Supporting Information

A large-scale, multi-center serum metabolite biomarkers identification study for the early detection of hepatocellular carcinoma

Ping Luo[1,8#], Peiyuan Yin[1#], Rui Hua[2#], Yexiong Tan[3], Gaokun Qiu[4], Zaifang Li[1,8], Zhenyu Yin[5], Xingwang Xie[6], Xiaomei Wang[2], Wenbin Chen[7], Lina Zhou[1], Xiaolin Wang[1], Yanli Li[1], Hongsong Chen[6], Ling Gao[7], Xin Lu[1], Tangchun Wu[4*], Hongyang Wang[3*], Junqi Niu[2*], Guowang Xu[1*]

1. CAS Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China

2. Department of Hepatology, First Hospital, Jilin University, Changchun, Jilin 130021, China.

3. International Cooperation Laboratory on Signal Transduction, Eastern Hepatobiliary Surgery Institute, the Second Military Medical University, Shanghai, 200031, China.

4. MOE Key Lab of Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science & Technology, Wuhan 430030, Hubei, China.

5. Zhongshan Hospital of Xiamen University, Xiamen, China.

6. Peking University People's Hospital, Beijing 100044, China.

7. Shangdong Provincial Hospital affiliated to Shandong University, Jinan, China.

8. University of Chinese Academy of Sciences, Beijing 100049, China.

*: Corresponding author:

Prof. Dr. Guowang Xu, e-mail: xugw@dicp.ac.cn.

Prof. Junqi Niu, e-mail: junqiniu@aliyun.com.

Prof. Hongyang Wang, e-mail: hywangk@vip.sina.com.

Prof. Tangchun Wu, e-mail: wut@mails.tjmu.edu.cn.

Table of contents

LC-MS analysis. A modified pseudotargeted method (18) was used to acquire the LC-MS spectra in the discovery and test sets to improve the stability of the long-term metabolomics study. The acquisition was performed with an ACQUITY UPLCTM system (Waters, Milford, MA, USA) coupled to a Q-Trap 5500 mass spectrometer system (AB SCIEX, Framingham, USA). A Waters ACQUITY BEH C8 column (100 mm × 2.1 mm, 1.7 μm) was used in ESI+ mode. Both mobile phases A and B contained 0.1% (v/v) formic acid in an aqueous solution and an acetonitrile solution, respectively. The elution gradient started with 10% B, and linearly increased to 40% B at 3 min, then to 100% B at 17 min and was held for 4 min. Another 4 min incubation was used to equilibrate the system with the initial conditions for the next injection. A Waters ACQUITY HSS T3 column (100 mm × 2.1 mm, 1.8 μm) was used in ESI- mode. Mobile phases C and D contained 6.5 mM ammonium bicarbonate in an aqueous solution and a 5% water/methanol solution, respectively. The elution gradient started with 2% D, linearly increased to 40% D at 4 min, then to 100% D at 12 min and was held for 4 min. Another 4 min incubation was used to equilibrate the system. For the two ESI modes, the injected volume was 10 μL, the flow rate was 0.35 mL/min and the column temperature was 50 °C. For the mass spectrometer system, the curtain gas, gas1 and gas2 were 0.241, 0.276 and 0.276 MPa, respectively. The electrospray ion source temperature and spray voltage were set to 500 °C, 5500 V and 450 °C, 4500 V for the ESI+ and ESI- modes, respectively.

A targeted analysis was used to quantify the identified biomarkers in the validation set. Separation and acquisition were performed using the LC (nexera x2)-MS (TQ8050) system (Shimadzu, Kyoto, Japan). A T3 column (100 mm × 2.1 mm, 1.8 μm) was employed for the LC separation in the validation set. The injected volume was 5 μL, and both mobile phases A and B contained 0.1% (v/v) formic acid

in an aqueous solution and an acetonitrile solution, respectively. The flow rate was 0.2 mL/min and the column temperature was 50 °C; the elution gradient started with 10% B, and linearly increased to 40% B at 3 min, then to 100% B at 17 min and was held for 4 min. Another 4 min incubation was used to equilibrate the system with the initial conditions for the next injection. For mass spectrometry, the flow of the heating gas, drying gas, and nebulizing gas were set at 10 L/min, 10 L/min, and 3 L/min, respectively. The temperatures of the DL, heat block and interface heater were set to 250 °C, 400 °C, and 300 °C, respectively.

