








Differences in Prediagnostic Serum Metabolomic and Lipidomic Profiles Between Cirrhosis Patients with and without Incident Hepatocellular Carcinoma

Hannah Powell ^{1,*}, Cristian Coarfa^{2-4,*}, Elisa Ruiz-Echartea^{2-4,*}, Sandra L Grimm ²⁻⁴, Omar Najjar ¹, Bing Yu⁵, Luis Olivares⁶, Michael E Scheurer⁶, Christie Ballantyne¹, Abeer Alsarraj¹, Emad Mohamed Salem ¹, Aaron P Thrift ^{1,2,7}, Hashem B El Serag ^{1,2}, Salma Kaochar ^{1,2,4}

¹Department of Medicine, Baylor College of Medicine, Houston, TX, USA; ²Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX, USA; ³Center for Precision Environmental Health, Baylor College of Medicine, Houston, TX, USA; ⁴Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX, USA; ⁵Department of Epidemiology, The University of Texas Health Science Center at Houston, Houston, TX, USA; ⁶Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; ⁷Section of Epidemiology and Population Sciences, Department of Medicine, Baylor College of Medicine, Houston, TX, USA

*These authors contributed equally to this work

Correspondence: Salma Kaochar; Hashem B El Serag, Email kaochar@bcm.edu; hasheme@bcm.edu

Background: Early detection of hepatocellular carcinoma (HCC) is crucial for improving patient outcomes, but we lack robust clinical biomarkers. This study aimed to identify a metabolite and/or lipid panel for early HCC detection.

Methods: We developed a high-resolution liquid chromatography mass spectrometry (LC-MS)-based profiling platform and evaluated differences in the global metabolome and lipidome between 28 pre-diagnostic serum samples from patients with cirrhosis who subsequently developed HCC (cases) and 30 samples from patients with cirrhosis and no HCC (controls). We linked differentially expressed metabolites and lipids to their associated genes, proteins, and transcriptomic signatures in publicly available datasets. We used machine learning models to identify a minimal panel to distinguish between cases and controls.

Results: Among cases compared with controls, 124 metabolites and 246 lipids were upregulated, while 208 metabolites and 73 lipids were downregulated. The top upregulated metabolites were glycothiocholic acid, 5-methyltetrahydrofolic acid, octanoyl-coenzyme A, and glycocholic acid. Elevated lipids comprised glycerol lipids, cardiolipin, and phosphatidylethanolamine, whereas suppressed lipids included oxidized phosphatidylcholine and lysophospholipids. There was an overlap between differentially expressed metabolites and lipids and previously published transcriptomic signatures, illustrating an association with liver disease severity. A panel of 12 metabolites that distinguished between cases and controls with an area under the receiver operating curve of 0.98 for the support vector machine (interquartile range, 0.9–1).

Conclusion: Using prediagnostic serum samples, we identified a promising metabolites panel that accurately identifies patients with cirrhosis who progressed to HCC. Further validation of this panel is required.

Keywords: lipid dysregulation, biomarker, fatty acids, machine learning

Introduction

Hepatocellular carcinoma (HCC) is the most common form of primary liver cancer. The major etiological risk factors for HCC, including hepatitis B and hepatitis C virus infections, alcoholic fatty liver disease, and non-alcoholic fatty liver disease, lead to development and progression of cirrhosis, which is the precursor lesion for 80–90% of HCC cases.¹ Survival of patients with HCC is poor except for the small proportion of patients whose disease was detected early and who receive curative therapies, such as liver transplant or resection.¹ Therefore, periodic surveillance of patients with cirrhosis using abdominal ultrasonography and serum alpha fetoprotein is recommended to detect early HCC.² However, this surveillance strategy is not particularly effective due to modest

test performance and poor utilization of ultrasound.² We need to identify reliable blood biomarkers with a high degree of sensitivity and specificity to effectively surveil and detect HCC in clinical practice.²

Over 8000 metabolites and 45,000 lipid structures have been identified in biological life.³ Serum-based metabolomic and lipidomic assays could be a less-invasive approach to detect HCC biomarkers in patient samples. Chronic hepatic inflammation and fibrosis create a pro-tumorigenic microenvironment that induces metabolic changes. For example, amplification or overexpression of the c-MYC gene, which was reported in advanced stages of liver fibrosis⁴ and the earliest events in HCC formation,⁵ can upregulate glycolytic enzymes, including hexokinase 2, phosphofructokinase 1, and pyruvate kinases type M2.⁶ Similarly, the expression of altered TP53, also observed in HCC, can lead to sustained glycolysis and decreased oxidative metabolism.⁷ In addition, metabolic dysfunction (eg, insulin resistance, high triglycerides) as part of fatty liver disease with or without other etiologies can promote the progression of chronic liver disease.⁸

Prior metabolomic clinical studies associated elevated bile acids, fatty acids, specific metabolites (eg, taurocholic acid, lysophosphoethanolamine, lysophosphatidylcholine) and reduced branched-chain amino acids with HCC development and prognosis.⁹ A systematic review identified 65 published studies through December 2023 that measured serum-based metabolites in patients with cirrhosis and HCC.⁹ More than half of these studies were conducted in China among patients with hepatitis B; 39 studies examined the differences between patients with HCC and cirrhosis, and 27 of the 39 used untargeted metabolomic and/or lipidomic assays. Of these, only six^{10–15} were US-based studies that compared HCC to cirrhosis; none examined prediagnostic serum samples collected from patients before their HCC diagnosis. Most studies reported changes in a handful of targeted metabolites, but the metabolic profiles were inconsistent and highly heterogeneous between studies.⁹ Only one of the six US-based studies measured fatty acid changes,¹³ the rest evaluated metabolites alone.^{11,12,15} Additional comprehensive, discovery-based studies are required to identify metabolite and lipid HCC biomarkers.

Here, we performed metabolic and lipidomic profiling using an unbiased high-resolution liquid chromatography mass spectrometry (LC-MS) platform and evaluated differences in matched pre-diagnostic serum samples between patients with cirrhosis who developed HCC and those who did not. We linked differentially expressed metabolites and lipids to their associated genes and proteins using the Human Metabolome Database (HMDB). We identified overlap between differentially expressed metabolites and lipids in our study with previously published transcriptomic signatures and showed that the overlapping signatures are associated with liver disease severity. Lastly, we used machine learning models to identify a minimal panel of serum metabolites and lipids that can distinguish between cirrhosis patients with and without HCC.

Methods

Study Subjects

Patients with cirrhosis were previously recruited from the hepatology clinic at the Michael E. DeBakey VA Medical Center in Houston, Texas between 08/2014 and 12/2018 as part of studies to examine risk factors and biomarkers of HCC development. Blood samples were collected in accordance with Institute Review Board-approved protocols, following written informed consent from participants. Serum was collected and frozen at -80°C until use. At the time of enrollment and blood collection, none of the participants had HCC based on the absence of liver masses on cross-sectional imaging. Follow up was performed as part of clinical HCC surveillance. We also searched the VA electronic health records and tumor registry for viral status and HCC development. For this study, we identified 30 patients with unfrozen serum samples who subsequently developed HCC (cases) and then identified sex-, race-, and etiology-matched cirrhosis patients who did not develop HCC or die as of 03/2021 (controls). We subsequently excluded two cases from the analysis due to ambiguous HCC diagnosis.

Unbiased Serum Metabolomics Analysis

Total metabolites were extracted from serum samples using a liquid-liquid extraction method and analyzed with high-throughput liquid chromatography with tandem mass spectrometry (LC-MS/MS). Lipids were extracted using water/methanol/dichloromethane (2:2:2 v/v) at room temperature after spiking internal standards (15:0–18:1(d7)

diacylglycerol, 15:0–18:1(d7) phosphatidylcholine, 15:0–18:1(d7) phosphatidylinositol (NH₄ Salt), 15:0–18:1(d7) phosphatidylserine (Na Salt), and 18:1(d7) lysophosphatidylcholine) and separated by reverse-phase chromatography (Acquity HSS UPLC T3 column (1.8 μm particle 50×2.1 mm, Waters, Milford, MA)) on a Shimadzu CTO-20A Nexera X2 UHPLC system.^{16–18} The spectrometry data were acquired in both positive and negative ionization modes via the data-dependent acquisition method using a TripleTOF 5600 mass spectrometer equipped with a Turbo VTM ion source (AB Sciex, Concord, Canada) coupled to Shimadzu CTO-20A Nexera X2 UHPLC system (Shimadzu, Maryland, USA) and analyzed using MultiQuant software (ver. 1.1.0.26, AB Sciex, Concord, Canada).¹⁸ We detected 1266 metabolites and 1311 lipid molecules.

