

## LETTER TO EDITOR

# A novel signature based on pairwise PD-1/PD-L1 signaling pathway genes for predicting the overall survival in patients with hepatocellular carcinoma

Dear Editor,

Hepatocellular carcinoma (HCC) is ranked as the most prevalent subgroup of liver malignancies in the world, representing about 90% of primary liver cancers.<sup>1</sup> To date, no widely accepted molecular biomarkers are available for survival stratification with HCC.<sup>2,3</sup> Recently, a novel algorithm was developed based on the relative orderings of mRNA expression patterns and had yielded excellent results.<sup>4,5</sup> Substantial breakthroughs have been founded in programmed death 1 (PD-1)/ programmed death 1 ligand (PD-L1) signaling pathway in various cancers.<sup>6</sup> Here, we used the TCGA cohort to develop a signature and other two databases to confirm the prognostic model based on pairwise PD-1/PD-L1 signaling pathway genes.

As shown in the flowchart (Figure 1A), the mRNA expression matrixes and their clinical characteristics were retrieved from the TCGA, GEO, and the International Cancer Genome Consortium (ICGC-JP cohort) databases. The clinicopathological information of the three databases was displayed in Table S1. Finally, 146 potential genes were acquired from the KEGG and Reactome pathway databases (Table S2). We filtered out the candidate genes computed by median absolute deviation (MAD) > 0.5. We created all possible pairs of genes based on the shared 34 genes. Among gene pairs, two genes (a and b) formed a gene pair in alphabetical order. If the expression values of  $gene_a > gene_b$  in a particular sample, the gene pair produced a score of 1; if the situation is opposite, the score equals 0. The association of these gene pairs was evaluated with univariate and LASSO Cox regression analysis after 1000 iterations (Figure 1B-C), and finally, 12 gene pairs, included 13 genes, were identified (Table S3). The risk score was computed via the expression patterns of these pairwise genes weighted by the respective coefficient obtained from LASSO algorithm. On the basis of time-dependent ROC curve, the ideal cutoff of the risk score was calculated at 0.573 to stratify individuals into the high- and

