LETTER TO THE EDITOR

Implications of pseudogenes for the prognosis of hepatocellular carcinoma

Dear Editor

Growing evidence has demonstrated that pseudogenes are intricately associated with tumour progression in hepatocellular carcinoma (HCC).[1] Nonetheless, pseudogenes related to tumour immune infiltration and their value in improving clinical outcomes remain largely
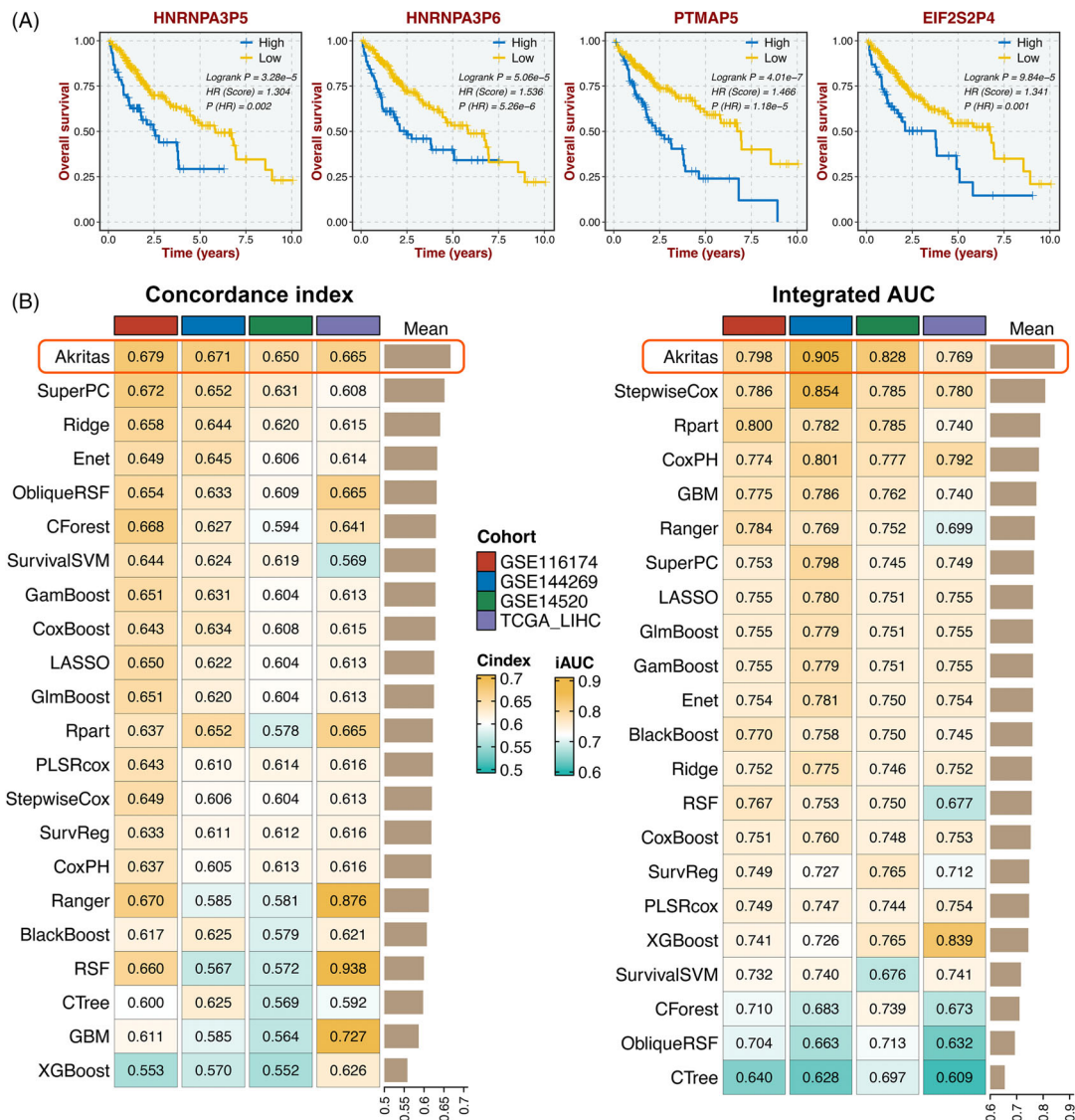

