Article

Dual tissue mRNA and serum protein signatures improve risk stratification in hepatocellular carcinoma

Check for updates

Ding-Fan Guo[1,3], Lin-Wei Fan[2,3], Qi Wen[1], Jin-Ke Wang[1], Yun-Hui Liang[1], Qi Feng[1], Ting Wang[1] ✉ & Kun-He Zhang[1] ✉

Hepatocellular carcinoma (HCC) is highly heterogeneous, making prognosis and treatment prediction challenging. Using multi-omics data from multiple HCC cohorts, we identified five biomarkers (AKR1B10, ANXA2, COL15A1, SPARCL1, and SPINK1) and developed dual serum and tissue signatures by machine learning. The tissue mRNA signature could stratify prognostic risk and reflect alterations in the tumor's genome, metabolism, and immune microenvironment. High-risk HCC responded poorly to sorafenib and transarterial chemoembolization (TACE) but sensitively to agent ABT-263 in silico, in vitro, and in vivo experiments. The serum protein signature outperformed the clinical tumor staging systems in predicting 24-month disease-free survival, with median time-dependent areas under the receiver operating characteristic curve (AUC(t)) of 0.79 and 0.75 in two postoperative cohorts, and the AUC was 0.90 for predicting treatment benefit in a TACE-treated cohort. Interpretability analysis revealed consistent biomarker contributions in both signatures. Conclusively, the dual signatures show promise for HCC risk stratification, pending external validation.

Liver cancer is the sixth most prevalent tumor and the fourth primary cause of cancer-related deaths, with hepatocellular carcinoma (HCC) constituting roughly 85–90% of cases[1]. Traditional HCC clinical staging, aimed at improving overall prognosis through stratified management, involves radical hepatectomy for early-stage cases, transarterial chemoembolization (TACE) for intermediate-stage cases, and systemic therapy for advanced-stage cases[1,2]. However, due to HCC heterogeneity, patients at the same stage exhibit varied responses to standardized treatments, highlighting the limitations of traditional clinical staging in treatment guidance[3,4].

Multi-omics sequencing techniques provide an opportunity to establish a molecular-driven HCC stratification method[5–8]. The standard process involves analyzing tissue samples to identify disease-driving and prognostic-associated biomarkers, followed by constructing a predictive model for prognosis assessment and treatment stratification. The Cancer Genome Atlas (TCGA) identified HCC molecular subtypes with prognostic differences through the integration of multi-dimensional genomic data[5]. The fatty acid metabolism-related classification developed by Li et al. was valuable for guiding individualized treatment[8]. However, these approaches rely on tissue-based sequencing, and their invasive nature impedes clinical translation.

Blood biomarkers offer the advantage of non-invasive sample collection and are thus ideal for disease evaluation[9,10]. Many studies have evaluated the prognostic value of traditional and emerging blood biomarkers for HCC. Classical biomarkers, such as alpha-fetoprotein (AFP), reflect tumor aggressiveness, with elevated AFP levels linked to higher recurrence risk and poorer survival[11]. Protein induced by vitamin K absence or antagonist-II (PIVKA-II) levels are correlated with radiological responses and patient outcomes, highlighting its potential for predicting treatment efficacy[12,13]. AFP and PIVKA-II have been incorporated into several prognostic models for recurrence risk stratification following liver transplantation, surgical resection, or local therapy, thereby supporting the design of long-term follow-up strategies[14–16]. Nonetheless, their predictive value varies across populations, and their value is limited when biomarker levels are low. Emerging blood biomarkers, including quantitative levels and mutation or methylation profiles of circulating tumor DNA (ctDNA) or cell-free DNA (cfDNA), as well as circulating non-coding RNAs, are gaining application in recurrence surveillance and treatment-response prediction[17–20]. Dynamic changes in these biomarkers during the postoperative period or systemic therapy can indicate minimal residual disease and signal an elevated risk of early recurrence. Moreover, specific DNA methylation signatures have been

[1]Department of Gastroenterology, Jiangxi Provincial Key Laboratory of Digestive Diseases, Jiangxi Clinical Research Center for Gastroenterology, Digestive Disease Hospital, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang, Jiangxi, China. [2]First Clinical Medical College, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang, Jiangxi, China. [3]These authors contributed equally: Ding-Fan Guo, Lin-Wei Fan. ✉e-mail: tingwang@ncu.edu.cn; khzhang@ncu.edu.cn

THE HORMEL INSTITUTE
UNIVERSITY OF MINNESOTA

associated with therapeutic responses to transarterial chemoembolization (TACE), targeted agents, and immunotherapy. Despite these advances, the clinical use of biomarkers for risk stratification remains hindered by limited reproducibility across cohorts, insufficient standardization of detection methods and diagnostic thresholds, and constraints related to high cost and accessibility[21]. Therefore, ongoing research is focusing heavily on developing and validating novel blood biomarkers to enhance precision stratification, recurrence monitoring, and efficacy prediction in HCC.

Proteins are the main type of blood biomarker, directly reflecting pathophysiological processes and benefiting from advanced detection technologies, which hold greater promise for clinical translation[10]. Xing et al. employed a "discovery-verification-validation" strategy to identify a combination of serum protein biomarkers for early HCC diagnosis[22]. However, identifying biomarkers solely from blood samples is limited by the difficulty of tracing their tissue origins[23,24]. Inflammatory interference can influence blood proteins originating from multiple tissues, making the identification of HCC-specific biomarkers challenging. By integrating bioinformatics data on secreted proteins and analyzing the conversion from organ-specific transcriptomes to circulating secreted proteomes, tissue-specific blood protein prediction models can be established[25]. The TexSEC method achieves this by transforming liver transcriptomes into secreted proteome prediction models, thereby enabling HCC risk stratification in cirrhosis patients[25].

In prior studies[26,27], we employed aptamers to identify a set of protein biomarkers elevated in the serum of HCC patients. Based on this finding, we analyzed the multi-omics expression profiles of these biomarkers and their correlation with patient prognosis in the present study. Using data from independent clinical cohorts, we developed and validated a tissue mRNA signature and explored the associated genomic variations, epigenetic regulation, metabolic reprogramming, and tumor immune microenvironment in signature-stratified groups. Further, we constructed a corresponding serum protein signature using machine learning and validated it in clinical settings, including curative resection and TACE. Results demonstrate that dual signatures hold potential clinical utility and may offer a novel tool for HCC clinical management.

Results

Candidate biomarkers identified through integrated multi-omics analysis

We identified biomarkers and systematically evaluated their expression consistency across multi-omics levels (Fig. 1A). Analyzing transcriptomic data from the TCGA-LIHC and ICGC (LIRI-JP) cohorts of HCC, we identified 3041 and 3004 differentially expressed genes that were upregulated in HCC tissues compared to adjacent normal tissues (ANT), respectively (Fig. 1B). By intersecting these findings with our previous aptamer-based serum proteomics dataset of HCC patients, eight preliminary biomarkers were identified: AKR1B10, ANXA2, COL15A1, SPARCL1, SPINK1, LYZ, MMP9, and PIP (Fig. 1C). Furthermore, we integrated 24 human multi-omics datasets to further validate the expression levels of the biomarkers (Supplementary Fig. 1A). The results showed that, compared with control samples, COL15A1, SPINK1, AKR1B10, ANXA2, LYZ, and SPARCL1 were upregulated in HCC tissue or serum in at least 80% of the cohorts and demonstrated diagnostic utility (Fig. 1D, Supplementary Table 1). Single-cell sequencing analysis demonstrated that AKR1B10, ANXA2, COL15A1, SPARCL1, and SPINK1 were upregulated across different cell types of HCC tissue (Fig. 1E). In particular, AKR1B10 exhibited a multicellular origin in the HCC tumor microenvironment. It was overexpressed in hepatocytes, fibroblasts, endothelial cells, and T/NK cells, which indicates that AKR1B10 as a serum biomarker may involve a multicellular secretory mechanism (Supplementary Fig. 1B). Animal model datasets demonstrated that in NASH-related (Fig. 1F), adenovirus-induced (Supplementary Fig. 1C), and HBV-related (Supplementary Fig. 1D) HCC mouse models, the expression levels of AKR1B10, ANXA2, COL15A1, SPARCL1, and SPINK1 increased with HCC development.

Next, we evaluated the prognostic value of the above biomarkers. Using the optimal cutoff value for transcriptomic expression, we divided samples from the TCGA-LIHC cohort into high and low-expression groups. Univariate Cox regression analysis revealed that AKR1B10, ANXA2, COL15A1, SPARCL1, and SPINK1 had statistically significant survival prediction efficacy (Fig. 2A). Moreover, after adjusting for age, sex, and Tumor Node Metastasis (TNM) stage, multivariate Cox analysis confirmed that high expression of AKR1B10, ANXA2, and SPINK1 was independently associated with a shorter overall survival (OS), whereas high expression of COL15A1 and SPARCL1 exhibited a protective prognostic effect (Fig. 2A). To verify the stability of the prognostic value, we performed log-rank survival analysis in the independent CLCA and CHCC-HBV cohorts, and the results indicated that these five biomarkers consistently maintained good prognostic value (Supplementary Fig. 2).

Based on the above analyses, we compiled a comparative analysis table of the eight candidate biomarkers and evaluated them using a decision matrix (Supplementary Tables 1 and 2). ANXA2, AKR1B10, SPINK1, COL15A1, and SPARCL1 exhibited significant prognostic value in tissue samples, consistent expression direction across cohorts (≥80%), and availability of commercial enzyme-linked immunosorbent assay (ELISA) kits, and were therefore selected for further investigation. MMP9 was eliminated due to low cross-cohort consistency in expression, LYZ due to a non-significant prognostic value, and PIP due to inadequate expression stability.

Considering the secretory characteristics of these candidate biomarkers, we verified the correlation between their serum and tissue expression levels. In 35 paired serum and tissue samples from HCC patients that were collected at the First Affiliated Hospital of Nanchang University (FAHNU), the results of immunohistochemistry (IHC) and ELISA indicated that serum levels of the five biomarkers were significantly positively correlated with tissue protein expression ($R > 0.8$, $p < 0.001$; Fig. 2B, C). Further analysis revealed a significant positive correlation between biomarker protein levels in the supernatants of six HCC cell lines and the corresponding mRNA expression levels ($R > 0.85$, $p < 0.001$; Supplementary Fig. 3A, B).

Additionally, we selected ANXA2 and AKR1B10, which are closely associated with a poor prognosis in HCC, and manipulated their expression using gene knockdown and overexpression (Supplementary Fig. 3C, D). Knockdown of either gene significantly suppressed the proliferation and invasion potential of Huh-7 and MHCC97H cells, whereas overexpression exerted the opposite effects (Supplementary Fig. 3E–H). These findings were consistent with the results of the survival analysis.

In summary, the selection of five biomarkers was based on: (1) consistent prognostic significance (HR > 1.5 or <0.67, $p < 0.05$) across ≥3 cohorts, (2) technical feasibility for ELISA detection, and (3) expression correlation between tissue and serum ($r > 0.8$).

