Development and validation of an ECM-related prognostic signature to predict the immune landscape of human hepatocellular carcinoma

Guozhi Wu[1,2,3†], Yuan Yang[1,2,3†], Rong Ye[4†], Hanxun Yue[1,2,3], Huiyun Zhang[1,2,3], Taobi Huang[1,2,3], Min Liu[2,3], Ya Zheng[2,3], Yuping Wang[2,3], Yongning Zhou[2,3*] and Qinghong Guo[2,3*]

Abstract

Background: The global burden of hepatocellular carcinoma (HCC) is increasing, negatively impacting social health and economies. The discovery of novel and valuable biomarkers for the early diagnosis and therapeutic guidance of HCC is urgently needed.

Methods: Extracellular matrix (ECM)-related gene sets, transcriptome data and mutation profiles were downloaded from the Matrisome Project and The Cancer Genome Atlas (TCGA)-LIHC datasets. Coexpression analysis was initially performed with the aim of identifying ECM-related lncRNAs ($r > 0.4$, $p < 0.001$). The screened lncRNAs were subjected to univariate analysis to obtain a series of prognosis-related lncRNA sets, which were incorporated into least absolute selection and shrinkage operator (LASSO) regression for signature establishment. Following the grouping of LIHC samples according to risk score, the correlations between the signature and clinicopathological, tumour immune infil-tration, and mutational characteristics as well as therapeutic response were also analysed. lncRNA expression levels used for modelling were finally examined at the cellular and tissue levels by real-time PCR. All analyses were based on R software.

Results: AL031985.3 and MKLN1-AS were ultimately identified as signature-related lncRNAs, and both were signifi-cantly upregulated in HCC tissue samples and cell lines. The prognostic value of the signature reflected by the AUC value was superior to that of age, sex, grade and stage. Correlation analysis results demonstrated that high-risk groups exhibited significant enrichment of immune cells (DCs, macrophages and Tregs) and increased expression levels of all immune checkpoint genes. Prominent differences in clinicopathological profiles, immune functions, tumour mutation burden (TMB) and drug sensitivity were noted between the two risk groups.

Conclusions: Our signature represents a valuable predictive tool in the prognostic management of HCC patients. Further validation of the mechanisms involved is needed.

Keywords: Hepatocellular carcinoma, Extracellular matrix, LncRNA, Immune infiltration, Tumour mutation burden

†Guozhi Wu, Yuan Yang and Rong Ye contributed equally to this work.

*Correspondence: yongningzhou@sina.com; guoza@sohu.com

2 Department of Gastroenterology, The First Hospital of Lanzhou University, No.1 West Donggang Road, Lanzhou 730000, Gansu, China
Full list of author information is available at the end of the article

Introduction

HCC is one of the subtypes of liver cancer, accounting for 90% of all cases [1, 2]. HCC has typically been associated with high mortality over the past decades due to the lack of effective approaches for early diagnosis and treatment.

Wu *et al. BMC Cancer* (2022) 22:1036

Page 2 of 18

Currently, HCC is the sixth most common cancer worldwide, and its incidence and mortality continue to increase [3]. Generally recognized risk factors attributed to the occurrence and progression of HCC include long-term alcohol consumption, nonalcoholic steatohepatitis, and hepatitis virus infection [4]. The prolonged presence of these factors makes individuals at high risk of developing cirrhosis, a state that predisposes them to progress to advanced liver cancer with a poor prognosis. Therefore, the earlier the biomarkers associated with cancer progression are identified, the more likely it is that early intervention in HCC will be possible.

The ECM is a complex and dynamic scaffold network of secreted macromolecular substances. In addition to supporting and connecting tissue structures, the ECM regulates many crucial cellular physiological activities, including growth, migration, differentiation and apoptosis [5]. Evidence supports the overview that alterations in the structure and composition of the ECM allow cancer cells to achieve numerous hallmarks of cancer [6]. Additionally, by interfering with the drug transport and delivery process, the ECM may be involved in drug resistance mechanisms to some degree [7]. Based on improvements in sequencing technology, an increasing number of lncRNAs that may affect gene expression at the transcriptional and posttranscriptional levels have been discovered. The pivotal roles that lncRNAs may play in maintaining the cellular microenvironment and in the interaction between malignant cells and other cells are also widely understood [8]. This interaction involves alterations in fibril production to regulate fibroblast viability and apoptosis [9–11]. Given their critical role in the ECM, lncRNAs have been identified as key targets for HCC research. Therefore, in this study, we developed a prediction model based on ECM-related lncRNAs to analyse sample information downloaded from TCGA database with the aim of predicting the probability of survival in HCC patients.

Materials and methods
Data download, preliminary processing, and screening of ECM-related LncRNAs (ECMrlncRNAs)
We downloaded RNA sequencing (RNA-seq) and clinical data of liver cancer from The Cancer Genome Atlas (TCGA)-LIHC dataset (https://gdc.cancer.gov/). Then, a systematic collation was performed for subsequent analysis. LncRNA sets were extracted in bulk from the transcriptome profiles using GTF files and then assembled into an expression matrix. GTF files were obtained from the Ensembl website (http://asia.ensembl.org) for annotation to distinguish lncRNAs from the RNA-seq data. A list of ECM-related genes was retrieved from the Matrisome Project (www.matrisomeproject.mit. edu/other-resources/human-matrisome). Coexpression analysis was subsequently applied to screen ECM-related lncRNAs (ECMrlncRNAs). Those with correlation coefficients greater than 0.4 and *p* values less than 0.001 were ultimately identified as ECMrlncRNAs. Corresponding calculations were primarily based on the "limma" package in R.

Construction of a prognostic signature using the LASSO algorithm
Hazard ratios (HR) of ECMrlncRNAs were calculated using univariate Cox hazard regression analysis. ECMrlncRNAs were considered eligible and incorporated into the LASSO algorithm for modelling if they were determined to have prognostic value based on HR values. LASSO Cox regression involves a penalized linear model with a shrinkage penalty that induces sparsity of predictors in the model. After completing the cross-validation, the LASSO results were optimized by multivariate regression analysis and then integrated into the model construction. Notably, the results of the cross-validation represent the number of lncRNAs that are deemed suitable for modelling by LASSO regression. The following formula was used for modelling: $\sum_{i=1}^{n} \left(\text{expression level of lncRNA} * \text{regression coefficient} \right)$ The R packages "survival" and "glmnet" were utilized for all calculations. The 1-, 3-, and 5-year ROC curves were subsequently plotted to show the prognostic performance of the signatures using the "survival", "survminer" and "timeROC" packages. When the model was determined to exhibit favourable prognostic value, total samples included in the analysis were divided into training and testing groups for the subsequent analyses. Each group was divided into a high-risk set and a low-risk set based on the median risk score. To further illustrate the distribution of the two risk groups, principal component analysis (PCA) was applied for visualization. In addition, Kaplan–Meier analysis was also performed with the "survival" package to reflect the difference in overall survival (OS). The relationship between the signature and clinicopathological characteristics was assessed primarily based on chi-square tests. A combined analysis of univariate regression and multivariate regression was conducted to determine the independent risk factors.

Functional enrichment exploration
To assess the contribution of differentially expressed genes (DEGs) to a phenotype, we adopted the "clusterProfiler" package for GO term and KEGG pathway analyses [12–14]. The thresholds were set as a *P* value <0.05 and FDR <0.05. The results are presented as bar plots.

Wu *et al. BMC Cancer*   (2022) 22:1036

Page 3 of 18

Immune correlation analysis

Using the median risk score as a cut-off, we divided the HCC patients into two subgroups: high- and low-risk groups. The relationships between the established signature and tumour immune features were analysed by ssGSEA and TIMER. The use of these algorithms in this procedure helped to reveal differences in the distribution of tumour-infiltrating immune cells and immune-related functions between the two risk groups. Moreover, we performed a correlation analysis of the risk score with cytokines and chemokines based on the gene expression level. These included cytokines and chemokines were obtained from Ozga et al. and Montanari et al. [15, 16]. A *P* value< 0.05 was considered a statistically significant threshold.

The relationship between the signature and TMB

TMB is defined as the total number of substitutions and insertions/deletions per megabase (Mut/Mb) in the exon coding region of the evaluated gene in one tumour sample. TMB is implicated in overall survival after immunotherapy in multiple cancer types, suggesting the importance of TMB as a predictive biomarker for the efficacy of immune checkpoint inhibitors (ICIs) [17]. To examine the predictive value of TMB in our model, we obtained masked somatic mutation (MSM) data from the TCGA-SNP category. After processing the data analytically, the R package "maftools" was used to analyse and visualize the mutation frequencies and types of genes between both risk groups.

The association between the signature and immune checkpoint genes

Linking the signature to the expression level of various immune checkpoint genes contributed to the identification of appropriate therapeutic options for HCC populations at different levels of risk. The Wilcoxon signed-rank test was applied in the comparison of gene expression levels in different risk groups using the R package "ggstatsplot", and the corresponding results were presented as box plots.

