Journal of Zhejiang University-SCIENCE B 2026 27(7):788-805
www.jzus.zju.edu.cn; www.springer.com/journal/11585
E-mail: jzus_b@zju.edu.cn

Research Article
https://doi.org/10.1631/jzus.B2500607

Check for updates

Machine learning-driven evaluation of protein kinase D3 as a co-diagnostic biomarker in hepatocellular carcinoma

Jing LI[1*], Yifan ZHAO[2*], Yicheng MA[1], Bei XIE[3], Li HUANG[4], Haitang YANG[1], Xingyuan MA[5], Haohua DENG[5], Shuaiyang WANG[5], Chanjuan SUN[6], Pengfei CAO[2✉], Linjing LI[1✉]

[1]*Department of Clinical Laboratory Center, Lanzhou University Second Hospital, Lanzhou 730030, China*
[2]*School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China*
[3]*Department of Immunology, School of Basic Medical Sciences, Lanzhou University, Lanzhou 730000, China*
[4]*Department of Pediatric Nephrology, Lanzhou University Second Hospital, Lanzhou 730030, China*
[5]*School of Second Clinical Medical, Lanzhou University, Lanzhou 730030, China*
[6]*School of Environment and Architecture, University of Shanghai for Science and Technology, Shanghai 200093, China*

Abstract: To elucidate the diagnostic value and clinical relevance of protein kinase D3 (PRKD3) in hepatocellular carcinoma (HCC), we analyzed data retrieved from The Cancer Genome Atlas (TCGA) database, which revealed high expression of *PRKD3* in HCC tissues. Subsequently, we collected a total of 392 clinical plasma samples from healthy individuals, patients with cirrhosis or decompensated cirrhosis, and patients with HCC. Plasma PRKD3 levels were then determined across HCC patients and individuals at high risk of developing the disease. The results revealed significantly elevated PRKD3 concentrations in patients with cirrhosis, decompensated cirrhosis, and HCC compared to healthy controls (P<0.01). The areas under the receiver operating characteristic (ROC) curve for these three groups were 0.8107, 0.7899, and 0.7177, respectively. To further evaluate the efficacy of PRKD3 as an adjunctive diagnostic biomarker for HCC, we employed a panel of machine learning algorithms as primary classifiers, including extra trees (ET), gradient boosting (GB), random forest (RF), and support vector machine (SVM). A multi-parameter joint diagnostic model was constructed by combining PRKD3 expression data with a set of clinical parameters, including gender, age, total bilirubin (TBIL), alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase (ALP), albumin (ALB), alpha-fetoprotein (AFP), and prothrombin induced by vitamin K absence-II (PIVKA-II). This integrated approach exhibited substantially improved diagnostic performance, achieving an accuracy of 0.861, sensitivity of 0.863, specificity of 0.925, and precision of 0.862. Collectively, these findings highlight the potential of PRKD3 as an integral component of a comprehensive diagnostic tool for the early identification of HCC.

Key words: Protein kinase D3 (PRKD3); Hepatocellular carcinoma (HCC); Machine learning; Combined diagnosis

1 Introduction

Hepatocellular carcinoma (HCC) is among the most lethal malignancies globally, imposing a particularly heavy burden worldwide. According to the Global Cancer Observatory (https://gco.iarc.fr), HCC is the sixth most prevalent cancer and the third leading cause of cancer-related deaths, with 906 000 new cases and 830 000 deaths in 2020 (Sung et al., 2021). Furthermore, the National Cancer Center of China reported 367 700 new primary HCC cases in China in 2022, ranking fourth among new cancer cases. With the fifth-highest incidence rate, the number of deaths attributed to primary liver cancer was 316 500 in China, placing it second in terms of cancer mortality and fatality rates (Han et al., 2024). Unlike in developed countries, the elevated HCC mortality rate in China is driven by challenges in the sensitivity, specificity, and precision of early diagnosis, resulting in diagnosis at advanced stages for most patients. Consequently, despite aggressive clinical interventions, these patients ultimately succumb to recurrence and metastasis.

✉ Linjing LI, lilinj@lzu.edu.cn
Pengfei CAO, caopf@lzu.edu.cn
* These two authors contributed equally to this work
Linjing LI, https://orcid.org/0000-0003-0864-4225
Pengfei CAO, https://orcid.org/0000-0001-8715-5002
Jing LI, https://orcid.org/0009-0006-2423-1964
Yifan ZHAO, https://orcid.org/0009-0002-0321-0110

Received Sept. 21, 2025; Revision accepted Jan. 18, 2026;
Crosschecked May 21, 2026; Published online June 20, 2026

© Zhejiang University Press 2026

International guidelines previously included alpha-fetoprotein (AFP), a biomarker commonly used in the clinical diagnosis of HCC. However, although the latest guidelines exclude AFP due to concerns about its diagnostic accuracy, its use remains a subject of ongoing debate (Bai et al., 2017; Razaghi and Björnstedt, 2024). The limitations of AFP may be mitigated by incorporating auxiliary diagnostic biomarkers. The 2024 Chinese Guidelines for the Diagnosis and Treatment of Primary Hepatocellular Carcinoma recommend prothrombin induced by vitamin K absence-II (PIVKA-II) (also known as des-gamma carboxyprothrombin (DCP)), plasma-free microRNA (Zhou et al., 2011), and serum AFP variants (lens culinaris agglutinin-reactive fraction of AFP (AFP-L3)) as early diagnostic markers for HCC, particularly for the serum AFP-negative population. Currently, both AFP and PIVKA-II are utilized in clinical practice to formulate diagnostic criteria for HCC. PIVKA-II is an abnormal prothrombin produced in individuals with either vitamin K deficiency or HCC, and patients with elevated PIVKA-II levels face a higher risk of HCC recurrence and metastasis (Debes et al., 2021; Tian et al., 2023). However, the efficacy of PIVKA-II, either as an independent biomarker or in combination with AFP, for prognostic prediction in HCC patients is also limited (Hiraoka et al., 2025). Therefore, there is an urgent need to identify biologically relevant biomarkers closely associated with HCC to enable synergistic diagnosis and improve the sensitivity and specificity of dynamic assessments of HCC progression.

The protein kinase D (PRKD) family of serine/threonine kinases, part of the Ca^{2+}/calmodulin-dependent protein kinase superfamily, possesses numerous cellular targets and is involved in a variety of biological processes, including cell growth, transcriptional regulation (Ha et al., 2008), angiogenesis (Evans and Zachary, 2011), protein trafficking (Yeaman et al., 2004), invasion and epithelial–mesenchymal transition (EMT) (Durand et al., 2016). Accumulating evidence indicates that PRKD3, a key member of this family, modulates cancer cell proliferation, growth, migration, and invasion across multiple tumor types. Our previous research demonstrated that PRKD3 is significantly overexpressed in gastric cancer tissues, where it promotes G2/M phase progression and tumor cell proliferation by regulating the expression of key cell cycle-related proteins, such as cyclin-dependent kinase 1 (CDK1), cyclin B1 (CCNB1), checkpoint kinase 1 (CHK1), and polo-like kinase 1 (PLK1) (Wang et al., 2025). Similarly, Zhang et al. (2019) demonstrated that PRKD3 initiates glycolysis and upregulates 6-phosphofructokinase/fructose-2,6-bisphosphatase 3 (PFKFB3), facilitating tumor development in gastric cancer. Liu Y et al. (2020) found that the PRKD3 likely promotes the cell proliferation in the breast cancer cells by activating extracellular signal-regulated kinase 1 (ERK1)–cellular myelocytomatosis oncogene (c-MYC) axis, while Huck et al. (2014) revealed that PRKD3 exerts a protumor effect by activating the mammalian target of rapamycin complex 1 (mTORC1)–ribosomal protein S6 kinase 1 (S6K1) pathway in triple-negative breast cancer cells. Additionally, Chen et al. (2008) suggested that while PRKD3 enhances the proliferation of prostate cancer cells by activating downstream protein kinase B (Akt) and ERK1/2 pathways, the absence of PRKD3 may induce G0/G1 phase cell cycle arrest in prostate cancer cell lines, and Li et al. (2019) demonstrated that PRKD3 promotes cancer progression by enhancing lipid production in prostate cancer cells. A positive feedback regulation between PRKD3 and programmed death-ligand 1 (PD-L1) was observed in oral squamous cell carcinoma, inducing EMT in tumors via the ERK/signal transducer and activator of transcription 1/3 (STAT1/3) pathway, thereby promoting tumor growth and metastasis (Cui et al., 2021). PRKD3 has been shown to exert oncogenic effects in HCC. Yang et al. (2017) proposed that elevated PRKD3 expression in HCC tissues is significantly associated with poor prognosis. Previous studies have reported a marked increase in PRKD3 protein expression in a highly malignant biological phenotype of insulin-resistant HCC cell models (Yan et al., 2022). Our team has also elucidated the inhibitory effect of PRKD3 knockdown on HCC proliferation and identified CDK4, serpin family E member 1 (SERPINE1), sequestosome 1 (SQSTM1), Ras-related protein Rab8A (RAB8A), and nuclear receptor-binding factor 2 (NRBF2) as potential key proteins in regulatory pathways with PRKD3. Collectively, PRKD3 has been reported to participate in tumor progression through multiple mechanisms, including cell cycle regulation, metabolic reprogramming, and the promotion of tumor growth, metastasis, and EMT, thereby attracting increasing attention. Notably, its effects are significantly tumor type-dependent (Tian et al., 2024). However, the diagnostic value of PRKD3 as a clinical serum marker for HCC warrants further investigation.