Development and multi-cohort validation of tissue mRNA signature

To prevent data leakage that could affect model construction, we excluded the CHCC-HBV and CLCA cohorts and used seven other independent cohorts to build a validation set ($n = 839$; Fig. 3A, B). Based on five candidate biomarkers (AKR1B10, ANXA2, COL15A1, SPARCL1, and SPINK1), a tissue mRNA signature for HCC was built using LASSO-Cox regression (Fig. 3C–E). Patients were stratified into high- and low-risk groups using the optimal cutoff value of the signature-derived risk score as the threshold (Supplementary Fig. 5A). Survival analysis of the training set showed that the high-risk group had significantly worse overall survival (HR 1.99, 95% CI 1.36–2.91) and disease-free survival (DFS; HR 1.61, 95% CI 1.12–2.32), as shown in Fig. 3F. Validation set confirmed the robust predictive performance of this signature for OS and DFS (Fig. 3G). The signature's risk score was significantly and positively correlated with pathological grading but not correlated with TNM staging (Supplementary Fig. 4B). Moreover, the signature retained its prognostic significance across various subgroups, including age (≥60 years and <60 years), gender (male and female),
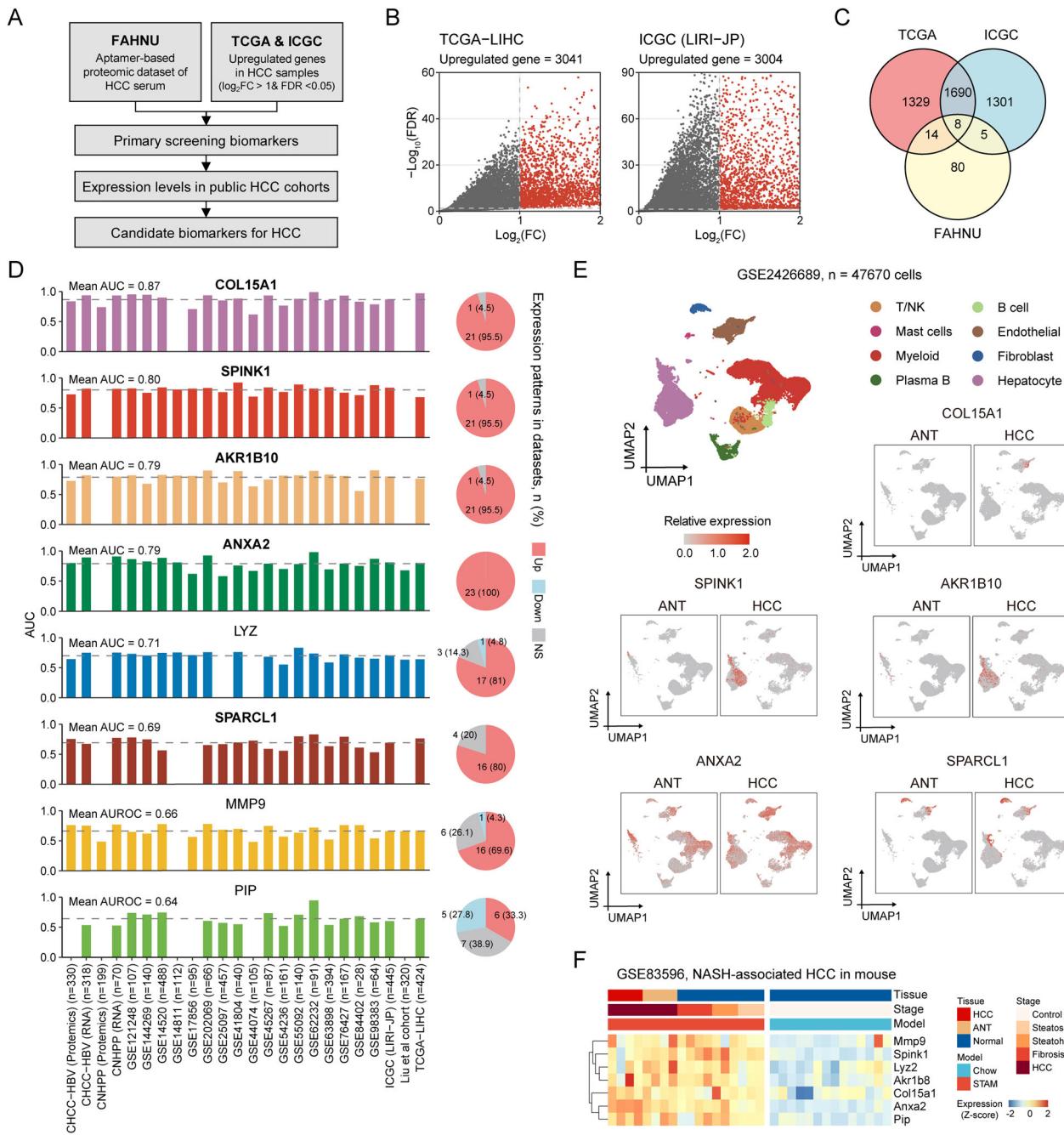
Fig. 1 | Multi-omics identification and analysis for candidate biomarkers.
A Biomarker screening flowchart. **B** Volcano plots depicting differentially expressed genes between HCC and ANT tissues in the TCGA-LIHC (left) and ICGC-LIRI-JP (right) datasets. Red dots indicate genes with statistically significant upregulation (FC ≥ 2, FDR < 0.05); gray dots indicate genes without statistically significant differential expression (FDR ≥ 0.05). **C** Venn diagram showing the intersection of upregulated biomarkers in the three datasets. **D** Bar charts indicating the diagnostic performance (AUC) of biomarkers (left); Pie charts illustrating the expression status of candidate biomarkers in the datasets (right): red indicates upregulation, blue indicates downregulation, and gray indicates no significant difference. The expression levels of biomarkers in HCC and control tissue were compared using the

Wilcoxon rank sum test with Benjamini–Hochberg multiple correction analysis to determine statistical significance. **E** Expression of candidate biomarkers at the single-cell level. UMAP plot showing cell type identification for 47670 high-quality single cells (upper left); Inset comparing expression levels of candidate biomarkers between ANT and HCC tissues (lower right). **F** Heatmap of NASH-related HCC progression in a mouse HCC model. ICGC International Cancer Genome Consortium, TCGA The Cancer Genome Atlas, ANT adjacent normal tissues, AUC area under the receiver operating characteristic curve, FC fold change, FDR false discovery rate, FAHNU the First Affiliated Hospital of Nanchang University, HCC hepatocellular carcinoma, NASH non-alcoholic steatohepatitis, UMAP uniform manifold approximation and projection.

pathological grading (G3–4 and G1–2), and clinical staging (III–IV and I–II) (Fig. 3H).

In using the signature to predict the response to non-surgical treatments, we integrated data from the GSE109211 ($n = 67$), GSE14520 ($n = 104$), and GSE104580 ($n = 147$) cohorts for analysis. Subgroup analysis based on the risk score showed that the treatment failure rate for sorafenib and TACE increased with higher risk scores, and non-responders exhibited significantly higher risk scores than responders ($p < 0.001$) (Supplementary Fig 4C, D). The high-risk group had a significantly higher treatment failure rate than the low-risk group (Fig. 3I). Moreover, the signature demonstrated

A

| Biomarker	Univariate Cox		Multivariate Cox		
	HR (95% CI)	*p* value	HR (95% CI)	*p* value	
AKR1B10	1.92 (1.28-2.86)	0.002	2.18 (1.40-3.41)	<0.001	
ANXA2	2.15 (1.5-3.06)	<0.001	1.95 (1.32-2.86)	<0.001	
SPINK1	1.65 (1.14-2.39)	0.008	1.74 (1.18-2.56)	0.005	
COL15A1	0.52 (0.37-0.74)	<0.001	0.52 (0.36-0.76)	<0.001	
SPARCL1	0.57 (0.40-0.81)	0.002	0.57 (0.39-0.84)	0.004	
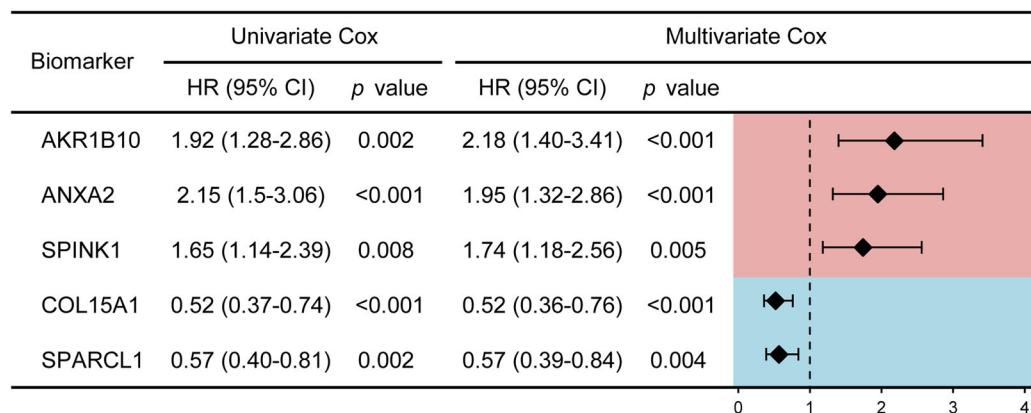
B
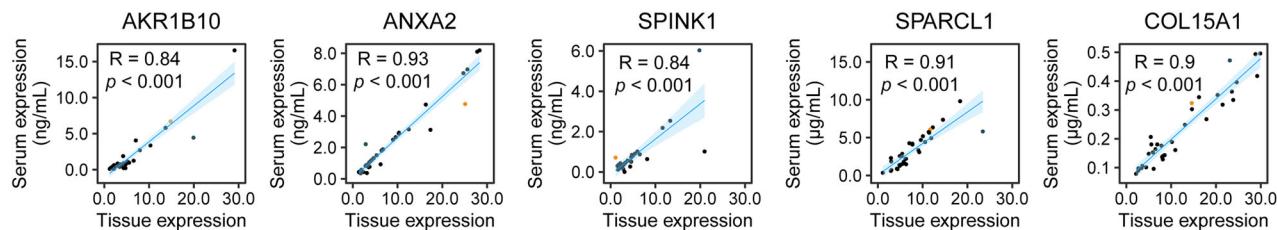
C
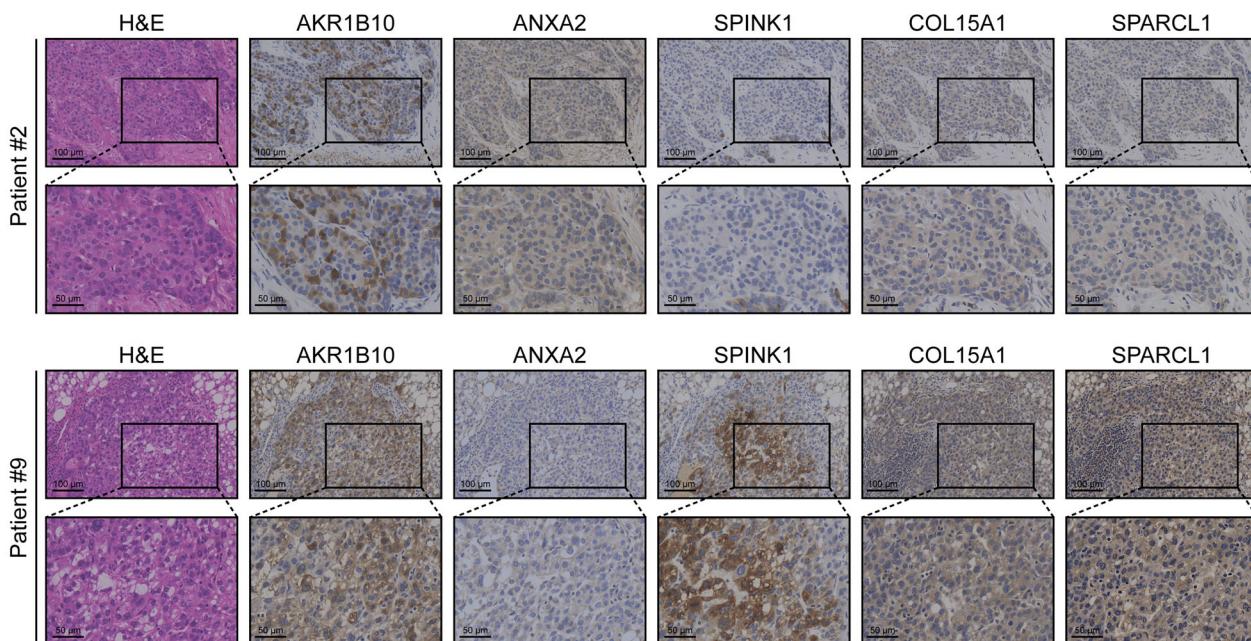
Fig. 2 | Prognostic value and expression correlations for the candidate biomarkers in HCC. A Univariate and multivariate Cox regression analysis based on biomarker gene expression levels in the TCGA-LIHC dataset. Sex, age, and TNM stage were the adjustment variables in the multivariate Cox regression. **B** Pearson correlation analysis between biomarker protein expression levels in HCC tissue samples and corresponding serum biomarker levels. **C** IHC images showing biomarker protein expression in HCC tissue samples. Scale bars: upper image, 100 μm; lower image, 50 μm. HCC hepatocellular carcinoma, IHC immunohistochemistry, TCGA The Cancer Genome Atlas, TNM tumor node metastasis.

significant predictive efficacy for sorafenib response (AUC = 0.850) and TACE response (AUC = 0.679) (Supplementary Fig. 4C, D).

