


RESEARCH

Open Access



Plasma lipidomic analysis reveals disruption of ether phosphatidylcholine biosynthesis and facilitates early detection of hepatitis B-related hepatocellular carcinoma

Yuyao Yuan^{1†}, Donghao Yin^{2†}, Xuemeng Yang², Di Liu², Hui Shan³, Juan Luo³, Xiuhui Li^{2*} and Yuxin Yin^{1,3*} 

Abstract

Background Hepatocellular carcinoma (HCC) is the third deadliest malignant tumor worldwide. Most patients are initially diagnosed as HCC at advanced stages and are too late for radical treatment by surgery, resulting in poor prognosis. Over 50% of the HCC patients are caused by hepatitis B virus (HBV) infection. Therefore, effective early identification of HCC in the high-risk population with HBV infection is crucial for early intervention of HCC.

Methods We employed plasma lipidomics to identify critical lipid classes associated with tumorigenesis in the high-risk population with HBV infection. Potential regulatory mechanisms are validated at multi-omic levels. A machine learning algorithm is used for feature selection and diagnostic modelling, and performance of the models is evaluated by ROC curves.

Results We unveiled varied profiles of plasma lipid metabolites in a cohort of 57 HBV-related HCC subjects, 57 HBV-related liver cirrhosis (LC) subjects and 61 chronic hepatitis B (CHB) subjects with matched age, sex and HBV status. We identified a correlation of the ether phosphatidylcholine (PC) synthesis with hepatocarcinogenesis in patients with HBV-related liver diseases. The diagnostic models achieved an area under ROC curve (AUC) of 0.849 for discriminating HCC from CHB and an AUC of 0.829 for discriminating HCC from LC.

Conclusions We illustrate the role of ether PC in hepatocarcinogenesis upon HBV infection and provide novel effective markers for early detection of HCC in a cohort with HBV infection.

Keywords Hepatitis B, Hepatocellular carcinoma, Plasma lipidomics, Ether phosphatidylcholine, Early detection, Machine learning

[†]Yuyao Yuan and Donghao Yin contributed equally to this work.

*Correspondence:

Xiuhui Li

lixihui@sohu.com

Yuxin Yin

yinyuxin@bjmu.edu.cn

¹ Institute of Systems Biomedicine, Department of Pathology, School of Basic Medical Sciences, Peking-Tsinghua Center for Life Sciences, Peking University Health Science Center, Beijing 100191, China

² Beijing Youan Hospital, Capital Medical University, Beijing 100069, China

³ Institute of Precision Medicine, Peking University Shenzhen Hospital, Shenzhen 518036, China



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Background

Hepatocellular carcinoma (HCC) is one of the most prevalent cancers worldwide, ranking as the third leading cause of cancer-related deaths globally in 2022 [1]. In China, HCC accounted for over 310,000 deaths in 2022, making it the second deadliest cancer type [2]. Major risk factors for HCC include alcohol consumption, diabetes, nonalcoholic steatohepatitis (NASH), and infection by hepatitis viruses, with hepatitis B virus (HBV) being the leading cause, responsible for over half of global HCC cases [3]. Particularly, the HBV infection rate among HCC patients in China reached 92.05% [4].

Despite significant advancements in HCC treatment, particularly with radical hepatectomy, over 70% of HCC patients are diagnosed at advanced stages, when curative treatment options are no longer feasible [3]. Current clinical screening methods for early HCC detection, including liver ultrasound imaging and serum alpha-fetoprotein (AFP) testing, are inadequate. Ultrasound has limited sensitivity for detecting small or early-stage HCC lesions, while AFP testing often fails to distinguish early HCC from other liver diseases such as cirrhosis or hepatitis [3, 5, 6]. For instance, studies have shown that up to 40% of HCC patients have normal AFP levels, and elevated AFP can also occur in non-malignant liver diseases [5–9]. Advanced imaging modalities like CT and MRI, although more sensitive, are costly and impractical for routine screening in high-risk populations [5, 6].

To overcome these challenges, several prediction scoring systems have been developed to assess HCC risk in HBV-infected populations by incorporating clinical and laboratory parameters, such as age, albumin, and AFP [7–9]. However, these scoring systems lack the sensitivity and specificity required for early detection. Similarly, multi-omics approaches, such as proteomics and metabolomics, have emerged as promising tools for HCC biomarker discovery. Proteomics studies [10–12] have provided valuable insights into the molecular mechanisms of HCC, while metabolomics studies [13–19] have demonstrated significant alterations in metabolic profiles, particularly in blood and tissue samples. However, many of these studies focus on general HCC populations and provide limited insights into the specific metabolic changes underlying HBV-related HCC.

Metabolic perturbation is a key feature of cancer. The liver, as the largest metabolic organ, plays a central role in nutrient metabolism, bile acid metabolism, and toxin clearance, making hepatic malignancies highly likely to affect systemic metabolic states [20]. Lipidomics, a rapidly advancing branch of metabolomics, offers a comprehensive view of lipid metabolism and its alterations in disease progression. Unlike previous studies that primarily focus on proteomics or general metabolomics, here we

employed lipidomics to analyze hydrophobic metabolites in HBV-infected patients, including those with chronic hepatitis B (CHB), HBV-related liver cirrhosis (LC), and HBV-related HCC. Through multi-omics integration, we identified significant elevations in ether phosphatidylcholines (ether PCs) and confirmed the dysregulation of ether PC biosynthesis pathways in HBV-HCC development. Finally, we developed a diagnostic model based on a panel of ether PCs selected by machine learning, which demonstrated high diagnostic accuracy for detecting early-stage HCC. This study provides a novel, non-invasive, and cost-effective strategy to facilitate early detection and intervention in HBV-HCC patients.

Methods

Study subjects and plasma collection

Subjects diagnosed with CHB, HBV-related LC and HBV-related HCC at Beijing Youan Hospital were enrolled from February, 2024 to April, 2024. Criteria for inclusion are listed as follows: i) Patients aged between 18 and 80 years. ii) Patients diagnosed with chronic hepatitis B (CHB), HBV-related cirrhosis (LC), or HBV-related hepatocellular carcinoma (HCC) based on the guidelines for the management of HBV-related diseases. iii) Patients with a history of chronic hepatitis B and meeting the diagnostic criteria for HBV-related cirrhosis or hepatocellular carcinoma. iv) Willingness to provide informed consent for participation in the study. Criteria for exclusion are listed as follows: i) Patients with severe systemic diseases involving the respiratory, cardiac, or central nervous systems, as well as those with autoimmune hepatitis, congenital, or hereditary liver diseases. ii) Patients with psychiatric or psychological disorders, including anxiety or depression. All HCC subjects were staged according to China liver cancer staging (CNLC) and over two thirds of the HCC subjects were at CNLC stage I.

Blood samples of enrolled subjects were collected by EDTA tubes and were kept at 4°C for less than 6 h before centrifugation to collect plasma. The plasma samples were stored at -80 °C until sample preparation for LC-MS analysis.

Sample collection and metabolite extraction

Quality control (QC) samples were obtained by forming a mixed pool of different samples. For untargeted lipidomic analysis, 100µL of the liquid-liquid extraction solution (chloroform-methanol 2:1, v/v) is added to 25µL of each serum sample including QC sample. Samples are vortexed for 30 s, vibrated at 1200 rpm for 8 min, and centrifuged at 12,000 rpm for 10 min. Lower organic phase containing hydrophobic metabolites are collected into new tubes and evaporated at room temperature under vacuum. The residue is dissolved in

25 μ L dissolving solution (chloroform–methanol 1:1, v/v), vortexed for 30 s and diluted by adding 75 μ L diluting solution (isopropanol–acetonitrile–H₂O 2:1:1, v/v/v). The mixture is then vortexed for 30 s, centrifuged at 12,000 rpm for 15 min, and supernatant transferred into vials for LC–MS analysis.

