Comprehensive evaluation of genes related to basement membrane in hepatocellular carcinoma

Guojing Wu[1,2,*], Fei Li[1,2,*], Danyan Guo[1,2,*], Kaiwen Xi[1,2], Dayong Zheng[1,2], Ruichao Huang[3], Xiuqiong Wu[1,2], Aimin Li[1,2], Xinhui Liu[1,2,3]

[1]Southern Medical University Hospital of Integrated Traditional Chinese and Western Medicine, Southern Medical University, Guangzhou 510315, China
[2]Cancer Center, Southern Medical University, Guangzhou 510315, China
[3]Heshan Hospital of Traditional Chinese Medicine, Jiangmen 529000, China
*Equal contribution and share first authorship

Correspondence to: Aimin Li, Xinhui Liu; email: liaimin@smu.edu.cn, jkkl@smu.edu.cn

ABSTRACT

In all mammals, the basement membrane serves as a pivotal extracellular matrix. Hepatocellular carcinoma (HCC) is a challenge among numerous cancer types shaped by basement membrane-related genes (BMGs). Our research established an innovative prognostic model that is highly accurate in its prediction of HCC prognoses and immunotherapy efficacy to summarize the crucial role of BMGs in HCC. We obtained HCC transcriptome analysis data and corresponding clinical data from The Cancer Genome Atlas (TCGA). To augment our dataset, we incorporated 222 differentially expressed BMGs identified from relevant literature. A weighted gene coexpression network analysis (WGCNA) of 10158 genes demonstrated four modules that were connected to HCC. Additionally, 66 genes that are found at the intersection of BMGs and HCC-related genes were designated as hub HCC-related BMGs. MMP1, ITGA2, P3H1, and CTSA comprise the novel model that was engineered using univariate and multivariate Cox regression analysis. Furthermore, the International Cancer Genome Consortium (ICGC) and Gene Expression Omnibus (GEO) datasets encouraged the BMs model's validity. The overall survival (OS) of individuals with HCC may be precisely predicted in the TCGA and ICGC databases utilizing the BMs model. A nomogram based on the model was created in the TCGA database at similar time, and displayed a favorable discriminating ability for HCC. Particularly, when compared to the patients at an elevated risk, the patients with a low-risk profile presented different tumor microenvironment (TME) and hallmark pathways. Moreover, we discovered that a lower risk score of HCC patients would display a greater response to immunotherapy. Finally, quantitative real-time PCR (qRT-PCR) experiments were used to verify the expression patterns of BMs model. In summary, BMs model demonstrated efficacy in prognosticating the survival probability of HCC patients and their immunotherapeutic responsiveness.

INTRODUCTION

Liver cancer ranks as the sixth most common cancer globally and stands third in cancer-related mortality, as reported by the World Health Organization [1]. Among primary liver malignancies, hepatocellular carcinoma (HCC) stands as the prevailing histological subtype, accounting for approximately 70-85% of all instances [2]. Despite the fact that HCC is a major health concern, there are very few effective interventions available, and

surgical resection remains the best treatment for HCC [3]. Over the past ten years, there have been only modest advancements in systemic therapies for liver cancer [4]. As a result, comprehending the biological mechanisms that drive HCC malignancy becomes imperative in order to devise more potent therapeutic approaches.

Basement membranes (BMs), an ancient extracellular matrix that appeared with the emergence of multicellular animals [5], are predominately composed of collagen, laminin, and non-collagenous glycoproteins [6]. Nidogen and heparan sulfate proteoglycans act as network-bridging proteins between type IV collagen molecules and laminin [7], and assist in the building of BMs along tissues by providing scaffolding structure [8]. Moreover, the BMs also incorporate "matricellular proteins" that serve specialized roles tailored to specific tissues, though they do not participate in the constitution or structural cohesion [9]. Functions of BMs including blood filtration, muscle homeostasis, controlling angiogenesis and tumor growth, mechanical stress resistant and storing growth factors and cytokines [7]. Studies have shown that germline variants in approximately 30 basement membrane-related genes (BMGs) can lead to genetic disorders [10]. Furthermore, emerging evidence suggests that disruptions in the synthesis and degradation of BMs proteins may be associated with fibrosis, diabetes, and cancer [11–13]. Additionally, collagen XIX, a lesser-known member of the FACIT family, has demonstrated the ability to suppress angiogenesis, thus inhibiting the new blood vessel formation and potentially impeding tumor cell invasion [14]. Additionally, the interaction between integrin α3β1 and laminins is crucial to tumor cell proliferation initiation and maintenance [15]. Studies have indicated that CD8[+] T cells infiltrate tumor lesions by secreting granzyme B, which promotes the remodeling of the BMs of tumor blood vessels, facilitating their migration [16] that means BMs may be related to the immune microenvironment in HCC. However, BMs are not well understood in HCC in terms of their role and status. Thus, it is quite crucial to explicate whether BMGs are effective at predicting HCC prognosis and immunotherapy response for HCC patients.

In the present study, an extensive investigation of characteristics and biomarkers associated with BMGs was carried out, as they will provide a means to explore HCC immune infiltration and monitor the response to immunotherapy. First, we constructed coexpression networks of tumor samples and identified the inter-secting genes of BMGs by weighted gene co-expression network analysis (WGCNA). Then a model based on BMGs was established and its prognostic value was assessed in predicting HCC patients' outcome and response to immunotherapy and chemotherapy. In conclusion, our research provides novel insights that could lead to improved prognosis and treatment outcomes for HCC patients in the coming years.

RESULTS

Identification and enrichment analysis of catheter holder modules associated with HCC

In the TCGA-LIHC cohort, the brown module (R^2 = -0.81, P = 8e−101), yellow module (R^2 = 0.5, P = 1e−27), turquoise module (R^2 = -0.49, P = 9e−27) and greenyellow module (R^2 = 0.32, P = 8e−12) were extremely correlated with tumor samples among the 11 modules (Figure 1A–1H). These four modules were identified as hub modules. Subsequently, there are 10158 genes that related to tumor were identified from the emergence of the above modules. Then, there are 66 common genes between hub module genes and BMGs for further study (Figure 1I).

In this part, 66 pivotal genes that could determine HCC progression and metastasis by BMs were explored for future research.

GO and KEGG enrichment analysis of BMGs in HCC

The "clusterProfiler" R package was employed for GO enrichment analysis to elucidate the relevant biological processes and molecular structures, and the correlated pathways were acquired applying KEGG enrichment analysis. As illustrated by GO enrichment analysis, the intersection of genes was closely related to the extra-cellular matrix and collagen-containing extracellular matrix in the Cellular Component (CC). The enriched Biological Processes (BP) primarily encompassed extracellular structure organization and extracellular matrix organization. In terms of Molecular Function (MF), changes were observed in extracellular structural constituent and cell adhesion molecule binding (Figure 2A–2C). KEGG enrichment analysis revealed significant gene intersections associated with extracellular matrix (ECM)-receptor interaction, Focal adhesion, Human papillomavirus infection, and PI3K-AKT signaling pathway (Figure 2D–2F).

Taken together, this part offered new ideas and clew to construct future exploration from CC, BP and MF by utilizing the GO enrichment analysis and the KEGG enrichment analysis.

Multi-omics landscape of BMs in HCC

We have shown Cox regression analysis of all 66 common genes (Supplementary Table 1) and identified
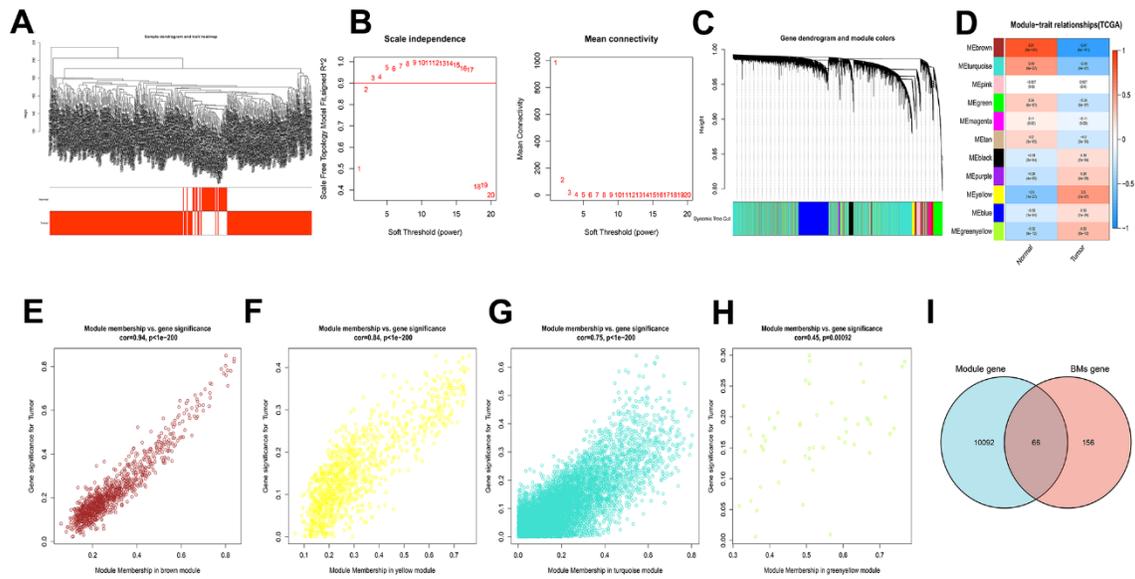
Figure 1. WGCNA analysis results of 10158 genes of HCC. (**A**) Clustering dendrogram of HCC samples. (**B**) The scale–free fit index for soft–thresholding powers. (**C**) A dendrogram of the differentially expressed genes clustering based on different metrics. (**D**) A heatmap showing the correlation between the gene module and associated traits. (**E**–**H**) Scatter plots of module eigengenes in brown, yellow, turquoise and greenyellow modules. (**I**) Venn diagram showed the intersection of genes of HCC and BMGs.
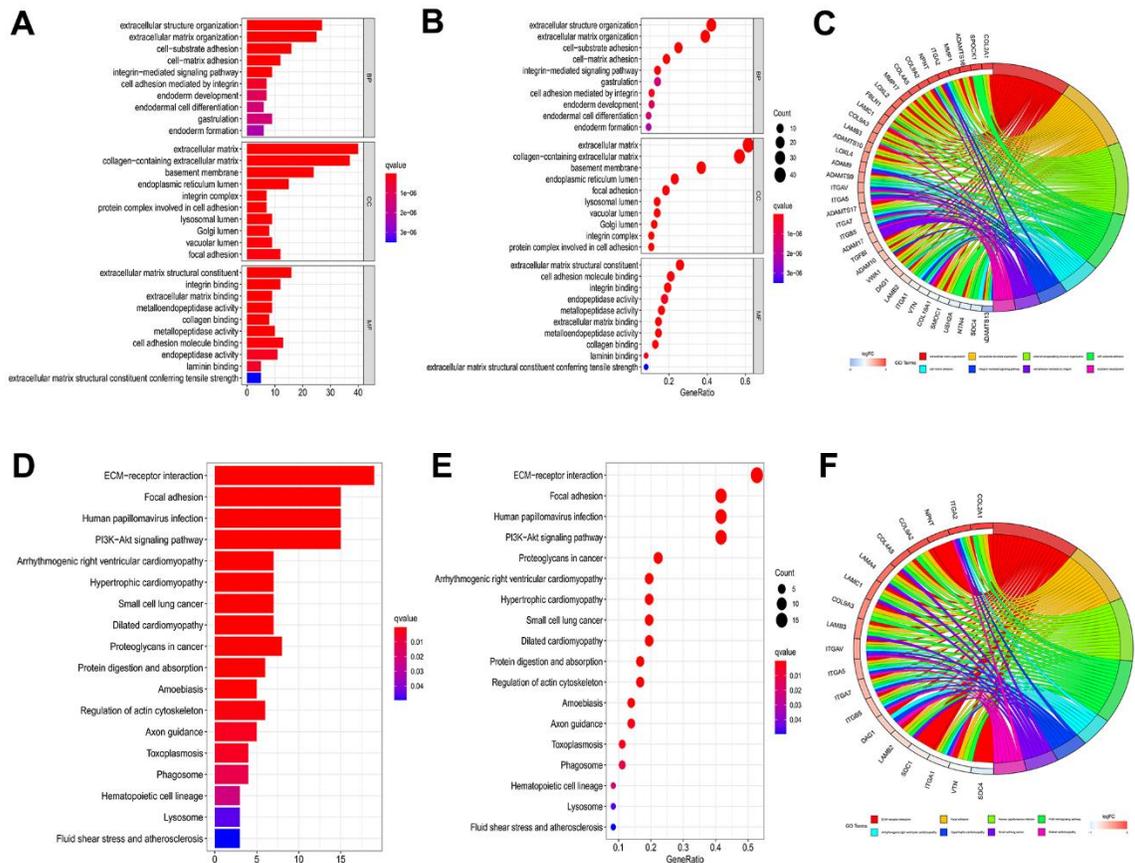
Figure 2. GO and KEGG enrichment analysis to 66 genes of intersection. Barplot (**A**), bubble plot (**B**) and chord diagram (**C**) of GO analyses of genes of intersection. Barplot (**D**), bubble plot (**E**) and chord diagram (**F**) of KEGG analyses of genes of intersection.

