

Metabolomic Characterization of Hepatocellular Carcinoma in Patients with Liver Cirrhosis for Biomarker Discovery

Cristina Di Poto¹, Alessia Ferrarini¹, Yi Zhao², Rency S. Varghese¹, Chao Tu¹, Yiming Zuo¹, Minkun Wang¹, Mohammad R. Nezami Ranjbar¹, Yue Luo¹, Chi Zhang¹, Chirag S. Desai³, Kirti Shetty⁴, Mahlet G. Tadesse⁵, and Habtom W. Ressom¹

Abstract

Background: Metabolomics plays an important role in providing insight into the etiology and mechanisms of hepatocellular carcinoma (HCC). This is accomplished by a comprehensive analysis of patterns involved in metabolic alterations in human specimens. This study compares the levels of plasma metabolites in HCC cases versus cirrhotic patients and evaluates the ability of candidate metabolites in distinguishing the two groups. Also, it investigates the combined use of metabolites and clinical covariates for detection of HCC in patients with liver cirrhosis.

Methods: Untargeted analysis of metabolites in plasma from 128 subjects (63 HCC cases and 65 cirrhotic controls) was conducted using gas chromatography coupled to mass spectrometry (GC-MS). This was followed by targeted evaluation of selected metabolites. LASSO regression was used to select a set of metabolites and clinical covariates that are associated with HCC. The

performance of candidate biomarkers in distinguishing HCC from cirrhosis was evaluated through a leave-one-out cross-validation based on area under the receiver operating characteristics (ROC) curve.

Results: We identified 11 metabolites and three clinical covariates that differentiated HCC cases from cirrhotic controls. Combining these features in a panel for disease classification using support vector machines (SVM) yielded better area under the ROC curve compared with alpha-fetoprotein (AFP).

Conclusions: This study demonstrates the combination of metabolites and clinical covariates as an effective approach for early detection of HCC in patients with liver cirrhosis.

Impact: Further investigation of these findings may improve understanding of HCC pathophysiology and possible implication of the metabolites in HCC prevention and diagnosis. *Cancer Epidemiol Biomarkers Prev*, 26(5); 675–83. ©2016 AACR.

Introduction

Hepatocellular carcinoma (HCC) is the fifth most common cancer in the world and the third leading cause of cancer mortality worldwide (1). The estimate of new cases of liver cancer (including intrahepatic bile duct cancers) expected to occur in the United States during the 2015 year was 35,660 with approximately three-fourths to be HCC (2). The survival rate of patients with HCC is still 5% (3), and it can only be significantly improved if the

diagnoses are made at earlier stages, when treatment is more effective. Ultrasonography, performed every 6 months, is the currently recommended screening and surveillance for patients with established liver cirrhosis (4). The diagnosis of HCC by imaging techniques requires availability of equipment and a correct interpretation of the results, which are limited in regions with high HCC burden (5). Other than liver imaging and histology, current diagnosis of HCC relies on measurement of level of the serum biomarker, α -fetoprotein (AFP). However, the sensitivity and specificity of AFP are not sufficient for diagnosis of HCC as elevated AFP levels may be seen in patients with cirrhosis or chronic hepatitis too (5). Different variants of AFP such as AFP-L1, AFP-L2, and AFP-L3 have been studied to improve its diagnostic performance (6). Des-gamma-carboxy-prothrombin (DCP) has also been investigated as a potential biomarker for HCC (7). However, reliable serologic biomarkers for early detection of HCC in high-risk population of cirrhotic patients are yet to be found and validated.

Metabolomics has been broadly used for biomarker discovery for many human diseases, including cancer (8). It provides simultaneous assessment of a broad range of metabolites. In this article, we evaluate the levels of plasma metabolites measured by gas chromatography coupled with selected ion monitoring mass spectrometry (GC-SIM-MS), combined with clinical covariates in detecting early-stage HCC cases in patients with liver cirrhosis recruited at MedStar Georgetown University

¹Department of Oncology, Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington DC. ²Department of Biostatistics, School of Public Health, Brown University, Providence, Rhode Island. ³MedStar Georgetown University Hospital and Georgetown University Medical Center, Washington, DC. ⁴Division of Gastroenterology & Hepatology, Johns Hopkins University, Sibley Memorial Hospital, Washington, DC. ⁵Department of Mathematics and Statistics, Georgetown University, Washington DC.