Prediction of therapeutic response

Comparison of drug sensitivity to targeted drugs and chemotherapeutic agents in the high- and low-risk groups was accomplished by calculating IC_{50} values using the R package "pRRophetic". The drugs included in the comparison are routinely applied in the treatment of HCC, such as sorafenib, sunitinib, erlotinib, bleomycin, cisplatin, doxorubicin, gemcitabine, mitomycin C, docetaxel, paclitaxel, temsirolimus and vinblastine. To establish a link between our signature and the immunotherapeutic response, we used the Tumour Immune Dysfunction and Exclusion Algorithm (TIDE) to model and integrate two key mechanisms of tumour immune evasion to provide predictive results for immunotherapy.
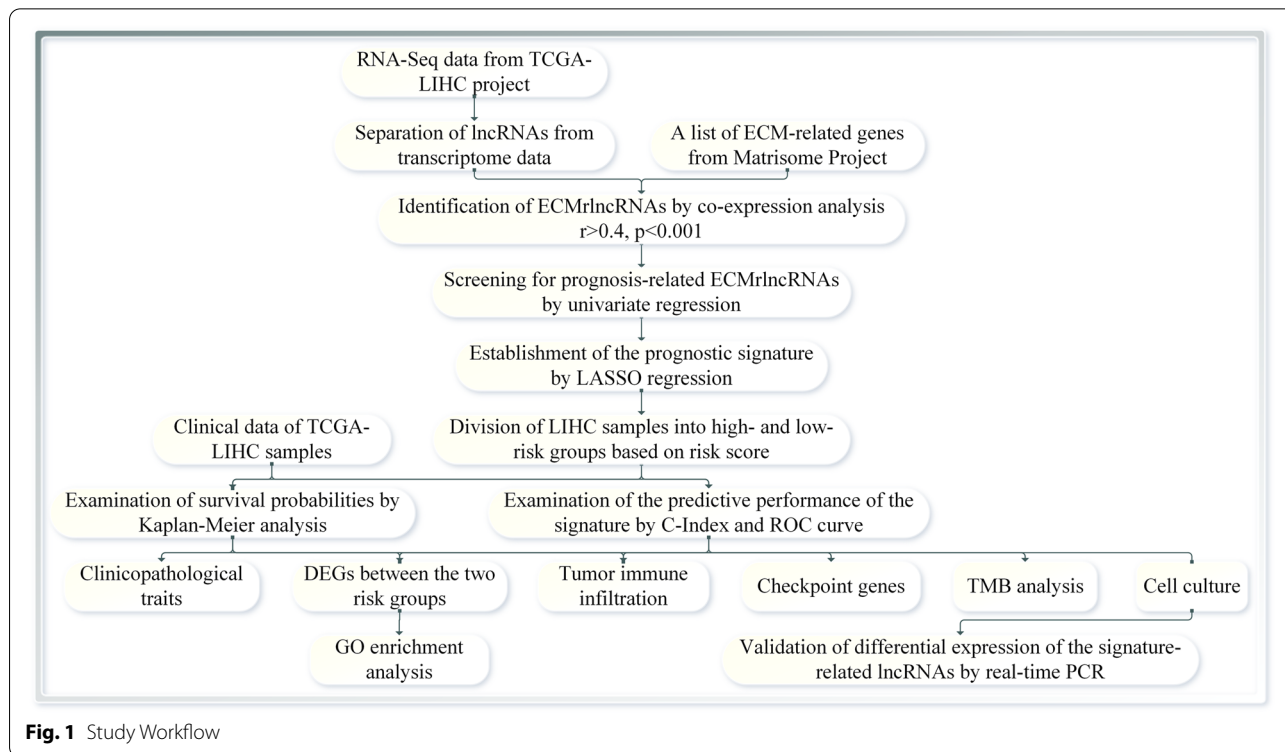

Fig. 1 Study Workflow

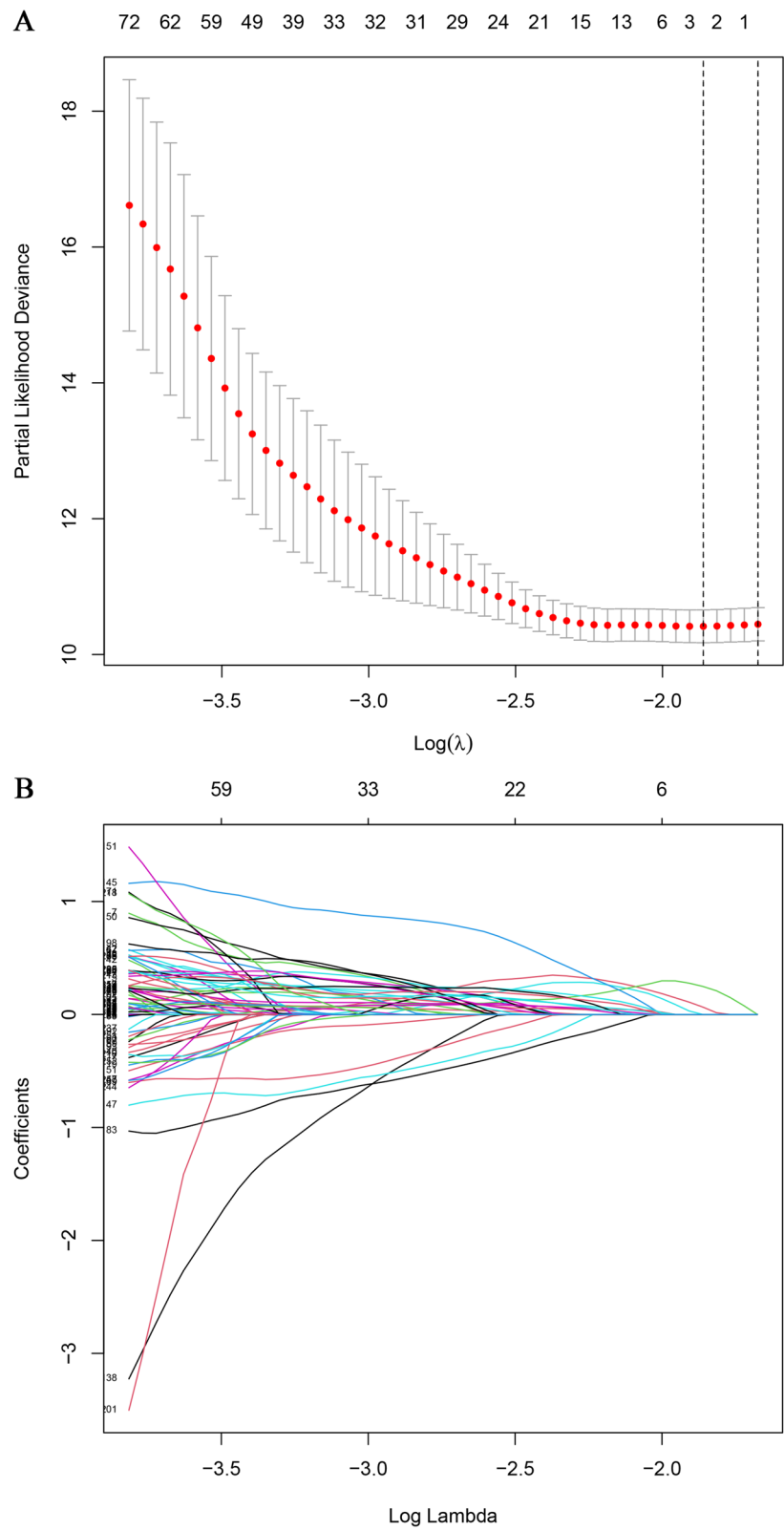Wu *et al. BMC Cancer*   (2022) 22:1036

Page 4 of 18

Fig. 2 Cross-validation and LASSO Regression. When the curve reaches the lowest point, the error of cross-validation is minimized. At this point, the corresponding figure represents the number of significant lncRNAs (**A**). LASSO regression plot (**B**)

Wu *et al. BMC Cancer*   (2022) 22:1036

Page 5 of 18

Table 1 Demographic and baseline disease characteristics of samples in the TCGA-LIHC project

Covariates	Total	Training set	Testing set	P value*
Age - no. (%)				1
<=65	227(62.19%)	114(61.96%)	113(62.43%)	
>65	138(37.81%)	70(38.04%)	68(37.57%)	
Gender - no. (%)				0.9091
Female	119(32.6%)	61(33.15%)	58(32.04%)	
Male	246(67.4%)	123(66.85%)	123(67.96%)	
Grade - no. (%)				0.8485
G1	55(15.07%)	27(14.67%)	28(15.47%)	
G2	175(47.95%)	92(50%)	83(45.86%)	
G3	118(32.33%)	58(31.52%)	60(33.15%)	
G4	12(3.29%)	5(2.72%)	7(3.87%)	
Unknow	5(1.37%)	2(1.09%)	3(1.66%)	
Stage - no. (%)				0.163
Stage I	170(46.58%)	79(42.93%)	91(50.28%)	
Stage II	84(23.01%)	41(22.28%)	43(23.76%)	
Stage III	83(22.74%)	44(23.91%)	39(21.55%)	
Stage IV	4(1.1%)	4(2.17%)	0(0%)	
Unknow	24(6.58%)	16(8.7%)	8(4.42%)	

*P value less than 0.05 is considered to be statistically significant

Elevated TIDE predicts that patients have suppressive cells that inhibit T-cell infiltration and fail to respond to immunotherapy [18].

Cell culture and real-time PCR

The normal cell line (L02) and HCC cell lines (SK-Hep-1, Hep-G2, Huh-7 and LM-3) were purchased from the Key Laboratory of Biotherapy and Regenerative Medicine, Gansu Province (Gansu, China). All cell lines were cultured in Dulbecco's modified Eagle's medium (DMEM, Solarbio, Beijing, China) with 10% foetal bovine serum (FBS, Solarbio) and cyan chain double antibodies (Beyotime, Jiangsu, China) in a 37°C humidified atmosphere that contained 5% CO_2. When the HCC cell lines achieved the indicated conditions, total RNA extraction was performed with TransZol (TransGen, Beijing, China) following the manufacturer's instructions. The isolation and purification of cytoplasmic and nuclear RNA was principally accomplished using an EasyPure RNA kit (TransGen). The concentration of purified RNA was then determined using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, MA, USA). To achieve cDNA synthesis, we employed a TransScript One-Step gDNA Removal and cDNA Synthesis SuperMix (TransGen) to reverse transcribe RNA that was previously extracted. HCC and adjacent tissue cDNA microarrays were obtained from Shanghai Outdo Biotech Company,

China. Finally, real-time PCR was performed in Trans-Start Top Green qPCR SuperMix (TransGen). Quantification of the RNA expression level was realized using the 2-ΔCt method (ΔCt = Ct (Target) - Ct (Reference)). The primer sequences are as follows: AL031985.3 forward primer AAATCCCATACCCCTTTCACC reverse primer TTTACTGAGTCCCTTCTGCGTG; MKLN1-AS forward primer GTGTTTCTCTCTGAAAGCAGCG, reverse primer TTCAAAAGTGACCAAAGCCAGG; GAPDH forward primer GTCAAGGCTGAGAACGGGAA, reverse primer AAATGAGCCCCAGCCTTCTC; β-actin (validated and provided by Outdo Biotech) forward primer GAAGAGCTACGAGCTGCCTGA, reverse primer CAGACAGCACTGTGTTGGCG. The ethics statement for this study was provided by Shanghai Otto Biotech.

Statistical analysis

All analyses were implemented based on R software (version 3.6.3). A *P* value < 0.05 was considered statistically significant.