Metabolomic and Lipidomic Data Analysis

Raw data in *wiff* format^{16–18} were converted to *mgf* format using proteoWizard.¹⁹ The NIST MS Pep software was used to compare *mgf* files against the NIST14 reference library for metabolites²⁰ and LipidBlast reference library for lipids.²¹ The *m/z* width was determined using the internal standard mass accuracy; 0.001 was used for positive mode and 0.005 for negative mode, targeting an overall mass error below 2 parts per million. MS/MS annotation was used to create a library for quantification, and raw data files were searched against this library based on their mass and retention time using the MultiQuant 1.1.0.26 software (AB Sciex, USA). The relative abundance of peak spectra was used for quantitative analyses. Identified metabolites and lipids were quantified by normalizing against the internal standards used in their respective methods.

Metabolomics data collected using the positive method were normalized to a tryptophan (¹⁵N₂) internal standard, and those using the negative method were normalized to a gibberellic acid internal standard. Lipidomics data collected using the positive method were normalized using a 18:1(d7) lysophosphatidylcholine internal standard, and those using negative data were normalized using the internal standard 18:1(d7) lysophosphatidylethanolamine. The data were log₂ transformed and batch corrected using the ComBat method.²² Linear regression in the *LieberInstitute/jaffelab* package²³ from the R statistical system was used to adjust for variation in metabolite or lipid expression.

Differentially expressed metabolites or lipids between cases and controls were determined using t-tests followed by multiple testing hypothesis correction using the Benjamini-Hochberg method with significance achieved at false discovery rate (FDR) of *p*<0.25; we followed previously established methodologies.^{24,25} Genes associated with differentially expressed metabolites and lipids were determined by mining the HMDB,²⁶ as described previously.^{16,18} Enriched pathways based on differentially expressed metabolites were determined using over-representation analysis implemented in MSigDB²⁷ compared with the Hallmark and the Gene Ontology databases. A hypergeometric distribution test was used with significance defined as FDR-adjusted *p*-value<0.05. Lipids were further aggregated at the lipid species level using the median value across all the structures or detection methods.

Integration of Metabolomics and Lipidomics with Publicly Available Transcriptomics Data

Two sets of transcriptomic data from liver tissue of patients with HCC and cirrhosis were downloaded from NCBI Gene Expression Omnibus (GEO): GSE25907²⁸ and GSE6764.²⁹ For GSE25907, we determined differentially expressed genes (DEGs) of 300 HCC cases compared with 40 cases with cirrhosis. For GSE6764, we determined DEGs between early HCC, advanced HCC, and all HCC compared with cirrhosis without HCC. DEGs were derived for each comparison using parametric t-tests with significance set at a >1.5- fold change and an FDR <0.05. The DEGs in HCC cases compared with cirrhosis only cases were cross-referenced to genes associated with differentially expressed serum metabolites or lipids in our study to create an overlapping signature.

Association of Metabolomics/Lipidomics Signature with HCC Datasets

As explained above, we developed three separate classes of signatures for HCC compared with cirrhosis: one derived from metabolomics and lipidomics, one from GEO liver transcriptomics datasets, and one from the overlap of the first two signatures. We integrated these signatures with gene expression datasets of HCC samples in the Cancer Genome Atlas Liver Hepatocellular Carcinoma (TCGA-LIHC):³⁰ GSE54238,³¹ GSE6764,²⁹ and GSE14520;³² non-alcoholic

steatohepatitis samples: GSE48452³³ and GSE33814,³⁴ and non-alcoholic fatty liver disease samples: GSE49541.³⁵ For each cohort, a gene signature activity score was derived by scaling every gene to a z-score and computing the summed z-score for all genes.^{36,37} Association of signature activity scores with HCC was assigned by encoding the disease states with numerical values in order of severity and computing a Pearson correlation coefficient. Multiple hypothesis testing correction using the Benjamini-Hochberg method was applied for each dataset and HCC severity clinical variable, with significance achieved at an FDR of <0.05.

Machine Learning Classification and Regression

Machine learning classification of cases versus controls was performed using metabolite, lipid species, metabolites and lipid species combined, and metabolites and lipid data for all structures and detection methods combined. Feature selection was performed using a t-test with an FDR of <0.05. Data were then subjected to three machine learning methods: k-nearest neighbor (KNN), Support Vector Machines (SVM),³⁸ and Random Forest (RF).³⁹ In cross-validation, data were randomly split between a training set containing 80% of the samples and a testing set containing 20% of samples for 100 iterations. Classifier performance was measured using the area under the receiver operating curve (AUROC). To identify a minimal set of features needed to discriminate cases and controls, machine learning classification models were created with increasing numbers of features, ranging from two to the number of all informative features in at least 50% of the cross-validation iterations.

Results

We analyzed 58 patients with cirrhosis, 28 who developed HCC during follow up (cases) and 30 who did not develop HCC (controls) (Figure 1). All patients were men, with White, Latino, and Black race represented almost equally between cases and controls (White: 31.2% vs 36.7%; Black 39.3% vs 30%; Latino 28.6% vs 33.3%). The median age was 56.8 years for cases and 61.7 years for controls. Hepatitis C virus (HCV) was the underlying etiology in all patients (Table 1). The mean duration between sample collection and HCC diagnosis date was 659.4 days (standard deviation [SD] 730.8), and for controls, the mean duration between sample collection and date of last follow up was 263 days (SD 193).

Differential Metabolomics and Lipidomics Between Cases and Controls

Associations between metabolomic and lipidomic markers were examined using logistic regression models adjusted for age and race/ethnicity. There were 124 upregulated and 208 downregulated metabolites in cases compared with controls (Figure 2A, Supplemental Table 1A). There were also 246 lipids upregulated and 73 downregulated in cases compared with controls (Figure 2B, Supplemental Table 1B).

The upregulated metabolites with the highest fold difference in cases included glycooursodeoxycholic acid (GUDCA, bile acid glycine conjugate; \log_2 fold change [\log_2 FC] 3.13), 5-methyltetrahydrofolic acid (methylated derivative of tetrahydrofolate; \log_2 FC 2.37), octanoyl-coenzyme A (medium-chain fatty acyl-CoA; \log_2 FC 2.24), and glycocholic acid (GCC; \log_2 FC 2.15). The downregulated metabolites with the highest fold difference in cases included 5-Nitrosopyrimidine-2,4,6-triamine/nitrosotriaminopyrimidine (\log_2 FC -2.44), and theobromine (\log_2 FC -1.98). Elevated lipids in cases included phosphocholines (PC 31:2 and PC 41:4; \log_2 FC 3.94 and 2.86, respectively) and triglycerides (TG 55:1; \log_2 FC 2.08), and reduced lipids were triglyceride 51:3 (\log_2 FC -4.29), sphingomyelin 42:1 (\log_2 FC -3.58), and phosphocholine 38:6 (\log_2 FC -2.70). Using HMDB taxonomy, the metabolite classes with the highest frequency of differentially expressed metabolites (DEMs) were fatty acyls (18 up/25 down), carboxylic acids and derivatives (12 up/20 down), and organooxygen compounds (4 up/15 down) (Figure 2C, Supplemental Table 1C). The lipid classes with the highest frequency of differentially expressed lipids (DELs) were oxidized phosphatidylcholine (69 up/53 down), triacylglycerols/glycerolipids (116 up/1 down), and cardiolipin (19 up/3 down) (Figure 2D, Supplemental Table 1C).