20 BMGs that related to prognostic by univariate Cox regression analysis (Figure 3A). Next, in TME of HCC, the relationship between BMGs and immune cell infiltration was investigated. As illustrated in Figure 3B, a significant positive correlation observed among the BMGs. Moreover, notable positive associations were observed between BMGs and the infiltration of immune cells in HCC individuals (Figure 3C). CNV is a DNA fragment in the human genome with a copy size of 1kB to 1MB. It is associated with tumorigenesis and tumor progression, including activation of oncogenes, inactivation of oncogenes, genomic heterogeneity and molecular phenotype [17, 18]. We analyzed BMGs for gene amplification and deletion frequencies. We displayed a waterfall plot illustrating a relatively low mutation rate of the 20 genes in the TCGA cohort (Figure 3D). ADAM17, ITGAV, ADAM9 and MEP1A had widespread CNV amplification, while PHF13, P3H1, ROBO3, LOXL2 presented prevalent CNV deletions (Figure 3E), and the locations of CNV alterations on chromosomes showed as Figure 3F.

In a word, 20 genes correlated with BMs were screened from 66 intersection of genes, which revealed a noteworthy and statistically significant positive correlation with immune cell infiltration and involved in ADAM17, ITGAV, ADAM9 and MEP1A that conducted widespread amplification and PHF13, P3H1, ROBO3, LOXL2 that implemented prevalent CNV deletions.

BMs model development and validation in HCC

We applied multivariate Cox regression analysis to identify 4 genes (MMP1, ITGA2, P3H1, CTSA) to establish a risk model. The excellent formula, Risk score = (0.238*MMP1) + (0.212*ITGA2) + (0.400*P3H1) + (0.240*CTSA), was utilized to evaluate the risk score. The midpoint risk score serves for a criterion that can categorized into high- and low-risk profiles. The ICGC queue serves as an external validation set. Both the TCGA cohort (Figure 4A) and the ICGC cohort (Figure 4G) demonstrated that patients
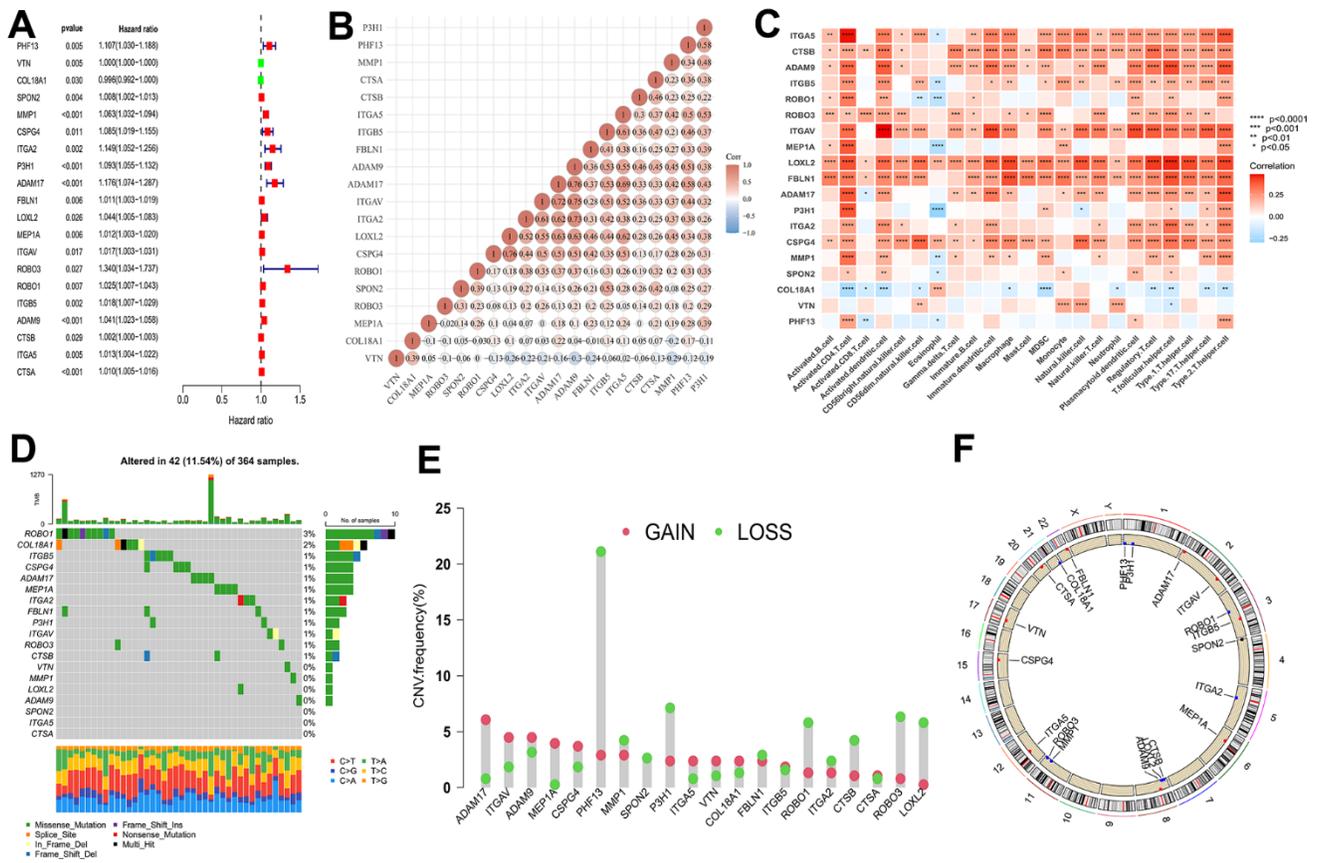
Figure 3. Multi-omics landscape by prognostic genes. (**A**) Univariate Cox regression analysis of genes. (**B**) Significant positive correlation between genes related to prognostic. (**C**) The correlation between genes related to prognostic and the infiltration level of 23 immune cells. (**D**) The mutation frequency of 20 BMGs in TCGA-LIHC cohort. Each column of the figure represents an individual patient. (**E**) The CNV variation frequency of BMGs (red and green plots separately represent CNV gain and CNV loss). (**F**) Locations of CNV alterations in BMGs on 23 chromosomes. * p < 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001.

with lower risk scores experienced more favorable outcomes, as depicted by the Kaplan-Meier curve. The four core risk genes MMP1, ITGA2, P3H1, and CTSA were utilized for PCA and t-SNE analysis which indicate that the BMs model shows bright differentiation (Figure 4B, 4C, 4H, 4I). The patients at an elevated risk were characterized by higher mortality and decreased survival from risk curves. The number of patients in the high-risk profile presented that increased number of deaths and the increased risk score were
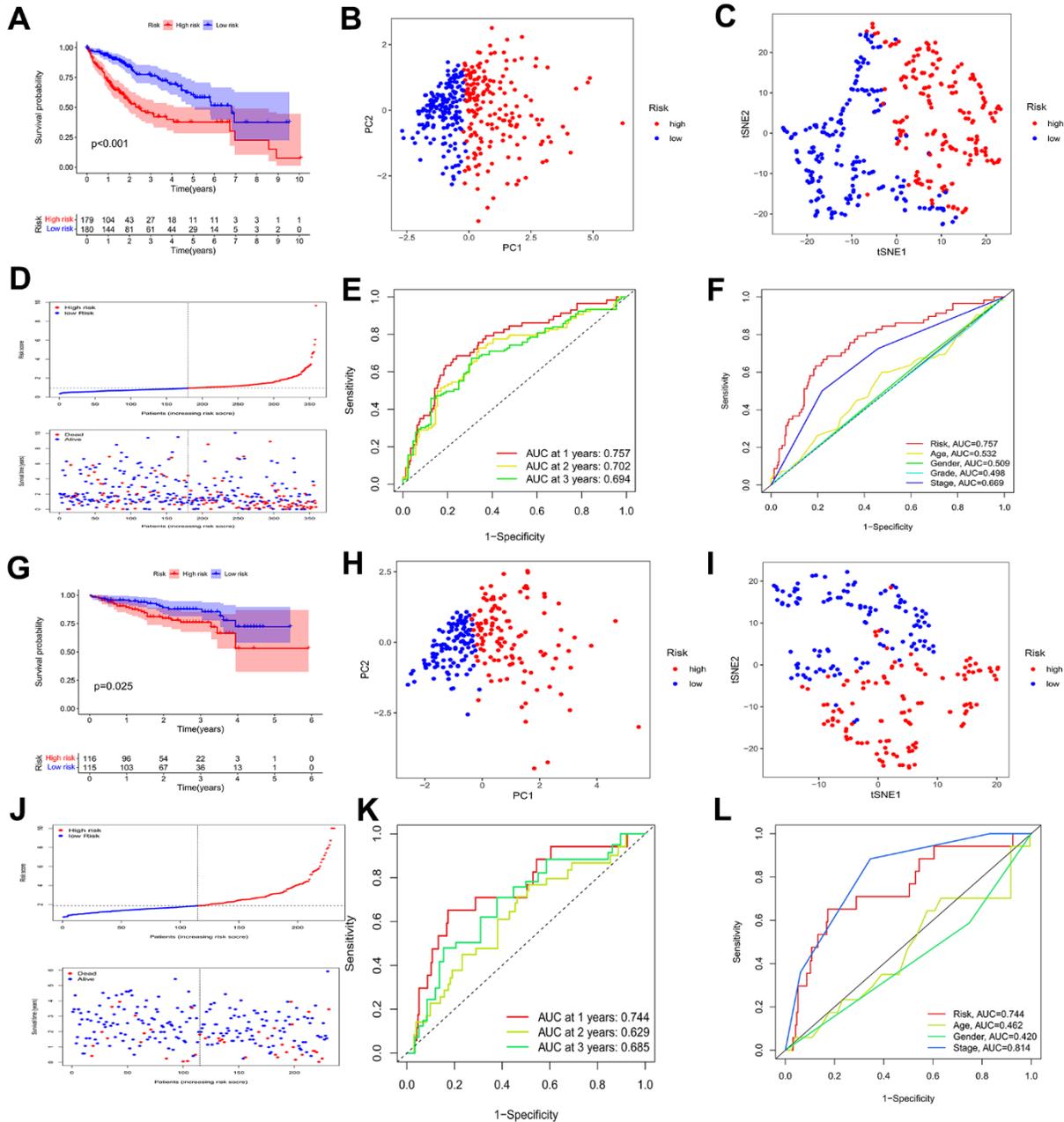
Figure 4. Identification of a prognostic risk model for HCC patients. (**A**) Kaplan–Meier curve of HCC patients in the TCGA cohort. (**B**, **C**) PCA and t-SNE analysis showing a remarkable difference in transcriptomes between the two risk categories in the TCGA cohort. (**D**) Scatter plots showing the risk score distribution and patient survival status in TCGA cohort. (**E**) ROC curves to predict the sensitivity and specificity of 1-, 2-, 3-year survival according to the risk score in the TCGA cohort. (**F**) Clinical ROC analysis in the TCGA cohort. (**G**) Kaplan–Meier curve of HCC patients in the ICGC cohort. (**H**, **I**) PCA and t-SNE analysis showing a remarkable difference in transcriptomes between the two risk categories in the ICGC cohort. (**J**) Scatter plots showing the risk score distribution and patient survival status in the ICGC cohort. (**K**) ROC curves to predict the sensitivity and specificity of 1-, 2-, 3-year survival according the risk score in the ICGC cohort. (**L**) Clinical ROC analysis in the ICGC cohort.

positive correlation (Figure 4D, 4J). In the TCGA cohort, we recognized that the area under curves (AUCs) in prognosticating 1-year, 2-year, and 3-year OS respectively were 0.757, 0.702, and 0.694 (Figure 4E). It is no difference with the results for the TCGA cohort, that AUC values for the 1-year, 2-year, and 3-year OS in the ICGC cohort were 0.744, 0.629, and 0.685, respectively (Figure 4K). In addition, in the TCGA-LIHC cohort, AUC values were significantly higher in BMs models than for age, sex, pathologic stage, and tumor stage (Figure 4F). Meanwhile, compared with the predictive power of age and gender, the BMs model exposed greater predictive power in the ICGC cohort (Figure 4L). The prognostic value of MMP1, ITGA2, P3H1, and CTSA in HCC are demonstrated using databases from TCGA-LIHC, ICGC, and GSE54236 HCC datasets (Supplementary Figures 1–3), which demonstrated that the four genes' predictive value for HCC prognosis to a certain extent, though not as significant as the risk model overall. Furthermore, we utilized GSE54236 HCC datasets to validate our risk model and the results have significant statistical significance (Supplementary Figure 4). Interestingly, the risk model was verified to have significantly predicted value not only for HCC prognosis, but also for the tumor doubling time, a critical event for tumor progression.

