



OPEN Programmed cell death-related genes define distinct molecular subtypes and risk profiles in hepatocellular carcinoma

Han Yang^{1,3}, Qi Liu^{1,2,3}, Shengli Cao¹, Hongbo Fang¹, Jinjin Li¹, Xiaoyi Shi¹, Chun Pang¹, Danyang Lu², Xiaofang Zhao², Jie Li¹✉, Senyan Wang^{1,2}✉ & Tianchun Wu¹✉

Hepatocellular carcinoma (HCC) is a biologically and clinically heterogeneous malignancy, whose initiation and progression are increasingly recognized to be driven by the aberrant regulation of programmed cell death (PCD) pathways. To elucidate this association, we systematically integrated gene signatures from 21 distinct PCD types to characterize their expression patterns in HCC and construct a prognostic model for survival and therapeutic response prediction. Based on the TCGA-LIHC, GSE14520, and GSE116174 datasets, 85 candidate genes were identified through differential expression analysis and random survival forest algorithms. A 10-gene PCD-based risk score model was developed using machine learning including key genes such as KIF20A (associated with ferroptosis) and SLC2A1 (associated with anoikis), which demonstrated robust prognostic performance across three independent cohorts by stratifying patients into high- and low-risk groups. The risk score significantly correlated with immune infiltration, immune evasion potential, and predicted sensitivity to multiple anticancer agents. Consensus clustering based on model gene expression revealed two molecular subtypes with distinct survival outcomes and immune characteristics. A nomogram integrating the risk score exhibited favorable calibration and clinical applicability. Collectively, these findings propose a novel PCD-based molecular framework for prognosis assessment and personalized therapy in HCC.

Keywords Hepatocellular carcinoma, Programmed cell death, Machine learning, Drug sensitivity prediction

Abbreviations

HCC	Hepatocellular carcinoma
PCD	Programmed cell death
DEGs	Differentially expressed genes
TCGA	The cancer genome atlas
GEO	Gene expression omnibus
GO	Gene ontology
KEGG	Kyoto encyclopedia of genes and genomes
ROC	Receiver operating characteristic
AUC	Area under the curve
GSVA	Gene set variation analysis
TIDE	Tumor immune dysfunction and exclusion
ICI	Immune checkpoint inhibitor

Hepatocellular carcinoma (HCC) is the dominant histological subtype of primary liver cancer on a global scale, and its rising incidence has emerged as a significant public health concern^{1,2}. The multifactorial nature of HCC often stems from chronic liver insults, including infections with HBV/HCV, alcohol overuse, and prolonged aflatoxin B1 exposure. In addition, obesity and metabolic syndrome, as metabolic disorders, accelerate the progression of nonalcoholic fatty liver disease (NAFLD) and its advanced form, nonalcoholic steatohepatitis

¹Department of Hepatobiliary and Pancreatic Surgery, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, China. ²Translational Medicine Centre, The First Affiliated Hospital of Zhengzhou University, Zhengzhou University, Zhengzhou 450052, China. ³These authors contributed to the work equally and should be regarded as co-first authors: Han Yang and Qi Liu. ✉email: doctorlj99@hotmail.com; senyanwang@163.com; 11618326@zju.edu.cn

(NASH), thereby promoting liver cancer^{3,4}. Despite advances in clinical therapies, HCC continues to exhibit high recurrence and mortality rates, representing a major clinical challenge⁵. Given the biological complexity and clinical unpredictability of HCC, creating reliable prognostic models is essential for directing individualized therapeutic decisions.

Programmed cell death (PCD) is essential for tissue development and cellular quality control, playing a crucial role in maintaining physiological balance⁶. In the context of cancer biology, PCD serves as both a defense mechanism to eliminate potentially malignant cells and a barrier often circumvented by tumor cells to promote survival and progression⁷. Traditional forms of PCD include apoptosis, autophagy, and necrosis⁸. However, the PCD spectrum has expanded to include novel forms such as cuproptosis, oxeiptosis, and disulfidptosis, which are increasingly recognized for their relevance to the tumor microenvironment^{9,10}. Dysregulation of PCD pathways is closely linked to tumor proliferation, invasion, metastasis, and therapy resistance. These aberrant processes exert multifaceted effects on both cancer cells and their surrounding microenvironment¹¹.

This research systematically compiled gene signatures for 21 PCD modalities and analyzed their expression patterns in HCC. Our objective is to develop a prognostic model capable of predicting patient outcomes and therapeutic responses. The expression patterns of these genes were analyzed in HCC and normal liver tissues to evaluate their potential involvement in tumorigenesis. Furthermore, we explored the prognostic significance of these genes and identified robust biomarkers for precise diagnosis and personalized treatment. We intend to investigate the relationship between PCD-associated genes and the tumor immune microenvironment to identify potential targets for immunotherapy in HCC.

Materials and methods

Acquisition and preprocessing of public datasets

Clinical annotations and transcriptomic profiles for 377 HCC cases were obtained from the TCGA-LIHC cohort through the Genomic Data Commons portal (<https://portal.gdc.cancer.gov>). In addition, two independent gene expression datasets, GSE14520 and GSE116174, were downloaded from the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>). The prediction of drug response utilized data from the Genomics of Drug Sensitivity in Cancer (GDSC) database (<https://www.cancerrxgene.org/>).

The PCD gene set was primarily derived from two key references^{12,13}, along with other reliable sources, including GSEA gene sets, KEGG pathways, review articles, and manual curation. After removing duplicates, a total of 2,701 PCD-related genes were included for subsequent analysis (Supplementary Table S1).

Selection of candidate PCD-related genes

Based on the TCGA-LIHC dataset, differential expression analysis was performed using four commonly used algorithms: “DESeq2”, “limma”, “edgeR”, and “Wilcoxon” (Supplementary Table S2). Genes with $|\log_2 \text{fold change} (\log_2 \text{FC})| \geq 1$ and adjusted $P < 0.05$ were considered differentially expressed, and the intersection of results from all methods was taken. The differentially expressed genes (DEGs) were compared with the curated PCD gene set to find those linked to programmed cell death.

A random survival forest (RSF) algorithm was then applied to each PCD subtype to select the top five genes with the highest importance scores. After duplicates were removed, 85 candidate genes were identified. The chromosomal positions of these genes were annotated through the ‘Rcircos’ R package. Protein-protein interaction data for the candidate genes was collected from the STRING database (<https://string-db.org/>).

Establishment and assessment of the PCD-based prognostic signature

Machine learning analysis was performed using ten algorithms: “randomForestSRC”, “glmnet”, “gbm”, “CoxBoost”, “survivalsvm”, “BART”, “compareC”, “ENET”, “Lasso”, and “Ridge”. A total of 117 algorithm combinations were evaluated, and the combination of “Stepwise Cox regression (backward) + RSF” was selected as the optimal strategy for prognostic gene identification based on its predictive performance. The RSF model constructed based on Stepwise Cox regression effectively predicted the relative mortality risk for individual patients. For each sample, a risk score was derived from the cumulative predicted mortality estimated by the RSF model, and patients in the GSE14520, GSE116174, and TCGA-LIHC cohorts were subsequently stratified into high- and low-risk groups according to the median risk score. Overall survival (OS) was assessed with the Kaplan-Meier method from the ‘survival’ R package. ROC curves and AUC metrics at 1, 3, and 5 years were generated using the ‘pROC’ package to evaluate prognostic accuracy over time.