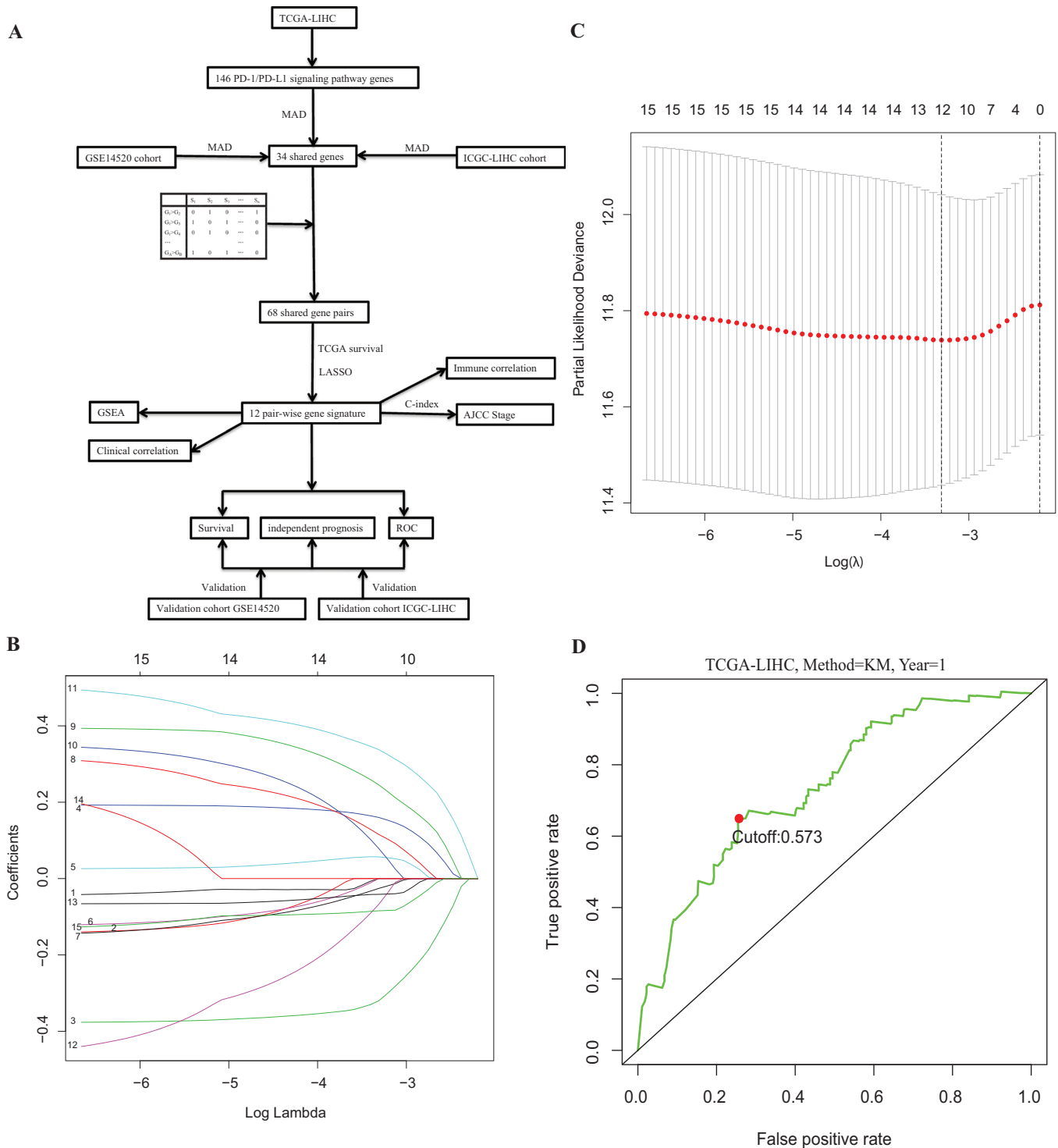
low-risk subgroups (Figure 1D). The corresponding signature risk score and individuals survival status between groups were illustrated in Figure 2A. The novel model could split individuals into low- and high-risk subgroups with different outcome (Figure 2B), and with the increased risk score, individuals have a greater risk of mortality ( $P = .00197$ ; Figure 2C). Additionally, the area under curve (AUC) of the new model for 1-, 3-, and 5-year outcome in TCGA group was 0.733, 0.716, and 0.722, respectively (Figure 2D). The performance of the model was successfully verified in the two external cohorts (Figure 2E-L). The risk score was clearly higher in individuals with increased AFP ( $\geq 300$  ng/mL), advanced Tumor, Node, Metastasis (TNM) staging (stage III-IV), advanced pathological grade (grades III-IV), patients without family history of cancer, and patients with vascular invasion (all  $P < .05$ ). However, there were no obvious differences in age, neoplasm status, gender, and history of prior cancer (Figure S1).

Multivariate Cox analysis demonstrated that the model risk score was an independent predictive indicator for OS (Figure 3A) after adjusting for multiple clinical and pathological parameters in TCGA portal. Moreover, the independent prognostic factor was verified in two validation cohorts (Figure 3B-C). A nomogram was therefore developed (Figure 3D). The C-index for the nomogram and TNM stage was 0.706 (95% CI: 0.648-0.764) and 0.551 (95% CI: 0.479-0.623), respectively. Calibration plots exhibited highly consistent with the actual observations (Figure 3E). According to decision curve analysis, the nomogram illustrated high clinical application potential and net benefits than the TNM stage examined (Figure 3F).

