

# Identification of Biomarkers of Hepatocellular Carcinoma Gene Prognosis Based on Immune-Related lncRNA Signature of Transcriptome Data

Andi Ma (✉ [andima.hunan@gmail.com](mailto:andima.hunan@gmail.com))

Institute of Physiology, Charité-Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany.

<https://orcid.org/0000-0002-8094-8322>

**Yukai Sun**

Max Delbrück Centrum für Molekulare Medizin Experimental and Clinical Research Center (MDC), AG Translational Oncology of Solid Tumors, Robert-Rössle-Straße 10, 13125 Berlin, Germany.

**Racheal O. Ogbodu**

Institute of Physiology, Charité-Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany.

**Ling Xiao**

Department of Histology and Embryology, School of Basic Medical Sciences, Central South University Xiangya School of Medicine, Changsha 410013, China.

**Haibing Deng**

Department of General Thoracic Surgery, Inselspital Universitätsspital Bern, Murtenstrasse 50, 3008, Swaziland.

**Hui Zhou**

Department of Lymphoma and Hematology, the Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University, Hunan Cancer Hospital, 410013 Changsha, Hunan, China.

---

## Research

**Keywords:** Long non-coding RNAs, hepatocellular carcinoma, signature, prognosis, immune-related

**Posted Date:** August 25th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-57303/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Identification of biomarkers of hepatocellular carcinoma  
gene prognosis based on immune-related lncRNA signature of transcriptome  
data**

Andi Ma<sup>1</sup>, Yukai Sun<sup>3</sup>, Racheal O. Ogbodu<sup>1</sup>, Ling Xiao<sup>5</sup>, Haibing Deng<sup>6</sup>, Hui Zhou<sup>\*2</sup>

Andi Ma

[andi.ma@charite.de](mailto:andi.ma@charite.de)

Yukai Sun

[Yukai.Sun@mdc-berlin.de](mailto:Yukai.Sun@mdc-berlin.de)

Racheal O. Ogbodu

[ogboduodiri@yahoo.com](mailto:ogboduodiri@yahoo.com)

Ling Xiao

[xiaolingcsu@csu.edu.cn](mailto:xiaolingcsu@csu.edu.cn)

Haibing Deng

[willd7324@gmail.com](mailto:willd7324@gmail.com)

Hui Zhou

[zhouhui9403@126.com](mailto:zhouhui9403@126.com)

- 1 . Institute of Physiology, Charité-Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany.
2. Department of Lymphoma and Hematology, the Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University, Hunan Cancer Hospital, 410013 Changsha, Hunan, China.
3. Max Delbrück Centrum für Molekulare Medizin Experimental and Clinical Research Center (MDC), AG Translational Oncology of Solid Tumors, Robert-Rössle-Straße 10, 13125 Berlin, Germany.
4. Department of Histology and Embryology, School of Basic Medical Sciences, Central South University Xiangya School of Medicine, Changsha 410013, China. E-mail: [xiaolingcsu@csu.edu.cn](mailto:xiaolingcsu@csu.edu.cn)
5. Department of General Thoracic Surgery, Inselspital Universitätsspital Bern, Murtenstrasse 50, 3008, Swaziland.

**Abstract**

**Background:**

It is well known that long non-coding RNAs (lncRNAs) play a vital role in cancer. We aimed to explore the prognostic value of potential immune-related lncRNAs in hepatocellular carcinoma (HCC).

**Methods:**

Validated the established lncRNA signature of 343 patients with HCC from The Cancer Genome Atlas (TCGA) and 81 samples from Gene Expression Omnibus (GEO). Immune-related lncRNAs for HCC prognosis were evaluated using Cox regression and Least Absolute Shrinkage and Selection Operator (LASSO) analyses. LASSO analysis was performed to calculate a risk score formula to explore the difference in overall survival between high- and low-risk groups in TCGA, which was verified using GEO, Gene Ontology (GO), and pathway-enrichment analysis. These analyses were used to identify the function of screened genes and construct a co-expression network of these genes.

**Results:**

Using computational difference algorithms and lasso Cox regression analysis, the differentially expressed and survival-related immune-related genes (IRGs) among patients with HCC were established as five novel immune-related lncRNA signatures (AC099850.3, AL031985.3, PRRT3-AS1, AC023157.3, MSC-AS1). Patients in the low-risk group showed significantly better survival than patients in the high-risk group ( $P = 3.033e-05$ ). The signature identified can be an effective prognostic factor to predict patient survival. The nomogram showed some clinical net benefits predicted by overall survival. In order to explore its underlying mechanism, several methods of enrichment were elucidated using Gene Set Enrichment Analysis.

**Conclusion:**

Identifying five immune-related lncRNA signatures has important clinical implications for predicting patient outcome and guiding tailored therapy for patients with HCC with further prospective validation.

**Keywords**

Long non-coding RNAs, hepatocellular carcinoma, signature, prognosis, immune-related

**Background**

Hepatocellular carcinoma (HCC) is the sixth most common cancer in the world and the fourth leading cause of cancer-related death, and its annual incidence is increasing rapidly (Bray *et al.*, 2018). Treatments for HCC include surgical resection, transplantation, ablation, transarterial chemoembolization, and tyrosine kinase inhibitors such as sorafenib, lenvatinib, and regorafenib have proven to be effective treatments for improving patient survival (Forner *et al.*, 2018). The latest research shows that through immune checkpoint treatment, cytotoxic T lymphocyte antigen 4 (CTLA-4) and programmed death 1 (PD-1) blocking antibodies, and chimeric antigen receptor (CAR) T cells can be used to neutralize cancer cells. Immunotherapy drugs have shown beneficial results in the treatment of other solid tumors (such as kidney and lung cancers, and melanoma) (Thoma, 2018, Rheinheimer *et al.*, 2020, Rodriguez-Cerdeira *et al.*, 2017). At the American Society of Clinical Oncology Annual Meeting 2019 (ASCO 2019), the phase I/II CheckMate-040 trial results showed that a combined objective response rate of nivolumab (PD-1 antibody) and ipilimumab (CTLA-4 antibody) was 31% and the longest duration of response was 17.5 months. These results indicate that immunotherapy has potentially beneficial effects on HCC.

In the past few years, long non-coding RNA (lncRNA) has attracted attention in tumor research. They are a class of non-coding RNAs (ncRNAs) longer than 200 nucleotides that have no protein-coding potential (Gao *et al.*, 2020). At present, there are more than 20 lncRNAs being studied for human cancer

and medical applications. They are differentially expressed in different tissues and cancers (including HCC)(Wu and Du Y, 2017). Although the role of most lncRNAs in liver cancer is still not understood fully, a small number of them have been extensively studied. For example, lncRNA HULC is upregulated in liver cancer and has been shown to promote liver cancer growth, metastasis, and drug resistance(Emmrich *et al.*, 2014, Lee *et al.*, 2015, Huang *et al.*, 2020, Chu *et al.*, 2011). These important roles and unique characteristics of lncRNA indicate that they have potential clinical value in the diagnosis and treatment of liver cancer.

In this study, a genome-wide analysis was performed for determining the lncRNA expression profile of patients with HCC from The Cancer Genome Atlas (TCGA) and the Gene Expression Omnibus (GEO) databases. This expression profile aided identification of potential immune-related lncRNA signals to predict the survival of patients with HCC. We determined the prognostic markers of five immune-related lncRNAs from the TCGA and GEO datasets. The results obtained in this study can help predict the prognosis of patients with HCC, provide tailored treatments, and further prospective verification.

## Results

### Differentially expressed immune genes co-expressed in TCGA and GEO databases and their interactions

After raw data pre-processing, samples from 343 patients with HCC (with a survival time longer than 30 days) and 50 healthy subjects contained in TCGA (using only patient samples) were used. An additional 81 HCC samples were retrieved from the GSE27150 data set. These samples were analyzed on the immune-related database through the Gene Set Enrichment Analysis (GSEA; <https://www.gsea-msigdb.org/gsea/index.jsp>). We used two datasets, IMMUNE\_RESPONSE (M13664) and IMMUNE\_SYSTEM\_PROCESS (M19817), to obtain immune-related genes (IRGs). A total of 331 IRGs were downloaded from the TCGA database. Then, by selecting the intersecting genes, 42 IRGs co-expressed in the TCGA and GEO databases were obtained. Subsequently, a total of 24 differentially expressed genes (DEGs) were identified, including five downregulated and 19 upregulated genes. The expression of each gene is displayed in the heat map (Fig. 1a), and the volcano map shows the distribution of genes (Fig. 1b).