Integrated multi-omics characterization in tissue mRNA signature risk stratification

Using the TCGA-LIHC genomic data, we further investigated the differences in genomic, epigenetic, metabolic, and tumor microenvironmental features between the high- and low-risk groups of the tissue mRNA signature. The high-risk group exhibited significantly higher mutation frequencies of TP53 (39% vs. 18%) and KEAP1 (7% vs. 2%) compared to the low-risk group (Fig. 4A, B). Although the intergroup difference in tumor mutation burden (TMB) was not statistically significant ($p = 0.06$), the high-risk group had a higher mean TMB (Supplementary Fig. 5A). Copy number variation (CNV) analysis indicated that the high-risk group displayed
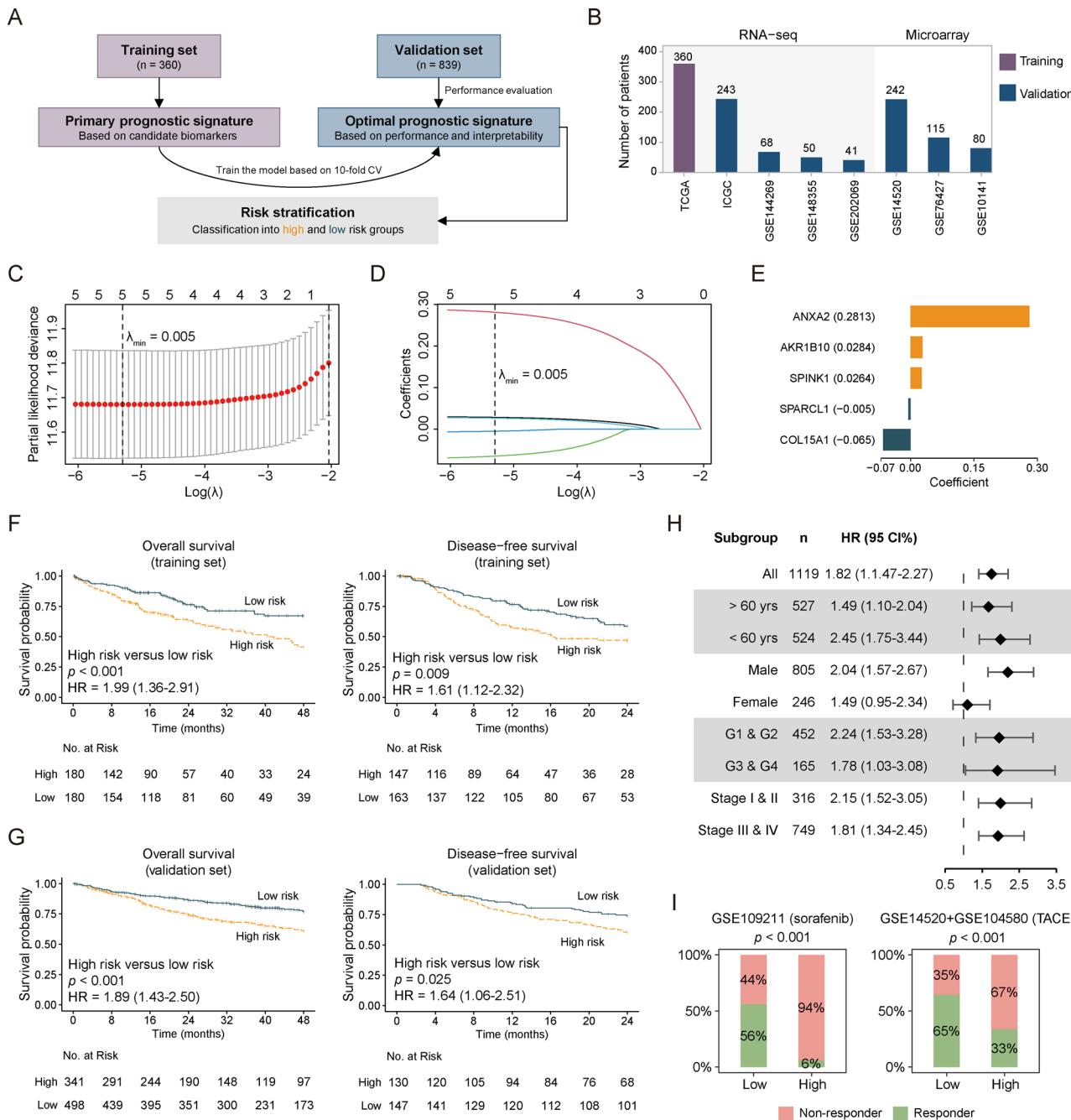
Fig. 3 | Construction and validation of a multi-cohort-based tissue mRNA signature. A Flowchart of tissue mRNA signature development. **B** Bar chart showing patient distribution across technical platforms. **C** Cross-validation for tuning parameter selection in the LASSO-Cox regression model. The x-axis shows $\log(\lambda)$; the y-axis shows partial likelihood deviance. The optimal λ was selected by minimizing the deviance through 10-fold cross-validation. **D** LASSO coefficient shrinkage path plot. The lower x-axis indicates λ values; the upper x-axis indicates the number of retained variables. **E** Bar plot displaying weight coefficients of biomarkers in the tissue mRNA signature. **F** and **G** Kaplan–Meier curves showing statistically significant differences in survival probability between the high-risk (yellow) and low-risk (blue) groups in the training set (**F**, $n = 360$) and validation set (**G**, $n = 839$). **H** Forest plot of univariate Cox regression analyses for specific subgroups. **I** Proportions of responders and non-responders to sorafenib (left) and TACE (right) treatments in high- and low-risk HCC patient groups. HCC hepatocellular carcinoma, LASSO least absolute shrinkage and selection operator, yrs years, TACE transarterial chemoembolization.

amplification patterns in the chromosomal regions 1q22, 2q35, 8q22, and 8q24 (Supplementary Fig. 5B). GSVA enrichment analysis with the MSigDB C1 gene sets revealed significant activation of these CNV regions at the transcriptome level in high-risk patients (Supplementary Fig. 5C).

Based on methods reported in previous studies, an analysis was conducted of the disparities in genomic alteration patterns associated with oncogenic signaling pathways between the high- and low-risk groups[28,29]. Compared to the low-risk group, the high-risk group had significantly higher amplification frequencies in the Hippo ($p < 0.05$) and RTK-RAS ($p < 0.01$) pathways, as well as a higher TP53 pathway mutation rate ($p < 0.001$, Fig. 4C). In contrast, the low-risk group exhibited a significantly higher PI3K pathway mutation rate ($p < 0.05$, Fig. 4C). No significantly different pathways for copy number deletion were identified between the two groups.

Epigenetic regulation analysis revealed that COL15A1 methylation levels were positively correlated with mRNA expression (Fig. 4D). In
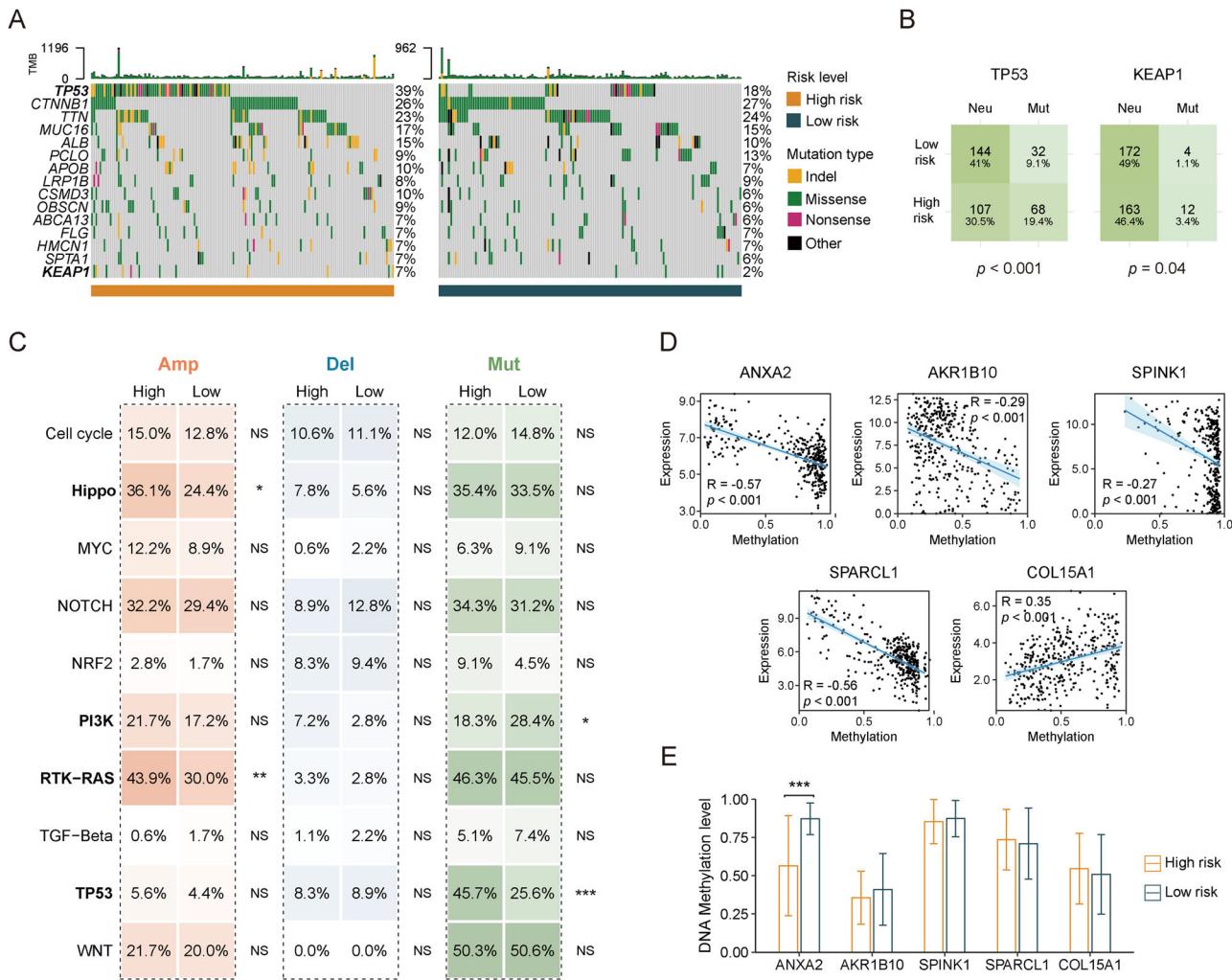
Fig. 4 | Genomic analysis between high- and low-risk groups. A Comparison of mutational landscapes between high- and low-risk groups. **B** Distribution of TP53 mutations and KEAP1 mutations in high- and low-risk groups. **C** Genomic alterations in ten oncogenic pathways between high- and low-risk groups. Color coding indicates alteration type (green: Mut; red: Amp; blue: Del); color intensity represents alteration frequency. **D** Scatter plot of methylation-gene expression correlation. **E** Methylation level comparison between high- and low-risk groups. ***p < 0.001; **p < 0.01; *p < 0.05; NS: *p* ≥ 0.05. Amp amplification, Mut mutation, Del deletion.

contrast, ANXA2, AKR1B10, SPINK1, and SPARCL1 exhibited negative correlations (Fig. 4D). Risk stratification analysis showed that ANXA2 methylation levels were significantly higher in the low-risk group than in the high-risk group (Fig. 4E).

Hallmark gene set enrichment analysis showed that the high-risk group had a significantly activated inflammatory response pathway (NES = 1.64, FDR < 0.001), while the low-risk group had more active bile acid metabolism-related pathways (NES = −2.54, FDR < 0.001) (Fig. 5A). KEGG metabolism-related pathway analyses across multiple HCC cohorts revealed that the risk score was positively correlated with glycan and nucleotide metabolism and negatively correlated with amino acid metabolism (Fig. 5C).

In the tumor microenvironment analysis, assessment using the Tumor Immune Dysfunction and Exclusion (TIDE) algorithm revealed a significant positive correlation between T cell exclusion levels and the risk score ($r = 0.27$, $p < 0.001$, Fig. 5D). The infiltration scores of cancer-associated fibroblasts (CAFs) and myeloid-derived suppressor cells (MDSCs), which possess immunosuppressive functions, were significantly higher in the high-risk group than in the control group (Fig. 5D, E). Analysis via the Scissor algorithm, which integrates single-cell sequencing data, demonstrated that after secondary clustering of the myeloid cell subset, the proportion of MDSCs in Scissor+ cells of the high-risk phenotype was higher than that in negative control cells (Supplementary Fig. 6). Immunophenoscore (IPS)

confirmed that the enrichment scores of MDSCs and regulatory T cells (Tregs) were significantly elevated in the high-risk subgroup compared to the control group, along with increased expression levels of the immune checkpoints PD-1 and CTLA-4 (Supplementary Fig. 7A, B). Furthermore, immune infiltration assessment using cytolytic activity score (CYT), TIMER, MCPcounter, and IPS showed no significant difference in the infiltration levels of CD8+ T cells between the two groups (Supplementary Fig. 7C, D), indicating that the immune feature differences in risk stratification mainly lie in the immunosuppressive microenvironment rather than the effector immune response. Preliminary validation based on the HCC immunotherapy GSE202069 cohort (n = 17) indicated that the high-risk subgroup exhibited a higher immunotherapy response rate, with a higher median risk score in the responder group than in the non-responder group (Supplementary Fig. 7E). These findings suggest that although the high-risk group exhibits immunosuppressive microenvironment characteristics, its higher TMB and preserved effector immune cells may offer immunotherapy opportunities for patients in the high-risk group.

