839806.  
doi: 10.3389/fmolb.2022.839806

**Background:** Accumulating evidence has suggested that the extracellular matrix (ECM) plays a vital role in the development and progression of cancer, and could be recognized as a biomarker of the response to immunotherapy. However, the effect of the ECM signature in hepatocellular carcinoma (HCC) is not well understood.

**Methods:** HCC patients derived from the TCGA-LIHC dataset were clustered according to the ECM signature. The differences in prognosis, functional enrichment, immune infiltration, and mutation characteristics between distinct molecular clusters were examined, and its predictive value on the sensitivities to chemotherapy and immunotherapy was further analyzed. Then, a prognostic model was built based on the ECM-related gene expression pattern.

**Results:** HCC patients were assigned into two molecular subtypes. Approximately 80% of HCC patients were classified into cluster A with poor prognosis, more frequent TP53 mutation, and lower response rate to immunotherapy. In contrast, patients in cluster B had better survival outcomes and higher infiltration levels of dendritic cells, macrophages, and regulatory T cells. The prognostic risk score model based on the expression profiles of six ECM-related genes (SPP1, ADAMTS5, MMP1, BSG, LAMA2, and CDH1) demonstrated a significant association with higher histologic grade and advanced TNM stage. Moreover, the prognostic risk score showed good performance in both the training dataset and validation dataset, as well as improved prognostic capacity compared with TNM stage.

**Conclusions:** We characterized two HCC subtypes with distinct clinical outcomes, immune infiltration, and mutation characteristics. A novel prognostic model based on the ECM signature was further developed, which may contribute to individualized prognostic prediction and aid in clinical decision-making.

**Keywords:** liver cancer, extracellular matrix, molecular subtype, tumor-infiltrating immune cell, risk score, prognostic evaluation, drug sensitivity

## INTRODUCTION

Hepatocellular carcinoma (HCC) accounts for approximately 80% of liver cancers and is the fourth leading cause of cancer-related death worldwide (Bray et al., 2018). Unfortunately, although surgery may effect a radical cure, as many as 70% of HCC patients would have tumor recurrence after surgery at 5 years (Villanueva, 2019). With an estimated survival time of approximately 1 year, sorafenib and lenvatinib remained the only effective systemic therapies for frontline therapy until immune checkpoint inhibitor (ICI)-based combination therapy showed desirable efficacy, but clinically available biomarkers to predict response to systemic therapies are still needed (Villanueva, 2019; Pinter et al., 2021). Therefore, it is urgent to further explore the underlying mechanisms of cancer development and detect novel prognostic and therapeutic targets of HCC.

Most biomarkers used for cancer classification and stratification are still “cancer-cell oriented” (such as TNM stage and tumor markers), leaving out key factors associated with cancer development, such as the tumor microenvironment (TME) and antitumor immunity (Lecchi et al., 2021). The TME is composed of three interrelated components, including stromal cells such as immune cells and fibroblasts, cytokines, and the extracellular matrix (ECM) (Théret et al., 2021). Previous studies have proven that HCC classification based on molecular features of immune infiltration could aid in prognosis evaluation and predicting response to ICIs (Sia et al., 2017; Kurebayashi et al., 2018). Accumulating evidence has suggested that the ECM is associated with tumor aggression, metastasis, treatment sensitivity, and prognosis (Deligne and Midwood, 2021). The ECM was also reported to not only provide a physical barrier, preventing interaction between immune effectors or drugs and tumor cells, but can also modulate immune cell proliferation, differentiation, motility, and activation (Zhang et al., 2021). However, the role of the ECM signature in HCC classification has not been estimated, although a dynamic ECM was associated with HCC carcinogenesis, progression, and prognosis (Jeng et al., 2015).

In the present research, we first classified patients from The Cancer Genome Atlas liver hepatocellular carcinoma (TCGA-LIHC) cohort into distinct molecular subtypes according to ECM-related gene expression. The relationships between molecular subtypes and clinicopathological features, prognosis, and drug sensitivity were further examined. Then, an ECM-related prognostic signature was developed and validated.

## MATERIALS AND METHODS

### Data Sources

Gene expression, somatic mutation, and corresponding clinicopathological data of 369 HCC samples were obtained from TCGA database (<https://portal.gdc.cancer.gov/>) in December 2021. One International Cancer Genome Consortium (ICGC) (<https://dcc.icgc.org/>) and two Gene Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/>

<https://www.ncbi.nlm.nih.gov/> geo/) HCC cohorts (there were 232, 80, and 81 HCC patients with complete follow-up information in ICGC-LIRI, GSE10141, and GSE76427 cohorts, respectively) were obtained for external validation. The gene expression values were transformed into transcripts per kilobase million (TPM) (Wagner et al., 2012). The mean value was reserved if one gene matched multiple probes. The ComBat method was used to remove the batch effects.

### Consensus Clustering Analysis

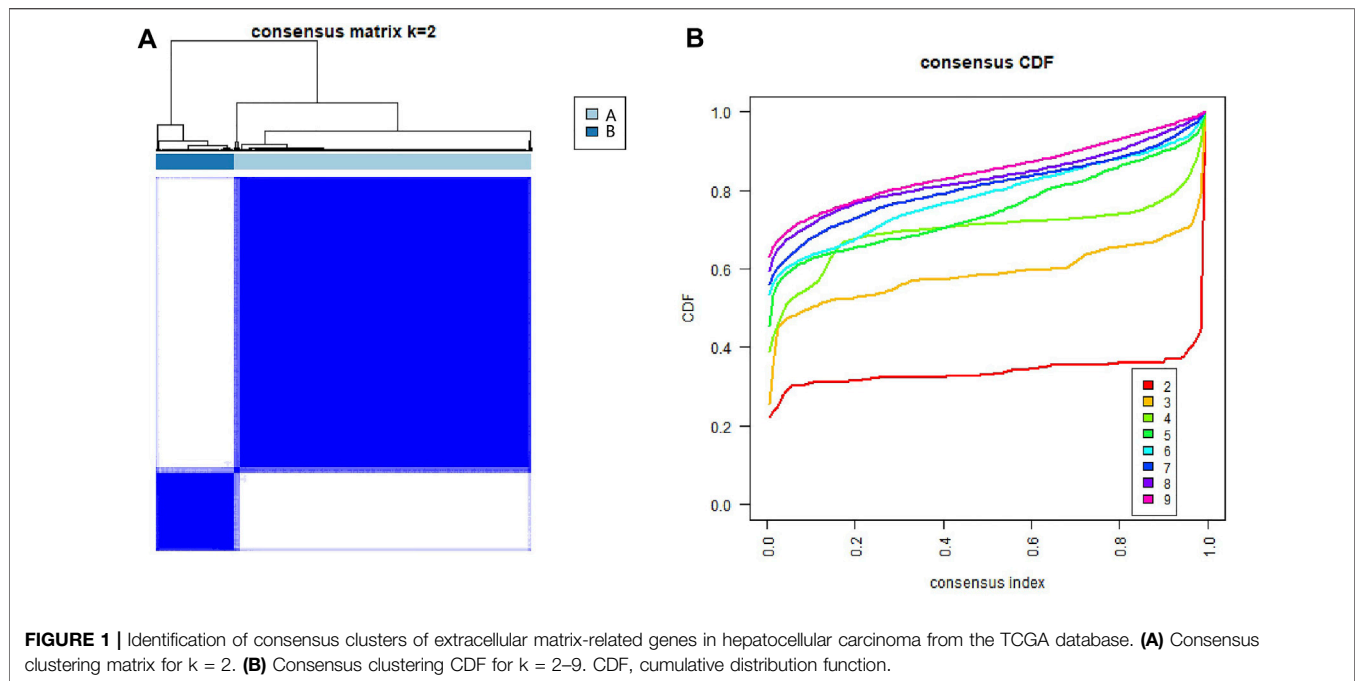
Three hundred and one ECM-related genes were retrieved from Reactome (<https://reactome.org/>) (Jassal et al., 2020). Consensus unsupervised clustering analysis was conducted by the R package “ConsensusClusterPlus” (Wilkerson and Hayes, 2010) (1,000 iterations, resample rate of 80%) to classify patients from the TCGA-LIHC cohort into distinct molecular subtypes according to ECM-related gene expression. This clustering was conducted, and the optimal cluster number was confirmed based on the following criteria: First, the cumulative distribution function (CDF) curve increased gradually and smoothly. Second, after clustering, the intracluster correlation increased, while the intercluster correlation decreased. Last, no clusters had a too small sample size.