DEMs and DELs were linked with related genes in the HMDB, which contains 220,945 genes for metabolites and lipids associated with 8610 protein sequences. We linked 332 DEMs to 1916 genes, 319 DELs to 321 genes, and 245 genes to both DEMs and DELs (Figure 2A and B) for a total of 1992 distinct genes associated with our DEMs and DELs. Pathway analysis of the 245 genes associated with both DEMs and DELs using the Hallmark geneset compendium

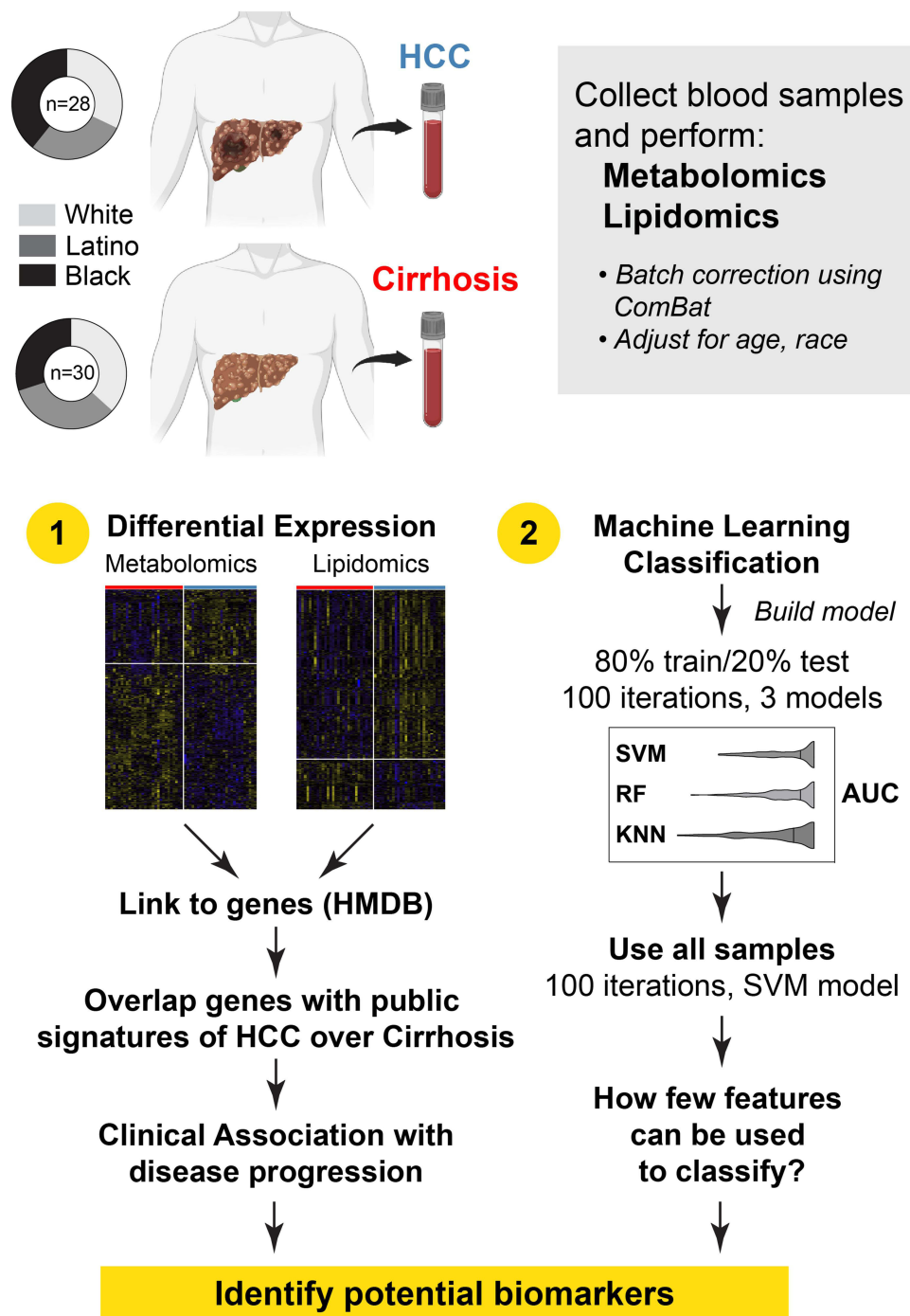


Figure 1 Study overview. Blood collected from patients with cirrhosis who either developed hepatocellular carcinoma (HCC) during follow up (cases) or did not (controls) was used to perform unbiased metabolomics and lipidomics. After correcting for analysis batches with ComBat, the data was adjusted for age, race, and etiology. Two approaches were used: differential expression and machine learning classification.

showed enrichment for *Complement*, *Adipogenesis*, *PI3K AKT MTOR Signaling*, and *Fatty Acid Metabolism* (Figure 2E, Supplemental Table 1D).

We examined whether our serum-derived metabolic/lipidomic signatures, assembled by linking DEMs and DELs with 1992 genes, was reflective of gene expression changes previously reported for cirrhosis patients with and without HCC in publicly available liver tissue transcriptomic datasets (Supplemental Table 1E). We derived four liver-based gene signatures: HCC over cirrhosis from dataset GSE25097 and all HCC over cirrhosis, advanced HCC over cirrhosis, and

Table 1 Demographics of Patients

	HCC	Cirrhosis
Number of samples	28	30
Avg age, years (SD)	61.7 (5.3)	56.8 (8.4)
Sex, male	100%	100%
HCV-positive	92.9%	100%
Race: White	32.1%	36.7%
Race: Latino	28.6%	33.3%
Race: Black	39.3%	30.0%

early HCC over cirrhosis from dataset GSE6764 (Figure 3A and B). These were designated as “full” signatures. The full HCC gene signatures were compared with the 1992 serum-based DEM and DEL genes, and the overlap ranged between 9.2% and 14.6% (Figure 3B, Supplemental Table 1E). The four sets of overlapping genes were designated as “overlap” signatures.

We evaluated the association of the four full HCC gene signatures as well as the four overlap HCC gene signatures (Figure 3A and B) with publicly available transcriptome datasets of HCC or disease states that increase risk for HCC (Figure 3C). The overlap HCC gene signatures performed as well as or better than the full HCC gene signatures in discriminating conditions of different liver disease severity. For example, the GSE6764 full signature (1294 genes) had a Pearson correlation coefficient of 0.73 (FDR <0.05) in the TCGA-LIHC dataset, whereas the GSE6764 overlap signature (177 genes that overlapped with our metabolomic/lipidomic linked genes) showed a Pearson correlation coefficient of 0.75 in TCGA-LIHC dataset (Figure 3C). Some overlapping genes identified via integration analysis (eg, *RACGAP1*, *BUB1B*, *UBE2T*, *ALDH3A1*) showed higher expression in HCC and were associated with overall poor survival in the TCGA-LIHC database.

Predictive Models for HCC

We compared the three methods of machine learning classification using metabolomics and lipidomics data. They all had similar robust performance with an AUROC of 0.95 (95% confidence interval [CI] 0.76, 1; interquartile range [IQR] 0.87–1) for KNN, 0.97 (95% CI 0.81, 1, IQR 0.9–1) for RF, and 0.97 (95% CI 0.81, 1; IQR 0.9–1) for SVM (Figure 4). Using lipidomics data aggregated at the species level, the median AUROC were 0.82 for KNN (95% CI 0.52, 1; IQR 0.7–0.87), 0.83 for RF (95% CI 0.56, 1; IQR 0.73–0.9), and 0.87 for SVM (95% CI 0.57, 1; IQR 0.77–0.93) (Figure 4). Using the full lipidomics profile, the AUROC were 0.87 for KNN (95% CI 0.66, 1; IQR 0.8–0.95), 0.87 for RF (95% CI 0.58, 1; IQR 0.77–0.93), and 0.9 for SVM (95% CI 0.68, 1; IQR 0.83–0.97) (Supplemental Figure 1A). A predictive model using both metabolites and lipid species had AUROCs of 0.97 for KNN (95% CI 0.8, 1; IQR 0.9–1), 0.97 for RF (95% CI 0.82, 1; IQR 0.9–1), and 0.98 for SVM (95% CI 0.79, 1; IQR 0.9–1) (Figure 4). Finally, for models using metabolites and the full lipid profile, the median AUROCs were 0.97 for KNN (95% CI 0.8, 1; IQR 0.9–1), 0.97 for RF (95% CI 0.76, 1; IQR 0.87–1), and 0.97 for SVM (95% CI 0.77, 1; IQR 0.9–1) (Supplemental Figure 1B).