Subsequently, the protein-protein interaction (PPI) network of the DEGs was constructed and visualized using the STRING (Search Tool for the Retrieval of Interacting Gene) database, as shown in (Fig. 1c). A total of 22 nodes and 47 edges were present in the PPI network, and the histogram shows the node number of each gene (Fig. 1d).

### The acquisition of risk genes and their correlation with clinical prognosis and their previous correlation

In order to obtain IRGs related to clinical prognosis from the TCGA database, univariate and lasso Cox regression analyses were performed.  $P < 0.001$  was set as statistically significant in univariate Cox regression analysis. The prognostic gene signature was calculated as: Risk Score =  $(0.0156018677981362 \times TRAF2 \text{ expression}) + (-0.021929917526684 \times MIA3 \text{ expression}) + (0.00172601388297704 \times CD24 \text{ expression}) + (0.0193704943364634 \times LYN \text{ expression})$

+ (0.111981892547848 × *MAP3K7* expression) + (0.0124416402682615 × *CCR1* expression) + (0.0235215834428141 × *CKLF* expression) + (0.101112508330369 × *HELLS* expression). The R package "survival" was used to explore the best dividing lines of patient risk score, and the Kaplan-Meier survival curves were constructed. Eight risk genes were obtained (Table 1), and the Kaplan-Meier survival curve (Fig. 2) demonstrates that these eight genes were related to prognosis in both TCGA and GEO databases and shows that the high-risk group had lower 5-year survival rate than the low-risk group of patients with HCC ( $P < 0.001$ ). Next, construction and analysis of the correlation matrix between different genes (Fig. 2) revealed that the IRG *MAP3K7* had the highest correlation with *HELLS* (helicase, lymphoid specific) gene, and the Spearman's rank correlation coefficient (rs) was calculated to be 0.46. All other rs values were  $\leq 0.4$ . The "CorrPlot" R package was utilized to create this figure.

**Table 1** Eight genes selected as prognosis-associated factors in hepatocellular carcinoma

Gene	Coef
<i>TRAF2</i>	0.0156018677981362
<i>MIA3</i>	-0.021929917526684
<i>CD24</i>	0.00172601388297704
<i>LYN</i>	0.0193704943364634
<i>MAP3K7</i>	0.111981892547848
<i>CCR1</i>	0.0124416402682615
<i>CKLF</i>	0.0235215834428141
<i>HELLS</i>	0.101112508330369

### Acquisition of immune-related risk lncRNAs and principal component analysis

To identify the immune-related lncRNAs, the lncRNAs in the TCGA database were co-expressed with eight previously related IRGs and checked for correlation using the "limma" R package.  $P$  values were considered statistically significant for the total score when  $< 0.001$ . Therefore, we obtained 678 differentially expressed lncRNAs, 11 of which were related to negative regulation, and the rest were related to positive regulation.

Subsequently, univariate and multivariate analyses of survival prognostic factors of these immune-related lncRNAs was performed using "survival" R package. In univariate Cox analysis, hazard ratio (HR; 95% confidence interval (CI))  $> 1$  indicated high risk, and  $HR < 1$  indicated low risk.  $P$  value (survival time)  $< 0.0002$  was considered statistically significant. A total of 14 related lncRNAs were obtained, of which one was low risk, and 13 were high risk (Fig. 3a). Then, multivariate Cox analysis was performed to obtain five immune-related risk lncRNAs (Table 2).

Then, we conducted principal component analysis (PCA) on immune genes, immune-related lncRNAs, and immune-related risk lncRNAs to test the ability to identify the prognosis of patients with HCC. The results show that immune genes (green and red circles) were clustered together, and immune-related lncRNAs (red-green and red circles) were more dispersed. Immune-related risk lncRNAs (red-green and red circles) were essentially separated (Fig. 3b).

**Table 2** Five immune-related lncRNAs selected as prognosis-associated factors in hepatocellular carcinoma

Gene symbol	Chromosome	P value	Hazard ratio	Coefficient
AC099850.3	Chr 17:59,197,566-59,215,226	0.019	1.3033	0.2649
AL031985.3	Chr 1: 40,464,319-40,466,767	0.045	1.7270	0.5464
PRRT3-AS1	Chr 3: 9,947,404-9,954,787	0.022	1.2646	0.2348
AC023157.3	Chr 12: 31,729,117-31,731,204	0.010	1.6056	0.4735
MSC-AS1	Chr 8 : 71,828,167-72,118,393	0.008	1.4034	0.3389

### Correlation between prognosis model and clinicopathological characteristics

The TCGA database included 343 patients with information on their clinical features such as gender, age, T (tumor), grade, clinical stage, lymph node metastasis, and survival time and status. Selecting 1.0 to be the median risk score for the entire discovery group as the cutoff value, 343 patients were classified as high-risk or low-risk patients (Fig. 4). The figure also shows the expression value, risk score, and survival time and status of the five lncRNAs of each patient. Further analysis of the data revealed that the Kaplan-Meier curve shows that the survival probability of the high-risk patients was significantly lower than that of the low-risk patients, and the *P* value of the log-rank test was  $3.033e-05$  (Fig. 5a). In addition, the area under the curve (AUC) of patients' survival time from the TCGA database revealed that the AUC of risk genes = 0.762, the AUC of stage = 0.743, and the AUC of tumor (T) stage = 0.752, indicating that immune-related lncRNA signatures have a moderate potential to monitor survival (Fig. 5b). Compared with other clinicopathological features, our model achieved the largest AUC value, which also reflects its excellent predictability.

We then analyzed the relationship between the expression of risk genes and clinical features such as T, grade (G), and stage (S). In the TCGA database, the expression of AC099850.3, AL031985.3, and PRRT3-AS1 was found to be associated with T, AL031985.3, PRRT3-AS1, and MSC-AS1 were found to be associated with grade, and AC099850.3 and AL031985.3 were found to be related to the stage ( $P < 0.05$ ; Fig. 5c).

### Building and validating the predictive nomogram

The TCGA data set included 343 patients with complete clinical information. Based on the stepwise Cox regression model, a prognostic nomogram that predicts the overall survival rate at 1, 2, and 3 years was established (Fig. 6). The parameters such as risk score, age, gender, grade, stage, and tumor, node, metastasis (TMN) were included in the nomogram.

### Gene set enrichment analyses

We further analyzed immune-related lncRNAs and the enrichment score from the GSEA analysis. This analysis used the permutation lists, and the leading-edge proteins were considered the hits before

the enrichment score for the gene set. GSEA was performed for elucidating Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and Gene Ontology (GO) analysis. For GO pathways, most enrichment pathways in high-risk patients were linked to cell cycle, gene signal, and RNA binding. In addition, in the low-risk patients, most protein-related pathways were found to be enriched. KEGG pathway analysis showed that in the high-risk group, most cell cycle-related pathways were found to be enriched, and in the low-risk patients, most metabolic pathways were found to be enriched (Fig. 7).

## Discussion

As a heterogeneous disease with poor prognosis and high mortality, it is crucial to investigate novel biomarkers that can effectively predict the survival of patients with HCC. In recent years, with the development of sequencing, many human transcripts without protein coding potential called non-coding RNAs (ncRNA), including microRNA (miRNA), small nucleolar RNA (snoRNA), long non-coding RNA (lncRNA), and circular RNA (circRNA) have been identified. Our understanding of lncRNAs has greatly expanded, and it is known that they are more abundant than standard protein-coding mRNA (Chen and Conn, 2017, Shi *et al.*, 2013). They are known to play a role in gene expression, gene imprinting, chromatin remodeling, transcription, post-transcription, and epigenetic regulation (Zhang *et al.*, 2019, Momen-Heravi and Bala, 2018). A variety of lncRNAs are abnormally expressed in HCC and participate in tumor metastasis, recurrence, prognosis, etc. This includes interactions with DNA, RNA, or protein to form a complex that regulates target gene expression through various mechanisms (Liu *et al.*, 2019, Lou *et al.*, 2019, Shi *et al.*, 2019, Wang *et al.*, 2019, Zhang *et al.*, 2019, Wei *et al.*, 2019). lncRNAs also have important roles in cancer immunity, such as antigen release, immune cell migration and filtration, antigen presentation, and immune activation (Chen *et al.*, 2017, Atianand *et al.*, 2017). Immunotherapy for liver cancer has also received extensive attention (Dal Bo *et al.*, 2020, Sim and Knox, 2018). TCGA is an open database that contains samples of hundreds of patients with various malignancies. The GEO database has been widely used in bioinformatics analysis because it is a comprehensive database that stores large amounts of gene expression data. In this study, we downloaded RNA sequencing data from TCGA and GEO databases, and obtained the lncRNA expression profiles of patients with HCC in the dataset for analysis.