Based on the modified pseudotargeted method, the ion-pairs of the all metabolites were used to collect the metabolic profiling data, the unidentified metabolites were also studied in the discovery cohort. At the begin of statistical analysis, all these detected known and unknown metabolites had pre-test, if the unknown metabolites were considered as important metabolites, they would be further identified one by one. Once they were identified, they were added to the group of known metabolites. Metabolites were identified based on our previous study, the searches in the home-developed database (containing MS, MS^2 and retention time information from more than 2,000 metabolite standards) and online databases (HMDB and Metlin), and confirmed with authentic standards if available.

The quantitation of GCA and Phe-Trp in the validation cohort was used by the corresponding isotope-labeled internal standards (0.6 μg/mL GCA-d5, and 0.0125 μg/mL Phe-Trp-13C9 in the acetonitrile). The methodology validation of these quantifications in serum samples were conducted as follow: replicates of QC samples were analyzed during consecutive three days (n=3*3) to characterized the repeatability. The coefficient of variation (CV%) based on corresponding replicates for each metabolites was applied to investigate the precision of intra-day and inter-day

(Supplementary Table S1). 11 concentration levels of GCA and Phe-Trp (0.0006, 0.001, 0.006, 0.01, 0.06, 0.1, 0.6, 1, 6, 11 and 111 μg/mL) were performed to test the linearity of the quantitation. Regression coefficients (R) of the linear regression based on these concentration levels were used to evaluate the linearity behavior (Supplementary Table S1). 10 μL standard solution of GCA and Phe-Trp at low, medium, and high levels (1 μg/mL, 10 μg/mL and 100 μg/mL for GCA; 2 μg/mL, 10 μg/mL and 20 μg/mL for Phe-Trp, respectively) was spiked into 50 μL serum sample prior to and after extraction to assessed the recovery (Supplementary Table S1). Three replicates were performed for each level (n = 3).