FIGURE 1 Integrative construction of a consensus signature. (A) Kaplan–Meier analysis of four immune-related intrinsic pseudogenes in TCGA-LIHC. (B) C-indices (left) and iAUC (right) of 22 survival machine-learning models in GSE116174, GSE144269, GSE14520, and TCGA-LIHC
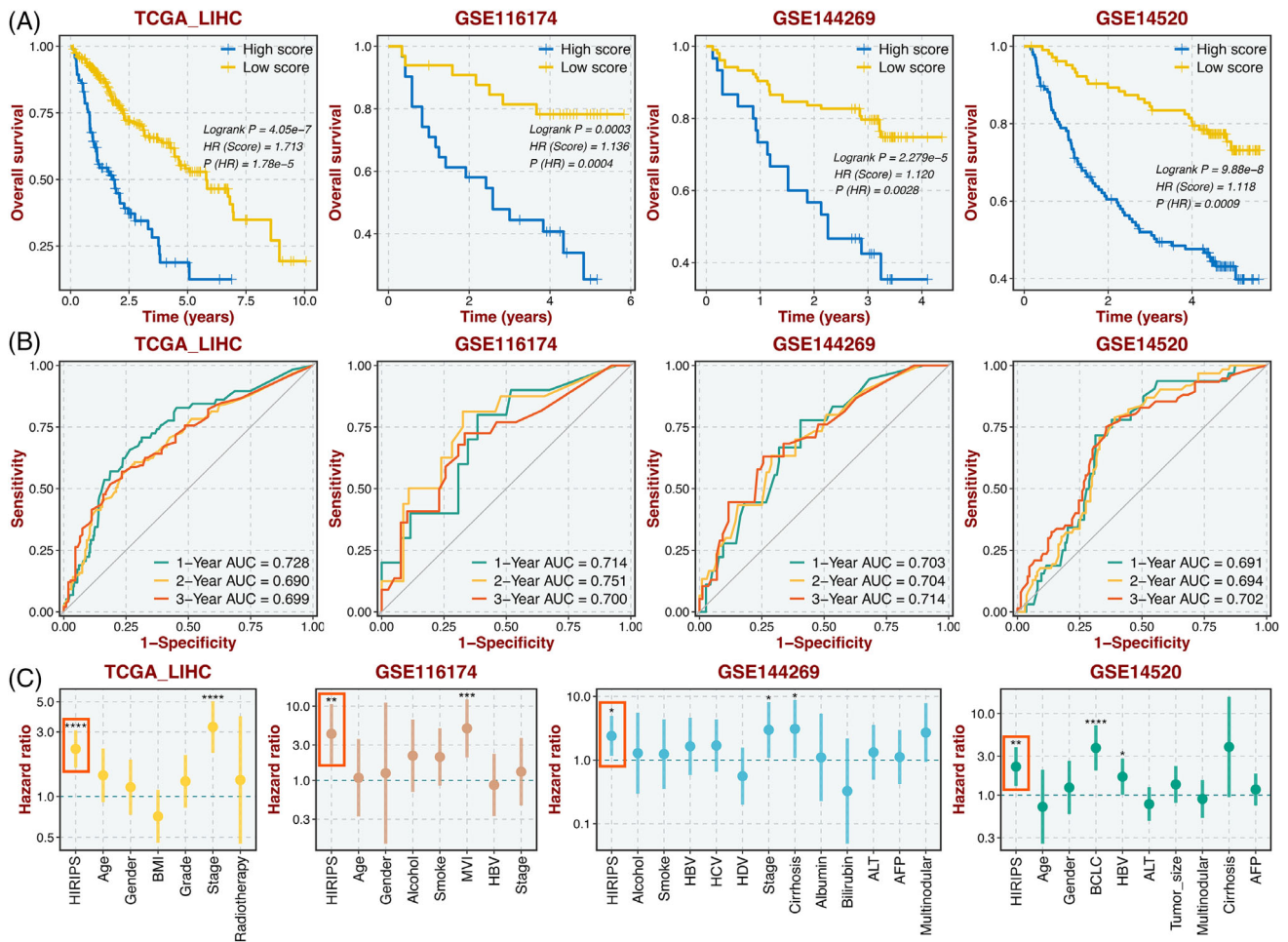
FIGURE 2 Evaluation of the performance of the HIRIPS model. (A) Kaplan–Meier curves of OS between high and low HIRIPS groups in TCGA-LIHC, GSE116174, GSE144269, and GSE14520. (B) Time-dependent ROC analysis for 1, 2 and 3 years in four datasets. (C) Multivariate Cox regression analysis of OS based on HIRIPS and other clinical traits in four datasets. *$p < .05$; **$p < .01$; ***$p < .001$; ****$p < .0001$