The rapid development of machine learning technology has driven significant advances in its application in medicine, particularly regarding clinical diagnosis, treatment decision-making, and medical resource management, where it has demonstrated great potential (Calderaro et al., 2022; Swanson et al., 2023). Interdisciplinary research on tumors has focused on integrating machine learning into tumor screening, diagnosis, treatment, patient care, and rehabilitation (Bagheri et al., 2017; Bi et al., 2019; Goldenberg et al., 2019; Ghosh et al., 2025). Machine learning not only overcomes the limitations of conventional statistical methods but also extracts critical features from vast quantities of data, revealing new potential biomarkers and thereby significantly improving diagnostic sensitivity and specificity (Moldogazieva et al., 2021; Liu et al., 2024). The main existing combined diagnostic models for HCC, along with their diagnostic efficiencies, are as follows. The GALAD model, incorporating gender, age, AFP-L3, AFP, and PIVKA-II, was recommended by the China Liver Cancer Guidelines for the Diagnosis and Treatment of Hepatocellular Carcinoma (2024 Edition) (Zhou et al., 2025). This model achieved a sensitivity of 85.6% and a specificity of 93.3% for the early diagnosis of HCC (Best et al., 2020). A phase III validation study by Fujiwara et al. (2025) demonstrated that GALAD outperforms AFP for HCC diagnosis and can detect cancer 12 months prior to a confirmed diagnosis. These findings provide high-quality, time-sensitive evidence supporting the utility of multi-analyte models. Furthermore, the simplified GAAD model (Piratvisuth et al., 2023), which incorporates gender, age, AFP, and PIVKA-II, and the ASAP model, which incorporates age, sex, AFP, and PIVKA-II (Yang et al., 2019), exhibited similar diagnostic efficacy. A detection method combining seven microRNAs achieved a sensitivity of 86.1% and a specificity of 76.8% for diagnosing HCC. A recent study by El-Serag et al. (2025) reported that HCC early detection screening (HES) V2.0, a diagnostic model, significantly outperformed GALAD and ASAP in the overall and early diagnosis of HCC, underscoring the diagnostic advances enabled by algorithm updates and optimized marker combinations. An international multi-center prospective comparative study (Hou et al., 2025) demonstrated that GAAD achieved diagnostic performance comparable to that of GALAD, with an area under the curve (AUC) of approximately 0.91 for early-stage HCC. Varghese et al. (2025) integrated metabolomics, proteomics, and glycoproteomics data with machine learning approaches to identify key biomarkers, including serpin peptidase inhibitor, clade A1 (SERPINA1) and branched-chain amino acids, which can effectively distinguish HCC from cirrhosis. Another study developed a machine learning model using whole-genome circulating DNA fragments from serum samples, which exhibited sensitivities of 88% and 85% for HCC detection in the general population and high-risk groups, respectively, alongside specificities of 98% and 80% (Foda et al., 2023). Notably, there remains room to improve the development of machine learning models that integrate AFP with newly identified biomarkers for early HCC diagnosis. Given PRKD3's oncogenic role in HCC, it is imperative to further evaluate its clinical utility as an auxiliary diagnostic marker. The present study therefore aimed to evaluate the role of PRKD3 in HCC and investigate its potential as a novel auxiliary biomarker for translational medicine applications. We adopted the method described by Xu et al. (2022) to analyze *PRKD3* messenger RNA (mRNA) expression data from HCC tissues and corresponding patient survival outcomes. Additionally, we integrated multiple traditional plasma indicators (including AFP and PIVKA-II) with clinical plasma PRKD3 levels across disease stages, using machine learning techniques to assess the diagnostic value of PRKD3.

2 Materials and methods

2.1 TCGA database analysis of PRKD3 in HCC

To investigate the role of PRKD3 in HCC, we downloaded RNA sequencing (RNA-seq) data on liver hepatocellular carcinoma (LIHC) and corresponding clinical information from The Cancer Genome Atlas (TCGA) database via the Genomic Data Commons (GDC) Data Portal (https://portal.gdc.cancer.gov) on May 26, 2024. This study analyzed *PRKD3* mRNA expression levels, along with corresponding clinical characteristics and survival outcomes. We extracted data from 371 confirmed HCC patients and 50 non-tumor control subjects within the downloaded dataset. Gene expression data were derived from log-transformed RNA Seq V2 RSEM (RNA-seq by expectation maximization) values, from which log-transformed mRNA

expression *z*-scores were calculated, using normal samples as references to ensure normalization and consistency of mRNA expression levels. The normality of the PRKD3 variable was assessed using the Kolmogorov-Smirnov test (KS distance=0.0267, $P>0.1$), which indicated that the data were normally distributed. All data processing and statistical analyses were performed using R software (version 4.3.3) and GraphPad Prism (version 9.5). PRKD3 expression levels in tumor and normal tissues were compared using an independent-samples *t*-test. One-way analysis of variance (ANOVA) was employed to assess differences in the pathological stages of HCC. Prior to ANOVA, the Levene's test was performed to evaluate the homogeneity of variances across groups. Because the homogeneity of variance assumption was met, Tukey's honest significant difference (HSD) test was used for post hoc pairwise comparisons to identify differences among pathological stages. Diagnostic accuracy was assessed using receiver operating characteristic (ROC) curve analysis. Furthermore, Kaplan-Meier survival analysis and log-rank tests were used to evaluate survival differences by PRKD3 expression levels. Statistical significance was defined as a two-sided *P*-value of <0.05.

2.2 Plasma specimen data sources and processing

2.2.1 Study participants

A total of 392 study samples were obtained from Lanzhou University Second Hospital: 102 from healthy individuals, 53 from patients with liver cirrhosis, 151 from patients with decompensated liver cirrhosis, and 86 from patients newly diagnosed with HCC who had not received any treatment or surgery. The samples encompassed various ethnic groups and regions, and their sources were diverse. We collected and summarized the baseline characteristics of these patients. Upon admission, initial ethylenediaminetetraacetic acid-K2 (EDTA-K2)-anticoagulated whole-blood samples were harvested, and plasma samples were centrifuged and collected within 4 h of acquisition. All samples were stored at −40 °C in a freezer prior to analysis. This work was approved by the Ethics Committee of Lanzhou University Second Hospital (No. 2023A-076).

2.2.2 Study criteria

Inclusion criteria for the newly diagnosed, treatment-naive HCC group were defined as follows: (1) age of ≥18 years with diagnosis confirmed through histopathological examination; (2) newly diagnosed HCC without any prior anti-tumor treatment interventions (including surgery, interventional therapy, radiotherapy, or chemotherapy); (3) complete and well-documented clinical data; (4) no concurrent benign or malignant tumors in the liver or other organs; (5) no evidence of vascular invasion (e.g., portal vein, hepatic vein, or bile duct involvement) or distant metastases. Exclusion criteria were as follows: (1) presence of metastatic tumors from other primary cancers; (2) history of other malignant neoplasms; (3) concurrent benign space-occupying lesions in the liver or other organs; (4) secondary HCC, autoimmune hepatitis, drug-induced liver injury, alcoholic liver disease, or nonalcoholic steatohepatitis; (5) prior history of surgery, radiotherapy, or chemotherapy for any malignancy; (6) death within 30 d following surgery; (7) incomplete or unavailable clinical data. Ultimately, a total of 86 eligible HCC patients were enrolled in this study. All healthy controls had normal complete blood counts and liver function parameters. All HCC cases and cirrhosis patients were pathologically confirmed following independent review by at least two qualified pathologists. The diagnosis of cirrhosis was established by a combination of serological tests and imaging examinations.

2.2.3 Data processing

We measured PRKD3 expression levels in the plasma of each group using an enzyme-linked immunosorbent assay (ELISA) kit specifically designed for human PRKD3 protein (see supplementary materials and methods). Statistical analysis was conducted using GraphPad Prism 9.5 to calculate the sensitivity, specificity, and corresponding 95% confidence intervals (CIs) for the positive and negative predictive values. PRKD3 expression was evaluated in patients with cirrhosis, decompensated cirrhosis, and HCC using AUC and ROC curves.

2.3 Data sources and processing for machine learning

Each of the 392 samples included in the analysis was characterized by ten baseline variables: demographic features (gender (male/female) and age) and laboratory parameters (total bilirubin (TBIL), alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase (ALP), albumin (ALB),

PRKD3, AFP, and PIVKA-II). Following rigorous data preprocessing to ensure data quality and consistency, all samples—together with their corresponding disease status annotations—were incorporated into subsequent statistical analyses. Finally, the processed samples were stratified into three mutually exclusive groups: healthy control, cirrhosis or decompensated cirrhosis, and HCC.

During the data preprocessing stage, the original dataset was first split into training and test sets at an 8:2 ratio to prevent information leakage. All subsequent preprocessing steps were performed exclusively on the training set. Missing values were handled using a label-agnostic feature-wise mean imputation strategy; imputation values were calculated solely from the training set without reference to any class or target information (the distribution of missing values in the original data is shown in Fig. S1). The same training-derived statistics were then consistently applied to the test set. Gender was encoded as a binary feature (0 for female; 1 for male). All continuous variables were normalized using parameters estimated from the training set to mitigate the influence of scale differences on model learning. To address class imbalance, the synthetic minority oversampling technique (SMOTE, random_state=0) was applied only to the training data to balance class distributions. The test set remained completely independent, retaining its original class proportions and undergoing no oversampling or data augmentation, thereby ensuring that the model evaluation reflects realistic clinical scenarios.

Data processing was performed in Python 3.11.8 based on Anaconda 24.9.2, primarily using fundamental data processing and machine learning libraries: Scikit-learn 1.2.2, Pandas 2.1.4, NumPy 1.26.4, Imblearn 0.12.4, XGBoost 2.0.3, Matplotlib 3.8.0, and Seaborn 0.12.2. Following these preprocessing steps, the HCC prediction model was developed using the rigorously cleaned and well-standardized dataset. The model data queue is shown in Table 1.