Enrichment analysis and somatic mutation analysis

Gene enrichment analysis was performed using the “clusterProfiler” package in R, based on the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases¹⁴. A significance threshold of $P < 0.05$ was applied. Gene set variation analysis (GSVA) was conducted to assess pathway activity differences between high- and low-risk groups¹⁵. Pathway enrichment analysis was executed using the ‘GSVA’ package in R. The TCGA-LIHC cohort’s somatic mutation profiles, including point mutations (single nucleotide variations, SNVs) and structural changes (copy number variations, CNVs), were examined using an online platform (<http://www.sxdyc.com/index>). CNVs were defined as “GAIN” if the copy number alteration value was greater than 0.2 and as “LOSS” if the value was less than -0.2 .

Consensus clustering analysis in HCC

Using the ‘ConsensusClusterPlus’ R package, unsupervised consensus clustering was applied to the expression profiles of 10 signature genes to identify molecular subtypes within the TCGA-LIHC cohort. This approach set the maximum cluster number to 10, applying 80% resampling across 100 iterations, with parameters ClusterAlg = “pam” and distance = “Minkowski”.

Construction of a nomogram

A prognostic nomogram was designed using the TCGA-LIHC dataset and the ‘rms’ and ‘replot’ packages in R to predict the overall survival of HCC patients. The accuracy and discrimination of the nomogram were assessed via calibration analysis using the ‘rms’, ‘caret’, and ‘replot’ packages. Calibration plots were created to compare the predicted survival probabilities at 1, 3, and 5 years with the observed outcomes. In these plots, the predicted survival is represented on the x-axis, and the observed survival rates are shown on the y-axis. This approach is commonly employed in Cox regression models to evaluate model calibration. In the calibration plot, predicted survival is shown along the x-axis, whereas observed survival rates are displayed on the y-axis.

Tumor microenvironment analysis and drug sensitivity prediction

To analyze the immune landscape, gene expression matrices were processed using tools such as TIMER, MCP-COUNTER, EPIC, and CIBERSORT¹⁶. The ESTIMATE algorithm was used to determine immune and stromal scores. The likelihood of benefiting from checkpoint inhibition and tumor immune evasion was also predicted using the TIDE (Tumor Immune Dysfunction and Exclusion) algorithm.

By using the ‘oncopredict’ R package, drug response profiles were inferred through the estimation of IC50 values from the CTRP and GDSC datasets, enabling the analysis of drug sensitivity comparisons across groups.

Quantitative real-time PCR (qRT-PCR)

Liver tissue specimens were obtained from patients who underwent surgical resection at the First Affiliated Hospital of Zhengzhou University. All diagnoses were confirmed through histopathological assessment. Ethical approval was granted by the Ethics Committee of the First Affiliated Hospital of Zhengzhou University (Approval No. 2022-KY-0826-001), and written informed consent was provided by all participants.

Total RNA was extracted from frozen liver samples using TRIzol™ reagent (Cat. No. 15596026, Life Technologies, USA) in accordance with the manufacturer’s protocol. Complementary DNA (cDNA) was synthesized from the extracted RNA using the SweScript All-in-One RT SuperMix for qPCR (Cat. No. G3337, Servicebio, China). qRT-PCR was carried out on a QuantStudio™ 5 Real-Time PCR System (Model A28138, Applied Biosystems, USA) utilizing the 2× Universal Blue SYBR Green qPCR Master Mix (Cat. No. G3326, Servicebio, China). Relative gene expression levels were calculated using the $2^{-\Delta\Delta Cq}$ method, with GAPDH employed as the internal control. Primer sequences are listed in Supplementary Table S3.

Statistical analysis

Appropriate statistical methods, including two-tailed t-tests, chi-square analysis, or one-way ANOVA, were used in R (version 4.1.2) to evaluate group comparisons. Correlation analyses were performed using Pearson correlation coefficients. The differences in survival between groups were analyzed using the log-rank test and Kaplan-Meier survival curves. Statistical significance was set at a P-value below 0.05. The significance levels were marked as: $P < 0.05$ (*), $P < 0.01$ (**), and $P < 0.001$ (***).

Results

Identification and analysis of PCD-related DEGs in HCC

We first analyzed the gene expression profiles of HCC and normal liver tissues from the TCGA database. DEGs were identified using four independent algorithms, and a total of 10,086 DEGs were obtained (Supplementary Fig. S1A). Among them, 7,924 genes were upregulated and 2,162 genes were downregulated in tumor tissues (Fig. 1A).

These DEGs were subsequently intersected with each of the 21 PCD gene sets. By extracting the top five genes for each subtype and excluding overlaps, 85 PCD-related gene candidates were determined (Fig. 1B, Supplementary Fig. S1B). The analysis of enrichment revealed that these genes were largely involved in pathways related to tumorigenesis and progression, regulation of cell death, immune and inflammatory responses, signal transduction, and metabolic and stress response pathways (Fig. 1C, F).

Gene expression analysis indicated a general upregulation of genes linked to PCD in tumor tissues (Fig. 1D). PPI network analysis revealed extensive interactions among these genes, with key nodes involving CASP3, CASP8, MAPK3, and HSP90AA1 (Fig. 1E). In the TCGA-LIHC cohort, LRRK2 had the highest occurrence among somatic mutations, found in 3% of patients (Fig. 1G). Furthermore, copy number variation (CNV) analysis demonstrated widespread CNV events among these genes (Fig. 1H).

Identification of prognostic genes in HCC based on machine learning

To pinpoint the ideal predictive model, we systematically tested 117 configurations across a range of machine learning algorithms. The integration of StepCox (backward) and RSF yielded the highest concordance index (C-index) across multiple datasets, including TCGA-LIHC, GSE14520, and GSE116174 (Fig. 2A). Based on variable importance scores, 10 key genes were selected from 85 candidates—KIF20A, FTL, SLC2A1, BMP6, PLA2G7, LARP1, CD8A, FAF1, PPAT, and NAT10—and the model’s stability and accuracy were confirmed using error rate curves (Fig. 2B–C).

We further analyzed the expression patterns and chromosomal locations of these 10 model genes in the TCGA-LIHC dataset. Apart from BMP6 and CD8A, the other genes were significantly upregulated in tumor tissues (Fig. 2D–E). Correlation analysis among the model genes suggested potential functional links (Fig. 2F), with a strong positive correlation found between LARP1 and NAT10, and a significant negative correlation between BMP6 and FTL. GO enrichment analysis demonstrated that these genes are chiefly enriched in basic biological processes, including protein binding and transport, as well as transcriptional and translational regulation. Moreover, they are involved in pathways concerning cell structure and function, metabolism and biosynthesis, and stress responses (Fig. 2G).