Broadly speaking, a new model of risk score about HCC was created to efficiently estimate individual prognosis by multivariate Cox regression analysis, which was validated by ICGC queue and was compared with other effective predictive clinicopathological indicators.

BMs model independence prognostic value, construction of nomograms and clinical correlation analysis

As depicted in Figure 5A, 5B, the TCGA-LIHC cohort revealed that BMs acted as an independent predictor of survival. This conclusion was drawn from the results obtained using both univariate and multivariate Cox regression models, indicating that BMs significantly influenced survival outcomes irrespective of other variables. Subsequently, nomograms of clinical adaptation were constructed by BMs models and other clinicopathological features, providing a visual method for prognosticating survival rates for the 1-year, 3-year, and 5-year time points for HCC (Figure 5C). Regarding predicting both short- and long-term survival, the nomogram has remarkable accuracy. The calibration plots of the column line plots revealed that the predictions of the column line plots and the actual observed probabilities were excellently coincident (Figure 5D). In addition, the net clinical benefit of BMs was higher than other clinicopathologic features in the

TCGA-LIHC cohort, as indicated by decision curve analysis (Figure 5E). Significantly, compared with other models, AUC value of the BMs possessed superior advantage (Figure 5F). Moreover, in the TCGA-LIHC cohort, we found out risk scores were extremely related with T status, pathologic grade, and tumor stage (Figure 6, P < 0.05).

Collectively, the nomogram suffering BMs exhibited remarkable forecast of survival outcomes in the short and long run in patients diagnosed with HCC, which could boast prognostic value and assist in clinical management. And the decision curve analysis contributed to clinical decision and received better benefit.

Analysis of tumor mutation burden and immune cell infiltration in different risk profiles

The waterfall plots depicted variations in mutations of the top 20 genes among distinct risk profiles in patients with HCC, and higher mutation frequencies belonged to the patients at an elevated risk. (Figure 7A, 7B). TP53 was significantly more frequently mutated regarding the patients at an elevated risk than in the patients at a decreased risk, in contrast CTNNB1 was greatly more frequently mutated regarding the patients with a low-risk profile, mainly consisting of frameshift deletions and nonsense mutations. Tumor mutation burden (TMB), known as a nonsynonymous variant, demonstrates a strong association with the penetration of immune cells and the elicitation of immune responses [19]. Individuals with HCC were divided into a lower TMB profile and a higher TMB profile thanks to the optimal threshold. Kaplan-Meier curves demonstrated that patients with lower TMB had better events (Figure 7C). Given TMB may make a big difference in clinical experiments, we sought to investigate the combined impact of TMB and BMs. TMB and BMs exhibited synergistic effects on survival outcomes of HCC patients (Figure 7D).

To investigate the attributes of the tumor immune microenvironment in the two risk categories, we investigated the expression profiles of 38 immune checkpoint inhibitor (ICI) genes, 20 HLA genes, and the frequency of infiltrating immune cell populations within the tumor. In the high-risk category (Figure 7E), we observed a substantial upregulation of immune checkpoint genes, indicative of heightened immunosuppressive signaling. Additionally, HLA-associated genes were prominently overexpressed in the same high-risk profile (Figure 7F), suggesting a potential role in modulating antigen presentation and immune recognition. Moreover, leveraging the "ssGSEA" algorithm, we quantified the abundance of penetration of immune cells, unveiling a significant

heterogeneity in the immune cell landscape between the two risk profiles, reflecting differential immune cell infiltration patterns and potentially distinct immune responses. Some immune cells levels were remarkably higher in the patients with a high-risk profile, which involved in activated CD4 T cells, activated dendritic cells, myeloid-derived suppressor cells (MDSCs), plasmacytoid dendritic cells, regulatory T cells,

T follicular helper cells, type 17 T helper cells, and type 2 T helper cells (Figure 7G), as found by the results of ssGSEA. Type I interferon (IFN) responses and type II IFN responses had an enrichment in the group of low risk, and antigen presenting cell (APC) costimulation, Chemokine receptors (CCR), major histocompatibility complex (MHC) class I, and para-inflammation showed enrichment in the group of high risk (Figure 7H).
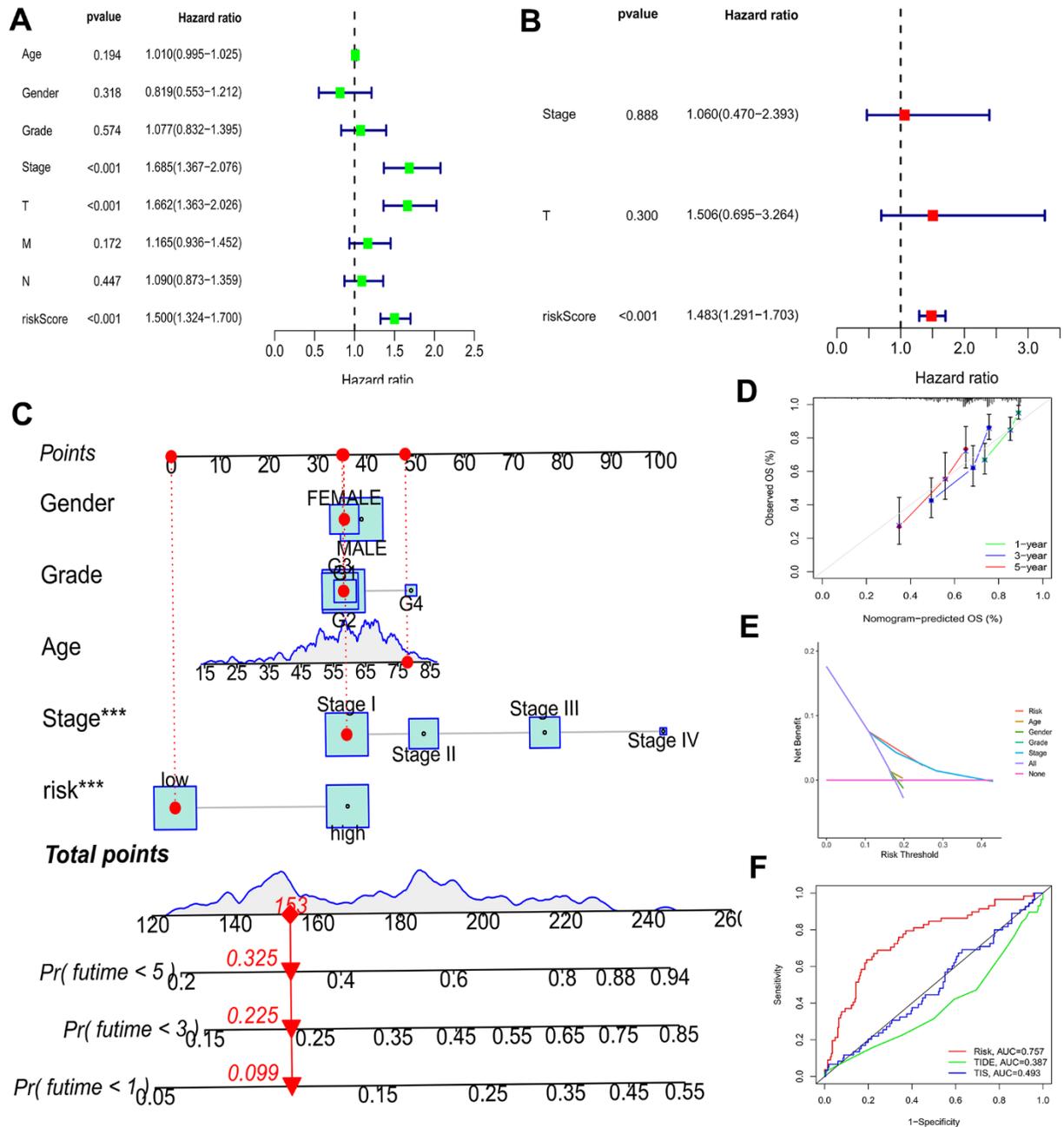
Figure 5. Efficacy estimate of prognostic model and nomogram. (**A**, **B**) Univariate Cox regression analysis of the model in the TCGA-LIHC cohort. (**C**) The nomogram consists of risk score and other clinicopathological parameters. (**D**) Calibration curves of the nomogram. (**E**) The decision curve analysis of the model in the TCGA-LIHC cohort. (**F**) AUC value of the BMs compared with other models in the TCGA cohort. *** p < 0.001.
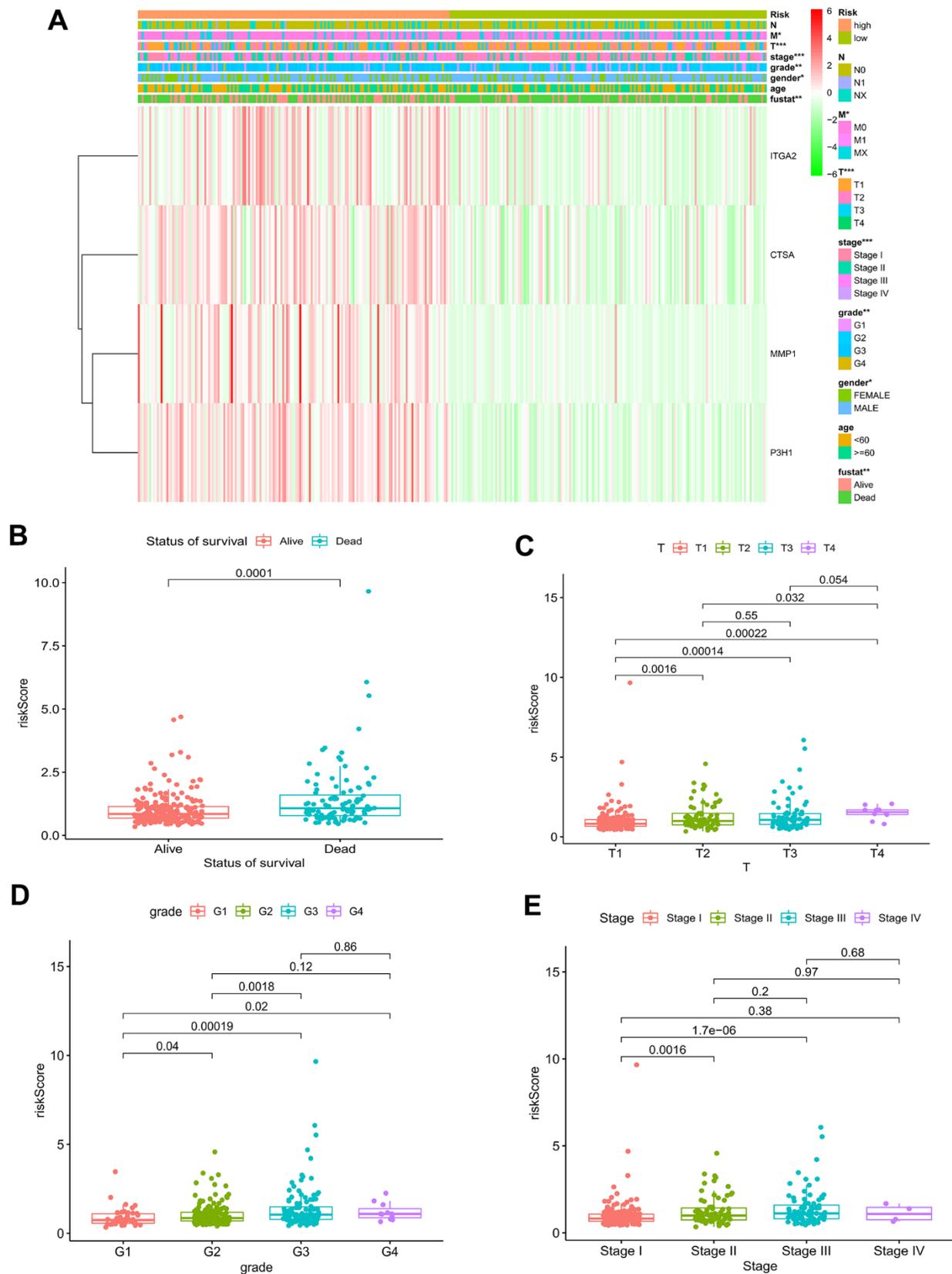
Figure 6. Relationship of risk model and clinical characteristics. (**A**) The heatmap of the model and clinical characteristics in the TCGA–LIHC cohort. Boxplot of risk score in HCC patients with different status of survival (**B**), T status (**C**), pathological grades (**D**), and tumor stages (**E**). * p < 0.05; ** p < 0.01; *** p < 0.001.
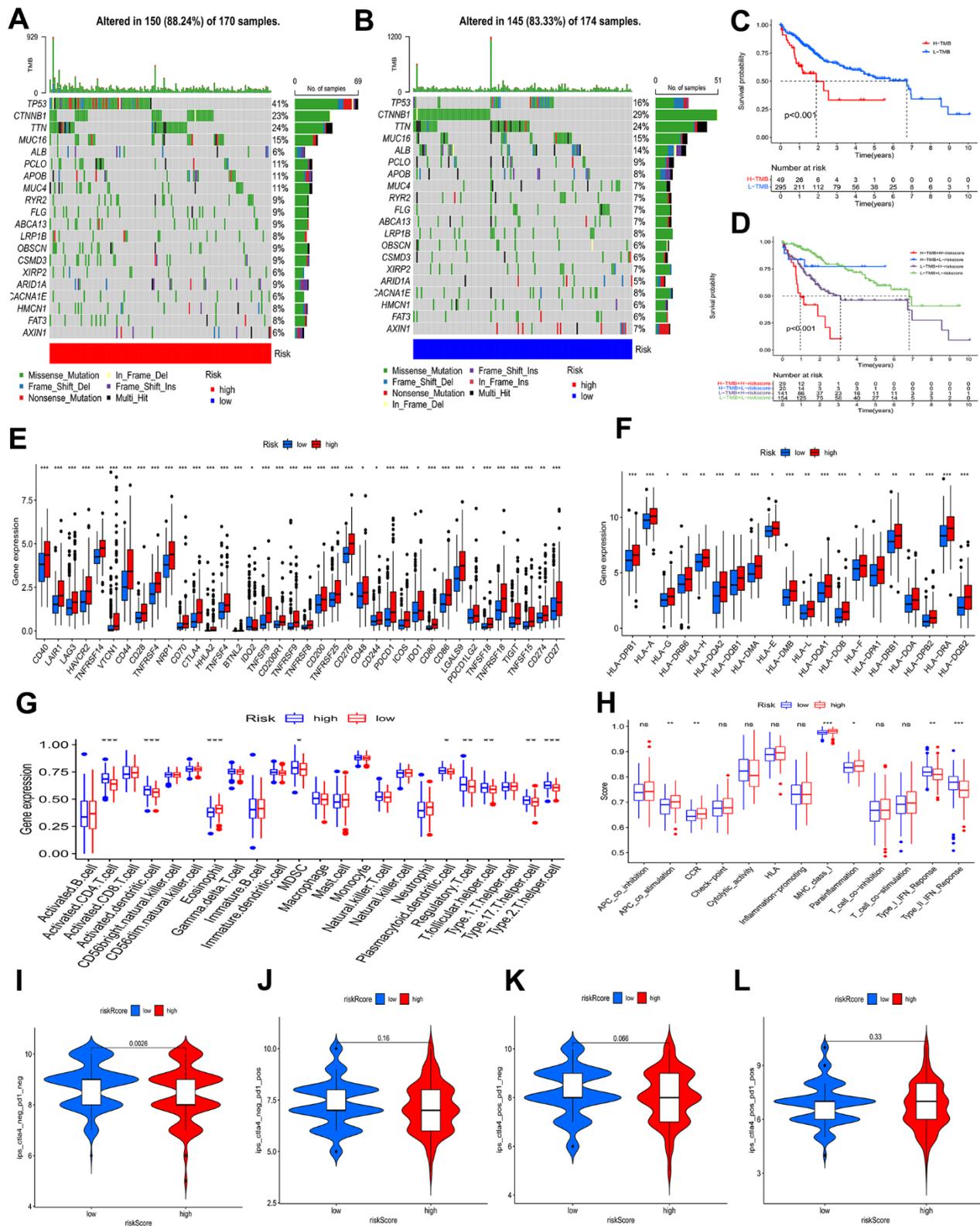
Figure 7. Analysis of tumor mutation and immune cell infiltration. (**A**, **B**) The waterfall plot showing the differences in somatic genomic mutation between high and low risk categories. (**C**, **D**) The Kaplan–Meier curve based on both TMB categories and the model for HCC patients. (**E**, **F**) Different risk categories expressed different levels of checkpoint genes and HLA genes. (**G**, **H**) Immune cell and immune function scores of different risk categories. Comparison of IPS between two risk categories. (**I**) CTLA4− PD1−, (**J**) CTLA4− PD1+, (**K**) CTLA4+ PD1−, and (**L**) CTLA4+ PD1+. ns p > 0.05; * p < 0.05; ** p < 0.01; *** p < 0.001.