2.4 Algorithm model

The present study adopted an ensemble machine learning strategy to develop and validate a predictive framework for assessing the diagnostic performance and feature importance of PRKD3. The workflow encompassed data preprocessing, base-learner construction, feature-importance evaluation, and meta-classifier integration. Predictive performance was systematically enhanced by integrating complementary models. Prior to model training, the dataset was partitioned into training and test subsets. To prevent information leakage, all preprocessing procedures—including feature standardization and class-imbalance correction—were conducted exclusively on the training set. Class imbalance was addressed using the SMOTE, while the test set was preserved in its original distribution to enable an unbiased and clinically realistic performance evaluation. During the base model training phase, four machine learning models were selected: extra trees (ET), gradient boosting (GB), random forest (RF), and support vector machine (SVM), all using a unified random seed (random_state=0) to ensure reproducibility. Hyperparameter optimization was performed using a grid search strategy: the core hyperparameters of each model are detailed in Table S1. During base model training and feature selection, a 10-fold cross-validation (random_state=0) was employed to assess the model's robustness and reliability. The feature importance scores for each model were first computed and then ranked. Guided by a forward feature selection strategy, the optimal subset of features that maximized accuracy was identified by incrementally expanding the feature set and documenting the corresponding accuracy change curves. Subsequently, weights were assigned to each base model according to their respective accuracies, and the feature importance scores were integrated using a weighted average method to generate the final ranking: $I_c=\sum_{i=1}^{n} w_i \times I_i$, where I_c denotes the comprehensive importance score, and w_i and I_i denote the weight coefficient and importance score of the ith base model, respectively. Within the stacking ensemble framework, an out-of-fold (OOF) strategy was adopted to generate meta-features. Specifically, predicted probabilities from each base classifier were obtained via 10-fold cross-validation on the training set, ensuring that meta-features were derived exclusively from fold-wise held-out samples rather than in-sample predictions. These OOF predictions were concatenated and used to train the secondary classifier extreme gradient boosting (XGBoost). Model performance was subsequently evaluated on an independent test set that was not involved in base-model training, feature selection, or meta-learner fitting. During the model evaluation phase, multidimensional performance indicators, including accuracy, sensitivity, specificity, precision, and

Table 1 Model data queue

Category	Healthy	Cirrhosis or decompensated cirrhosis	HCC	Total
Gender				
Male	47	131	68	246
Female	55	73	18	146
Total	102	204	86	392
Age (years)				
Median (range)	47.5 (18.0–74.0)	55.0 (20.0–83.0)	58.0 (31.0–92.0)	54.0 (18.0–92.0)
Mean±SD	45.5±14.7	53.8±11.1	59.3±11.3	52.8±13.1
IQR	35.00–56.75	48.00–60.00	52.25–66.00	46.00–60.25
Mode	51	59	55	59
Ethnic group				
Han	90	157	73	320
Hui	7	34	4	45
Tibetan	2	4	6	12
Dongxiang	2	7	2	11
Other ethnic groups	1	2	1	4
Region				
The Hexi Corridor (Wuwei, Zhangye, Jiayuguan, and Dunhuang)	12	18	16	46
The mountainous areas of Longnan (Longnan and Tianshui)	10	47	9	66
The Gannan Plateau	5	7	6	18
The Loess Plateau of eastern Gansu (Qingyang and Pingliang)	8	9	8	25
The Loess Plateau of western Gansu (Lanzhou, Dingxi, and Linxia)	65	118	45	228
Other provinces	2	5	2	9

HCC: hepatocellular carcinoma. Age: Median (range) represents the median age and its range for each category; Mean±SD indicates the mean age±standard deviation for each category; IQR denotes the interquartile range of age for each category; Mode represents the most frequently occurring age in each category.

F1-score, were computed from the confusion matrix. In addition, ROC curves and AUC values were used to quantify classification confidence across different prediction tasks, thereby providing a comprehensive assessment of its predictive effectiveness. The model's workflow is illustrated in Fig. 1.

3 Results

3.1 Expression, diagnosis, and prognosis of *PRKD3* based on TCGA database

The study analyzed *PRKD3* mRNA expression in HCC and normal tissues using TCGA database. The findings indicated that *PRKD3* expression was markedly elevated in HCC tissues (P<0.01; Fig. 2a). One-way ANOVA revealed a significant difference in *PRKD3* expression levels among the four HCC stages (P<0.01; Fig. 2b). To further determine the exact stage pairs with statistical differences, Levene's test was conducted to verify the homogeneity of variance, and the result satisfied the homogeneity assumption, thereby justifying the use of the HSD test for post hoc analysis. Post-hoc pairwise comparisons demonstrated that *PRKD3* expression was significantly higher in stage TIII than in stage TI (mean difference=0.321, P<0.01), and marginally higher in stage TII than in stage TI (mean difference=0.216, P=0.09). No statistically significant differences in *PRKD3* expression were observed between other stage pairs. A ROC curve analysis of the diagnostic accuracy of *PRKD3* in HCC revealed an AUC of 68.2% in TCGA database (P< 0.01; Fig. 2c). The association between *PRKD3* expression levels in HCC patients and overall survival

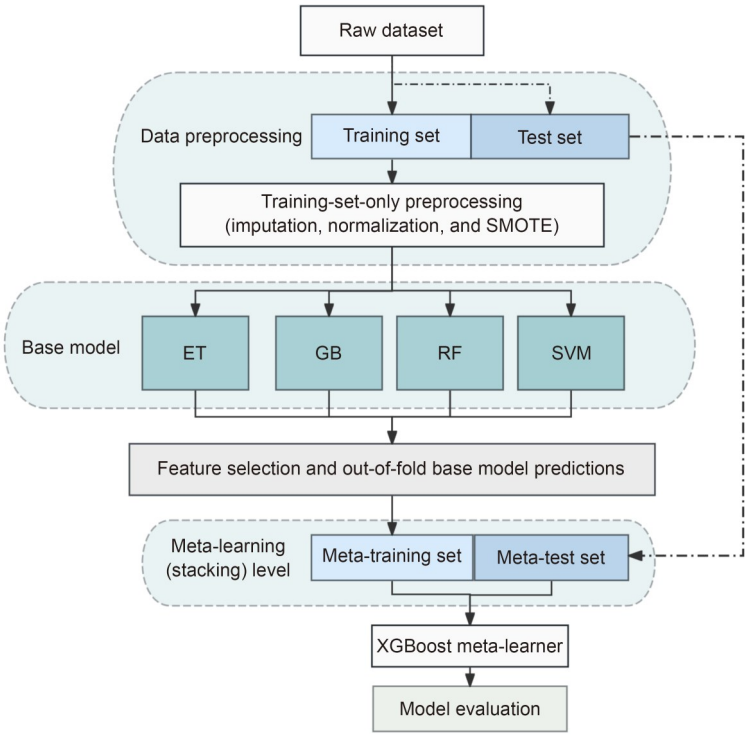

Fig. 1 Flowchart of model construction. Solid lines represent training data/main workflows; dashed lines indicate the data paths of the test set. SMOTE: synthetic minority oversampling technique; ET: extra trees; GB: gradient boosting; RF: random forest; SVM: support vector machine; XGBoost: extreme gradient boosting.

(OS) was examined using Kaplan-Meier curves and the Gene Expression Profiling Interactive Analysis (GEPIA) database. The results demonstrated that patients with lower *PRKD3* expression had a significantly longer OS than those with higher expression (P<0.05; Fig. 2d).

3.2 Expression levels of plasma PRKD3

The expression levels of plasma PRKD3 in the liver cirrhosis, decompensated liver cirrhosis, and HCC groups were significantly different from those in the healthy control group (P<0.01 for all groups; Fig. 3a). As shown in Fig. 3b, the AUC for the liver cirrhosis and healthy control groups was 0.8107, and the difference was statistically significant (P<0.01). Compared with the healthy control group, the AUC for the decompensated liver cirrhosis group was 0.7899, and the difference was statistically significant (P<0.01). Compared with the healthy control group, the AUC for the HCC group was 0.7177, and the difference was still statistically significant (P< 0.01). This indicates that PRKD3 cannot be used as an independent diagnostic marker for HCC; however, its expression level can increase across different stages of liver disease, suggesting its potential as an auxiliary diagnostic indicator.

3.3 Machine learning in data model result analysis and feature analysis

3.3.1 Data model result analysis

This study conducted a detailed analysis of the model's performance across feature combinations, with a focus on the role of the newly identified PRKD3 feature in HCC classification. Confusion matrices were employed to evaluate the model's accuracy, sensitivity, specificity, and other performance metrics across various feature inputs. Furthermore, the diagnostic value of the PRKD3 biomarker was substantiated by a combined analysis that included additional features and a feature importance ranking.

Multiple feature combinations were employed for model training and testing. The model's performance was evaluated using the confusion matrix and associated metrics. In the confusion matrix and ROC curve, categories 0, 1, and 2 represented healthy individuals, patients with cirrhosis or decompensated cirrhosis, and

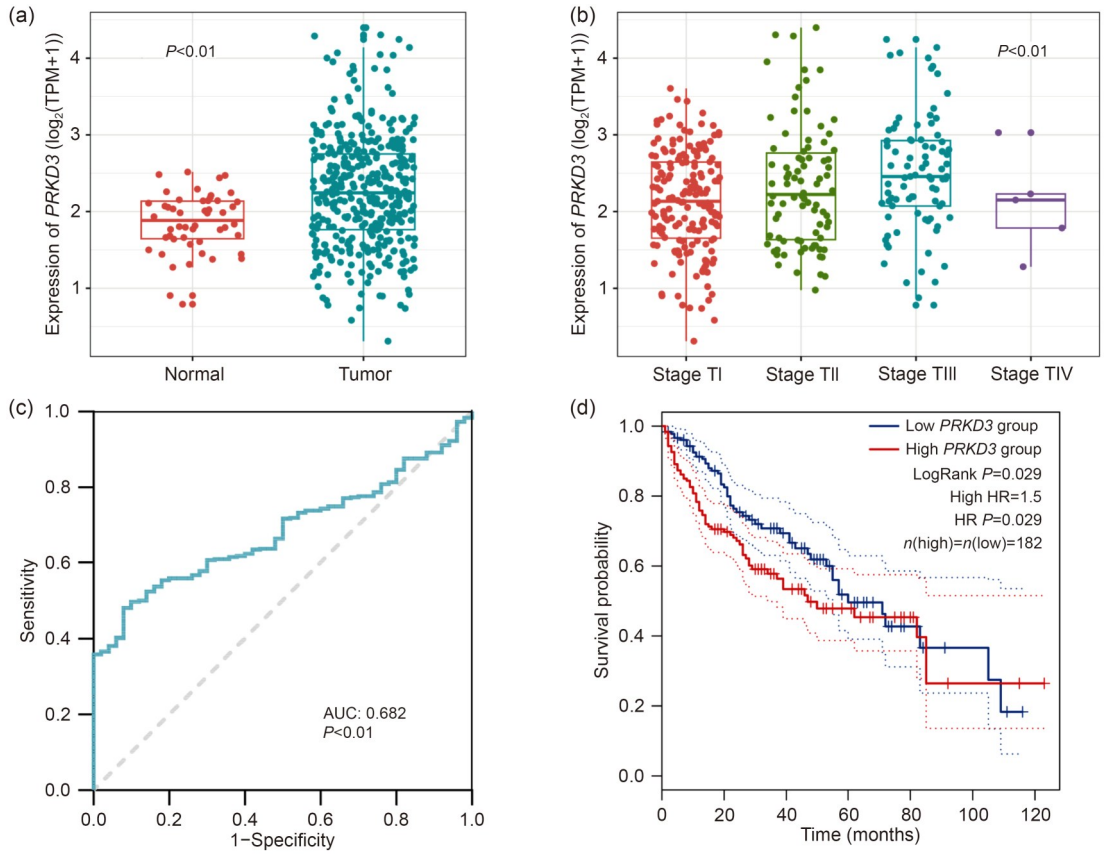

Fig. 2 Expression, diagnosis, and prognosis of protein kinase D3 (*PRKD3*) based on The Cancer Genome Atlas (TCGA) database. (a) Messenger RNA (mRNA) expression of *PRKD3* in hepatocellular carcinoma (HCC) tissues compared to normal tissues. (b) Correlation between *PRKD3* expression and clinical pathological staging in HCC patients. (c) Diagnostic accuracy of *PRKD3* in HCC via receiver operating characteristic (ROC) curve. (d) Kaplan-Meier survival analysis of *PRKD3* expression. TPM: transcripts per million; AUC: area under the curve; HR: hazard ratio.