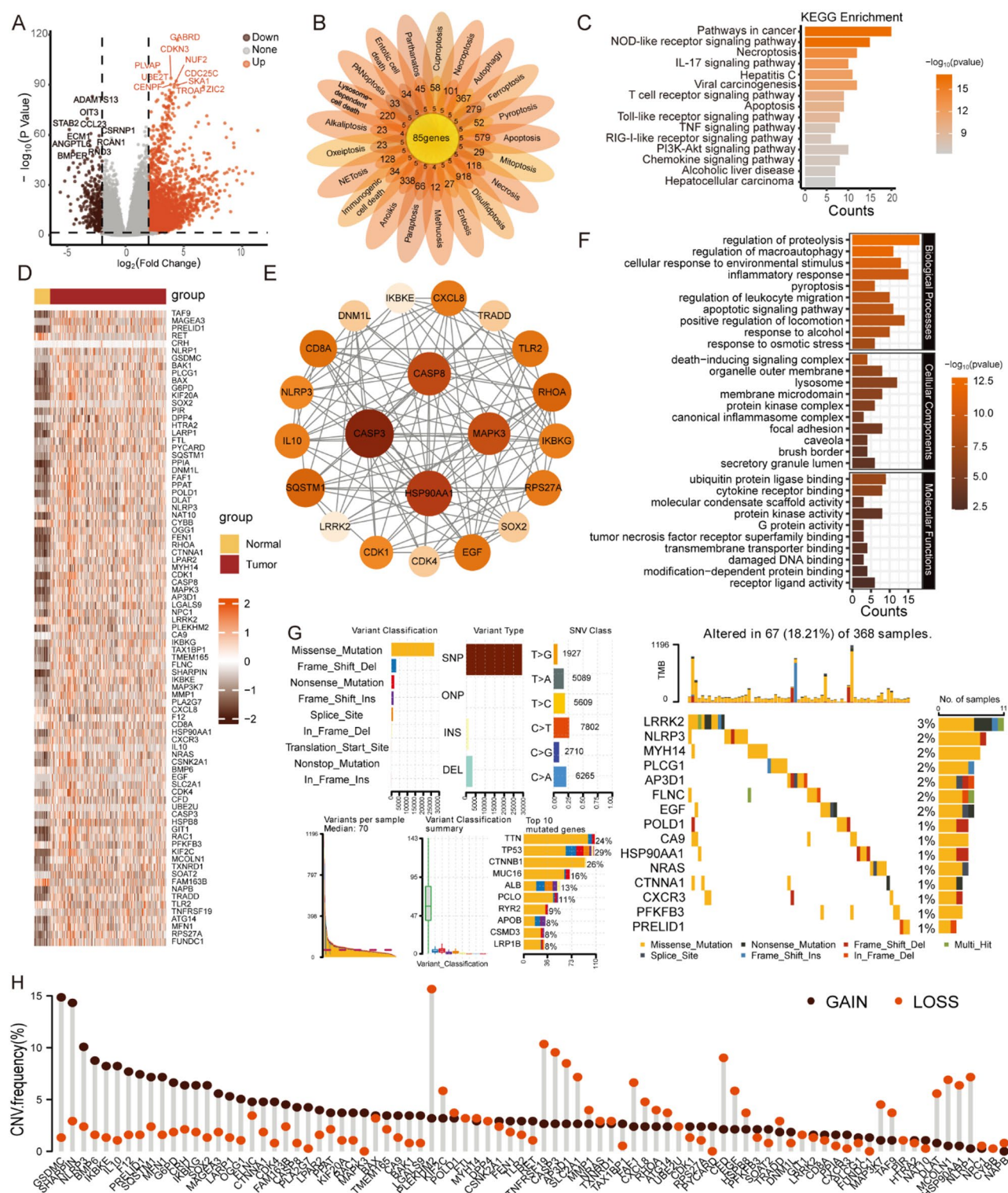


Fig. 1. Identification and Functional Insights of candidate genes in Hepatocellular Carcinoma (HCC). **(A)** Volcano plot of DEGs between tumor and normal tissues in the TCGA-LIHC dataset. **(B)** Petal diagram of PCD-related gene selection, illustrating the distribution of genes across 21 PCD types and the number of selected genes for each type. The central number represents the total number of unique genes after deduplication. **(C)** GO functional enrichment analysis of candidate genes. **(D)** Heatmap showing the expression of PCD-related DEGs between normal and tumor tissues in the TCGA-LIHC dataset. **(E)** PPI network of the 85 selected candidate genes. **(F)** KEGG pathway enrichment analysis of candidate genes. **(G)** SNV analysis of candidate genes. **(H)** CNV analysis of candidate genes.

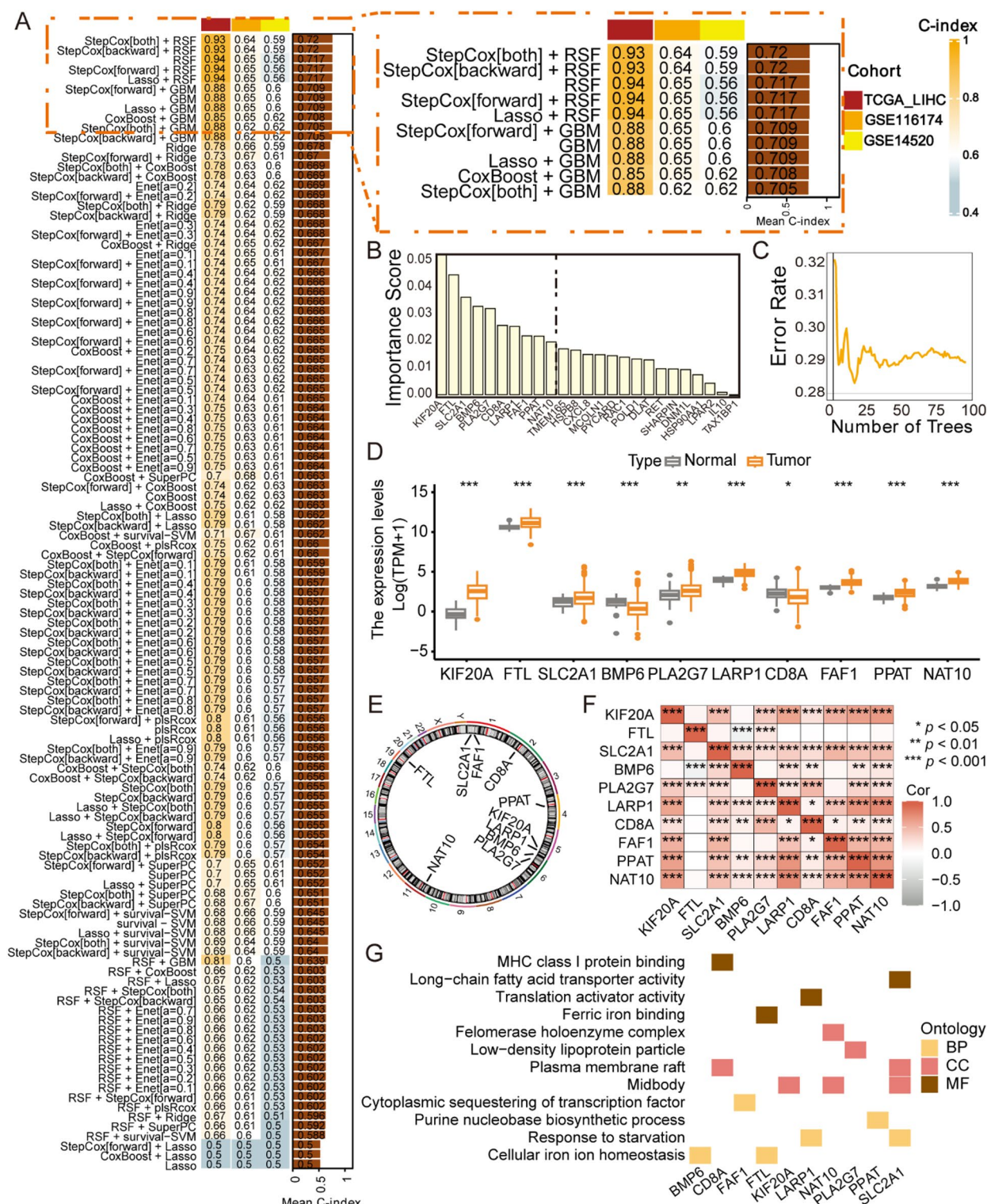


Fig. 2. Machine Learning-Based Selection and Key Feature Analysis of Model Genes. (A) C-index distribution of 117 predictive models based on 10-fold cross-validation. (B) Feature importance scores of the selected model genes identified by the random forest algorithm. (C) Error rate curve for model performance. (D) Boxplots showing differential expression of selected model genes between normal and tumor tissues. (E) Chromosomal localization of the 10 model genes. (F) Correlation analysis of model genes. (G) GO and KEGG pathway enrichment analysis of model genes. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Risk stratification and validation of the prognostic model