### Exploring the Differences Between Distinct Molecular Clusters of HCC Patients

The differences in overall survival (OS) among different subtypes were assessed using Kaplan–Meier curves generated by the R packages “survival” and “survminer”. The top 20 highest mutational frequencies in the TCGA-LIHC cohort were recognized and visualized via the R package “maftools”. Tumor mutation burden (TMB) was defined as the total gene mutation number per million base pair, which was calculated using a Perl script and corrected by dividing by the total length of exons. The immune cell infiltration abundance data in the TCGA database were retrieved from xCell (<https://xcell.ucsf.edu/>) (Aran et al., 2017), and the CIBERSORT algorithm was used to evaluate the immune and stromal scores of each patient (Newman et al., 2015). xCell is a transcriptome-based method learned from thousands of pure cell types from various sources that performs cell type enrichment analysis from gene expression data for 64 immune cell types (Aran et al., 2017). Furthermore, gene set enrichment analysis (GSEA) was conducted to investigate the functional enrichment of differentially expressed genes in distinct molecular clusters using the R packages “limma”, “org.Hs.eg.db”, “clusterProfiler”, and “enrichplot”. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) gene sets (c5.all.v7.1 and c2.cp.kegg.v7.1) were downloaded from the MSigDB database (<https://www.gsea-msigdb.org/gsea/index.jsp>).

### Drug Sensitivity Analysis

The half-maximal inhibitory concentration (IC<sub>50</sub>) of several chemotherapy or targeted drugs in each HCC sample from TCGA database was estimated via Genomics of Drug Sensitivity in Cancer (GDSC; <https://www.cancerrxgene.org/>) (Yang et al., 2013) using the R package “pRRophetic”



(Geeleher et al., 2014), which is extensively utilized in studies evaluating drug sensitivity in cancers (Xiao et al., 2021; Liu et al., 2021a). Each sample's IC50 value was evaluated by ridge regression, and 10-fold cross-validation was used to ensure prediction accuracy based on the GDSC training model. The Tumor Immune Dysfunction and Exclusion (TIDE) (<https://tide.dfci.harvard.edu/>) algorithm was implemented to predict the ICI therapy response of each patient (Jiang et al., 2018). TIDE is a prevalent algorithm used to assess immune evasion mechanism and predict the immunotherapeutic response (Jiang et al., 2018; Liu et al., 2021b). TIDE scores and estimated immunotherapeutic responses were obtained after uploading the input data as described in the instructions, and higher TIDE scores mean a lower likelihood of response to immunotherapy.

### Construction of an ECM-Related Prognostic Signature

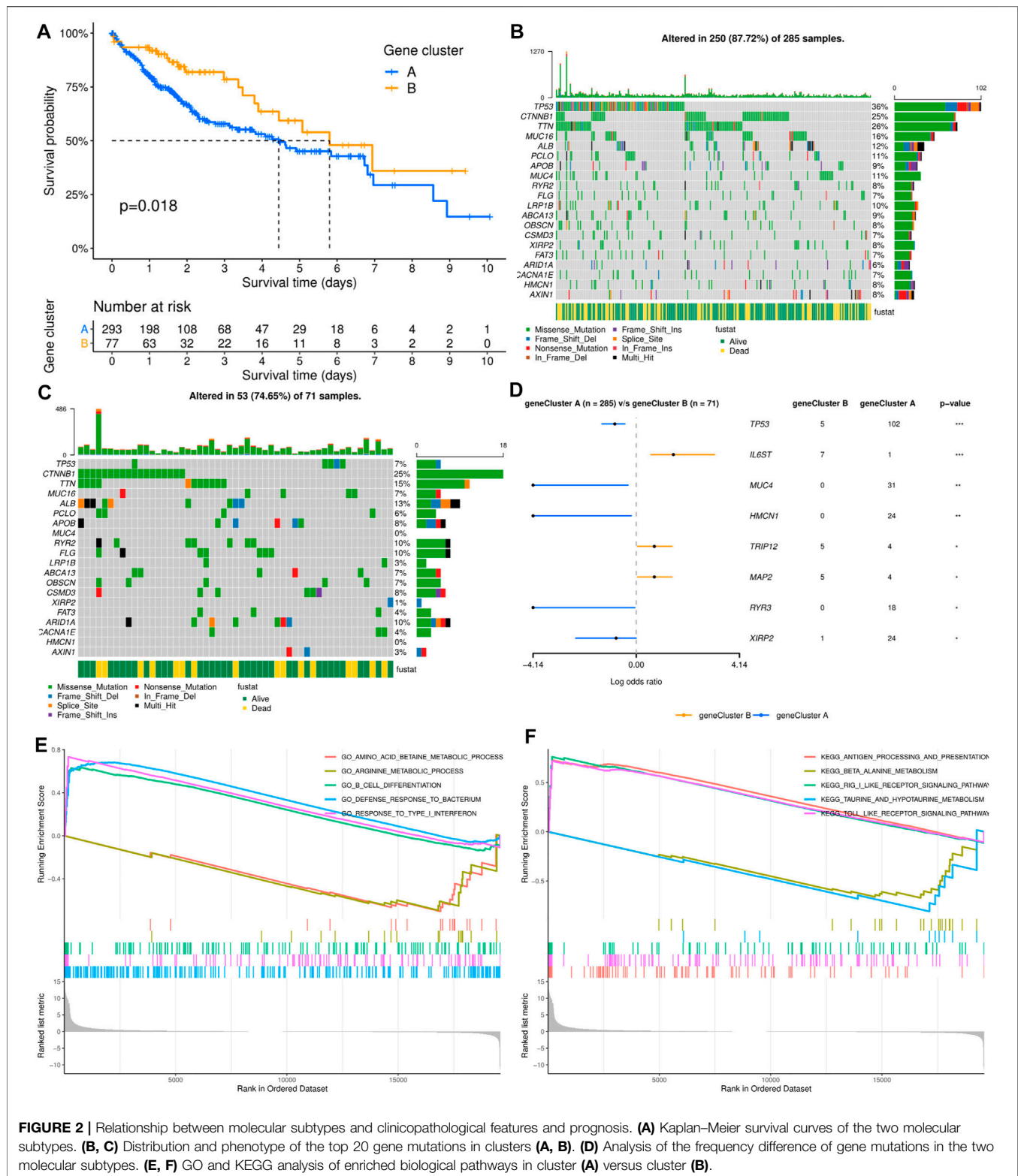
Least absolute shrinkage and selection operator (LASSO) regression analysis with the R package “glmnet” was used to identify the candidate genes of the ECM-related prognostic signature. Based on the common genes and corresponding expression profiles in the four independent cohorts mentioned above, six ECM-related genes were selected to build the risk signature according to the optimal lambda value and the corresponding coefficients using the TCGA-LIHC dataset. The risk score of the prognostic signature for each patient was calculated as follows: risk score = (exp Gene1 × coef Gene1) + (exp Gene2 × coef Gene2) + ... + (exp GeneN × coef GeneN).

### Validation of the ECM-Related Prognostic Signature

Patients were divided into two risk groups (low/high) according to their risk score and based on the optimal cutoff value of OS calculated by the R package “survminer”. Propensity-score matching (PSM) was used in the TCGA-LIHC dataset to minimize the impact of confounding factors. The propensity scores were calculated by the R package “MatchIt” using multivariable logistic regression based on age, TNM stage, histologic grade, Child–Pugh grade, vascular invasion, alpha fetoprotein, and residual tumor after surgery. Then, the risk scores of each patient in three validation cohorts were calculated and divided using the same formula and optimal cutoff value. The OS of the two risk score groups was compared using Kaplan–Meier curves. Univariate and multivariate Cox proportional hazard models were performed by the R package “finalfit” to calculate the hazard ratios (HR) with 95% confidence intervals (CI) of variables associated with OS in HCC patients. Based on the results of multivariate Cox analysis, a nomogram was constructed by the R package “rms” and assessed by the receiver operating characteristic (ROC) curve, concordance index (C-index), and calibration curves. Ultimately, decision curve analysis (DCA) was performed by the R package “rmda” to compare the prognostic capacity of the nomogram model and TNM stage.