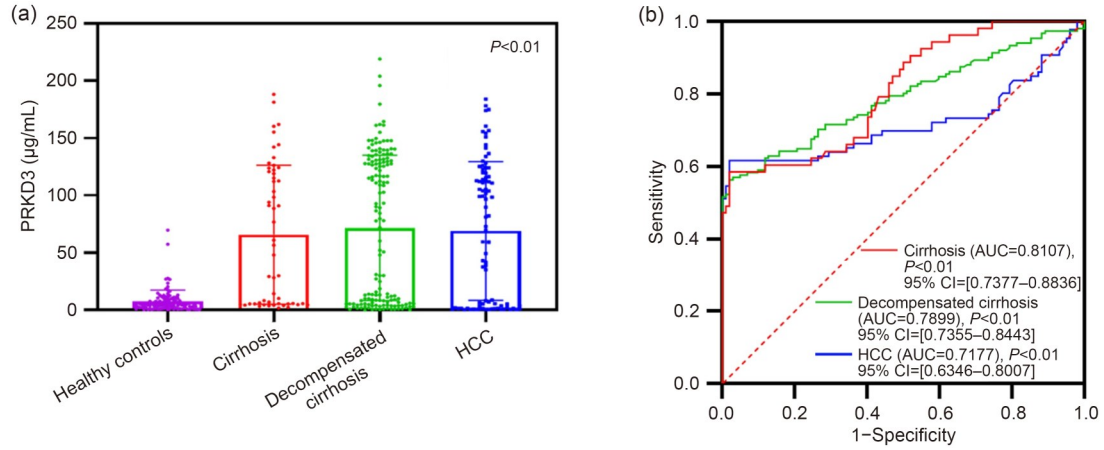

Fig. 3 Expression levels of plasma protein kinase D3 (PRKD3). (a) Scatter plot of PRKD3 expression levels in the cirrhosis, decompensated cirrhosis, and hepatocellular carcinoma (HCC) groups compared to the healthy control group. A significance threshold for between-group comparisons was set at $P<0.01$. The data are expressed as mean±standard deviation (SD), with sample sizes of 53, 151, 86, and 102 for the liver cirrhosis, decompensated liver cirrhosis, HCC, and healthy control groups, respectively. (b) Diagnostic accuracy of PRKD3 in cirrhosis, decompensated cirrhosis, and HCC via receiver operating characteristic (ROC) curve. All *P*-values were <0.01. AUC: area under the curve; CI: confidence interval.

HCC patients, respectively. For multi-class ROC analysis, we employed the One-vs-Rest (OvR) strategy and all ROC curves included 95% CIs derived from 1000 bootstrap resampling iterations.

The model performance when using a single feature PRKD3 is shown in Fig. 4a. The model achieved an accuracy of 0.570, a sensitivity of 0.551, a specificity of 0.781, a precision of 0.603, and an F1 score of 0.582. However, the performance of PRKD3 in independent diagnosis was relatively limited; its high specificity indicated that it was advantageous for identifying non-HCC patients.

The model results when using the single feature AFP are shown in Fig. 4b, with an accuracy of 0.532, a sensitivity of 0.515, a specificity of 0.746, a precision of 0.561, and an F1 score of 0.536. Although AFP has been widely used in clinical practice, its performance as a single biomarker was similarly inadequate across multiple evaluation metrics.

Overall, these results indicate that neither PRKD3 nor AFP alone meets clinical diagnostic requirements, underscoring the need to integrate multiple biomarkers and machine learning-based models to improve diagnostic accuracy. The model performance significantly improved when combining the three features PRKD3, AFP, and PIVKA-II, as shown in Fig. 4c. At this point, the model achieved an accuracy of 0.722, a sensitivity of 0.747, a specificity of 0.862, a precision of 0.751, and an F1 score of 0.720. This suggested that the model could improve diagnostic accuracy and classify more effectively by integrating multiple biomarkers. Furthermore, the model performance was highest when

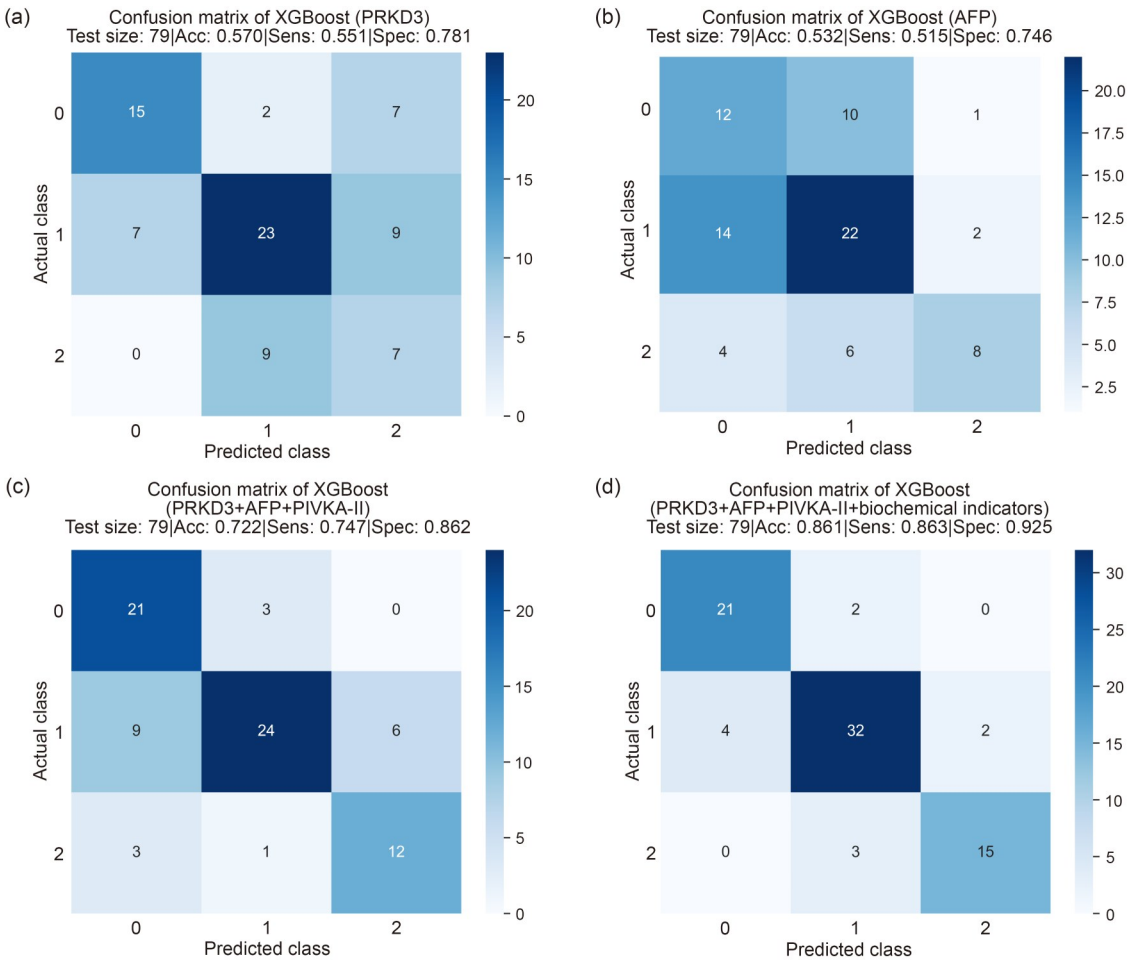

Fig. 4 Model performance via the single or multiple features. (a) Protein kinase D3 (PRKD3) features; (b) Alpha-fetoprotein (AFP) features; (c) Features of PRKD3 combined with AFP and prothrombin induced by vitamin K absence-II (PIVKA-II); (d) Features of PRKD3 combined with AFP, PIVKA-II, and biochemical indicators. Diagnostic classification is categorized as healthy individuals (0), patients with cirrhosis or decompensated cirrhosis (1), and hepatocellular carcinoma (HCC) patients (2). XGBoost: extreme gradient boosting; Acc: accuracy; Sens: sensitivity; Spec: specificity.

biochemical indicators (ten features) were integrated for combined diagnosis, as shown in Fig. 4d. With an F1 score of 0.861, specificity of 0.925, precision of 0.862, sensitivity of 0.863, and accuracy of 0.861, all indicators improved. To ensure the transparency of classification results, confusion matrices and ROC curves for the four base models (ET, GB, RF, and SVM) are presented in Fig. S2. These findings underscore the efficacy of combining features, indicating that integrating multiple features may considerably enhance model performance, particularly in terms of specificity and overall accuracy.

As shown in Fig. 5a, model accuracy consistently increased with the inclusion of additional features, with the ET and RF models exhibiting particularly strong performance. The highest accuracy was achieved when incorporating ten features. As illustrated in Fig. 5b, the model integrating PRKD3, AFP, PIVKA-II, and routine biochemical indicators demonstrated excellent discriminative ability, with AUC values of 0.94, 0.86, and 0.93 across the three classification tasks (healthy, cirrhosis or decompensated cirrhosis, and HCC), indicating robust and stable performance.