In this study, we used a TCGA training cohort including 343 patients with HCC, examined the relationship between the expression of immune-related lncRNAs and the prognosis of patients with HCC, and verified their effectiveness in the GSE27150 dataset. First, by co-expressing genes from two databases and performing Cox regression analysis, 24 genes related to immunity were obtained. Univariate and lasso Cox regression analyses were used to identify related immune genes to predict and construct prognostic gene signatures. Eight genes related to immunity were generated as prognostic factors for HCC. Then, through co-expression, univariate, and multivariate analyses, five immune-related lncRNAs affecting the prognosis were obtained. The overall survival rate of patients was found to be significantly correlated with lncRNAs. Through multiple Cox regression analysis and risk-scoring methods, five immune-related lncRNA signals were obtained, and patients with HCC were divided into high-risk and low-risk groups. Survival analysis showed that the overall survival rates of the two groups were significantly different. Since there was a possibility of overturning or five lncRNA signatures being false positive, we used a separate cohort of 343 patients with HCC to further validate the prognostic value of the five lncRNA signatures. The results of the independent verification showed that the five

lncRNA signatures had good repeatability and robustness in predicting the prognosis of patients with HCC. However, due to molecular heterogeneity, traditional clinical predictors had limited success in predicting the survival of patients with HCC. The results of multivariate analysis indicate that the five lncRNA signatures did not depend on traditional clinical risk factors and molecular characteristics.

In addition, in order to detect the correlation of the survival time of patients with HCC using the AUC, we found that the AUC of five immune-related lncRNAs = 0.762, then the AUC at S = 0.743, and the AUC at T = 0.752, indicating that the related lncRNA signal has medium potential. Compared with other clinicopathological features, our model obtains the largest AUC value, which also reflects its excellent predictability. In addition, we found a close relationship between dangerous lncRNA expression and clinicopathological factors (T, G, S). In order to make lncRNAs more effective and clinically meaningful, we established and validated the predicted nomogram, which was able to effectively determine the patient's risk. Finally, the results of GSEA showed that risk lncRNAs are associated with multiple pathways in KEGG and GO. On the other hand, the results further demonstrate a strong link between the signature and the immune system, and indicate the potential of signatures in predicting the survival of patients with HCC.

## **Conclusions**

In summary, this data analysis study identified five immune-related lncRNA signals that can predict the prognosis of patients with HCC. Our findings also provide a basis for the selection of therapeutic targets and prognostic molecular markers.

## **Material and Methods**

### **Patient datasets**

The study employed the gene expression profiles and clinical data from TCGA (<https://cancergenome.nih.gov/>, January 06, 2020), enrolled 50 normal and 343 patients with HCC with survival data, and GSE27150 dataset from the GEO database (<https://www.ncbi.nlm.nih.gov/>), including 81 patients with HCC.

### **Identification of DEGs**

The immune-related gene set in the GSEA database (gene set: IMMUNE\_RESPONSE, M19817, and IMMUNE\_SYSTEM\_PROCESS, M13664) was identified, then the IRGs were obtained using genetic data from TCGA through the co-expression method (co-data moderated random forest (CoRF) > 0.4,  $P < 0.001$ ). Then, using the intersection, 24 IRGs co-expressed in the TCGA and GEO databases were obtained. Subsequently, DEGs were identified using “limma” package (<http://www.bioconductor.org/>) in R software version 3.6.2. The threshold value of DEGs was set as log-fold change (FC) > 1.0 and  $P$  value < 0.05. Further, “pheatmap” package in R software was used for clustering analysis.

### **PPI network analysis**

The PPI network was constructed based on online search databases (STRING; <http://string-db.org>)

(version 11.0) (Franceschini *et al.*, 2013) for identifying interactive genes to make predictions. Analyzing functional interactions between proteins may provide insights into disease development or development mechanisms, including direct (physical) and indirect (functional) associations. In this study, the STRING database was used to evaluate their functional associations and construct a PPI network, and interactions with a combined score  $> 0.4$  were considered statistically significant. Subsequently, R software was employed to obtain the histogram of node number.

### **Construction of the prognostic immune gene signature**

Univariate and lasso Cox regression analyses (O'Connor, 2000) were used to identify relevant immune genes to predict and construct prognostic gene signatures.  $P < 0.001$  was considered statistically significant in univariate Cox regression analysis. The prognostic gene signature was calculated as: Risk Score = (coefficient of gene<sub>1</sub> × gene<sub>1</sub> expression) + (coefficient of gene<sub>2</sub> × gene<sub>2</sub> expression) + ... + (coefficient of gene<sub>N</sub> × Gene<sub>N</sub> expression). The R package "survival" was used to explore the best dividing line of patient risk score, and draw Kaplan-Meier survival curve ( $P < 0.001$ ). Next, the "CorrPlot" R package was used to construct the correlation matrix between different genes.

### **Acquisition of immune-related risk lncRNAs and PCA**

The "limma" R software package was used to screen differential genes between lncRNA and the eight genes obtained, and the absolute value of cor  $> 0.4$ . When cor is positive, it is related to positive regulation and vice versa.  $P < 0.001$  indicates statistical significance.

Then, we conducted univariate and multivariate analyses on the survival prognostic factors of these immune-related lncRNAs using the "survival" R package. In univariate Cox analysis, HR (95% CI)  $> 1$  indicated high risk, and HR  $< 1$  indicated low risk.  $P$  value (survival time)  $< 0.0002$  indicated statistical significance.

### **Correlation between prognosis model and clinicopathological characteristics**

Using the "pheatmap" package for analysis, the median risk score of the entire discovery group was set to 1.0 as the cutoff value, and 343 patients with HCC were divided into low-risk and high-risk patients. The relationship between the expression value of five lncRNAs and risk score, survival time, and status of each patient was analyzed. The "survival" and "survminer" packages were used to obtain Kaplan-Meier curves and  $P < 0.01$  was considered to be statistically significant. The R software package "survivalROC" was used to study the prognostic value of gene signatures over time (Lambert and Chevret, 2016). Then, the R software package "ggpubr" was used to analyze the relationship between risk gene expression and clinical characteristics (such as T, G, and S).

### **Building and validating a predictive nomogram**

Nomogram (Iasonos *et al.*, 2008) was established by including all the prognostic factors. The calibration chart and the consistency index were used to study the calibration and identification of the nomogram (using the bootstrap method with 1,000 replicates), respectively. Based on the stepwise Cox regression model, the prognostic nomograms predicting the total survival time of 1, 2, and 3 years were established.

## **GSEA**

GSEA is supported by the Broad Institute website (<https://www.gsea-msigdb.org/gsea/index.jsp>)(Croken *et al.*, 2014), which includes versions compatible with Java and R. All GSEA analysis was performed using Java GSEA.

### **Statistical analysis**

R (version 3.6.2) was used to analyze the data. The “limma” package was used to identify DEGs. The threshold value of DEGs was set to  $\log_2$  fold change (FC) > 1.0 and  $P$  value < 0.05. The "survival" and "survminer" R packages were used to explore the best dividing line of risk score, and draw the Kaplan-Meier survival curve ( $P < 0.001$ ). The “CorrPlot” R package was used to construct the correlation matrix between different genes. In univariate Cox analysis, HR (95% CI) > 1 indicated high risk, and HR < 1 indicated low risk.  $P$  value (survival time) < 0.0002 indicated statistical significance. The R software package "survivalROC" was used to study the prognostic value of gene signatures over time. The R software package "ggpubr" was used to analyze the relationship between risk gene expression and clinical characteristics.

### **Abbreviations**

(lncRNAs) long non-coding RNAs; (HCC) hepatocellular carcinoma; (TCGA) The Cancer Genome Atlas; (GEO) Gene Expression Omnibus; (LASSO) least Absolute Shrinkage and Selection Operator; (GO) Gene Ontology; (IRGs) immune-related genes; (CAR) chimeric antigen receptor; (ncRNAs) non-coding RNAs; (DEGs) differentially expressed genes; (PPI) protein-protein interaction; (CI) confidence interval; (PCA) principal component analysis; (AUC) area under the curve; (KEGG) Kyoto Encyclopedia of Genes and Genomes; (miRNA) microRNA; (snRNA) small nucleolar RNA; (lncRNA) long non-coding RNA; (circRNA) circular RNA.

### **Declarations**

**Ethics approval and consent to participate:** Ethical approval was waived since we used only publicly available data and materials in this study.