In silico prediction and experimental validation of risk-stratified HCC therapeutics

Numerous studies have shown that drug repositioning based on molecular characteristics has a reliable biological basis[30–32]. To explore the potential
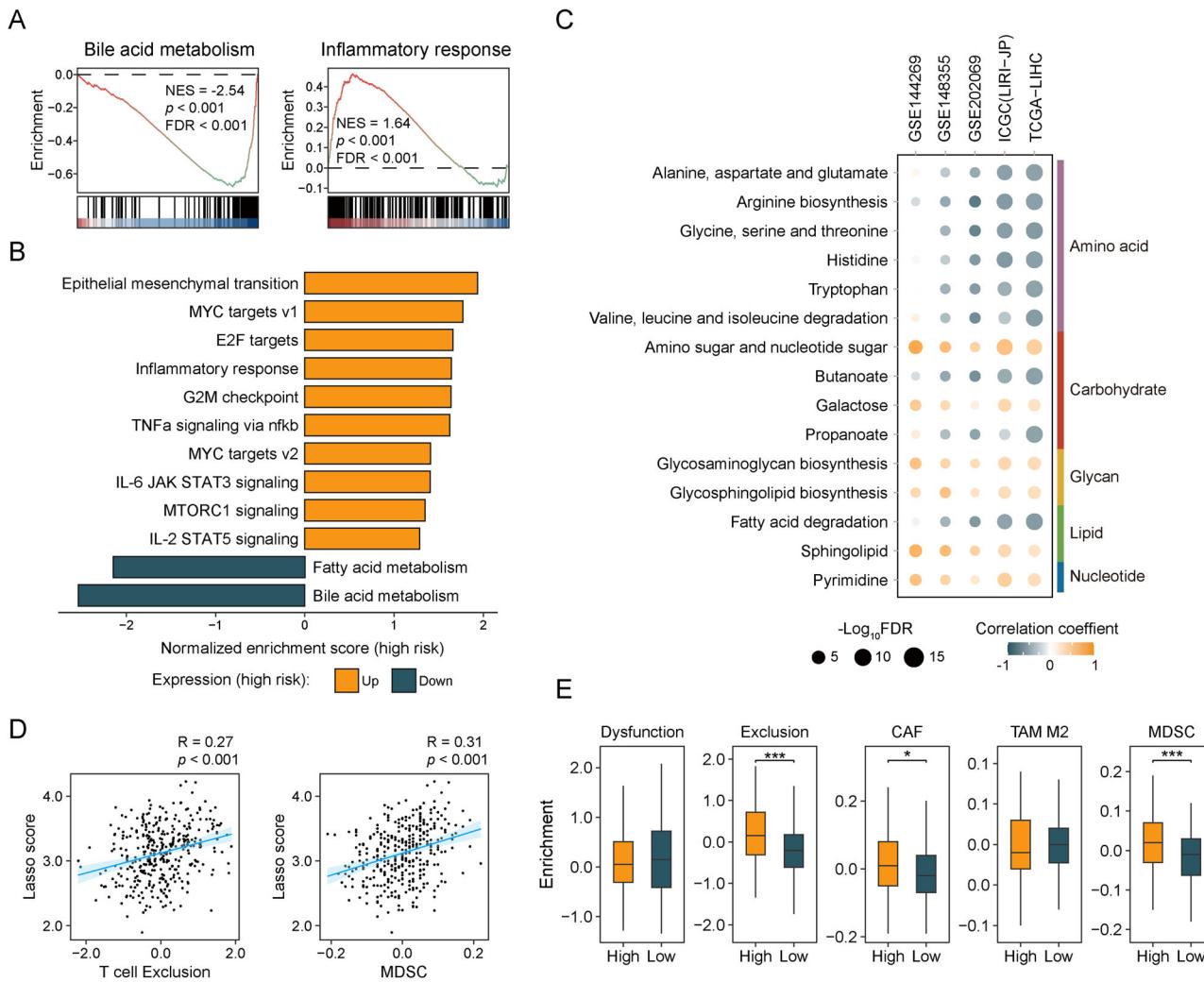
Fig. 5 | Pathway and tumor immune microenvironment analyses between high-and low-risk groups. A GSEA enrichment plot displays pathways significantly downregulated or upregulated in the high-risk group. The statistical results were corrected using the Benjamini–Hochberg method. **B** Bar plot shows upregulated (yellow) and downregulated (blue) Hallmark gene sets in the high-risk group. **C** Dot plot illustrating correlations between risk scores and metabolic pathway activity levels across cohorts. Yellow denotes positive correlation; blue denotes negative correlation. The statistical results were corrected using the Benjamini–Hochberg method. **D** Assessment of the tumor immune microenvironment using the TIDE algorithm, illustrated by scatter plots demonstrating correlations of risk scores with T-cell exclusion and MDSC infiltration. **E** Box plot compares immune features between risk groups. ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$. FDR false discovery rate, GSEA gene set enrichment analysis, NES normalized enrichment score, TIDE tumor immune dysfunction and exclusion, MDSC myeloid-derived suppressor cells.

sensitivity of high-risk HCC patients to existing drugs, we conducted a two-stage drug screening using bioinformatics approaches (Fig. 6A). In the first stage, we used the transcriptional profiles of HCC cell lines from the LIMORE to perform risk stratification through the mRNA signature (Supplementary Fig. 8A). By integrating drug responses, we identified that ABT-263, Afatinib, and Ibrutinib exhibited efficacy differences between risk groups (Fig. 6B). In the second stage, we combined the CTRP, GDSC, and PRISM pharmacogenomic datasets and applied ridge regression to predict drug sensitivity in the TCGA-LIHC cohort. Correlation analysis revealed that the AUC values of Afatinib and ABT-263 were negatively correlated with the risk scores (Fig. 6C). The integrated analysis demonstrated that these two drugs have selective effects on high-risk HCC patients (Fig. 6D, Supplementary Fig. 9A, B). Additionally, the target genes of ABT-263 (BCL2L1/BCL2L2) exhibited significantly elevated expression in the high-risk patients ($p < 0.001$, Fig. 6E, Supplementary Fig. 9C), while Afatinib's target genes showed no significant expression differences between groups (Supplementary Fig. 9D).

Based on the above findings, we selected ABT-263 for experimental validation. Using HCC cell line transcriptomic data from the CCLE, we applied the mRNA signature to conduct risk stratification of HCC cell lines

and integrated stratification information from the LIMORE for verification (Supplementary Fig. 8A, B). Ultimately, we selected 10 cell lines with distinct risk stratification for further study. As shown by colony formation assays, high-risk cells demonstrated significantly reduced colony formation capacity compared to low-risk cells following ABT-263 treatment (Fig. 6F, Supplementary Fig. 10A). CCK-8 assays confirmed that the IC50 of ABT-263 was lower in high-risk cells compared to low-risk cells (Fig. 6G). Flow cytometry analysis revealed that after 24 h of treatment with 2.5 μM ABT-263, high-risk HCC cells showed significantly higher rates of early apoptosis, late apoptosis, and total apoptosis compared to low-risk cells (Fig. 6H, Supplementary Fig. 10B).

In vivo experiments utilized subcutaneous xenograft models of Huh-7 (low-risk) and MHCC97H (high-risk) cell lines in nude mice. Figure 5I indicates that the ABT-263 treatment significantly inhibited tumor growth in the MHCC97H model, with tumor volumes significantly smaller in the treatment group than those in the control group. Conversely, no statistically significant difference in tumor volume was observed between the treatment and control groups in the Huh-7 model. After 20 days of treatment, mice were anesthetized and euthanized, and the tumors were harvested for further analysis. Figure 5J and K indicate that tumor weights were significantly
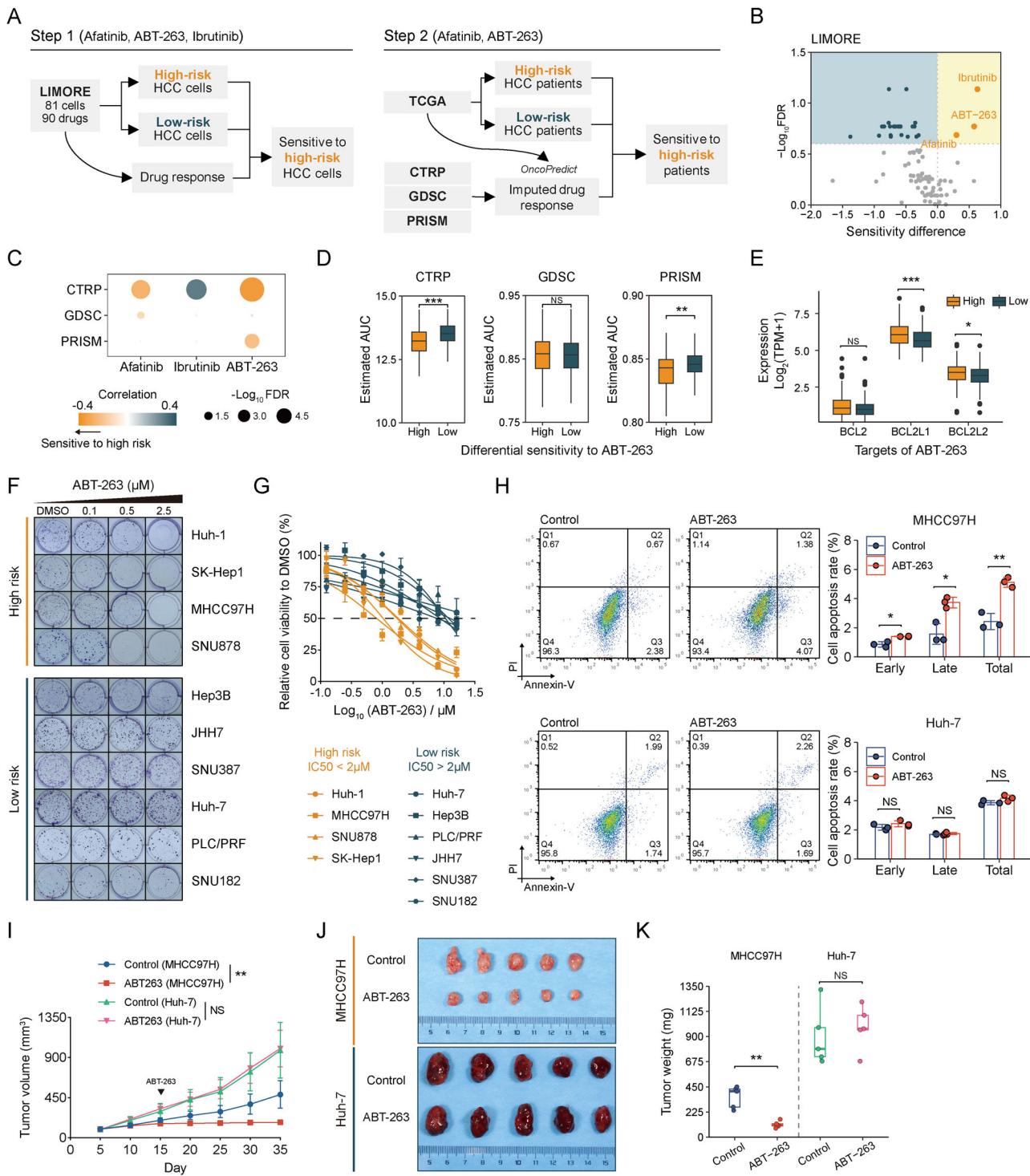
Fig. 6 | Identification of high-risk group-specific therapeutics for HCC.
A Flowchart for in silico drug screening. **B** Differential drug sensitivity analysis between high- and low-risk HCC cell lines based on the LIMORE dataset. Yellow dots denote drugs with higher sensitivity in high-risk cell lines; blue dots denote drugs with higher sensitivity in low-risk cell lines. The statistical results were corrected using the Benjamini–Hochberg method. **C** Correlation of pharmacogenomic database-predicted drug AUC values with risk scores in TCGA-LIHC cohort. Negative correlation (yellow) implies greater drug sensitivity in high-risk patients. The statistical results were corrected using the Benjamini–Hochberg method. **D** Differences in predicted ABT-263 AUC values between risk groups across pharmacogenomic databases for TCGA-LIHC patients. **E** Expression levels of ABT-263 targets (from Drug Repurposing Hub) in high- vs. low-risk groups of TCGA-LIHC. **F** Colony formation assay evaluating responses to ABT-263 in high- and low-