### Statistical Analysis

The Mann–Whitney  $U$  test and Pearson's chi-square test or Fisher's exact test were applied to the comparison of the difference between the continuous and categorical variables

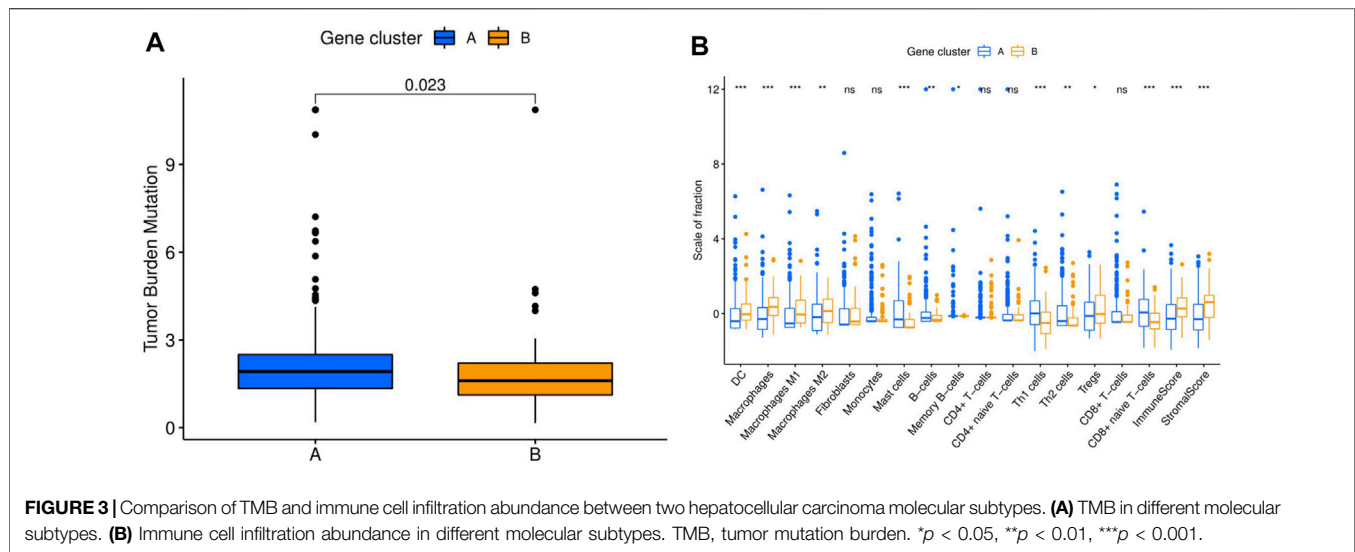


**FIGURE 2 |** Relationship between molecular subtypes and clinicopathological features and prognosis. **(A)** Kaplan–Meier survival curves of the two molecular subtypes. **(B, C)** Distribution and phenotype of the top 20 gene mutations in clusters **(A, B)**. **(D)** Analysis of the frequency difference of gene mutations in the two molecular subtypes. **(E, F)** GO and KEGG analysis of enriched biological pathways in cluster **(A)** versus cluster **(B)**.

between two groups, respectively. Strawberry Perl (version 5.30.0, <https://strawberryp Perl.com/>) was used to extract gene expression and mutation data from downloaded datasets and transfer them into a data matrix for subsequent analyses. All statistical analyses

and visualization were performed using R software (version 3.6.1, <https://www.r-project.org/>). A two-tailed value of P or FDR (false discovery rate)  $q < 0.05$  was considered to be statistically significant.





## RESULT

### Consensus Clustering of ECM-Related Genes Identified two Clusters of HCC

To examine the roles of ECM-related genes in HCC, 371 patients with transcriptome data who were retrieved from the TCGA-LIHC dataset were categorized using a consensus clustering algorithm based on the expression profiles of 301 ECM-related genes. Unsupervised consensus clustering analysis indicated that  $k = 2$  was the optimal selection; thus, patients in the entire cohort were sorted into subtypes A ( $n = 294$ , 79.2%) and B ( $n = 77$ , 20.8%) (Figures 1A,B).

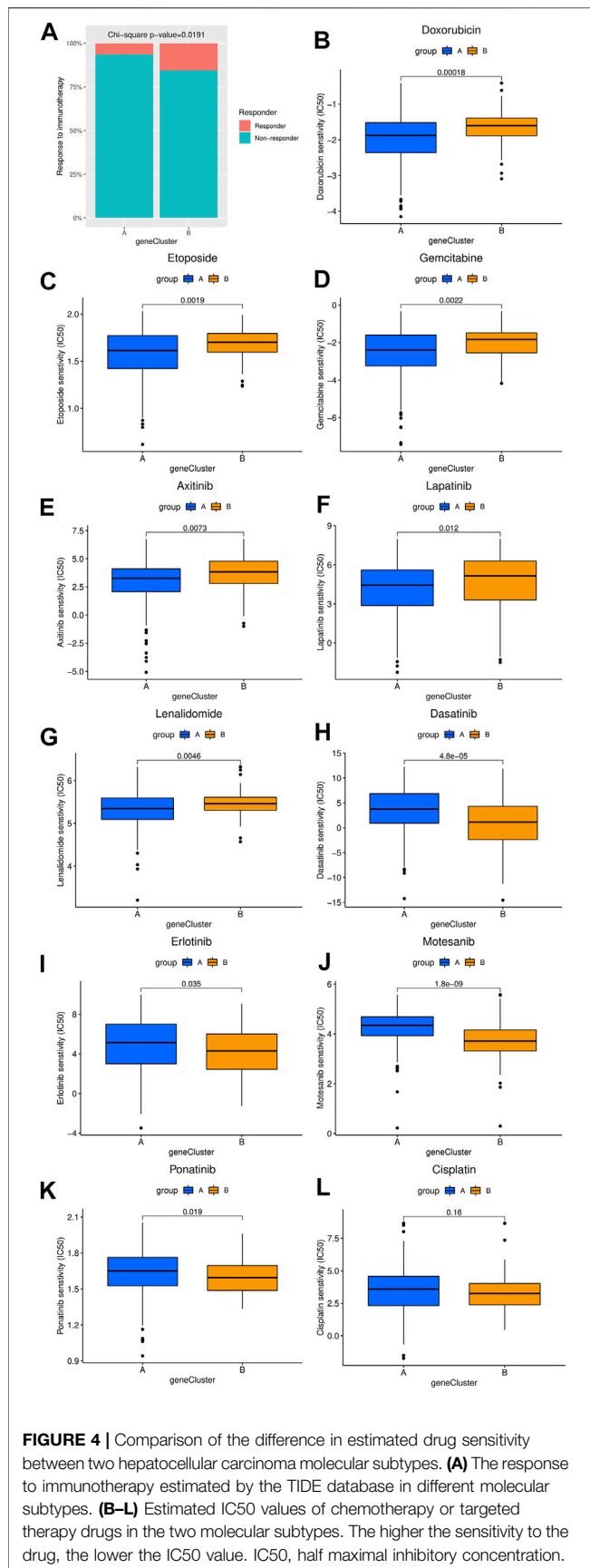
### Relationship Between Molecular Subtypes and Clinicopathological Features and Prognosis

As shown in Figure 2A, the results of survival analysis showed that the patients' prognosis was significantly different between the two subtypes. This suggested that there may be certain distinct differences apart from clinical outcomes between the two subtypes. However, there was no significant difference in clinical characteristics between the two molecular subtypes, including age, sex, TNM stage, histologic grade, Ishak score, Child–Pugh grade, and vascular invasion (Supplementary Table S1). To further explore the differences between the two molecular subgroups, three hundred and sixty-one HCC patients with somatic mutation data were included for mutational feature analysis. Mutation profile features indicated that missense mutation was the most common type in both molecular subtypes (Figures 2B,C). The top 20 mutated genes were visualized by a horizontal histogram, and the results showed that TP53 (35.8 vs. 7.0%), MUC4 (10.9 vs. 0%), XIRP2 (8.4 vs. 1.4%), HMCN1 (8.4 vs. 0%), and RYR3 (6.3 vs. 0%) mutations were more frequent in cluster A, while IL6ST (0.4 vs. 9.9%), TRIP12 (1.4 vs. 7.0%), and MAP2 (1.4 vs.

7.0%) mutations were more frequent in cluster B (Figure 2D). Furthermore, mutant CTNBN1, TTN, and ALB were the most common mutations in both subtypes.