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (<http://cebp.aacrjournals.org/>).

C. Di Poto and A. Ferrarini contributed equally to this article.

Corresponding Author: Habtom W. Ressom, Department of Oncology, Lombardi Comprehensive Cancer Center; Georgetown University Medical Center, Suite 173, Building D, 4000 Reservoir Rd. NW, Washington, DC 20057. Phone: 202-687-2283; Fax: 202-687-0227; E-mail hwr@georgetown.edu

doi: 10.1158/1055-9965.EPI-16-0366

©2016 American Association for Cancer Research.

Hospital (MGUH, Washington, DC). Metabolites and clinical covariates relevant for detecting HCC in cirrhotic patients were selected through least absolute shrinkage and selection operator (LASSO) logistic regression (9). We observed that the combination of LASSO-selected metabolites and AFP, Child–Pugh score, and etiologic factors leads to improved area under the ROC curve compared with AFP. We used correlation and network analyses to evaluate any associations among the selected metabolites and clinical covariates. Finally, we performed pathway enrichment analysis to examine the biological meaning of the results.

Materials and Methods

Study cohort and sample collection

Adult patients were recruited from the hepatology clinic at MGUH. The characteristics of 128 patients investigated in this study are summarized in Table 1. All participants provided informed consent to the study approved by the Institutional Review Board at Georgetown University. The patients were diagnosed to have liver cirrhosis on the basis of established clinical, laboratory, and/or imaging criteria. Cases were diagnosed to have HCC based on well-established diagnostic imaging criteria and/or histology. Clinical stages for HCC cases were determined on the basis of the tumor–node–metastasis (TNM) staging system. Controls were required to be HCC free for at least 6 months from the time of study entry.

Through peripheral venipuncture, single blood sample was drawn into 10 mL BD Vacutainer sterile vacuum tube in the presence of EDTA anticoagulant. The blood was immediately centrifuged at $1,000 \times g$ for 10 minutes at room temperature. The plasma supernatant was carefully collected and centrifuged at $2,500 \times g$ for 10 minutes at room temperature. After aliquoting, plasma was kept frozen at -80°C until use.

Chemical and reagents

Deuterium-labeled internal standards were purchased from CDN isotopes. These include Tyrosine-d2 (D-1611), L-glutamic-2,3,3,4,4-d5 acid (D-899), L-alanine-2,3,3,3-d4 (D-1488), and L-phenyl-d5-alanine-2,3,3,-d3 (D-1241). Glycine-d5 (175838), Myristic acid-d27 (366889), Alkane standard mixture (68281), fatty acid methyl ester standards (FAME), C8 (260673), C9 (245895), C10 (299030), C12 (234591), C14 (P5177), C16 (P5177), C18 (S5376), C20 (10941), C22 (11940), C24 (87115), C26 (H6389), C28 (74701), except for the C30 purchased from TCI chemicals (T0812), methoxyamine hydrochloride (226904), and pyridine (360570) were purchased from Sigma Aldrich. MSTFA (TS-48910) was purchased from Thermo Scientific. HPLC grade 2 propanol, acetonitrile, and water were used for the extraction of metabolites. Helium was purchased from Robert Oxygen.

Experimental design and quality assessment

Among the 128 participants, plasma samples collected from 120 subjects (60 HCC cases and 60 patients with liver cirrhosis) were used for untargeted analysis and plasma samples from 84 subjects (40 HCC cases and 44 patients with liver cirrhosis) were used for targeted analysis, with an overlap of 74 participants between the two analyses. Plasma samples were divided into batches, with balanced proportions of cases and controls by clinical covariates, age, sex, and ethnicity, to allow