**Consent for publication:** Not applicable

**Availability of data and material:** The GSE27150 gene expression profiles came from GEO (<https://www.ncbi.nlm.nih.gov/geo>). The TCGA data came from GDC data portal (<https://portal.gdc.cancer.gov/>). The immune-related genes (IRGs) data came from the ImmPort database (<https://import.niaid.nih.gov>). The statistical analyses used the R software (<https://www.r-project.org/>).

**Competing interest:** All authors have read and approved to submit it to your journal. There are no conflicts of interest of any authors in relation to the submission.

**Funding:** No funding

**Authors' contributions:** Andi Ma designed the research plan, analyzed datasets and wrote the manuscript. Yukai Sun, Racheal O. Ogbodu and LingXiao provided meaningful

discussion of key points. HaibingDeng and HuiZhou revised the manuscript and assisted the production of figures. All authors read and approved the final manuscript.

**Acknowledgements:** Thank for contributors of TCGA and GEO for sharing the melanoma data.

## References:

- ATIANAND, M. K., CAFFREY, D. R. & FITZGERALD, K. A. (2017), "Immunobiology of Long Noncoding RNAs", *Annu Rev Immunol*, Vol. 35:177-198.
- BRAY, F., FERLAY, J., SOERJOMATARAM, I., SIEGEL, R. L., TORRE, L. A. & JEMAL, A. (2018), "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries", *CA Cancer J Clin*, Vol. 68 No. 6, pp. 394-424.
- CHEN, J. A. & CONN, S. (2017), "Canonical mRNA is the exception, rather than the rule", *Genome Biol*, Vol. 18 No. 1, pp. 133.
- CHEN, Y. G., SATPATHY, A. T. & CHANG, H. Y. (2017), "Gene regulation in the immune system by long noncoding RNAs", *Nat Immunol*, Vol. 18 No. 9, pp. 962-972.
- CHU, C., QU, K., ZHONG, F. L., ARTANDI, S. E. & CHANG, H. Y. (2011), "Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions", *Mol Cell*, Vol. 44 No. 4, pp. 667-78.
- CROKEN, M. M., QIU, W., WHITE, M. W. & KIM, K. (2014), "Gene Set Enrichment Analysis (GSEA) of *Toxoplasma gondii* expression datasets links cell cycle progression and the bradyzoite developmental program", *BMC Genomics*, Vol. 15:515.
- DAL BO, M., DE MATTIA, E., BABOCI, L., MEZZALIRA, S., CECCHIN, E., ASSARAF, Y. G. & TOFFOLI, G. (2020), "New insights into the pharmacological, immunological, and CAR-T-cell approaches in the treatment of hepatocellular carcinoma", *Drug Resist Updat*, Vol. 51:100702.
- EMMRICH, S., STRELTSOV, A., SCHMIDT, F., THANGAPANDI, V. R., REINHARDT, D. & KLUSMANN, J. H. (2014), "LincRNAs MONC and MIR100HG act as oncogenes in acute megakaryoblastic leukemia", *Mol Cancer*, Vol. 13:171.
- FORNER, A., REIG, M. & BRUIX, J. (2018), "Hepatocellular carcinoma", *Lancet*, Vol. 391 No. 10127, pp. 1301-1314.
- FRANCESCHINI, A., SZKLARCZYK, D., FRANKILD, S., KUHN, M., SIMONOVIC, M., ROTH, A., LIN, J., MINGUEZ, P., BORK, P., VON MERING, C. & JENSEN, L. J. (2013), "STRING v9.1: protein-protein interaction networks, with increased coverage and integration", *Nucleic Acids Res*, Vol. 41 No. Database issue, pp. D808-15.
- GAO, F., CAI, Y., KAPRANOV, P. & XU, D. (2020), "Reverse-genetics studies of lncRNAs-what we have learnt and paths forward", *Genome Biol*, Vol. 21 No. 1, pp. 93.
- HUANG, Z., ZHOU, J. K., PENG, Y., HE, W. & HUANG, C. (2020), "The role of long noncoding RNAs in hepatocellular carcinoma", *Mol Cancer*, Vol. 19 No. 1, pp. 77.
- IASONOS, A., SCHRAG, D., RAJ, G. V. & PANAGEAS, K. S. (2008), "How to build and interpret a nomogram for cancer prognosis", *J Clin Oncol*, Vol. 26 No. 8, pp. 1364-70.

- LAMBERT, J. & CHEVRET, S. (2016), "Summary measure of discrimination in survival models based on cumulative/dynamic time-dependent ROC curves", *Stat Methods Med Res*, Vol. 25 No. 5, pp. 2088-2102.
- LEE, J. H., DAUGHARTHY, E. R., SCHEIMAN, J., KALHOR, R., FERRANTE, T. C., TERRY, R., TURCZYK, B. M., YANG, J. L., LEE, H. S., AACH, J., ZHANG, K. & CHURCH, G. M. (2015), "Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues", *Nat Protoc*, Vol. 10 No. 3, pp. 442-58.
- LIU, D., GAO, M., WU, K., ZHU, D., YANG, Y. & ZHAO, S. (2019), "LINC00152 facilitates tumorigenesis in esophageal squamous cell carcinoma via miR-153-3p/FYN axis", *Biomed Pharmacother*, Vol. 112108654.
- LOU, Y., YU, Y., XU, X., ZHOU, S., SHEN, H., FAN, T., WU, D., YIN, J. & LI, G. (2019), "Long non-coding RNA LUCAT1 promotes tumourigenesis by inhibiting ANXA2 phosphorylation in hepatocellular carcinoma", *J Cell Mol Med*, Vol. 23 No. 3, pp. 1873-1884.
- MOMEN-HERAVI, F. & BALA, S. (2018), "Emerging role of non-coding RNA in oral cancer", *Cell Signal*, Vol. 42134-143.
- O'CONNOR, H. J. (2000), "'Lasso' technique for retrieval of impacted biliary stent", *Gastrointest Endosc*, Vol. 51 No. 5, pp. 640-1.
- RHEINHEIMER, S., HEUSSEL, C. P., MAYER, P., GAISSMAIER, L., BOZORGMEHR, F., WINTER, H., HERTH, F. J., MULEY, T., LIERSCH, S., BISCHOFF, H., KRIEGSMANN, M., EL, S. R., STENZINGER, A., THOMAS, M., KAUCZOR, H. U. & CHRISTOPOULOS, P. (2020), "Oligoprogressive Non-Small-Cell Lung Cancer under Treatment with PD-(L)1 Inhibitors", *Cancers (Basel)*, Vol. 12 No. 4, pp.
- RODRIGUEZ-CERDEIRA, C., CARNERO, G. M., LOPEZ-BARCENAS, A., SANCHEZ-BLANCO, E., SANCHEZ-BLANCO, B., FABBROCINI, G., BARDHI, B., SINANI, A. & GUZMAN, R. A. (2017), "Advances in Immunotherapy for Melanoma: A Comprehensive Review", *Mediators Inflamm*, Vol. 20173264217.
- SHI, L., HONG, X., BA, L., HE, X., XIONG, Y., DING, Q., YANG, S. & PENG, G. (2019), "Long non-coding RNA ZNF1-AS1 promotes the tumor progression and metastasis of colorectal cancer by acting as a competing endogenous RNA of miR-144 to regulate EZH2 expression", *Cell Death Dis*, Vol. 10 No. 3, pp. 150.
- SHI, X., SUN, M., LIU, H., YAO, Y. & SONG, Y. (2013), "Long non-coding RNAs: a new frontier in the study of human diseases", *Cancer Lett*, Vol. 339 No. 2, pp. 159-66.
- SIM, H. W. & KNOX, J. (2018), "Hepatocellular carcinoma in the era of immunotherapy", *Curr Probl Cancer*, Vol. 42 No. 1, pp. 40-48.
- THOMA, C. (2018), "Kidney cancer: Combining targeted and immunotherapy", *Nat Rev Urol*, Vol. 15 No. 5, pp. 263.
- WANG, Y., YANG, L., CHEN, T., LIU, X., GUO, Y., ZHU, Q., TONG, X., YANG, W., XU, Q., HUANG, D. & TU, K. (2019), "A novel lncRNA MCM3AP-AS1 promotes the growth of hepatocellular carcinoma by targeting miR-194-5p/FOXA1 axis", *Mol Cancer*, Vol. 18 No. 1, pp. 28.
- WEI, L., WANG, X., LV, L., LIU, J., XING, H., SONG, Y., XIE, M., LEI, T., ZHANG, N. & YANG, M. (2019), "The emerging role of microRNAs and long noncoding RNAs in drug resistance of hepatocellular carcinoma", *Mol Cancer*, Vol. 18 No. 1, pp. 147.
- WU, T. & Du Y (2017), "LncRNAs: From Basic Research to Medical Application", *Int J Biol Sci*, Vol. 13 No. 3, pp. 295-307.