To elucidate the potential influence of the ECM-related subtypes on the expression profiles of HCC, GSEA was applied to compare clusters A and B. Functional enrichment analysis showed that immune-related pathway terms (GO\_B\_CELL\_DIFFERENTIATION, GO\_DEFENSE\_RESPONSE\_TO\_BACTERIUM, GO\_RESPONSE\_TO\_TYPE\_I\_INTERFERON, KEGG\_ANTIGEN\_PROCESSING\_AND\_PRESENTATION, KEGG\_RIG\_I\_LIKE\_RECEPTOR\_SIGNALING\_PATHWAY, and KEGG\_TOLL\_LIKE\_RECEPTOR\_SIGNALING\_PATHWAY) were significantly enriched in cluster A, while some pathway terms relevant to amino acid metabolism were significantly enriched in cluster B (GO\_AMINO\_ACID\_BETAINE\_METABOLIC\_PROCESS, GO\_ARGININE\_METABOLIC\_PROCESS, KEGG\_BETA\_ALANINE\_METABOLISM, and KEGG\_TAURINE\_AND\_HYPOTAURINE\_METABOLISM) (Figures 2E,F).

ICI-based combination therapy is one of the most important breakthroughs in the treatment of HCC in recent years, while TMB and tumor-infiltrating immune cells may be potential biomarkers to predict the response to ICIs (Pinter et al., 2021). Therefore, we examined the differences in TMB and immune microenvironment between the two molecular subtypes. For TMB, the results showed that patients in cluster A had significantly higher TMB, but almost all HCC patients derived from TCGA database had low TMB (<10 mutations per megabase) (Figure 3A). In contrast, patients in cluster B had higher immune and stromal scores. Furthermore, the infiltration levels of dendritic cells (DCs), macrophages, M1 and M2 macrophages, and regulatory T cells (Tregs) were significantly higher in cluster B, while mast cells, B cells, memory B cells, Th1 and Th2 cells, and CD8<sup>+</sup> naive T cells had significantly lower infiltration in cluster B than in cluster A. In addition, there was no significant difference in the infiltration levels of fibroblasts, monocytes, CD4<sup>+</sup> T cells,



and CD8<sup>+</sup> T cells between the two molecular subtypes (**Figure 3B**).

## Drug Sensitivity Estimation

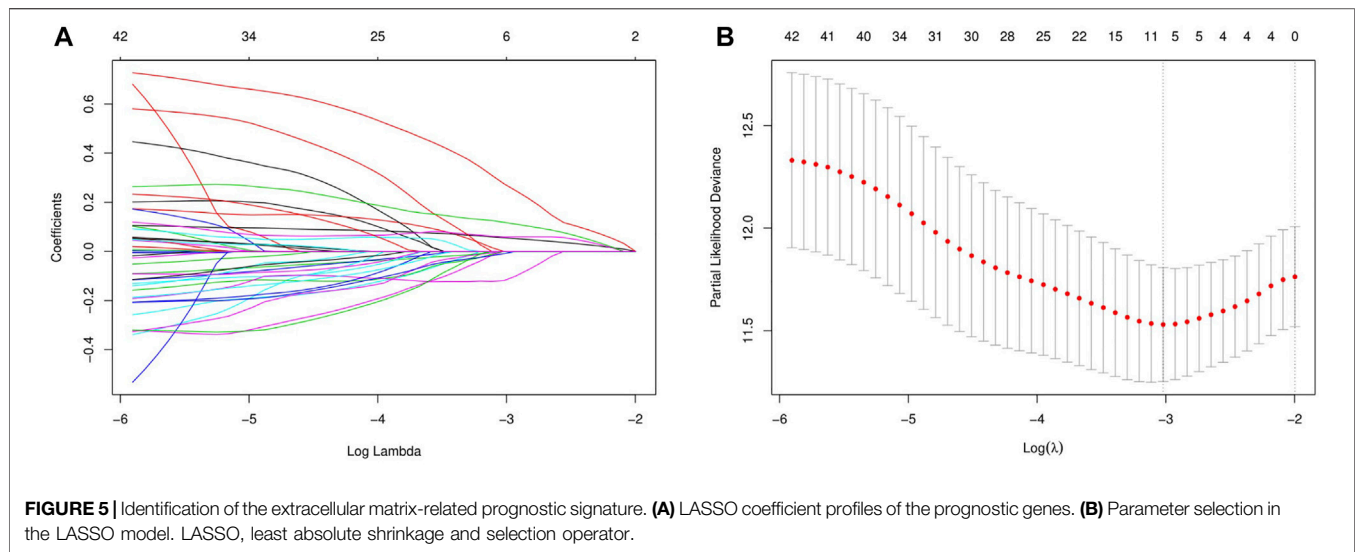
The differences in the sensitivity of immunotherapy, several chemotherapy and targeted therapy drugs between the two HCC clusters were analyzed using the TIDE and GDSC databases. As shown in **Figure 4A**, the response rate to ICIs predicted by the TIDE database was significantly higher in cluster B patients. Furthermore, the estimated IC50 values of doxorubicin, etoposide, gemcitabine, axitinib, lapatinib, and lenalidomide were significantly lower in cluster A samples than in cluster B samples, while the estimated IC50 values of dasatinib, erlotinib, motesanib, and ponatinib were significantly lower in cluster B samples (**Figures 4B–K**). However, the estimated IC50 value of cisplatin was not significantly different between clusters A and B (**Figure 4L**).

## Development and Validation of the ECM-Related Prognostic Signature

Three hundred forty-one HCC patients with complete TNM stage and follow-up information were used for the development of an ECM-related prognostic signature. Following LASSO regression analysis, six genes remained according to the minimum partial likelihood deviance (**Figures 5A,B**). According to the results of LASSO regression, the ECM-related prognostic risk score was constructed as follows: Risk score = (0.055493666\* expression of SPP1) + (0.275597331\* expression of ADAMTS5) + (0.058747561\* expression of MMP1) + (0.118350341\* expression of BSG) + (-0.119029579\* expression of LAMA2) + (-0.009262772\* expression of CDH1). The risk score of each patient was calculated using this formula. According to the optimal cutoff value of the risk score, patients were divided into high- and low-risk groups. As shown in **Table 1**, the risk score was not significantly associated with sex, Ishak score, alpha fetoprotein (AFP), or residual tumor status. However, the high-risk group was significantly correlated with higher histologic grade, advanced age and TNM stage, indicating that a higher risk score may be correlated with HCC progression. Kaplan–Meier survival analyses showed that high-risk patients had remarkably reduced OS compared with low-risk patients in the TCGA-LIHC dataset, even if confounding factors (age, TNM stage, histologic grade, Child–Pugh grade, vascular invasion, AFP, and residual tumor status after surgery) were adjusted after PSM analysis (**Figures 6A,B**). Furthermore, the prognostic value of the risk score was robust when validated in three external validation datasets (**Figures 6C–E**).

## Construction and Evaluation of the ECM-Related Prognostic Nomogram

Univariate analysis demonstrated that risk score, vascular invasion status, and TNM stage were significantly associated with the OS of HCC patients. Further multivariate analysis confirmed that a higher risk score (HR: 4.45, 95% CI: 2.75–7.23,  $p < 0.001$ ) and advanced TNM stage were



independent indicators of unfavorable OS (Table 2). According to the results of multivariate Cox analyses, a nomogram predicting the OS of HCC patients was constructed based on TNM stage and risk score (Figure 7A). The C-index of the nomogram for OS prediction was 0.722 (95% CI: 0.697–0.748). As shown in the calibration plot (Figure 7B), the nomogram demonstrated excellent agreement between the predicted and actual survival outcomes (1-, 3-, and 5-years OS) after surgery. The areas under the curve (AUC) for 1-, 3-, and 5-years OS were 0.794, 0.754, and 0.726, respectively (Figure 7C). In addition, the DCA curve demonstrated that the signature–stage nomogram showed better prognostic capacity than TNM stage (Figure 7D).

## DISCUSSION

With advancements in high-throughput sequencing technology, accumulated prognostic biomarkers and therapeutic targets have been identified and have promoted our understanding of cancer. Previous studies have proven that the ECM-related signature is associated with prognosis and the immune microenvironment in breast cancer and esophageal cancer (Lecchi et al., 2021; Zhang et al., 2021). However, reliable biomarkers for the immunotherapy response and prognosis in HCC based on the ECM are still very rare. Considering the influence of ECM alteration in the TME shares a common molecular mechanism in carcinogenesis and progression, it can be regarded as a pancancer effect (Yu et al., 2021). In addition, liver cirrhosis is closely related to liver cancer, which is also a well-known pathological condition linked to ECM stiffening (Piersma et al., 2020). It is understandable that the ECM plays an important role in HCC.

In the current study, using data from the TCGA-LIHC dataset, we first found that HCC patients could be categorized into two subtypes by the expression profile of ECM-related genes.