We determined the minimal complement of metabolite/lipid features that had similar performance to classifiers derived using all features (Figure 4). Data were input as 1266 metabolites only, 475 lipid species only, or 1741 metabolites and lipids species combined. We determined the informative features in at least 50% of the cross-validation iterations and identified 242 metabolites, 93 lipid species, and 327 metabolites and lipid species (Figure 5A, Supplemental Table 1F). Performance of models with increasing number of features were evaluated (Figure 5A, Supplemental Figure 1C and D). These molecules were ranked by their median importance score across all cross-validation iterations (top 20 shown in Figure 5B). The minimum number of the features that differentiated between cases and controls were (1) the top 12 metabolites, (2) the top 20 lipid species, (3) the top 50 from the full lipid profile, (4) the top 12 combined metabolites and lipid species, and (5) the top 12 combined metabolites and full lipid profile. The 12 predictive metabolites alone were all downregulated, whereas the top 20 predictive lipid species or full lipid profile were upregulated in cases compared with controls (Figure 5B, Supplemental Figure 1E). The top 12 features for combined

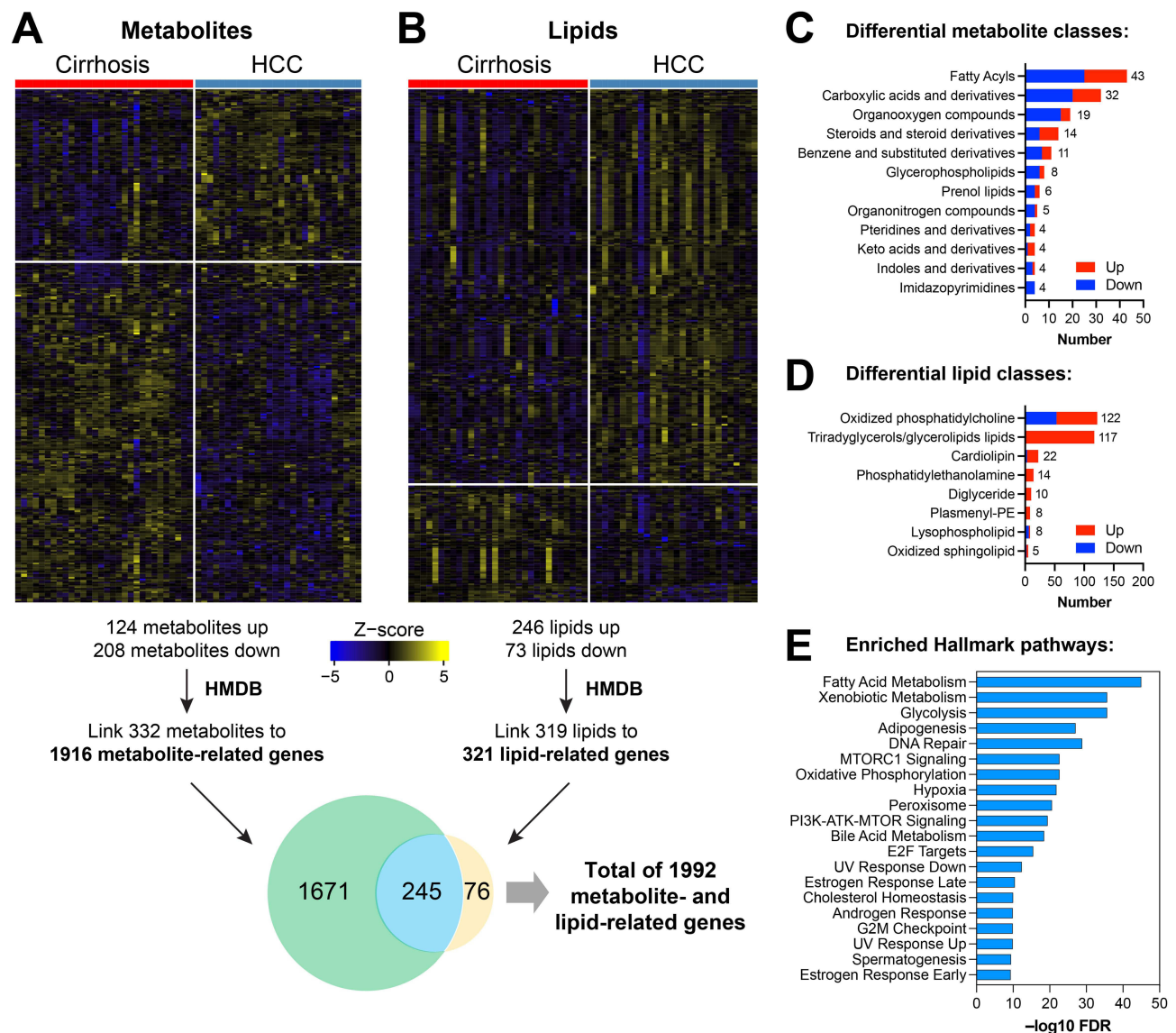


Figure 2 Differential metabolomic and lipidomic analysis. Heatmaps showing differential expression of metabolites (A) or lipids (B) using a cutoff of $FDR < 0.25$, comparing serum samples collected before diagnosis for cases (patients with cirrhosis who developed HCC) and controls (patients with cirrhosis and no HCC). Top classes of differentially expressed (C) metabolites and (D) lipids were determined and stratified by direction of regulation in HCC. Differentially expressed metabolites or lipids were linked to related genes using the Human Metabolome Database (HMDB). Genes associated with metabolites or lipids were determined, as shown in the Venn diagram. (E) Pathway analysis of the genes associated with both DEMs and DELs using the Hallmark geneset compendium.

metabolite and lipid species (Figure 5B) and the top 12 features for combined metabolite and full lipid profile (Supplemental Figure 1F) were all metabolites that were downregulated in cases compared with controls. The top 12 predictive metabolites were three branched amino acid (L-isoleucine, L-lysine, and L-valine), 2-fluorobenzyl alcohol, N-formyl-L-methionine, Tris(hydroxymethyl)aminomethane (THAM), 2-Aminopentanoic acid/DL-Norvaline/DL-valeric acid, 3-amino-2-methyl-propanoic acid/ 3-Aminoisobutyric Acid, N(epsilon)-Acetyl-L-lysine, L-pipecolic acid, alpha-D-(+)-Talose, and 3-methoxytyramine(3-MT).

Discussion

We used an unbiased high-resolution LC-MS-based platform to assay and compare pre-diagnostic serum metabolomic and lipidomic profiles between cases with cirrhosis who developed HCC and controls with cirrhosis and no HCC. Our analysis highlighted major differences in serum levels of fatty acyls, carboxylic acid, and derivatives and stark