Untargeted lipidomic analysis

Untargeted lipidomic analysis was performed using liquid chromatography–mass spectrometry (LC–MS). The Ultimate 3000 UHPLC system (Thermo) and Acquity CSH C18 column (100 \times 2.1 mm i.d., 2.5 μ m, Waters) were used for reversed phase liquid chromatographic separation. Column temperature was set to 50 $^{\circ}$ C. Acetonitrile (LC–MS grade, Fisher Scientific, USA)–water (60/40, v/v) with 10 mM ammonium acetate (Sigma–Aldrich, St. Louis, MO, USA) and 0.1% formic acid (Sigma–Aldrich, St. Louis, MO, USA) was used as mobile phase A, and isopropanol (LC–MS grade, Fisher Scientific, USA)–acetonitrile (90/10, v/v) with 10 mM ammonium acetate and 0.1% formic acid was used as mobile phase B. The flow rate was set to 0.3 mL/min. The gradient of liquid phase was set as follows: 0 min–40% B; 2 min–43% B; 2.1 min–50% B; 10 min–60% B; 10.1 min–75% B; 16 min – 99% B; 17 min–99% B; 18 min–40% B; and 19 min–40% B.

Q-Exactive (hybrid quadrupole–Orbitrap mass spectrometer) coupled with heated electrospray ionization (HESI) source (Thermo Fisher Scientific) was used for mass analysis. Data dependent acquisition (DDA) mode was used. Each acquisition cycle consists of one survey scan (MS¹ scan) at 35,000 resolution from 190 to 1200 m/z, followed by ten MS/MS scans in HCD mode at 17,500 resolution. MS/MS parameters were set as follows: Automatic gain control target (AGC), 5e6 (maximum injection time 80 ms) for MS¹ scan and 1e5 (maximum injection time 70 ms) for MS/MS scan; Fixed first mass, 50 m/z; Dynamic exclusion, 8 s; Stepped normalized collision energy (step–NCE) to 15, 30, and 45. HESI ion source parameters were set as follows: spray voltage, 3.3 kV for positive ion mode and 3.0 kV for negative ion mode; ion source sheath gas, 40; aux gas, 10; capillary temperature, 320 $^{\circ}$ C; probe heater temperature, 300 $^{\circ}$ C; S–lens RF level, 55.5. QC samples were analyzed repeatedly in the batch of sample acquisition to evaluate the stability of the LC–MS instrument. All samples were acquired in the positive–negative switching ion mode.

Peak extraction, alignment, identification and quantification from raw data files were performed using the MS–DIAL software (version 4.70). Specifically, internal MS/MS spectra lipid libraries in MS–DIAL software was used for lipid identification. Characteristic fragments of

selected lipid features were manually checked in MS/MS spectra from raw data files.

SVM-based feature selection

SVM model (liblinear 2.20) was built to classify categories of enrolled subjects and select important features as previously reported [21, 22]. SVM was employed to build classification models in 1000 experiments of fourfold cross validation and to generate the weight representing importance in classification for all features.

$$\min_{w,b} \frac{\|w\|^2}{2}, s.t. (w^T x_i + b) y_i \geq 1$$

As shown in above equation, the inferred w could be regarded as the importance weight for each feature. A validation operation of feature selection was conducted to select top-ranking important features with the highest classification accuracy. Top 50 important features were analyzed to generate predictive models for feature selection, which was performed by increasing and selecting from the top-ranking feature one by one, e.g., selecting the Top 1 feature as the first model, Top 2 features as the second model, and then iterated to Top–N features as the N–th model. For performance evaluation, mean accuracies for each model (N=50) in feature selection were calculated after 100 times iterations of fourfold cross-validation.

Measurement of AFP

Serum AFP levels were measured using an automated chemiluminescence immunoassay (CLIA) system according to the manufacturer's protocol (Abbott ARCHITECT i2000SR, Abbott Laboratories, Chicago, IL, USA). The detection range of the assay was 0.5–350 ng/mL. All samples were processed in duplicate, and strict quality control was maintained throughout the testing process.

Multi-omic analysis

Datasets of HCC from The Cancer Genome Atlas (TCGA) and Clinical Proteomic Tumor Analysis Consortium (CPTAC) databases were used for RNA and protein analysis, respectively, via the integrated online analysis platform at <https://ualcan.path.uab.edu/index.html> [23]. The RNA levels and protein levels of HCC patients in primary tumor tissue compared to normal adjacent tissue were analyzed. Basic patient characteristics of original datasets in TCGA (TCGA–LIHC) and CPTAC (PDC000198) which we used were exported from official websites and summarized in Table S7 of Additional file 2.

Statistical analysis

Metaboanalyst (<https://www.metaboanalyst.ca/>) was used for preliminary identification of differential lipid features by ANOVA, PCA, PLSDA and hierarchy cluster analysis. Features with $FDR < 0.05$ (Benjamini–Hochberg method) in ANOVA analysis were subjected to further analysis. Corresponding R2Y and Q2Y values of PLSDA analysis were listed in Table S6 of Additional file 2.

The MATLAB software (R2022b) was used to perform SVM-based feature selection, SVM modeling and receiver operator characteristic (ROC) analysis.

The GraphPad Prism 9.0.0 and R software were used for data analysis and visualization, and R packages including “Mfuzz” “psych” “Rtsne” were used for mfuzz clustering, correlation analysis and t-SNE visualization, respectively. Modeling by multiple ML algorithms was performed using the following functions in the tidymodels framework of R software, including XGBoost (boost_tree()), Decision Tree (decision_tree()), Logistic Regression (logistic_deg()), KNN (nearest_neighbor()), Random Forest (rand_forest()), and SVM (svm_linear()).

Results

Characteristics of study subjects

To uncover the circulatory metabolic changes that occur in cancer development in high-risk population with HBV infection, subjects diagnosed with CHB, HBV-related LC and HBV-related HCC were enrolled in this study. Summarized and detailed characteristics of enrolled subjects are listed in Table 1 and Table S1-2, respectively. Basic information including age, sex and HBV infection were of no statistical significance among the three groups. To obtain a full picture of the metabolic disturbance in the

process of tumorigenesis in HBV-infected population, we sought to characterize the global plasma lipid metabolites in CHB, LC and HCC participants by lipidomic analysis. We identified 1728 features in the positive electrospray ionization mode (ESI+) and 939 features in the negative electrospray ionization mode (ESI-).

Plasma lipidomic profiling unveils differential lipids from multiple lipid classes

For a global view of the obtained lipidomic data, we performed PLSDA analysis and t-SNE analysis (Fig. S1) to visualize all lipids identified. As shown by our data, CHB illustrated obvious separation from the other two groups, while the group of LC and HCC exhibited less difference. Two-group comparisons were then performed to visualize global difference between CHB/HCC and LC/HCC, respectively (Fig. 1A-B). Corresponding R2Y and Q2Y values in PLSDA analysis are listed in Table S6 of Additional File 2, illustrating a good predictive power for discriminating HCC from CHB ($R2Y$ and $Q2Y > 0.3$) while a poor predictive power for discriminating HCC from LC (the difference between $R2Y$ and $Q2Y$ was larger than 0.3).

Next, ANOVA analysis was performed to identify the differential lipid features ($FDR < 0.05$) among the three groups, and 508 features in ESI+ mode and 359 features in ESI- mode were subjected to further analysis. Lipid classes of all differential features were then observed. The number of differential lipid features in each lipid class detected in ESI+ or ESI- mode is listed in Table S3, and lipid classes with the top 5 most features in each ion mode are demonstrated in Fig. 2.