The BMs Models predicted tumor immune microenvironment (TIME), with the patients at an elevated risk showing higher expression of 38 ICI genes, 20 HLA genes, and increased tumor-infiltrating immune cells. The high-risk profile also exhibited enrichment in immune processes like APC costimulation, CCR signaling, MHC class I pathway activation, and para-inflammation, potentially explaining the significant difference in prognosis between the subgroups.

The role of BMs in immunotherapy

An IPS analysis was conducted to further test the BMs model's ability to prognosticate the ICI efficacy and to determine immunotherapy sensitivity in HCC patients separately. A significantly higher IPS-CTLA4-neg-PD1-neg score, in the low-risk profile than in the high-risk profile, as shown in Figure 7I–7L, demonstrated that the immunogenic phenotype is stronger in the patients with a low-risk profile and therefore they may acquire benefits from immunotherapy.

In addition, the ratio of HCC immune subtypes in the different risk groups was compared. It was a statistically significant difference that the immunophenotype analysis presented by comparing the different groups (Figure 8A). It was further evaluated whether BMs could be utilized as a predictive factor of immunotherapy in HCC individuals. The patients with a low-risk profile exhibited increased rates of remission than the
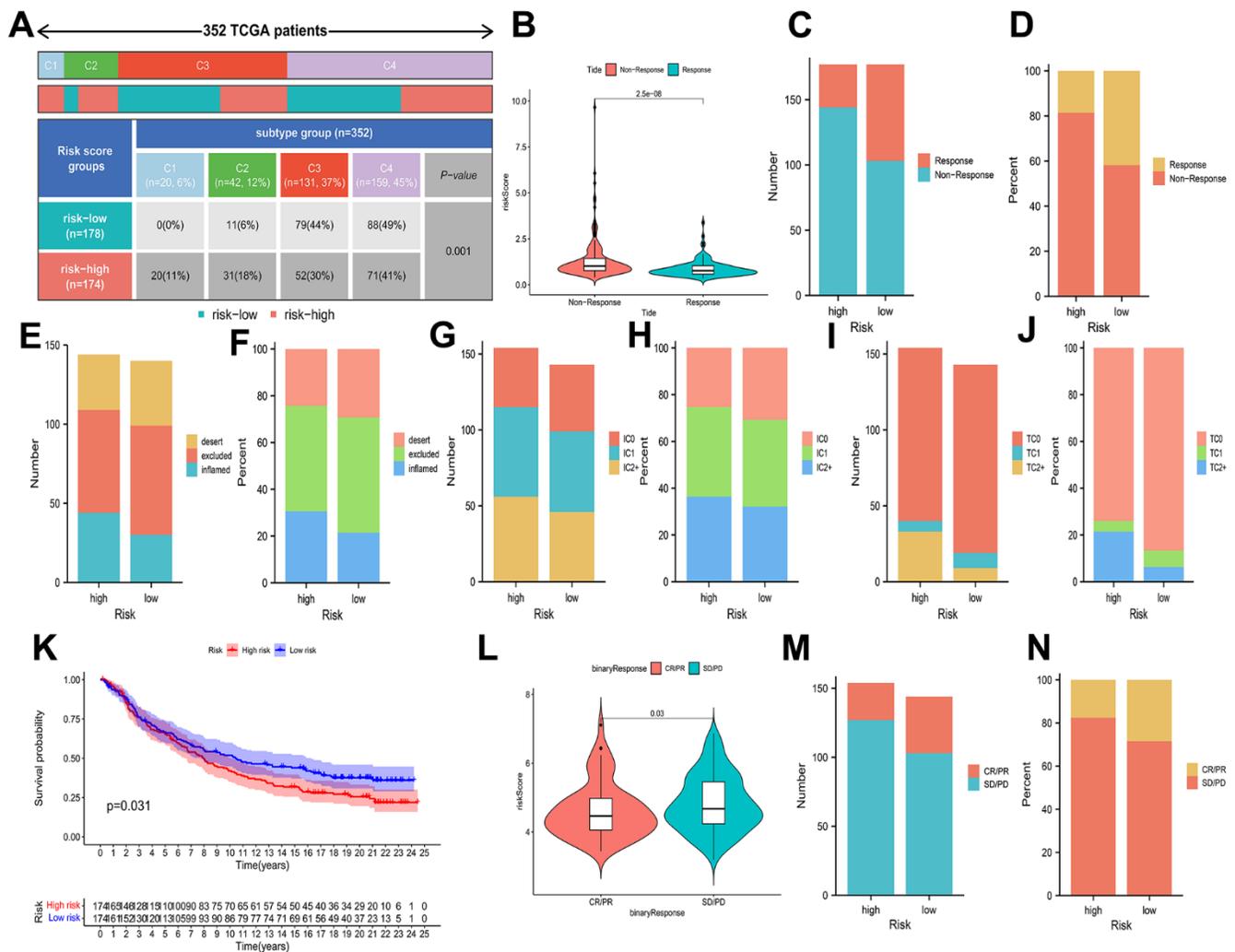
Figure 8. Associations of prognostic risk model and HCC immunotherapy. (**A**) A comparison of immune subtypes between different risk categories. (**B**) A comparison of risk scores between two response categories. (**C–J**) Stacked bar plot of rates of response, inflammatory immune subtypes, tumor-infiltrating immune cells, and PD-L1-expressing tumor tissue samples. (**K**) Kaplan-Meier curves of different risk categories. (**L**) A comparison of risk scores between the CR/PR category and stable disease (SD)/progressive disease (PD) category. (**M**, **N**) Stacked bar plot of CR/PR and SD/PD.

individuals with a high-risk profile (Figure 8B–8D). In the IMvigor210 cohort, 348 patients were split into high-risk and low-risk profiles, with metastatic uroepithelial carcinoma after the treatment of anti-PD-L1 drugs. Between the two profiles, statistically remarkable differences were observed in the proportions of inflammatory immune subtypes, tumor-infiltrating immune cells, and PD-L1-expressing tumor tissue samples (Figure 8E–8J). Kaplan-Meier curves presented that individuals with low-risk scores had better prognosis (Figure 8K). In comparison to the patients at an elevated risk, the patients at a decreased risk possessed a higher percentage of complete remission (CR)/partial remission (PR) (Figure 8L–8N).