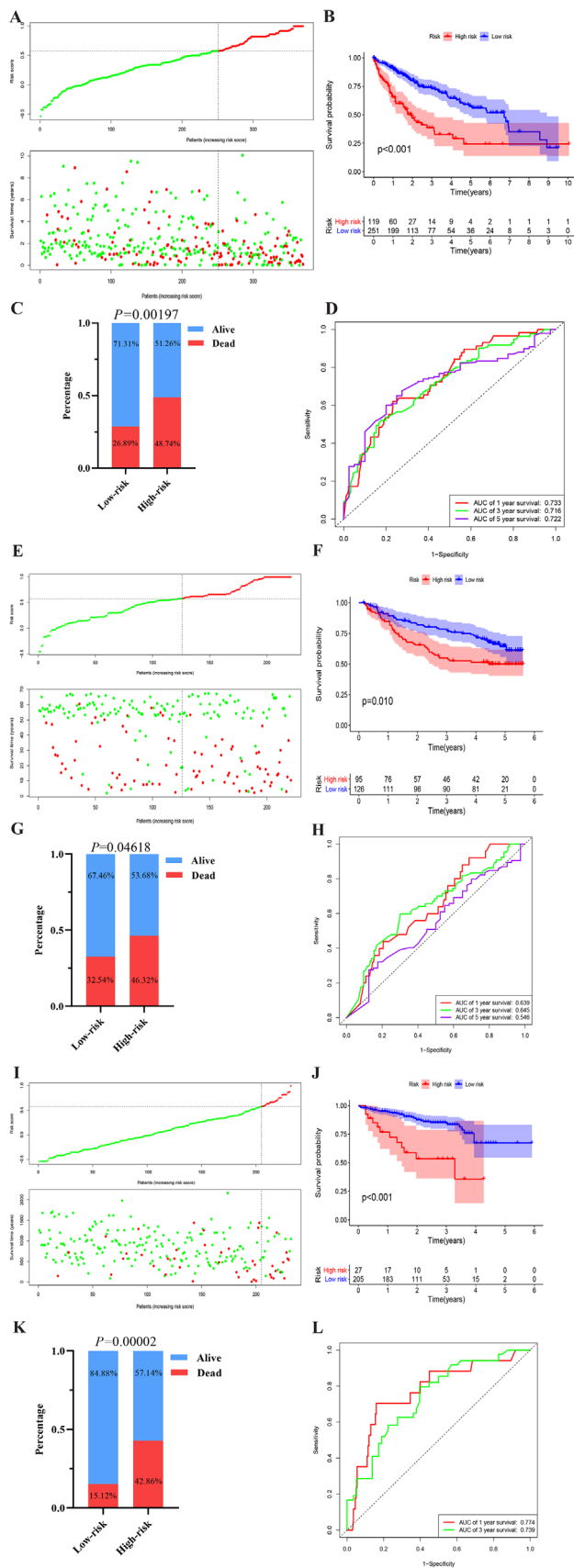
The outline of immune cell abundance between two subgroups measured by CIBERSORT is illustrated in Figure 4A. It was revealed that memory B cells, resting mast cells, were elevated, and CD8 T cells were reduced in high-risk subgroup (Figure 4B-D). The model risk score was positively associated with M0 macrophage, Tregs, neutrophil,

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Clinical and Translational Medicine* published by John Wiley & Sons Australia, Ltd on behalf of Shanghai Institute of Clinical Bioinformatics



**FIGURE 1** (A) A flow diagram depicting the forming and verifying of the 12-gene pair signature. (B) LASSO coefficient profiles of the 15 survival-related gene pairs; (C) choosing the tuning parameter ( $\lambda$ ) in the LASSO algorithm by 10-fold cross-validation; and (D) time-dependent ROC curve illustrated the cutoff of score was determined to stratify patients into low- or high-risk subgroups. Abbreviations: LASSO, least absolute shrinkage and selection operator; PD-1, programmed death 1; ROC, area under curve; TCGA, The Cancer Genome Atlas; PD-L1, programmed death 1 ligand