ZHANG, P., DONG, Q., ZHU, H., LI, S., SHI, L. & CHEN, X. (2019), "Long non-coding antisense RNA GAS6-AS1 supports gastric cancer progression via increasing GAS6 expression", *Gene*, Vol. 6961-9.

ZHANG, Y., LI, Z., CHEN, M., CHEN, H., ZHONG, Q., LIANG, L. & LI, B. (2019), "Identification of a New Eight-Long Noncoding RNA Molecular Signature for Breast Cancer Survival Prediction", *DNA Cell Biol*, Vol. 38 No. 12, pp. 1529-1539.

**Fig. 1 Immune genes co-expressed in The Cancer Genome Atlas (TCGA) Gene Expression Omnibus (GEO).** Differentially expressed immune-related genes (IRGs) and their interactions. Heatmap (a) and volcano map show differentially expressed IRGs with red and green dots representing differentially expressed genes (DEGs). (b) Visualization of the protein-protein interaction (PPI) network of DEGs. (c) Histogram showing the number of nodes of each gene (d)

**Fig. 2 Kaplan-Meier curves and interaction with eight genes:** Kaplan-Meier curves were used to analyze the overall survival rate of patients with hepatocellular carcinoma in high-risk and low-risk groups from the two databases, TCGA (The Cancer Genome Atlas; a) and GEO (Gene Expression Omnibus; b). (c) Terpene correlation plot. Visual representation of the correlation matrix; darker blue and larger dot size indicate a stronger positive correlation, while darker red indicates a stronger negative correlation

**Fig. 3 Acquisition of immune-related risk long non-coding RNAs (lncRNAs) and principal component analysis (PCA).** (a) A forest map of the hazard ratio (HR) of immune-related lncRNAs. (b) PCA for all immune genes, immune lncRNAs, and risk lncRNAs; red dots represent high-risk genes, and green dots represent low-risk genes

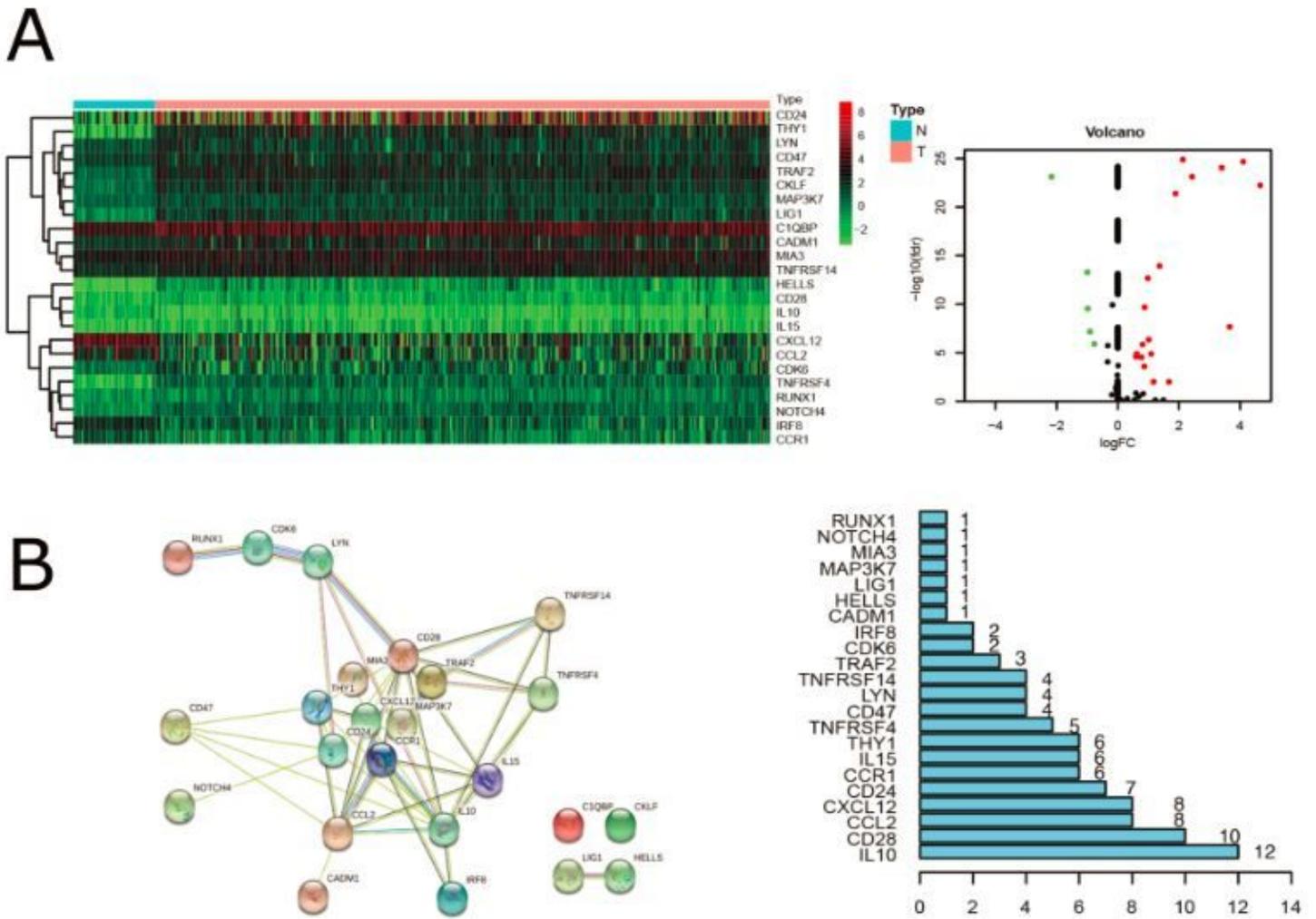
**Fig. 4 Characteristic of prognosis model:** Long non-coding RNA (lncRNA) expression (a), risk score (b) and survival time. (c) Patterns for patients in high-group and low-risk group determined using the five-immune-related lncRNA signature

**Fig. 5 Clinicopathological characteristics of model:** (a) Kaplan-Meier analysis of the overall survival of The Cancer Genome Atlas (TCGA) in high-risk ( $n = 171$ ) and low-risk subgroups ( $n = 173$ ). (b) Receiver operating characteristic curve (ROC) analysis of risk scores were used to predict the overall survival of the test set. (c) Association between clinical features (T (tumor), G (grade), and S (stage)) based on the five risk genes

**Fig. 6 Nomogram.** (a) Prognostic nomogram predicting 1-, 2-, and 3-year overall survival of patients with hepatocellular carcinoma

**Fig. 7 Gene set enrichment analyses (GSEA).** Significantly enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways by GSEA. (a) Ten representatives of KEGG pathways in high-risk and low-risk patients with hepatocellular carcinoma. (b) Ten representatives of Gene Ontology (GO) pathways in high-risk and low-risk patients with hepatocellular carcinoma

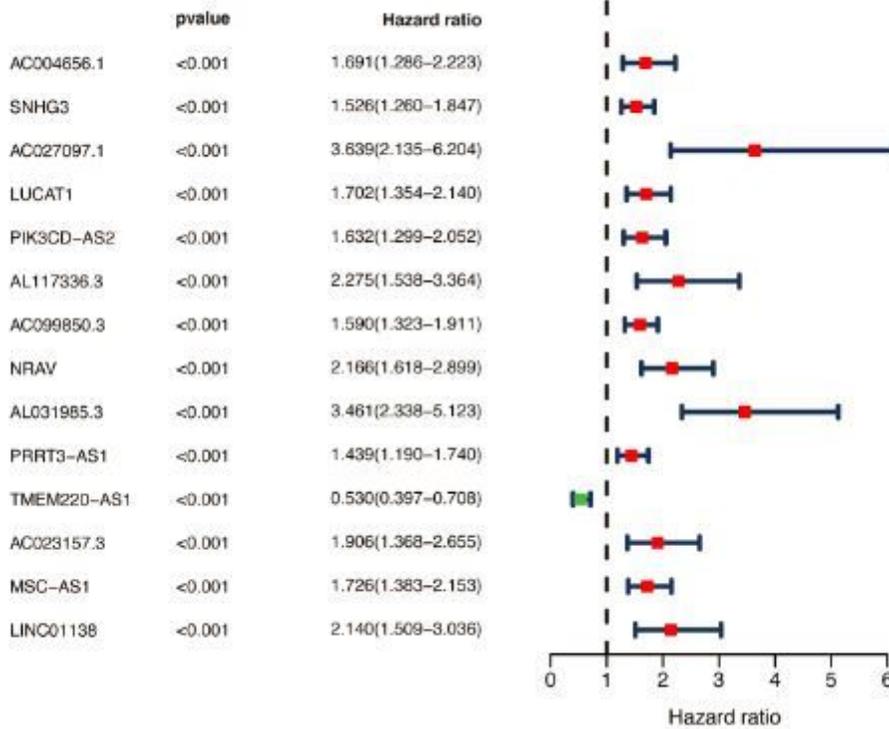
# Figures



**Figure 1**

Immune genes co-expressed in The Cancer Genome Atlas (TCGA) Gene Expression Omnibus (GEO). Differentially expressed immune-related genes (IRGs) and their interactions. Heatmap (a) and volcano map show differentially expressed IRGs with red and green dots representing differentially expressed genes (DEGs). (b) Visualization of the protein-protein interaction (PPI) network of DEGs. (c) Histogram showing the number of nodes of each gene (d)