risk HCC cell lines. **G** Cell viability curves and IC_{50} values of ABT-263 in high- and low-risk HCC cell lines. **H** Flow cytometry analysis of apoptosis: proportions of early apoptotic (Annexin V+/PI−) and late apoptotic/necrotic (Annexin V+/PI+) cells in low-risk (Huh7) and high-risk (MHCC-97H) HCC cell lines, untreated or treated with ABT-263 (2 μM, 48 h). **I** Line graph illustrating tumor volume changes in the ABT-263 treatment group and the control group within Huh-7 and MHCC97H xenograft mouse models. **J** Images of tumors collected at the experimental endpoint. **K** Box plot illustrating the comparison of tumor weights at the experimental endpoint. $***p < 0.001$; $**p < 0.01$; $*p < 0.05$; NS: $p \geq 0.05$. FDR false discovery rate, CTRP Cancer Therapeutics Response Portal, GDSC Genomics of Drug Sensitivity in Cancer, HCC hepatocellular carcinoma, IC_{50} half-maximal inhibitory concentration, LIMORE liver cancer model repository, PI propidium iodide, PRISM profiling relative inhibition simultaneously in mixtures, TCGA The Cancer Genome Atlas.
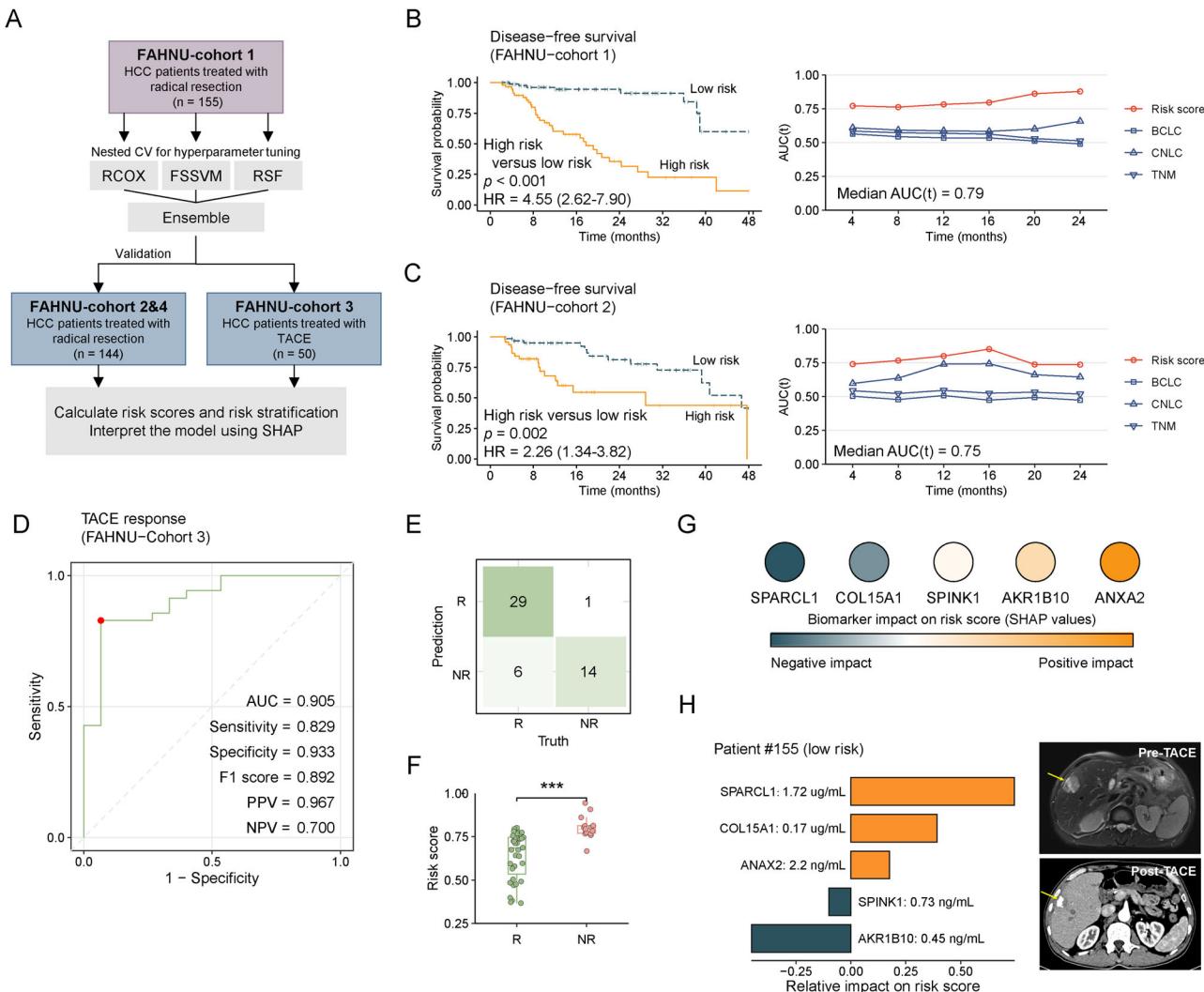
Fig. 7 | Development, validation, and clinical evaluation of the serum protein signature. **A** Flowchart for serum protein signature development and validation. **B** Kaplan–Meier DFS curves for cohort 1 (left); AUC(t) analysis comparing DFS prediction by the serum protein signature versus conventional clinical staging systems (right). **C** Kaplan–Meier DFS analysis for cohort 2 (left); AUC(t) comparison (right). **D** ROC curve assessing the signature's predictive accuracy for TACE clinical benefit. **E** The confusion matrix illustrates the correspondence between the signature-based predictions and the actual clinical outcomes at the optimal cutoff value. **F** Box plot of risk scores in TACE responders vs. non-responders. **G** Dot plot of aggregated SHAP values for global interpretability of the serum protein signature. Elevated protein expression with negative SHAP values (blue) associates with lower risk scores; elevated expression with positive SHAP values (yellow) associates with higher risk scores. **H** Individualized SHAP explanations for representative TACE-responsive HCC cases. Left: Bar plot visualizing direction and magnitude of serum protein contributions to risk predictions. Right: Pre-/post-treatment CT or MRI. ***$p < 0.001$. FAHNU the First Affiliated Hospital of Nanchang University, RCOX regularized Cox, FSSVM fast survival support vector machine, ROC receiver operating characteristic, AUC area under the curve, PPV positive predictive value, NPV negative predictive value, BCLC Barcelona Clinic Liver Cancer, CNLC China Liver Cancer Staging, TNM tumor node metastasis, SHAP Shapley Additive exPlanations, CT computed tomography, MRI magnetic resonance imaging, TACE transarterial chemoembolization.

lower in the ABT-263 treatment group compared to the control group in the MHCC97H model, while no significant reduction was observed in the Huh-7 model.

Collectively, these results indicate that ABT-263 may serve as a promising therapeutic candidate for high-risk HCC patients.

Development, validation, and interpretability analysis of machine learning-driven serum protein signature

As described in the "Methods" section, we established four HCC cohorts based on serum samples from our hospital. Cohort 1 ($n = 155$) was used to develop a serum protein signature, cohort 2 ($n = 109$) for validation, cohort 3 ($n = 50$) to predict the efficacy of TACE, and cohort 4 ($n = 35$) for temporal validation (Fig. 7A). Supplementary Table 3 presents the baseline characteristics of each cohort.

First, we evaluated the prognostic value of each biomarker in predicting postoperative recurrence in cohort 1 (Supplementary Fig. 11A). Univariate Cox regression analysis revealed that ANXA2 (HR = 5.73, 95% CI: 2.70–12.15), AKR1B10 (HR = 4.11, 95% CI: 2.05–8.24), and SPINK1 (HR = 3.09, 95% CI: 1.56–6.24) exhibited a better prognostic capacity than AFP (HR = 2.48, 95% CI: 1.23–5.05). Meanwhile, COL15A1 (HR = 0.44, 95% CI: 0.22–0.88) and SPARCL1 (HR = 0.36, 95% CI: 0.16–0.83) demonstrated significant predictive performance.

Next, we integrated three sub-models based on data from cohort 1 to construct a serum protein biomarker signature and evaluated the predictive performance of each sub-model as well as the protein signature (Fig. 7A). The results showed that (Supplementary Fig. 11B), in cohort 1, the random survival forest (RSF) model exhibited the best predictive performance (AUC: 0.908, 95% CI: 0.845–0.972), followed by the signature (AUC: 0.877,

95% CI: 0.800–0.955). However, the signature outperformed the RSF model in cohort 2 (AUC: 0.736, 95% CI: 0.585–0.888 vs. 0.719, 95% CI: 0.571–0.867).

Then, patients were stratified into high- and low-risk groups using the optimal cutoff value of the risk score from cohort 1 (Supplementary Fig. 12A). Survival analysis demonstrated that in the two postoperative cohorts (cohorts 1 and 2), the high-risk group had significantly shorter DFS than the low-risk group (Fig. 7B, C). Time-dependent receiver operating characteristic curve analysis confirmed that the risk score outperformed clinical tumor staging systems, including Barcelona Clinic Liver Cancer (BCLC), TNM, and China Liver Cancer Staging (CNLC), as well as prognostic scores such as Albumin-Bilirubin (ALBI), Platelet-Albumin-Bilirubin (PALBI), and Aspartate Aminotransferase-to-Platelet Ratio Index (APRI), in predicting early postoperative recurrence of HCC (Fig. 7B, C and Supplementary Fig. 12C, D). Multivariate Cox regression analysis revealed that, after adjusting for confounding factors such as age, sex, microvascular invasion (MVI), and tumor stage, risk stratification was an independent predictor of early HCC recurrence (Supplementary Fig. 12B).

In cohort 3, the risk score exhibited good performance in distinguishing TACE treatment responders, with an AUC of 0.903 (Fig. 7D). At the optimal threshold, 43 out of 50 patients were correctly classified (Fig. 7E), with the sensitivity, specificity, PPV, and NPV of 0.829, 0.933, 0.967, and 0.700, respectively (Fig. 7D). Risk scores were significantly higher in the non-responder group than in the responder group (Fig. 7F). Analysis of the SMOTE-adjusted cohort 3 data confirmed that the predictive performance of the risk score for the TACE treatment response remained stable (Supplementary Fig. 12E).

In cohort 4, we performed temporal validation of the protein signature using prospectively collected samples. The high-risk group exhibited a higher hazard ratio (HR = 2.41) compared to the low-risk group, although there was no significant difference, likely due to the short follow-up period and the relatively few postoperative recurrence events (Supplementary Fig. 12F). Nonetheless, patients who experienced recurrence had significantly higher risk scores than those without recurrence (Supplementary Fig. 12G).

Using cohort 1 data, we applied SHAP to assess the contribution of each biomarker within the serum protein signature. SHAP beeswarm plots for sub-models (Supplementary Fig. 13, left) revealed that biomarkers with identical concentration ranges exhibited heterogeneous SHAP value distributions across individuals, suggesting that their effects are modulated by interactions with other biomarkers. Bootstrap resampling of each sub-model revealed the overall direction of effect for each biomarker at the sub-model level (Supplementary Fig. 13, right). The integration of SHAP analyses across all sub-models demonstrated directional consistency between the biomarkers in the serum protein signature and the tissue mRNA signature (Fig. 7G). Local interpretability analysis further characterized the direction and magnitude of each biomarker's contribution to the signature's prediction in a representative low-risk patient (Fig. 7H).

Discussion

Although tumor tissue-based multi-omics molecular subtyping has provided insights for the precision therapy of HCC, its clinical application remains limited by invasive sampling procedures and cost-prohibitive diagnostic assays. Here, we developed dual signatures as research tools for risk stratification by integrating publicly available multi-omics datasets and in-house serum protein data, using a combination of bioinformatics and machine learning algorithms.

Firstly, we integrated serum protein biomarkers with multi-omics data to identify HCC-related prognostic biomarkers across multiple omics levels. The identified biomarkers (AKR1B10, SPINK1, ANXA2, COL15A1, and SPARCL1) exhibited prognostic significance, which was validated in three independent cohorts, with consistent risk or protective effects across multicenter datasets. Mechanistically, these biomarkers exert diverse regulatory functions in cancer. AKR1B10 promotes a chemotherapy resistance phenotype through metabolic reprogramming, a mechanism validated within the SMARCA4-IRAK1 signaling pathway[33], consistent with our finding that

it is associated with poor prognosis. SPINK1 mediates radiotherapy resistance via the EGFR/Nrf2 pathway[34]. ANXA2 drives MEK/ERK pathway activation via plasma membrane localization and mediates autophagy-related radiotherapy resistance[35,36], reinforcing its clinical significance as a prognostic risk factor. SPARCL1 inhibits invasion and metastasis by mediating the mesenchymal–epithelial transition in colorectal cancer[37], whereas in prostate cancer, it is epigenetically suppressed by androgen receptor signaling[38]; its loss disrupts the extracellular matrix–cytoskeleton mechanical coupling, driving tumor progression. Collectively, these findings support SPARCL1's role as a tumor suppressor, consistent with our observation that its high expression correlates with a favorable prognosis. Clinically, the diagnostic advantage of AKR1B10 in early-stage and AFP-negative HCC cases[39,40], combined with the potential of SPINK1 as a biomarker for radiotherapy resistance[34], demonstrates the potential value of these biomarkers for monitoring tumors in diagnosis, treatment, and prognosis.