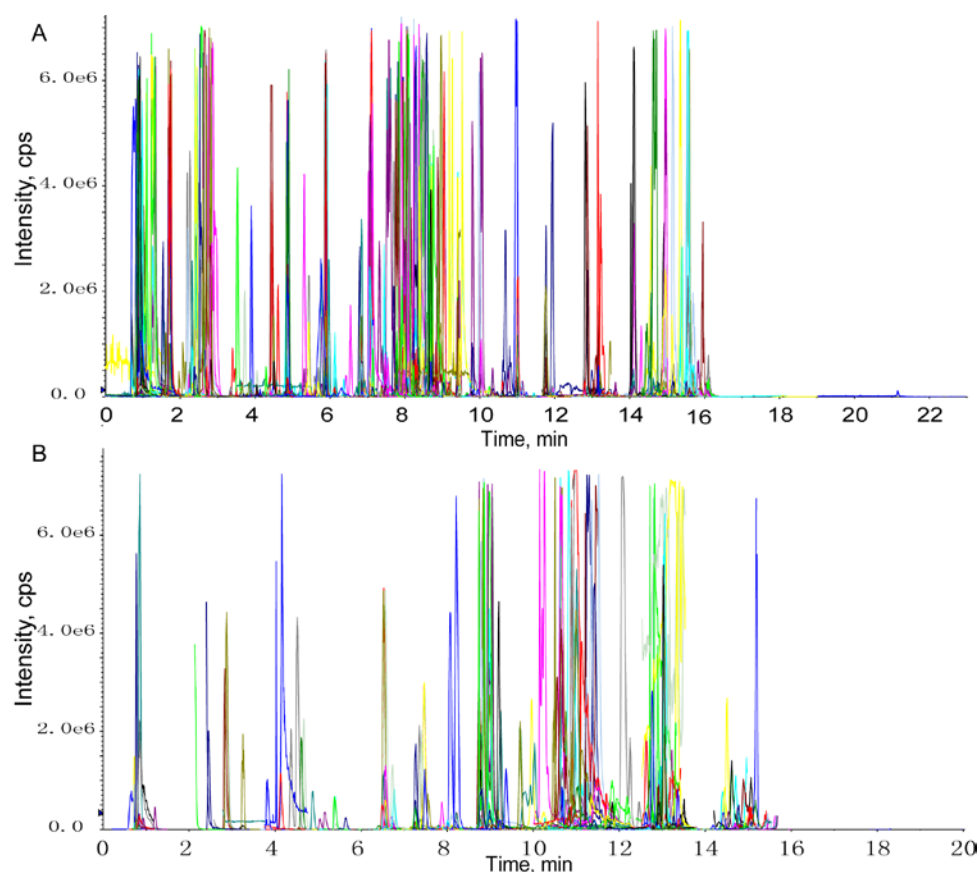
Figure S1. Typical EIC of the ESI+ (A) and ESI- (B) modes.
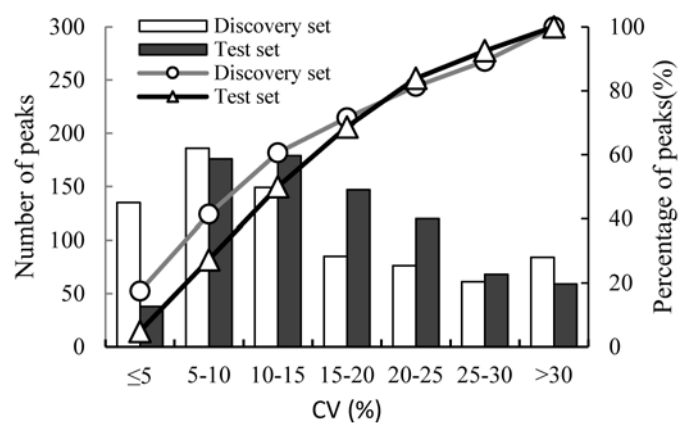
Figure S2. CV distribution of peaks in combinational dataset of ESI+ and ESI- modes from the discovery and validation sets, respectively. The columns and lines represent the number of peaks and accumulative percentage of peaks in corresponding CV interval, respectively. After removing the missing values using the 80% rule, and signal correction using the local linear regression (LoReg) method, 61% and 89% peaks had coefficients of variation (CV) below 15% and 30%, respectively
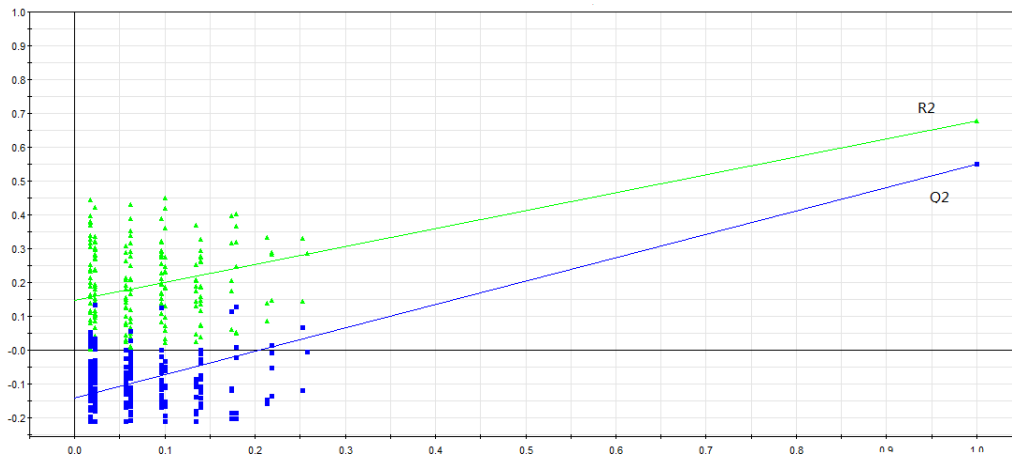
Figure S3. Cross-validation plot with a permutation test repeated 200 times. The intercepts of R2=(0.0, 0.15) and Q2=(0.0, -0.15) illustrate the PLS-DA model is not over-fitting.