Results

ECMrlncRNA screening and prognostic signature construction

The study scheme is presented in Fig. 1. We initially obtained a total of 1068 ECM-related gene sets from the Matrisome Project. Following further coexpression analysis between the ECM-related genes and lncRNA sets, 1961 lncRNAs were eventually identified as ECMrlncRNAs. Through subsequent univariate Cox regression, lncRNAs with prognostic value were formally included in the LASSO regression analysis (cross-validation and regression plots are shown in Fig. 2A and B, respectively). Two ECMrlncRNAs (MKLN1-AS and AL031985.3) that demonstrated distinct expression differences between tumour and normal tissues were included in the risk model. The model formula is described as follows: *risk score = 0.942362440899263* mRNA expression level of MKLN1-AS+ 0.756996665090891* mRNA expression level of AL031985.3.* LIHC samples were divided into training and testing groups in our analysis. However, comparative analysis concerning the baseline characteristics of the three groups revealed no significant differences (Table 1). Figure 3A represents the expression levels of two lncRNAs in the total sample group. Dimensionality degradation via PCA allowed for a clearer view of the distribution of patients from the total LIHC sample with different risk scores (Fig. 3B) and the corresponding differences in patient survival status (alive or dead) (Fig. 3C). Immediately afterwards, we performed Kaplan–Meier analysis to assess the survival differences, the results of
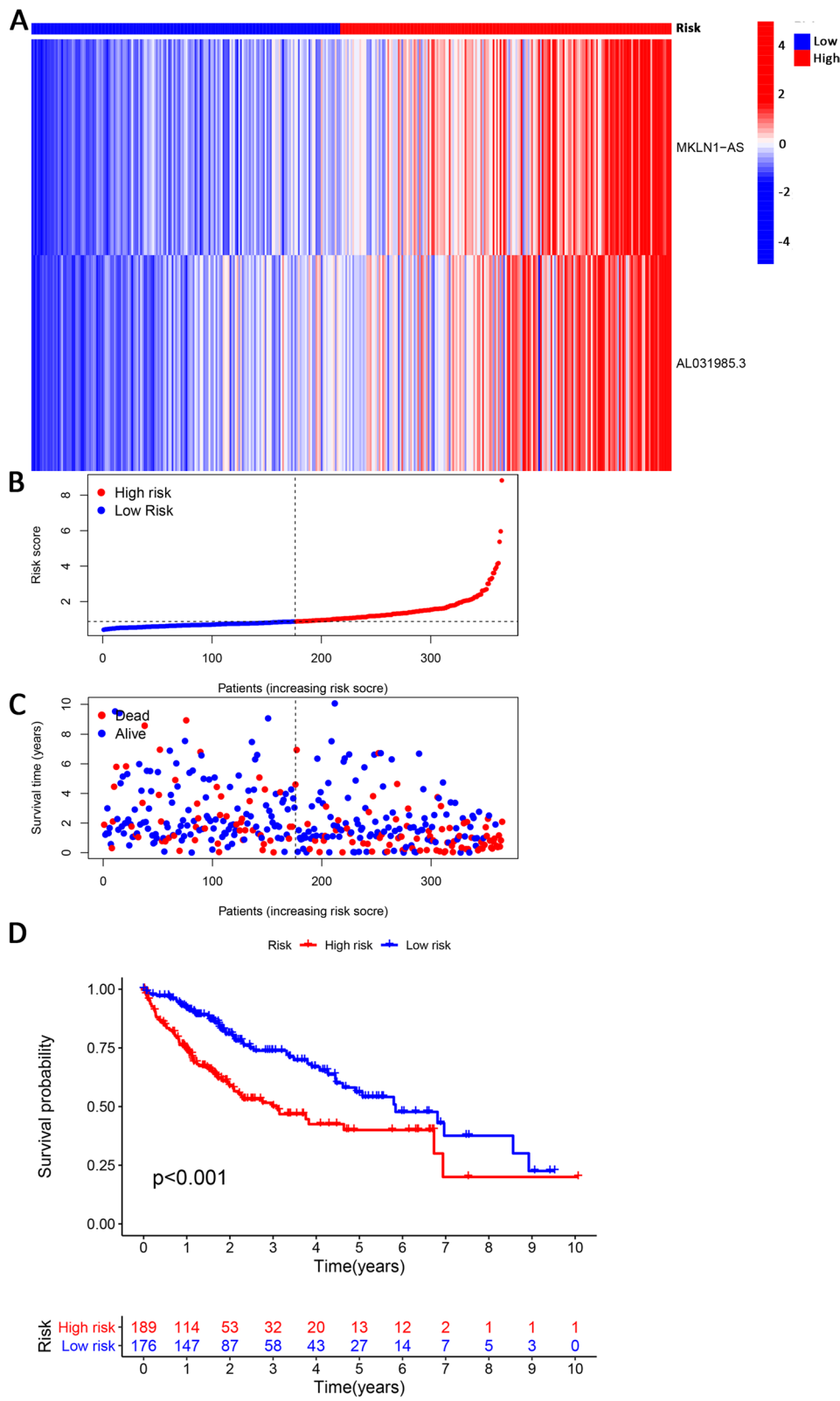
Fig. 3 Identification of Prognosis-Related ECMrlncRNAs and Signature Development. Heatmap visualizing the expression levels of MKLN1-AS and AL031985.3 in the total LIHC cohort (**A**). PCA displays the distribution of different-risk patients and differences in survival status (alive or dead) (**B** and **C**). Survival curves revealed the prognostic differences between the high- and low-risk HCC groups (**D**)

Wu *et al. BMC Cancer*   (2022) 22:1036

Page 7 of 18

(See figure on next page.)
Fig. 4 Examination of the Predictive Properties of the Signature. Univariate (**A**) and multivariate (**B**) regression analyses indicated that the stage and risk score were independent risk factors. The predictive value of the signature was illustrated by the concordance index (**C**) and ROC curves. ROC curves comparing the signature with age, sex, grade and stage showed the superiority of the risk score compared with other indicators (**D**). ROC curves at 1, 3 and 5 years were also plotted to reflect the long-term predictive value (**E**)

which are presented as survival curves (Fig. 3D). The findings indicated that patients with higher risk scores have lower OS (*P* < 0.001). Similar lncRNA expression (Fig. S1A and E), patient distribution (Fig. S1B and F), patient survival status (Fig. S1C and G) and OS (Fig. S1D and H) results were observed in the total LIHC group as well as the training and testing groups.

Assessment of the predictive value of the signature

The rationale for signature construction based on the risk score was derived mainly from the results of univariate and multivariate regression analyses, which suggested that the risk score could serve as an independent risk factor for liver cancer (Fig. 4A and B). To further illustrate the use of the risk score as an important reference for model construction, we compared it with commonly available clinical predictors. The risk score showed a comparatively superior diagnostic ability compared with age, sex, grade and stage (Fig. 4C and D). The AUC values representing the prognostic predictions at 1, 3 and 5 years were 0.746, 0.683 and 0.670, respectively (Fig. 4E).

Application of the signature to TCGA-LIHC datasets

The constructed model employed a risk score-based cut-off criterion, namely, the median risk score, to divide HCC samples into high- and low-risk groups. Analysis of the relationship between the signature and clinicopathological profiles illustrated significant differences between the two risk groups in terms of age, sex, grade and TNM stage (Fig. S2A-O). Subsequently, we employed the PCA approach to confirm the strength of the risk score in separating patients with different risks, demonstrating the superiority of the risk score in accurately distinguishing samples compared with other indicators (all genes, ECM-related genes and ECMrlncRNAs) (Fig. 5A-D).

Functional enrichment analysis

The biological functions of the 636 DEGs identified between the two risk groups were assessed. GO terms are classified into three subtypes: BP, CC and MF. BP primarily included antigen binding, cell adhesion molecule binding, immunoglobulin receptor binding, and extracellular matrix structural constituent. CC mainly included collagen-containing extracellular matrix, the external side of the plasma membrane, and immunoglobulin complexes.

MF comprised phagocytosis, humorall immune response, and protein activation cascade (Fig. 6A). Using KEGG analysis, 17 pathways were ultimately identified, including ECM-receptor interaction, focal adhesion, and PI3K-Akt signalling pathway (Fig. 6B and C).

Relationship between the signature and tumour immune-infiltrating features

Detailed differences in immune-related functions, immune cell subtypes and checkpoint genes were analysed between the two groups. The results suggested that MHC class I, APC costimulation, CCR, parainflammation and cytolytic activity were more significantly correlated with high risk scores (Fig. 7A and B). In contrast to NK cells, DCs, macrophages and Treg cells were prominently enriched in the high-risk group (Fig. 7C). The TIMER algorithm also yielded similar results (Fig. S3A). Furthermore, all of the checkpoint genes of interest showed considerable overexpression in the high-risk group (Fig. 7D). Correlation analysis of the risk score with cytokines and chemokines revealed that the risk score was positively correlated with CXCL1, CXCL3, CXCL5, CXCL8, CXCL9, CXCL11, CCL2, CCL4, CCL20, CCL26, CCL28, IL6, IL10, TGFB2, FGF2, VEGFA and HGF and negatively correlated with CCL14 and CCL15 (Fig. S3B).

Correlation of the signature and TMB

Given the crucial role that TMB plays in cancer progression, we report the mutation frequencies of the top 20 driver genes in the high- and low-risk groups, separately (Fig. 8A and B). Accordingly, the low-TMB group had a comparatively higher survival probability when grouped exclusively by TMB, whereas the low-TMB and low-risk group had a superior prognosis when grouped according to a criterion based on TMB in combination with the risk score (Fig. 8C and D).