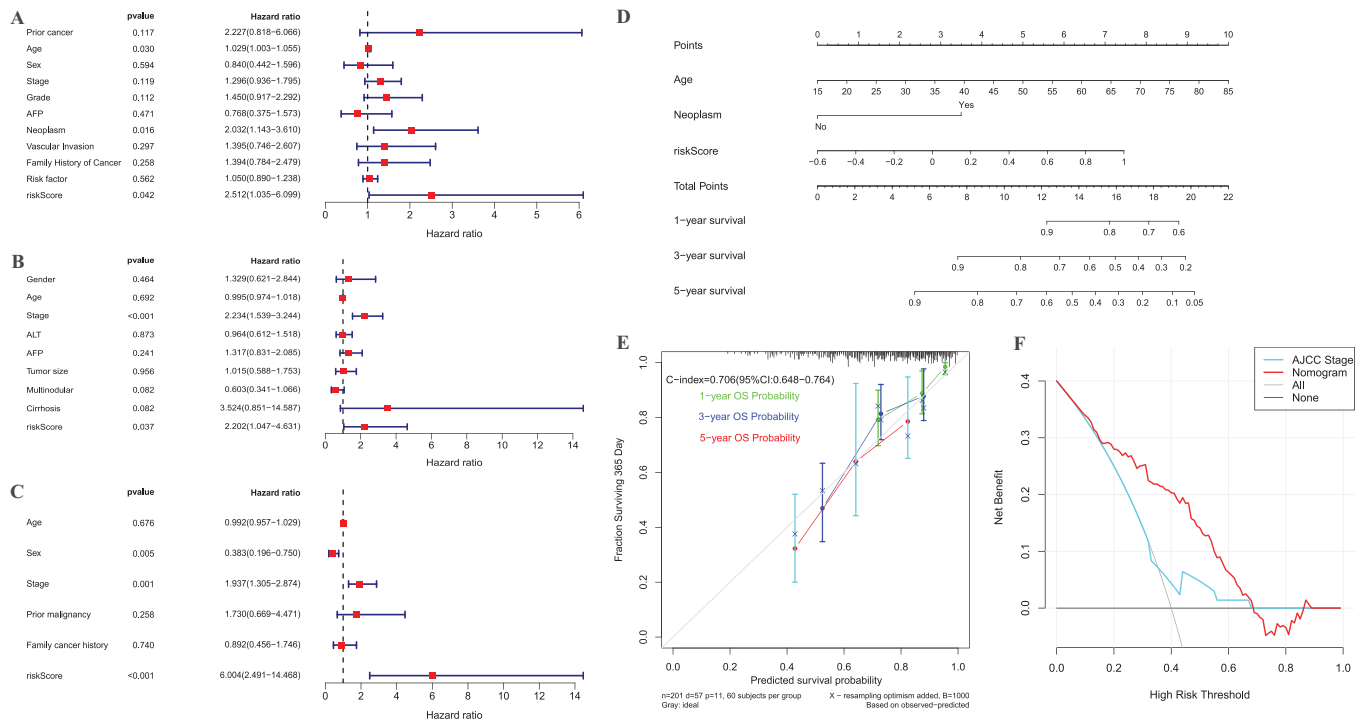


memory B cell, eosinophil, and macrophage (all coefficient  $> 0$  &  $P < .05$ ) and negatively associated with monocyte, M1 macrophage, CD8+ T cell, and activated mast cell (Figure 4E). The scores of immunophenoscore (IPS), PD1/PD-L1, CTLA4, and CTLA4 + PD1/PD-L1 inhibitors were applied to explore the potential clinical application of immune checkpoint inhibitors. As manifested in Figure 4F-I, the IPS, PD-1/PD-L1\_positive scores, CTLA4\_positive, and CTLA4\_positive plus PD-1/PD-L1\_positive scores were notably increased in the low-risk subgroup, indicating more immunogenicity on immune checkpoint inhibitors in the low-risk group.

Gene ontology (GO) analysis of the 13 signaling pathway genes illustrated that the signature was mainly enriched in immune response processes (Figure S2A). KEGG enrichment analysis demonstrated that the signature was mainly involved in PD-1/PD-L1 checkpoint pathway in cancer, Fc epsilon RI signaling pathway, and B-/T-cell receptor signaling pathways (Figure S2B). GSEA performed between two risk subgroups exhibited that multiple gene sets were markedly involved in high-risk group (Figure S2C) and several metabolism pathways were mainly involved in low-risk group (Figure S2D).

Traditional approaches that previous studies used were required scaling and appropriate normalization among various measurement platforms.<sup>2,7</sup> We designed a novel method for the identification of functional associations between pairwise genes based on relative expression orderings of two genes within-sample, without the need of standardization, batch correction and scaling among different platforms. Besides, the cutoff of the risk score can be

**FIGURE 2** (A) The 12 pairwise genes signature corresponding risk score distribution and individuals outcome; (B) K-M analysis curves of individuals' OS between two groups in the TCGA cohort; (C) comparison of risk of mortality between the two groups in the TCGA cohort; and (D) time-dependent ROC analyses of the model risk score in forecasting 1-, 3-, and 5-year OS in the TCGA dataset. (E) The 12 pairwise genes signature corresponding risk score distribution and individuals outcome; (F) K-M analysis curves of individuals' OS between two groups in the GSE14520 dataset; (G) comparison of risk of mortality between the two groups in GSE14520 dataset; (H) time-dependent ROC analyses of the model risk score in forecasting 1-, 3-, and 5-year OS in GSE14520 dataset; (I) the 12 pairwise genes signature corresponding risk score distribution and individuals outcome in the ICGC cohort; (J) K-M analysis curves of individuals' OS between two groups in the ICGC portal; (K) comparison of risk of mortality between the two groups in the ICGC portal; and (L) time-dependent ROC analyses of the model risk score in forecasting 1-, and 3-year survival in the ICGC portal. Abbreviations: ICGC, International Cancer Genome Consortium; K-M, Kaplan-Meier; OS, overall survival; ROC, area under curve; TCGA, The Cancer Genome Atlas