A



B

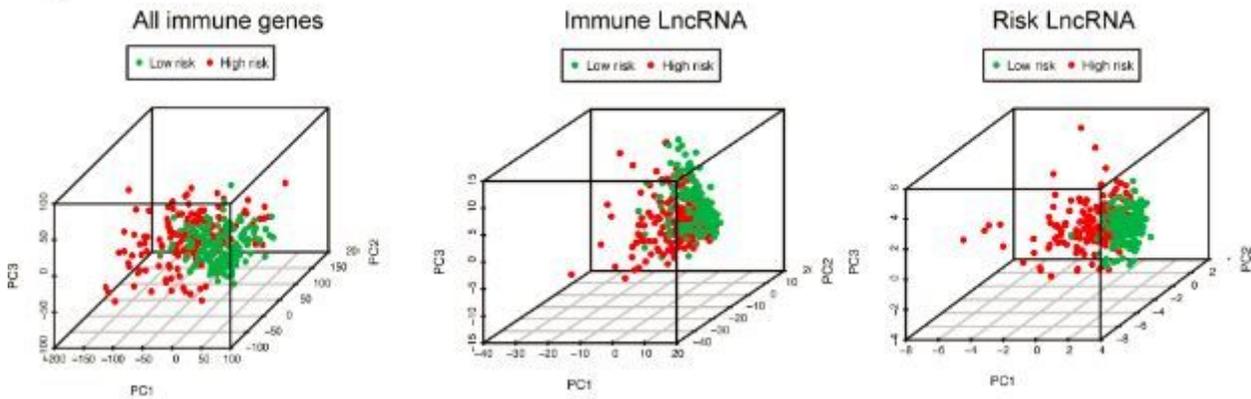
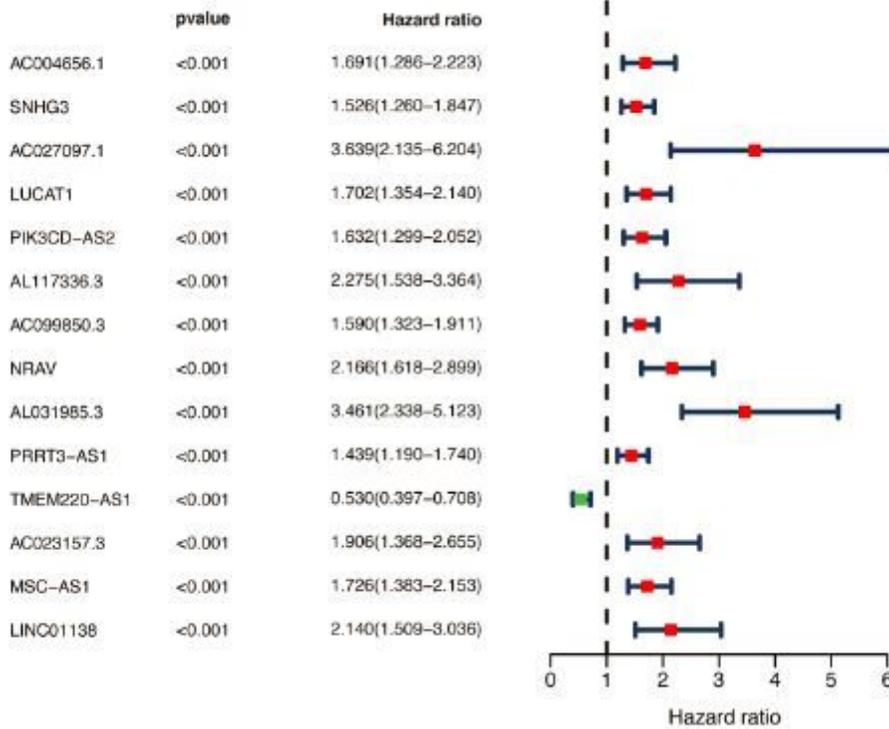


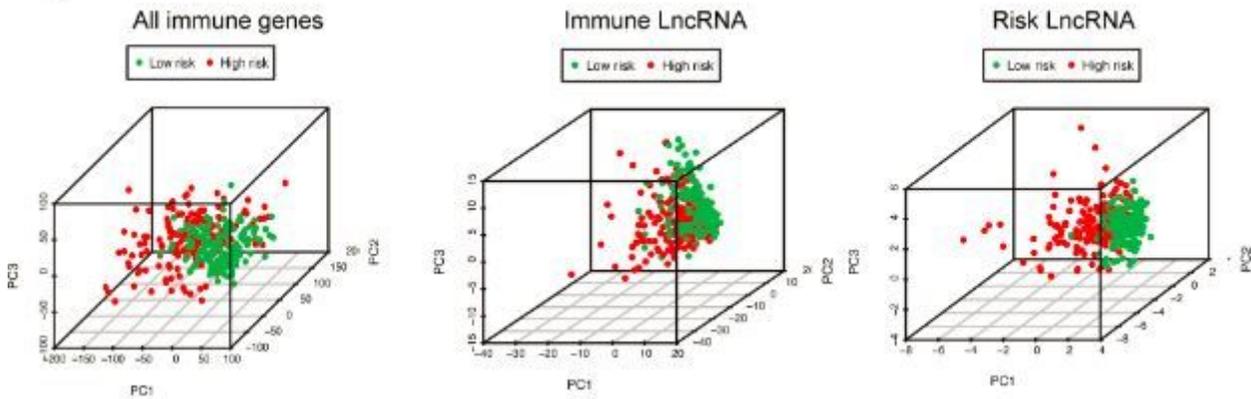
Figure 2

Kaplan-Meier curves and interaction with eight genes: Kaplan-Meier curves were used to analyze the overall survival rate of patients with hepatocellular carcinoma in high-risk and low-risk groups from the two databases, TCGA (The Cancer Genome Atlas; a) and GEO (Gene Expression Omnibus; b). (c) Terpene correlation plot. Visual representation of the correlation matrix; darker blue and larger dot size indicate a stronger positive correlation, while darker red indicates a stronger negative correlation