Consequently, the response to immunotherapy varied between the two risk profiles defined by BMs, with patients in the low-risk category potentially exhibiting heightened sensitivity to immunotherapy, leading to more favorable clinical outcomes.

Drug sensitivity of BMs in HCC

With the aim of investigating the possible use of BMs in HCC personalized treatment and further facilitating the clinical value of BMs for dealing with HCC, we evaluated several chemotherapeutic agents for their IC_{50} values between different risk profiles. The IC_{50} value demonstrated an inverse relationship with the responsiveness of the chemotherapy drugs. The drug sensitivity demonstrated that the IC_{50} values of bleomycin, cisplatin, gemcitabine, mitomycin C, and paclitaxel in the elevated-risk category exhibited lower IC_{50} values than those in the decreased-risk category, implying that high risk individuals is likely to receive more beneficial effects from the above chemotherapies, while bosutinib, cyclopamine, dasatinib, docetaxel, metformin, methotrexate and rapamycin in the low-risk profile were lower, indicating that the decreased-risk category may benefit more from the above chemotherapies (Figure 9A–9L).
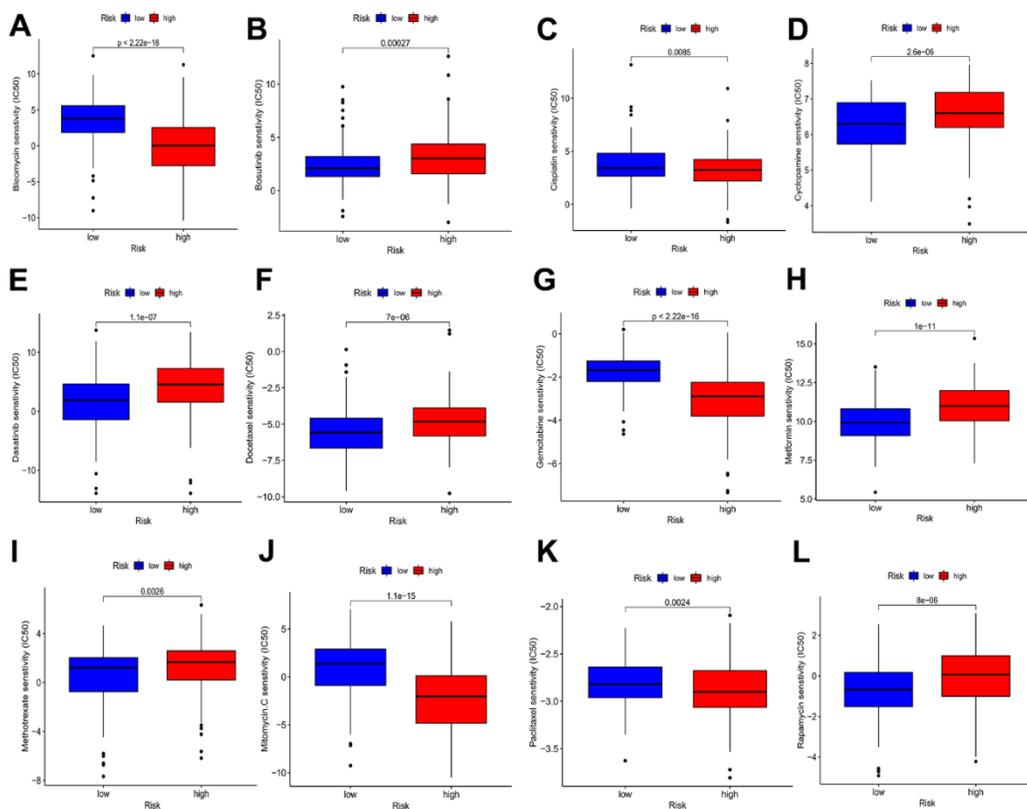
Figure 9. Drug sensitivity analyses among high- and low-risk categories. (**A**) Estimated IC_{50} of bleomycin in the elevated-risk category were lower than in the decreased-risk category. (**B**) Estimated IC_{50} of bosutinib in the elevated-risk category were higher than in the decreased-risk category. (**C**) Estimated IC_{50} of cisplatin in the elevated-risk category were lower than in the decreased-risk category. (**D**, **E**, **F**) Estimated IC_{50} of cyclopamine, dasatinib and docetaxel in the elevated-risk category were higher than in the decreased-risk category. (**G**) Estimated IC_{50} of gemcitabine in the elevated-risk category were lower than in the decreased-risk category. (**H**, **I**) Estimated IC_{50} of metformin and methotrexate in the elevated-risk category were higher than in the decreased-risk category. (**J**, **K**) Estimated IC_{50} of mitomycin C and paclitaxel in the elevated-risk category were lower than in the decreased-risk category. (**L**) Estimated IC_{50} of rapamycin in the elevated-risk category were higher than in the decreased-risk category.

Based on these results, the risk score has the potential to guide patients towards receiving more tailored chemotherapies.

The differential expression of BMs model in cell lines of HCC

In the preliminary work, a risk model based on BMs-related gene was constructed. Now, we further verified these results *in vitro*, 4 genes of the model were subjected to qRT-PCR. Compared with LO2 cells, the expression levels of 4 genes in five types of HCC cells were significantly differential (Figure 10A–10D).

DISCUSSION

Considered as a major global health burden, liver cancer is expected to affect more than one million people by 2025 [20]. Generally, cancer of the liver develops from chronic hepatitis with infiltration by various immune cells. Its aggressiveness, heterogeneity

and usually advanced stage of diagnosis contribute to its poor prognosis [21]. Currently, western immune checkpoint inhibitors offer great hope for patients with HCC. Therefore, new and reliable models and a better understanding of the link between models and cancer immunity are essential for patients with HCC.

Greater understanding of BMs has helped to make them a hot topic of research over the past years. However, no studies have investigated the predictive value of BMs in cancer. Taking advantage of the genetic characteristics of genes associated with BMs, our novel model was developed for predicting survival and treatment decisions in HCC patients.

In our research, amount to 222 BMGs were first acquired from the available literature. Then, we obtained the tumor-associated genes of HCC by WGCNA. Identification of 66 BMGs intersected with tumor-associated genes for further study. When employing GO and KEGG enrichment analysis, we
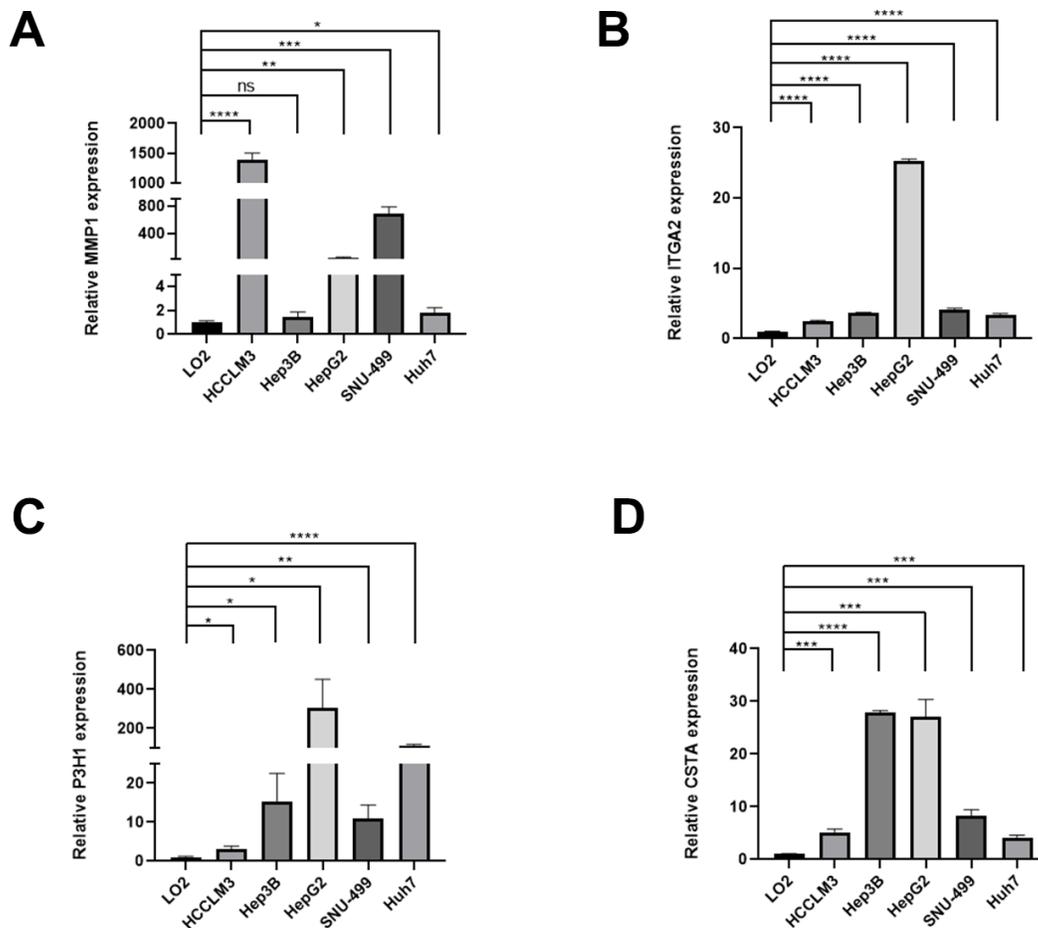
Figure 10. The expression differences of normal liver and HCC cell lines. The expression differences of MMP1 (**A**), ITGA2 (**B**), P3H1 (**C**), and CTSA (**D**) in normal liver and HCC cell lines.

determined these genes enriched in ECM-receptor interactions, adherent spots, and the PI3K-Akt signaling pathway. The prognostic significance of these genes was evaluated, and subsequently, we recognized 20 genes associated with prognosis through a univariate Cox regression analysis. Utilizing a multivariate Cox regression analysis, we constructed a BMs correlation model and validated its performance using the ICGC cohort. Risk Score = 0.238*MMP1 + 0.212*ITGA2 + 0.400*P3H1 + 0.240*CTSA is the result of this algorithm. According to their median risk score, patients with HCC in the TCGA and ICGC cohorts were separated into elevated-risk and decreased-risk categories. Higher risk scores were corresponding to poorer individuals results, as determined by Kaplan-Meier curves. The ROC curves demonstrated the model's accuracy. The BMs model was a predictor of survival in univariate and multivariate Cox regression investigations for the TCGA-LIHC cohort. As a result, we additionally implemented the risk model to determine the connection between tumor mutational load and immune cell infiltration. The waterfall plot illustrates that the high-risk patient profile manifested a greater prevalence of mutations in the top 20 genes. The elevated-risk category possessed significantly more TP53 mutations than the decreased-risk category, whereas the low-risk profile featured significantly more CTNNB1 mutations. The expression of immune checkpoint genes appeared to be greater in the high-risk profile. 20 HLA genes and 38 ICI genes encountered high levels of expression in the low-risk individuals. More intriguingly, IPS-CTLA4-neg-PD1-neg scores in the profile with a lower risk and the profile with a higher risk significantly differed, indicating a more immunogenic phenotype in those in the low-risk category. In contrast, the proportion of immune cells between the two categories. However, several limitations of our study do exist. First, more thorough research should be done to confirm the model's predictive utility in a clinical environment. As our research was only based on retrospective data from public databases, prospective studies will be required in the future to substantiate our discoveries.

MMP1, ITGA2, P3H1 and CTSA were recruited in this BMs risk model of HCC. As we know, MMP1 is identified as a matrix metalloproteinase capable of degrading various components of the extracellular matrix (ECM), facilitating tumor cell invasion and metastasis. It promotes ECM remodeling, assisting tumor cells in breaching the basement membrane and surrounding tissue, which in turn supports angiogenesis and tumor growth [22]. ITGA2, a member of the integrin family, plays a role in cell-cell and cell-matrix interactions. Wang et al. reported ITGA2 may promote HCC progression by activating cell adhesion, migration,

and survival pathways, such as FAK/PI3K/Akt pathway and modulating the interactions between tumor cells and their microenvironment [23]. The specific role of P3H1 in HCC is not as well elucidated as MMP1 or ITGA2, which deserved the further explored in the future [24]. As reported, CTSA is overexpressed in various types of cancer and is linked to poor clinical outcomes. The high protein level of CTSA was significantly correlated to the poor clinicopathological parameters, such as TNM stage, vascular invasion, tumor recurrence, and patient death of HCC [25]. Overall, further experiments need to be carried out to better understand the mechanism of these genes in HCC.