To assess the impact of SMOTE, we compared model performance on the SMOTE-balanced dataset with that on the original imbalanced dataset. The confusion matrix and ROC curves obtained from the original imbalanced dataset are shown in Fig. S3, and the corresponding performance metrics are summarized in Table S2. To further evaluate model stability under different data partitioning scenarios, four independent random train-test splits were performed on the SMOTE-processed dataset. As summarized in Table S3, the model achieved a mean accuracy of 0.855±0.034 and a mean F1 score of 0.874±0.037 across repeated experiments. The mean AUC values for the healthy, cirrhosis/decompensated cirrhosis, and HCC groups were 0.953±0.027, 0.909±0.042, and 0.881±0.044, respectively, indicating that the proposed model maintains robust and reliable performance across varying data distributions.

Fig. 6 depicts the feature importance analysis of the optimal model integrating PRKD3, AFP, PIVKA-II, and routine biochemical indicators. ALB (0.197) was the primary contributor, followed by PIVKA-II (0.169), AFP (0.161), and PRKD3 (0.136), with all four features exceeding a contribution threshold of 0.10. These results indicate that liver function-related indicators and established HCC biomarkers jointly drive model performance. Notably, PRKD3 ranked among the top contributors, supporting its added diagnostic value when combined with conventional markers. Additional biochemical indicators, including AST, TBIL, and ALP, provided complementary information, thereby further enhancing the model's discriminative capacity. Overall, this feature ranking highlights the benefit of multi-marker integration for improving the robustness and accuracy of HCC prediction.

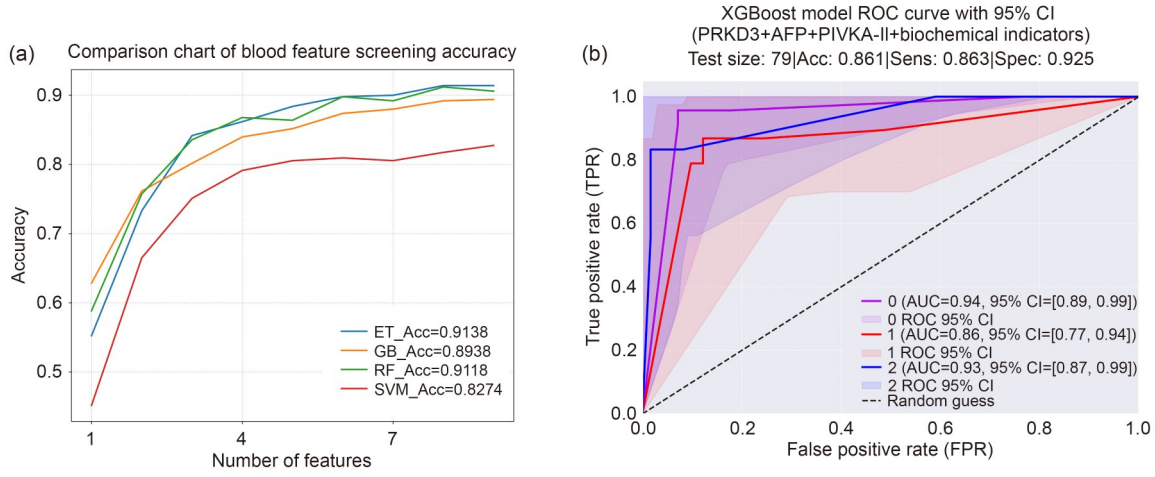

Fig. 5 Features of the accuracy and model receiver operating characteristic (ROC) curves under combined features. (a) Feature accuracy curve; (b) Model ROC curve under combined features. Diagnostic classification is categorized as healthy individuals (0), patients with cirrhosis or decompensated cirrhosis (1), and hepatocellular carcinoma (HCC) patients (2). ET: extra trees; GB: gradient boosting; RF: random forest; SVM: support vector machine; XGBoost: extreme gradient boosting; CI: confidence interval; PRKD3: protein kinase D3; AFP: alpha-fetoprotein; PIVKA-II: prothrombin induced by vitamin K absence-II; Acc: accuracy; Sens: sensitivity; Spec: specificity; AUC: area under the curve.

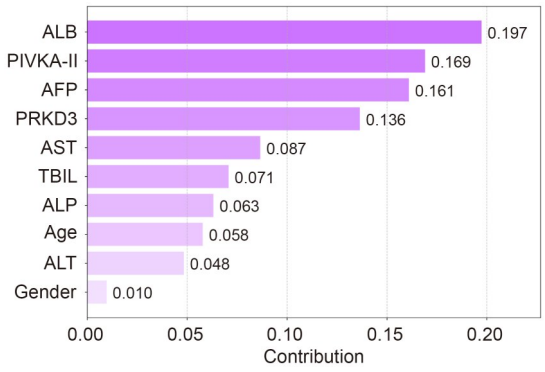

Fig. 6 Feature importance ranking under the optimal combined model. ALB: albumin; PIVKA-II: prothrombin induced by vitamin K absence-II; AFP: alpha-fetoprotein; PRKD3: protein kinase D3; AST: aspartate aminotransferase; TBIL: total bilirubin; ALP: alkaline phosphatase; ALT: alanine aminotransferase.

3.3.2 Feature result analysis

This study focused on the performance of the newly identified PRKD3 feature in the early diagnosis of HCC and compared it with traditional biomarkers such as AFP and PIVKA-II. We investigated the role and significance of PRKD3 in combined diagnostics by employing a multi-feature combination classification model. Our findings indicated that PRKD3 contributed significantly when combined with other markers. This suggests that its potential as an auxiliary diagnostic marker should not be underestimated, although its effectiveness appears to be limited when utilized independently.

3.3.3 Interpretability analysis

Boxplots of PRKD3, AFP, and PIVKA-II across disease states (healthy, cirrhosis or decompensated cirrhosis, and HCC) revealed their potential roles in the diagnosis of HCC. PRKD3 expression levels in patients with cirrhosis/decompensated cirrhosis and those with HCC were substantially higher than those in the healthy group (P<0.01 for both comparisons; Fig. 7). This suggests that PRKD3 may be considered an auxiliary diagnostic marker for HCC. In HCC patients, AFP, a commonly used HCC marker, was markedly elevated, further supporting its status as a classic marker. Furthermore, PIVKA-II levels were highest in HCC patients, underscoring its crucial role in HCC diagnosis. The feature correlation heatmap (Fig. 8) indicated complex correlations among these markers and biochemical indicators such as ALB,

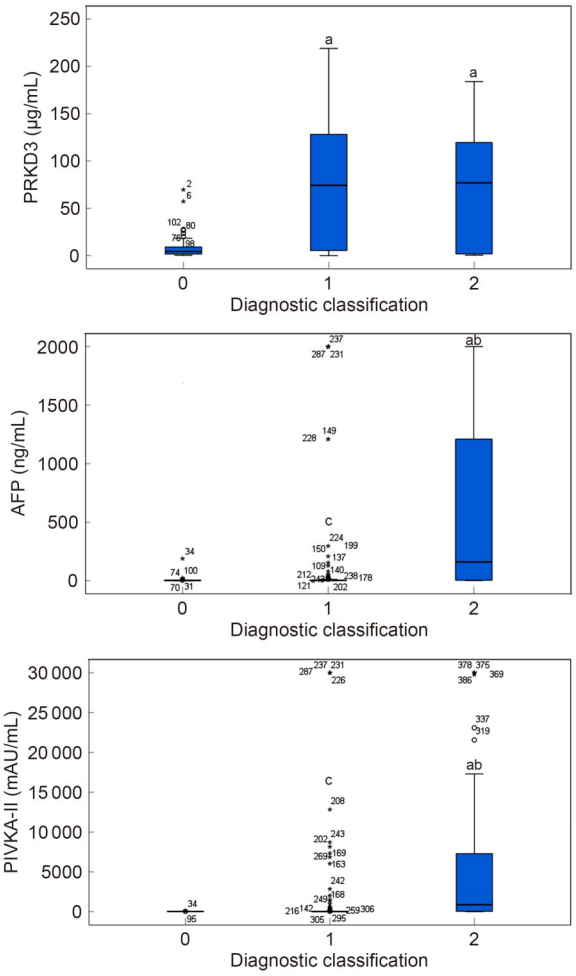

Fig. 7 Boxplots of protein kinase D3 (PRKD3), alpha-fetoprotein (AFP), and prothrombin induced by vitamin K absence-II (PIVKA-II) features. Diagnostic classification is categorized as healthy individuals (0), patients with cirrhosis or decompensated cirrhosis (1), and hepatocellular carcinoma (HCC) patients (2). The numbers adjacent to outliers represent sample case numbers; hollow circles (○) indicate mild outliers and asterisks (*) represent extreme outliers. [a] P<0.01, vs. healthy group; [b] P<0.01, vs. cirrhosis or decompensated cirrhosis group; [c] P<0.01, vs. HCC group.

particularly the positive correlation between AFP and PIVKA-II, suggesting that the combined use of these markers may further enhance diagnostic accuracy.

3.4 Clinical performance under clinically relevant binary classification tasks

To address the limited clinical relevance of multi-class classification, the proposed model was further evaluated with two clinically meaningful binary tasks using the XGBoost classifier. The first task focused on identifying HCC within a high-risk population, namely,

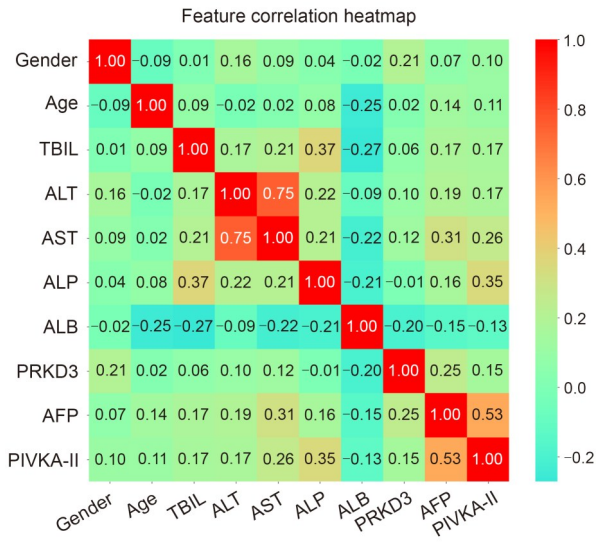

Fig. 8 Heatmap depicting the correlations between features. TBIL: total bilirubin; ALT: alanine aminotransferase; AST: aspartate aminotransferase; ALP: alkaline phosphatase; ALB: albumin; PRKD3: protein kinase D3; AFP: alpha-fetoprotein; PIVAKA-II: prothrombin induced by vitamin K absence-II.

patients with cirrhosis or decompensated cirrhosis. The second task evaluated the ability to distinguish HCC from non-HCC individuals. For both tasks, clinically interpretable decision thresholds were determined by maximizing the Youden index, thereby explicitly balancing sensitivity and specificity.