Table 1 Clinical characteristics of enrolled subjects

		Overall	CHB	LC	HCC	P
n		175	61	57	57	
Age (mean (SD))		57.74 (9.80)	57.79 (8.94)	56.39 (9.85)	59.05 (10.60)	0.35
Sex, n (%)	Male	131(74.86)	47(77.05)	40(70.18)	44(77.19)	0.611
	Female	44(25.14)	14(22.95)	17(29.82)	13(22.81)	
HBsAg (%)	missing	1 (0.6)	0 (0.0)	1 (1.8)	0 (0.0)	0.489
	-	14 (8.0)	3 (4.9)	5 (8.8)	6 (10.5)	
	+	160 (91.4)	58 (95.1)	51 (89.5)	51 (89.5)	
HBcAb (%)	missing	1 (0.6)	0 (0.0)	1 (1.8)	0 (0.0)	0.271
	-	5 (2.9)	0 (0.0)	2 (3.5)	3 (5.3)	
	+	169 (96.6)	61 (100.0)	54 (94.7)	54 (94.7)	
AFP (ng/mL), median [IQR]		2.45 [1.60, 4.64]	2.29 [1.39, 2.73]	2.50 [1.80, 4.60]	3.30 [1.68, 34.40]	0.007
CNLC stage	I				38(66.7)	
	II				12(21.1)	
	III				6(10.5)	
	IV				1(1.8)	

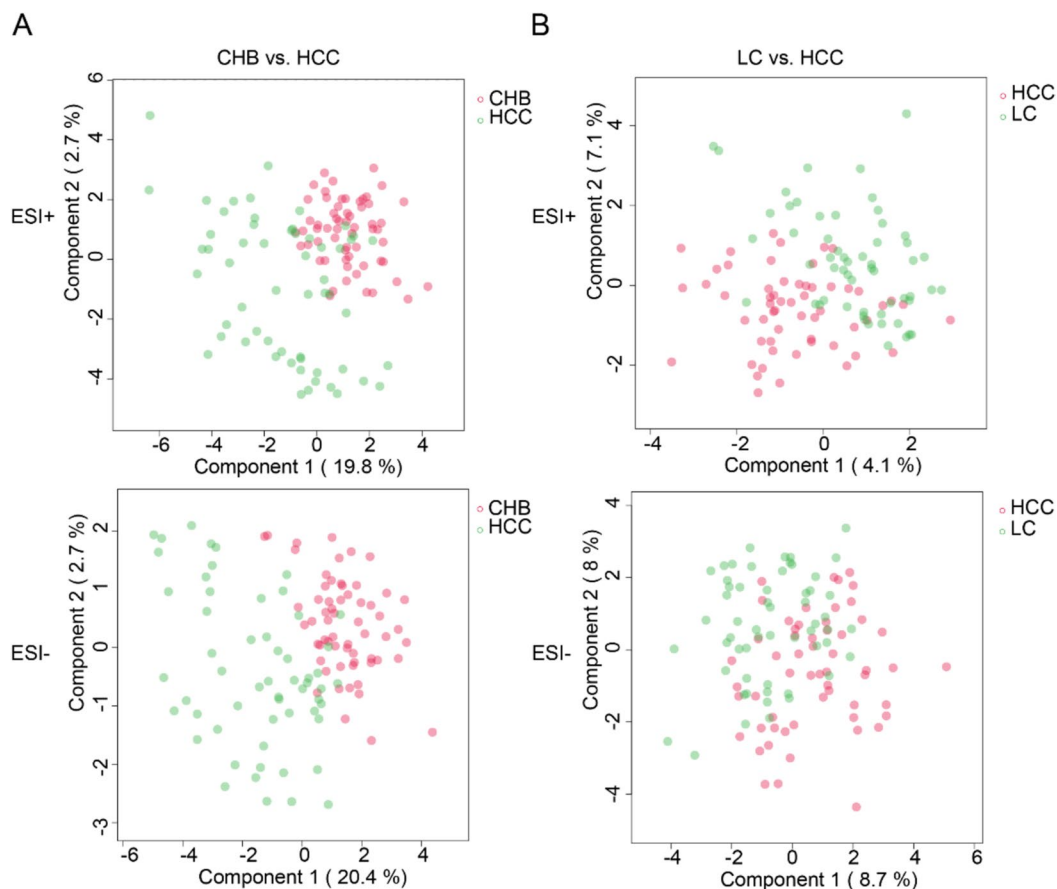


Fig. 1 Overview of untargeted lipidomic data by PLS-DA visualization in groups of CHB/HCC (A) or LC/HCC (B)

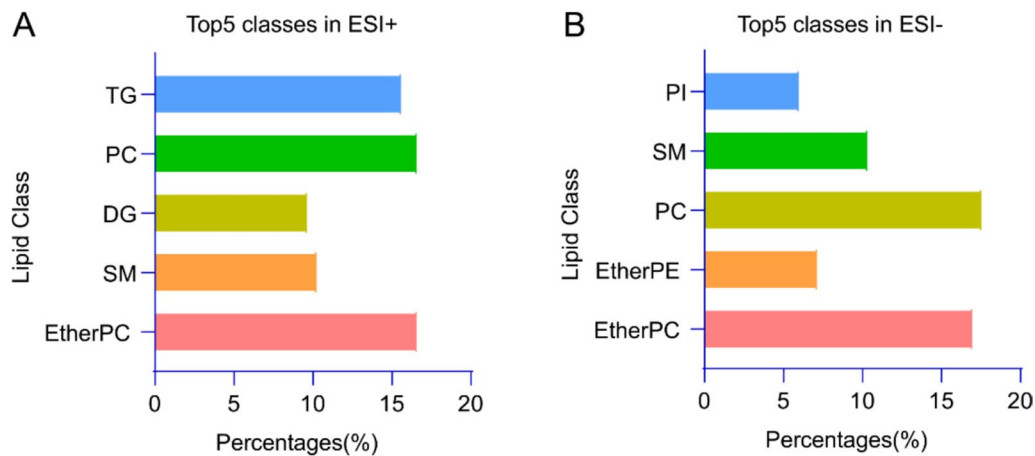


Fig. 2 Lipid classes with the top 5 most differential lipid features of ESI+ (A) or ESI- (B) mode

Plasma levels of ether PCs closely correlate with hepatic carcinogenesis

To identify the crucial metabolic changes during hepatic carcinogenesis in subjects with HBV infection, we

thereby performed mfuzz clustering using the differential lipids to observe the variation patterns of lipids in different clusters. Five clusters were obtained in ESI+ or ESI- mode (Fig. 3A-B). In clusters illustrating consecutive