Moreover, there was a significant difference in survival outcome between the two molecular subtypes, which confirmed the role of the ECM in HCC prognosis. However, it is difficult to speculate which subtype a patient belongs to based only on clinical features. Furthermore, mutation characteristics analysis revealed that cluster A, with poor prognosis and higher TMB, had higher mutation frequencies of TP53, MUC4, XIRP2, HMCN1, and RYR3, while cluster B had higher mutation frequencies of IL6ST, TRIP12, and MAP2. This is consistent with previous reports that mutant MUC4 was correlated with higher TMB and potentially associated with prognosis in pancancer (Yang Y. et al., 2020), while XIRP2 mutation was potentially associated with metastasis in breast cancer (Paul et al., 2020). Previous studies have reported that genetic mutations in HCC have something to do with risk factors and centered on CTNNB1 (alcohol) and TP53 (HBV) (Schulze et al., 2015), which may be one of the underlying mechanisms of grouping. In addition, mutant TP53 has been consistently associated with poor prognosis in a wide variety of cancers, including HCC (Olivier et al., 2010; Villanueva, 2019). IL6ST is recognized as an oncogene involved in tumorigenesis and associated with inflammatory hepatocellular adenomas (Sun et al., 2014). Functional enrichment analysis demonstrated that immune-relevant pathways were significantly enriched in cluster A, while some pathways related to amino acid metabolism were significantly enriched in cluster B. This is consistent with a previous study that the HCC subgroup with enriched amino acid metabolism-relevant pathways showed a good prognosis (Yang C. et al., 2020). Moreover, inflating immune cells are involved in various steps of antitumor immunity, high infiltration levels of DCs and high-PD-L1-expressing Tregs were reported to be indicators of favorable prognosis in patients treated by ICIs (Wu et al., 2018; Huang et al., 2022). Thus, downregulated immune-related pathways, high infiltration levels of DCs and Tregs may partly explain why patients in cluster B had a higher response to ICIs in our

**TABLE 1** | Association between risk score and the clinicopathological variables in HCC patients ( $n = 341$ ).

	Risk group			<i>p</i>
	Total (N = 341)	High (N = 125)	Low (N = 216)	
Age				
<60	159 (46.6%)	49 (39.2%)	110 (50.9%)	<b>0.048</b>
≥60	182 (53.4%)	76 (60.8%)	106 (49.1%)	
Gender				
Female	109 (32.0%)	45 (36.0%)	64 (29.6%)	0.273
Male	232 (68.0%)	80 (64.0%)	152 (70.4%)	
Family history of cancer				
No	196 (57.5%)	69 (55.2%)	127 (58.8%)	0.214
Yes	100 (29.3%)	43 (34.4%)	57 (26.4%)	
Unknown	45 (13.2%)	13 (10.4%)	32 (14.8%)	
TNM stage				
I	170 (49.9%)	44 (35.2%)	126 (58.3%)	<b>&lt;0.001</b>
II	84 (24.6%)	35 (28.0%)	49 (22.7%)	
III	83 (24.3%)	45 (36.0%)	38 (17.6%)	
IV	4 (1.2%)	1 (0.8%)	3 (1.4%)	
Histologic grade				
G1–G2	212 (62.2%)	69 (55.2%)	143 (66.2%)	<b>0.032</b>
G3–G4	127 (37.2%)	54 (43.2%)	73 (33.8%)	
Unknown	2 (0.6%)	2 (1.6%)	0 (0%)	
Ishak score				
0–4	124 (36.4%)	39 (31.2%)	85 (39.4%)	0.253
5–6	74 (21.7%)	27 (21.6%)	47 (21.8%)	
Unknown	143 (41.9%)	59 (47.2%)	84 (38.9%)	
Child–Pugh grade				
A	207 (60.7%)	60 (48.0%)	147 (68.1%)	<b>0.001</b>
B–C	21 (6.2%)	9 (7.2%)	12 (5.6%)	
Unknown	113 (33.1%)	56 (44.8%)	57 (26.4%)	
Vascular invasion				
Macro	16 (4.7%)	9 (7.2%)	7 (3.2%)	<b>0.001</b>
Micro	84 (24.6%)	33 (26.4%)	51 (23.6%)	
None	193 (56.6%)	56 (44.8%)	137 (63.4%)	
Unknown	48 (14.1%)	27 (21.6%)	21 (9.7%)	
Alpha fetoprotein				
Negative	87 (25.5%)	24 (19.2%)	63 (29.2%)	0.057
Positive	254 (74.5%)	101 (80.8%)	153 (70.8%)	
Residual tumor				
R0	301 (88.3%)	105 (84.0%)	196 (90.7%)	0.068
R1–R2	14 (4.1%)	5 (4.0%)	9 (4.2%)	
Unknown	26 (7.6%)	15 (12.0%)	11 (5.1%)	

The bold values indicate they are statistically significant ( $p < 0.05$ ).

study. However, tumor-infiltrating immune cells are highly heterogeneous. For instance, different DC subsets have significantly different effects on immunity and tolerance in cancer settings (Wculek et al., 2020). The prognosis of colorectal cancer patients with increased numbers of Tregs has also been controversial, which may be attributed to an improper interpretation of heterogeneous FOXP3+ cells as a single population of Tregs (Tanaka and Sakaguchi, 2017). More coming studies based on single-cell sequencing may help elucidate this issue.

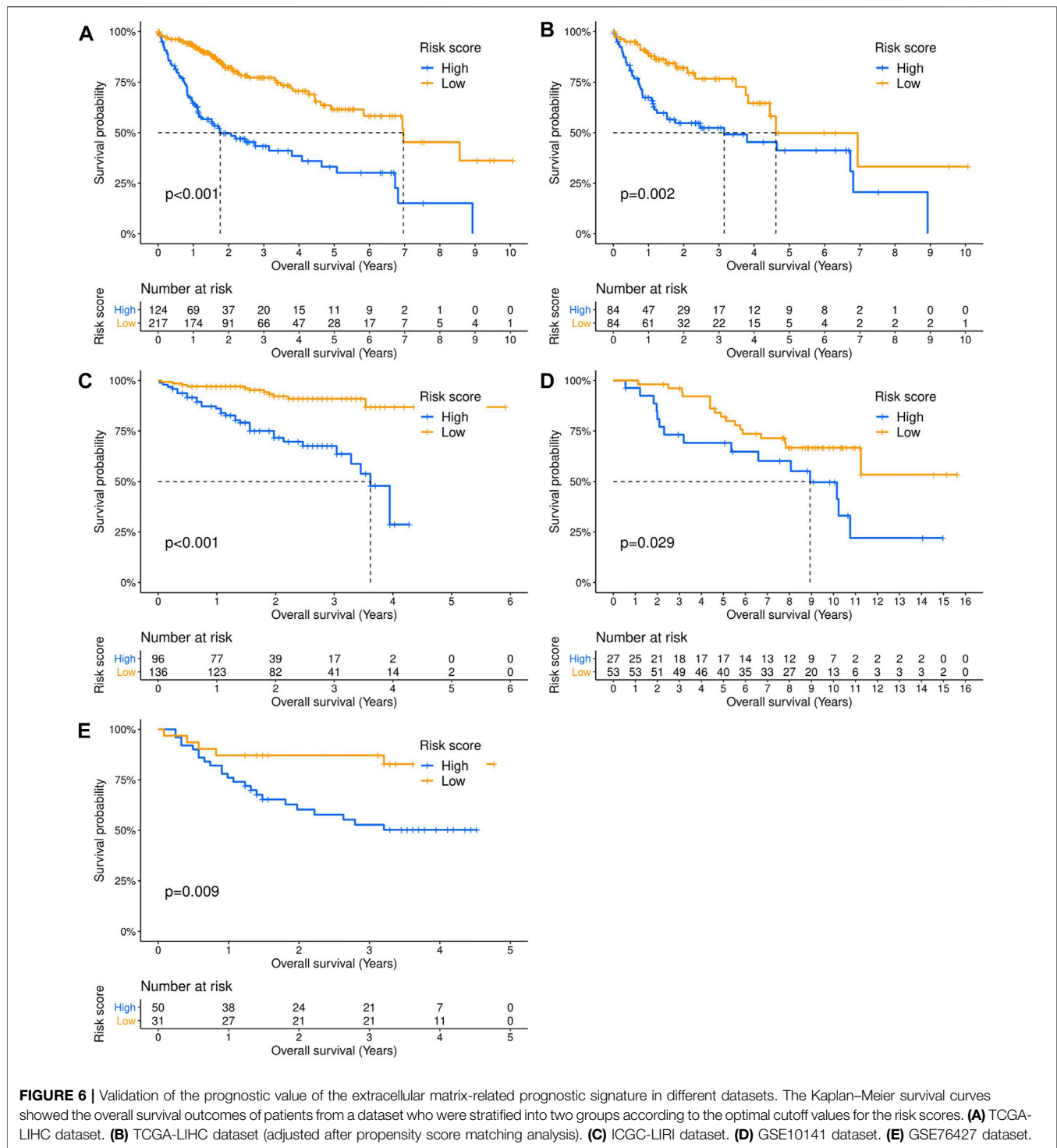
The IMbrave150 phase 3 study reported that ICI plus antiangiogenesis-based therapies could improve the survival outcomes of HCC patients over sorafenib alone in the first-line setting, but not all patients can benefit from ICIs (Pinter et al., 2021). Intriguingly, our results suggested that HCC patients in ECM-related cluster A had significantly