According to the cumulative predicted mortality estimated by the Stepwise Cox + RSF model, patients in the three independent cohorts (TCGA-LIHC, GSE14520, and GSE116174) were assigned risk scores. Using the median risk score as the cutoff, patients were stratified into high- and low-risk groups accordingly (Fig. 3A). To evaluate the discriminative ability of the risk model, principal component analysis (PCA) demonstrated clear separation between the two risk groups (Fig. 3B–C). Kaplan–Meier survival analysis demonstrated that the overall survival for patients in the high-risk group was significantly inferior to that of patients in the low-risk group (Fig. 3D). The model's accuracy was confirmed by time-dependent ROC analysis, which demonstrated robust discriminative capacity with elevated AUCs in all three cohorts (Fig. 3E).

We conducted GSVA to assess functional variation across distinct risk groups (Fig. 3F–H). Several biological programs linked to cell proliferation—such as DNA damage repair, MYC/E2F targets, G2M phase transition, and protein folding stress—were predominantly enriched in the high-risk classification group. Conversely, patients in the low-risk category demonstrated upregulation of metabolic processes such as fatty acid oxidation, bile acid biosynthesis, and detoxification across all datasets. The analysis underscores that patients with varying risk scores possess both similar and divergent biological features.

We further assessed protein abundance of model genes by examining staining strength and image data available in the HPA resource (Fig. 3I). The analysis demonstrated that the majority of core genes exhibited elevated expression in tumor tissues compared to adjacent non-tumorous counterparts, aligning with transcriptomic-level observations.

Identification of HCC patient subtypes and construction of a prognostic nomogram

Unsupervised consensus clustering was applied to TCGA-LIHC transcriptome profiles to define HCC molecular subtypes in relation to the model genes. Based on the observed shift in Delta area values from $k = 2$ to 9, the most stable clustering outcome corresponded to $k = 2$ (Fig. 4A–C). Survival analysis based on Kaplan–Meier estimates demonstrated a clear prognostic divergence, where the group corresponding to C2 showed a better clinical outcome (Fig. 4D–E).

According to GSVA findings, subtype C1 demonstrated pronounced signaling in proliferation- and repair-associated networks, while C2 was mainly associated with metabolic regulation and cell differentiation (Fig. 4F). In addition, both clusters showed enrichment in signaling and stress response pathways. To validate functional differences between the clusters, GSEA was employed, highlighting divergent signaling landscapes between the two clustering subgroups (Supplementary Fig. S1C). Evaluation of immune features showed that C1 was characterized by increased immune pathway scores and elevated TIDE scores, suggesting both enhanced immune activation and immune evasion potential (Fig. 4G–H). Immune cell composition varied notably across the two clusters (Supplementary Fig. S1D). Accordingly, analysis of risk scores indicated elevated values in C1 compared to C2, aligning with its poorer prognosis. Moreover, after adjusting for confounding variables, both univariate and multivariate Cox regression analyses confirmed the risk score as an independent prognostic factor for HCC patients (Fig. 4J–K).

Based on the risk score and clinical staging information, we constructed a prognostic nomogram to predict overall survival at 1, 3, and 5 years (Fig. 4L). Calibration results showed a close agreement with the 45-degree reference line, indicating reliable predictive accuracy (Fig. 4M). Temporal ROC assessment further demonstrated that the nomogram had strong discriminative ability for survival prediction (Fig. 4N–O).

Immune characteristics analysis

Quantitative profiling of immune infiltrates from TCGA-LIHC demonstrated considerable differences in cellular prevalence linked to risk status (Fig. 5A). We correlated the expression levels of each model gene with the relative abundance of immune cell types estimated by CIBERSORT (Fig. 5B). CD8A was found to be positively linked to multiple immune cell types, whereas FAF1 exhibited negative correlations.

Immune score evaluation suggested that patients in the low-risk category had more pronounced immune activation (Fig. 5C). To explore associations between model gene profiles and clinical attributes, we visualized their expression patterns alongside major clinical indicators using a heatmap in the TCGA-LIHC dataset (Fig. 5D). Analysis revealed elevated expression of most model genes in individuals classified as high-risk, a group further characterized by advanced T staging, progressive clinical stages, and increased mortality.

Notable variations in immune composition, stromal characteristics, ESTIMATE outputs, and TIDE predictions were identified across the stratified risk groups (Fig. 5E–H). Notably, patients in the lower-risk category demonstrated elevated stromal content, immune infiltration indices, and ESTIMATE-derived values, reflective of a more immunologically active tumor microenvironment. In contrast, a higher TIDE score was characteristic of the high-risk group, pointing to more pronounced immune escape capacity.

Finally, we evaluated immune checkpoint gene profiles across distinct risk categories (Fig. 5I). Elevated levels of various checkpoint molecules were observed in individuals classified as low risk. Combined with the higher TME-related scores in this group, these findings suggest that individuals with lower risk scores could exhibit increased responsiveness to ICI-based treatments.

Drug sensitivity analysis of anticancer agents

Taking into account the observed immune differences between the risk groups, we further investigated their potential implications for drug treatment. Initially, we compared the sensitivity of high-risk and low-risk groups to commonly used HCC therapeutic agents. The results revealed substantial differences in their responses to the drugs (Fig. 5J). Except for Selumetinib and Vincristine, the high-risk group exhibited greater sensitivity to most drugs.