unexplored. Pseudogenes, newly discovered non-coding homologs of protein-coding genes,[1] have been regarded as non-functional evolutionary relics. Nevertheless, increasing evidence has demonstrated that pseudogenes play vital roles in tumourigenesis as regulators of coding genes. For example, *HMGA1P6*, a pseudogene transcriptionally activated by oncogene *MYC*, could contribute to oncogenesis in ovarian cancer.[2] Besides, pseudogenes also have profound impacts on anti-tumour responses via involvement in regulating tumour-immune interactions. *BRCA1* Pseudogene 1 (*BRCA1P1*) has been reported to weaken immune response by suppressing innate immune defense mechanisms.[3] Through glioma-derived exosomes, *TMEM198B* promoted macrophage lipid accumulation and increased fatty acid oxidation, further inducing macrophages to M2 polarization.[4] However, the clinical significance of immune-related pseudogenes remains largely unexplored in HCC.

Here, we enrolled four public datasets with abundant pseudogene expression profiles and complete prognos-

tic information, including: The Cancer Genome Atlas Liver Hepatocellular Carcinoma (TCGA-LIHC) ($n = 340$), GSE116174 ($n = 64$), GSE144269 ($n = 67$) and GSE14520 ($n = 242$) (Table S1). To systematically evaluate candidate intrinsic pseudogene modulators of immune cells, we introduced a novel framework to identify immune-related housekeeping pseudogenes (Supplementary Materials). In TCGA-LIHC, the relative abundance of 28 immune cells was first deciphered using the ssGSEA algorithm.[5] Considering that tumour purity may obscure links in tumour microenvironment,[6] we thus calculated the first-order partial correlation coefficient (PCC) between pseudogenes and immune cells by removing the effect of tumour purity.[6,7] Pseudogenes with the top 5% PCC were extracted as candidate immune-related pseudogenes for each immune cell. A hypothesis is that if a specific pseudogene has strong correlations with all immune cell types, it may execute a housekeeping role in HCC immune microenvironment,[8] which is also defined as HCC immune-related intrinsic pseudogene (HIRIP) in
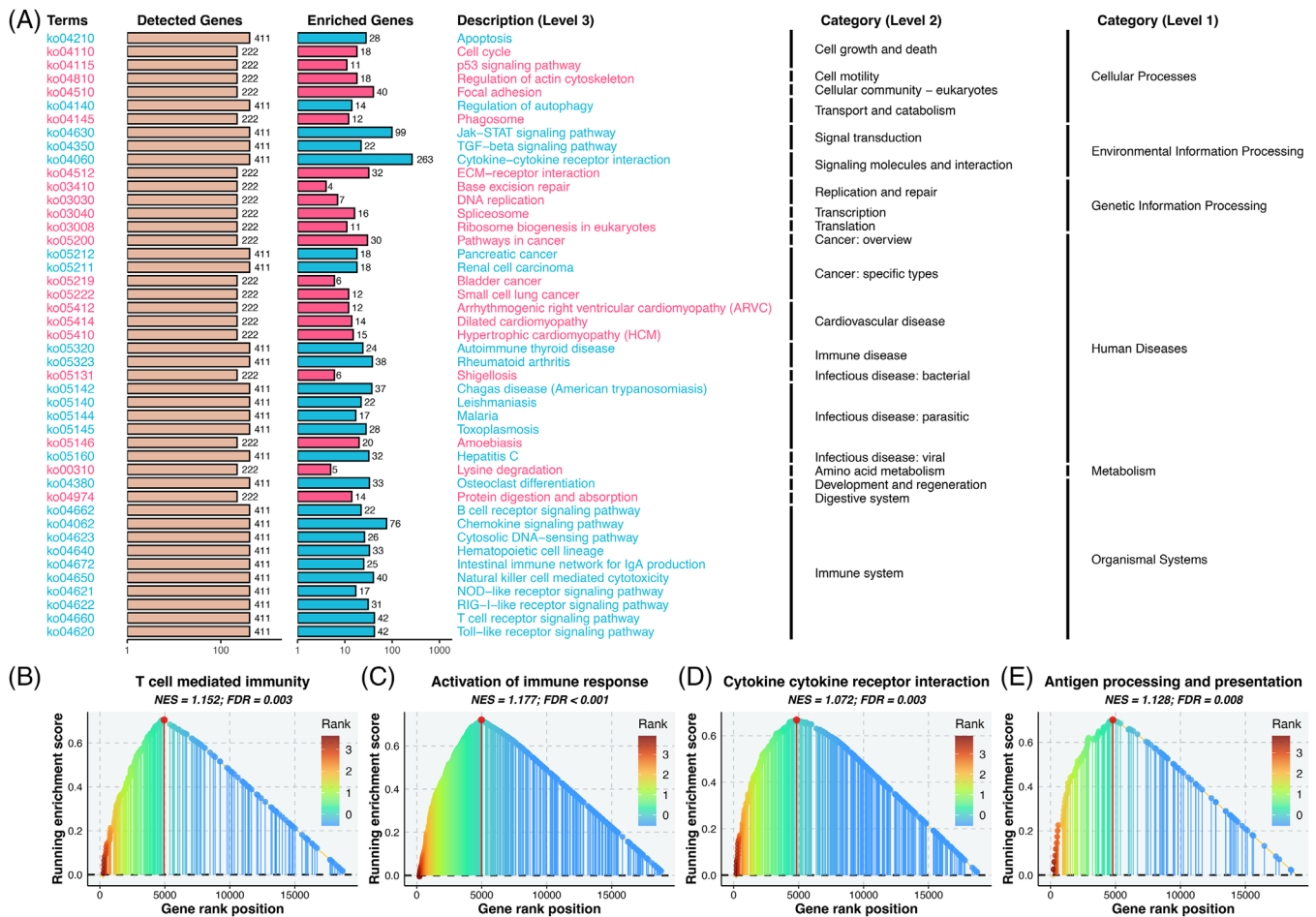
FIGURE 3 Underlying biological pathways and distinct microenvironment patterns related to HIRIPS. (A) KEGG pathways that were significantly enriched in high HIRIPS are shown in red; those significantly enriched in low HIRIPS are shown in blue. (B–E) Gene set enrichment analysis (GSEA) of immune-related terms for HIRIPS, including "T cell-mediated immunity" (B), "activation of immune response" (C), "cytokine cytokine-receptor interaction" (D) and "antigen processing and presentation" (E)