**FIGURE 3** Multivariate Cox regression of the correlation between clinicopathologic factors and outcome of individuals with HCC in the TCGA portal (A), GSE14520 dataset (B), and the ICGC portal (C). Nomogram forecasting 1-, 3-, and 5-year OS possibility for individuals with HCC in the TCGA portal (D); calibration curves for the nomogram (E); and DCA curves showing the comparison between the model and AJCC stage for predicting overall survival in HCC (F). Abbreviations: AJCC, American Joint Committee on Cancer; DCA, decision curve analysis; HCC, hepatocellular carcinoma; OS, overall survival; TCGA, The Cancer Genome Atlas

applied to other validation cohort in the future because it was computed as expression levels (0 or 1) of these pairwise genes weighted by the coefficients, presenting some significant advantages than previous signatures. Moreover, to identify reliable signature of HCC prognosis, we developed a signature for patients with HCC and confirmed its performance in other two external validation cohorts with large cancer samples for biomarker mining through multidimensional bioinformatics methods. The signature was accurate and robust, giving that the data resource was adequate and the analytical procedures were comprehensive. Therefore, all findings reveal that the new model can present a precise overall survival (OS) prediction of HCC patients and can measure HCC OS in an individualized way and can be readily translated into clinical utility according to the cutoff value determined from previous formula.

In conclusion, we proposed and validated a pairwise PD-1/PD-L1 gene panel that can accurately predict OS of patients with HCC. The signature can help to the effect of immunotherapy for HCC patients and be used as a promising marker to evaluate individuals who could benefit from immunotherapy.

## ETHICS APPROVAL AND CONSENT TO PARTICIPATE


The data used were publicly available to all, so the ethics approval was waived.

## DATA AVAILABILITY STATEMENT

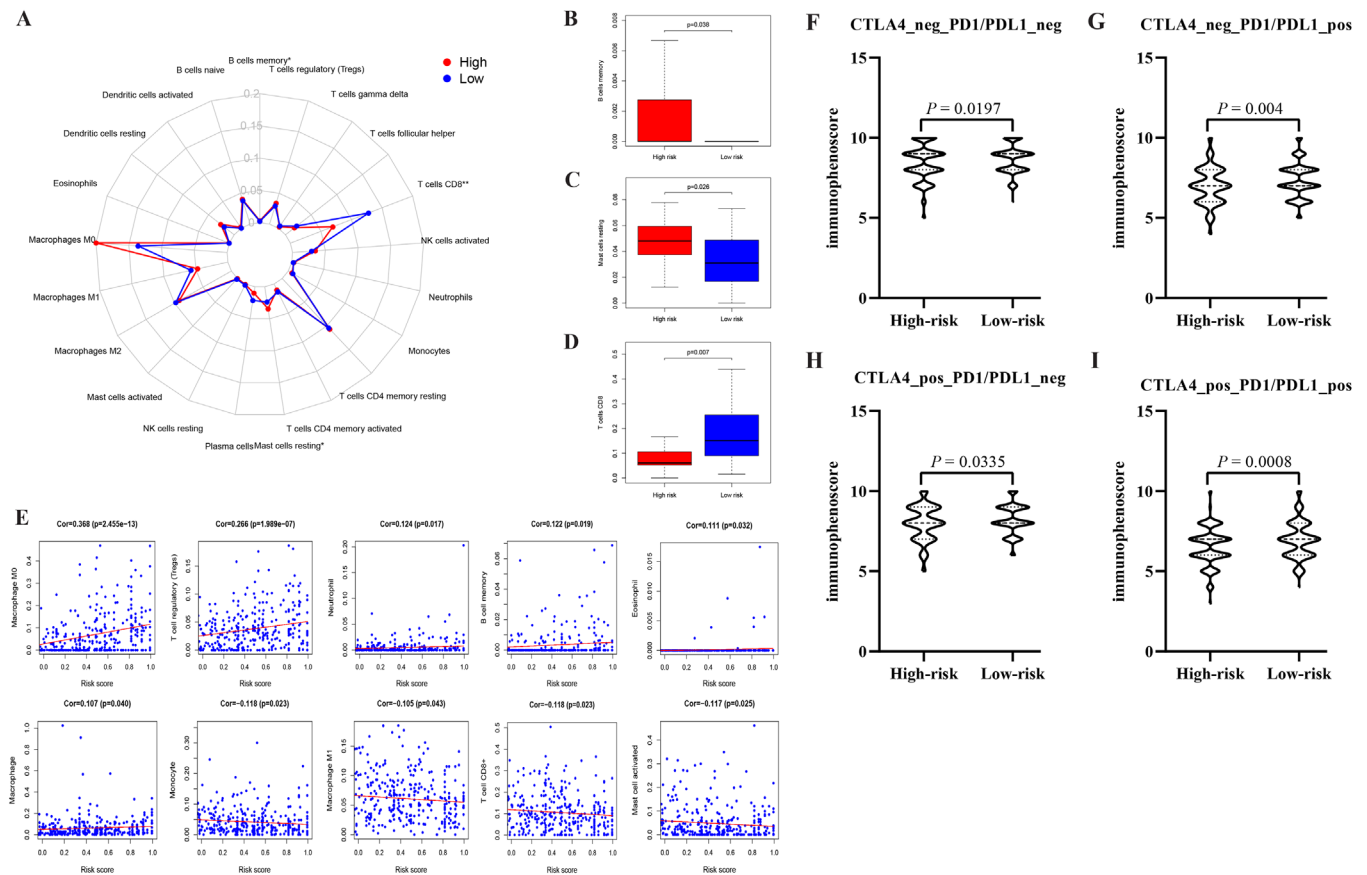
All datasets that support the present study are publicly available to all at GEO (<https://www.ncbi.nlm.nih.gov/geo/>) database, ICGC portal, and the TCGA portal.