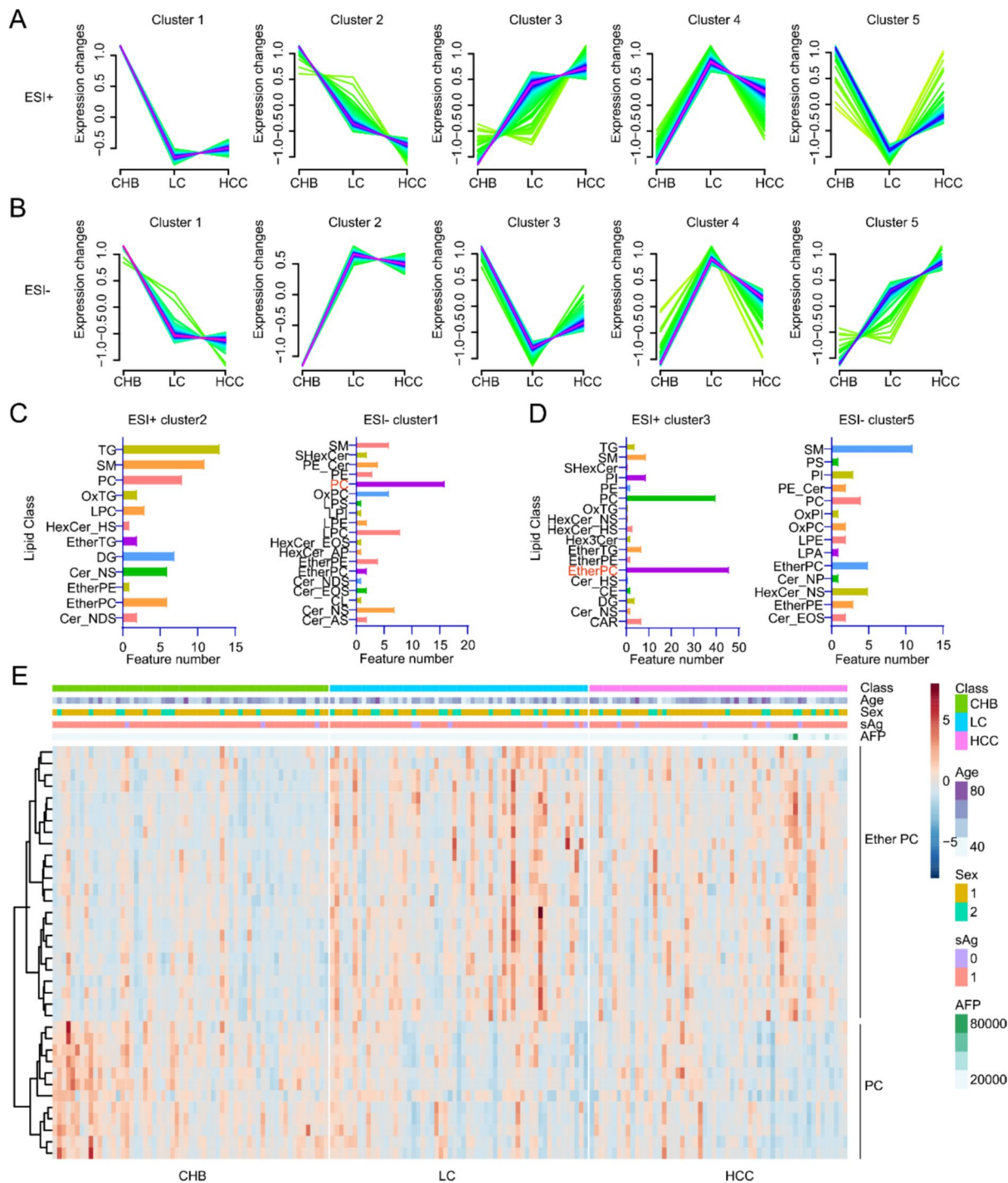


Fig. 3 Major lipid classes potentially correlated with hepatocarcinogenesis. Mfuzz clustering for variation pattern along the trajectory of CHB-LC-HCC using all identified lipid features in ESI + mode (A) or ESI- mode (B). Bar plot illustration of the feature number of each lipid class in clusters with consecutive decline (C) and elevation (D) in the trajectory of CHB-LC-HCC and heatmap illustration of ether PC and PC levels in selected clusters, along with clinical characteristics of Age, sex, sAg and AFP (E)

decline along the trajectory of CHB-LC-HCC as disease development, including cluster 2 in ESI+ mode and cluster 1 in ESI- mode, phosphatidylcholine (PC)

was identified as the lipid class with the most features (Fig. 3C). While in clusters illustrating consecutive elevation, including cluster 3 in ESI+ mode and cluster 5 in

ESI- mode, ether PC was identified as the lipid class with the most features (Fig. 3D). Abundances of PCs and ether PCs in ESI- cluster 1 and ESI+ cluster 3, respectively, are presented in Fig. 3E. Collectively, these data indicate that increased ether PC levels and decreased PC levels in plasma possibly account for the malignant conversion in hepatitis B-related liver diseases.

Integrated analysis by multi-omic data illustrates dysregulation of the ether PC biosynthetic pathway

To further validate the role of ether PC biosynthesis in hepatocarcinogenesis, we then investigated the expression of key enzymes modulating biosynthesis of ether PC. GNPAT and AGPS, catalyzing key steps of ether bond formation in liver peroxisomes [23], illustrated obvious upregulation in HCC tissue samples compared to normal controls in both RNA levels (Fig. 4A) and protein levels (Fig. 4B). PEMT, accounting for conversion from ether PE to ether PC, markedly declined in HCC samples, while CHPT1, accounting for conversion from ether DG to ether PC, significantly increased in HCC samples (Fig. 4A-B). FAR1 and FAR2, which have been previously known to show negative responses to levels of ether PC, showed decreased RNA levels in HCC while no obvious change in protein levels (Fig. 4A-B). The alterations of these key enzymes in ether PC biosynthesis are summarized and illustrated in Fig. 4C. Together, these data suggest that biosynthesis of ether PC is indeed enhanced in the process of hepatocarcinogenesis.

Ether PCs show superior classification performances in machine-learning models compared to alpha fetoprotein

Given that our data demonstrated a critical role of ether PC in hepatocarcinogenesis, we developed SVM-based machine-learning models using abundances of all identified ether PCs to detect HCC subjects from CHB or LC subjects. Firstly, a feature selection strategy was performed to determine the importance of individual ether PC lipid feature and the optimal number of lipid feature combination in the models. The average accuracies of fourfold cross validation models in 100 tests using 1–50 top-weighted features are illustrated in Fig. 5A–B. For discriminating HCC from CHB, using 11 top-weighted features, the model reached the highest accuracy of 81.83% (Fig. 5A). For discriminating HCC from LC, using 31 top-weighted features, the model reached the highest accuracy of 77.86% (Fig. 5B). The selected features and corresponding weight in SVM models are listed in Table S4. The ROC curve for classification of HCC from CHB using 11 selected features for SVM modeling showed an area under ROC curve (AUC) of 0.849 (Fig. 5C), and the ROC curve for classification of HCC

from LC using 31 selected features for SVM modeling showed an AUC of 0.829 (Fig. 5D). To further determine the best ML algorithm for modeling, we built machine learning models based on several ML algorithms and depicted ROC curves for comparison. The performances of modeling using selected features were compared by AUC (Fig. S2 of Additional file 1). We can see that compared to other algorithms, the SVM model illustrated the second highest AUC for discriminating CHB vs. HCC and the highest AUC for discriminating LC vs. HCC. The diagnostic performance of alpha fetoprotein (AFP) was also evaluated for CHB vs. HCC (Fig. 5E) and LC vs. HCC (Fig. 5F) by ROC curves, respectively, demonstrating limited value of AFP to detect HCC in our study cohorts. Together, these results indicate that SVM modeling using selected ether PCs shows better performance to detect HBV-related HCC than the conventional marker AFP.

Levels of plasma ether PCs are significantly correlated with indicators of liver function

To further investigate the biological significance of ether PCs, the correlations between paired plasma ether PC levels and clinical results of examination or lab tests were determined by spearman correlation. The ether PCs in ESI+ cluster3 mentioned earlier with an elevating trend in the process of hepatocarcinogenesis and clinical indicators for liver function were included for analysis (Fig. 6 and Table S5). The ether PCs that most significantly associated with indicators of liver function included PC O-33:1, PC O-32:2, PC O-36:4, PC O-36:2, PC O-32:0, PC O-34:1, PC O-32:1 and PC O-38:6. Our data revealed massive correlation between ether PCs and indicators of liver function reflecting liver cell injury, liver excretion function, liver reserve function and liver interstitial changes. We unveil the value of plasma ether PCs in evaluating the liver function, which may ultimately facilitate early detection of hepatocarcinogenesis in HBV-infected population.

Discussion

In this study, we employed a lipidomic analytic approach to delineate the global circulatory lipid metabolites associated with hepatocarcinogenesis in a cohort with HBV infection. Through mfuzz clustering, we identified marked variations in plasma lipids, particularly in the class of ether phosphatidylcholines (ether PCs), and further verified enhanced biosynthesis of ether PCs by integrating multi-omic datasets, including transcriptomic and proteomic data from HCC tissue samples. Ultimately, we developed a diagnostic model utilizing ether PCs selected via a machine-learning-based feature selection strategy, which demonstrated high efficiency in differentiating HCC patients from CHB and LC subjects.