this study. Here, we calculated the tissue specificity index (TSI)[9] to identify pseudogenes generally correlated with different immune cell types (Table S2). As previously reported,[9] pseudogenes with a lower TSI score were strongly associated with all immune cell types, suggesting their vital biological functions in immunity. According to the criteria described in previous studies,[8,9] we settled the threshold of TSI <.2 and identified 23 HIRIP essential for immune regulation (Figure S1A).

Subsequently, the 23 HIRIP were further filtered to develop an integrative HIRIP signature (HIRIPS). Initially, univariate Cox regression analysis determined four pseudogenes with prognostic potential, including *HNRNPA3P5*, *HNRNPA3P6*, *PTMAP5* and *EIF2S2P4* (Figure S1B). Kaplan–Meier analysis demonstrated that high expression of these four pseudogenes suggested an unfavorable prognosis in HCC (Figure 1A). Afterward, to identify an optimal machine-learning algorithm for assessing prognosis, we developed 22 types of survival

machine-learning models (Supplementary Materials) based on the four pseudogene expression profiles of TCGA-LIHC. As mentioned above, A final HIRIPS fitted by the Akritas algorithm with the highest average mean C-index (.667) and integrated areas under the curve (iAUC) (.825) in the TCGA training cohort and three validation cohorts (GSE116174, GSE144269 and GSE14520) was considered the optimal one (Figure 1B).

To validate the robust performance of HIRIPS, Kaplan-Meier and Cox regression survival analyses were applied. Notably, patients with high HIRIPS possessed adverse prognosis in TCGA-LIHC (HR = 1.713; log-rank test: $P < .001$), and similar tracks were observed in three validation datasets, GSE116174 (HR = 1.136; log-rank test: $P < .001$), GSE144269 (HR = 1.120; log-rank test: $P < .001$), and GSE14520 (HR = 1.118; log-rank test: $P < .001$) (Figure 2A). Subsequently, time-dependent receiver operating characteristic (ROC) curves for OS (1-, 2-, and 3-year) exhibited excellent discrimination of

HIRIPS (Figure 2B) in TCGA-LIHC (AUC values: .728, .690, and .699), GSE116174 (AUC values: .714, .751, and .700), GSE144269 (AUC values: .703, .704, and .714), and GSE14520 (AUC values: .691, .694, and .702). Multivariate Cox regression analysis was performed and demonstrated HIRIPS remained statistical significance after adjusting confounding factors across all cohorts (Figure 2C). Overall, HIRIPS presented a clear superior performance in predicting prognosis.