Subsequently, we constructed an mRNA signature for HCC based on these biomarkers. We enhanced the signature's generalizability by incorporating multi-etiology and multi-center cohorts and expanded its prognostic prediction function to evaluate responses to TACE and sorafenib therapies. By integrating multi-omics data, we analyzed the association between risk stratification and biological behavior. Our results revealed that high-risk HCC was characterized by TP53 high-frequency mutations, MYC pathway hyperactivation, and tumor-specific metabolic reprogramming. These findings align with the observation by Woo et al. that stem-like gene expression profiles and aggressive phenotypes are present in TP53-mutated HCC[41], suggesting that TP53 mutations may drive malignant progression by preserving tumor cell stemness. Furthermore, reduced fatty acid metabolism in high-risk patients was correlated with an unfavorable prognosis, aligning with the report that it is associated with adverse clinical features, such as advanced clinical stages, larger tumor size, and elevated AFP levels[8]. These findings suggest that HCC may enhance tumor adaptability of cells to microenvironmental stress through lipid metabolic reprogramming, thereby accelerating disease progression.

In tumor microenvironment analysis, the high-risk group displayed an immunosuppressive microenvironment, marked by increased infiltration of T cell exclusion, MDSC, and Treg immune suppressor cells, alongside upregulated immune checkpoint molecules. High MDSC and Treg infiltration has been connected to clinical benefits from the T + A (atezolizumab plus bevacizumab) regimen[42], implying that high-risk HCC patients may better respond to this combination therapy. The biological basis lies in bevacizumab blocking the VEGF pathway to inhibit MDSC tumor homing and inducing vascular normalization to boost T cell infiltration[43]. The IMbrave050 study validated the T + A regimen's clinical worth, cutting post-hepatectomy HCC patients' disease progression risk by 28%[44]. Although we have not validated direct links between risk stratification and therapeutic response, this study offers preliminary guidance for identifying patients with HCC likely to benefit from the T + A regimen.

In drug repositioning, we found that high-risk HCC cells are more sensitive to the Bcl-2/Bcl-xl multitarget inhibitor ABT-263. While ABT-263's single-agent use is limited by thrombocytopenia, it inhibits HCC progression by selectively clearing senescent hepatic stellate cells, providing a rationale for senescence-targeting therapies aimed at the tumor microenvironment[45]. This suggests that mRNA signature-based patient stratification may inform the development of ABT-263-related drug.

To validate the clinical translational potential of these biomarkers, we performed a tissue origin investigation of serum biomarkers and assessed the correlation between circulating protein levels and their corresponding tissue expression in HCC. Paired tissue pathological (IHC) and circulating protein (ELISA) levels demonstrated a significant positive correlation between tissue and serum biomarker expression. This finding confirms the reliability of computational methodologies for inferring serum protein origins from tissue gene expression profiles[25], while underscoring its translational significance: the concordance between tissue and serum biomarker expression establishes a basis for tissue and serum signatures; additionally, the inherent secretory nature of these proteins supports their

translational potential into liquid biopsy applications, providing a basis for future development of minimally invasive detection.

To develop the serum protein signature, we used ELISA to measure the serum protein concentrations and applied machine learning algorithms. ELISA is a classic clinical laboratory assay. Its absolute concentration data meet protein quantification requirements and satisfy the accuracy needs of machine learning algorithms. We developed the protein signature through ensemble learning methodology, wherein the risk scores from three machine learning algorithms were averaged. This retains the predictive efficacy of the single algorithms, enhances generalization through collaborative optimization of multiple algorithms, and shows excellent risk stratification ability. Also, circulating proteins, carrying genomic encoding information, and reflecting environmental exposure-related risk changes, can dynamically monitor individuals' molecular health changes[46–48]. These explain the advantage of the protein signature over the mRNA signature at the biological level. Besides, the SHAP value explanation analysis of the protein signature reveals that the contribution direction of biomarkers is consistent with that in the mRNA signature, confirming the biological rationality of the dual signatures.

Previous studies report costs for common HCC surveillance modalities[49]: abdominal ultrasound $142, AFP $19, and non-contrast CT $274. The standard semiannual ultrasound and AFP surveillance incurs $322/patient/year. In contrast, the five biomarkers evaluated here can be assayed via ELISA for ≤$42/sample, substantially below imaging costs (e.g., CT is about 6.5-fold more expensive than ELISA). However, prospective clinical studies are required for a comprehensive cost-effectiveness assessment. Such studies should evaluate potential benefits, including reducing the frequency of unnecessary imaging and lowering advanced-stage treatment costs through early intervention in high-risk patients.

Although the serum protein signature demonstrated good predictive performance for TACE response, several issues merit consideration before clinical application. The signature showed favorable specificity and PPV, suggesting that patients predicted to benefit from TACE are likely to respond. However, its relatively low sensitivity and NPV indicate a risk of false-negative predictions, whereby patients who might respond to TACE could be incorrectly excluded from this treatment and thus lose the opportunity for potential survival benefits. Therefore, for patients classified as non-responders by the model, clinical decision-making regarding TACE should be made with caution, taking into account other relevant clinical factors. Given that TACE response prediction represents a secondary application of this signature and considering limitations such as its single-center design, limited sample size, and class imbalance, further validation in large-scale multicenter studies is required.

In clinical practice, the serum protein signature enables preoperative risk stratification, thereby informing treatment planning and prognosis evaluation. If a patient is considered suitable for biopsy or surgery, tissue samples can be obtained for further evaluation using the tissue mRNA signature. Integrating the assessments from both signatures (Supplementary Fig. 12I) into clinical decision-making may enable more precise clinical strategies.

This study has several limitations. First, the serum protein signature was developed and validated in single-center cohorts, and its generalizability requires prospective multicenter validation. Second, we mainly focused on biomarker selection and their application, without investigating the mechanism by which the biomarkers are secreted from tissue to blood. Future studies should investigate the specific secretion pathways (e.g., the exosome-mediated pathway). Third, due to the limited availability of HCC tissue samples, we were unable to perform a paired analysis linking tissue mRNA with circulating protein. Previous studies[23,24] have demonstrated that paired tissue–serum samples are critical for tracing the tissue origin of circulating biomarkers. Moreover, multimodal models have been shown to provide complementary prognostic information and achieve better predictive performance compared with single-modality models[50]. Considering the potential complementarity between serum protein signatures (enabling minimally invasive monitoring) and tissue mRNA signatures (providing mechanistic insights), paired tissue-serum studies should be performed in the future. Such integration may not only improve predictive accuracy and clinical utility but also advance understanding of the mechanisms underlying treatment response.

In conclusion, we developed dual signatures for HCC prognostic assessment by integrating novel tissue multi-omics data with serum protein biomarkers. The dual signatures enable risk stratification while revealing preliminary molecular correlates of risk strata, offering a potentially actionable tool for precision oncology in HCC. While our dual signature approach shows promise, several limitations need to be addressed before clinical implementation, including external validation, standardization protocols, and prospective evaluation.

Methods
Clinical sample acquisition
Serum samples were obtained from the First Affiliated Hospital of Nanchang University (FAHNU) between January 2020 and May 2025, and their serum biomarker levels were determined using the ELISA method. The study cohorts were constructed as follows: patients who underwent curative resection for HCC were selected from FAHNU. Cohort 1 ($n = 155$) was used to develop the serum protein signature (including 35 patients with paraffin-embedded tumor tissue sections for immunohistochemical analysis). Cohort 2 ($n = 109$) was used to validate the protein signature. Cohort 3 ($n = 50$, received TACE) was used to evaluate the value of the protein signature for the treatment response. Cohort 4 ($n = 35$) served as a prospectively collected temporal validation cohort. Cohorts 1–3 were collected from January 2020 to October 2024, while cohort 4 was collected from October 2024 to May 2025. Clinical characteristics and baseline information for all cohorts are presented in Supplementary Tables 3 and 4.

This study was approved by the Medical Research Ethics Committee of the First Affiliated Hospital of Nanchang University (ethics approval number: (2023)CDYFYYLK(01-070)) and was conducted in accordance with the Declaration of Helsinki and relevant guidelines and regulations. All participants provided written informed consent to participate in the study. Samples were anonymized and identified only by study codes for analysis.

Patient treatment response assessment
TACE treatment response was evaluated according to the Response Evaluation Criteria in Solid Tumors (RECIST 1.1), assessed by comparing the maximum diameter of lesions in pre- and post-treatment CT or MRI scans. Two researchers independently evaluated the medical imaging data, categorizing patients into response group (complete response or partial response) and non-response group (stable disease or progressive disease) based on the objective response rate. The classifications were validated by a radiologist. Of the 50 patients included in the study, 35 demonstrated a treatment response, while 15 showed no response.

Public data collection and processing
We re-analyzed a total of 36 publicly available datasets, comprising 21 microarray datasets, 11 RNA-seq datasets, 3 proteomics datasets, and 1 single-cell RNA-seq dataset. The samples were derived from tumor tissues, adjacent normal tissues (ANT) of HCC patients, normal liver tissues, serum samples, as well as tissues from mouse HCC models, normal mouse tissues, and human HCC cell lines.

Data from TCGA and ICGC were downloaded via the Xena platform (https://xenabrowser.net/datapages/). The remaining datasets were sourced from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/), ArrayExpress (https://www.ebi.ac.uk/arrayexpress/), or supplementary materials of relevant publications. Detailed information for all datasets is provided in Supplementary Table 5.

For the tissue mRNA signature, we normalized the gene expression. For batch effect correction, we applied the ComBat function (R package sva v.3.48.0)[51] to five tumor microarray datasets (GSE14520, GSE76427, GSE109211, GSE104580, and GSE10141), with the parameters set as: par.prior = TRUE, mean.only = FALSE. The batch-corrected microarray

datasets were subsequently used as the reference set for cross-platform normalization. Specifically, we employed the training distribution matching (TDM) method (R package TDM v.0.3)[52] to normalize seven RNA-seq datasets, including TCGA-LIHC, ICGC (LIRI-JP), GSE144269, GSE148355, GSE202069, CCLE, and LIMORE, using the following parameter settings: filter_p = FALSE, inv_reference = TRUE, log_target = TRUE, tdm_factor = 1. This process yielded a cross-platform normalized expression matrix. Normalized expression matrices have been deposited in the Zenodo repository (doi.org/10.5281/zenodo.16469089).

Human HCC cell line culture
We utilized ten human HCC cell lines in this study. Huh-1, JHH7, and SNU878 were obtained from Shanghai Zhongqiao Xinzhou Biotechnology Co., Ltd. (China); SK-Hep1, PLC/PRF/5, SNU387, and SNU182 were obtained from Zhejiang BaiDi Biotechnology Co., Ltd. (China); and Hep3B, MHCC97H, and Huh-7 were obtained from the National Collection of Authenticated Cell Cultures (China). All cell lines were authenticated using short tandem repeat profiling to ensure their identity. The cell lines were cultured in basal medium supplemented with 10% fetal bovine serum and 0.1% penicillin–streptomycin mixture (100 U/mL penicillin and 100 μg/mL streptomycin). The specific culture media used were as follows: high-glucose DMEM for SK-Hep1, Huh-1, Huh-7, JHH7, and MHCC-97H; RPMI 1640 for SNU878, SNU387, and SNU182; and MEM for Hep3B and PLC/PRF/5. Cultures were maintained in a cell culture incubator at 37 °C with 5% CO_2.

RNA extraction and quantitative real-time PCR (qPCR)
Total RNA was extracted from cells using the MolPure Cell Total RNA Kit (#19221ES60; YEASEN, China) following the manufacturer's instructions. Reverse transcription to cDNA was performed with Hifair RT SuperMix (#11151ES60; YEASEN, China). qPCR assays were conducted using SYBR Green Master Mix (#11184ES08; YEASEN, China) on a StepOnePlus Real-Time PCR System (Thermo Fisher Scientific, USA). Relative gene expression was normalized to GAPDH and calculated via the $2^{(-\Delta\Delta C)}$ method. Primer sequences are provided in Supplementary Table 6.

Western blot analysis
Total protein was extracted following cell lysis in RIPA buffer supplemented with protease inhibitors and quantified using a BCA assay kit. Samples were mixed with loading buffer, denatured by boiling, resolved via sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and transferred to Nitrocellulose membranes. After blocking with 5% skimmed milk, primary antibody incubation was performed overnight at 4 °C using anti-ANXA2 (1:20,000; #11256-1-AP) and anti-AKR1B10 (1:2000; #18252-1-AP) (Proteintech, China). Membranes were then incubated with horseradish peroxidase (HRP)-conjugated secondary antibodies. Protein bands were visualized by enhanced chemiluminescence using an iBright FL1000 imaging system (Thermo Fisher Scientific, USA).