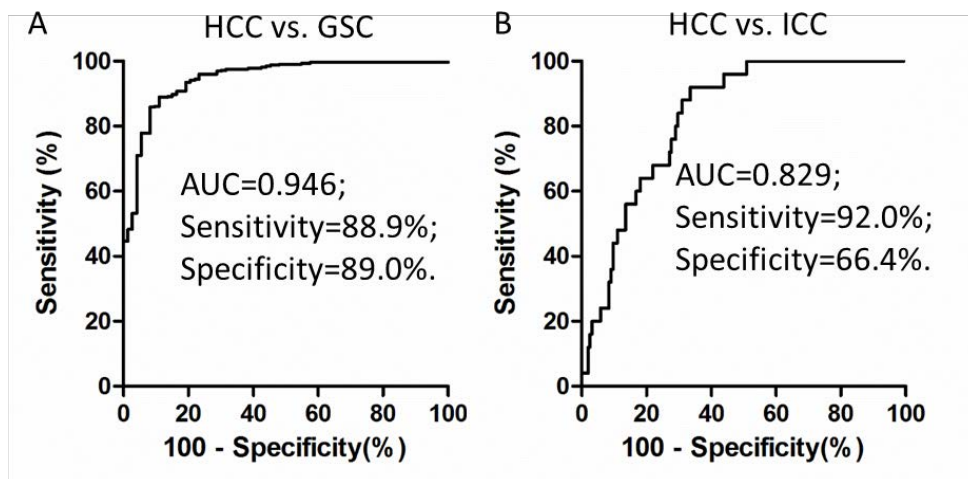
Figure S4. Diagnostic performances of the serum metabolite panel for the discrimination of HCC from other two cancers. ROC curve of this panel for patients with HCC versus patients with GSC(A), ICC(B) in the test set and validation set, respectively.
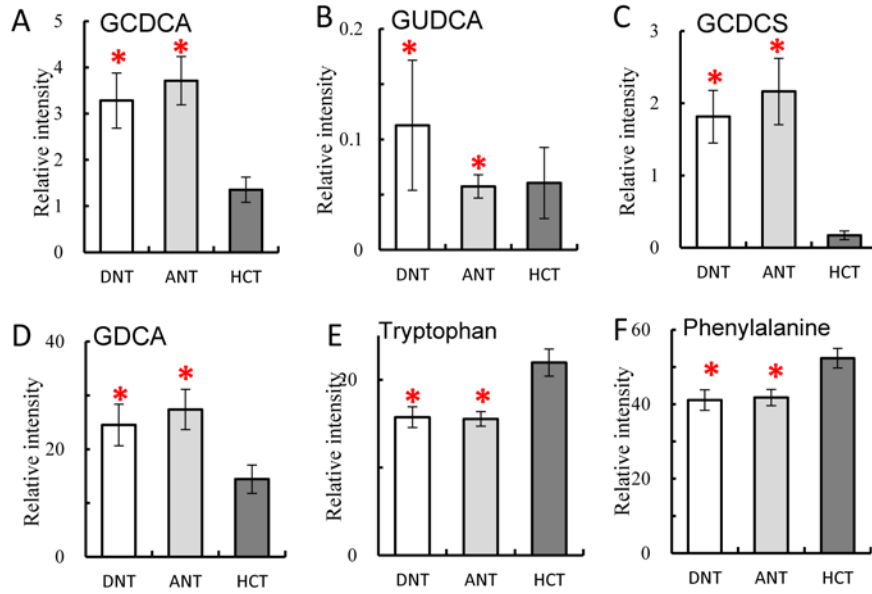
Figure S5. Histograms of differential metabolites related to bile acids and Phe-Trp in DNT, ANT and HCT liver tissues. *: $p < 0.05$ when compared with HCT. All data are presented as mean ± SEM. GCA, GCDCA, glycochenodeoxycholate; GUDCA, glycoursodeoxycholate; GCDCS, glycoursodeoxycholate sulfate; and GDCA, glycodeoxycholate.
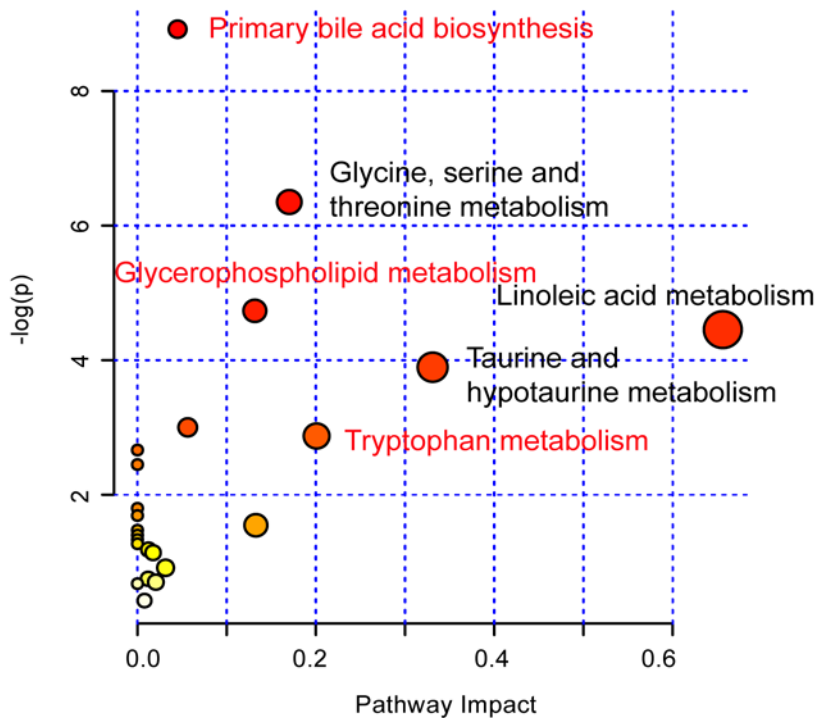
Figure S6. Pathway analysis related to the differential metabolites of HCC group. Global metabolite pathways related to the perturbations of HCC were performed by the website of MetaboAnalyst based on all the differential metabolites listed in table S1.