A

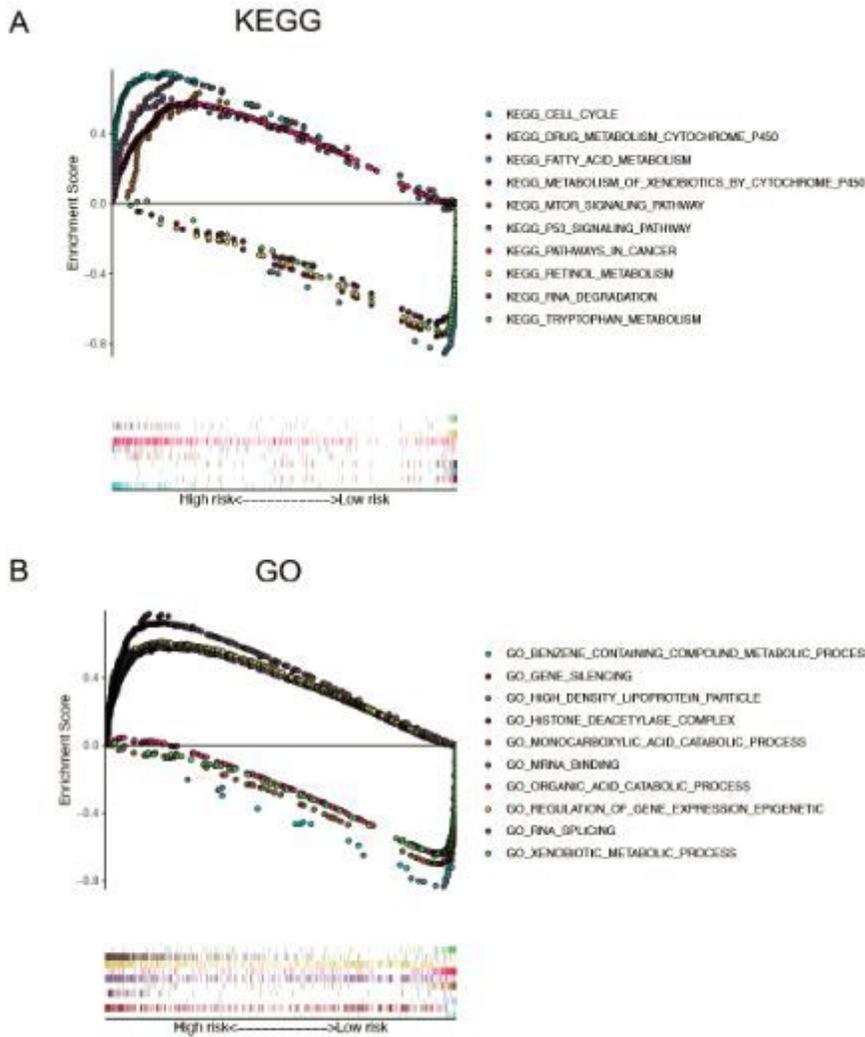


B



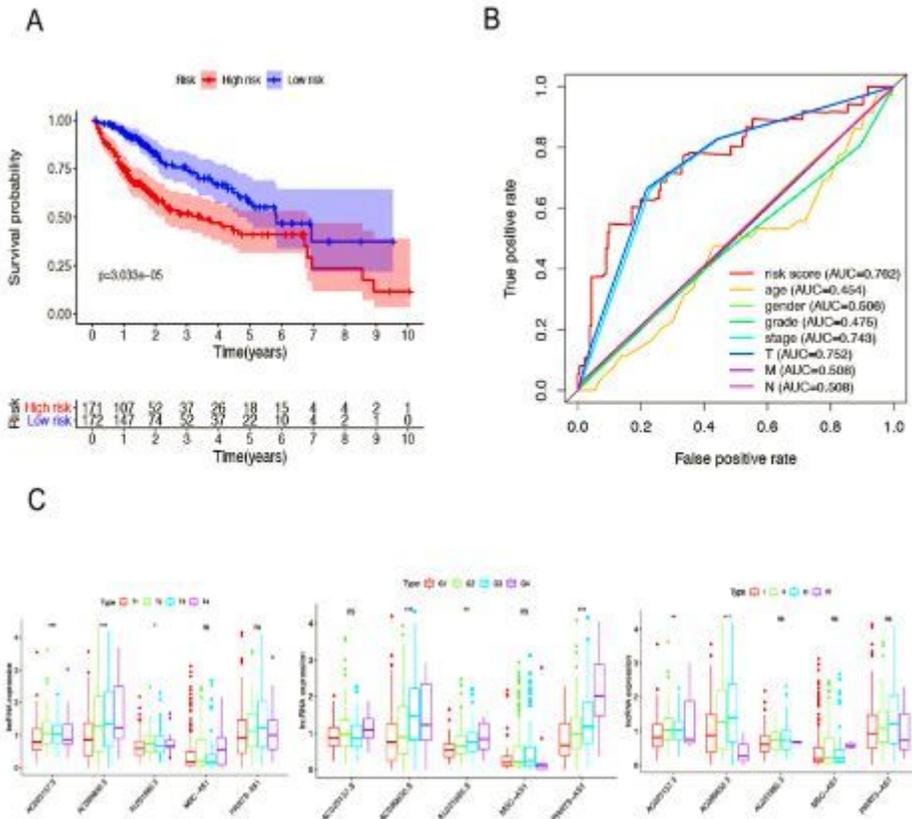
**Figure 3**

Acquisition of immune-related risk long non-coding RNAs (lncRNAs) and principal component analysis (PCA). (a) A forest map of the hazard ratio (HR) of immune-related lncRNAs. (b) PCA for all immune genes, immune lncRNAs, and risk lncRNAs; red dots represent high-risk genes, and green dots represent low-risk genes



**Figure 4**

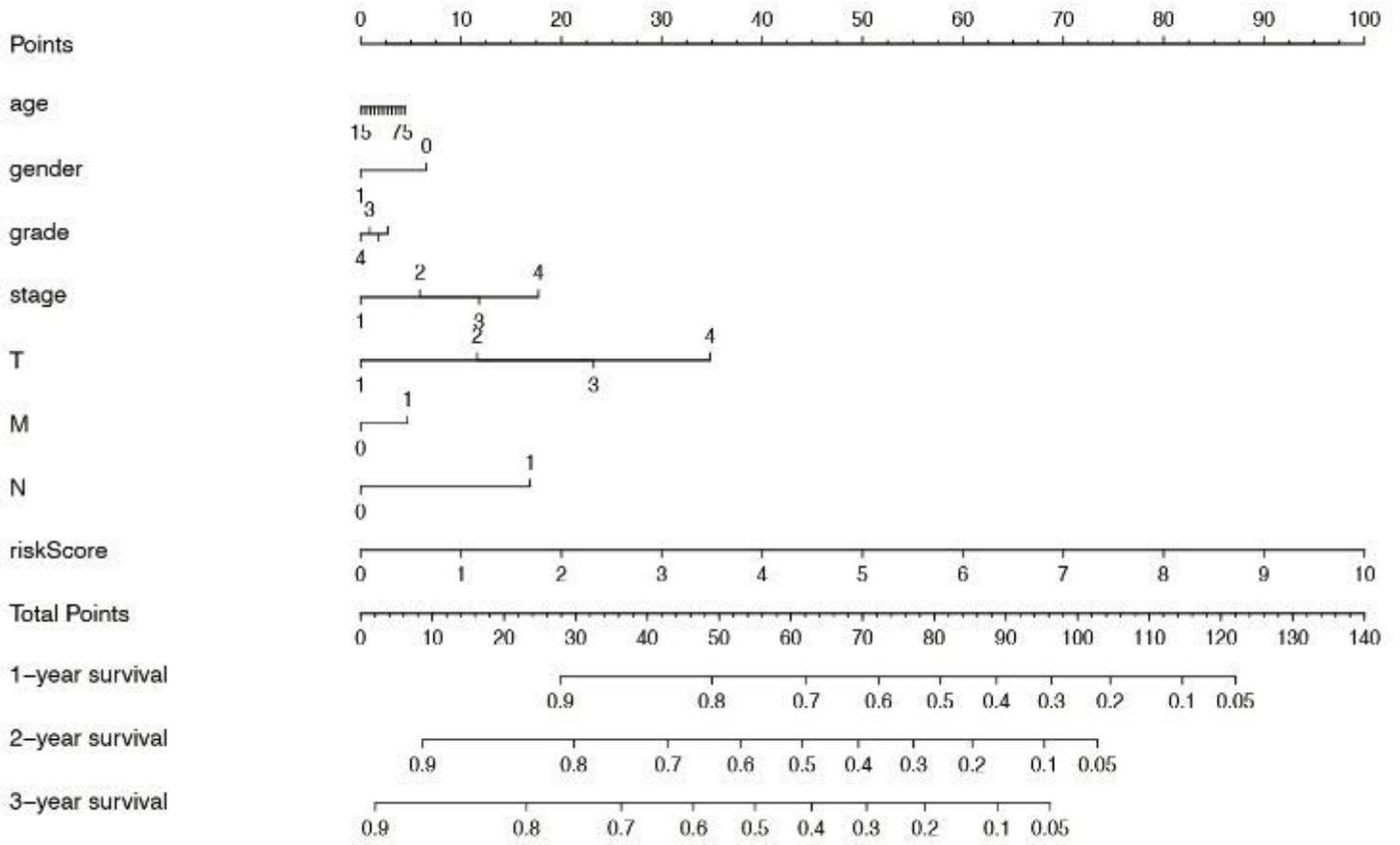
Characteristic of prognosis model: Long non-coding RNA (lncRNA) expression (a), risk score (b) and survival time. (c) Patterns for patients in high-group and low-risk group determined using the five-immune-related lncRNA signature