In conclusion, we identified a BMs-related model for HCC patients in the present investigation. What's more, the same model was validated to forecast HCC patients' OS and demonstrated excellent function of prediction. We also investigated disparities in immunotherapy response and sensitivity to chemotherapeutic medications among BMs risk categories. Our study contributes to the advancement of understanding immune infiltration characteristics and innovative approaches for personalized treatment.

MATERIALS AND METHODS

Data source

In this work, we obtained 222 BMGs from existing literature [26]. Additionally, we downloaded clinical data, genomic mutation profiles, and RNA-seq expression data regarding HCC patients from the Cancer Genome Atlas (TCGA) database. Additionally, we leveraged the gene expression data from the IMvigor210 dataset and the corresponding clinical information of HCC patients undergoing immune checkpoint inhibitor therapy. Utilizing the "CoreBiologies" R package, we corroborated that our risk model had the capability to anticipate the response of HCC patients to immunotherapy.

WGCNA

WGCNA can effectively identify gene set modules that display a strong association. The parameter setting and steps in WGCNA analysis were as followed. Initialization and data preparation: utilization of the WGCNA and limma packages, specifying samples. Data pre-processing: numerical conversion of gene expression data, application of avereps function for log2 transformation, and filtering genes with a standard deviation less than 0.5. Sample clustering: hierarchical clustering based on expression data, using average linkage method, and detection of outliers through sample dendrogram. Network construction:

selection of an appropriate soft threshold (power) with pickSoftThreshold, construction of adjacency and TOM matrices, followed by gene clustering. Module identification: dynamic tree cutting for gene module identification, calculation of module eigengenes (MEs), and module similarity clustering. Gene significance and module membership: calculation of gene significance for clinical traits and module membership, with visualization through Risk Score Triad Diagrams for each clinical feature and module.

In this study, we used WGCNA to calculate correlation weights of coexpression and identify coexpressed genes associated with tumors [27]. Rather than solely relying on differential gene expression, WGCNA allowed us to pinpoint relevant gene sets and conduct rigorous association analyses with phenotypic traits. This approach effectively shifted the challenge of correcting for multiple hypothesis testing from handling correlations between thousands of individual genes and phenotypes to examining the associations of a few gene sets with the phenotypes. A scale-free network was generated by specifying the optimal soft threshold power. To continue, we computed gene dendrograms and modules utilizing the 50-min cluster size based on the topological overlap matrix TOM dissimilarity. We identified the most significant module associated with HCC and used it for further analysis. We found the intersection genes of the module genes that related with HCC and BMGs by Venn plot.

Functional and pathway enrichment analysis

To assess the characteristic functional attributes of the two risk groups, we conducted an analysis of the shared genes between the two groups and subsequently annotated them using the "clusterProfiler" R package with information from the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases [28].

Prognostic, correlation and mutation analysis

The prognosis-related BMGs were checked out using univariate Cox proportional hazard regression. The prognosis-related BMGs correlation pheatmap is displayed by the R software package. Furthermore, we utilized the "ggplot" R package to explore the relationship between BMGs and the abundance of immune cells within the tissue. We established a significance threshold of P<0.05 for statistical significance. The R package "maftools" then showed the mutation status of prognosis-related BMGs [29]. In addition, the frequency of copy number variations (CNVs) in prognosis-related BMGs was calculated.

The chromosome positions of these genes were visualized through "RCircos" software package.

Construction and validation of BMs model

Using multivariate Cox regression, we developed a robust prognostic model for BMGs. External validation was performed on the International Cancer Genome Consortium (ICGC) cohort. Patients were categorized into high- and low-risk profiles based on the median BMGs expression score. Kaplan-Meier curves were employed for precise comparison of overall survival (OS) between the two risk profiles, offering critical insights for tailored treatment approaches in HCC patients. The classification results were validated by using principal component analysis (PCA) and t-distribution random neighborhood embedding (t-SNE) to obtain low-dimensional clustering distributions from high-dimensional gene sets. "survivalROC" and "timeROC" R packages were used to quantify the prognostic model's forecast value. To better understand the prognostic nomogram using the "regplot" package in this article, we showed the constructed details. Cox proportional hazards model construction: survival data were analyzed using the Cox proportional hazards model, with risk scores and clinical variables serving as covariates. Nomogram generation: utilizing the Cox model's results, nomograms were created via the regplot function to visualize the impact of risk scores on survival probabilities. Calibration curve generation: calibration curves were produced using the calibrate function, comparing the predicted survival probabilities from nomograms against observed survival rates to evaluate the model's predictive performance. This study involved constructing a prognostic nomogram using the "regplot" package, incorporating clinical factors and gene signatures to estimate 1-, 3-, and 5-year OS probabilities. To assess the predictive capacity of column line graphs, 3-year and 5-year calibration curves were generated. Furthermore, decision curve analysis was utilized to assess the clinical applicability of the risk score model by computing the net benefit across different risk thresholds [30]. Additionally, we also compared 1-year receiver operating characteristic (ROC) value of the model with other models.

Exploration of immune infiltration

Tumor microenvironment (TME) cells, a crucial component of tumor tissue, hold significant clinicopathological importance in prognosticating outcomes and treatment effectiveness, supported by emerging evidence. Further research was completed on the expression of human leukocyte antigen (HLA) genes and common immunological checkpoints in various clusters. To compare the immune infiltration

Table 1. Sequences of primers of BMs model and GAPDH.

Gene	Sequences of primers
GAPDH	Forward Primer: ACAACTTTGGTATCGTGGAAGG
	Reverse Primer: GCCATCACGCCACAGTTTC
MMP1	Forward Primer: AAAATTACACGCCAGATTTGCC
	Reverse Primer: GGTGTGACATTACTCCAGAGTTG
ITGA2	Forward Primer: CCTACAATGTTGGTCTCCCAGA
	Reverse Primer: AGTAACCAGTTGCCTTTTGGATT
P3H1	Forward primer: CAGCTCGAGCGGGACAG
	Reverse primer: AGGTCCATCTCTTCTGGGCT
CTSA	Forward Primer: GTCGCCCAGAGCAATTTTGAG
	Reverse Primer: TCTCCCCGGTCAGGAAAAGTT

and immunological function of each group, a Wilcoxon test with two samples was performed. The single-sample gene set enrichment analysis (ssGSEA) algorithm was applied for immunological inverse fold product analysis of 23 immune cell type gene sets, enabling exploration of TME infiltration in high-risk and low-risk profiles [31]. The 23 different TME cell types and the risk category were correlated via Spearman correlation analysis.

Estimation of immunotherapeutic response prediction

The liver hepatocellular carcinoma (LIHC) project of The Cancer Immunome Atlas is where the immunophenoscore (IPS) of HCC samples was developed. CTLA4 and PD-1 blockers are part of this project's method for forecasting the effectiveness of immunotherapies [31]. Moreover, the immune subtypes of HCC in the TCGA database have been described in a previous study [32] and we analyzed immune subtype proportions between the two risk groups based on UCSC-Xena database while using the "RColorBrewer" package to provide color schemes and create visualizations of immune clustering results. Additionally, we confirmed the link between immune checkpoint inhibitors and HCC risk markers thanks to the IMvigor210 dataset.

Chemotherapeutic drug sensitivity analysis for different risk groups

We conducted a pharmacore sensitivity analysis using the "prophecy" and "ggplot2" packages and compared the sensitivity of various chemotherapeutic agents between the groups of high risk and groups of low risk utilizing the Wilcoxon signed-rank test in the absence of official biomarkers to accurately predict the reaction of chemotherapeutic agents in patients with HCC. The half-maximal inhibitory concentrations (IC_{50}) of different chemotherapeutic drugs were compared between HCC high-risk category and low-risk category [33].

Cell culture

Human hepatic cell line LO2 and five human HCC cell lines (HCCLM3, Hep3B, HepG2, SNU-499 and Huh7 cells) were flashed frozen in liquid nitrogen with 2 ml tubes and stored at −80° C. All cell lines were cultured in 10% fetal bovine serum (FBS; Invitrogen, Carlsbad, CA, USA) in DMEM medium. All cell lines grew in a humid environment of 37° C, 5% CO_2, 99% relative humidity and did not contain antibiotics. The cells were subcultured at a ratio of 1:2 or 1:3 when they reached 80% confluence.

RNA extraction and quantitative real-time PCR (qRT-PCR)

TRIzol reagent kit (Invitrogen) was used to extract the total RNA from logarithmic growth cells. cDNA was synthesized using PrimeScript RT reagent kit (Takara Biotechnology, Dalian, China). QRT-PCR analysis was conducted using TB Green Premix Ex Taq II kit (Takara Biotechnology, Dalian, China) according to the instructions, three replicates were set in each well. All operations were carried out on ice. The $2^{-\Delta\Delta Ct}$ method was used for qRT-PCR analysis. The primers were designed using the Primer Bank (https://pga.mgh.harvard.edu/primerbank/) and the NCBI primer-BLAST tool (https://www.ncbi.nlm.nih.gov/). The sequences of primers were shown as Table 1, GAPDH was used as reference genes.

Statistical analysis

Spearman's rank correlation was employed to investigate the relationship between TIP scores and immune traits. Kaplan-Meier analyses and log-rank tests, facilitated by the "survival" package, assessed survival outcomes. Additionally, the Pearson Chi-square test explored the relationship between the TIPRGPI group and clinico-pathological variables. For univariate and multivariate determination of independent prognostic indicators, the

"survival" package was studied. The optimal threshold for survival analysis was determined using the "survminer" R package. All statistical analyses were conducted using R software (version 3.6.1). $p < 0.05$ was accepted as indicative of significant differences if there is no other case.

Data availability statement

The original contributions presented in the study are included in the article. Further inquiries can be directed to the corresponding author.

Abbreviations

HCC: hepatocellular carcinoma; TCGA: the Cancer Genome Atlas Database; ICGC: International Cancer Genome Consortium; BMGs: basement membrane-related genes; BMs: Basement membranes; TME: tumor micro-environment; LIHC: liver hepatocellular carcinoma; ssGSEA: single-sample gene set enrichment analysis; ROC: receiver operating characteristic; CNVs: copy number alterations; PCA: principal component analysis; GO: Gene ontology; KEGG: The Kyoto Encyclopedia of Genes and Genomes; OS: overall survival; APC: antigen presenting cell; MHC: major histocompatibility complex; HLA: human leukocyte antigen; AUC: area under curve; ICI: immune checkpoint inhibitors; IC_{50}: the half-maximal inhibitory concentrations; t-SNE: t-distribution random neighborhood embedding; IPS: immunophenoscore; CC: cellular component; BP: biological processes; MF: molecular function; TMB: tumor mutation burden; CCR: chemokine receptors; MDSCs: myeloid-derived suppressor cells; CR: complete response; PR: partial response; SD: stable disease; PD: progressive disease; IFN: interferon; TIME: tumor immune microenvironment; ECM: extracellular matrix.

AUTHOR CONTRIBUTIONS

XL and AL designed and monitored the research. GW initially screened and analyzed the data. GW, FL and DG performed the qRT-PCR experiment and drafted the manuscript. KX performed statistical analyses. DZ, RH and XW revised the manuscript. All authors approved the submitted version.

ACKNOWLEDGMENTS

For free use of TCGA-LICH, ICGC, GEO dataset and R software, we would like to express our thanks.

CONFLICTS 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.