Table 1. Characteristic of the study population

		Case	Control	<i>P</i> ^a
		<i>N</i> = 63 (%)	<i>N</i> = 65 (%)	
		63.0	65.0	
Age	Mean (SD)	60.0 (6.4)	58.6 (7.2)	0.2561
Race	African American	17.0 (27.0)	15.0 (23.1)	0.3938
	White	33.0 (52.4)	43.0 (66.2)	
	Asian	6.0 (9.5)	2.0 (3.1)	
	Hispanic/Latino	4.0 (6.3)	2.0 (3.1)	
	Other	3.0 (4.8)	3.0 (4.6)	
Gender	Female	18.0 (28.6)	19.0 (29.2)	1.0000
	Male	45.0 (71.4)	46.0 (70.8)	
Smoker	Current	14.0 (22.2)	15.0 (23.1)	1.0000
	Former	31.0 (49.2)	32.0 (49.2)	
	None	18.0 (28.6)	17.0 (26.2)	
Alcohol	Current	8.0 (12.7)	11.0 (16.9)	0.7809
	Former	33.0 (52.4)	34.0 (52.3)	
	None	21.0 (33.3)	19.0 (29.2)	
BMI	Mean (SD)	30.2 (6.6)	29.2 (6.3)	0.3838
Diabetes	No	39.0 (61.9)	40.0 (61.5)	1.0000
	Yes	24.0 (38.1)	23.0 (35.4)	
Family history of cancer	No	25.0 (39.7)	24.0 (36.9)	0.6129
	Unknown	2.0 (3.2)	5.0 (7.7)	
	Yes	36.0 (57.1)	35.0 (53.8)	
Etiology	Alcohol	17.0 (27.0)	25.0 (38.5)	0.0210
	Autoimmune	2.0 (3.2)	1.0 (1.5)	
	Cryptogenic	1.0 (1.6)	0.0 (0.0)	
	HBV	9.0 (14.3)	1.0 (1.5)	
	HCV	39.0 (61.9)	29.0 (44.6)	
	NAFLD	4.0 (6.3)	3.0 (4.6)	
	Other	2.0 (3.2)	6.0 (9.2)	
	PBC	0.0 (0.0)	3.0 (4.6)	
	PSC	2.0 (3.2)	4.0 (6.2)	
	HCV Ab	Negative	24.0 (38.1)	
	Positive	37.0 (58.7)	27.0 (41.5)	
Anti HBc	Negative	30.0 (47.6)	40.0 (61.5)	0.1822
	Positive	27.0 (42.9)	19.0 (29.2)	
	Unknown	1.0 (1.6)	1.0 (1.5)	
HBsAg	Negative	49.0 (77.8)	55.0 (84.6)	0.1247
	Positive	8.0 (12.7)	3.0 (4.6)	
Ascites	No	37.0 (58.7)	21.0 (33.3)	0.0038
	Yes	24.0 (38.1)	42.0 (66.7)	
AST	Median (IQR)	83.0 (74.0)	67.0 (52.5)	0.1216
ALT	Median (IQR)	70.0 (67.0)	49.5 (36.8)	0.0009
AFP	Median (IQR)	28.8 (102.1)	4.5 (11.0)	0.0000
MELD	Median (IQR)	10.0 (5.0)	14.0 (7.0)	0.0000
	≤10	30.0 (47.6)	10.0 (15.4)	
	>10	28.0 (44.4)	54.0 (83.1)	
	Mean (SD)	11.4 (4.1)	16.2 (13.8)	0.0087
Stage	I	37.0 (58.7)		
	II	20.0 (31.7)		
	III	6.0 (9.5)		
HCV RNA	Median (IQR)	350800.0 (856449.0)	293900.0 (878781.4)	0.7622
	>281	27.0 (42.9)	16.0 (24.6)	
	≤281	5.0 (7.9)	7.0 (10.8)	
Child–Pugh score	Median (IQR)	7.0 (3.0)	9.0 (3.0)	0.0001
	Mean (SD)	7.1 (2.1)	8.8 (2.4)	
Child–Pugh grade	A	24.0 (38.1)	9.0 (13.8)	0.0011
	B	23.0 (36.5)	34.0 (52.3)	
	C	7.0 (11.1)	18.0 (27.7)	

^aFisher exact test was used for categorical variables. Wilcoxon rank sum test was used for continuous variables not symmetrical distributed.

adequate time intervals between sample derivatization and GC-MS instrument calibration prior to each analysis. For the untargeted experiment, samples were split into three batches of