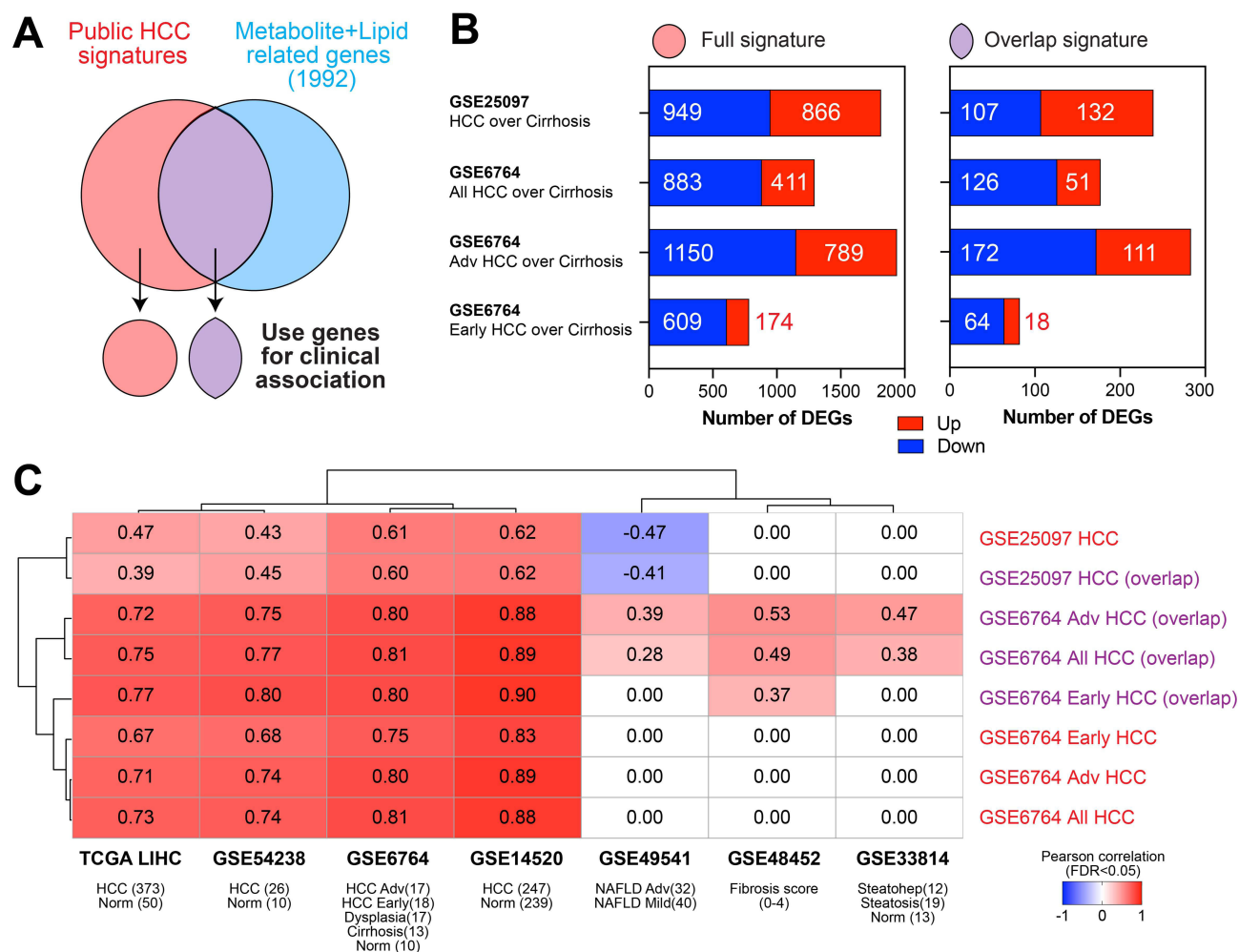


Figure 3 Clinical association of metabolite/lipid signatures with disease progression. **(A)** We compared genes differentially expressed in HCC over cirrhosis from publicly available databases (full signatures) with the serum metabolite/lipid-related genes identified in our analysis to generate an overlap signature. **(B)** The number of differentially expressed genes (DEGs) of the four comparison groups from two public datasets (fold change >2, FDR < 0.05) with the full signature on the left and the overlap signature on the right. **(C)** Full or overlap signatures were correlated with transcriptomics from cohorts of liver disease progression using a Pearson correlation coefficient (FDR < 0.05 per dataset).

upregulation of glycerolipids in cases who developed HCC. Using integrative bioinformatics analysis, we linked DEMs and DELs to their respective genes, which are involved in multiple metabolic pathways. Many of these genes were differentially regulated in published transcriptomic datasets from clinical cohorts of patients with cirrhosis or HCC. We leveraged machine learning models to identify a minimal panel of metabolites and lipids that accurately distinguish between cases and controls. The best-performing classifier of 1741 metabolites and lipids was derived using SVM, achieving a median AUROC of 0.98 over 100 cross-validation iterations. We also showed that a panel of only 12 metabolites had similar discrimination power as the entire panel. Most of the 12 metabolites were amino acids (L-isoleucine, L-lysine, L-valine, L-methionine, DL-Norvaline) and related biproducts (L-pipecolic acid, and aldehyde sugar (alpha-D-Talose).

The 12 metabolites identified in our Phase 2 biomarker study outperformed the most common biomarker for detecting HCC, alpha-fetoprotein (AFP); this is a promising start. A meta-analysis showed that AFP assays for HCC diagnosis had a pooled sensitivity of only 51.9% and a specificity of 94%, yielding an AUROC value of 0.81.⁴⁰ More recently, we developed and validated an HCC risk stratification model in patients with cirrhosis that combined AFP with demographics (age and sex), clinical (liver disease etiology (HCV) vs other etiologies), alanine aminotransferase, platelet count, albumin) and lifestyle features (alcohol use, smoking, body mass index (BMI)). Even this enhanced model with

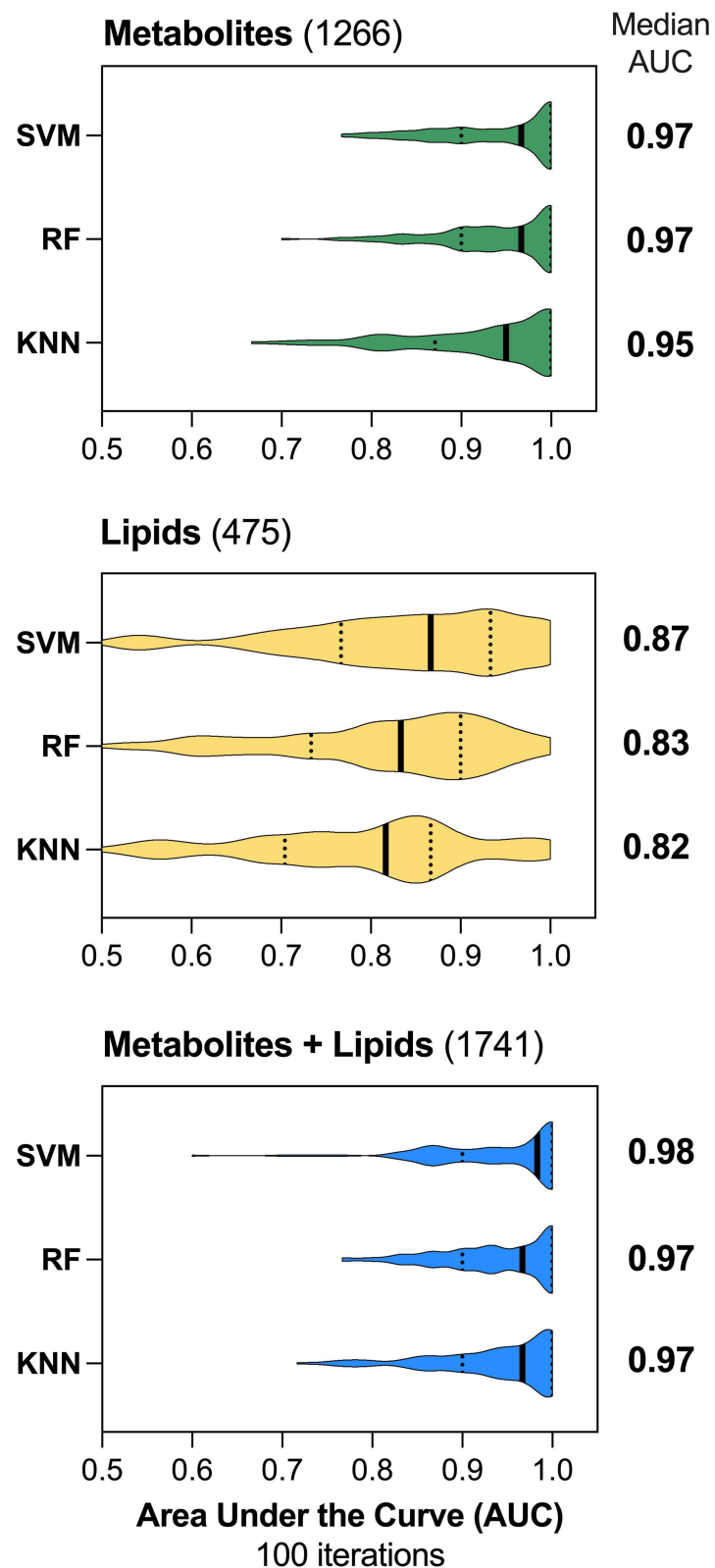


Figure 4 Machine Learning Classification. Three methods of machine learning classification, Support Vector machines (SVM linear), Random Forest (RF), or K-nearest neighbor (KNN) were used to predict HCC vs cirrhosis over 100 cross-validation iterations using 80% of the data as training and 20% as testing. Metabolites alone, lipids species alone, or metabolites and lipid species combined (blue) were used for classification. The median area under the curve (AUC) of each model is listed on the right.