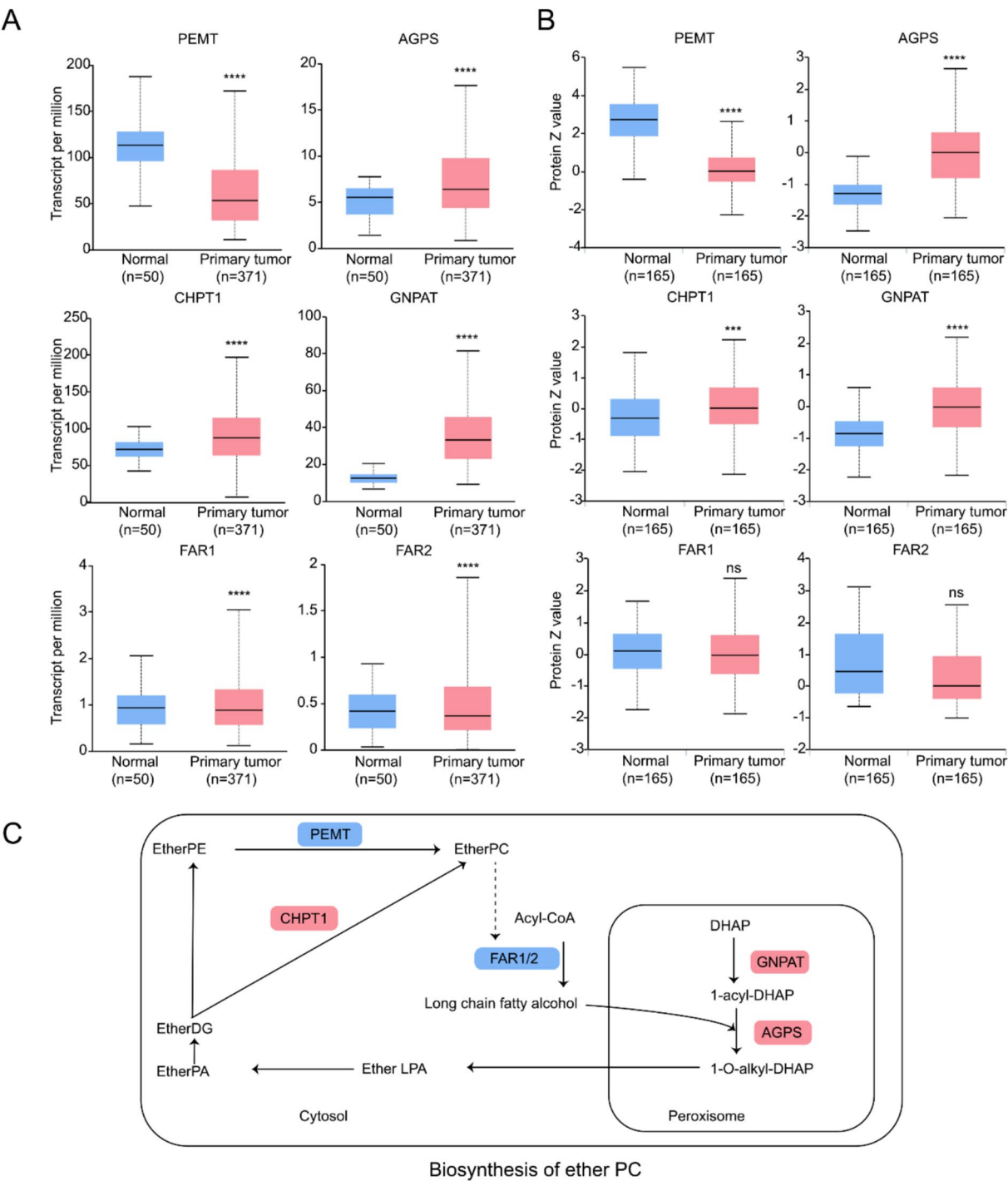


Fig. 4 Validation of altered ether PC biosynthesis by multi-omic analysis. Comparison of RNA levels (**A**) and protein levels (**B**) of key enzymes involved in ether PC biosynthesis between primary HCC tissue samples and normal controls in TCGA datasets and CPTAC datasets, respectively (***, $P < 0.001$; ****, $P < 0.0001$). Graphic illustration of molecular dysregulation in ether PC biosynthesis in HCC (**C**). Upregulated enzymes are marked in red and downregulated are marked in blue. PEMT, phosphatidylethanolamine N-methyltransferase; AGPS, alkyl DHAP synthase; CHPT1, choline phosphotransferase 1; GNPAT, DHAP acyltransferase; Far1/2, fatty acyl-CoA reductase 1 or 2

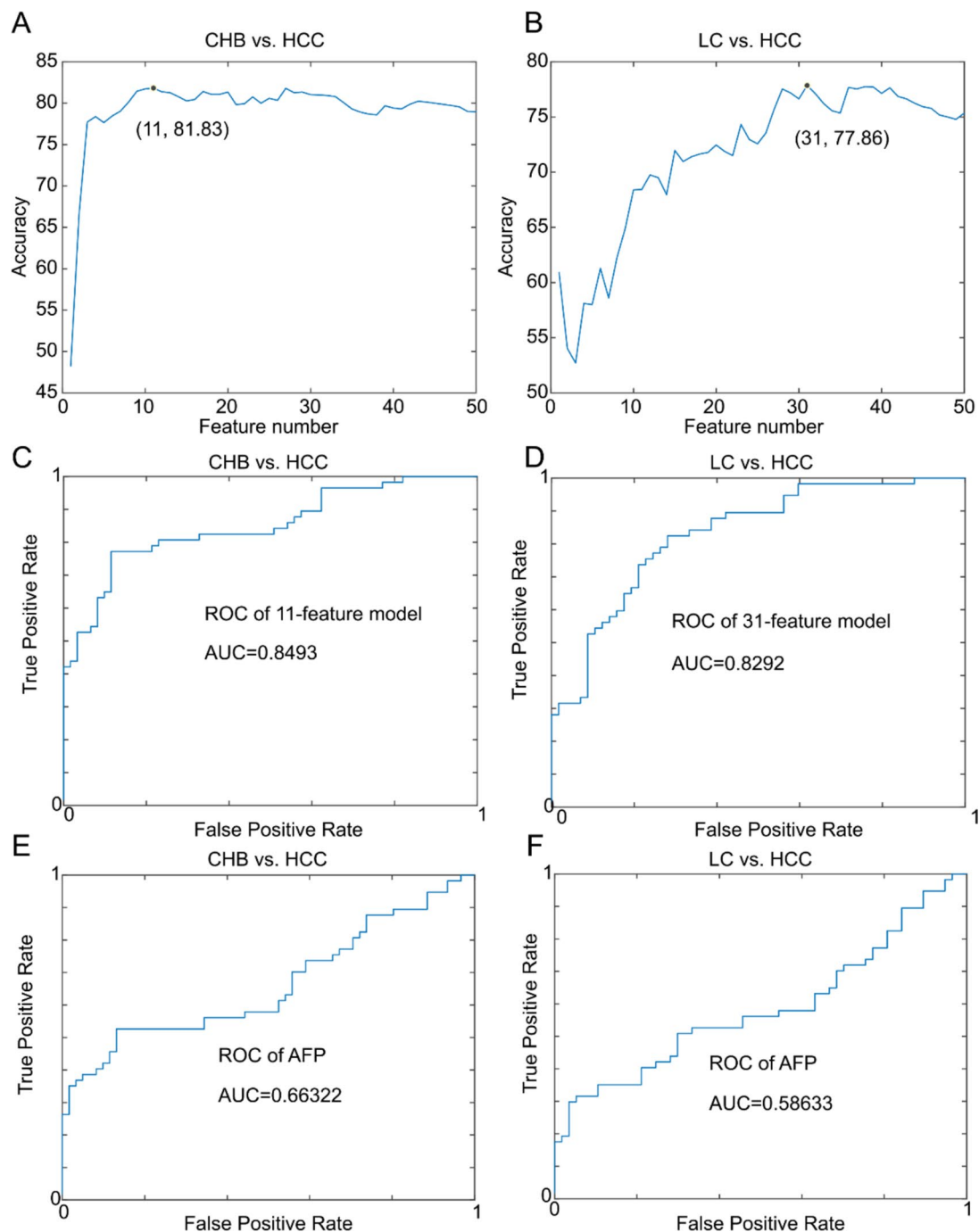


Fig. 5 Diagnostic performance of ML-selected lipid features. Illustration of average accuracies of SVM models for classifying HCC from CHB (A) or classifying HCC from LC (B) using 1–50 top-weighted features in 100 fourfold cross validations. Labeled coordinate indicates the optimal feature number achieving the highest accuracy. Illustration of ROC curves to classify CHB vs. HCC using 11 selected features (C) or AFP concentrations (E). Illustration of ROC curves to classify LC vs. HCC using 31 selected features (D) or AFP concentrations (F)