40 samples each, while for targeted analysis, samples were divided into two batches of 41 and 43 samples, respectively. To monitor the system's stability and performance, quality assurance procedures were applied as follows. First, a retention index (RI) standard mixture was run at the beginning and the end of each batch for retention index calibration. The standard mixture was prepared by mixing a series of fatty acid methyl esters (FAME, C8-C30) and Alkanes (C10-C40), as described previously (10). Then, blank samples were prepared together with the patient samples by adding the derivatization reagents to an empty tube and following the same steps, to monitor possible contaminations and background ions introduced by the derivatization process. Finally, quality control (QC) samples were prepared by taking 10 μ L of each derivatized sample within the batch and run at the beginning of each batch for system equilibrium, in between runs and at the end for quality assessment.

Metabolite extraction

Plasma metabolites were extracted by adding 1 mL of a working solution composed of acetonitrile, isopropanol, and water (3:3:2) containing isotope-labeled internal standards at a concentration of 1.25 μ g/mL (Tyrosine-d2, L-glutamic-2,3,3,4,4-d5 acid, L-alanine-2,3,3,3-d4, L-phenyl-d5-alanine-2,3,3-d3, Glycine-d5, Myristic acid d27) to 30 μ L of plasma to evaluate the quality of metabolite extraction. After vortexing, samples were centrifuged at 14,500 \times g for 15 minutes at room temperature. The supernatant was then split into aliquotes of 460 μ L for subsequent untargeted and targeted analyses by GC-MS. Each supernatant was then concentrated to dryness in speedvac. The dried samples were kept at -20° C until derivatization prior to analysis by GC-MS. Twenty microliters of a 20 mg/mL methoxyamine hydrochloride in pyridine was added to the dried extracts, vortexed, and incubated at 30° C for 90 minutes. After returning the samples at room temperature, 80 μ L of MSTFA was added, vortexed, and incubated at 30° C for 30 minutes. Samples were then centrifuged at 14,500 rpm for 15 minutes, and 60 μ L of the supernatant was transferred into 250 μ L clear glass autosampler vials.

Data acquisition and preprocessing

Untargeted metabolomic data were acquired by analyzing metabolites extracted from the plasma samples. The analysis was carried out using two GC-MS systems operated at full scan: a GC-qMS (Agilent Technologies 5975C MSD coupled to an Agilent Technologies 7890A GC) and a GC-TOFMS (LECO Pegasus TOF coupled to an Agilent Technologies 7890A GC). The GC-MS data acquisition and preprocessing were performed following the methods we reported previously (10). For targeted analysis, 46 metabolites from the following three sources were considered: (i) metabolites with statistically significant changes in the untargeted analysis of the samples derived from 120 participants of the same cohort (see ref. 10 for details on the statistical method), (ii) metabolites selected from our previous GC-MS study conducted on an Egyptian cohort (10), and (iii) metabolites retrieved from the literature by text mining. Targeted quantification was performed in selected ion monitoring (SIM) mode by using the GC-qMS platform, as described previously (10). For each analyte, four ions were selected on the basis of their specificity, where one ion was used as a quantifier for intensity calculation and the other three as qualifiers for confirmation. The fragments were manually

selected on the basis of the uniqueness across coeluting analytes and their relative intensity compared with the base peak in the spectrum. Time segments were set up to allow at least 10 msec dwell time for each ion monitored. The complete list of the targeted metabolites (together with the IS) is shown in Supplementary Table S1. The GC-SIM-MS data were pre-processed by the Automated Mass Spectral Deconvolution and Identification System (AMDIS) for peak detection, deconvolution, and identification (11). The resulting peaks were aligned using Mass Profiler Professional (MPP) from Agilent Technologies.