Drug sensitivity and immune response prediction

Sensitivity analysis of several chemotherapeutic and targeted therapeutic agents, including bleomycin, cisplatin, doxorubicin, and mitomycin C, was performed. Patients in the high-risk group were more sensitive to
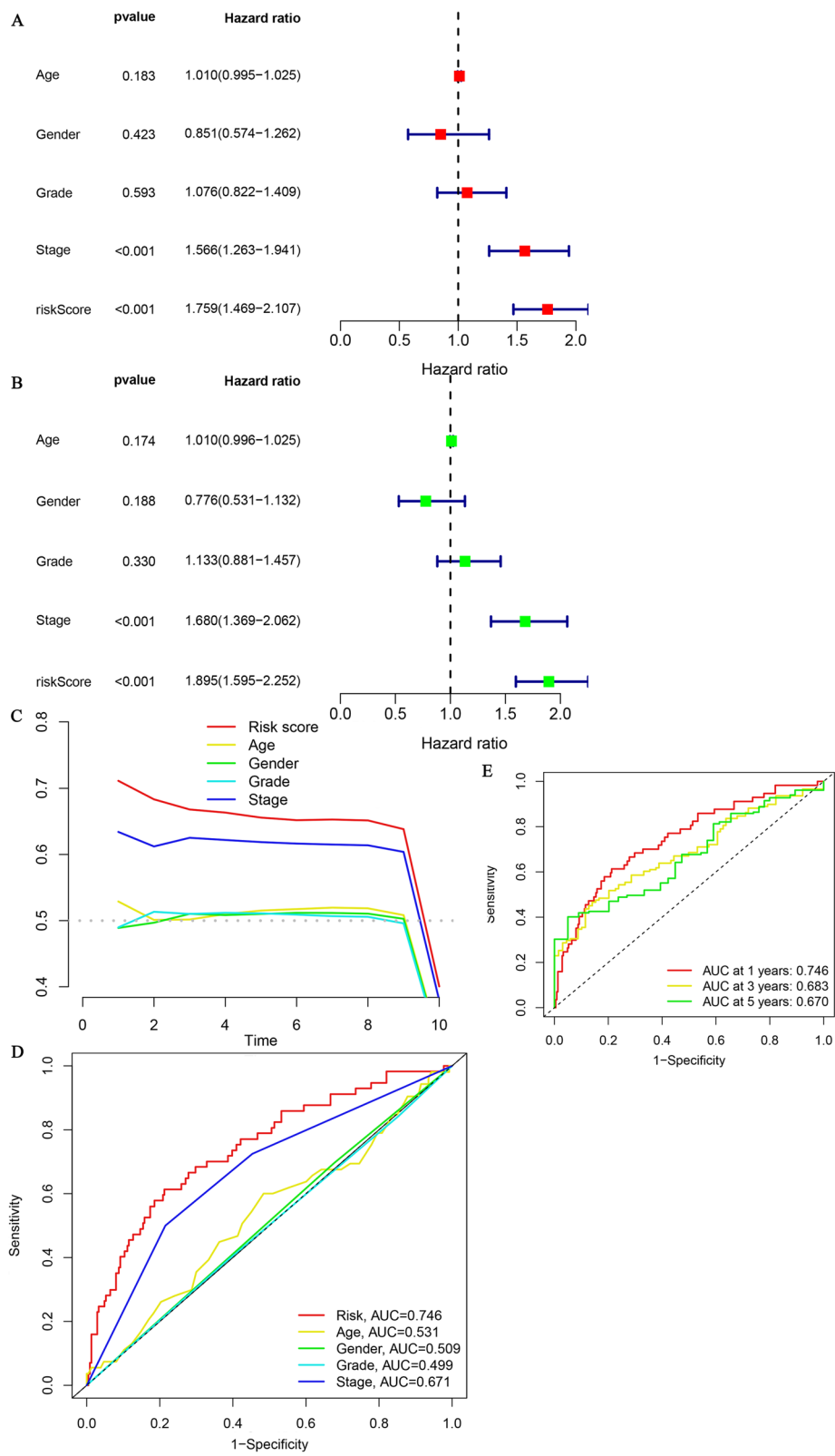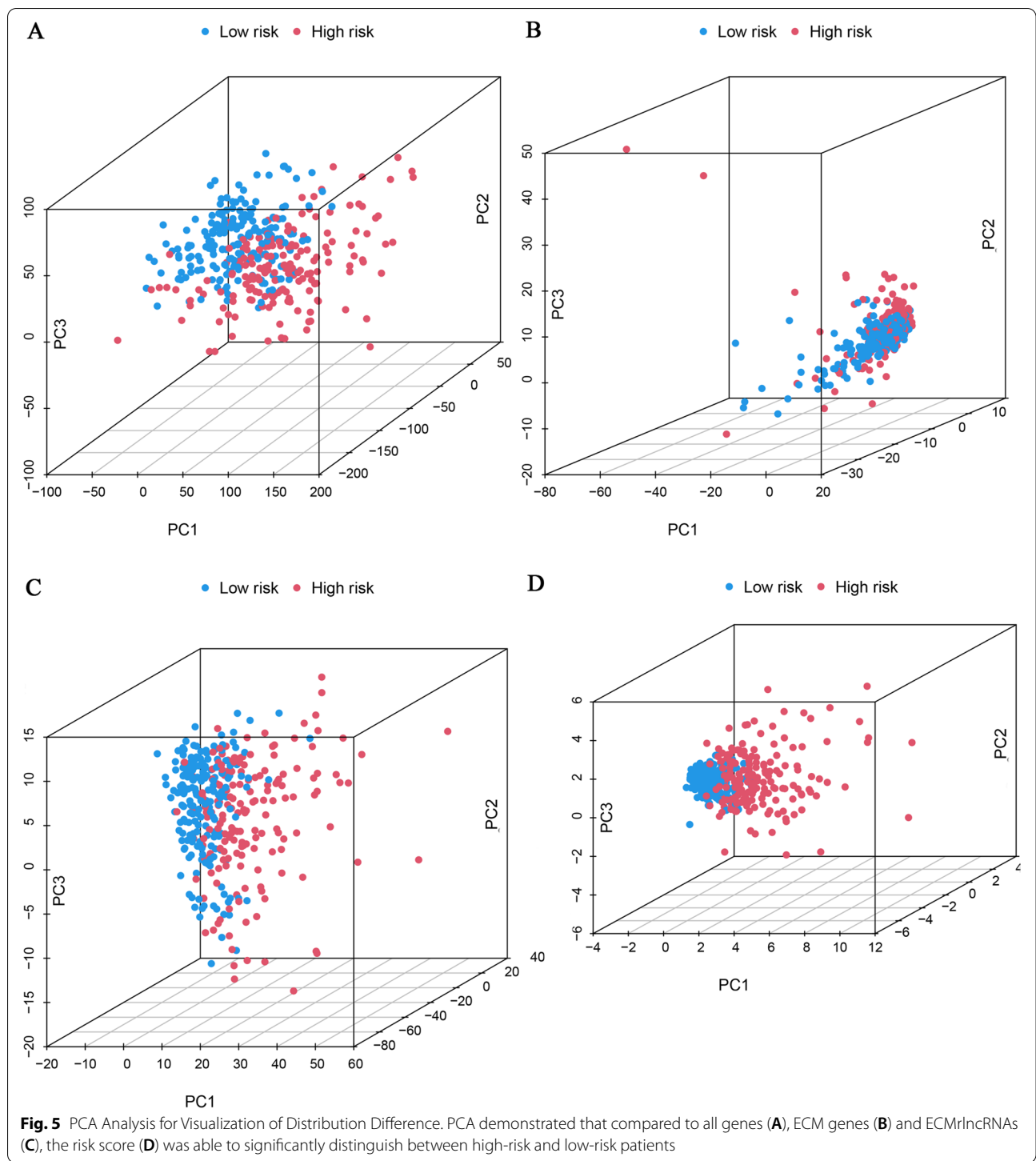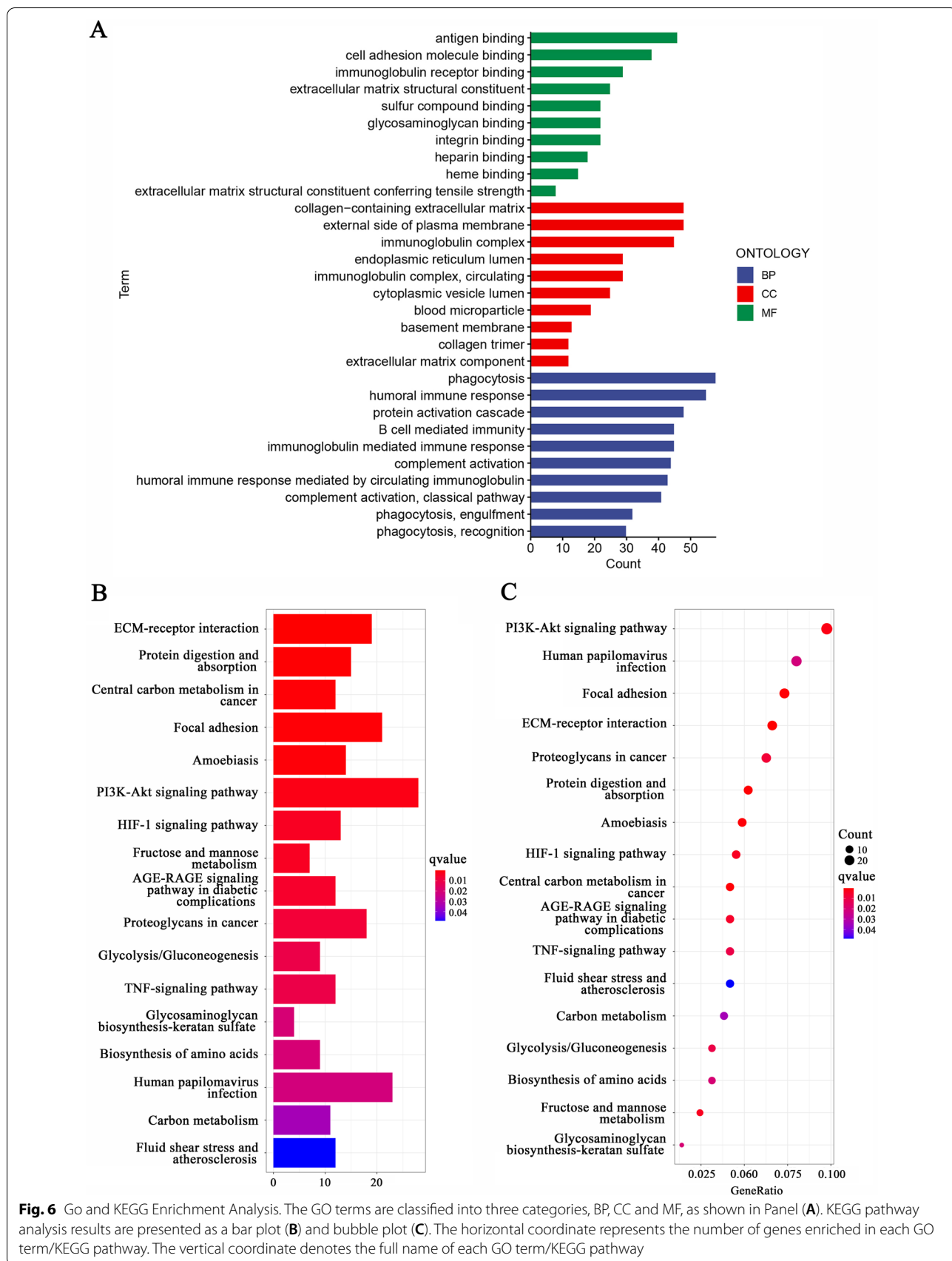
Wu *et al. BMC Cancer*   (2022) 22:1036

Page 8 of 18

Fig. 4 (See legend on previous page.)