In the field of HCC diagnosis, various methods have been explored. Traditional methods based on serum biomarkers, such as alpha-fetoprotein (AFP), are widely used in clinical practice but have significant limitations

in sensitivity and specificity, especially for early HCC detection. Other clinical methods including ultrasound imaging, CT and MRI, also demonstrated deficiency in sensitivity, cost and availability for routine screening,

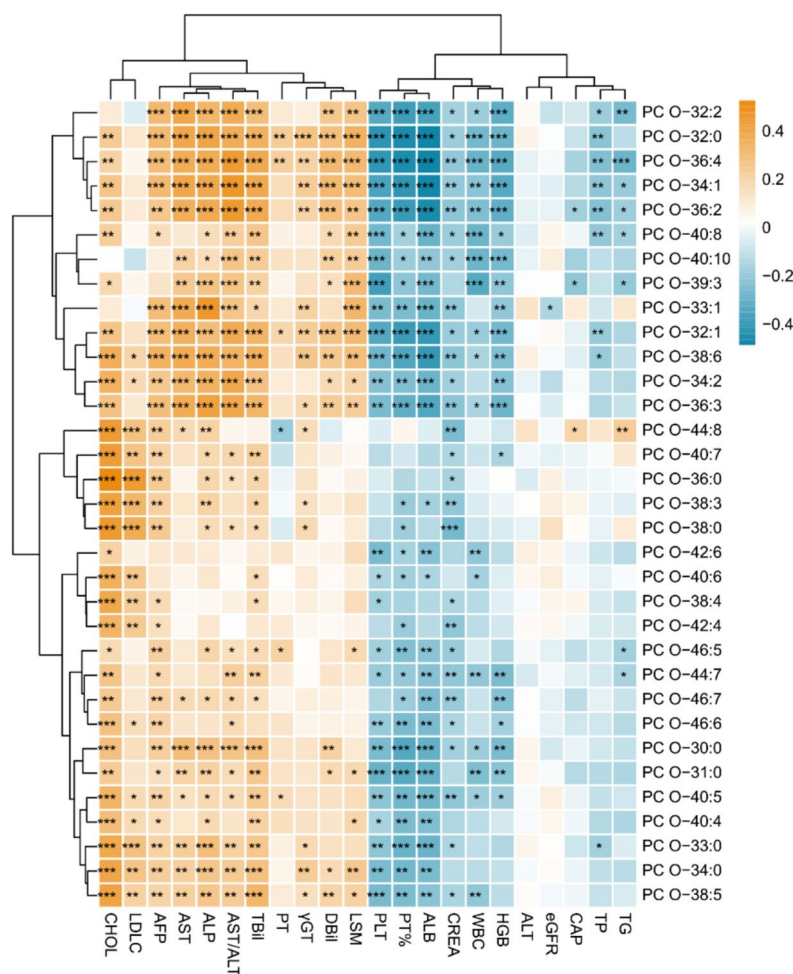


Fig. 6 Heatmap illustration of the correlation between ether PCs and selected clinical indicators. The color represents the correlation coefficient, red as positively correlated and blue as negatively correlated (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$). TG, triglycerides; TP, total protein; CAP, controlled attenuation parameter; eGFR, glomerular filtration rate; ALT, alanine aminotransferase; HGB, hemoglobin; WBC, white blood cell count; CREA, creatinine; ALB, albumin; PT%, prothrombin activity; PLT, platelet; LSM, liver stiffness measurement; DBil, direct bilirubin; γGT, γ-glutamyl transpeptidase; PT, prothrombin time; TBil, total bilirubin; ALP, alkaline phosphatase; AST, aspartate aminotransferase; AFP, alpha fetoprotein; LDL-C, low density lipoprotein cholesterol; CHOL, total cholesterol

highlighting the urgency for novel biomarkers and diagnostic methods. The role of metabolic drivers in tumorigenesis have been recently reviewed [24], that metabolic remodeling may begin at stages of early tumorigenesis, providing the theoretical basis for identifying metabolite markers for early detection of cancer. Multiple previous metabolomic studies have identified potential markers of HCC [14, 25–27], yet limited studies focused on HBV-related HCC in well-designed schemes. Here we uncover new markers of HBV-related HCC via lipidomic analysis in patients with HBV infection. The samples we used for this study were collected within a short time interval, and subjects in groups of CHB, LC and HCC were matched in age, sex and HBV infection, excluding potential confounding factors to ensure the identification of truly

important metabolite features. Moreover, the majority of subjects in the HCC group are at early stages, underscoring the value of selected markers for early HCC detection.

Studies on HCC diagnosis by machine learning have been recently reviewed [28], in which various datatypes were used, including clinical data, imaging data, pathology data, and gene sequencing data. Despite that multiple studies have used plasma lipidomics to explore novel markers of early HCC [29–31], few studies have used combined methods of plasma lipidomics and machine learning to identify novel diagnostic markers. Lewinska et. al [32] employed lipidomics and ML-based feature selection in detection of NAFLD-HCC and identified increased fatty acid uptake in NAFLD-HCC. Powell et.

al [33] analyzed serum metabolomics and lipidomics in a small cohort of 28 HCC and 30 cirrhosis with ML algorithms, while limited coverage of metabolites and lipids and sample numbers might lead to over-optimistic performances of ML classifications due to model overfitting. Here in this study, aiming to identify lipid characteristics and potential diagnostic markers of HBV-related HCC, a relatively larger cohort of 61 CHB, 57 LC and 57 HCC was established and ML-based feature selection was performed to enhance the robustness of selected markers.

Regarding the role of lipid metabolism in HCC development, previous studies have revealed disrupted lipid metabolism and significant alteration of multiple lipid classes [29–31, 34, 35]. Decrease of circulating polyunsaturated PCs was identified in HBV-related HCC [29], which is in consistency with our results that most PCs in ESI- cluster 1 with consecutive decline were polyunsaturated. Moreover, declined levels of LPC were observed in HCV-related HCC [30], which is also the case in HBV-related HCC as our data suggested, indicating that down-regulation of LPC might be a common feature of hepatitis virus-related HCC. Here in this study, we specifically targeted HBV-related HCC and controlled for confounding factors such as age, sex, and HBV status. Our results are consistent with these findings that lipid remodeling occurs during liver disease progression, yet we provide novel evidence linking ether PC dysregulation with hepatocarcinogenesis and we uniquely highlight ether PCs as potential biomarkers with significant diagnostic implications for HBV-HCC.

Ether PC, also termed as 1-O-alkyl-2-acyl-GPC, is a type of peroxisome-derived glycerophospholipid with the acyl chain at the sn-1 position attached to the backbone by an ether bond [36, 37], playing important roles in membrane structuring and cell signaling. Significant elevation of ether PC has been reported in obesity and systemic inflammation [38], raising the possibility that elevated ether PC levels reflect general poor health rather than cancer-specific metabolic remodeling. To address this, we propose future studies to control confounding factors such as BMI, inflammation, and metabolic comorbidities. Additionally, the role of plasmalogens, a subclass of ether glycerophospholipids containing a vinyl-ether linked alkyl chain at the sn-1 position, have been recently identified to play a role in neurodegenerative and cardiometabolic diseases [36]. Given that the role of ether lipids in membrane trafficking has been proposed [39], combined with previous evidence that impaired membrane trafficking in hepatocytes may lead to HCC [40] and increased ether PC in HCC patients according to our data, we hypothesize that increased ether PC might promote tumorigenesis via dysregulated membrane trafficking. Considering that the specific role of ether PCs in

hepatocarcinogenesis remains poorly understood, our data may provide future directions for investigations into the biological roles of ether PCs.