Selection of metabolites and clinical covariates by LASSO

Two LASSO regression models were applied to select a set of metabolites and clinical covariates, respectively, based on their association with HCC or cirrhotic disease status. For the metabolites, the data matrix was obtained by preprocessing the GC-SIM-MS runs of the 84 plasma samples on the basis of the quantifier ion's intensity selected for each of the 46 metabolite targets. The metabolite intensities were log-transformed and the batch effect was removed by using R Combat package. For the clinical covariates, AFP measurements (reported in ng/mL) were also log-transformed to satisfy the linearity assumption with the log-odds of HCC status. For both LASSO models, the tuning parameter was chosen by (leave-one-out) cross-validation with deviance as the loss function. A univariate logistic regression model was also fit on each individual metabolite to examine its association with the risk of HCC. Adjusted *P* values were calculated following the Benjamini-Hochberg procedure (12). To further investigate the performance of the metabolites for early detection of HCC, a multinomial logistic regression model was fit considering the HCC stage I & II combined as a group and using the cirrhotic controls as a reference group.

Performance evaluation of predictors

Logistic regression models and support vector machines (SVM) were built to evaluate the performance of the predictors selected by LASSO. We evaluated the performances of four sets of predictors: (i) AFP measurements only; (ii) clinical covariates selected by LASSO; (iii) metabolites selected by LASSO; and (iv) the combination of (ii) and (iii). Receiver operating characteristics (ROC) curves and 95% confidence interval (CI) of area under the ROC curve (AUC) calculated on the basis of leave-one-out cross-validation were used for performance evaluation.

Correlation analysis

Pearson correlation coefficients were calculated to investigate associations between the LASSO-selected metabolites. Separate correlation graphs were obtained for the HCC and cirrhotic groups by using the R corplot package. The *P* values were adjusted for multiple comparison by the Benjamini & Hochberg procedure (12). The significance cutoff was set to be 0.05. Associations between LASSO-selected metabolites and a subset of clinical covariates were also investigated excluding 17 patients who had missing values for the clinical covariates.

Metabolites ID verification by standards

Identities of the majority of the metabolites selected by LASSO were confirmed by analysis of authentic compounds purchased from Sigma-Aldrich: L-valine (PHR1172), glycine (G7403), D-, L-isoleucine (298689), creatinine (C4255), L-pyroglutamic acid

(83160)/[L-glutamic acid (95436)], alpha-D-glucosamine 1-phosphate (G9753), tagatose (T2751) [sorbose (S4887)], linoleic acid (L1376), lauric acid (61609). Individual 0.25 mg/mL stock standard solutions were prepared in the appropriate solvent and stored at -20°C until the analysis. Working standard solutions, at the concentration of 1.25 $\mu\text{g/mL}$, were prepared by appropriate dilution of the stock standard solutions in acetonitrile, isopropanol, and water (3:3:2). Standards were then concentrated to dryness and derivatized following the same procedure applied for plasma metabolites as described in the "Metabolite extraction" section. Each standard was analyzed by both GC-qMS and GC-TOF-MS platform, following the same GC and MS methods as described in Ranjbar and colleagues. (10). Acquired spectra of the individual standards were cross matched with the corresponding ones extracted from the analysis of the plasma samples.

Network and pathway analysis

To investigate the association among LASSO-selected metabolites, group-specific networks were built for HCC cases and cirrhotic controls through a Gaussian graphical model (GGM) and graphical LASSO algorithm implemented in the R Glasso package (13). In a GGM network, the connection between two nodes indicates a conditional independence between them given all the others. The sparsity parameter was tuned on the basis of the result of a 10-fold cross-validation applying the one standard error rule (14). The shared and group-specific connections between the HCC and cirrhotic GGM networks were also investigated. Further evaluation on the shared connections between two GGM networks was conducted by recovering the metabolites that are not detected in our experiment but are reported in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database to have interaction with a pair of nodes in the GGM networks. This was accomplished by using the Matlab MetaboNetworks toolbox to discover the shortest path between each metabolite pairs from the KEGG database (15). To get a deeper understanding of the biological relation between the

metabolites we detected in our study and those derived from KEGG in HCC and cirrhotic patients, we performed pathway enrichment analysis using MetaboAnalyst 3.0 (16).