Wu *et al. BMC Cancer* (2022) 22:1036

Page 9 of 18

Fig. 5 PCA Analysis for Visualization of Distribution Difference. PCA demonstrated that compared to all genes (**A**), ECM genes (**B**) and ECMrlncRNAs (**C**), the risk score (**D**) was able to significantly distinguish between high-risk and low-risk patients

paclitaxel and gemcitabine based on lower IC_{50} values, whereas patients in the low-risk group were sensitive to sorafenib, docetaxel, erlotinib, and temsirolimus based on higher IC_{50} values. No significant differences in sensitivity to sunitinib and vincristine were observed between the two groups (Fig. S4A-L). To

compare the difference in response to immunotherapy between the high-risk and low-risk groups, we then performed an analysis by employing the TIDE algorithm (Fig. 9A). The TIDE score is composed of two components: a dysfunction score and exclusion score. Here, the high-risk group exhibits a higher dysfunction

Fig. 6 Go and KEGG Enrichment Analysis. The GO terms are classified into three categories, BP, CC and MF, as shown in Panel (**A**). KEGG pathway analysis results are presented as a bar plot (**B**) and bubble plot (**C**). The horizontal coordinate represents the number of genes enriched in each GO term/KEGG pathway. The vertical coordinate denotes the full name of each GO term/KEGG pathway

Wu *et al. BMC Cancer* (2022) 22:1036

Page 11 of 18

score (Fig. 9B), and the low-risk group exhibits a higher exclusion score (Fig. 9C). Overall, the high-risk group had a higher TIDE score (Fig. 9A), implying that it may be associated with poorer immune checkpoint inhibition therapeutic efficacy. Next, we compared the predictive value of our model with the TIDE score, and the results are shown in Fig. 9D.

Exploration and verification of signature-related LncRNA expression profiles

Both overall and pairwise comparisons demonstrated significantly higher expression levels of AL031985.3 in the tumour group than in the normal group (Fig. 10A and C). Validation results at the cellular level also provided support for this finding (Fig. 11). In addition, based on our bioinformatics analysis, MKLN1-AS was also predicted to be overexpressed in tumour tissues (Fig. 10B and D). Verification results have been presented in a previous study. To confirm the aforementioned findings, 20 pairs of HCC samples (containing 20 tumour tissues and 32 adjacent tissues) were assembled and utilized for quantitative analysis of lncRNAs. Validation of these samples showed that AL031985.3 (Fig. 10E and G) and MKLN1-AS (Fig. 10F and H) were significantly upregulated in HCC in both pairwise and unpaired analyses.

Discussion

Given the pivotal role that ECM components play in tumour onset and progression, a comprehensive understanding of their biophysical and biochemical effects and remodelling processes seems particularly imperative to uncover promising biomarkers for diagnostic and therapeutic applications. Database mining using various algorithms and accurate prediction models in the context of big data are of great help in the discovery of lncRNAs that are relevant to ECM. Therefore, this study identified ECMrlncRNAs to provide a prognostic prediction approach for various risk populations to rationally guide clinical decision-making.

Despite the existence of several valuable signatures for the prognostic analysis of HCC patients, [19–28], our study differs from previous studies based on various features, such as the prognostic value of the signature, the ability to accurately differentiate samples and the correlation with immune landscape. The principal cause of this difference was based on our selection of a lncRNA set that was coexpressed with ECM-related genes. Two signatures consisting of ECM-related genes have been previously reported [24, 25]; however, these signatures were not equivalent to our lncRNA signature. Indeed, comprehensive studies were performed to evaluate the prognostic value of the models generated using these signatures, and these signatures also predicted the immune landscape as well as drug sensitivity. In contrast to the study by Tang et al., our study focused on assessing the differences in drug sensitivity to sorafenib in the risk subgroup obtained from LIHC samples. Given that sorafenib is currently the first-line drug for HCC treatment, it would be more clinically useful to construct a model that demonstrates a better association with this drug. More importantly, following a series of screening processes according to the developed criteria, two lncRNAs, AL031985.3 and MKLN1-AS, were eventually identified and incorporated into the signature. We subsequently validated these two lncRNAs at the cellular and tissue levels and further examined their prognostic value in HCC. Thus, our findings are more reliable and valuable. The upregulation of AL031985.3 and MKLN1-AS in HCC was identified through database mining but was also confirmed in a series of in vitro experiments. Few experimental studies have been performed on AL031985.3, and its function in HCC requires further exploration. Comparatively, several recent studies demonstrated phenotypic alterations mediated by MKLN1-AS. MKLN1-AS is potentially involved in the regulation of cell proliferation, angiogenesis, migration and invasion [26–28] and implicated in poor survival [28]. Mechanistically, MKLN1-AS may function as a competing endogenous RNA to induce pro-oncogenic effects during HCC progression [28]. Overall, the signature had favourable prognostic performance, as evidenced by the ROC and corresponding AUC values at 1, 3 and 5 years. Accordingly, we divided HCC populations into high- and low-risk groups based on the risk score cut-off value. Our signature was also applicable to HCC patients with specific clinicopathological characteristics. Patients with lower risk scores generally had a better prognosis, regardless of stratification by age, sex, grade or TNM stage. In addition, the DEGs identified between the two risk groups were then entered into the GO analysis, and these genes had robust associations with ECM and immune functions.

(See figure on next page.)

Fig. 7 Differential Analysis of Immune function, Tumour Infiltrating Immune Cells and Immune Checkpoint Genes. The distinctions in immune-related functions between the high- and low-risk groups are presented as a heatmap (**A**) and box plots (**B**). The differences in tumour-infiltrating immune cells between the two groups are reported as box plots using the ssGSEA algorithm (**C**). Differences in the expression of immune checkpoint genes in the two groups are presented as box plots (**D**). "*", "**" and "***" represent $P < 0.05$, $P < 0.01$ and $P < 0.001$, respectively
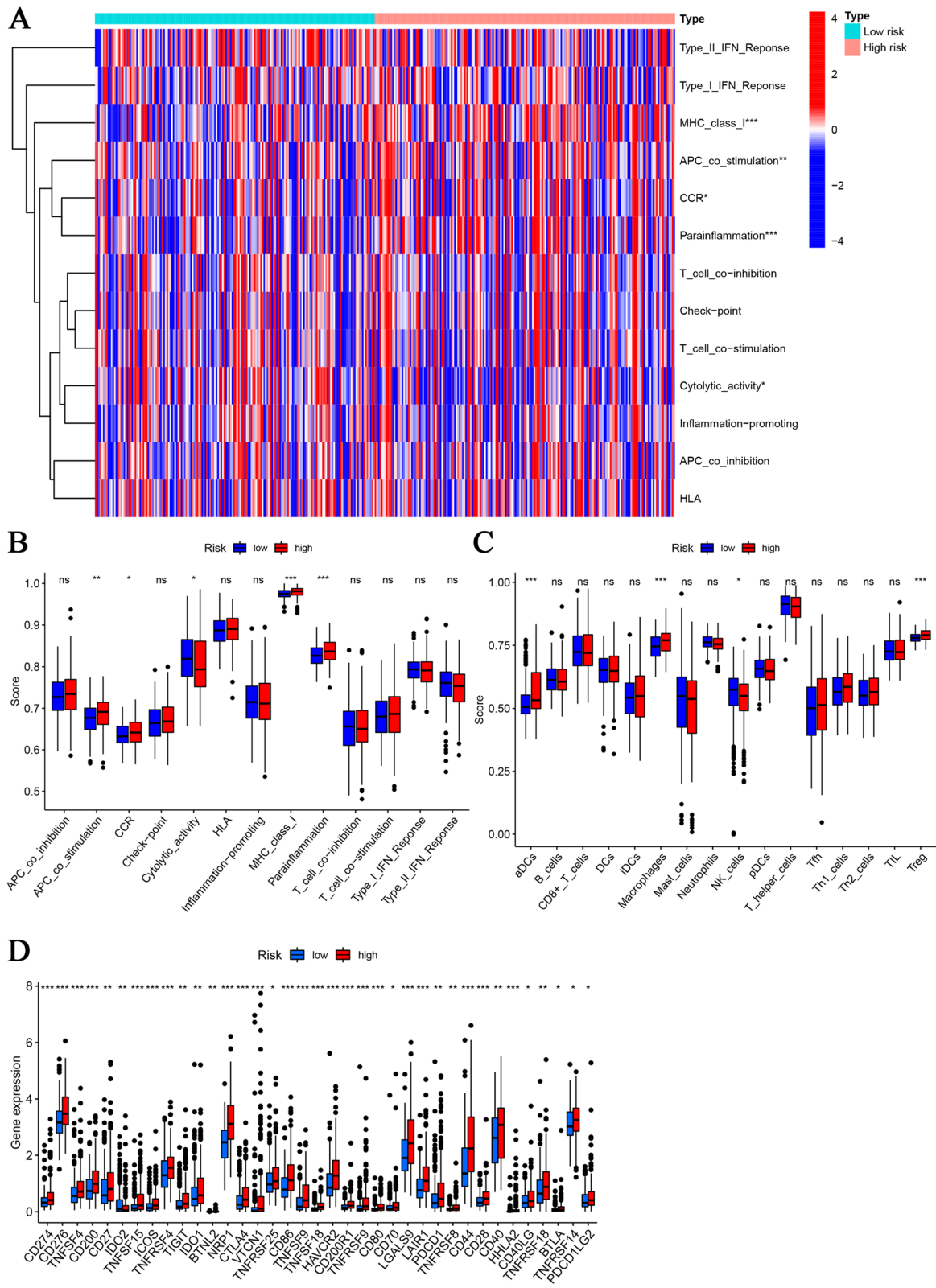
Wu *et al. BMC Cancer*   (2022) 22:1036

Page 12 of 18

Fig. 7 (See legend on previous page.)