Cell transfection
Plasmids for the overexpression of ANXA2 and AKR1B10 were constructed by the Public Protein Plasmid Library (China), while siRNAs for gene knockdown were synthesized by GenePharma (China). siRNA transfection was performed using Lipofectamine RNAiMAX (Invitrogen, USA), and overexpression plasmids were transfected using jetPRIME (Polyplus Transfection, USA), following the manufacturer's instructions. Human HCC cells were transfected accordingly, with empty vector-transfected cells used as controls. The siRNA sequences used for gene knockdown are listed in Supplementary Table 7.

Cell invasion assay
24-well Transwell chambers with polycarbonate membranes pre-coated with Matrigel (Corning, #356237, USA) were placed into cell culture plates. A total of 600 μl complete medium was added to the lower chamber, while 4×10^4 cells suspended in 200 μl serum-free medium were seeded into the upper chamber. Cells were incubated at 37 °C in a humidified atmosphere

with 5% CO_2 for 48 h. After incubation, non-invading cells on the upper surface of the membrane were gently removed with a cotton swab. Cells that had invaded the lower surface were fixed with 4% paraformaldehyde and stained with 0.1% crystal violet. Images were captured under an inverted phase-contrast microscope (Nikon Eclipse Ti, Japan) at 200× magnification, and cells were counted in three randomly selected fields.

Colony formation assay
Cells were seeded at a density of $1–2 \times 10^3$ cells per well in a six-well plate and cultured in complete medium for 24 h. Subsequently, the medium was replaced with a complete medium containing 0.1, 0.5, or 2.5 μM of ABT-263 (#S1001, Selleck, China) or DMSO. The medium containing the corresponding drug concentration was refreshed every 3 days for a total treatment period of 10–14 days. At the end of the experiment, the medium was removed, and the cells were fixed with 4% paraformaldehyde. The cells were then stained with a 0.1% crystal violet solution. After staining, the background dye was gently washed away with running water. Colonies were counted after drying.

Cell viability assay
Cells were seeded at a density of 5×10^3 cells per well in a 96-well plate and cultured overnight. On the following day, the medium was replaced with complete medium containing 0.125, 0.25, 0.5, 1, 2, 4, 8, or 16 μM of ABT-263 (#S1001, Selleck, China). The cells were then incubated for 72 h. After the drug treatment, CCK-8 reagent was added, and the plate was incubated at 37 °C in the dark for 3 hours. The absorbance at 450 nm was measured using a microplate reader. A four-parameter logistic regression model was used to fit the dose-response curve and calculate the half-maximal inhibitory concentration (IC_{50}).

Flow cytometry
Cell apoptosis was evaluated using an apoptosis detection kit (#BB-4101, BestBio, China) following the manufacturer's instructions. After treatment with 2 μM ABT-263 for 48 h, cells were harvested, washed with pre-cooled PBS, and resuspended to form a single-cell suspension. The cell suspension was then mixed with FITC-Annexin V and propidium iodide and incubated at room temperature in the dark. Apoptosis was assessed using a FACS-Canto II (BD Biosciences, USA) flow cytometer. Data analysis was performed using FlowJo (v.10.4) to distinguish live cells (Annexin V⁻/PI⁻), early apoptotic cells (Annexin V⁺/PI⁻), and late apoptotic/necrotic cells (Annexin V⁺/PI⁺).

Immunohistochemistry (IHC)
Paraffin-embedded HCC tissue samples were obtained from the Department of Pathology at the First Affiliated Hospital of Nanchang University. The corresponding paraffin blocks and pathological diagnostic reports were retrieved through the hospital's pathology information system. The paraffin-embedded tissue blocks were sectioned continuously and mounted onto slides. The sections were processed sequentially for dewaxing and rehydration, antigen retrieval, blocking of endogenous peroxidase activity, and blocking. Subsequently, the sections were incubated overnight at 4 °C with the following primary antibodies: AKR1B10 antibody (1:100; #121558, Zenbio, China), SPINK1 antibody (1:100; #R389023, Zenbio, China), ANXA2 antibody (1:100; #861473, Zenbio, China), COL15A1 antibody (1:100; #PK13762, Abmart, China), and SPARCL1 antibody (1:200; #13517-1-AP, Proteintech, China). The next day, the sections were incubated with biotinylated secondary antibodies, developed with 3,3'-diaminobenzidine, counterstained with hematoxylin for nuclei, and mounted with neutral gum.

Animal experiments
Huh-7 or MHCC97H cells (2.5×10^6 cells per mouse) were subcutaneously injected into the right axillary region of 6-week-old female BALB/c nude mice. When tumor volumes reached ~80 mm³, the mice were randomized to receive either ABT-263 (100 mg/kg/day administered by oral gavage) or vehicle control for 20 days. Tumor volumes were measured using the

formula: volume = (width² × length)/2. The mice were anesthetized with an intraperitoneal injection of sodium pentobarbital (50 mg/kg) and euthanized by cervical dislocation after unconsciousness was confirmed. All animal experiments were conducted under specific pathogen-free conditions at the Experimental Animal Center of the First Affiliated Hospital of Nanchang University and in accordance with protocols approved by the Institutional Animal Care and Use Committee of the hospital (ethics approval number: CDYFY-IACUC-202507GR023).

Enzyme-linked immunosorbent assay (ELISA)

Serum biomarker quantification was performed using ELISA kits (AKR1B10 #EH1225, ANXA2 #EH2012, SPINK1 #EH0276, SPARCL1 #EH0275, and COL15A1 #EH0540; FineTest, China) according to the manufacturer's protocols. The validation data from each lot of the kits, along with the serum dilution ratios for each biomarker, are detailed in the Supplementary Table 8. All samples were assayed in duplicate. Each batch incorporated quality control (QC) samples, and the inter-assay coefficient of variation (CV) was required to be ≤15%. Batches were reassayed if any QC value exceeded predefined limits. Each 96-well plate contained an 8-point calibration curve, QC samples, blank controls, and diluted serum samples. Absorbance at 450 nm was measured using a Varioskan Flash multimode reader (Thermo Scientific, USA), and concentrations were calculated via four-parameter logistic curve fitting. Data were corrected for batch effects using the ELISAtools (R package v.0.1.8)[53]. Additional quality control measures included pre-experimental validation of kit performance (standard curve $R^2 \geq 0.99$) and re-assaying absorbance values outside the linear range of the calibration curve.

Construction of the tissue mRNA signature

The tissue mRNA signature was constructed using normalized transcriptome data and clinical survival data from the TCGA-LIHC cohort. The expression matrices of five biomarkers (AKR1B10, ANXA2, COL15A1, SPARCL1, and SPINK1) were selected as input variables. A LASSO-Cox signature was developed using the glmnet (R package v.4.1-8)[54,55]. The signature parameters were optimized through 10-fold cross-validation, with deviance serving as the evaluation metric to determine the optimal regularization parameter (lambda.min). This parameter was used to derive the following formula:

$$\text{Risk score} = \sum_{j=1}^{k} \text{coef}(j) \times \text{Exp}(j) \qquad (1)$$

The formula (1) was then applied to calculate the risk scores for each patient. The risk score was used to stratify patients into high- and low-risk groups using the survminer (R package v.0.4.8), and its optimal cutoff value for grouping was determined by the maximally selected rank statistics algorithm. The HCC cohorts used for the analysis included six cohorts with prognostic information (TCGA-LIHC, ICGC (LIRI-JP), GSE14520, GSE76427, GSE10141, and GSE144269), two cohorts with treatment information (GSE148355 and GSE202069), and two cohorts with both treatment and prognostic information (GSE109211 and GSE104580). Risk stratification of HCC cell lines was performed using data from the CCLE and LIMORE.

Construction of the serum protein signature

Serum concentrations of five biomarkers (AKR1B10, ANXA2, COL15A1, SPARCL1, and SPINK1) quantified via ELISA were utilized as input data to develop three survival sub-models: regularized Cox (RCOX)[56], fast survival support vector machine (FSSVM)[57], and random survival forest (RSF)[58], using the scikit-survival (Python package v.0.24.1)[59], with disease-free survival (DFS) and survival status as clinical endpoints. During the data preprocessing phase, the input features for the RCOX model were standardized using Z-score normalization, while the other models used the original serum concentration data. Then, we performed nested cross-validation in cohort 1

to tune the hyperparameters for each sub-model. The outer loop employed 5-fold cross-validation to evaluate model performance, and the inner loop used 4-fold cross-validation to select hyperparameters. In each outer fold, the dataset was split into training and testing subsets. Within the outer training set, inner cross-validation was conducted using the concordance index (C-index) as the evaluation metric. The average C-index across the inner validation folds was then calculated for each hyperparameter combination, and the combination with the highest average C-index was selected. The model was then trained on the entire outer training set using the optimal hyperparameters, and its performance was evaluated on the corresponding outer test set using the C-index. Finally, the optimal hyperparameter was determined by evaluating C-index values from the inner and outer loops.

Given the differences in the scales of risk scores among the sub-models, we normalized the risk scores of each cohort using min–max normalization based on the extreme values of cohort 1, as referenced in previous studies[60,61]. The arithmetic mean of the normalized scores from the three sub-models was calculated as the final risk score. The optimal cutoff value for risk stratification in cohort 1 was determined using the maximally selected rank statistics algorithm, which was then applied to stratify patients in all cohorts into high- and low-risk groups.

When predicting TACE therapeutic outcomes using the risk score in cohort 3, we reported the area under the receiver operating characteristic curve (AUC) value, sensitivity, specificity, F1 score, positive predictive value (PPV), and negative predictive value (NPV) at the optimal threshold to account for class imbalance in treatment response. Furthermore, the synthetic minority over-sampling technique (SMOTE) was applied to process cohort 3 data and to mitigate potential evaluation bias induced by imbalanced class distribution[62].

Model interpretability

To analyze the interpretability of the serum protein signature, we employed the SHapley Additive exPlanations (SHAP) framework (Python package v.0.45.1)[63]. This framework elucidates feature importance and influence direction through global and local explanations. For the three sub-models of the signature, we matched corresponding SHAP explainers. Specifically, the Explainer function was applied to RCOX and RSF, while the KernelExplainer function was used for FSSVM.

In global interpretation, SHAP analysis was implemented with 50-iteration bootstrap resampling for each model. For each bootstrap sample, the corresponding model's adapted SHAP interpreter was applied to compute the SHAP values for all instances. The mean absolute SHAP value of each feature was calculated for each iteration. The results from all 50 iterations were aggregated to estimate the average feature importance, along with the corresponding 95% CI. Then, we applied a two-step normalization process to the SHAP values output by each algorithm. First, we performed min-max normalization based on the extreme values of the training set. Second, we determined the influence direction based on the Spearman correlation coefficient between the original feature values and the SHAP values (two-sided test, with $p < 0.05$ considered significant)[61,64]. If the correlation was significant and negative, we inverted the normalized values. The aggregated SHAP values of the integrated model were obtained by integrating the normalized SHAP values of the three algorithms.

In local interpretation, we performed sample-level min-max normalization based on the extreme values of cohort 1 for individual patient samples. The direction of feature contribution was quantified by combining the signs of the SHAP values. The sample-level interpretation results of the three sub-models were integrated using the arithmetic mean method.

All normalizations and direction determinations mentioned above were implemented using native functions of the SHAP library.

Bioinformatics analysis

We conducted bioinformatics analyses using R (v.4.3.1) and Bioconductor (v.3.20) packages (e.g., maftools v2.16.0, clusterProfiler v.4.8.3, DESeq2 v.1.40.2) to assess features and heterogeneity related to genomic variation,

epigenetic regulation, metabolic reprogramming, and the tumor micro-environment. Additionally, we identified potential therapeutic drugs for high-risk HCC patients by integrating pharmacogenomic data. The specific analysis workflow adhered to the standard guidelines outlined in the official documentation of each software package. Detailed analysis parameter settings are provided in the Supplementary Notes 1–9.

Statistical analysis

All experiments were conducted at least three times. Quantitative data were presented as mean ± standard deviation, two-group comparisons were analyzed using Student's t-test or the Mann–Whitney U test, and multiple-group comparisons were performed using the Kruskal–Wallis H test. Categorical variables were analyzed using the Pearson χ^2 test or Fisher's exact test. Correlation analysis was performed using Pearson's or Spearman's correlation coefficients. Survival curves were plotted using the Kaplan–Meier method, and differences between groups were assessed using the log-rank test. Statistical calculations and model construction were performed using R (v.4.3.1), GraphPad Prism (v.9.6.2), or Python (v.3.12.0). P values obtained from differential expression analysis during biomarker selection, drug sensitivity prediction, and pathway enrichment analysis were corrected for multiple testing using the Benjamini–Hochberg false discovery rate (FDR). A two-sided P value of less than 0.05 was considered statistically significant.