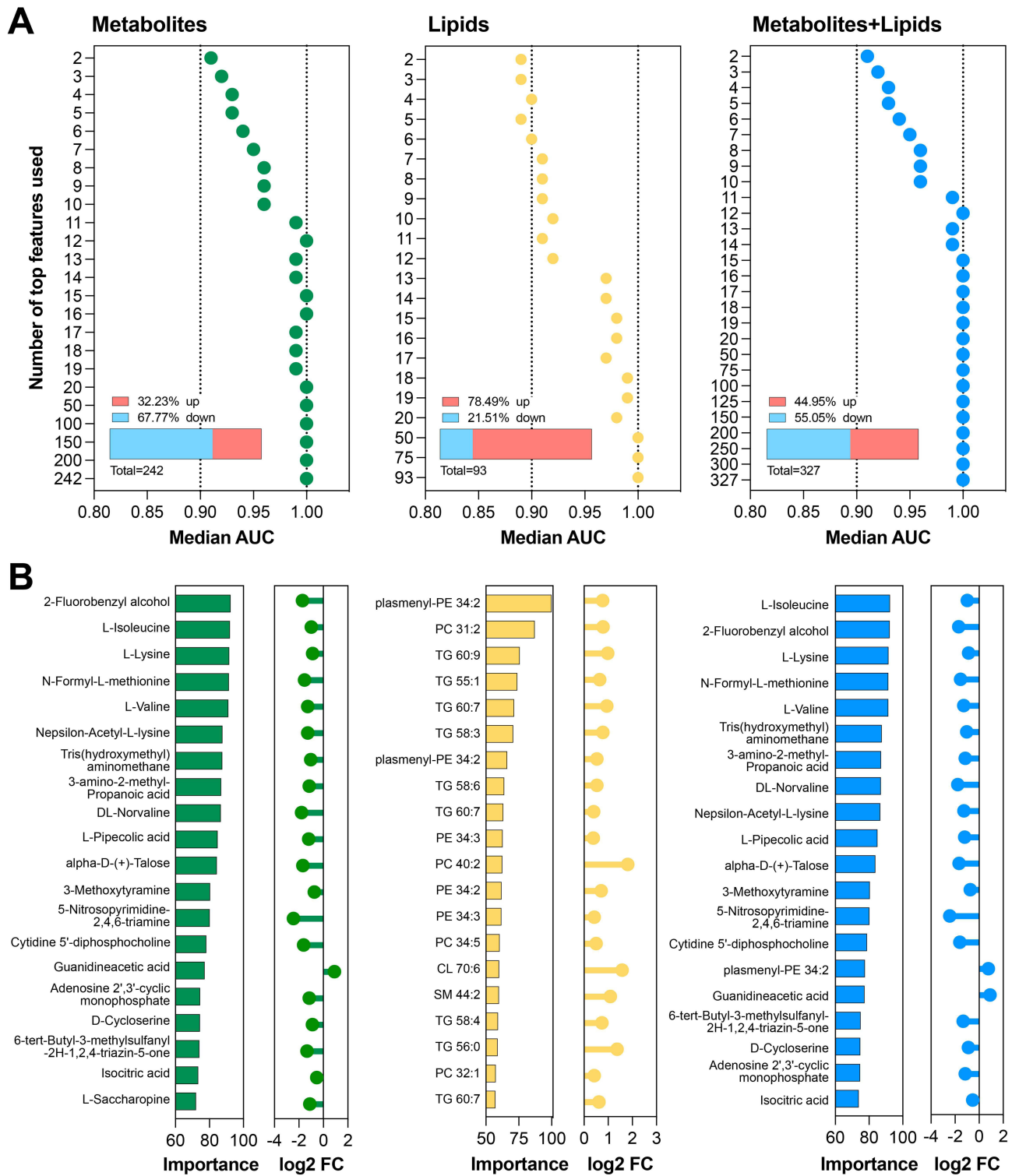


Figure 5 Machine Learning Classification: parsimonious models. **(A)** To determine minimal classifiers, we selected features (metabolites or lipids) informative in at least 50% of the 100 iterations. Using all samples with an increasing number of features, SVM machine learning classification was performed to predict HCC vs cirrhosis. Each dot indicates the median AUC for metabolites alone (green), lipids alone (yellow), or both combined (blue). Bar charts show the direction of change of all informative features in HCC. **(B)** The top 20 features for each data set are graphed, sorted by importance values, together with the \log_2 fold change in HCC versus cirrhosis.

added features had moderate discriminatory ability with AUROC of 0.75 for predicting HCC at 1-year follow-up, lower than the 12 lipids identified in this study.⁴¹

Biomarker studies that use MS-based metabolomics and lipidomics are often hampered by heterogeneity and inconsistency. A meta-analysis of 68 clinical studies revealed that approximately 600 metabolites have been identified as potential diagnostic or prognostic HCC biomarkers.⁹ However, only 20 of these metabolites have been reported in three or more studies,⁹ and only 9 were conducted in the US.^{10–15,42–44} Among these nine studies, six employed untargeted assays^{10–15} and only three adjusted for multiple comparisons.^{10,11,15} The variability is related to diverse population characteristics, quality control procedures, sample handling methods, data acquisition techniques, statistical approaches, and metabolite identification processes. Most studies examined samples obtained at or after HCC diagnosis, and many did not address or adjust for HCC clinical treatment. In our study, we aimed to reduce variations between cases and controls by conducting comparisons on samples matched based on sex, race/ethnicity, and etiology. Furthermore, we utilized samples collected prior to HCC diagnosis, targeting a clinically relevant window between cirrhosis and HCC, and avoiding issues related to reverse causality.

We found that fatty acid metabolism exhibited the most pronounced alterations when comparing patients with and without HCC (Figure 2E, Supplemental Table 1D) and substantial enrichment of kinase signaling pathways related to lipid metabolism, such as the “PI3K-AKT-MTOR signaling” pathway. This observation aligns with earlier research that identified certain lipid classes, particularly phospholipids, as predictive markers of HCC risk, particularly in individuals with metabolic-associated fatty liver disease.^{45,46} Other studies indicated that upregulation of fatty acid transporters in NAFLD-HCC tumors depletes serum fatty acids.^{45,46} In a Chinese case-control study, six serum phospholipids (out of 44 metabolites measured) were significantly different between normal and HCC individuals.⁴⁷ Differential levels of citrulline have further been associated with inflammation in plasma Th17.²⁴ Similarly, a European case-control study of cirrhotic patients with and without HCC found that two phosphocholines (PC (30:2) and PC (30:1)) were associated with increased risk of tumor burden following curative treatment.⁴⁸ We observed a similar enrichment of polyunsaturated fatty acids in the serum of cirrhosis patients with HCC compared with those without HCC. These findings are in line with those of a previous study that reported upregulation of sphingosine-1-phosphate and lysophosphatidylcholine (lysoPC 17:0) in sera of HCC patients⁴⁹ and another study that found higher levels of lysophosphatidylcholines and ceramides in exosomes from HCC patients compared with those from cirrhosis patients without HCC.⁴⁹ Lastly, stearoyl CoA desaturase 1 and fatty acid synthase, play pivotal roles in the synthesis of unsaturated fatty acids from saturated fatty acids. Pharmacological inhibition of these enzymes inhibits tumor growth and induces apoptosis in preclinical models of HCC.^{50,51} A Phase 2 clinical trial investigated the FASN inhibitor TVB 2640 in patients with non-alcoholic steatohepatitis and reported significant reductions in liver fat and improvements in biochemical, inflammatory, and fibrotic biomarkers.⁵² These studies reaffirm the potential of targeting this pathway and the importance of gaining a deeper understanding of the changes in lipid metabolism during HCC development and progression.