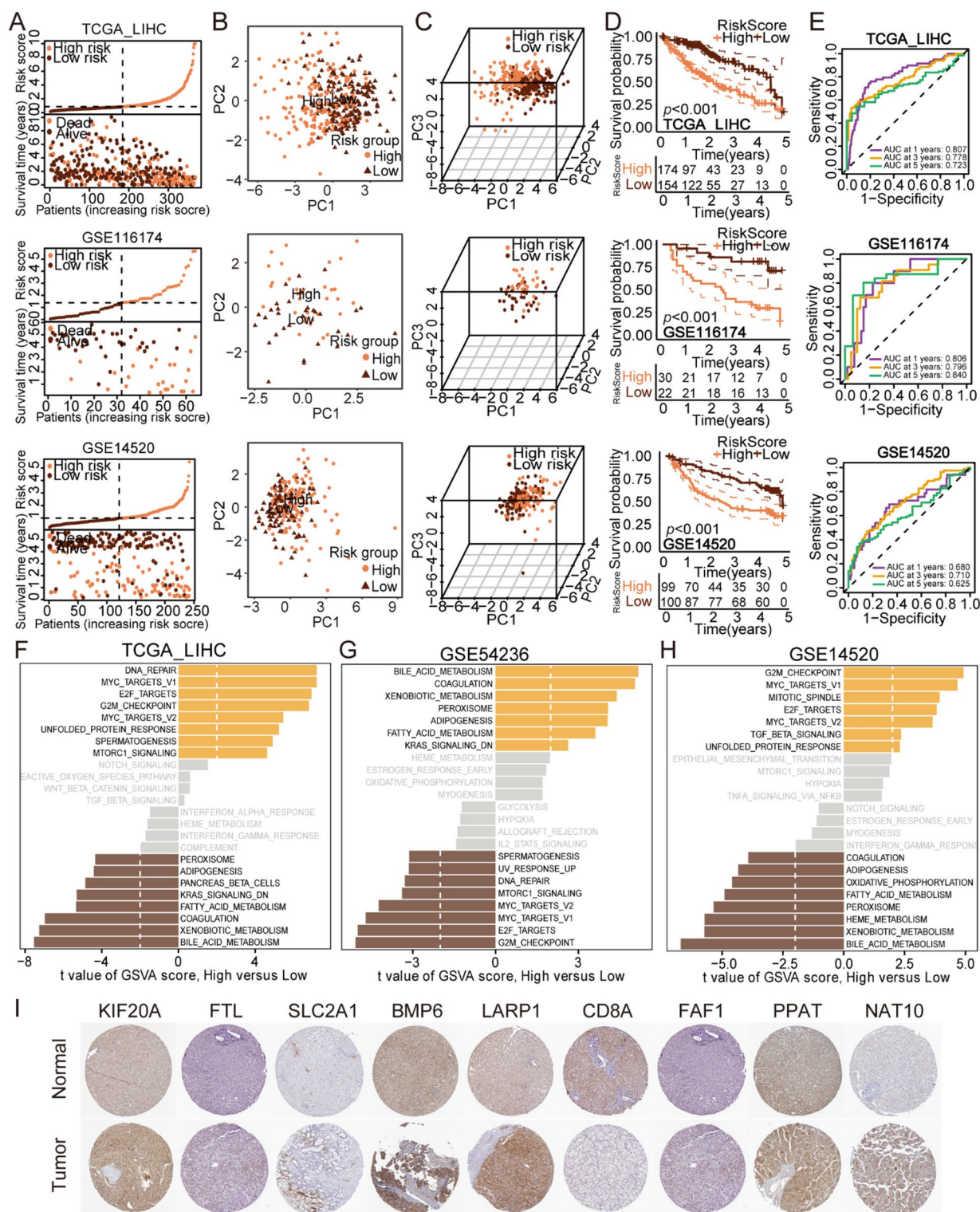
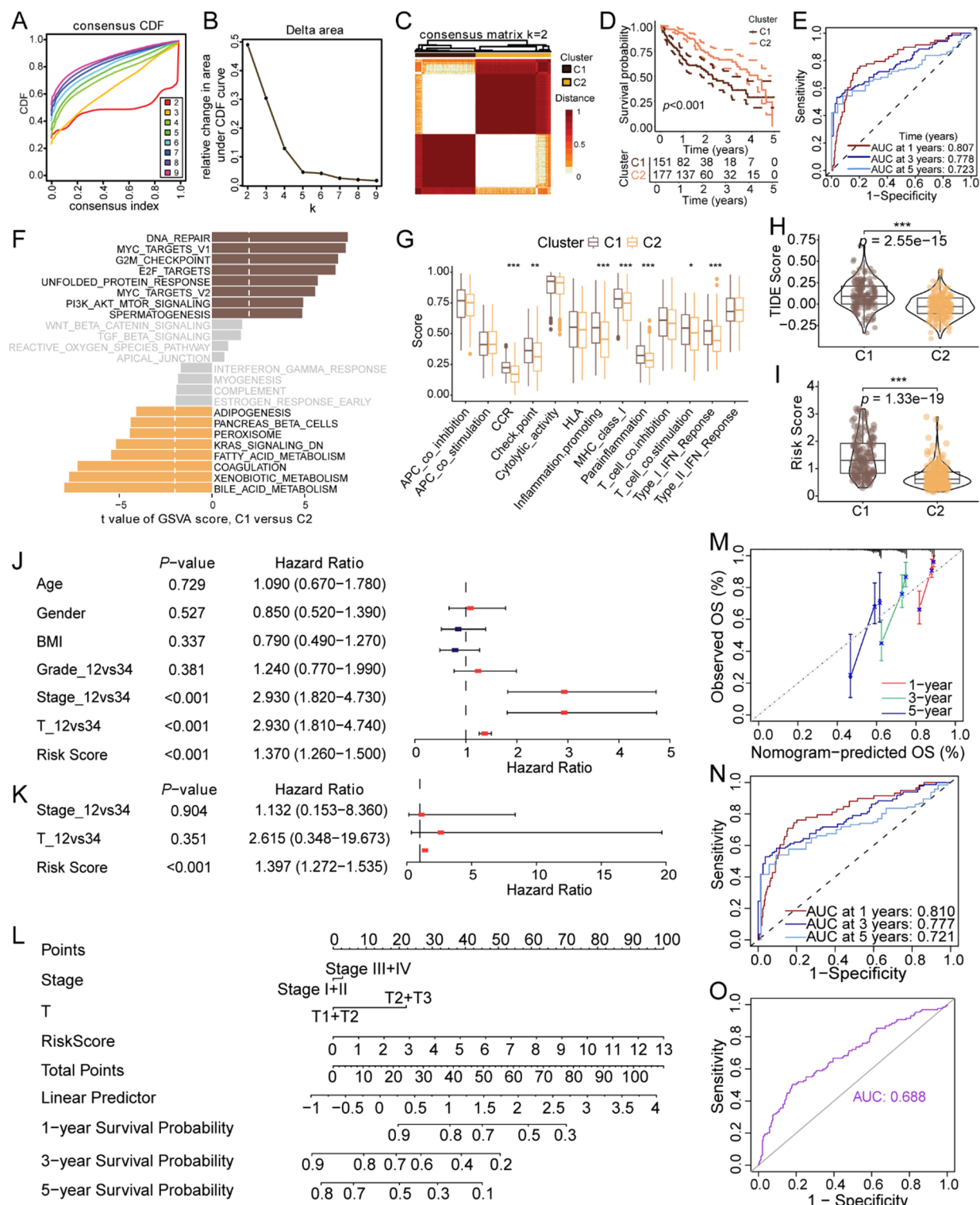


Fig. 3. Construction of a Prognostic Model for HCC Patients, Risk Stratification, and Functional Characterization. (A) Risk stratification analysis in TCGA-LIHC, GSE116174, and GSE14520 cohorts. (B) PCA of high- and low-risk groups. (C) 3D PCA of high- and low-risk groups. (D) Comparison of survival probabilities between high- and low-risk groups. (E) ROC curves for risk prediction in TCGA-LIHC, GSE116174, and GSE14520 cohorts. (F-H) GSEA of high- and low-risk groups in the three HCC datasets (TCGA-LIHC, GSE116174, and GSE14520). (I) Immunohistochemical (IHC) staining images from the HPA showing the expression differences of model genes between normal and tumor tissues.