Table S1. The results of methodology validation for the quantification of Phe-Trp and GCA in serum.

	precisions（%）		linear range	regression	recovery (%)		
	intar-day	inter-day	（μg/mL)	coefficient (r)	low	medium	high
Phe-Trp	7.2	7.1	0.0006-5.6	0.999	104.9	102.7	89.3
GCA	3.4	9.2	0.001-111	0.999	92	91.5	105.4

Table S2. Differential metabolites between HCC and the other two groups (normal controls and cirrhosis) in the discovery and test sets.

Metabolite	ESI Mode	Precursor Ion (Da)	Product Ion (Da)	t_R (min)	CE (V)	DP (V)	VIP1	VIP2	fold change H/N	fold change H/C
dimethyl sulfone[#]	+	95.0	63	0.7	15	90	1.0	0.8	0.5*	0.9
L-serine[#]	+	106.0	60	0.7	15	50	0.7	0.6	1.3*	1.0
Creatinine[a#]	+	114.0	44	0.8	20	100	1.0	1.0	0.7*	0.9
DMG[a#]	+	104.0	58	0.8	20	90	2.0	1.5	0.7*	1.0
Choline[abc#&]	+	104.5	60	0.8	30	110	2.3	2.0	0.5*	0.8*
GPC[a#]	+	258.2	104	0.8	25	90	1.2	1.4	0.8*	2.2
3-(1-pyrazolyl)-L-alanine[#]	+	156.0	88	0.9	17	80	0.5	0.8	0.2*	0.3
Serotonin[a#&]	+	177.3	160	0.9	10	90	1.8	1.4	0.4*	1.6
Hypoxanthine[#]	+	137.1	119	0.9	40	100	0.3	1.3	3.0	4.0*
Arginine[#]	+	175.0	158	1.0	20	100	0.8	1.7	1.0	0.6*
Taurine[abc#&]	+	126.0	44	1.1	20	90	1.2	1.2	0.8*	1.3*
L-kynurenine[#&]	+	192.2	146	1.6	20	130	0.3	1.8	1.2	1.2*
ACN C4:0[#]	+	232.2	85	1.6	40	100	1.0	0.9	0.8*	1.3*
Indole-3-carbinol[a#]	+	130.4	103	1.7	27	140	1.5	1.2	0.8*	1.0
ACN C5:0[%]	+	246.2	85	2.1	20	100	0.2	0.6	0.8*	0.8
DHC[#]	+	388.0	89	2.1	20	90	0.6	1.6	0.6	0.1*
Hippuric acid[#&]	+	180.1	105	2.1	20	100	0.9	0.7	0.5*	1.1
Phe-phe[#]	+	313.2	120	2.6	40	100	0.1	1.7	0.7*	0.6*
Phe-Trp[abc#&]	+	352.2	120	2.7	40	100	2.2	1.9	0.1*	0.4*
ACN-OH[ab%]	+	310.2	85	3.6	40	100	1.0	1.1	1.9*	0.7*
ACN C8:0[%]	+	288.2	85	3.8	40	100	0.9	1.0	0.5*	0.7
2,6 dimethylheptanoyl carnitine[a#]	+	302.2	85	4.1	40	100	1.8	1.4	0.3*	0.8
TCA[abc#&]	+	480.3	462	4.2	20	100	1.1	1.3	40.3*	0.4*
N-Acetyl-L-tyrosine[a#&]	+	223.1	181	4.3	40	100	1.5	1.1	0.3*	2.2
ACN C10:1[#]	+	314.2	85	4.4	40	100	0.8	0.9	0.6*	0.8
Inositol[ab#&]	+	181.1	81	4.6	20	100	2.3	2.0	0.4*	5.2*
GCA[abc#&]	+	488.3	470	4.7	40	100	1.3	1.5	5.8*	0.5*
ACN C10:0[%]	+	316.3	85	4.8	40	100	1.0	0.9	0.6*	0.9
ACN C12:1[%]	+	342.3	85	5.4	40	100	0.2	0.8	0.6*	0.6
GCDCA[a#&]	+	472.3	454	5.6	40	100	1.3	1.5	3.5	0.5*
LPC 12:0[#]	+	440.3	184	5.7	30	100	0.6	0.7	0.6*	0.8
ACN C12:0[%]	+	344.3	85	5.9	40	100	0.7	0.9	0.5*	0.7
ACN C14:2[%]	+	368.3	85	5.9	40	100	0.6	0.9	0.6*	0.7
GCDCS[#&]	+	528.3	448	5.9	-40	-100	0.7	0.6	3.7*	0.8
ACN C14:1[%]	+	370.3	85	6.5	40	100	0.7	0.8	0.5*	0.7
Aldosterone[#]	+	361.2	163	6.8	40	100	0.8	0.6	1.1*	1.0