Amino acids (isoleucine, lysine, and valine) were represented in our predictive metabolite/lipid panel. Leucine and valine were reported among 13 serum metabolites that distinguished HCC patients (n=20) from healthy subjects (n=20) in a case-control study from China.⁵³ In a Chinese nested case-control study of patients with HCC vs healthy controls, 18 serum metabolites (including 5 amino acid-related metabolites: hydroxyphenyllactic acid, cystathionine, citrulline, arginine, and sarcosine) were associated with the risk of developing HCC.⁴⁷ Similarly, a US study used untargeted analysis of metabolites in plasma from 63 patients with HCC and 65 patients with cirrhosis and no HCC¹¹ identified 11 metabolites including the valine and isoleucine that were part of our 12-metabolite panel.

Our study has few limitations. Although ethnically diverse, the sample size in our study was relatively small, which could potentially result in type 2 errors. Furthermore, all the participants in our study were male with hepatitis C infection, and the generalizability of our findings to women and individuals with other etiologies could be limited. Lastly, while we performed internal validation, an external validation is required.

In conclusion, using prediagnostic serum samples, we identified a promising metabolites panel that accurately identifies patients with cirrhosis who progressed to HCC. Further validation of this panel is required. It will be essential to validate the diagnostic and prognostic value of our 12-metabolite panel within a larger cohort and assess their performance when combined with other clinical factors.

Data Sharing Statement

All data is available upon request.

Statements

The abstract of this paper was presented at the 2024 American Association of Cancer Research (AACR) Annual Conference as a poster with interim findings. The poster's abstract was published in "Poster Abstracts" in AACR's *Cancer Research* Volume 84, Issue 6 supplement in March 2024.

Statement of Ethics

Study approval statement: This study protocol was reviewed and approved by Baylor College of Medicine IRB proposal number H-22934. Our study complies with the Declaration of Helsinki.

Consent to participate statement: Informed consent was obtained from all research subjects.

Funding

This work was supported by the Cancer Prevention & Research Institute of Texas (CPRIT) grants RP220119, RP210227, and RP200504, NCI (NCI P01 CA263025), NIH P30 shared resource grant CA125123, and NIEHS grant P30 ES030285, and in part by Center for Gastrointestinal Development, Infection, and Injury (NIDDK P30 DK 56338). Data analysis was performed on the HPC cluster supported by the award S10 OD032185. The funders had no role in the design, data collection, data analysis, and reporting of this study. Research reported in this publication was also supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number T32GM136554. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Disclosure

The authors have no conflict of interest to declare.

References

1. McGlynn KA, Petrick JL, El-Serag HB. Epidemiology of hepatocellular carcinoma. *Hepatology*. 2021;73(Suppl 1):4–13. doi:10.1002/hep.31288
2. McMahon B, Cohen C, Brown Jr RS, et al. Opportunities to address gaps in early detection and improve outcomes of liver cancer. *JNCI Cancer Spectr*. 2023;7(3). doi:10.1093/jncics/pkad034.
3. Conroy MJ, Andrews RM, Andrews S, et al. LIPID MAPS: update to databases and tools for the lipidomics community. *Nucleic Acids Res*. 2024;52(D1):D1677–D1682. doi:10.1093/nar/gkad896
4. Nevzorova YA, Hu W, Cubero FJ, et al. Overexpression of c-myc in hepatocytes promotes activation of hepatic stellate cells and facilitates the onset of liver fibrosis. *Biochim Biophys Acta*. 2013;1832(10):1765–1775. doi:10.1016/j.bbadis.2013.06.001
5. Lin CP, Liu C-R, Lee C-N, et al. Targeting c-Myc as a novel approach for hepatocellular carcinoma. *World J Hepatol*. 2010;2(1):16–20. doi:10.4254/wjh.v2.i1.16
6. Feng J, Li J, Wu L, et al. Emerging roles and the regulation of aerobic glycolysis in hepatocellular carcinoma. *J Exp Clin Cancer Res*. 2020;39(1):126. doi:10.1186/s13046-020-01629-4
7. Hussain SP, Schwank J, Staib F, et al. TP53 mutations and hepatocellular carcinoma: insights into the etiology and pathogenesis of liver cancer. *Oncogene*. 2007;26(15):2166–2176. doi:10.1038/sj.onc.1210279
8. Jinjuvadia R, Patel S, Liangpunsakul S. The association between metabolic syndrome and hepatocellular carcinoma: systemic review and meta-analysis. *J Clin Gastroenterol*. 2014;48(2):172–177. doi:10.1097/MCG.0b013e3182a030c4
9. Anh NH, Long NP, Min YJ, et al. Molecular and metabolic phenotyping of hepatocellular carcinoma for biomarker discovery: a meta-analysis. *Metabolites*. 2023;13(11):1112. doi:10.3390/metabo13111112
10. Bowers J, Hughes E, Skill N, et al. Detection of hepatocellular carcinoma in hepatitis C patients: biomarker discovery by LC-MS. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2014;966:154–162. doi:10.1016/j.jchromb.2014.02.043
11. Di Poto C, Ferrarini A, Zhao Y, et al. Metabolomic characterization of hepatocellular carcinoma in patients with liver cirrhosis for biomarker discovery. *Cancer Epidemiol Biomarkers Prev*. 2017;26(5):675–683. doi:10.1158/1055-9965.EPI-16-0366
12. Fitian AI, Nelson DR, Liu C, et al. Integrated metabolomic profiling of hepatocellular carcinoma in hepatitis C cirrhosis through GC / MS and UPLC / MS - MS. *Liver Int*. 2014;34(9):1428–1444. doi:10.1111/liv.12541
13. Khan IM, Gjukan D, Jiao J, et al. A novel biomarker panel for the early detection and risk assessment of hepatocellular carcinoma in patients with cirrhosis. *Cancer Prev Res (Phila)*. 2021;14(6):667–674. doi:10.1158/1940-6207.CAPR-20-0600
14. Patterson AD, Maurhofer O, Beyoğlu D, et al. Aberrant lipid metabolism in hepatocellular carcinoma revealed by plasma metabolomics and lipid profiling. *Cancer Res*. 2011;71(21):6590–6600. doi:10.1158/0008-5472.CAN-11-0885