While our findings demonstrate the diagnostic potential of ether PCs, we acknowledge the challenges and limitations inherent to lipidomics, as well as the need for further validation. Firstly, as a single-centered study, our study is limited in sample size, and the conclusions and the utility of the selected markers derived from this cohort still need further validation in external cohorts from other medical centers. Secondly, the absence of chemical standard or isotope-labeled standard for each lipid feature might limit the precision of ID confirmation and quantitation of selected ether PCs. Additionally, the identification of lipids heavily depends on the software and libraries used for data analysis. Although we utilized the MS-DIAL software with integrated lipid libraries, the variability among software platforms could introduce uncertainties. These limitations underscore the need for cross-validation across platforms and the inclusion of robust standards in future studies [41–43]. Furthermore, novel separation strategies and techniques such as ion mobility and ozone-mediated cleavage and derivatization may identify the structures of indicated markers more precisely [43], thus the translational opportunity of selected markers in varied cohorts would be eventually enhanced.

Conclusions

Taken together, we identify a role of ether PC in hepatocarcinogenesis and show the value of plasma ether PCs in early detection of HCC in patients with HBV-related liver diseases in this cohort. Our data may provide novel targets for early intervention of malignant conversion and a novel method for early detection of HCC in the high-risk population with HBV infection.

Abbreviations

HCC	Hepatocellular carcinoma
HBV	Hepatitis B virus
PC	Phosphatidylcholine
ROC	Receiver operating characteristic
AUC	Area under ROC curve
LC	Liver cirrhosis
CHB	Chronic hepatitis B
AFP	Alpha fetoprotein
LC-MS	Liquid chromatography-mass spectrometry

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12944-025-02475-z>.

Additional file 1: Fig. S1. Overview of the results of untargeted lipidomic analysis by PLSDA visualization (A) and T-SNE visualization (B) in the groups of CHB, HCC and LC. Fig. S2. Comparison of ROC curves of classifying (A) CHB from HCC and (B) LC from HCC by ML algorithms including extreme gradient boosting (XGboost), decision tree (DT), logistic

regression (LR), neural network (NNet), K-nearest neighbor (KNN), random forest (RF) and support vector machine (SVM).

Additional file 2: Table S1. Summary of clinical characteristics and lab tests results of enrolled subjects. Table S2. Clinical characteristics and lab tests of individual study subjects. Table S3. Number of differential lipids in each lipid class detected in ESI+ or ESI- ion mode. Table S4. Selected ether PCs and corresponding weights for SVM modeling of CHB vs. HCC or LC vs. HCC. Table S5. Correlation coefficients between ether PCs and clinical lab tests reflecting liver function. Table S6. Parameters of PLS-DA analysis. Table S7. Summary of characteristics of HCC subjects from TCGA and CPTAC datasets.

Acknowledgements

Not applicable.

Authors' contributions

Y.Yuan, D.Y., X.Y., D.L., H.S., J.L., X.L. and Y.Yin designed the study; Y.Yuan performed LC-MS analysis; Y.Yuan and D.Y. collected clinical data; D.Y., X.Y., D.L., H.S., J.L., X.L. and Y.Yin provided clinical samples; Y.Yuan and D.Y. analyzed the data; Y.Yuan and D.Y. wrote the original draft; X.L. and Y.Yin revised the manuscript; All authors read and approved the final manuscript.

Funding

This work was supported by grants to Y.Yin including the National Key Research and Development Program of China (2021YFA1300601), National Natural Science Foundation of China (key grant 82030081) and the Lam Chung Nin Foundation for Systems Biomedicine.

Data availability

Data is provided within the manuscript or supplementary information files.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of Beijing Youan Hospital (No. LL-2024-130-K). Informed consents were obtained from all study subjects. All methods were performed in accordance with relevant guidelines and regulations.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 27 November 2024 Accepted: 10 February 2025

Published online: 24 February 2025

References

- Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2024;74(3):229–63.
- Han B, Zheng R, Zeng H, Wang S, Sun K, Chen R, Li L, Wei W, He J. Cancer incidence and mortality in China, 2022. *J Natl Cancer Cent*. 2024;4(1):47–53.
- Llovet JM, Kelley RK, Villanueva A, Singal AG, Pikarsky E, Roayaie S, Lencioni R, Koike K, Zucman-Rossi J, Finn RS. Hepatocellular carcinoma. *Nature Rev Dis Prim*. 2021;7(1):6.
- Cao WH, Zhang YQ, Li XX, Zhang ZY, Li MH. Advances in immunotherapy for hepatitis B virus associated hepatocellular carcinoma patients. *World J Hepatol*. 2024;16(10):1158–68.
- Colli A, Fraquelli M, Casazza G, Massironi S, Colucci A, Conte D, Duca P. Accuracy of ultrasonography, spiral CT, magnetic resonance, and alpha-fetoprotein in diagnosing hepatocellular carcinoma: a systematic review. *Am J Gastroenterol*. 2006;101(3):513–23.
- Colli A, Nadarevic T, Miletic D, Giljaca, Fraquelli M, Stimac D, Casazza G. Abdominal ultrasound and alpha-fetoprotein for the diagnosis of hepatocellular carcinoma in adults with chronic liver disease. *Cochrane Database Syst Rev*. 2021(4):CD013346.
- Bouchard MJ, Navas-Martin S. Hepatitis B and C virus hepatocarcinogenesis: lessons learned and future challenges. *Cancer Lett*. 2011;305(2):123–43.
- Shen Y, Liu J, Han Z, Jiang W, Cui H, Xun Y. Risk prediction models for hepatocellular carcinoma in chronic hepatitis B patients on antiviral therapy: A meta-analysis. *Clin Res Hepatol Gastroenterol*. 2022;46(6):101930.
- Wang HW, Chen CY, Lai HC, Hu TH, Su WP, Lu SN, Hung CH, Chuang PH, Wang JH, Chen CH, Peng CY. Prediction model of hepatocellular carcinoma in patients with hepatitis B virus-related compensated cirrhosis receiving antiviral therapy. *Am J Cancer Res*. 2023;13(2):526–37.
- Jiang Y, Sun A, Zhao Y, Ying W, Sun H, Yang X, Xing B, Sun W, Ren L, Hu B, et al. Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. *Nature*. 2019;567(7747):257–61.
- Gao Q, Zhu H, Dong L, Shi W, Chen R, Song Z, Huang C, Li J, Dong X, Zhou Y, et al. Integrated Proteogenomic Characterization of HBV-Related Hepatocellular Carcinoma. *Cell*. 2019;179(2):561–577 e522.
- Xu M, Xu K, Yin S, Chang C, Sun W, Wang G, Zhang K, Mu J, Wu M, Xing B, et al. In-depth serum proteomics reveals the trajectory of hallmarks of cancer in hepatitis B virus-related liver diseases. *Mol Cell Proteomics*. 2023;22(7):100574.
- Liu Y, Hong Z, Tan G, Dong X, Yang G, Zhao L, Chen X, Zhu Z, Lou Z, Qian B, et al. NMR and LC/MS-based global metabolomics to identify serum biomarkers differentiating hepatocellular carcinoma from liver cirrhosis. *Int J Cancer*. 2014;135(3):658–68.
- Luo P, Yin P, Hua R, Tan Y, Li Z, Qiu G, Yin Z, Xie X, Wang X, Chen W, et al. A Large-scale, multicenter serum metabolite biomarker identification study for the early detection of hepatocellular carcinoma. *Hepatology*. 2018;67(2):662–75.
- Chen T, Xie G, Wang X, Fan J, Qiu Y, Zheng X, Qi X, Cao Y, Su M, Wang X, et al. Serum and urine metabolite profiling reveals potential biomarkers of human hepatocellular carcinoma. *Mol Cell Proteomics*. 2011;10(7):M110004945.
- Xiao JF, Varghese RS, Zhou B, Nezami Ranjbar MR, Zhao Y, Tsai TH, Di Poto C, Wang J, Goerlitz D, Luo Y, et al. LC-MS based serum metabolomics for identification of hepatocellular carcinoma biomarkers in Egyptian cohort. *J Proteome Res*. 2012;11(12):5914–23.
- Gong ZG, Zhao W, Zhang J, Wu X, Hu J, Yin GC, Xu YJ. Metabolomics and eicosanoid analysis identified serum biomarkers for distinguishing hepatocellular carcinoma from hepatitis B virus-related cirrhosis. *Oncotarget*. 2017;8(38):63890–900.
- Pan HY, Wu QQ, Yin QQ, Dai YN, Huang YC, Zheng W, Hui TC, Chen MJ, Wang MS, Zhang JJ, et al. LC/MS-based global metabolomic identification of serum biomarkers differentiating hepatocellular carcinoma from chronic hepatitis b and liver cirrhosis. *ACS Omega*. 2021;6(2):1160–70.
- Zhang L, Wu GY, Wu YJ, Liu SY. The serum metabolic profiles of different Barcelona stages hepatocellular carcinoma associated with hepatitis B virus. *Oncol Lett*. 2018;15(1):956–62.
- Reinke H, Asher G. Circadian clock control of liver metabolic functions. *Gastroenterology*. 2016;150(3):574–80.
- Wang G, Yao H, Gong Y, Lu Z, Pang R, Li Y, Yuan Y, Song H, Liu J, Jin Y, et al. Metabolic detection and systems analyses of pancreatic ductal adenocarcinoma through machine learning, lipidomics, and multi-omics. *Sci Adv*. 2021;7(52):eabh2724.
- Yuan Y, Zhao Z, Xue L, Wang G, Song H, Pang R, Zhou J, Luo J, Song Y, Yin Y. Identification of diagnostic markers and lipid dysregulation in oesophageal squamous cell carcinoma through lipidomic analysis and machine learning. *Br J Cancer*. 2021;125(3):351–7.
- Chandrashekar DS, Karthikeyan SK, Korla PK, Patel H, Shovon AR, Athar M, Netto GJ, Qin ZS, Kumar S, Manne U, et al. UALCAN: An update to the integrated cancer data analysis platform. *Neoplasia*. 2022;25:18–27.
- Zhang S, Xiao X, Yi Y, Wang X, Zhu L, Shen Y, Lin D, Wu C. Tumor initiation and early tumorigenesis: molecular mechanisms and interventional targets. *Signal Transduct Target Ther*. 2024;9(1):149.