Wu *et al. BMC Cancer*   (2022) 22:1036
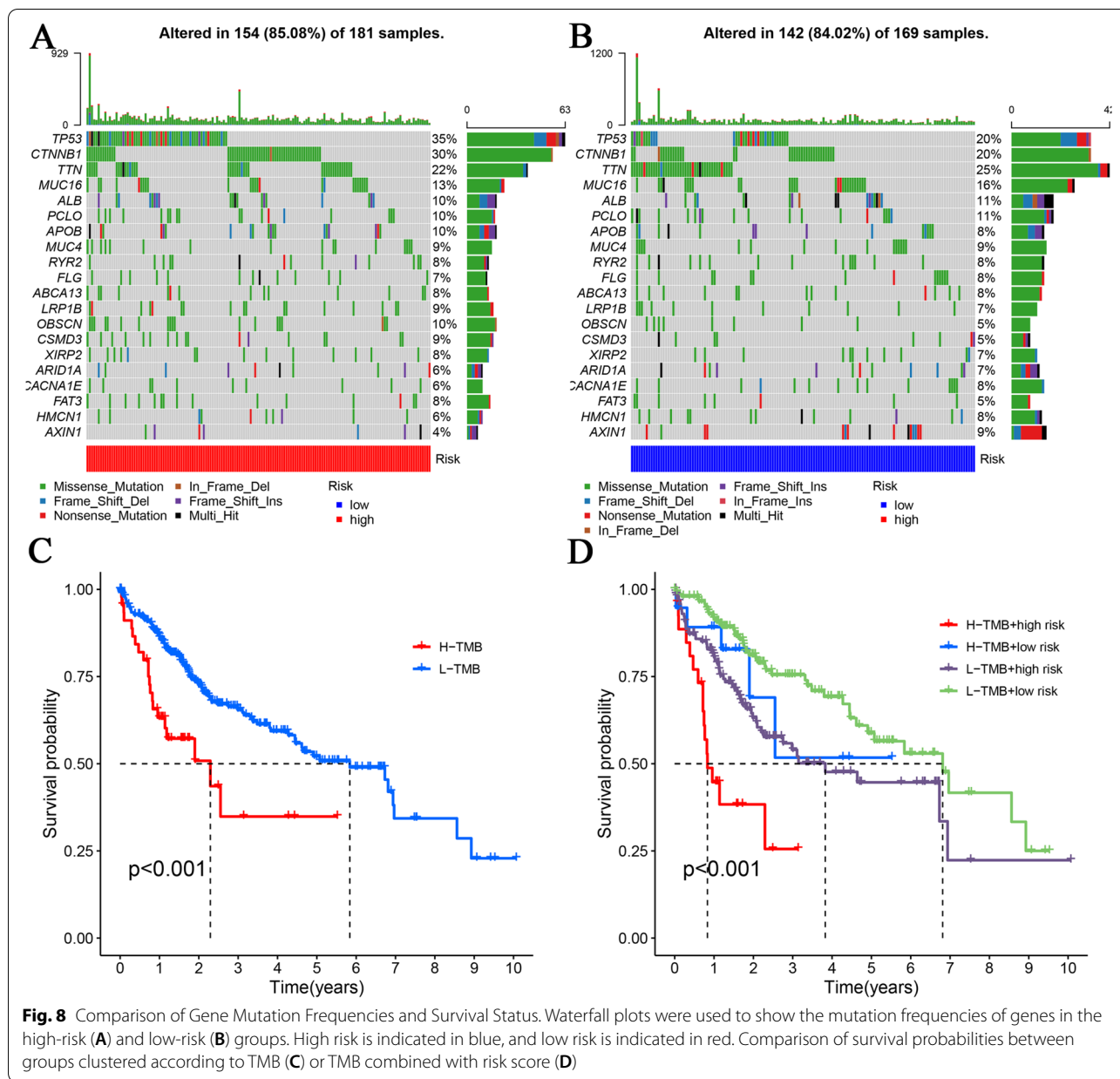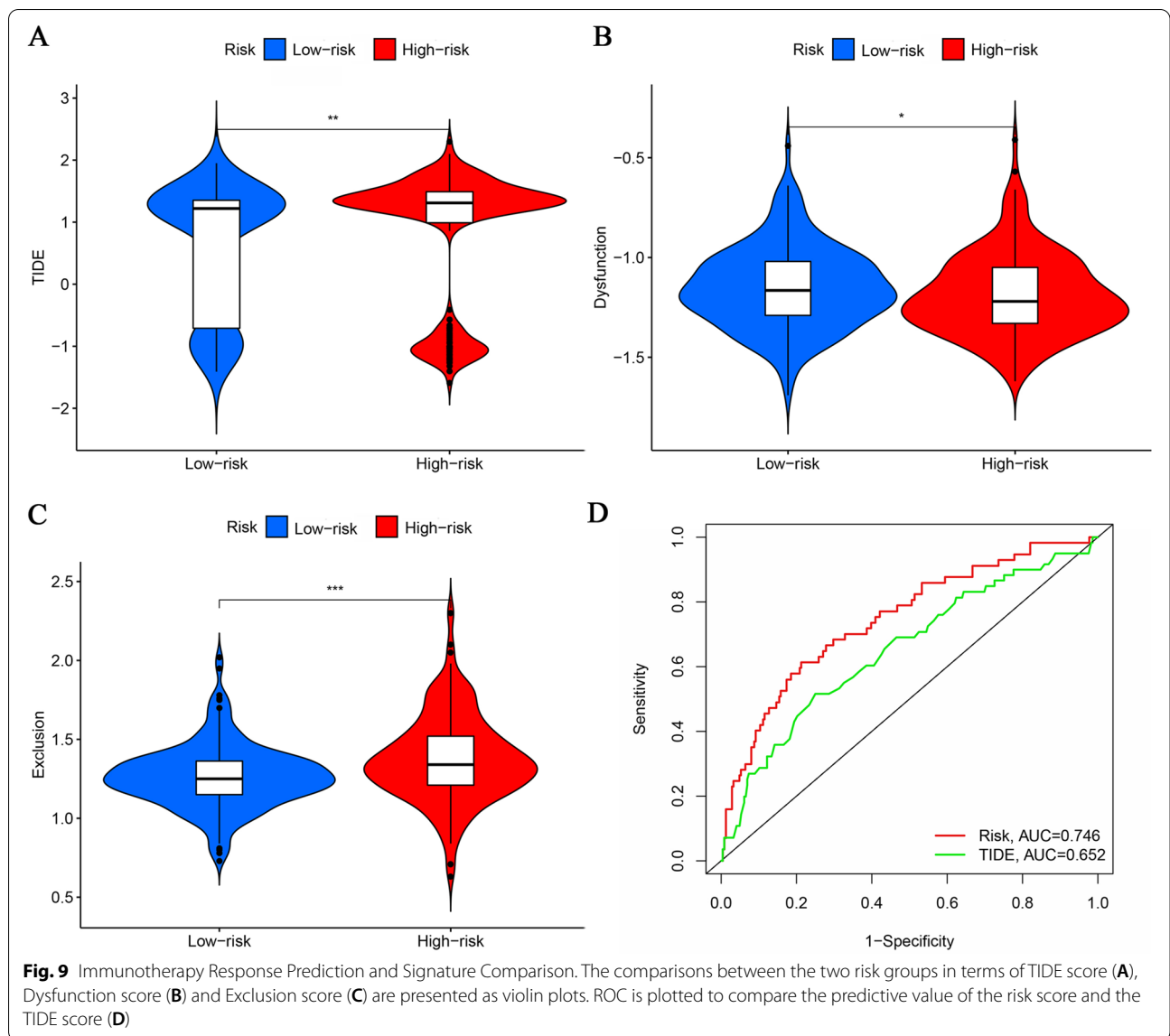
Page 13 of 18

Fig. 8 Comparison of Gene Mutation Frequencies and Survival Status. Waterfall plots were used to show the mutation frequencies of genes in the high-risk (**A**) and low-risk (**B**) groups. High risk is indicated in blue, and low risk is indicated in red. Comparison of survival probabilities between groups clustered according to TMB (**C**) or TMB combined with risk score (**D**)

Considering the poorly characterized immune microenvironment in HCC [29], the risk score determined in this study is a valuable marker in the differentiation of populations, particularly with respect to tumour-infiltrating immune cells, immune-related functions and immune checkpoint genes. We noted that the high-risk group was generally accompanied by a significant enrichment of immune cells associated with more immune functions and, more importantly, positively correlated with the expression level of immune checkpoint genes. Here, we identified DCs, Tregs, macrophages and CAFs as four hallmark immune cells given their remarkable

abundance in the high-risk set. DCs in the tumour microenvironment (TME) may be associated with T-cell dysfunction through mediator release or checkpoint ligand expression, a mechanism by which tumours cooperate with their microenvironment to eclipse immune surveillance [30, 31]. Specifically, in HCC, Treg cells can be recruited into tumour tissues, and this accumulation is mainly mediated by the chemokine CCL22, which is secreted by intratumoural DCs [32]. Tumour-infiltrating DCs and Tregs promote immunotolerance and immune escape by suppressing effector T-cell responses, but high intratumoral numbers of these immune cells may

Wu *et al. BMC Cancer*   (2022) 22:1036

Page 14 of 18

Fig. 9 Immunotherapy Response Prediction and Signature Comparison. The comparisons between the two risk groups in terms of TIDE score (**A**), Dysfunction score (**B**) and Exclusion score (**C**) are presented as violin plots. ROC is plotted to compare the predictive value of the risk score and the TIDE score (**D**)

be associated with poor prognosis [32, 33]. This notion is closely linked to our finding that increased levels of DCs and Tregs accumulated in the high-risk group (a high risk score was associated with a negative impact on clinical outcomes). Moreover, macrophages also showed increased enrichment in the high-risk group. The characterized enrichment of macrophages is notably linked to the survival inferiority of LIHC [30], indicating that this type of tumour-infiltrating macrophage can serve as

a potential candidate cell for cancer therapies. Cancer-associated fibroblasts (CAFs) are also key players in the pathogenesis of HCC. In the HCC TME, CAF-derived CLCF1 enhances the secretion of the chemokines CXCL6 and TGF-β in HCC cells, which are subsequently capable of activating the ERK 1/2 signalling pathway in CAFs and stimulating the increased production of CLCF1. This positive feedback mechanism undoubtedly accelerates HCC progression and predicts a poor prognosis [34].