Subsequently, we utilized data from the GDSC database to investigate the relationship between risk scores and drug response, and to predict drug sensitivity differences across risk groups (Fig. 5K). The analysis showed that the high-risk group was more sensitive to Irinotecan, Oxaliplatin, Olaparib, and Niraparib, whereas it showed lower sensitivity to Axitinib, Crizotinib, Cediranib, and Lapatinib. This evidence indicates that stratifying HCC cases based on risk may facilitate precision treatment approaches.

qRT-PCR validation of model gene expression in HCC tissues

To validate the expression patterns of the model genes, we performed qRT-PCR analysis on 16 paired HCC and adjacent normal liver tissue samples. As shown in Fig. 6, nine out of the ten model genes exhibited significantly higher mRNA expression levels in tumor tissues compared to normal tissues. Specifically, FTL, FAF1, LARP1, SLC2A1, PPAT, PLA2G7, NAT10, and KIF20A were markedly upregulated in HCC samples. With the exception

Fig. 4. Consensus Clustering Based on Marker Genes and Its Associations with Survival, Immunity, and Functional Characteristics. (A) Consensus cumulative distribution function (CDF) curves for different k values. (B) Relative change in the Delta area when k varies from 2 to 9. (C) Consensus matrix heatmap and clustering heatmap for $k = 2$. (D) Kaplan-Meier survival curves comparing the two consensus clusters (C1 vs. C2). (E) ROC curves and AUC evaluation for 1-, 3-, and 5-year survival predictions. (F) GSVA between the two clusters. (G) Comparison of GSEA results between the two clusters. (H) Comparison of TIDE scores between consensus clusters. (I) Risk score distribution across the two consensus clusters. (J) Univariate Cox regression analysis of clinicopathological characteristics and risk scores in TCGA-LIHC patients. (K) Multivariate Cox regression analysis of clinicopathological characteristics and risk scores in TCGA-LIHC patients. (L) Nomogram predicting 1-, 3-, and 5-year survival probabilities. (M) Calibration curves for the Nomogram. (N) Time-dependent ROC curves assessing the predictive performance of the Nomogram. (O) AUC scores validating the predictive accuracy of the Nomogram. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ indicate statistical significance.

of BMP6 and CD8A, the expression trends of the remaining genes were consistent with the transcriptomic data shown in (Fig. 2D).

Discussion

PCD plays a pivotal role in preserving tissue equilibrium and preventing malignant transformation. In HCC, dysregulation of PCD is increasingly recognized as a critical contributor to tumor initiation and progression¹⁷. Beyond classical forms such as apoptosis, necrosis, and autophagy, emerging PCD subtypes—such as pyroptosis and ferroptosis—have also been implicated in the pathogenesis of HCC. While prior investigations have addressed how PCD mechanisms relate to HCC, most have focused on a single PCD mechanism, lacking comprehensive integration across multiple modalities. Moreover, the robustness and applicability of previously proposed prognostic models remain to be thoroughly validated^{18,19}. Against this backdrop, our study is the first to systematically integrate 21 types of PCD, resulting in the development of a robust and efficient risk score model.

Through differential expression analysis based on the TCGA-LIHC cohort, we identified 10,086 DEGs, from which 85 PCD-related candidate genes were obtained via intersection with curated PCD gene sets. These genes were broadly involved in tumorigenesis, apoptosis, and inflammatory responses. To enhance model performance, we applied a combined feature selection strategy using Stepwise Cox regression (backward) and the RSF algorithm. This hybrid approach leverages the interpretability of linear models and the predictive power of nonlinear models, resulting in improved model stability and accuracy while minimizing complexity. The final 10-gene model encompassed multiple PCD subtypes, including NETosis, lysosome-dependent cell death, immunogenic cell death, anoikis, parthanatos, ferroptosis, autophagy, necroptosis, and cuproptosis. Collectively, the model captures the diverse contributions of PCD mechanisms in HCC progression and demonstrates consistent prognostic performance across multiple independent datasets.

Individuals were successfully stratified into distinct risk categories, which exhibited marked heterogeneity in molecular features, immune microenvironment, metabolic pathways, and drug sensitivity. According to GSVA, patients with higher risk scores showed pathway activation patterns associated with cell cycle progression and stress tolerance, notably involving DNA repair, MYC, and E2F modules. In contrast, low-risk patients showed enrichment in metabolic pathways, including fatty acid, bile acid, and xenobiotic metabolism, suggesting higher metabolic activity. Immune analysis demonstrated that high-risk patients had lower tumor microenvironment (TME) scores and elevated TIDE scores, reflecting enhanced immune evasion capacity²⁰. Conversely, the low-risk group exhibited upregulation of multiple immune checkpoint genes and higher TME scores, suggesting a greater likelihood of benefiting from ICI. These findings are consistent with the survival trends observed between risk groups and highlight immune escape as a potential mechanism underlying poor prognosis in patients with higher risk scores.

Drug response profiling identified notable disparities across the two risk categories, with most agents closely linked to PCD-related mechanisms. The high-risk subset demonstrated increased sensitivity to irinotecan, oxaliplatin, olaparib, and niraparib—agents that primarily induce DNA damage and interfere with DNA repair to trigger apoptotic responses. For instance, irinotecan inhibits topoisomerase I, causing single-strand DNA breaks and subsequent apoptosis, while PARP inhibitors like olaparib and niraparib block DNA repair, leading to damage accumulation and apoptosis induction^{21,22}. Conversely, individuals classified as low-risk exhibited increased susceptibility to agents like axitinib, crizotinib, cediranib, and lapatinib, which act through anti-angiogenic or tyrosine kinase signaling pathways to induce apoptosis or autophagy. Axitinib, for example, suppresses VEGFR signaling to inhibit angiogenesis and promote apoptosis, while lapatinib targets EGFR/HER2 to trigger apoptotic and autophagic responses^{23,24}. These distinct drug responses underscore the molecular and mechanistic heterogeneity between risk groups and highlight the model's potential utility in guiding personalized therapy decisions.

Additionally, unsupervised clustering based on the model genes successfully identified biologically and prognostically distinct molecular subtypes of HCC, further validating the stratification capacity of the gene set. The nomogram incorporating both risk score and clinical variables demonstrated favorable accuracy in short-term survival prediction, providing a practical tool for individualized prognostic assessment in clinical settings. These results extend the applicability of the model and enhance our understanding of molecular heterogeneity in HCC.