To further explore the biological mechanisms underlying the HIRIPS, we proposed a novel pipeline to perform functional enrichment, which maximized information retention and comprehensively considered gene ordering across all cohorts (Supplementary Materials). The over-representing analysis revealed patients with high HIRIPS exhibited predominant enrichment in pathways related to malignant progressions, such as cell cycle and p53 signaling (Figure 3A and Figure S2). Whereas patients with low HIRIPS were particularly evident in immune-related pathways, for example, cytokine activity and T cell receptor pathways (Figure 3A and Figure S2). Subsequently, another bioinformatics algorithm, GSEA, also confirmed that low HIRIPS was significantly linked to immune response (Figure 3B–E). Overall, remarkable differences regarding biological function were identified between the high and low HIRIPS, which may account for the discrepancy in prognosis. Patients with high HIRIPS were featured by cellular proliferation and lower immune status, in line with their unfavorable prognosis. Low HIRIPS was predominantly distinguished by high immune activity and abundant immune cell infiltration, which indicated patients with low HIRIPS harbored more reserves of immunization resources for immunotherapy.

In conclusion, we comprehensively investigated pseudogenes associated with HCC-infiltrating immune cells and systematically identified an optimal pseudogene signature (termed HIRIPS) from 22 survival machine-learning algorithms. This signature displayed a robust and stable performance for predicting prognosis and might also serve as a latent biomarker for assessing immunotherapy response. Overall, our study provided a promising platform for optimizing precise treatment and improving clinical outcomes in HCC.

CONFLICT OF INTEREST
The authors declare that they have no competing interests.

Zaoqu Liu[1,2,3]
Yuyuan Zhang[1]
Siyuan Weng[1]
Hui Xu[1]
Xinwei Han[1,2,3]

[1]Department of Interventional Radiology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China
[2]Interventional Institute of Zhengzhou University, Zhengzhou, China
[3]Interventional Treatment and Clinical Research Center of Henan Province, Zhengzhou, China

Correspondence
Xinwei Han, Department of Interventional Radiology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan 450052, China.
Email: fcchanxw@zzu.edu.cn

Zaoqu Liu and Yuyuan Zhang have contributed equally to this work and share the first authorship.

ORCID
Zaoqu Liu https://orcid.org/0000-0002-0452-742X
Xinwei Han https://orcid.org/0000-0003-4407-4864

REFERENCES
1. Chen X, Wan L, Wang W, Xi WJ, Yang AG, Wang T. Re-recognition of pseudogenes: from molecular to clinical applications. *Theranostics*. 2020;10(4):1479-1499. doi:10.7150/thno.40659
2. Tian X, Song J, Zhang X, et al. MYC-regulated pseudogene HMGA1P6 promotes ovarian cancer malignancy via augmenting the oncogenic HMGA1/2. *Cell Death Dis*. 2020;11(3):167. doi:10.1038/s41419-020-2356-9
3. Han YJ, Zhang J, Lee JH, et al. The BRCA1 pseudogene negatively regulates antitumor responses through inhibition of innate immune defense mechanisms. *Cancer Res*. 2021;81(6):1540-1551. doi:10.1158/0008-5472.Can-20-1959
4. Zhan Y, Qiao W, Yi B, et al. Dual role of pseudogene TMEM198B in promoting lipid metabolism and immune escape of glioma cells. *Oncogene*. 2022;41(40):4512-4523. doi:10.1038/s41388-022-02445-0
5. Charoentong P, Finotello F, Angelova M, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep*. 2017;18(1):248-262. doi:10.1016/j.celrep.2016.12.019
6. Li Y, Jiang T, Zhou W, et al. Pan-cancer characterization of immune-related lncRNAs identifies potential oncogenic biomarkers. *Nat Commun*. 2020;11(1):1000. doi:10.1038/s41467-020-14802-2
7. Liu Z, Liu L, Weng S, et al. Machine learning-based integration develops an immune-derived lncRNA signature for improving outcomes in colorectal cancer. *Nat Commun*. 2022;13(1):816. doi:10.1038/s41467-022-28421-6
8. Sun J, Zhang Z, Bao S, et al. Identification of tumor immune infiltration-associated lncRNAs for improving prognosis and immunotherapy response of patients with non-small cell lung cancer. *J Immunother Cancer*. 2020;8(1):e000110. doi:10.1136/jitc-2019-000110

9. Yanai I, Benjamin H, Shmoish M, et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics*. 2005;21(5):650-659. doi:10.1093/bioinformatics/bti042

SUPPORTING INFORMATION

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