**Figure 5**

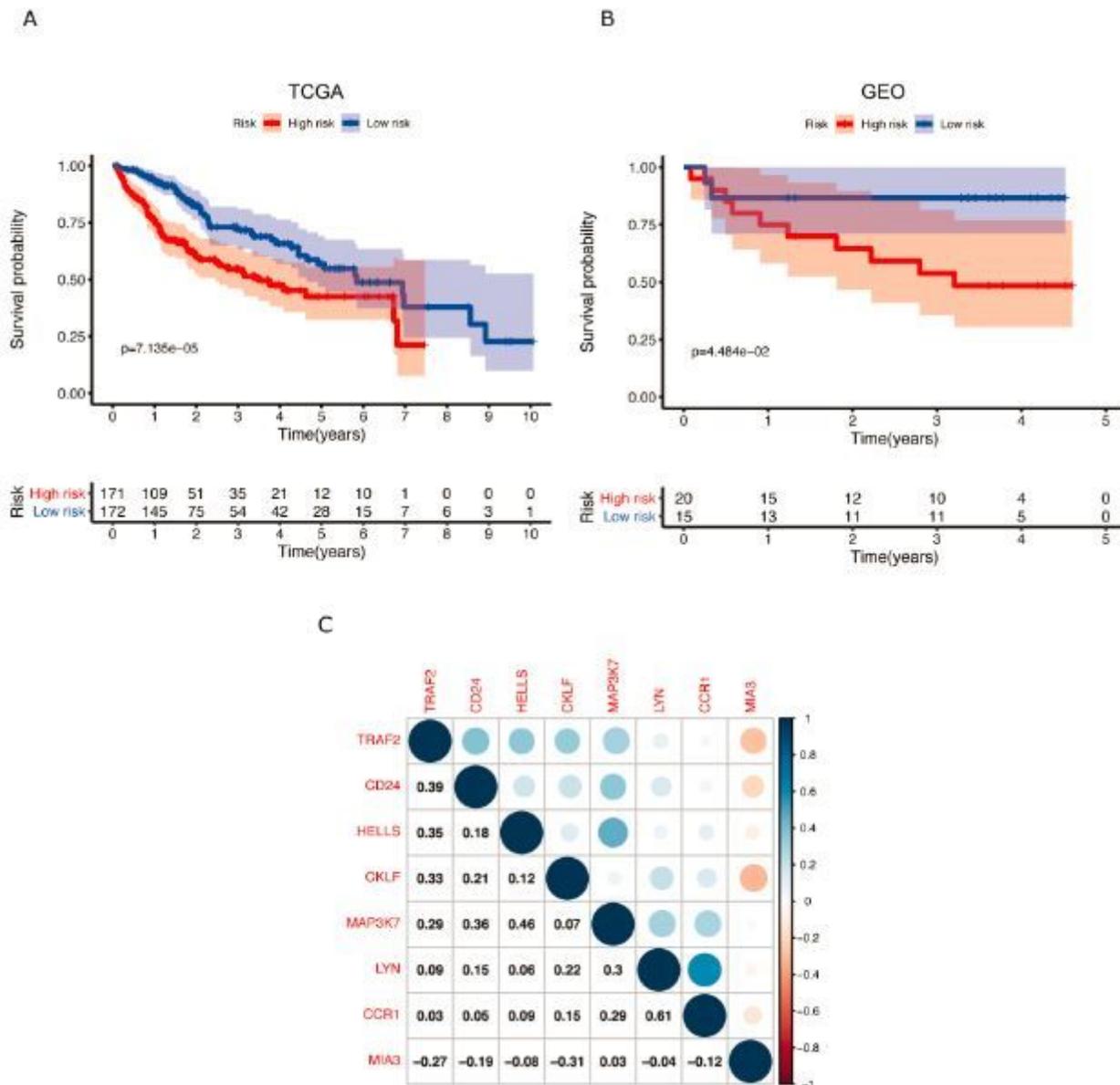
Clinicopathological characteristics of model: (a) Kaplan-Meier analysis of the overall survival of The Cancer Genome Atlas (TCGA) in high-risk ( $n = 171$ ) and low-risk subgroups ( $n = 173$ ). (b) Receiver operating characteristic curve (ROC) analysis of risk scores were used to predict the overall survival of the test set. (c) Association between clinical features (T (tumor), G (grade), and S (stage)) based on the five risk genes

A



**Figure 6**

Nomogram. (a) Prognostic nomogram predicting 1-, 2-, and 3-year overall survival of patients with hepatocellular carcinoma



**Figure 7**

Gene set enrichment analyses (GSEA). Significantly enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways by GSEA. (a) Ten representatives of KEGG pathways in high-risk and low-risk patients with hepatocellular carcinoma. (b) Ten representatives of Gene Ontology (GO) pathways in high-risk and low-risk patients with hepatocellular carcinoma

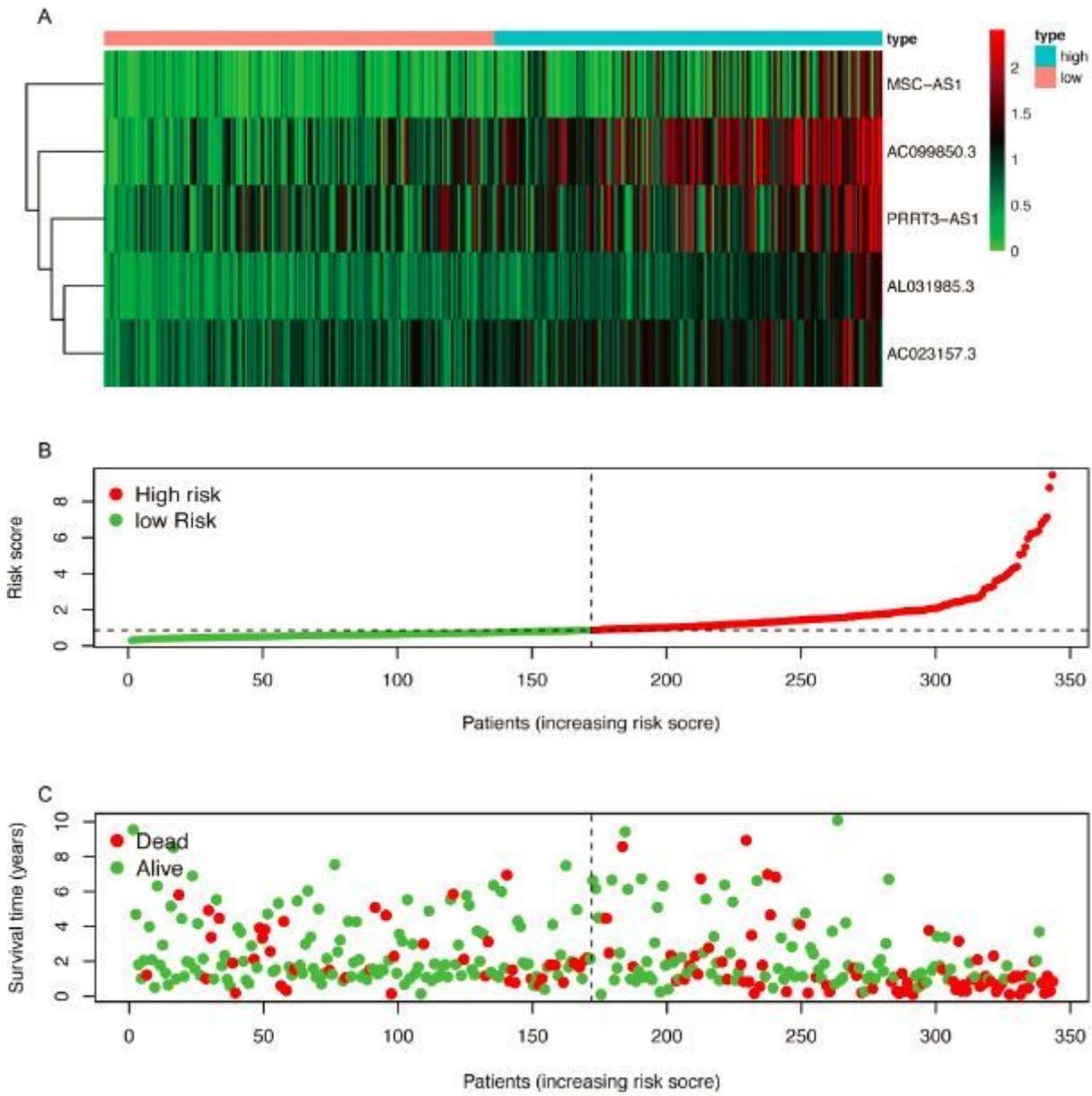


Figure 8

Figure 8