## CONFLICT OF INTEREST

None

Enfa Zhao<sup>1</sup>   
Shimin Chen<sup>2</sup>  
Ying Dang<sup>3</sup>

<sup>1</sup> Department of Structural Heart Disease, the First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China



**FIGURE 4** Tumor infiltrating immune cells between two subgroups and the correlation with the model risk score. Twenty two types immune cell subtypes between two subgroups (A); the abundance distribution of memory B cells (B), resting mast cells (C), and CD8 T cells between high- and low-risk subgroups; (D) correlation determined by Person correlation analysis between the signature risk score and infiltrating immune cells in patients with HCC ( $P < .05$ ). IPS comparison between two subgroups in patients with HCC in CTLA4\_neg\_PD1/PDL1\_neg (E), CTLA4\_negative + PD-1/PD-L1\_positive (F), CTLA4\_positive + PD-1/PD-L1\_negative (G), and CTLA4\_positive + PD-1/PD-L1\_positive (H); PD-1/PD-L1\_pos or CTLA4\_pos indicated anti- PD-1/PD-L1 or anti-CTLA4 therapy, respectively. Abbreviations: CTLA4, cytotoxic T-lymphocyte associated protein 4; IPS, immunophenoscore; PD-1, programmed death 1; PD-L1, programmed death 1 ligand

<sup>2</sup> Department of Gastroenterology, Traditional Chinese Medical Hospital of Taihe Country, Taihe, China

<sup>3</sup> Department of Ultrasound Medicine, the First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China

### Correspondence

Ying Dang, MD, Department of Ultrasound Medicine, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi 710061, China  
Email: [986dangying@163.com](mailto:986dangying@163.com)

Enfa Zhao and Shimin Chen contributed equally to this work and should be considered cofirst authors.

### ORCID

Enfa Zhao  <https://orcid.org/0000-0002-8028-2997>

### REFERENCES

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. 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.
- Dai Y, Qiang W, Lin K, Gui Y, Lan X, Wang D. An immune-related gene signature for predicting survival and immunotherapy efficacy in hepatocellular carcinoma. *Cancer Immunol Immunother CII.* 2021;70:967-979.
- Yan Y, Lu Y, Mao K, et al. Identification and validation of a prognostic four-genes signature for hepatocellular carcinoma: integrated ceRNA network analysis. *Hepatol Int.* 2019;13(5):618-630.
- Li B, Cui Y, Diehn M, Li R. Development and validation of an individualized immune prognostic signature in early-stage nonsquamous non-small cell lung cancer. *JAMA Oncol.* 2017;3(11):1529-1537.
- Zhao E, Zhou C, Chen S. A signature of 14 immune-related gene pairs predicts overall survival in gastric cancer. *Clin Transl Oncol.* 2021;23(2):265-274.

6. Tian M, Yang J, Han J, He J, Liao W. A novel immune checkpoint-related seven-gene signature for predicting prognosis and immunotherapy response in melanoma. *Int Immunopharmacol.* 2020;87:106821.
7. Chen Q, Li F, Gao Y, Xu G, Liang L, Xu J. Identification of energy metabolism genes for the prediction of survival in hepatocellular carcinoma. *Front Oncol.* 2020;10:1210.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.