higher TMB, but almost all HCC patients derived from TCGA database had low TMB. In contrast, patients in cluster B had higher immune and stromal scores. Importantly, we also found that cluster B patients may have a higher response rate to ICIs. Besides, 10 chemotherapy and targeted drugs had great differences in estimated IC50 values between the two molecular subtypes, further indicating that this signature could play a pivotal role in the prognosis and therapeutic responses of HCC patients.

After that, for the convenience of clinical application, we constructed an ECM-related prognostic risk score model using the expression profiles of six genes (SPP1, ADAMTS5, MMP1, BSG, LAMA2, and CDH1), which was internally and externally validated using four independent cohorts. Our results suggested that the expression levels of SPP1, ADAMTS5, MMP1, and BSG were associated with poor prognosis of HCC, while LAMA2 and CDH1 were related to longer survival. Intriguingly, SPP1 overexpression was reported to be associated with HCC progression and immune escape in lung adenocarcinoma (Zhang et al., 2017; Wang et al., 2019). MMP1 and BSG overexpression are also well-known poor prognostic markers of HCC (Liu H. et al., 2021; Ma et al., 2021). High CDH1 mRNA expression was also significantly correlated with better survival outcomes in HCC patients (Wu et al., 2019). However, the prognostic effect of ADAMTS5 in hepatocellular carcinoma remains controversial (Theret et al., 2021). Furthermore, the ECM-related prognostic nomogram demonstrated excellent agreement between the predicted and actual survival outcomes, as well as better prognostic capacity than TNM stage.

In addition to being a prognostic marker, targeting the ECM-mediated immunosuppressive stromal microenvironment and physical barrier in combination with other systemic therapies could promisingly ameliorate the response to those drugs. For example, ECM deposition was related to immunosuppression and gemcitabine resistance in pancreatic cancer (Peran et al., 2021), while the proteolysis of ECM proteoglycans was associated with T cell infiltration in colorectal cancer (Hope et al., 2017). Although various studies targeting different components of the ECM (such as degradation of stromal collagen and hyaluronic acid) have reported desirable success, this strategy has repeatedly failed in clinical studies (Abyaneh et al., 2020), which may be partly explained by the nonspecificity of drugs and complex context-specific roles of the ECM (Mushtaq et al., 2018). Excessive depletion of ECM would compromise or even worsen the outcomes. This was shown in transgenic mouse models exhibiting reduced stromal content. Both fibroblast-depleted tumors and hedgehog-inhibited tumors showed more aggressive behaviors (Rhim et al., 2014; Özdemir et al., 2014). Moreover, excessive removal of ECM components may result in matrix collapse and decreased drug penetration (Abyaneh et al., 2020). Therefore, normalization of the ECM rather than its depletion may be the better goal (Abyaneh et al., 2020).





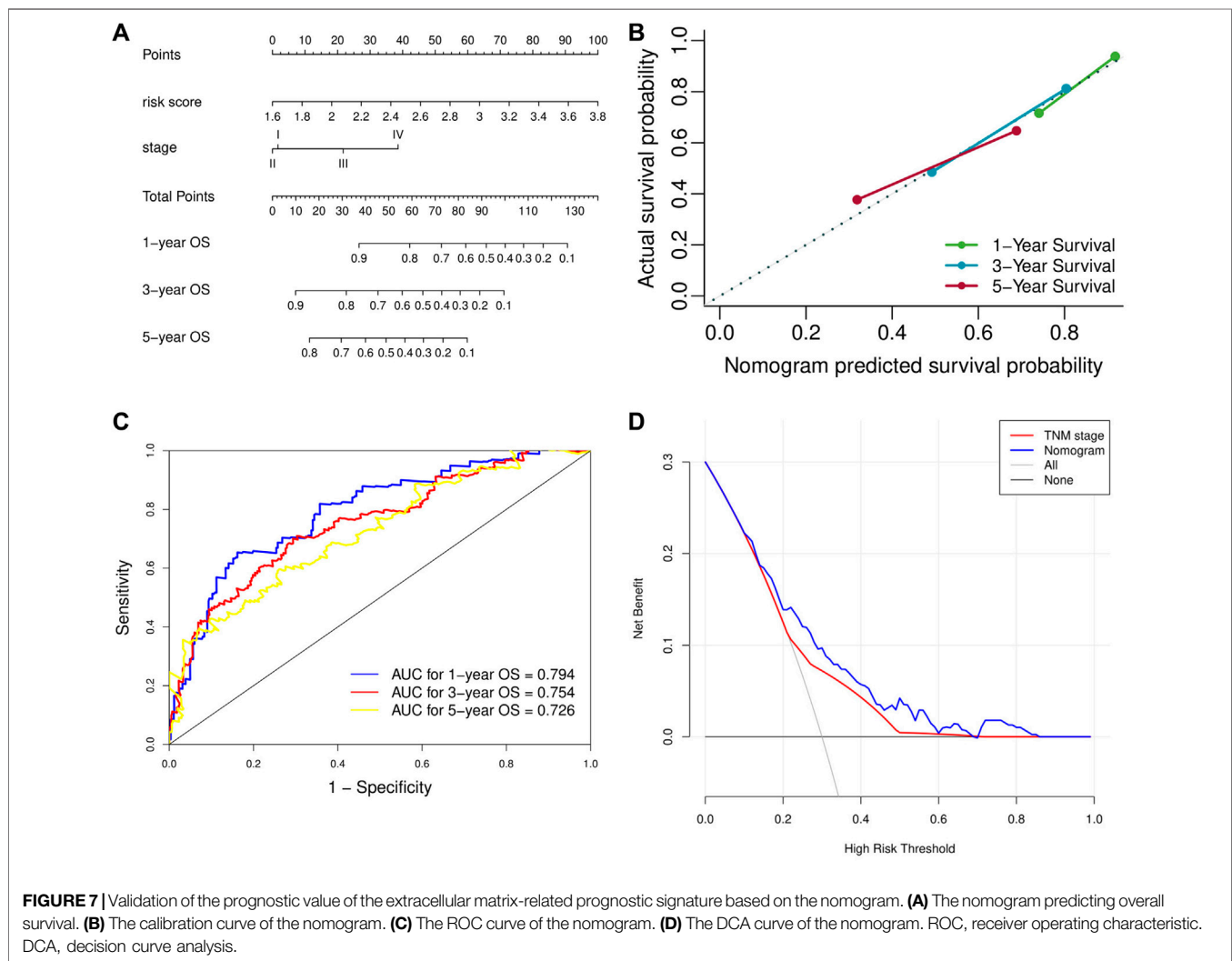
On the other hand, targeting the ECM for drug delivery is another promising therapeutic strategy. For example, matrix-binding ICIs could enhance antitumor efficacy and reduce adverse events in a preclinical melanoma model (Ishihara et al., 2017). However, ECM-targeting drug delivery strategies are mainly explored in animal studies and deserve further validation in clinical trials.