ACN C16:2%	+	396.3	85	6.9	40	100	0.6	0.8	0.5*	0.7
LPC 14:0a%	+	468.3	184	6.9	40	100	1.5	1.2	0.6*	0.9
LPE 16:1#	+	452.3	311	7.3	40	100	0.6	1.1	0.6*	0.7*
LPC 16:1a#	+	494.3	184	7.3	40	100	1.5	1.2	0.7*	1.0
LPC 18:3a%	+	518.3	459	7.4	40	100	1.9	1.4	0.4*	1.1
LPC 15:0a%	+	482.3	184	7.5	40	100	1.8	1.3	0.5*	1.1
LPC 18:2 %	+	520.3	184	7.5	40	100	0.3	0.6	1.1*	1.0
LPE 20:5abc%	+	500.3	457	7.7	40	100	1.3	1.4	0.5*	0.7*
LPC 22:6a%	+	568.3	184	7.7	40	100	1.9	1.5	0.4*	1.0
LPE 18:2abc%	+	478.3	337	7.7	40	100	1.4	1.4	0.4*	0.7*
LPC 20:5 %	+	542.3	483	7.8	40	100	0.9	2.0	0.8*	0.8*
LPE 22:4abc%	+	530.3	471	7.9	40	100	1.6	1.5	0.5*	0.8*
LPE 16:0a#	+	454.3	313	8.1	40	100	1.4	1.3	0.5*	0.8
LPC 16:0a#	+	496.3	184	8.2	40	100	1.9	1.5	0.6*	0.9
LPC 18:1 #	+	522.4	184	8.3	40	100	1.2	0.9	0.8*	1.0
LPC 20:3a%	+	546.4	184	8.3	40	100	1.8	1.4	0.6*	0.9
LPC O-16:0a%	+	482.4	104	8.4	40	100	1.6	1.4	0.5*	0.8
LPE 18:1#	+	480.3	339	8.5	40	100	0.8	1.3	0.6*	0.7*
LPC 22:5a%	+	570.4	184	8.5	40	100	2.1	1.6	0.3*	0.9
LPC O-18:0a%	+	510.4	104	8.7	40	100	2.1	1.5	0.4*	1.1
LPC P-18:0a%	+	508.4	104	8.8	40	100	1.5	1.4	0.5*	0.8
LPE 18:0a#	+	482.3	341	9.2	40	100	1.5	1.2	0.6*	0.9
MG 16:0a%	+	331.3	57	9.2	40	100	1.6	1.3	0.7*	1.0
LPE 20:2a%	+	506.4	163	9.3	40	100	1.8	1.5	0.6*	0.9
LPC 18:0 #	+	524.4	184	9.3	40	100	1.2	0.9	0.8*	1.0
SM d18:1a%	+	466.4	294	9.6	40	100	1.4	1.1	0.6*	1.2
Mesterolonea#	+	305.3	93	10.8	40	100	2.0	1.6	0.4*	0.9
FAAD C20:1%	+	310.3	69	11.6	40	100	0.1	1.7	0.5*	0.4*
FAAD C22:2%	+	336.3	95	11.9	60	100	0.2	1.2	0.6	0.6*
FAAD C20:0%	+	312.3	74	12.4	60	100	0.5	0.9	0.6	0.7*
PC(32:4)%	+	726.5	184	12.6	30	100	1.0	0.9	0.6*	0.9
SM18:2%	+	727.6	184	13.3	30	100	0.9	0.7	0.7*	1.1
PC(36:6)%	+	778.5	184	13.9	30	100	1.2	0.9	0.5*	1.1
PC(39:2)%	+	828.6	184	13.9	30	100	0.5	0.6	0.7*	0.9
PC(34:4)a%	+	754.5	184	14.2	30	100	1.6	1.2	0.4*	1.1
PC(38:7)a%	+	804.6	184	14.3	30	100	1.4	1.1	0.6*	1.3
PC(36:5)a%	+	780.6	184	14.5	30	100	1.4	1.0	0.5*	1.1
PC(36:4)%	+	782.6	184	14.7	30	100	0.2	1.6	0.8*	0.7
PC(32:1)%	+	732.6	184	15.0	30	100	1.1	0.9	1.8*	0.8
PC(38:5)%	+	808.6	184	15.4	30	100	1.0	0.7	0.7*	1.0
PC(32:0)ab#	+	734.6	184	15.6	30	100	1.4	1.5	1.8*	0.7*
PC(38:4)%	+	810.6	184	15.8	30	100	0.5	0.5	0.8*	0.9
PC(31:0)a%	+	720.6	184	15.9	30	100	1.4	1.3	2.2*	0.8
Uridine#	−	243.1	153	1.0	-30	-100	0.3	0.3	0.9*	1.0