15. Resson HW, Xiao JF, Tuli L, et al. Utilization of metabolomics to identify serum biomarkers for hepatocellular carcinoma in patients with liver cirrhosis. *Anal Chim Acta*. 2012;743:90–100. doi:10.1016/j.aca.2012.07.013
16. Piyarathna DWB, Rajendiran TM, Putluri V, et al. Distinct lipidomic landscapes associated with clinical stages of urothelial cancer of the bladder. *Eur Urol Focus*. 2018;4(6):907–915. doi:10.1016/j.euf.2017.04.005
17. Purwaha P, Gu F, Piyarathna DWB, et al. Unbiased lipidomic profiling of triple-negative breast cancer tissues reveals the association of sphingomyelin levels with patient disease-free survival. *Metabolites*. 2018;8(3):41. doi:10.3390/metabo8030041
18. Vantaku V, Dong J, Ambati CR, et al. Multi-omics integration analysis robustly predicts high-grade patient survival and identifies CPT1B effect on fatty acid metabolism in bladder cancer. *Clin Cancer Res*. 2019;25(12):3689–3701. doi:10.1158/1078-0432.CCR-18-1515
19. Chambers MC, Maclean B, Burke R, et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol*. 2012;30(10):918–920. doi:10.1038/nbt.2377
20. Kind T, Meissen JK, Yang D, et al. Qualitative analysis of algal secretions with multiple mass spectrometric platforms. *J Chromatogr A*. 2012;1244:139–147. doi:10.1016/j.chroma.2012.04.074
21. Kind T, Liu K-H, Lee DY, et al. LipidBlast in silico tandem mass spectrometry database for lipid identification. *Nat Methods*. 2013;10(8):755–758. doi:10.1038/nmeth.2551
22. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007;8(1):118–127. doi:10.1093/biostatistics/kxj037
23. Collado-Torres L, Burke EE, Peterson A, et al. Regional heterogeneity in gene expression, regulation, and coherence in the frontal cortex and hippocampus across development and schizophrenia. *Neuron*. 2019;103(2):203–216e8. doi:10.1016/j.neuron.2019.05.013
24. Bouch RJ, Zhang J, Miller BC, et al. Distinct inflammatory Th17 subsets emerge in autoimmunity and infection. *J Exp Med*. 2023;220(10). doi:10.1084/jem.20221911.
25. Song Q, Ruiz J, Xing F, et al. Single-cell sequencing reveals the landscape of the human brain metastatic microenvironment. *Commun Biol*. 2023;6(1):760. doi:10.1038/s42003-023-05124-2
26. Wishart DS, Tzur D, Knox C, et al. HMDB: the human metabolome database. *Nucleic Acids Res*. 2007;35(Database issue):D521–6. doi:10.1093/nar/gkl923
27. Liberzon A, Birger C, Thorvaldsdóttir H, et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1(6):417–425. doi:10.1016/j.cels.2015.12.004
28. Lamb JR, Zhang C, Xie T, et al. Predictive genes in adjacent normal tissue are preferentially altered by sCNV during tumorigenesis in liver cancer and may rate limiting. *PLoS One*. 2011;6(7):e20090. doi:10.1371/journal.pone.0020090
29. Wurmbach E, Chen Y-B, Khitrov G, et al. Genome-wide molecular profiles of HCV-induced dysplasia and hepatocellular carcinoma. *Hepatology*. 2007;45(4):938–947. doi:10.1002/hep.21622
30. Ally A, Balasundaram M, Carlsen R, Cancer Genome Atlas Research Network. Electronic address, w.b.e. and N. cancer genome atlas research, comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell*. 2017;169(7):1327–1341e23. doi:10.1016/j.cell.2017.05.046
31. Yuan S, Wang J, Yang Y, et al. The prediction of clinical outcome in hepatocellular carcinoma based on a six-gene metastasis signature. *Clin Cancer Res*. 2017;23(1):289–297. doi:10.1158/1078-0432.CCR-16-0395
32. Roessler S, Jia H-L, Budhu A, et al. A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. *Cancer Res*. 2010;70(24):10202–10212. doi:10.1158/0008-5472.CAN-10-2607
33. Ahrens M, Ammerpohl O, von Schönfels W, et al. DNA methylation analysis in nonalcoholic fatty liver disease suggests distinct disease-specific and remodeling signatures after bariatric surgery. *Cell Metab*. 2013;18(2):296–302. doi:10.1016/j.cmet.2013.07.004
34. Starmann J, Fälth M, Spindelböck W, et al. Gene expression profiling unravels cancer-related hepatic molecular signatures in steatohepatitis but not in steatosis. *PLoS One*. 2012;7(10):e46584. doi:10.1371/journal.pone.0046584
35. Moylan CA, Pang H, Dellinger A, et al. Hepatic gene expression profiles differentiate presymptomatic patients with mild versus severe nonalcoholic fatty liver disease. *Hepatology*. 2014;59(2):471–482. doi:10.1002/hep.26661
36. Taylor BS, Schultz N, Hieronymus H, et al. Integrative genomic profiling of human prostate cancer. *Cancer Cell*. 2010;18(1):11–22. doi:10.1016/j.ccr.2010.05.026
37. Geng C, Kaochar S, Li M, et al. SPOP regulates prostate epithelial cell proliferation and promotes ubiquitination and turnover of c-MYC oncoprotein. *Oncogene*. 2017;36(33):4767–4777. doi:10.1038/ncr.2017.80
38. Cortes C, Vapnik V. V.V. Support-vector networks. *Machine Learning*. 1995;20(3):273–297. doi:10.1007/BF00994018
39. L. B. Random Forests. *Machine Learning*. 2001;45(1):5–32. doi:10.1023/A:1010933404324
40. Xu C, Yan Z, Zhou L, et al. A comparison of glypican-3 with alpha-fetoprotein as a serum marker for hepatocellular carcinoma: a meta-analysis. *J Cancer Res Clin Oncol*. 2013;139(8):1417–1424. doi:10.1007/s00432-013-1458-5
41. Kanwal F, Khaderi S, Singal AG, et al. Risk stratification model for hepatocellular cancer in patients with cirrhosis. *Clin Gastroenterol Hepatol*. 2023;21(13):3296–3304e3. doi:10.1016/j.cgh.2023.04.019
42. Di Poto C, He S, Varghese RS, et al. Identification of race-associated metabolite biomarkers for hepatocellular carcinoma in patients with liver cirrhosis and hepatitis C virus infection. *PLoS One*. 2018;13(3):e0192748. doi:10.1371/journal.pone.0192748
43. Baniyasadi H, Gowda GAN, Gu H, et al. Targeted metabolic profiling of hepatocellular carcinoma and hepatitis C using LC - MS / MS. *Electrophoresis*. 2013;34(19):2910–2917. doi:10.1002/elps.201300029
44. Xiao J, Zhao Y, Varghese RS, et al. Evaluation of metabolite biomarkers for hepatocellular carcinoma through stratified analysis by gender, race, and alcoholic cirrhosis. *Cancer Epidemiol Biomarkers Prev*. 2014;23(1):64–72. doi:10.1158/1055-9965.EPI-13-0327
45. Lu H, George J, Eslam M, et al. Discriminatory changes in circulating metabolites as a predictor of hepatocellular cancer in patients with metabolic (dysfunction) associated fatty liver disease. *Liver Cancer*. 2023;12(1):19–31. doi:10.1159/000525911
46. Lewinska M, Santos-Laso A, Arretxe E, et al. The altered serum lipidome and its diagnostic potential for Non-Alcoholic Fatty Liver (NAFL)-associated hepatocellular carcinoma. *EBioMedicine*. 2021;73:103661. doi:10.1016/j.ebiom.2021.103661
47. Hang D, Yang X, Lu J, et al. Untargeted plasma metabolomics for risk prediction of hepatocellular carcinoma: a prospective study in two Chinese cohorts. *Int J Cancer*. 2022;151(12):2144–2154. doi:10.1002/ijc.34229
48. Nenu I, Stefanescu H, Procopet B, et al. Navigating through the lipid metabolism maze: diagnosis and prognosis metabolites of hepatocellular carcinoma versus compensated cirrhosis. *J Clin Med*. 2022;11(5):1292. doi:10.3390/jcm11051292

49. Sanchez JI, Jiao J, Kwan S-Y, et al. Lipidomic profiles of plasma exosomes identify candidate biomarkers for early detection of hepatocellular carcinoma in patients with cirrhosis. *Cancer Prev Res (Phila)*. 2021;14(10):955–962. doi:10.1158/1940-6207.CAPR-20-0612
50. Wang H, Zhou Y, Xu H, et al. Therapeutic efficacy of FASN inhibition in preclinical models of HCC. *Hepatology*. 2022;76(4):951–966. doi:10.1002/hep.32359
51. Bansal S, Berk M, Alkhouri N, et al. Stearoyl-CoA desaturase plays an important role in proliferation and chemoresistance in human hepatocellular carcinoma. *J Surg Res*. 2014;186(1):29–38. doi:10.1016/j.jss.2013.07.001
52. Loomba R, Mohseni R, Lucas KJ, et al. TVB-2640 (FASN Inhibitor) for the treatment of nonalcoholic steatohepatitis: FASCINATE-1, a randomized, placebo-controlled phase 2a trial. *Gastroenterology*. 2021;161(5):1475–1486. doi:10.1053/j.gastro.2021.07.025
53. Xue R, Lin Z, Deng C, et al. A serum metabolomic investigation on hepatocellular carcinoma patients by chemical derivatization followed by gas chromatography/mass spectrometry. *Rapid Commun Mass Spectrom*. 2008;22(19):3061–3068. doi:10.1002/rcm.3708

Journal of Hepatocellular Carcinoma

Dovepress

Publish your work in this journal

The Journal of Hepatocellular Carcinoma is an international, peer-reviewed, open access journal that offers a platform for the dissemination and study of clinical, translational and basic research findings in this rapidly developing field. Development in areas including, but not limited to, epidemiology, vaccination, hepatitis therapy, pathology and molecular tumor classification and prognostication are all considered for publication. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/journal-of-hepatocellular-carcinoma-journal>