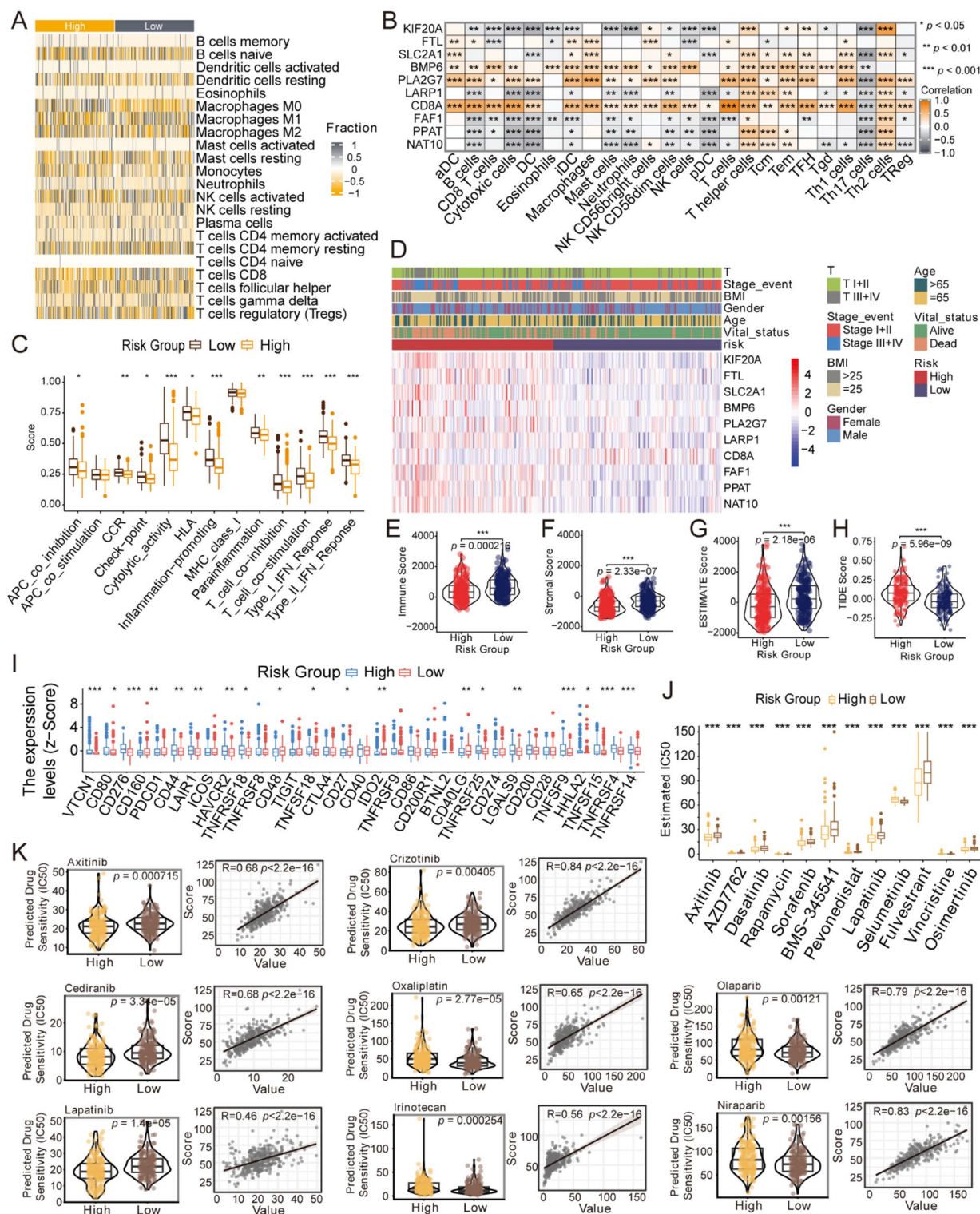


Fig. 5. Analysis of the Tumor Immune Microenvironment, Immune Checkpoints, and Drug Sensitivity in HCC. **(A)** Heatmap comparison of immune cell infiltration levels between risk groups. **(B)** Correlation analysis between immune cell infiltration and the expression of model genes. **(C)** Comparison of immune-related scores between high- and low-risk groups. **(D)** Distribution of clinical characteristics and model gene expression patterns between high- and low-risk groups. Comparison of immune **(E)**, stromal **(F)**, and ESTIMATE **(G)** scores between high- and low-risk groups. **(H)** Comparison of TIDE scores between different risk groups. **(I)** Expression analysis of immune checkpoint genes between high- and low-risk groups. **(J)** Drug sensitivity analysis of commonly used HCC therapeutic agents across different risk groups. **(K)** Comparison of drug sensitivity differences between different risk groups. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ indicate statistical significance.

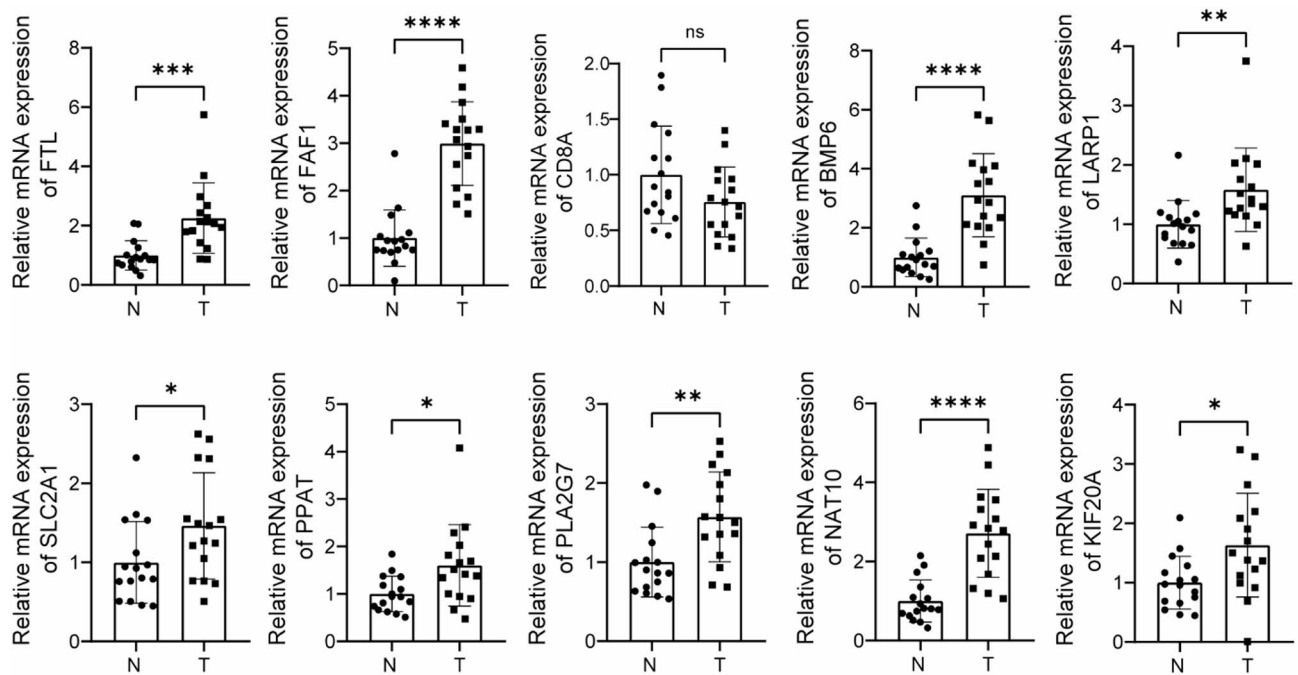


Fig. 6. qRT-PCR-based validation of model gene expression. The relative mRNA expression levels of representative model genes were assessed by qRT-PCR in paired normal (N, $n = 16$) and tumor (T, $n = 16$) tissue samples. Statistical differences between groups were evaluated, with significance indicated as follows: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$; ns, not significant.

To verify the reliability of the transcriptomic analysis, we further examined the expression of model genes in liver tissue samples using qRT-PCR. With the exception of a few genes, most exhibited upregulated expression in tumor tissues consistent with the transcriptome data, supporting their potential roles in HCC and validating the robustness of our selection strategy. Notably, We observed that CD8A gene expression did not show a significant difference between the control and tumor groups in qRT-PCR, although its expression trend was consistent with the results from transcriptomic analysis. In contrast, BMP6 gene expression was downregulated in the tumor group in transcriptomic analysis, but was significantly upregulated at the mRNA level in the same group. This discrepancy may be attributed to clinical and molecular heterogeneity between the HCC patient samples we collected and the HCC samples from the TCGA-LIHC dataset. Samples from different sources may exhibit variations in tumor microenvironment, immune cell infiltration, and other factors, which could influence gene expression measurements. Additionally, although RNA-seq technology is highly sensitive, its accuracy in quantifying low-expression genes may be affected by various factors. These factors may collectively contribute to the observed differences in gene expression results.