When comparing HCC and cirrhosis or decompensated cirrhosis, the model achieved a sensitivity of 0.880 and a specificity of 0.848 at an optimal threshold of 0.1246. In the HCC versus non-HCC task, the optimal threshold was 0.1233, yielding a sensitivity of 0.833 and a specificity of 0.918. For each task, the sensitivity–specificity tradeoff curve, the variation of sensitivity and specificity across decision thresholds, and the confusion matrix at the optimal threshold are shown in Fig. 9. Collectively, these results demonstrate that the proposed model maintains robust diagnostic performance in clinically relevant screening scenarios and provides explicit decision thresholds suitable for real-world HCC surveillance.

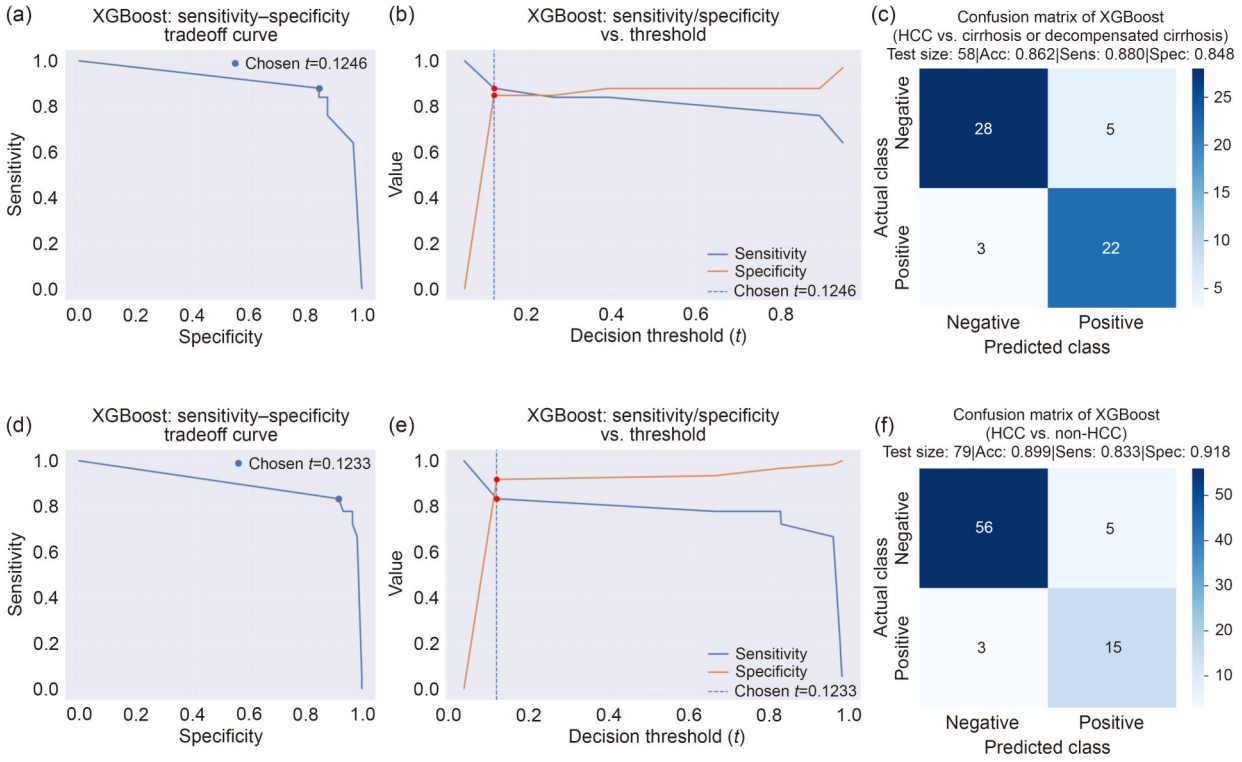

Fig. 9 Clinical performance of the extreme gradient boosting (XGBoost) model for discriminating hepatocellular carcinoma (HCC) vs. cirrhosis or decompensated cirrhosis (a–c) and HCC vs. non-HCC (d–f). (a, d) Sensitivity–specificity tradeoff curves with the optimal operating point determined by the Youden index; (b, e) Sensitivity and specificity as functions of the decision threshold (*t*), with dashed lines indicating the optimal thresholds; (c, f) Confusion matrices summarizing classification outcomes at the selected thresholds. Acc: accuracy; Sens: sensitivity; Spec: specificity.

4 Discussion

The innovation and core advantages of machine learning algorithms are mainly reflected in their breakthroughs in traditional diagnostic models. These algorithms ushered in a paradigm shift from "experience-driven" to "data-driven" and, through multi-dimensional data integration, automated feature mining, and accurate predictive capabilities, they have provided novel solutions for the early diagnosis and prognosis of HCC (Huang et al., 2024; Li et al., 2025). In the present study, although the sensitivity and specificity of PRKD3 alone as a diagnostic biomarker are insufficient, machine learning models can compensate for this limitation by integrating multi-source data, thereby highlighting the advantages of a combined diagnostic strategy involving PRKD3 and machine learning. By employing a combination analysis of machine learning and ten core features (including gender, age, TBIL, ALT, AST, ALP, ALB, PRKD3, AFP, and PIVKA-II), our diagnostic model achieved an F1 score of 0.861, a specificity of 0.925, a precision of 0.862, a sensitivity of 0.863, and an accuracy of 0.861. Our integrated machine learning model demonstrates comparable performance in early HCC screening.

PRKD3, a new diagnostic indicator for HCC, was incorporated into the machine learning model used in this study. PRKD3 drives tumor progression through multidimensional mechanisms, including cell cycle regulation and metabolic reprogramming, and its effect is significantly tumor-type-dependent. Regarding the core mechanism by which PRKD3 drives HCC development, the analysis of TCGA database revealed that PRKD3 was highly expressed in liver cancer tissues and cell lines, and that elevated PRKD3 expression correlated with poor patient prognosis. High PRKD3 expression was associated with multiple tumor nodules, poor differentiation, vascular invasion, advanced American Joint Committee on Cancer (AJCC) staging, and a poor clinical prognosis and clinical manifestations in patients with HCC after hepatectomy (Yang et al., 2017). Tian et al. (2024) demonstrated that knockdown of PRKD3 significantly inhibits the proliferation of Huh7 cells while promoting apoptosis. Conversely, PRKD3 overexpression enhanced these malignant phenotypes. CDK4, SERPINE1, SQSTM1, RAB8A, and NRBF2 may be key proteins in PRKD3 regulatory pathways.

Serological experiments revealed that PRKD3 alone exhibits limited diagnostic efficacy for HCC (AUC= 0.7177). Notably, its AUC value is higher in patients with liver cirrhosis or decompensated cirrhosis than in those with HCC. The core biological mechanism underlying this phenomenon may be closely linked to the functional heterogeneity of PRKD3 within the chronic liver injury–fibrosis–carcinogenesis axis. HCC originates from hepatocytes, whereas liver cirrhosis and decompensated cirrhosis are well-established high-risk factors for HCC (Foda et al., 2023). The pathogenesis of HCC involves multiple molecular defects, and its specific mechanisms vary depending on the underlying etiologies. The typical disease progression follows a cascade: from liver injury, chronic inflammation, fibrosis, and cirrhosis to the eventual development of HCC (Ding et al., 2025). Therefore, we hypothesize that PRKD3 may be a key molecular player in the progression of chronic liver injury to cancer. Its sustained high expression in liver cirrhosis or decompensated cirrhosis suggests that PRKD3 is not only a potential auxiliary diagnostic biomarker for HCC but also a driver of the malignant transformation of chronic liver diseases. The underlying mechanism may involve PRKD3 playing a critical role in liver fibrosis (LF) by regulating the activation of profibrotic macrophages. PRKD3 is highly expressed in hepatic macrophages, and PRKD3 knockout skews the polarization of mouse macrophages toward a profibrotic phenotype; advanced LF can progress to liver cirrhosis and HCC (Zhang et al., 2020). In conclusion, PRKD3 is more suitable for assessing the severity of chronic liver injury than for serving as a standalone diagnostic biomarker for HCC. To utilize PRKD3 for HCC diagnosis, it should be combined with other HCC-specific biomarkers, such as AFP and PIVKA-II, to compensate for its limited diagnostic performance in isolation. Future studies should investigate the dynamic expression changes of PRKD3 during the transition from liver cirrhosis to HCC and clarify its functional role in the initiation and progression of HCC.

In the final diagnostic model, this study integrated gender, age, and other biochemical indicators with PRKD3, AFP, and PIVKA-II, improving diagnostic performance. This showed that considering multiple biomarkers and clinical indicators could significantly enhance the accuracy and reliability of early HCC diagnosis. Biochemical indicators, such as ALB, AST, TBIL, ALP, and ALT, play important roles in the

occurrence and progression of HCC. The specific mechanism is as follows: ALB inhibits inflammatory responses by binding to pro-inflammatory factors, including interleukin-6 (IL-6) and tumor necrosis factor-alpha (TNF-α) (Kuwano et al., 2021). A decrease in ALB leads to excessive activation of the nuclear factor-kappa B (NF-κB) pathway, driving abnormal hepatocyte proliferation (Zhu et al., 2025). AST is primarily distributed in the mitochondria of hepatocytes. In chronic liver diseases, an AST/ALT ratio of >1 indicates mitochondrial damage and the progression of LF. Mitochondrial dysfunction increases reactive oxygen species (ROS) production, activates the Hippo–Yes-associated protein (YAP) pathway, and promotes the activation and fibrosis of hepatic stellate cells, which constitute the precancerous microenvironment of HCC (Liu YC et al., 2020). Elevated TBIL (>2 mg/dL) reflects hepatocyte excretory dysfunction or biliary obstruction. Bile acids activate the epidermal growth factor receptor (EGFR)/mitogen-activated protein kinase (MAPK) pathway via Takeda G protein-coupled receptor 5 (TGR5), promoting tumor cell migration and EMT (Lee et al., 2022). The increase in ALP indicates bile duct cell proliferation. The bile duct responds by secreting IL-6 and hepatocyte growth factor (HGF), which stimulate the proliferation of hepatoblasts through paracrine action, activate the Notch and Wnt pathways, and increase the risk of canceration (Maemura et al., 2014). In the multi-index combined diagnostic model of this study, ALB and TBIL reflected liver reserve function, AST and ALT excluded acute injury, and age and gender were adjusted for confounding factors, thereby achieving a multidimensional assessment. The core feature that distinguished PRKD3 from traditional markers was that AFP/PIVKA-II reflected hepatocyte differentiation status, whereas PRKD3 was directly involved in carcinogenic signal transduction and may indicate a malignant tendency earlier.