FUNDING

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021; 71:209–49.
https://doi.org/10.3322/caac.21660 PMID:33538338

2. Woodrell CD, Hansen L, Schiano TD, Goldstein NE. Palliative Care for People With Hepatocellular Carcinoma, and Specific Benefits for Older Adults. Clin Ther. 2018; 40:512–25.
https://doi.org/10.1016/j.clinthera.2018.02.017 PMID:29571567

3. de Lope CR, Tremosini S, Forner A, Reig M, Bruix J. Management of HCC. J Hepatol. 2012; 56 Suppl 1:S75–87.
https://doi.org/10.1016/S0168-8278(12)60009-9 PMID:22300468

4. Qiu Z, Li H, Zhang Z, Zhu Z, He S, Wang X, Wang P, Qin J, Zhuang L, Wang W, Xie F, Gu Y, Zou K, et al. A Pharmacogenomic Landscape in Human Liver Cancers. Cancer Cell. 2019; 36:179–93.e11.
https://doi.org/10.1016/j.ccell.2019.07.001 PMID:31378681

5. Ozbek S, Balasubramanian PG, Chiquet-Ehrismann R, Tucker RP, Adams JC. The evolution of extracellular matrix. Mol Biol Cell. 2010; 21:4300–5.
https://doi.org/10.1091/mbc.E10-03-0251 PMID:21160071

6. Sanes JR. The basement membrane/basal lamina of skeletal muscle. J Biol Chem. 2003; 278:12601–4.
https://doi.org/10.1074/jbc.R200027200 PMID:12556454

7. Pozzi A, Yurchenco PD, Iozzo RV. The nature and biology of basement membranes. Matrix Biol. 2017; 57–8:1–11.
https://doi.org/10.1016/j.matbio.2016.12.009 PMID:28040522

8. Yurchenco PD. Basement membranes: cell scaffoldings and signaling platforms. Cold Spring Harb Perspect Biol. 2011; 3:a004911.

https://doi.org/10.1101/cshperspect.a004911
PMID:21421915

9. Murphy-Ullrich JE, Sage EH. Revisiting the matricellular concept. Matrix Biol. 2014; 37:1–14.
https://doi.org/10.1016/j.matbio.2014.07.005
PMID:25064829

10. Nyström A, Bornert O, Kühl T. Cell therapy for basement membrane-linked diseases. Matrix Biol. 2017; 57–8:124–39.
https://doi.org/10.1016/j.matbio.2016.07.012
PMID:27609402

11. Naba A, Clauser KR, Whittaker CA, Carr SA, Tanabe KK, Hynes RO. Extracellular matrix signatures of human primary metastatic colon cancers and their metastases to liver. BMC Cancer. 2014; 14:518.
https://doi.org/10.1186/1471-2407-14-518
PMID:25037231

12. Tsilibary EC. Microvascular basement membranes in diabetes mellitus. J Pathol. 2003; 200:537–46.
https://doi.org/10.1002/path.1439
PMID:12845621

13. Randles MJ, Lausecker F, Kong Q, Suleiman H, Reid G, Kolatsi-Joannou M, Davenport B, Tian P, Falcone S, Potter P, Van Agtmael T, Norman JT, Long DA, et al. Identification of an Altered Matrix Signature in Kidney Aging and Disease. J Am Soc Nephrol. 2021; 32:1713–32.
https://doi.org/10.1681/ASN.2020101442
PMID:34049963

14. Oudart JB, Monboisse JC, Maquart FX, Brassart B, Brassart-Pasco S, Ramont L. Type XIX collagen: A new partner in the interactions between tumor cells and their microenvironment. Matrix Biol. 2017; 57–8:169–77.
https://doi.org/10.1016/j.matbio.2016.07.010
PMID:27491275

15. Ramovs V, Te Molder L, Sonnenberg A. The opposing roles of laminin-binding integrins in cancer. Matrix Biol. 2017; 57–8:213–43.
https://doi.org/10.1016/j.matbio.2016.08.007
PMID:27562932

16. Prakash MD, Munoz MA, Jain R, Tong PL, Koskinen A, Regner M, Kleifeld O, Ho B, Olson M, Turner SJ, Mrass P, Weninger W, Bird PI. Granzyme B promotes cytotoxic lymphocyte transmigration via basement membrane remodeling. Immunity. 2014; 41:960–72.
https://doi.org/10.1016/j.immuni.2014.11.012
PMID:25526309

17. Shlien A, Malkin D. Copy number variations and cancer. Genome Med. 2009; 1:62.
https://doi.org/10.1186/gm62
PMID:19566914

18. Clifford RJ, Zhang J, Meerzaman DM, Lyu MS, Hu Y, Cultraro CM, Finney RP, Kelley JM, Efroni S, Greenblum SI, Nguyen CV, Rowe WL, Sharma S, et al. Genetic variations at loci involved in the immune response are risk factors for hepatocellular carcinoma. Hepatology. 2010; 52:2034–43.
https://doi.org/10.1002/hep.23943 PMID:21105107

19. Tang B, Zhu J, Zhao Z, Lu C, Liu S, Fang S, Zheng L, Zhang N, Chen M, Xu M, Yu R, Ji J. Diagnosis and prognosis models for hepatocellular carcinoma patient's management based on tumor mutation burden. J Adv Res. 2021; 33:153–65.
https://doi.org/10.1016/j.jare.2021.01.018
PMID:34603786

20. Llovet JM, Kelley RK, Villanueva A, Singal AG, Pikarsky E, Roayaie S, Lencioni R, Koike K, Zucman-Rossi J, Finn RS. Hepatocellular carcinoma. Nat Rev Dis Primers. 2021; 7:6.
https://doi.org/10.1038/s41572-020-00240-3
PMID:33479224

21. Dang H, Takai A, Forgues M, Pomyen Y, Mou H, Xue W, Ray D, Ha KC, Morris QD, Hughes TR, Wang XW. Oncogenic Activation of the RNA Binding Protein NELFE and MYC Signaling in Hepatocellular Carcinoma. Cancer Cell. 2017; 32:101–14.e8.
https://doi.org/10.1016/j.ccell.2017.06.002
PMID:28697339

22. Kim E, Kim D, Lee JS, Yoe J, Park J, Kim CJ, Jeong D, Kim S, Lee Y. Capicua suppresses hepatocellular carcinoma progression by controlling the ETV4-MMP1 axis. Hepatology. 2018; 67:2287–301.
https://doi.org/10.1002/hep.29738
PMID:29251790

23. Wang L, Gao Y, Zhao X, Guo C, Wang X, Yang Y, Han C, Zhao L, Qin Y, Liu L, Huang C, Wang W. HOXD3 was negatively regulated by YY1 recruiting HDAC1 to suppress progression of hepatocellular carcinoma cells via ITGA2 pathway. Cell Prolif. 2020; 53:e12835.
https://doi.org/10.1111/cpr.12835
PMID:32557953

24. Wang J, Xu X, Liu Z, Wei X, Zhuang R, Lu D, Zhou L, Xie H, Zheng S. LEPREL1 Expression in Human Hepatocellular Carcinoma and Its Suppressor Role on Cell Proliferation. Gastroenterol Res Pract. 2013; 2013:109759.
https://doi.org/10.1155/2013/109759
PMID:24319452

25. Luo L, Wang X, Wang H, Yang C, Zhang Y, Li X, Xu Z. High cathepsin A protein expression predicts poor prognosis and tumor recurrence of hepatocellular carcinoma patients after curative hepatectomy. Am J Cancer Res. 2022; 12:3843–56.

PMID:36119821

26. Jayadev R, Morais MR, Ellingford JM, Srinivasan S, Naylor RW, Lawless C, Li AS, Ingham JF, Hastie E, Chi Q, Fresquet M, Koudis NM, Thomas HB, et al, and Genomics England Research Consortium. A basement membrane discovery pipeline uncovers network complexity, regulators, and human disease associations. Sci Adv. 2022; 8:eabn2265. https://doi.org/10.1126/sciadv.abn2265 PMID:35584218

27. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9:559. https://doi.org/10.1186/1471-2105-9-559 PMID:19114008

28. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012; 16:284–7. https://doi.org/10.1089/omi.2011.0118 PMID:22455463

29. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. 2018; 28:1747–56. https://doi.org/10.1101/gr.239244.118 PMID:30341162

30. Van Calster B, Wynants L, Verbeek JF, Verbakel JY, Christodoulou E, Vickers AJ, Roobol MJ, Steyerberg EW. Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators. Eur Urol. 2018; 74:796–804. https://doi.org/10.1016/j.eururo.2018.08.038 PMID:30241973

31. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, Hackl H, Trajanoski Z. Pancancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. Cell Rep. 2017; 18:248–62. https://doi.org/10.1016/j.celrep.2016.12.019 PMID:28052254

32. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, Porta-Pardo E, Gao GF, Plaisier CL, Eddy JA, Ziv E, Culhane AC, Paull EO, et al, and Cancer Genome Atlas Research Network. The Immune Landscape of Cancer. Immunity. 2018; 48:812–30.e14. https://doi.org/10.1016/j.immuni.2018.03.023 PMID:29628290

33. Geeleher P, Cox N, Huang RS. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. PLoS One. 2014; 9:e107468. https://doi.org/10.1371/journal.pone.0107468 PMID:25229481

Supplementary Figures
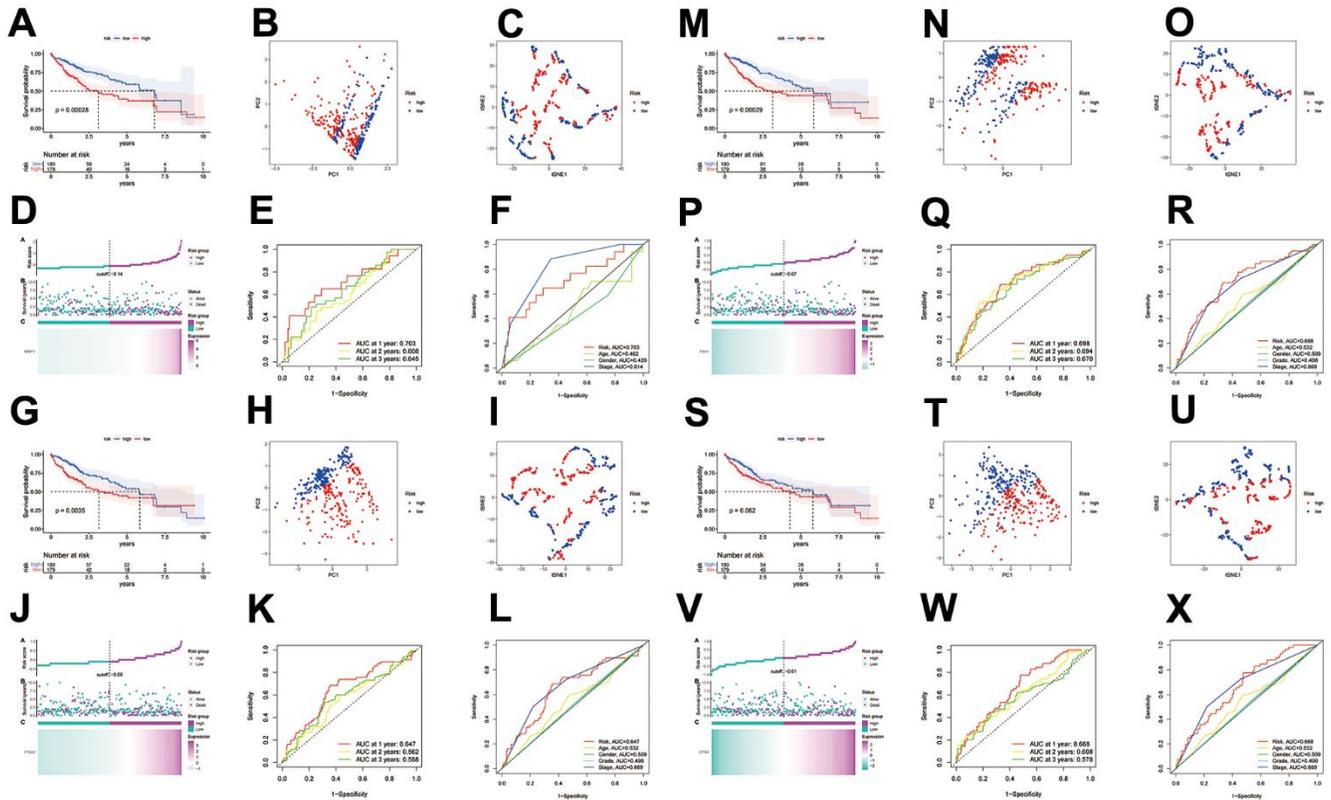
Supplementary Figure 1. The TCGA database was utilized to individually validate the prognostic value of MMP1, ITGA2, P3H1, and CTSA in hepatocellular carcinoma. (**A**) Kaplan-Meier curve of HCC patients in the MMP1. (**B**, **C**) PCA and t-SNE analysis showing a remarkable difference in transcriptomes between the two risk categories in the MMP1. (**D**) Risk Score Triad Diagrams showing the risk score distribution and patient survival status in the MMP1. (**E**) ROC curves to predict the sensitivity and specificity of 1-, 2-, 3-year survival according the risk score in the MMP1. (**F**) Clinical ROC analysis in the MMP1. (**G**) Kaplan-Meier curve of HCC patients in the ITGA2. (**H**, **I**) PCA and t-SNE analysis showing a remarkable difference in transcriptomes between the two risk categories in the ITGA2. (**J**) Risk Score Triad Diagrams showing the risk score distribution and patient survival status in the ITGA2. (**K**) ROC curves to predict the sensitivity and specificity of 1-, 2-, 3-year survival according the risk score in the ITGA2. (**L**) Clinical ROC analysis in the ITGA2. (**M**) Kaplan-Meier curve of HCC patients in the P3H1. (**N**, **O**) PCA and t-SNE analysis showing a remarkable difference in transcriptomes between the two risk categories in the P3H1. (**P**) Risk Score Triad Diagrams showing the risk score distribution and patient survival status in the P3H1. (**Q**) ROC curves to predict the sensitivity and specificity of 1-, 2-, 3-year survival according the risk score in the P3H1. (**R**) Clinical ROC analysis in the P3H1. (**S**) Kaplan-Meier curve of HCC patients in the CTSA. (**T**, **U**) PCA and t-SNE analysis showing a remarkable difference in transcriptomes between the two risk categories in the CTSA. (**V**) Risk Score Triad Diagrams showing the risk score distribution and patient survival status in the CTSA. (**W**) ROC curves to predict the sensitivity and specificity of 1-, 2-, 3-year survival according the risk score in the CTSA. (**X**) Clinical ROC analysis in the CTSA.