25. Wu T, Zheng X, Yang M, Zhao A, Li M, Chen T, Panee J, Jia W, Ji G. Serum lipid alterations identified in chronic hepatitis B, hepatitis B virus-associated cirrhosis and carcinoma patients. *Sci Rep*. 2017;7:42710.
26. Rashid MM, Varghese RS, Ding Y, Ressom HW. Biomarker discovery for hepatocellular carcinoma in patients with liver cirrhosis using untargeted metabolomics and lipidomics studies. *Metabolites*. 2023;13(10):1047.
27. Feng N, Yu F, Yu F, Feng Y, Zhu X, Xie Z, Zhai Y. Metabolomic biomarkers for hepatocellular carcinoma: A systematic review. *Medicine (Baltimore)*. 2022;101(3):e28510.
28. Feng SJ, Wang JH, Wang LH, Qiu QX, Chen DD, Su H, Li XL, Xiao Y, Lin CY. Current status and analysis of machine learning in hepatocellular carcinoma. *J Clin Transl Hepatol*. 2023;11(5):1184–91.
29. Buechler C, Aslanidis C. Role of lipids in pathophysiology, diagnosis and therapy of hepatocellular carcinoma. *Biochim Et Biophys Acta-Mol Cell Biol Lipids* 2020;1865(5):158658.
30. Caponigro V, Tornesello AL, Merciai F, La Gioia D, Salviati E, Basilicata MG, Musella S, Izzo F, Megna AS, Buonaguro L, et al. Integrated plasma metabolomics and lipidomics profiling highlights distinctive signature of hepatocellular carcinoma in HCV patients. *J Transl Med*. 2023;21(1):918.
31. Liu QB, Zhang XY, Qi JJ, Tian XC, Dovjak E, Zhang JQ, Du HH, Zhang N, Zhao J, Zhang YM, et al. Comprehensive profiling of lipid metabolic reprogramming expands precision medicine for HCC. *Hepatology* 2024. <https://doi.org/10.1097/HEP.0000000000000962>.
32. Lewinska M, Santos-Laso A, Arretxe E, Alonso C, Zhuravleva E, Jimenez-Agüero R, Eizaguirre E, Pareja MJ, Romero-Gómez M, Arrese M, et al. The altered serum lipidome and its diagnostic potential for Non-Alcoholic Fatty Liver (NAFL)-associated hepatocellular carcinoma. *Ebiomedicine*. 2021;73:103661.
33. Powell H, Coarfa C, Ruiz-Echartea E, Grimm SL, Najjar O, Yu B, Olivares L, Scheurer ME, Ballantyne C, Alsarraj A, et al. Differences in prediagnostic serum metabolomic and lipidomic profiles between cirrhosis patients with and without incident hepatocellular carcinoma. *J Hepatocell Carcinoma*. 2024;11:1699–712.
34. Li BH, Li YZ, Zhou HJ, Xu YC, Cao YJ, Cheng CX, Peng J, Li H, Zhang LZ, Su K, et al. Multiomics identifies metabolic subtypes based on fatty acid degradation allocating personalized treatment in hepatocellular carcinoma. *Hepatology*. 2024;79(2):289–306.
35. Azbazzar Y, Demirci Y, Heger G, Ipekçil D, Karabicici M, Ozhan G. Comparative membrane lipidomics of hepatocellular carcinoma cells reveals diacylglycerol and ceramide as key regulators of Wnt/ β -catenin signaling and tumor growth. *Mol Oncol*. 2023;17(11):2314–36.
36. Paul S, Lancaster GI, Meikle PJ. Plasmalogens: A potential therapeutic target for neurodegenerative and cardiometabolic disease. *Prog Lipid Res*. 2019;74:186–95.
37. Dean JM, Lodhi IJ. Structural and functional roles of ether lipids. *Protein Cell*. 2018;9(2):196–206.
38. Donovan EL, Pettine SM, Hickey MS, Hamilton KL, Miller BF. Lipidomic analysis of human plasma reveals ether-linked lipids that are elevated in morbidly obese humans compared to lean. *Diabetol Metab Syndr*. 2013;5(1):24.
39. Jiménez-Rojo N, Riezman H. On the road to unraveling the molecular functions of ether lipids. *FEBS Lett*. 2019;593(17):2378–89.
40. Schulze RJ, Schott MB, Casey CA, Tuma PL, McNiven MA. The cell biology of the hepatocyte: a membrane trafficking machine. *J Cell Biol*. 2019;218(7):2096–112.
41. von Gerichten J, Saunders K, Bailey MJ, Gethings LA, Onoja A, Geifman N, Spick M. Challenges in lipidomics biomarker identification: avoiding the pitfalls and improving reproducibility. *Metabolites*. 2024;14(8):461.
42. Liebisch G, Ahrends R, Arita M, Arita M, Bowden JA, Ejsing CS, Griffiths WJ, Holcapek M, Köfeler H, Mitchell TW, et al. Lipidomics needs more standardization. *Nat Metab*. 2019;1(8):745–7.
43. Köfeler HC, Ahrends R, Baker ES, Ekroos K, Han XL, Hoffmann N, Holcapek M, Wenk MR, Liebisch G. Recommendations for good practice in MS-based lipidomics. *J Lipid Res*. 2021;62:100138.

Publisher's Note

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