Results

Metabolites selected by LASSO regression

LASSO regression selected 11 metabolites whose expression levels jointly differentiate HCC cases from cirrhotic controls. Table 2 presents statistical results for these variables based on multivariable analysis and univariate logistic regression (P values and FDR adjusted P values) along with fold changes (average and median values calculated on the raw intensities after correcting the batch effect). Their fold change ranged between +2.48 (alpha-D-glucosamine 1-phosphate) and -2.18 (tagatose). Figure 1 depicts their dot-plots. The selected metabolites include amino acids and their derivatives (valine, serine, glycine, isoleucine, creatinine and pyroglutamic acid/glutamic acid), sugars, and their derivatives (furanose sugar and alpha-D-glucosamine 1-phosphate), fatty acids (linoleic acid and lauric acid), and one inorganic acid (phosphoric acid). Among the metabolites selected by LASSO, those found to be significant in the multinomial logistic regression model as discriminating stage I and II HCC cases from cirrhotic controls are indicated in Table 2. While we were able to confirm with high confidence (match value >850) the identity of nine of the selected metabolites by comparing their fragmentation patterns with the ones from commercial and/or in-house libraries, we were unable to determine with certainty the identity of the metabolite belonging to the class of furanose sugars. However, based on the similarity with the RI of the standard, we determined tagatose as the likely identification. Also we could not distinguish between pyroglutamic acid and glutamic acid, based on our data. Although in the following sections of the article, both names (pyroglutamic and glutamic acid) have been kept as the identification of the selected metabolite, only glutamic acid was chosen for investigation by literature search

Table 2. Metabolites and clinical covariates selected by LASSO

	Metabolites	Fold change raw		Multivariable analysis Coefficient (P values)	Univariate logistic regression Coefficient (P values, adj. P values ^a)
		Mean	Median		
Amino acids and derivatives	Valine ^{b,c}	↑ +1.42	↑ +1.46	1.249 (0.036)	1.016 (0.007-0.091)
	Serine	↑ +1.23	↑ +1.13	0.958 (0.042)	0.428 (0.152-0.381)
	Glycine ^b	↓ -1.31	↓ -1.21	-1.821 (0.073)	-2.117 (0.002-0.091)
	Isoleucine ^{b,c}	↑ +1.29	↑ +1.28	1.251 (0.138)	1.301 (0.021-0.168)
	Creatinine	↓ -1.58	↓ -1.53	-0.782 (0.138)	-0.508 (0.086-0.330)
	Pyroglutamic acid/glutamic acid ^{b,c}	↑ -1.24	↓ -1.19	-2.101 (0.079)	-2.101 (0.006-0.0914)
Sugars and alcohols	Alpha-D-glucosamine 1-phosphate	↑ +1.49	↑ +2.48	0.521 (0.012)	0.146 (0.160-0.381)
	Tagatose (furanose sugar) ^{b,c}	↓ -2.18	↓ -1.97	0.092 (0.633)	-0.359 (0.011-0.102)
Fatty acids	Linoleic acid ^b	↑ +1.42	↑ +1.77	1.808 (0.007)	0.863 (0.005-0.091)
	Lauric acid	↓ -1.31	↓ -1.35	-1.344 (0.024)	-0.543 (0.123-0.355)
Inorganic acid	Phosphoric acid	↓ -1.03	↓ -1.01	-1.530 (0.140)	-0.331 (0.514-0.685)
Clinical covariates					
	AFP	↑ +1.99	↑ +2.53	0.405 (0.032)	0.547 (0.003-0.065)
	Child-Pugh score	—	—	-0.287 (0.142)	-0.469 (0.005-0.065)
	Etiology (alcohol vs. HCV)	—	—	-1.725 (0.168)	-2.287 (0.042-0.283)
	Etiology (HBV vs. HCV)	—	—	16.929 (0.993)	17.358 (0.994-0.994)
	Etiology (NAFLD vs. HCV)	—	—	-0.378 (0.827)	-0.901 (0.482-0.765)
	Etiology (Other vs. HCV)	—	—	-0.851 (0.370)	-1.055 (0.179-0.479)

^aMultiple testing adjusted P values.

^bMetabolites found to be significant in the multinomial logistic regression model as discriminating HCC - Stage I and II versus cirrhotic controls.

^cMetabolites previously reported in our GC-MS study on an Egyptian cohort (10).