Supplementary Figure 2. The ICGC database was utilized to individually validate the prognostic value of MMP1, ITGA2, P3H1, and CTSA in hepatocellular carcinoma. (**A**) Kaplan-Meier curve of HCC patients in the MMP1. (**B**, **C**) PCA and t-SNE analysis showing a remarkable difference in transcriptomes between the two risk categories in the MMP1. (**D**) Risk Score Triad Diagrams showing the risk score distribution and patient survival status in the MMP1. (**E**) ROC curves to predict the sensitivity and specificity of 1-, 2-, 3-year survival according the risk score in the MMP1. (**F**) Clinical ROC analysis in the MMP1. (**G**) Kaplan-Meier curve of HCC patients in the ITGA2. (**H**, **I**) PCA and t-SNE analysis showing a remarkable difference in transcriptomes between the two risk categories in the ITGA2. (**J**) Risk Score Triad Diagrams showing the risk score distribution and patient survival status in the ITGA2. (**K**) ROC curves to predict the sensitivity and specificity of 1-, 2-, 3-year survival according the risk score in the ITGA2. (**L**) Clinical ROC analysis in the ITGA2. (**M**) Kaplan-Meier curve of HCC patients in the P3H1. (**N**, **O**) PCA and t-SNE analysis showing a remarkable difference in transcriptomes between the two risk categories in the P3H1. (**P**) Risk Score Triad Diagrams showing the risk score distribution and patient survival status in the P3H1. (**Q**) ROC curves to predict the sensitivity and specificity of 1-, 2-, 3-year survival according the risk score in the P3H1. (**R**) Clinical ROC analysis in the P3H1. (**S**) Kaplan-Meier curve of HCC patients in the CTSA. (**T**, **U**) PCA and t-SNE analysis showing a remarkable difference in transcriptomes between the two risk categories in the CTSA. (**V**) Risk Score Triad Diagrams showing the risk score distribution and patient survival status in the CTSA. (**W**) ROC curves to predict the sensitivity and specificity of 1-, 2-, 3-year survival according the risk score in the CTSA. (**X**) Clinical ROC analysis in the CTSA.

Supplementary Figure 3. The GEO database was utilized to individually validate the prognostic value of MMP1, ITGA2, P3H1, and CTSA in hepatocellular carcinoma. (**A**) Kaplan-Meier curve of HCC patients in the MMP1. (**B**, **C**) PCA and t-SNE analysis showing a remarkable difference in transcriptomes between the two risk categories in the MMP1. (**D**) Risk Score Triad Diagrams showing the risk score distribution and patient survival status in the MMP1. (**E**) ROC curves to predict the sensitivity and specificity of 1-, 2-, 3-year survival according the risk score in the MMP1. (**F**) Clinical ROC analysis in the MMP1. (**G**) Kaplan-Meier curve of HCC patients in the ITGA2. (**H**, **I**) PCA and t-SNE analysis showing a remarkable difference in transcriptomes between the two risk categories in the ITGA2. (**J**) Risk Score Triad Diagrams showing the risk score distribution and patient survival status in the ITGA2. (**K**) ROC curves to predict the sensitivity and specificity of 1-, 2-, 3-year survival according the risk score in the ITGA2. (**L**) Clinical ROC analysis in the ITGA2. (**M**) Kaplan-Meier curve of HCC patients in the P3H1. (**N**, **O**) PCA and t-SNE analysis showing a remarkable difference in transcriptomes between the two risk categories in the P3H1. (**P**) Risk Score Triad Diagrams showing the risk score distribution and patient survival status in the P3H1. (**Q**) ROC curves to predict the sensitivity and specificity of 1-, 2-, 3-year survival according the risk score in the P3H1. (**R**) Clinical ROC analysis in the P3H1. (**S**) Kaplan-Meier curve of HCC patients in the CTSA. (**T**, **U**) PCA and t-SNE analysis showing a remarkable difference in transcriptomes between the two risk categories in the CTSA. (**V**) Risk Score Triad Diagrams showing the risk score distribution and patient survival status in the CTSA. (**W**) ROC curves to predict the sensitivity and specificity of 1-, 2-, 3-year survival according the risk score in the CTSA. (**X**) Clinical ROC analysis in the CTSA.

Supplementary Figure 4. The GEO-GSE54236 dataset was used to validate the risk analysis model. (**A**–**F**) Using the GEO-GSE54236 dataset, we demonstrated through Kaplan-Meier curve, PCA and t-SNE analysis, Risk Score Triad Diagrams, ROC curves and Clinical ROC analysis that our risk model has good predictive value for the prognosis of hepatocellular carcinoma. (**G**–**K**) We further illustrated the predictive value of the model using tumor doubling time as the endpoint event.

Supplementary Table

Supplementary Table 1. Cox regression analysis of all 66 common genes.

Gene	HR	z	P-value	Lower	Upper
CTSA	1.208736081	1.953452323	0.050766021	0.999368267	1.461966486
MEP1A	1.141371898	2.39743743	0.016510201	1.024422359	1.271672565
CSPG4	1.012235234	0.102779146	0.918138257	0.802723461	1.276429828
ECM1	1.14633304	2.098923067	0.035823683	1.009082487	1.302251754
NPNT	1.007033479	0.100735389	0.919760518	0.878658105	1.154164994
SMOC1	0.798636763	-2.523830863	0.011608377	0.670680112	0.951005803
SERPINF1	0.911291972	-1.658511915	0.097214185	0.81654932	1.017027433
LAMC1	1.055998131	0.793713324	0.427362342	0.923060811	1.20808081
MPZL2	1.183457869	1.813153702	0.069808118	0.986454024	1.419805174
SEMA3B	0.95708511	-0.6029272	0.546557121	0.829899768	1.103762098
LAMA4	1.062529559	0.580777097	0.561390693	0.865859198	1.303871422
LAMB2	0.914523064	-0.89420185	0.371213903	0.751863239	1.11237309
ADAMTS9	1.15147831	1.203592992	0.228746929	0.9151772	1.448792975
MATN2	0.895918432	-1.444349248	0.148640759	0.771786861	1.040014903
COL18A1	0.824274663	-2.053736731	0.040001179	0.685450377	0.991215035
ADAMTS10	0.914344947	-0.853764883	0.393235276	0.744443173	1.123022836
GPC5	0.905059229	-0.773022665	0.439508967	0.702802825	1.165522076
NTN1	0.879202179	-1.181412318	0.237438967	0.710120348	1.088542913
CTSD	1.105473087	1.028377246	0.303772407	0.913167749	1.338276288
NTN4	1.033783681	0.374077212	0.708346872	0.868612205	1.230363438
LOXL2	1.040852407	0.406449874	0.68441208	0.858096209	1.262531779
COL9A3	1.027544405	0.273398916	0.784546586	0.84567506	1.248526243
LAMB3	1.030025767	0.465960975	0.641243434	0.909505611	1.166516257
LAD1	1.036862133	0.807651529	0.419291214	0.949664335	1.132066397
COL4A5	1.076875164	0.847288172	0.396834518	0.907318578	1.278117903
LOXL4	1.061932196	1.250305779	0.211187868	0.966468678	1.166825179
VTN	0.907834236	-1.998720916	0.045638559	0.825709732	0.998126786
ITGA7	0.836451544	-1.476105543	0.139915561	0.659869621	1.060287007
SPOCK1	1.073046995	0.665152555	0.505952923	0.871758887	1.320812291
DAG1	1.017861896	0.152430931	0.878847063	0.810634681	1.278063799
ITGAV	1.209718378	2.625821436	0.008644015	1.049462999	1.394445115
SDC1	1.025940946	0.256446377	0.797606181	0.843561124	1.247751698
ADAMTS16	1.10684722	0.971784005	0.331158015	0.901920945	1.35833498
FBLN1	1.072513695	1.211194296	0.225820948	0.957645396	1.201160293
SDC4	1.022396952	0.268471839	0.788336149	0.869746271	1.20183962
ANG	0.891216004	-2.290837663	0.021972804	0.807588427	0.983503403
TINAG	1.122051271	2.110226018	0.034838892	1.008233759	1.248717417
MEGF9	1.060015695	0.542360358	0.587570283	0.858696231	1.308534069
SPON2	1.074107391	1.027796918	0.304045366	0.937218997	1.230989438
ADAM9	1.305910556	3.447748995	0.000565279	1.122068525	1.519873647
EVA1A	0.918881431	-1.289501607	0.197223758	0.808008841	1.04496763
ADAM10	1.018417794	0.135188957	0.892462478	0.781655637	1.326894802
ADAMTS17	0.951728773	-0.505998903	0.61285742	0.785751184	1.15276652
P3H1	1.511152022	3.397823996	0.000679241	1.190907368	1.917513063
ADAMTS13	1.209832757	1.87216184	0.061184217	0.991106401	1.476829631

MMP17	1.067408965	0.542758362	0.587296185	0.843383003	1.350942448
ITGA5	1.205334624	2.114436612	0.034477993	1.013737267	1.433144073
ITGB5	1.263085673	2.205451578	0.027422422	1.026338049	1.554444385
VWA1	1.054276615	0.609218937	0.54237933	0.889417543	1.249693341
CD151	1.152078406	1.842717162	0.065370317	0.991032904	1.339294234
TGFBI	0.939343121	-0.787057822	0.431248019	0.803803891	1.097737283
BCAN	1.063229498	0.491762915	0.622886956	0.832727138	1.357535879
ROBO3	1.347567377	1.728408144	0.083915076	0.96082445	1.889978795
COL2A1	1.030761316	0.543729797	0.586627425	0.924118199	1.149711034
ITGA1	0.888351324	-0.785825061	0.431969983	0.661222345	1.193498798
GPC3	0.992137493	-0.332147775	0.739777676	0.946984416	1.039443509
ITGA2	1.168161502	1.547026388	0.121856883	0.959360641	1.422407003
CTSB	1.377725522	3.074126954	0.002111196	1.123148191	1.690006386
ACHE	0.939268943	-0.781950065	0.43424392	0.802763817	1.098985938
ADAM17	1.398506601	2.582043115	0.009821731	1.084162164	1.803992777
USH2A	0.721946612	-1.907458477	0.056461243	0.51655351	1.009008556
MMP1	1.449655656	4.648097194	3.35E-06	1.23955064	1.695373673
ROBO1	1.091411473	1.475357006	0.140116588	0.97167718	1.225899947
COL9A2	1.097038218	1.039747457	0.298457246	0.921302386	1.306295165
PHF13	1.378771189	2.223279867	0.026196934	1.038773513	1.83005243
TENM1	1.030695222	0.214083474	0.830481963	0.781485423	1.359376142