To the best of our knowledge, this is the first and comprehensive study classifying HCC patients based on ECM-related gene expression profiles. Furthermore, an ECM-related prognostic signature and nomogram were developed and validated. However, there were still some limitations in our study, although the prognostic signature was validated by three independent cohorts and demonstrated

**TABLE 2** | Univariate and multivariate analyses of overall survival.

Variables	Univariate analysis		Multivariate analysis	
	HR (95% CI)	p	HR (95% CI)	p
Age ( $\geq 60$ vs. $< 60$ )	1.25(0.87,1.81)	0.234	–	–
Gender (Female vs. Male)	1.28(0.88,1.86)	0.197	–	–
Family history of cancer (Yes vs. No)	1.13(0.77,1.68)	0.532	–	–
TNM stage (II vs. I)	1.42(0.87,2.32)	0.158	1.02(0.57,1.82)	0.944
TNM stage (III vs. I)	2.68(1.75,4.08)	<b>&lt;0.001</b>	1.76(1.07,2.88)	<b>0.025</b>
TNM stage (IV vs. I)	5.5(1.7,17.82)	<b>0.005</b>	3.84(1.17,12.57)	<b>0.026</b>
Histologic grade (G3–G4 vs. G1–G2)	1.14(0.79,1.67)	0.481	–	–
Ishak score (5–6 vs. 0–4)	0.87(0.5,1.5)	0.613	–	–
Child–Pugh grade (B–C vs. A)	1.66(0.82,3.36)	0.159	–	–
Vascular invasion (Micro vs. None)	0.51(0.23,1.14)	0.101	0.77(0.33,1.82)	0.556
Vascular invasion (Macro vs. None)	0.44(0.21,0.93)	<b>0.032</b>	0.87(0.39,1.93)	0.735
Alpha fetoprotein (Positive vs. Negative)	1.2(0.78,1.83)	0.412	–	–
Residual tumor (R1–R2 vs. R0)	1.16(0.47,2.86)	0.74	–	–
Risk score (High vs. Low)	5.29(3.35,8.35)	<b>&lt;0.001</b>	4.45(2.75,7.23)	<b>&lt;0.001</b>

The bold values indicate they are statistically significant ( $p < 0.05$ ).



**FIGURE 7** | Validation of the prognostic value of the extracellular matrix-related prognostic signature based on the nomogram. **(A)** The nomogram predicting overall survival. **(B)** The calibration curve of the nomogram. **(C)** The ROC curve of the nomogram. **(D)** The DCA curve of the nomogram. ROC, receiver operating characteristic. DCA, decision curve analysis.

good accuracy. First, HCC patient classification and the prognostic signature were conducted based on retrospective data, and prospective validation is needed. Moreover, considering ECM components and structures are quite dynamic being subject to protein changes, further proteomics studies are also warranted to better evaluate the impact of ECM on HCC. Second, due to methodological limitations, we were unable to analyze and predict the sensitivity of some essential drugs for HCC patients, such as sorafenib and lenvatinib, which may be addressed in other databases or prospective studies. Third, the candidate genes enrolled in our study were restricted to the ECM-related signature, while the immune microenvironment in tumors is a complex consisting of the ECM and a variety of cells. Thus, the prognostic predictive power of the signature may be limited. Nonetheless, the ECM-related signature provides abundant information about the immune microenvironment and demonstrates good accuracy, which proves to some extent that developing a prognostic model based on an ECM-related signature is rational.

## CONCLUSION

Collectively, based on ECM-related gene expression profiles, two molecular subtypes in HCC patients were characterized, with distinct clinical outcomes, somatic mutation profiles, and drug sensitivity. Ultimately, an ECM-related prognostic signature was developed and validated. These findings may improve our understanding of the ECM signature in HCC and pave a new path for the assessment of prognosis and drug sensitivity.

## REFERENCES

- Abyaneh, H. S., Regenold, M., McKee, T. D., Allen, C., and Gauthier, M. A. (2020). Towards Extracellular Matrix Normalization for Improved Treatment of Solid Tumors. *Theranostics* 10 (4), 1960–1980. doi:10.7150/thno.39995
- Aran, D., Hu, Z., and Butte, A. J. (2017). xCell: Digitally Portraying the Tissue Cellular Heterogeneity Landscape. *Genome Biol.* 18 (1), 220. doi:10.1186/s13059-017-1349-1
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer J. Clinicians* 68 (6), 394–424. doi:10.3322/caac.21492
- Deligne, C., and Midwood, K. S. (2021). Macrophages and Extracellular Matrix in Breast Cancer: Partners in Crime or Protective Allies. *Front. Oncol.* 11, 620773. doi:10.3389/fonc.2021.620773
- Geeleher, P., Cox, N., and Huang, R. S. (2014). pRRophetic: an R Package for Prediction of Clinical Chemotherapeutic Response from Tumor Gene Expression Levels. *PLoS One* 9 (9), e107468. doi:10.1371/journal.pone.0107468
- Hope, C., Emmerich, P. B., Papadas, A., Pagenkopf, A., Matkowskyj, K. A., Van De Hey, D. R., et al. (2017). Versican-Derived Matrikines Regulate Batf3-Dendritic Cell Differentiation and Promote T Cell Infiltration in Colorectal Cancer. *J. Immunol.* 199 (5), 1933–1941. doi:10.4049/jimmunol.1700529
- Huang, T.-X., Tan, X.-Y., Huang, H.-S., Li, Y.-T., Liu, B.-L., Liu, K.-S., et al. (2022). Targeting Cancer-Associated Fibroblast-Secreted WNT2 Restores Dendritic Cell-Mediated Antitumor Immunity. *Gut* 71 (2), 333–344. doi:10.1136/gutjnl-2020-322924
- Ishihara, J., Fukunaga, K., Ishihara, A., Larsson, H. M., Potin, L., Hosseinchi, P., et al. (2017). Matrix-binding Checkpoint Immunotherapies Enhance

## 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.

## ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## AUTHOR CONTRIBUTIONS

HT conceived of and designed the study. HT and TY performed literature search. HT and ZS analyzed the data. HT and TY generated the figures and tables. HT wrote the manuscript, and ZS and YW critically reviewed the manuscript. YW and CB supervised the research. All authors have read and approved the manuscript.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2022.839806/full#supplementary-material>

**Supplementary Table S1** | Association between molecular clusters and clinicopathological variables in HCC patients ( $n = 341$ ).

- Antitumor Efficacy and Reduce Adverse Events. *Sci. Transl. Med.* 9 (415). doi:10.1126/scitranslmed.aan0401
- Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., et al. (2020). The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* 48 (D1), D498–d503. doi:10.1093/nar/gkz1031
- Jeng, K.-S., Chang, C.-F., Jeng, W.-J., Sheen, I.-S., and Jeng, C.-J. (2015). Heterogeneity of Hepatocellular Carcinoma Contributes to Cancer Progression. *Crit. Rev. Oncology/Hematology* 94 (3), 337–347. doi:10.1016/j.critrevonc.2015.01.009
- 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
- Kurebayashi, Y., Ojima, H., Tsujikawa, H., Kubota, N., Maehara, J., Abe, Y., et al. (2018). Landscape of Immune Microenvironment in Hepatocellular Carcinoma and its Additional Impact on Histological and Molecular Classification. *Hepatology* 68 (3), 1025–1041. doi:10.1002/hep.29904
- Lecchi, M., Verderio, P., Cappelletti, V., De Santis, F., Paolini, B., Monica, M., et al. (2021). A Combination of Extracellular Matrix- and Interferon-associated Signatures Identifies High-grade Breast Cancers with Poor Prognosis. *Mol. Oncol.* 15 (5), 1345–1357. doi:10.1002/1878-0261.12912
- Liu, H., Lan, T., Li, H., Xu, L., Chen, X., Liao, H., et al. (2021c). Circular RNA circDLC1 Inhibits MMP1-Mediated Liver Cancer Progression via Interaction with HuR. *Theranostics* 11 (3), 1396–1411. doi:10.7150/thno.53227
- Liu, Z., Lu, T., Wang, L., Liu, L., Li, L., and Han, X. (2021b). Comprehensive Molecular Analyses of a Novel Mutational Signature Classification System with Regard to Prognosis, Genomic Alterations, and Immune Landscape in Glioma. *Front. Mol. Biosci.* 8, 682084. doi:10.3389/fmolb.2021.682084
- Liu, Z., Wang, L., Guo, C., Liu, L., Jiao, D., Sun, Z., et al. (2021a). TTN/OBSCN 'Double-Hit' Predicts Favourable Prognosis, 'immune-hot' Subtype and