Research on the molecular mechanisms of gender difference in HCC has found that men generally have a higher incidence than women. Men also tend to have a worse prognosis in terms of survival period and pathological characteristics. Recent research has focused on elucidating the roles of sex hormones, DNA damage and repair pathways, immune microenvironments, and genetic epigenetic factors in driving gender-specific disparities. For instance, estrogen receptor signaling has been shown to suppress HCC progression, whereas androgen receptor signaling promotes tumor development (Xu et al., 2025). The combined use of multiple biomarkers and clinical characteristics not only improved the model's diagnostic precision but also enhanced its adaptability to individual patient differences, making the diagnostic results more aligned with clinical needs. The high importance score of PRKD3 in this multi-feature combination further supports its potential for use in HCC diagnosis.

In recent research on traditional biomarkers, Tian et al. (2023) demonstrated that combined detection using both AFP and PIVKA-II is superior to either test alone. Higher serum PIVKA-II levels are associated with more advanced HCC and a poorer prognosis, whereas AFP levels were not correlated with prognosis. In addition, these two markers also play an important role in liver transplantation. A study by Wang et al. (2023) confirmed that PIVKA-II can be integrated into transplant criteria in HCC patients. Similarly, Lai et al. (2024) demonstrated that the combination of PIVKA-II and AFP has significant clinical utility for predicting the risk of post-transplant tumor recurrence and identifying suitable candidates for liver transplantation. In this study, diagnostic performance was significantly enhanced when PRKD3 was combined with AFP and PIVKA-II. This model exhibits strong performance in specificity, precision, sensitivity, and accuracy, thereby compensating for the limitations of single-marker diagnosis. This result indicates that the combined use of multiple biomarkers could significantly improve the efficiency of identifying patients with HCC, reducing the possibility of misdiagnosis and missed diagnosis. Specifically, the potential value of PRKD3 as an auxiliary diagnostic marker should not be overlooked. Despite its limited effectiveness in isolation, its contribution was substantially enhanced when combined with other markers. Further investigations are warranted to explore the other potential roles of PRKD3.

This study has the following limitations. First, the data used in this study are single-center data from Gansu Province with an initial class imbalance (balanced via the SMOTE method), lacking external multicenter validation, and the generalizability of the model across different regions and ethnicities has not been confirmed. Furthermore, constrained by data collection conditions, this study did not include sufficient AFP-negative HCC samples, thereby precluding a systematic exploration of the diagnostic value and potential applications of PRKD3 in this subgroup.

5 Conclusions

This study explored the potential of PRKD3 as a combined diagnostic biomarker for HCC. The integrated model combining PRKD3 with AFP, PIVKA-II, and ALB achieved satisfactory performance in our single-center cohort. PRKD3 is expected to serve as a core biomarker in the comprehensive system for early HCC diagnosis. Meanwhile, leveraging machine learning in the diagnosis and treatment of HCC can facilitate the development of more accurate and individualized clinical strategies.

Data availability statement

The dataset used or analyzed in this study contains sensitive patient information, which is subject to the ethical approval (the Ethics Committee of Lanzhou University Second Hospital (No. 2023A-076)) and Chinese data privacy regulations to protect patient privacy. Therefore, raw data cannot be publicly shared. However, to ensure the reproducibility of the study, the de-identified feature matrix, including ten core features (gender, age, total bilirubin (TBIL), alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase (ALP), albumin (ALB), protein kinase D3 (PRKD3), alpha-fetoprotein (AFP), and prothrombin induced by vitamin K absence-II (PIVKA-II)) and disease labels (0=healthy, 1=cirrhosis/decompensated cirrhosis, and 2=hepatocellular carcinoma (HCC)), is available from the corresponding authors upon reasonable request. Requests should include a brief description of the research purpose, an institutional affiliation certificate (if applicable), and a commitment not to re-identify individuals or use the data for non-research purposes. We will respond within seven working days.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 82573290, 82272405, 82060531, and 81602622), the General Projects of Gansu Provincial Joint Research Fund (No. 23JRRA1503), the Key Research and Development Program of Gansu Province (No. 25YFFA052), the Natural Science Foundation of Gansu Province (Nos. 23JRRA0984 and 24JRRA431), the Lanzhou Talent Innovation and Entrepreneurship Project (No. 2023RC28), the Lanzhou Science and Technology Plan Project (No. 2025-2-88), the Chengguan District Talent Innovation and Entrepreneurship Project in Lanzhou City (No. 2024-rc-2), the Cuiying Scientific and Technological Innovation Program of Lanzhou University Second Hospital (No. CY2024-ZD-01), and the Key Incubation Project Funds of the Second Hospital and Clinical Medical School, Lanzhou University (Nos. 2025-22-zdfy-001 and 2025-25-zdfy-002), China.

Author contributions

Linjing LI and Pengfei CAO were responsible for project administration and supervision, with Linjing LI additionally undertaking funding acquisition. Jing LI and Linjing LI conducted the investigation and designed the study protocol for HCC research. Pengfei CAO and Yifan ZHAO developed the machine learning algorithms and performed the related data analyses. Bei XIE and Yicheng MA were responsible for the collection and organization of clinical samples. Li HUANG, Xingyuan MA, and Haohua DENG conducted the analysis of PRKD3 in HCC using The Cancer Genome Atlas (TCGA) database. Shuaiyang WANG, Haitang YANG, and Chanjuan SUN performed the plasma PRKD3 experiments and data analyses. Jing LI and Yifan ZHAO drafted the initial manuscript. All authors contributed to the revision and refinement of the paper. All the authors have read and approved the final version of the manuscript, and therefore, have full access to all the data in the study and take responsibility for the integrity and security of the data.

Compliance with ethics guidelines

Jing LI, Yifan ZHAO, Yicheng MA, Bei XIE, Li HUANG, Haitang YANG, Xingyuan MA, Haohua DENG, Shuaiyang WANG, Chanjuan SUN, Pengfei CAO, and Linjing LI declare that they have no conflicts of interest.

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. This work was approved by the Ethics Committee of Lanzhou University Second Hospital (No. 2023A-076). This work is a retrospective analysis, and individual consent from the participants has been waived.

Declaration on the use of generative AI tools

ChatGPT was used during the preparation of this manuscript to polish the language, improve readability, and correct grammatical errors. The authors have reviewed and revised all content thereafter, and assume full responsibility for the final manuscript.

References

Bagheri MH, Ahlman MA, Lindenberg L, et al., 2017. Advances in medical imaging for the diagnosis and management of common genitourinary cancers. *Urol Oncol Semin Orig Invest*, 35(7):473-491. https://doi.org/10.1016/j.urolonc.2017.04.014

Bai DS, Zhang C, Chen P, et al., 2017. The prognostic correlation of AFP level at diagnosis with pathological grade, progression, and survival of patients with hepatocellular carcinoma. *Sci Rep*, 7:12870. https://doi.org/10.1038/s41598-017-12834-1

Best J, Bechmann LP, Sowa JP, et al., 2020. GALAD score

detects early hepatocellular carcinoma in an international cohort of patients with nonalcoholic steatohepatitis. *Clin Gastroenterol Hepatol*, 18(3):728-735.e4. https://doi.org/10.1016/j.cgh.2019.11.012

Bi WL, Hosny A, Schabath MB, et al., 2019. Artificial intelligence in cancer imaging: clinical challenges and applications. *CA Cancer J Clin*, 69(2):127-157. https://doi.org/10.3322/caac.21552

Calderaro J, Seraphin TP, Luedde T, et al., 2022. Artificial intelligence for the prevention and clinical management of hepatocellular carcinoma. *J Hepatol*, 76(6):1348-1361. https://doi.org/10.1016/j.jhep.2022.01.014

Chen J, Deng F, Singh SV, et al., 2008. Protein kinase D3 (PKD3) contributes to prostate cancer cell growth and survival through a PKCε/PKD3 pathway downstream of Akt and ERK1/2. *Cancer Res*, 68(10):3844-3853. https://doi.org/10.1158/0008-5472.CAN-07-5156

Cui BM, Chen J, Luo M, et al., 2021. PKD3 promotes metastasis and growth of oral squamous cell carcinoma through positive feedback regulation with PD-L1 and activation of ERK-STAT1/3-EMT signalling. *Int J Oral Sci*, 13:8. https://doi.org/10.1038/s41368-021-00112-w

Debes JD, Romagnoli PA, Prieto J, et al., 2021. Serum biomarkers for the prediction of hepatocellular carcinoma. *Cancers*, 13(7):1681. https://doi.org/10.3390/cancers13071681

Ding ZX, Wang LS, Sun JT, et al., 2025. Hepatocellular carcinoma: pathogenesis, molecular mechanisms, and treatment advances. *Front Oncol*, 15:1526206. https://doi.org/10.3389/fonc.2025.1526206

Durand N, Borges S, Storz P, 2016. Protein kinase D enzymes as regulators of EMT and cancer cell invasion. *J Clin Med*, 5(2):20. https://doi.org/10.3390/jcm5020020

El-Serag HB, Jin QC, Tayob N, et al., 2025. HES V2.0 outperforms GALAD for detection of HCC: a phase 3 biomarker study in the United States. *Hepatology*, 81(2):465-475. https://doi.org/10.1097/HEP.0000000000000953

Evans IM, Zachary IC, 2011. Protein kinase D in vascular biology and angiogenesis. *IUBMB Life*, 63(4):258-263. https://doi.org/10.1002/iub.456

Foda ZH, Annapragada AV, Boyapati K, et al., 2023. Detecting liver cancer using cell-free DNA fragmentomes. *Cancer Discov*, 13(3):616-631. https://doi.org/10.1158/2159-8290.CD-22-0659