(See figure on next page.)
Fig. 10 Assessment of Differential Expression of LncRNAs Used for Modelling. Comparison of the differential expression of AL031985.3 (**A**) and MKLN1-AS (**B**) between tumour and normal samples from the TCGA-LIHC project. Paired comparison of the differential expression of AL031985.3 (**C**) and MKLN1-AS (**D**) between tumour and normal samples from TCGA-LIHC project. Unpaired comparison of the differential expression of AL031985.3 (**E**) and MKLN1-AS (**F**) between 32 HCC tissues and 20 adjacent tissues. Paired comparison of the differential expression of AL031985.3 (**G**) and MKLN1-AS (**H**) between 20 HCC tissues and 20 adjacent tissues
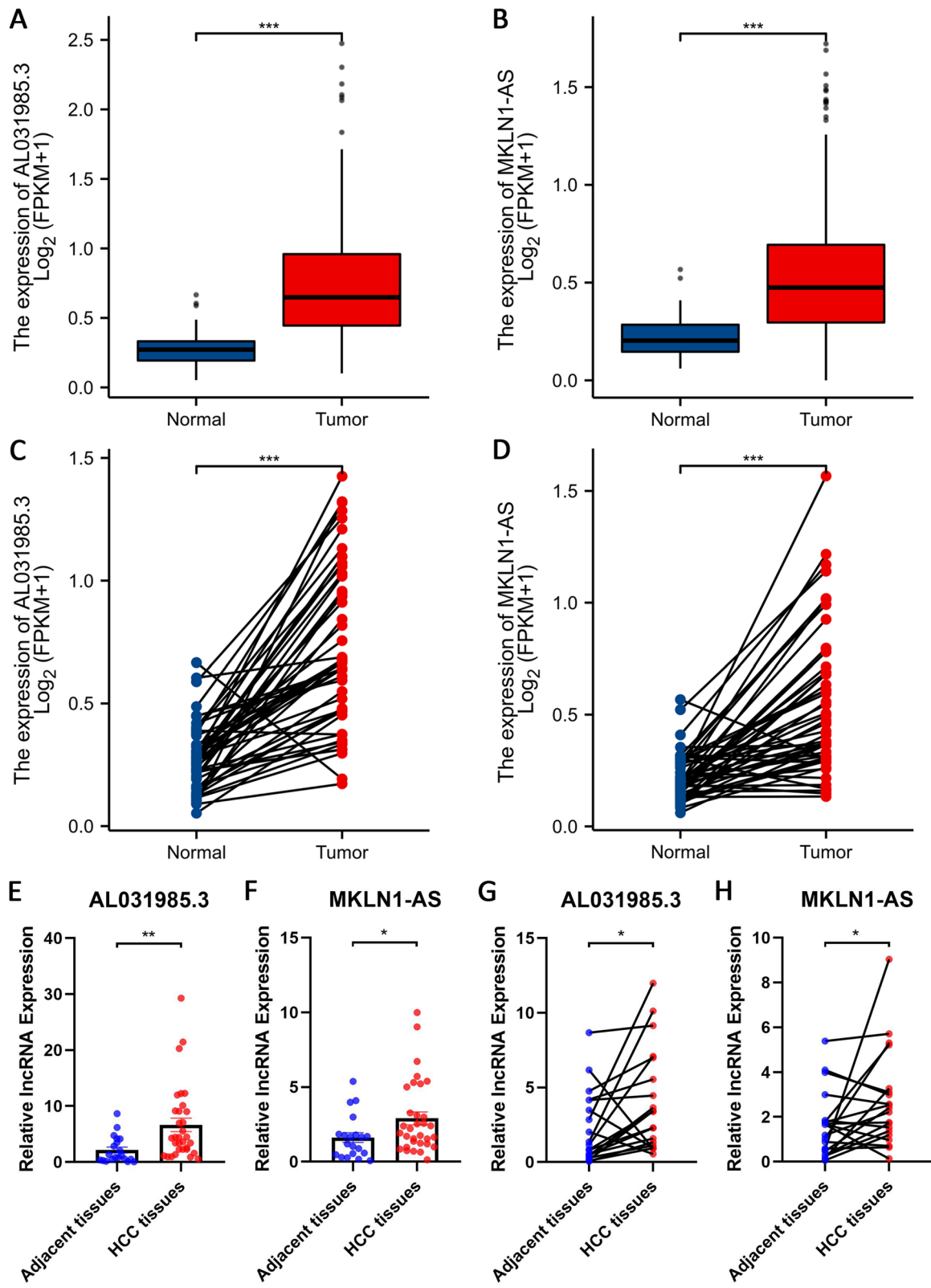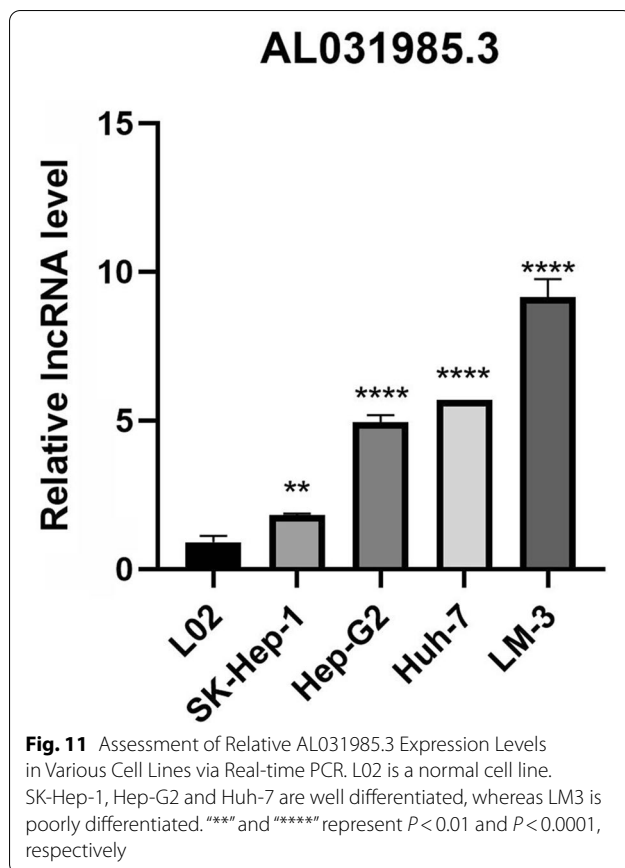
Wu *et al. BMC Cancer*   *(2022) 22:1036*

Page 15 of 18

Fig. 10 (See legend on previous page.)

Wu *et al. BMC Cancer* (2022) 22:1036

Page 16 of 18

Fig. 11 Assessment of Relative AL031985.3 Expression Levels in Various Cell Lines via Real-time PCR. L02 is a normal cell line. SK-Hep-1, Hep-G2 and Huh-7 are well differentiated, whereas LM3 is poorly differentiated. "**" and "****" represent $P < 0.01$ and $P < 0.0001$, respectively

Additional findings also revealed a positive feedback loop between CAFs and the FOXQ1/NDRG1 axis that drives the initiation of HCC in tumour cells [35]. Furthermore, loss of CAF-derived exosomal miR-320a may partly contribute to HCC tumour proliferation and metastasis [36]. These observations objectively confirmed our assumption that high-risk HCC patients might have an advanced cancer stage with reduced survival probability. NK cells were present at significantly decreased levels in high-risk samples in our analysis, and these cells were also decreased in HCC tissues [37]. Of note, the number of intrahepatic NK cells is positively associated with poor clinical outcomes [38]. In addition to the abovementioned immune cells, differential expression of all included immune checkpoint genes was observed between the two groups. On the one hand, our results demonstrated that the accuracy and importance of patient clustering based on this criterion (risk score) is self-evident. In addition, the differential expression levels of these genes in the two risk sets revealed insight into the precise management of HCC. The most prominent immune checkpoint gene is cytotoxic T lymphocyte protein 4 (CTLA-4), which mediates immunosuppression and inhibits T-cell proliferation and the release of cytotoxic mediators [39–42].

Furthermore, it is also particularly important to perform correlation analysis between our models and chemokines and cytokines. Chemokines comprise a large family of at least 47 structurally related small cytokines, and their interaction with the extracellular matrix is specifically essential in controlling the directional migration of cells. In contrast, cytokines are key mediators of cellular communication in the TME and are closely associated with tumour development, progression and metastasis as well as the response to tumour therapy [42]. Overall, analysis of LIHC samples revealed underlying migration patterns of immune cells and regulatory mechanisms of immune surveillance. Given their crucial role in the regulation of immune responses, infiltrating immune cells, immune checkpoint genes, chemokines and cytokines represent attractive targets for tumour immunotherapy.

A critical and novel project introduced in this study was the evaluation of the correlation between TMB and the model. ICIs have revolutionized the field of cancer management by providing a new paradigm. Several attempts have been made to determine valuable predictive biomarkers, the most intriguing of which is TMB [17]. In this study, we examined the mutation frequencies of the top genes and focused on the survival probability of patients based on various groupings. Through our comparative analysis of TMB in different risk groups along with exploration of differences in the expression levels of immune checkpoint genes, precise and effective management of ICIs in HCC treatment will be possible. Moreover, the findings of the drug sensitivity analysis of chemotherapeutic agents, targeted agents and immunotherapeutic agents also provide a reference for the rational choice of HCC treatment. More importantly, regarding the evaluation and prediction of immunotherapy efficacy, researchers found that the TIDE score was the best predictor of immune checkpoint inhibitor therapy compared to all candidate biomarkers as well as two single indicators of dysfunction and rejection as reported in the literature [18]. Based on the scoring results presented here, the high TIDE scores in the high-risk group strongly indicated that an increase in risk was also accompanied by increased resistance to immunotherapy to the extent that it might be associated with lower patient survival under ICI treatment.