Tryptophan[#&]	−	203.1	116	3.4	-30	-100	0.9	0.7	1.2*	1.0
ANDS[#]	−	369.2	97	7.7	-50	-100	0.8	0.7	0.7*	1.3
DHTS[a#]	−	369.2	97	7.7	-50	-100	1.4	1.1	0.4*	1.2
Eicosenoic acid[#]	−	307.7	308	8.0	-25	-100	0.5	0.4	3.7	0.8*
LPC 20:2[a%]	−	548.4	184	8.9	40	100	2.0	1.5	0.4*	1.0
TCDCA[a#&]	−	498.3	124	9.0	-60	-100	1.1	1.2	20.6*	0.5
TDCA[#&]	−	498.3	124	9.0	-60	-100	0.6	0.8	13.8*	0.3
FFA 18:3[#]	−	277.2	277	10.6	-12	-100	1.0	0.9	1.6*	0.8
FFA 20:5[a%]	−	301.2	301	10.8	-12	-100	1.8	1.4	0.5*	1.0
FFA 22:7[ab%]	−	325.2	325	11.0	-12	-100	1.9	1.6	0.4*	2.2*
FFA 22:5[a%]	−	329.2	285	12.3	-20	-100	1.5	1.2	0.3*	0.9
FFA 19:0[%]	−	297.3	297	13.0	-12	-100	0.8	0.6	2.0*	1.0
FFA 24:2[%]	−	363.3	331	13.3	-30	-100	0.8	1.0	1.3*	1.1
FFA 24:1[a%]	−	365.3	332	13.4	-30	-100	1.3	1.0	1.2*	1.0
FFA 24:4[a%]	−	359.3	359	13.5	-12	-100	1.7	1.5	0.5	1.8*

DMG, N,N-dimethylglycine; GPC, Glycerophosphate choline; DHC, dihydrocholesterol; Phe-Phe, PhenylalanylPhenylalanine; Phe-Trp, Phenylalanyltryptophan; DMHC, dimethylheptanoylcarnitine; ANDS, androsterone sulfate; DHTS, dihydrotestosterone sulfate; TCA, taurocholate; GCA, glycocholate; GCDCA, glycochenodeoxycholate; TCDCA, taurochenodeoxycholate.; GCDCS, glycoursodeoxycholate sulfate; ACN, acetylcarnitine; LPC, lysophosphatidylcholine; LPE, lysophosphatidylethanolamine; SM, shpingomyelin; PC, phosphatidylcholine; FFA, free fatty acid; MG, Monoacylglycerol, FAAD, fatty amide.

*; p value less than 0.05 in the corresponding comparison. The value of fold change >1 or <1 with significance, means up-regulated or down-regulated in HCC group when compared with normal or cirrhosis group.

Marker of [a], [b], [c] means the metabolite came from these 57 metabolites (in figure 3B) with both VIP values larger than 1, these 17 metabolites (in figure 3B) with both VIP values larger than 1 and p-value less than 0.05, and these 8 identified candidates which were further confirmed by test cohort in above mentioned 17, respectively.

[#]: identification by the offline database developed by our laboratory, which contained the MS1, MS2 and retention time of more than 2000 metabolite standards.. [%]: identification by our previous studies; [&]: confirmed with authentic standards.