Data availability

TCGA and ICGC datasets can be accessed via the Xena platform (https://xenabrowser.net/datapages/), while other public datasets are downloadable from GEO (http://www.ncbi.nlm.nih.gov/geo/) and ArrayExpress (https://www.ebi.ac.uk/arrayexpress/). All accession numbers and sources for these public datasets are detailed in Supplementary Table 5. The ELISA data and cross-platform normalized data used to train the dual signatures have been deposited in Zenodo (https://doi.org/10.5281/zenodo.16469089).

Code availability

The code used for signature development is publicly available in the Zenodo repository (https://doi.org/10.5281/zenodo.16469089). Additional code for bioinformatics analyses and data processing is available from the corresponding author (K.-H.Z.) upon reasonable request via email.

References

1. Sung, H. et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
2. Llovet, J. M. et al. Hepatocellular carcinoma. *Nat. Rev. Dis. Prim.* **7**, 6 (2021).
3. Alizadeh, A. A. et al. Toward understanding and exploiting tumor heterogeneity. *Nat. Med.* **21**, 846–853 (2015).
4. Caruso, S., O'Brien, D. R., Cleary, S. P., Roberts, L. R. & Zucman-Rossi, J. Genetics of hepatocellular carcinoma: approaches to explore molecular diversity. *Hepatology* **73**, 14–26 (2021).
5. Ally, A. et al. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell* **169**, 1327–1341.e23 (2017).
6. Chen, L. et al. Deep whole-genome analysis of 494 hepatocellular carcinomas. *Nature* **627**, 586–593 (2024).
7. Gao, Q. et al. Integrated proteogenomic characterization of HBV-related hepatocellular carcinoma. *Cell* **179**, 561–577.e22 (2019).
8. Li, B. et al. Multiomics identifies metabolic subtypes based on fatty acid degradation allocating personalized treatment in hepatocellular carcinoma. *Hepatology* **79**, 289–306 (2024).
9. Vellanki, P. J. et al. Regulatory implications of ctDNA in immuno-oncology for solid tumors. *J. Immunother. Cancer* **11**, e005344 (2023).
10. Ding, Z., Wang, N., Ji, N. & Chen, Z.-S. Proteomics technologies for cancer liquid biopsies. *Mol. Cancer* **21**, 53 (2022).
11. Yao, L.-Q. et al. Prognostic value of serum α-fetoprotein level as an important characteristic of tumor biology for patients undergoing liver resection of early-stage hepatocellular carcinoma (BCLC Stage 0/A): a large multicenter analysis. *Ann. Surg. Oncol.* **31**, 1219–1231 (2024).
12. Chen, S.-C. et al. PIVKA-II as a surrogate biomarker for therapeutic response in non-AFP-secreting hepatocellular carcinoma. *BMC Cancer* **25**, 199 (2025).
13. Wang, S.-Y. et al. Prothrombin induced by vitamin K absence or antagonist-II (PIVKA-II) predicts complete responses of transarterial chemoembolization for hepatocellular carcinoma. *J. Formos. Med. Assoc.* **121**, 1579–1587 (2022).
14. Costentin, C. et al. R3-AFP score is a new composite tool to refine prediction of hepatocellular carcinoma recurrence after liver transplantation. *JHEP Rep.* **4**, 100445 (2022).
15. Lee, J.-H. et al. Serum tumor markers provide refined prognostication in selecting liver transplantation candidate for hepatocellular carcinoma patients beyond the Milan criteria. *Ann. Surg.* **263**, 842–850 (2016).
16. Mehta, N. et al. Validation of a risk estimation of tumor recurrence after transplant (RETREAT) score for hepatocellular carcinoma recurrence after liver transplant. *JAMA Oncol.* **3**, 493–500 (2017).
17. Xu, R.-H. et al. Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. *Nat. Mater.* **16**, 1155–1161 (2017).
18. Guo, D.-Z. et al. Early detection and prognosis evaluation for hepatocellular carcinoma by circulating tumour DNA methylation: a multicentre cohort study. *Clin. Transl. Med* **14**, e1652 (2024).
19. Sogbe, M. et al. Prognostic value of ultra-low-pass whole-genome sequencing of circulating tumor DNA in hepatocellular carcinoma under systemic treatment. *Clin. Mol. Hepatol.* **30**, 177–190 (2024).
20. Chen, V. L., Xu, D., Wicha, M. S., Lok, A. S. & Parikh, N. D. Utility of liquid biopsy analysis in detection of hepatocellular carcinoma, determination of prognosis, and disease monitoring: a systematic review. *Clin. Gastroenterol. Hepatol.* **18**, 2879–2902.e9 (2020).
21. Park, J. et al. Liquid biopsy in hepatocellular carcinoma: challenges, advances, and clinical implications. *Clin. Mol. Hepatol.* **31**, S255–S284 (2025).
22. Xing, X. et al. Proteomics-driven noninvasive screening of circulating serum protein panels for the early diagnosis of hepatocellular carcinoma. *Nat. Commun.* **14**, 8392 (2023).
23. Niu, L. et al. Noninvasive proteomic biomarkers for alcohol-related liver disease. *Nat. Med.* **28**, 1277–1287 (2022).
24. Govaere, O. et al. A proteo-transcriptomic map of non-alcoholic fatty liver disease signatures. *Nat. Metab.* **5**, 572–578 (2023).
25. Fujiwara, N. et al. A blood-based prognostic liver secretome signature and long-term hepatocellular carcinoma risk in advanced liver fibrosis. *Med* **2**, 836–850.e10 (2021).
26. Wang, T. et al. A universal strategy designed for selecting bench-to-bedside aptamers to serum and validated in hepatocellular carcinoma diagnosis. *Microchem. J.* **150**, 104152 (2019).
27. Zhou, J.-M. et al. Aptamer-based triple serum fluorescence intensity assay: a novel and feasible method for the clinical diagnosis of primary hepatic carcinoma. *Front. Oncol.* **12**, 897775 (2022).
28. Sanchez-Vega, F. et al. Oncogenic signaling pathways in the Cancer Genome Atlas. *Cell* **173**, 321–337.e10 (2018).
29. Gong, Y. et al. Metabolic-pathway-based subtyping of triple-negative breast cancer reveals potential therapeutic targets. *Cell Metab.* **33**, 51–64.e9 (2021).
30. Li, Y. et al. UPP1 promotes lung adenocarcinoma progression through the induction of an immunosuppressive microenvironment. *Nat. Commun.* **15**, 1200 (2024).

31. Fu, S. et al. Development of a stemness-related prognostic index to provide therapeutic strategies for bladder cancer. *NPJ Precis. Oncol.* **8**, 14 (2024).

32. Pushpakom, S. et al. Drug repurposing: progress, challenges and recommendations. *Nat. Rev. Drug Discov.* **18**, 41–58 (2019).

33. Kim, S. Y. et al. SMARCA4 oncogenic potential via IRAK1 enhancer to activate Gankyrin and AKR1B10 in liver cancer. *Oncogene* **40**, 4652–4662 (2021).

34. Suwa, T. et al. SPINK1 as a plasma marker for tumor hypoxia and a therapeutic target for radiosensitization. *JCI Insight* **6**, e148135 (2021).

35. Liu, Z.-Y. et al. FBXW10-S6K1 promotes ANXA2 polyubiquitination and KRAS activation to drive hepatocellular carcinoma development in males. *Cancer Lett.* **566**, 216257 (2023).

36. Gao, Y. et al. Integrin β6/Annexin A2 axis triggers autophagy to orchestrate hepatocellular carcinoma radioresistance. *Cell Death Differ.* **32**, 689–701 (2025).

37. Hu, H. et al. Secreted protein acidic and rich in cysteines-like 1 suppresses aggressiveness and predicts better survival in colorectal cancers. *Clin. Cancer Res.* **18**, 5438–5448 (2012).

38. Hurley, P. J. et al. Androgen regulated SPARCL1 in the tumor microenvironment inhibits metastatic progression. *Cancer Res.* **75**, 4322–4334 (2015).

39. Xie, C., Ye, X., Zeng, L., Zeng, X. & Cao, D. Serum AKR1B10 as an indicator of unfavorable survival of hepatocellular carcinoma. *J. Gastroenterol.* **58**, 1030–1042 (2023).

40. Ye, X. et al. A Large-scale multicenter study validates aldo-keto reductase family 1 member B10 as a prevalent serum marker for detection of hepatocellular carcinoma. *Hepatology* **69**, 2489–2501 (2019).

41. Woo, H. G. et al. Association of TP53 mutations with stem cell-like gene expression and survival of patients with hepatocellular carcinoma. *Gastroenterology* **140**, 1063–1070 (2011).

42. Zhu, A. X. et al. Molecular correlates of clinical response and resistance to atezolizumab in combination with bevacizumab in advanced hepatocellular carcinoma. *Nat. Med.* **28**, 1599–1611 (2022).

43. Llovet, J. M. et al. Immunotherapies for hepatocellular carcinoma. *Nat. Rev. Clin. Oncol.* **19**, 151–172 (2022).

44. Qin, S. et al. Atezolizumab plus bevacizumab versus active surveillance in patients with resected or ablated high-risk hepatocellular carcinoma (IMbrave050): a randomised, open-label, multicentre, phase 3 trial. *Lancet* **402**, 1835–1847 (2023).

45. Li, F. et al. FBP1 loss disrupts liver metabolism and promotes tumourigenesis through a hepatic stellate cell senescence secretome. *Nat. Cell Biol.* **22**, 728–739 (2020).

46. Geyer, P. E. et al. Plasma proteome profiling to assess human health and disease. *Cell Syst.* **2**, 185–195 (2016).

47. Williams, S. A. et al. Plasma protein patterns as comprehensive indicators of health. *Nat. Med.* **25**, 1851–1857 (2019).

48. Polygenic Risk Score Task Force of the International Common Disease Alliance et al. Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps. *Nat. Med.* **27**, 1876–1884 (2021).

49. Parikh, N. D., Singal, A. G., Hutton, D. W. & Tapper, E. B. Cost-Effectiveness of hepatocellular carcinoma surveillance: an assessment of benefits and harms. *Am. J. Gastroenterol.* **115**, 1642–1649 (2020).

50. Boehm, K. M. et al. Multimodal data integration using machine learning improves risk stratification of high-grade serous ovarian cancer. *Nat. Cancer* **3**, 723–733 (2022).

51. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).

52. Thompson, J. A., Tan, J. & Greene, C. S.(T. D. M. Cross-platform normalization of microarray and RNA-seq data for machine learning applications. *PeerJ* **4**, e1621 (2016).

53. Feng, F. et al. A computational solution to improve biomarker reproducibility during long-term projects. *PLoS ONE* **14**, e0209060 (2019).

54. Friedman, J. H., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).

55. Simon, N., Friedman, J. H., Hastie, T. & Tibshirani, R. Regularization paths for Cox's proportional hazards model via coordinate descent. *J. Stat. Softw.* **39**, 1–13 (2011).

56. Verweij, P. J. M. & Van Houwelingen, H. C. Penalized likelihood in Cox regression. *Stat. Med.* **13**, 2427–2436 (1994).

57. Polsterl, S., Navab, N. & Katouzian, A. Fast training of support vector machines for survival analysis. *Lect. Notes Artif. Int.* **9285**, 243–259 (2015).

58. Ishwaran, H., Kogalur, U. B., Blackstone, E. H. & Lauer, M. S. Random survival forests. *Ann. Appl. Stat.* **2**, 841–860 (2008).

59. Pölsterl, S. scikit-survival: a library for time-to-event analysis built on top of scikit-learn. *J. Mach. Learn. Res.* **21**, 1–6 (2020).

60. Sammut, S.-J. et al. Multi-omic machine learning predictor of breast cancer therapy response. *Nature* **601**, 623–629 (2022).

61. Yoo, S.-K. et al. Prediction of checkpoint inhibitor immunotherapy efficacy for cancer using routine blood tests and clinical data. *Nat. Med* **31**, 869–880 (2025).

62. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: synthetic minority over-sampling technique. *JAIR* **16**, 321–357 (2002).

63. Lundberg, S. M. et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat. Biomed. Eng.* **2**, 749–760 (2018).

64. Zhang, J. D., Xue, C., Kolachalama, V. B. & Donald, W. A. Interpretable machine learning on metabolomics data reveals biomarkers for Parkinson's disease. *ACS Cent. Sci.* **9**, 1035–1045 (2023).

Acknowledgements

Author contributions

D.-F.G. wrote the code, conducted bioinformatics analysis and machine learning modeling, and drafted the manuscript. D.-F.G., L.-W.F., and J.-K.W. performed the cell and animal experiments. D.-F.G., L.-W.F., and W.Q. collected and processed the human serum samples and pathological tissue sections. D.-F.G., Y.-H.L., and Q.F. collated the clinical and follow-up data. D.-F.G., K.-H.Z., and T.W. revised the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information