- Potentially Better Immunotherapeutic Efficacy in Colorectal Cancer. *J. Cel. Mol. Med.* 25 (7), 3239–3251. doi:10.1111/jcmm.16393
- Ma, Y., Sun, W., Zhang, Q., Gao, B., Cai, W., Liu, Q., et al. (2021). lncRNA BSG-AS1 Is Hypoxia-Responsive and Promotes Hepatocellular Carcinoma by Enhancing BSG mRNA Stability. *Biochem. Biophysical Res. Commun.* 566, 101–107. doi:10.1016/j.bbrc.2021.06.002
- 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. Immunotherapy Cancer* 6 (1), 65. doi:10.1186/s40425-018-0376-0
- 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
- Olivier, M., Hollstein, M., and Hainaut, P. (2010). TP53 Mutations in Human Cancers: Origins, Consequences, and Clinical Use. *Cold Spring Harbor Perspect. Biol.* 2 (1), a001008. doi:10.1101/cshperspect.a001008
- Özdemir, B. C., Pentcheva-Hoang, T., Carstens, J. L., Zheng, X., Wu, C.-C., Simpson, T. R., et al. (2014). Depletion of Carcinoma-Associated Fibroblasts and Fibrosis Induces Immunosuppression and Accelerates Pancreas Cancer with Reduced Survival. *Cancer Cell* 25 (6), 719–734. doi:10.1016/j.ccr.2014.04.005
- Paul, M. R., Pan, T.-c., Pant, D. K., Shih, N. N. C., Chen, Y., Harvey, K. L., et al. (2020). Genomic Landscape of Metastatic Breast Cancer Identifies Preferentially Dysregulated Pathways and Targets. *J. Clin. Invest.* 130 (8), 4252–4265. doi:10.1172/jci129941
- Peran, I., Dakshanamurthy, S., McCoy, M. D., Mavropoulos, A., Allo, B., Sebastian, A., et al. (2021). Cadherin 11 Promotes Immunosuppression and Extracellular Matrix Deposition to Support Growth of Pancreatic Tumors and Resistance to Gemcitabine in Mice. *Gastroenterology* 160 (4), 1359–1372. e1313. doi:10.1053/j.gastro.2020.11.044
- Piersma, B., Hayward, M. K., and Weaver, V. M. (2020). Fibrosis and Cancer: A Strained Relationship. *Biochim. Biophys. Acta (Bba) - Rev. Cancer* 1873 (2), 188356. doi:10.1016/j.bbcan.2020.188356
- Pinter, M., Jain, R. K., and Duda, D. G. (2021). The Current Landscape of Immune Checkpoint Blockade in Hepatocellular Carcinoma. *JAMA Oncol.* 7 (1), 113–123. doi:10.1001/jamaoncol.2020.3381
- Rhim, A. D., Oberstein, P. E., Thomas, D. H., Mirek, E. T., Palermo, C. F., Sastra, S. A., et al. (2014). Stromal Elements Act to Restrain, rather Than Support, Pancreatic Ductal Adenocarcinoma. *Cancer Cell* 25 (6), 735–747. doi:10.1016/j.ccr.2014.04.021
- Schulze, K., Imbeaud, S., Letouze, E., Alexandrov, L. B., Calderaro, J., Rebouissou, S., et al. (2015). Exome Sequencing of Hepatocellular Carcinomas Identifies New Mutational Signatures and Potential Therapeutic Targets. *Nat. Genet.* 47 (5), 505–511. doi:10.1038/ng.3252
- Sia, D., Jiao, Y., Martinez-Quetglas, I., Kuchuk, O., Villacorta-Martin, C., Castro de Moura, M., et al. (2017). Identification of an Immune-specific Class of Hepatocellular Carcinoma, Based on Molecular Features. *Gastroenterology* 153 (3), 812–826. doi:10.1053/j.gastro.2017.06.007
- Sun, L., Sui, L., Cong, X., Ma, K., Ma, X., Huang, Y., et al. (2014). Low Incidence of IL6ST (Gp130) Mutations in Exon 6 in Lung Cancer of a Chinese Cohort. *Cancer Genet.* 207 (7–8), 291–298. doi:10.1016/j.cancergen.2014.07.003
- Tanaka, A., and Sakaguchi, S. (2017). Regulatory T Cells in Cancer Immunotherapy. *Cell Res.* 27 (1), 109–118. doi:10.1038/cr.2016.151
- Théret, N., Bouezzeddine, F., Azar, F., Diab-Assaf, M., and Legagneux, V. (2021). ADAM and ADAMTS Proteins, New Players in the Regulation of Hepatocellular Carcinoma Microenvironment. *Cancers* 13 (7), 1563. doi:10.3390/cancers13071563
- Villanueva, A. (2019). Hepatocellular Carcinoma. *N. Engl. J. Med.* 380 (15), 1450–1462. doi:10.1056/NEJMra1713263
- Wagner, G. P., Kin, K., and Lynch, V. J. (2012). Measurement of mRNA Abundance Using RNA-Seq Data: RPKM Measure Is Inconsistent Among Samples. *Theor. Biosci.* 131 (4), 281–285. doi:10.1007/s12064-012-0162-3
- Wang, J., Hao, F., Fei, X., and Chen, Y. (2019). SPP1 Functions as an Enhancer of Cell Growth in Hepatocellular Carcinoma Targeted by miR-181c. *Am. J. Transl. Res.* 11 (11), 6924–6937.
- Wculek, S. K., Cueto, F. J., Mujal, A. M., Melero, I., Krummel, M. F., and Sancho, D. (2020). Dendritic Cells in Cancer Immunology and Immunotherapy. *Nat. Rev. Immunol.* 20 (1), 7–24. doi:10.1038/s41577-019-0210-z
- Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: a Class Discovery Tool with Confidence Assessments and Item Tracking. *Bioinformatics* 26 (12), 1572–1573. doi:10.1093/bioinformatics/btq170
- Wu, S.-P., Liao, R.-Q., Tu, H.-Y., Wang, W.-J., Dong, Z.-Y., Huang, S.-M., et al. (2018). Stromal PD-L1-Positive Regulatory T Cells and PD-1-Positive CD8-Positive T Cells Define the Response of Different Subsets of Non-small Cell Lung Cancer to PD-1/pd-L1 Blockade Immunotherapy. *J. Thorac. Oncol.* 13 (4), 521–532. doi:10.1016/j.jtho.2017.11.132
- Wu, X., Yao, X., Cao, Q., Wu, Z., Wang, Z., Liu, F., et al. (2019). Clinicopathological and Prognostic Significance of CDH1 Hypermethylation in Hepatocellular Carcinoma: a Meta-Analysis. *Cmar* Vol. 11, 857–864. doi:10.2147/cmar.S179710
- Xiao, Z., Li, J., Yu, Q., Zhou, T., Duan, J., Yang, Z., et al. (2021). An Inflammatory Response Related Gene Signature Associated with Survival Outcome and Gemcitabine Response in Patients with Pancreatic Ductal Adenocarcinoma. *Front. Pharmacol.* 12, 778294. doi:10.3389/fphar.2021.778294
- Yang, C., Huang, X., Liu, Z., Qin, W., and Wang, C. (2020b). Metabolism-associated Molecular Classification of Hepatocellular Carcinoma. *Mol. Oncol.* 14 (4), 896–913. doi:10.1002/1878-0261.12639
- Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., Forbes, S., et al. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a Resource for Therapeutic Biomarker Discovery in Cancer Cells. *Nucleic Acids Res.* 41, D955–D961. doi:10.1093/nar/gks1111
- Yang, Y., Zhang, J., Chen, Y., Xu, R., Zhao, Q., and Guo, W. (2020a). MUC4 , MUC16 , and TTN Genes Mutation Correlated with Prognosis, and Predicted Tumor Mutation burden and Immunotherapy Efficacy in Gastric Cancer and pan-cancer. *Clin. Translational Med.* 10 (4), e155. doi:10.1002/ctm2.155
- Yu, C., You, M., Zhang, P., Zhang, S., Yin, Y., and Zhang, X. (2021). A Five-gene Signature Is a Prognostic Biomarker in pan-cancer and Related with Immunologically Associated Extracellular Matrix. *Cancer Med.* 10 (13), 4629–4643. doi:10.1002/cam4.3986
- Zhang, H., Shi, Q., Yang, Z., Wang, K., Zhang, Z., Huang, Z., et al. (2021). An Extracellular Matrix-Based Signature Associated with Immune Microenvironment Predicts the Prognosis and Therapeutic Responses of Patients with Oesophageal Squamous Cell Carcinoma. *Front. Mol. Biosci.* 8, 598427. doi:10.3389/fmolb.2021.598427
- Zhang, Y., Du, W., Chen, Z., and Xiang, C. (2017). Upregulation of PD-L1 by SPP1 Mediates Macrophage Polarization and Facilitates Immune Escape in Lung Adenocarcinoma. *Exp. Cel. Res.* 359 (2), 449–457. doi:10.1016/j.yexcr.2017.08.028

**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:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Tang, You, Sun, Bai and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.