Fujiwara N, Lopez C, Marsh TL, et al., 2025. Phase 3 validation of PAaM for hepatocellular carcinoma risk stratification in cirrhosis. *Gastroenterology*, 168(3):556-567.e7. https://doi.org/10.1053/j.gastro.2024.10.035

Ghosh S, Zhao X, Alim M, et al., 2025. Artificial intelligence applied to 'omics data in liver disease: towards a personalised approach for diagnosis, prognosis and treatment. *Gut*, 74(2):295-311. https://doi.org/10.1136/gutjnl-2023-331740

Goldenberg SL, Nir G, Salcudean SE, 2019. A new era: artificial intelligence and machine learning in prostate cancer. *Nat Rev Urol*, 16(7):391-403. https://doi.org/10.1038/s41585-019-0193-3

Ha CH, Wang WY, Jhun BS, et al., 2008. Protein kinase D-dependent phosphorylation and nuclear export of histone deacetylase 5 mediates vascular endothelial growth factor-induced gene expression and angiogenesis. *J Biol Chem*, 283(21):14590-14599. https://doi.org/10.1074/jbc.M800264200

Han BF, Zheng RS, Zeng HM, et al., 2024. Cancer incidence and mortality in China, 2022. *J Natl Cancer Cent*, 4(1):47-53. https://doi.org/10.1016/j.jncc.2024.01.006

Hiraoka A, Kudo M, Tada T, et al., 2025. The current status of tumor markers as biomarkers in the era of immunotherapy for hepatocellular carcinoma: alpha-fetoprotein alone is not sufficient. *Oncology*, 104(1):79-91. https://doi.org/10.1159/000543405

Hou JL, Berg T, Vogel A, et al., 2025. Comparative evaluation of multimarker algorithms for early-stage HCC detection in multicenter prospective studies. *JHEP Rep*, 7(2):101263. https://doi.org/10.1016/j.jhepr.2024.101263

Huang HR, Wu FF, Yu Y, et al., 2024. Multi-transcriptomics analysis of microvascular invasion-related malignant cells and development of a machine learning-based prognostic model in hepatocellular carcinoma. *Front Immunol*, 15:1436131. https://doi.org/10.3389/fimmu.2024.1436131

Huck B, Duss S, Hausser A, et al., 2014. Elevated protein kinase D3 (PKD3) expression supports proliferation of triple-negative breast cancer cells and contributes to mTORC1-S6K1 pathway activation. *J Biol Chem*, 289(6):3138-3147. https://doi.org/10.1074/jbc.M113.502633

Kuwano A, Tanaka M, Suzuki H, et al., 2021. Upregulated expression of hypoxia reactive genes in peripheral blood mononuclear cells from chronic liver disease patients. *Biochem Biophys Rep*, 27:101068. https://doi.org/10.1016/j.bbrep.2021.101068

Lai Q, Ito T, Iesari S, et al., 2024. Role of protein induced by vitamin-K absence-II in transplanted patients with HCC not producing alpha-fetoprotein. *Liver Transpl*, 30(5):472-483. https://doi.org/10.1097/LVT.0000000000000259

Lee J, Hong EM, Kim JH, et al., 2022. Ursodeoxycholic acid inhibits epithelial-mesenchymal transition, suppressing invasiveness of bile duct cancer cells: an *in vitro* study. *Oncol Lett*, 24(6):448. https://doi.org/10.3892/ol.2022.13568

Li HQ, Zhang JY, Shi Y, et al., 2025. Identification of matrix stiffness-related molecular subtypes in HCC via integrating multi-omics analysis and machine learning algorithms. *J Transl Med*, 23:716. https://doi.org/10.1186/s12967-025-06733-7

Li L, Hua L, Fan HH, et al., 2019. Interplay of PKD3 with SREBP1 promotes cell growth via upregulating lipogenesis in prostate cancer cells. *J Cancer*, 10(25):6395-6404. https://doi.org/10.7150/jca.31254

Liu Y, Song H, Yu SY, et al., 2020. Protein kinase D3 promotes the cell proliferation by activating the ERK1/c-MYC axis in breast cancer. *J Cell Mol Med*, 24(3):2135-2144. https://doi.org/10.1111/jcmm.14772

Liu YC, Wang XH, Yang YZ, 2020. Hepatic Hippo signaling inhibits development of hepatocellular carcinoma. *Clin Mol Hepatol*, 26(4):742-750. https://doi.org/10.3350/cmh.2020.0178

Liu ZK, Wu YC, Khan AA, et al., 2024. Deep learning-based radiomics allows for a more accurate assessment of sarcopenia as a prognostic factor in hepatocellular carcinoma. *J Zhejiang Univ-Sci B (Biomed & Biotechnol)*, 25(1):83-90. https://doi.org/10.1631/jzus.B2300363

Maemura K, Natsugoe S, Takao S, 2014. Molecular mechanism of cholangiocarcinoma carcinogenesis. *J Hepato-Biliary Pancreat Sci*, 21(10):754-760. https://doi.org/10.1002/jhbp.126

Moldogazieva NT, Mokhosoev IM, Zavadskiy SP, et al., 2021. Proteomic profiling and artificial intelligence for hepatocellular carcinoma translational medicine. *Biomedicines*, 9(2):159. https://doi.org/10.3390/biomedicines9020159

Piratvisuth T, Hou JL, Tanwandee T, et al., 2023. Development and clinical validation of a novel algorithmic score (GAAD) for detecting HCC in prospective cohort studies. *Hepatol Commun*, 7(11):e0317. https://doi.org/10.1097/HC9.0000000000000317

Razaghi A, Björnstedt M, 2024. Exploring selenoprotein P in liver cancer: advanced statistical analysis and machine learning approaches. *Cancers*, 16(13):2382. https://doi.org/10.3390/cancers16132382

Sung H, Ferlay J, Siegel RL, et al., 2021. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, 71(3):209-249. https://doi.org/10.3322/caac.21660

Swanson K, Wu E, Zhang A, et al., 2023. From patterns to patients: advances in clinical machine learning for cancer diagnosis, prognosis, and treatment. *Cell*, 186(8):1772-1791. https://doi.org/10.1016/j.cell.2023.01.035

Tian S, Chen YY, Zhang YM, et al., 2023. Clinical value of serum AFP and PIVKA-II for diagnosis, treatment and prognosis of hepatocellular carcinoma. *J Clin Lab Anal*, 37(1):e24823. https://doi.org/10.1002/jcla.24823

Tian Y, Xie B, Wang SY, et al., 2024. PRKD3 promotes proliferation of liver cancer cells: a downstream proteomics profiling study. *Am J Transl Res*, 16(11):6384-6398. https://doi.org/10.62347/YLJE5332

Varghese RS, Zhang XR, Giridharan S, et al., 2025. Multiomics feature selection to identify biomarkers for hepatocellular carcinoma. *Metabolites*, 15(9):575. https://doi.org/10.3390/metabo15090575

Wang K, Dong LB, Lu Q, et al., 2023. Incorporation of protein induced by vitamin K absence or antagonist-II into transplant criteria expands beneficiaries of liver transplantation for hepatocellular carcinoma: a multicenter retrospective cohort study in China. *Int J Surg*, 109(12):4135-4144. https://doi.org/10.1097/JS9.0000000000000729

Wang SY, Xie B, Deng HH, et al., 2025. Effect of PRKD3 on cell cycle in gastric cancer progression and downstream regulatory networks. *Med Oncol*, 42(5):135. https://doi.org/10.1007/s12032-025-02663-y

Xu SJ, Ding N, Zheng S, et al., 2022. RNA binding protein-based risk score model for prognosis prediction of patients with hepatocellular carcinoma. *Chin Med J*, 135(23):2890-2892. https://doi.org/10.1097/CM9.0000000000002232

Xu ZQ, Luo SQ, Wu ZJ, et al., 2025. Current status and perspectives of molecular mechanisms of gender difference in hepatocellular carcinoma: the tip of the iceberg? *BioSci Trends*, 19(3):266-280. https://doi.org/10.5582/bst.2025.01103

Yan J, Xie B, Tian Y, et al., 2022. iTRAQ-based proteome profiling of differentially expressed proteins in insulin-resistant human hepatocellular carcinoma. *Front Cell Dev Biol*, 10:836041. https://doi.org/10.3389/fcell.2022.836041

Yang HY, Xu M, Chi XF, et al., 2017. Higher PKD3 expression in hepatocellular carcinoma (HCC) tissues predicts poorer prognosis for HCC patients. *Clin Res Hepatol Gastroenterol*, 41(5):554-563. https://doi.org/10.1016/j.clinre.2017.02.005

Yang T, Xing H, Wang GQ, et al., 2019. A novel online calculator based on serum biomarkers to detect hepatocellular carcinoma among patients with hepatitis B. *Clin Chem*, 65(12):1543-1553. https://doi.org/10.1373/clinchem.2019.308965

Yeaman C, Ayala MI, Wright JR, et al., 2004. Protein kinase D regulates basolateral membrane protein exit from *trans*-Golgi network. *Nat Cell Biol*, 6(2):106-112. https://doi.org/10.1038/ncb1090

Zhang JW, Zhang YJ, Wang JQ, et al., 2019. Protein kinase D3 promotes gastric cancer development through p65/6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3 activation of glycolysis. *Exp Cell Res*, 380(2):188-197. https://doi.org/10.1016/j.yexcr.2019.04.022

Zhang SY, Liu H, Yin MM, et al., 2020. Deletion of protein kinase D3 promotes liver fibrosis in mice. *Hepatology*, 72(5):1717-1734. https://doi.org/10.1002/hep.31176

Zhou J, Yu L, Gao X, et al., 2011. Plasma microRNA panel to diagnose hepatitis B virus–related hepatocellular carcinoma. *J Clin Oncol*, 29(36):4781-4788.

https://doi.org/10.1200/JCO.2011.38.2697

Zhou J, Sun H, Wang Z, et al., 2025. China Liver Cancer Guidelines for the Diagnosis and Treatment of Hepatocellular Carcinoma (2024 Edition). *Liver Cancer*, 14(6): 779-835.
https://doi.org/10.1159/000546574

Zhu LP, Lv B, Gao YF, et al., 2025. Lactucin alleviates liver fibrosis by regulating the TLR4-MyD88-MAPK/NF-κB signaling pathway through intestinal flora. *Arch Biochem Biophys*, 12:110341.
https://doi.org/10.1016/j.abb.2025.110341

Supplementary information

Materials and methods; Figs. S1–S3; Tables S1–S3