However, some limitations associated with this study should be noted. First, all analyses were performed by mining public databases. Although we validated of the expression levels of lncRNAs associated with model construction, studies on their mechanistic roles and associated phenotypes are lacking. Moreover, practical application of our model remains an issue that needs to

Wu *et al. BMC Cancer* (2022) 22:1036

Page 17 of 18

be addressed, emphasizing the desperate need for in-depth work in the real world.

Conclusion

In summary, our series of comprehensive analyses contributed to the development of a valuable signature that exhibited promising prognostic performance and strongly correlated with immune cell infiltration characteristics and TMB. This signature will provide us with in-depth biological insights to facilitate the identification of prognostically and therapeutically relevant biomarkers for HCC.

Abbreviations

HCC: Hepatocellular Carcinoma; ECM: Extracellular Matrix; LASSO: Least Absolute Selection and Shrinkage Operator; TMB: Tumour Mutation Burden; ECMrl-ncRNAs: ECM-related LncRNAs; RNA-seq: RNA Sequencing; HR: Hazard Ratios; PCA: Principal Component Analysis; OS: Overall Survival; DEGs: Differentially Expressed Genes; ICIs: Immune Checkpoint Inhibitors; MSM: Masked Somatic Mutation; TME: Tumour Microenvironment; CTLA-4: Cytotoxic T Lymphocyte Protein 4.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12885-022-10049-w.

Additional file 1.

Declarations

Ethics approval and consent to participate

All data used in this paper were obtained from the publicly available TCGA database and the Matisome project. Ethical approval for the development of these public datasets has been obtained by other researchers, and approval for tissue cDNA microarrays was obtained by Shanghai Outdo Biotech Company. All of our methods in this study were performed in accordance with the Declaration of Helsinki.

Consent for publication

All authors consent for publication of the paper.

Competing interests

The authors declare no conflicts of interest.

Author details

[1]The First Clinical Medical College, Lanzhou University, Lanzhou 730000, Gansu, China. [2]Department of Gastroenterology, The First Hospital of Lanzhou University, No.1 West Donggang Road, Lanzhou 730000, Gansu, China. [3]Gansu Key Laboratory of Gastroenterology, Lanzhou University, Lanzhou 730000, Gansu, China. [4]Department of Radiology, the First Hospital of Lanzhou University, Lanzhou 730000, Gansu, China.

References

1. Villanueva A. Hepatocellular Carcinoma. N Engl J Med. 2019;380(15):1450–62.
2. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424.
3. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015;136(5):E359–86.
4. Kim HS, El-Serag HB. The epidemiology of hepatocellular carcinoma in the USA. Curr Gastroenterol Rep. 2019;21(4):17.
5. Theocharis AD, Skandalis SS, Gialeli C, et al. Extracellular matrix structure. Adv Drug Deliv Rev. 2016;97:4–27.
6. Karamanos NK, Theocharis AD, Neill T, et al. Matrix modeling and remodeling: a biological interplay regulating tissue homeostasis and diseases. Matrix Biol. 2019;75-76:1–11.
7. Seo BR, DelNero P, Fischbach C. In vitro models of tumor vessels and matrix: engineering approaches to investigate transport limitations and drug delivery in cancer. Adv Drug Deliv Rev. 2014;69-70:205–16.
8. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. Nat Rev Genet. 2009;10(3):155–9.
9. Wang Z, Feng C, Song K, et al. lncRNA-H19/miR-29a axis affected the viability and apoptosis of keloid fibroblasts through acting upon COL1A1 signaling. J Cell Biochem. 2020;121(11):4364–76.
10. Jiang R, Tang J, Chen Y, et al. The long noncoding RNA lnc-EGFR stimulates T-regulatory cells differentiation thus promoting hepatocellular carcinoma immune evasion. Nat Commun. 2017;8:15129.
11. Pachera E, Assassi S, Salazar GA, et al. Long noncoding RNA H19X is a key mediator of TGF-β-driven fibrosis. J Clin Invest. 2020;130(9):4888–905.
12. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.
13. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. Protein Sci. 2019;28(11):1947–51.
14. Kanehisa M, Furumichi M, Sato Y, et al. KEGG: integrating viruses and cellular organisms. Nucleic Acids Res. 2021;49(D1):D545–51.
15. Ozga AJ, Chow MT, Luster AD. Chemokines and the immune response to cancer. Immunity. 2021;54(5):859–74.
16. Rico MN, Anugwom CM, Boonstra A, et al. The role of cytokines in the different stages of hepatocellular carcinoma. Cancers (Basel). 2021;13(19):4876.
17. Addeo A, Friedlaender A, Banna GL, et al. TMB or not TMB as a biomarker: that is the question. Crit Rev Oncol Hematol. 2021;163:103374.
18. Jiang P, Gu S, Pan D, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat Med. 2018;24(10):1550–8.
19. Hong W, Liang L, Gu Y, et al. Immune-related lncRNA to construct novel signature and predict the immune landscape of human hepatocellular carcinoma. Mol Ther Nucleic Acids. 2020;22:937–47.
20. Liu T, Wu H, Qi J, et al. Seven immune-related genes prognostic power and correlation with tumor-infiltrating immune cells in hepatocellular carcinoma. Cancer Med. 2020;9(20):7440–52.

Wu *et al. BMC Cancer*   (2022) 22:1036

Page 18 of 18

21. Wu ZH, Li ZW, Yang DL, et al. Development and validation of a Pyroptosis-related long non-coding RNA signature for hepatocellular carcinoma. Front Cell Dev Biol. 2021;9:713925.

22. Huang C, Zhao J, Zhu Z. Prognostic Nomogram of prognosis-related genes and Clinicopathological characteristics to predict the 5-year survival rate of Colon Cancer patients. Front Surg. 2021;8:681721.

23. Zhang S, Li X, Zhang X, et al. The Pyroptosis-related gene signature predicts the prognosis of hepatocellular carcinoma. Front Mol Biosci. 2021;8:781427.

24. Tang H, You T, Sun Z, et al. Extracellular matrix-based gene expression signature defines two prognostic subtypes of hepatocellular carcinoma with different immune microenvironment characteristics. Front Mol Biosci. 2022;9:839806.

25. Zhang X, Jiang M, Zhang X, et al. An extracellular matrix-based signature associated with immune microenvironment predicts the prognosis of patients with hepatocellular carcinoma. Clin Res Hepatol Gastroenterol. 2022;46(4):101877.

26. Chen X, Ye Q, Chen Z, et al. Long non-coding RNA muskelin 1 antisense RNA as a potential therapeutic target in hepatocellular carcinoma treatment. Bioengineered. 2022;13(5):12237–47.

27. Pan G, Zhang J, You F, et al. ETS proto-oncogene 1-activated muskelin 1 antisense RNA drives the malignant progression of hepatocellular carcinoma by targeting miR-22-3p to upregulate ETS proto-oncogene 1. Bioengineered. 2022;13(1):1346–58.

28. Gao W, Chen X, Chi W, et al. Long non-coding RNA MKLN1-AS aggravates hepatocellular carcinoma progression by functioning as a molecular sponge for miR-654-3p, thereby promoting hepatoma-derived growth factor expression. Int J Mol Med. 2020;46(5):1743–54.

29. Ruf B, Heinrich B, Greten TF. Immunobiology and immunotherapy of HCC: spotlight on innate and innate-like immune cells. Cell Mol Immunol. 2021;18(1):112–27.

30. Zhang Q, He Y, Luo N, et al. Landscape and dynamics of single immune cells in hepatocellular carcinoma. Cell. 2019;179(4):829–45.e20.

31. Lurje I, Hammerich L, Tacke F. Dendritic cell and T cell crosstalk in liver Fibrogenesis and Hepatocarcinogenesis: implications for prevention and therapy of liver Cancer. Int J Mol Sci. 2020;21(19):7378.

32. Wiedemann GM, Knott MM, Vetter VK, et al. Cancer cell-derived IL-1α induces CCL22 and the recruitment of regulatory T cells. Oncoimmunology. 2016;5(9):e1175794.

33. Ouyang FZ, Wu RQ, Wei Y, et al. Dendritic cell-elicited B-cell activation fosters immune privilege via IL-10 signals in hepatocellular carcinoma. Nat Commun. 2016;7:13453.

34. Song M, He J, Pan QZ, et al. Cancer-associated fibroblast-mediated cellular crosstalk supports hepatocellular carcinoma progression. Hepatology. 2021;73(5):1717–35.

35. Luo Q, Wang CQ, Yang LY, et al. FOXQ1/NDRG1 axis exacerbates hepatocellular carcinoma initiation via enhancing crosstalk between fibroblasts and tumor cells. Cancer Lett. 2018;417:21–34.

36. Zhang Z, Li X, Sun W, et al. Loss of exosomal miR-320a from cancer-associated fibroblasts contributes to HCC proliferation and metastasis. Cancer Lett. 2017;397:33–42.

37. Cai L, Zhang Z, Zhou L, et al. Functional impairment in circulating and intrahepatic NK cells and relative mechanism in hepatocellular carcinoma patients. Clin Immunol. 2008;129(3):428–37.

38. Chew V, Chen J, Lee D, et al. Chemokine-driven lymphocyte infiltration: an early intratumoural event determining long-term survival in resectable hepatocellular carcinoma. Gut. 2012;61(3):427–38.

39. Iñarrairaegui M, Melero I, Sangro B. Immunotherapy of hepatocellular carcinoma: facts and hopes. Clin Cancer Res. 2018;24(7):1518–24.

40. Prieto J, Melero I, Sangro B. Immunological landscape and immunotherapy of hepatocellular carcinoma. Nat Rev Gastroenterol Hepatol. 2015;12(12):681–700.

41. Llovet JM, Montal R, Sia D, et al. Molecular therapies and precision medicine for hepatocellular carcinoma. Nat Rev Clin Oncol. 2018;15(10):599–616.

42. Propper DJ, Balkwill FR. Harnessing cytokines and chemokines for cancer therapy. Nat Rev Clin Oncol. 2022;19(4):237–53.

Publisher's Note