Despite these advances, several limitations should be noted. First, the model was primarily constructed using transcriptomic data, lacking validation from proteomic analyses or functional experiments, which may limit its biological interpretability. Additionally, drug sensitivity predictions require further validation in larger, multicenter cohorts. Our drug sensitivity analysis primarily relied on cell line data, which did not fully account for the influence of the tumor microenvironment and interpatient variability on drug responses. Therefore, future studies should prioritize the use of patient-derived models to more accurately assess the clinical relevance of drug sensitivity predictions in clinical practice. Additionally, although our risk score shows a certain correlation with TIDE predictions and immune checkpoint profiles, inferring treatment responses using drug sensitivity predictions and immune checkpoint gene expression profiles still presents challenges. The clinical utility of our model in predicting actual immune therapy responses requires validation in prospective cohorts to confirm its feasibility in clinical practice. Integrating cellular-resolution techniques like scRNA-seq or spatial transcriptomics could help elucidate the contribution of PCD pathways in defined microenvironments, thus improving model precision and generalizability. Expanding this model to other cancer types may also facilitate pan-cancer risk prediction and therapeutic guidance. Ultimately, this work enhances current insights into the role of PCD in cancer progression and lays the groundwork for future precision oncology interventions.

Conclusion

Our analysis systematically integrated genes associated with 21 types of PCD and identified a 10-gene risk score model with strong prognostic relevance in HCC. The model was validated across multiple independent cohorts, effectively stratifying patients into high- and low-risk groups. Furthermore, the risk score showed significant correlations with the tumor immune microenvironment, immunotherapy responsiveness, and drug sensitivity. These findings offer valuable insights and a practical tool for risk-based stratification and individualized treatment in HCC.

Data availability

The datasets analysed during the current study are publicly available. TCGA-LIHC data were obtained from the TCGA Research Network portal (<https://portal.gdc.cancer.gov/>). GEO datasets GSE14520 and GSE116174 were retrieved from the Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>).

Received: 26 June 2025; Accepted: 18 September 2025

Published online: 23 October 2025

References

- Villanueva, A. & Hepatocellular Carcinoma *N Engl. J. Med.* **380**, 1450–1462 (2019).
- Huang, D. Q., Mathurin, P., Cortez-Pinto, H. & Loomba, R. Global epidemiology of alcohol-associated cirrhosis and HCC: trends, projections and risk factors. *Nat. Rev. Gastroenterol. Hepatol.* **20**, 37–49 (2023).
- Yang, J. D. et al. A global view of hepatocellular carcinoma: trends, risk, prevention and management. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 589–604 (2019).
- Levero, M. & Zucman-Rossi, J. Mechanisms of HBV-induced hepatocellular carcinoma. *J. Hepatol.* **64**, S84–s101 (2016).
- Park, J. W. et al. Global patterns of hepatocellular carcinoma management from diagnosis to death: the BRIDGE study. *Liver Int.* **35**, 2155–2166 (2015).
- Bedoui, S., Herold, M. J. & Strasser, A. Emerging connectivity of programmed cell death pathways and its physiological implications. *Nat. Rev. Mol. Cell. Biol.* **21**, 678–695 (2020).
- Wang, M., Yu, F., Zhang, Y. & Li, P. Programmed cell death in tumor immunity: mechanistic insights and clinical implications. *Front. Immunol.* **14**, 1309635 (2023).
- Nikoletopoulou, V., Markaki, M., Palikaras, K. & Tavernarakis, N. Crosstalk between apoptosis, necrosis and autophagy. *Biochim. Biophys. Acta.* **1833**, 3448–3459 (2013).
- Ye, T., Shan, P. & Zhang, H. Progress in the discovery and development of small molecule methuosis inducers. *RSC Med. Chem.* **14**, 1400–1409 (2023).
- Pashootan, P. et al. Metal-based nanoparticles in cancer therapy: exploring photodynamic therapy and its interplay with regulated cell death pathways. *Int. J. Pharm.* **649**, 123622 (2024).
- Wang, H. et al. Cell death affecting the progression of gastric cancer. *Cell. Death Discov.* **8**, 377 (2022).
- Qin, H. et al. Integrated machine learning survival framework develops a prognostic model based on inter-crosstalk definition of mitochondrial function and cell death patterns in a large multicenter cohort for lower-grade glioma. *J. Transl. Med.* **21**, 588 (2023).
- Wang, S. et al. Machine learning reveals diverse cell death patterns in lung adenocarcinoma prognosis and therapy. *NPJ Precis. Oncol.* **8**, 49 (2024).
- Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
- Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform.* **14**, 7 (2013).
- Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M. & Alizadeh, A. A. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol. Biol.* **1711**, 243–259 (2018).
- Peng, F. et al. Regulated cell death (RCD) in cancer: key pathways and targeted therapies. *Signal. Transduct. Target. Ther.* **7**, 286 (2022).
- Zhou, X. et al. O-GlcNAcylation regulates the stability of transferrin receptor (TFRC) to control the ferroptosis in hepatocellular carcinoma cells. *Redox Biol.* **73**, 103182 (2024).
- Gu, X., Pan, J., Li, Y. & Feng, L. A programmed cell death-related gene signature to predict prognosis and therapeutic responses in liver hepatocellular carcinoma. *Discov. Oncol.* **15**, 71 (2024).
- Zhou, Y., Zhu, J., Gu, M. & Gu, K. Prognosis and characterization of microenvironment in cervical cancer influenced by fatty acid Metabolism-Related genes. *J. Oncol.* **2023**, 6851036 (2023).
- Petković, M., Henis, M., Heese, O. & Relógio, A. Chronotherapy in glioblastoma: state of the Art and future perspectives. *EBioMedicine* **89**, 104470 (2023).
- Kim, J. W. et al. TDP1 and TOP1 modulation in Olaparib-Resistant cancer determines the efficacy of subsequent chemotherapy. *Cancers (Basel)* **12**, 334 (2020).
- Jiang, H. et al. The multikinase inhibitor axitinib in the treatment of advanced hepatocellular carcinoma: the current clinical applications and the molecular mechanisms. *Front. Immunol.* **14**, 1163967 (2023).
- Qu, N. et al. INPP4B suppresses HER2-induced mesenchymal transition in HER2+ breast cancer and enhances sensitivity to lapatinib. *Biochem. Pharmacol.* **226**, 116347 (2024).

Author contributions

H.Y. and D.L. drafted the main manuscript text. Q.L. and S.W. carried out the methodology and data curation. T.W. developed the software and contributed to project supervision, administration, and funding acquisition. S.C. performed the validation and was responsible for funding acquisition. J.J.L. conducted the formal data analysis. H.F. carried out the investigation. H.Y. and S.C. provided resources essential to the study. S.W. and J.L. critically reviewed and revised the manuscript. C.P. and T.W. supervised the overall project. X.S. coordinated project administration. X.Z. contributed to funding acquisition, supervision, and writing – review and editing. All authors read and approved the final version of the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (grant number U2004122); the Henan Charity Federation Fund (grant number GDXZ2023014); the Henan Provincial Natural Science Foundation (grant number 242300420395); and the Science and Technology Research Project of Henan Province (grant number 252102311286).

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

This study was conducted in accordance with the principles of the Declaration of Helsinki. Human cholangiocarcinoma tissue samples were collected at the First Affiliated Hospital of Zhengzhou University (Zhengzhou, China). The study was approved by the Ethics Committee of the First Affiliated Hospital of Zhengzhou University (Approval No. 2022-KY-0826-001). Written informed consent was obtained from all participants prior to sample collection.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-21123-1>.

Correspondence and requests for materials should be addressed to J.L., S.W. or T.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025