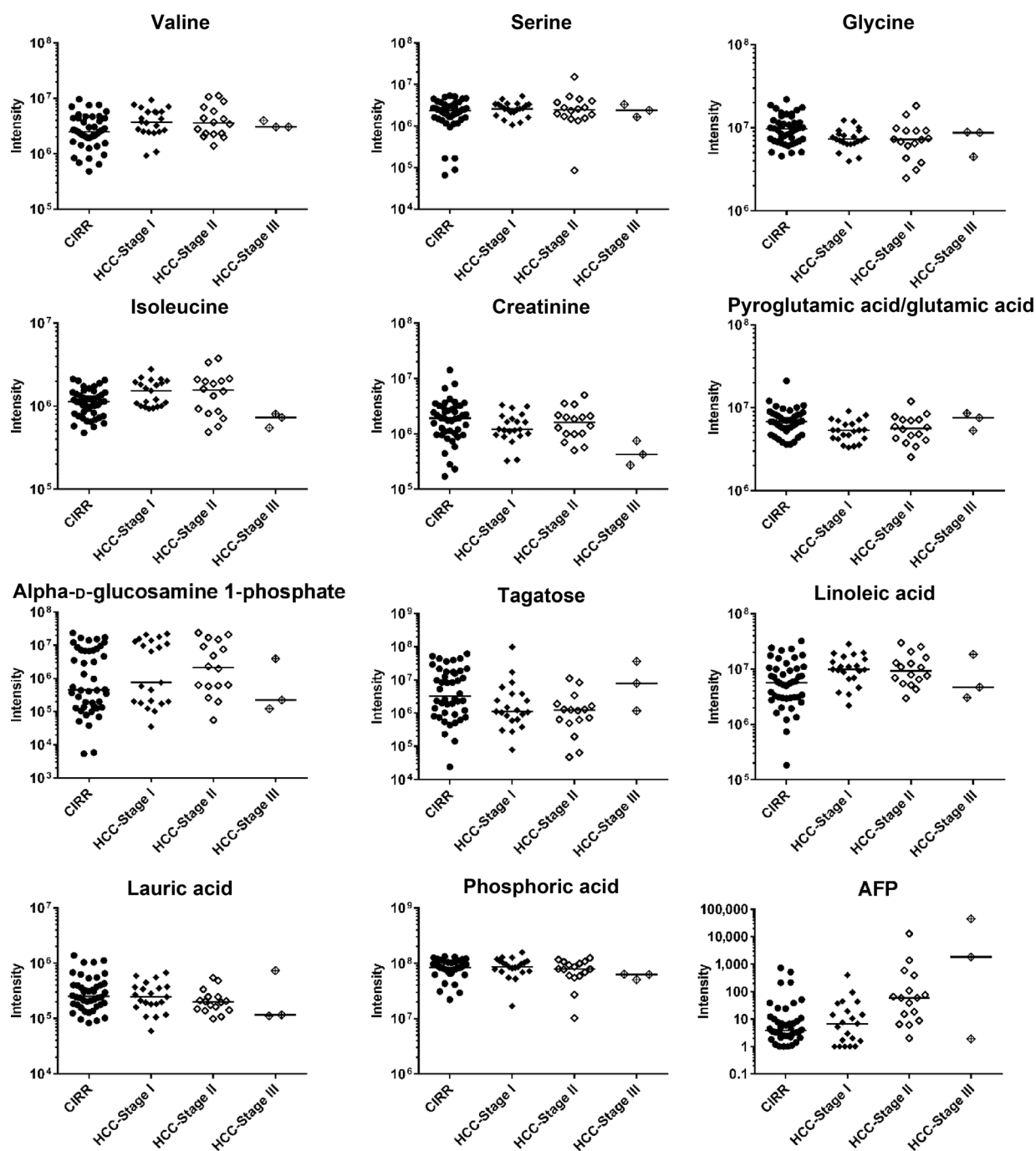
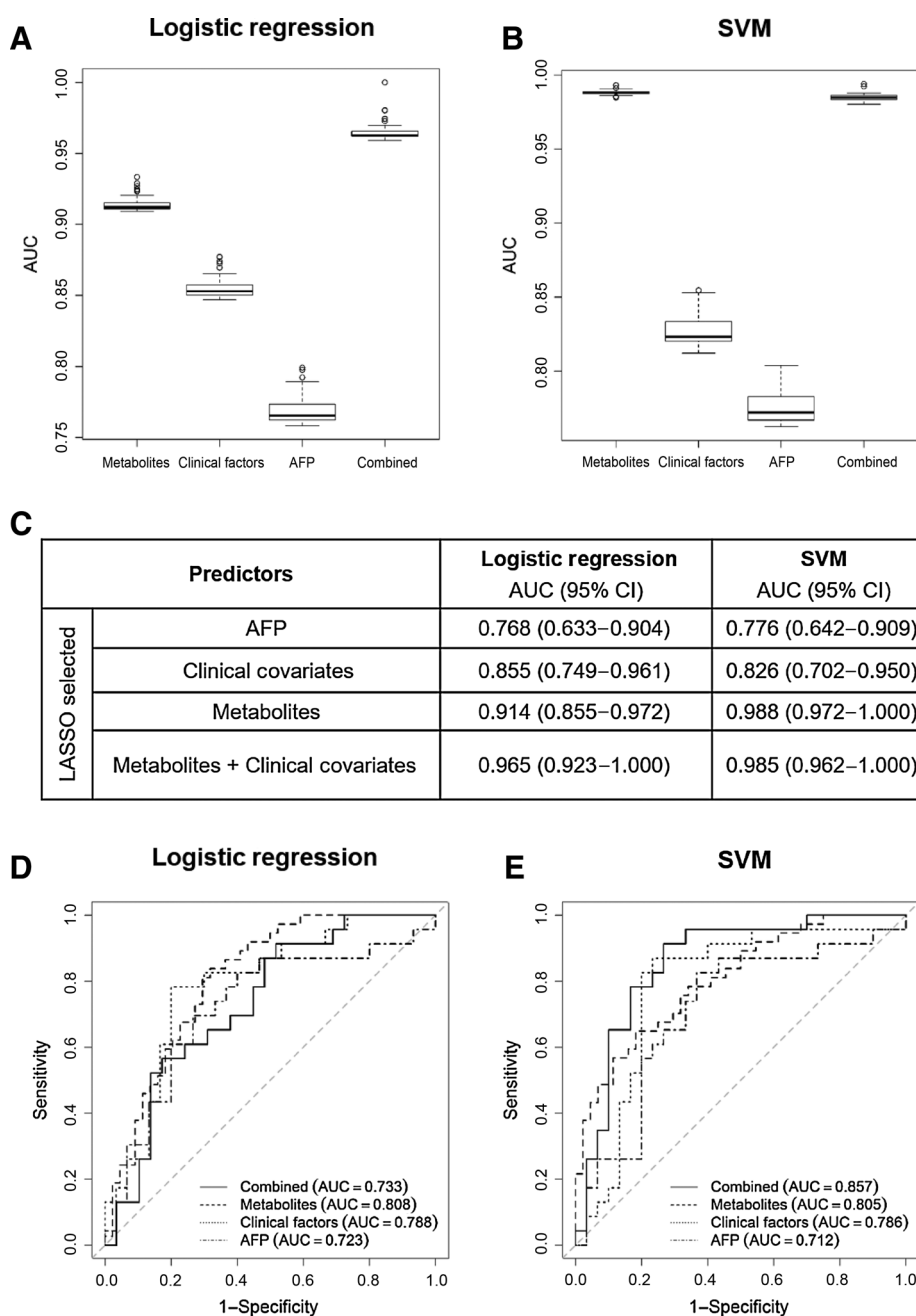


Figure 1. Individual dot plots of LASSO-selected metabolites and AFP. Horizontal lines represent median.

and pathway analysis. This is because pyroglutamic acid is most likely a product of glutamate cyclization during the chemical derivatization process (17). Higher levels of valine, serine, isoleucine, alpha-D-glucosamine 1-phosphate, and linoleic acid were found in HCC cases, while glycine, creatinine, glutamic acid, tagatose, lauric acid, and phosphoric acid were found elevated in cirrhotic controls.

Clinical covariates selected by LASSO regression

LASSO regression analysis, applied on the clinical variables selected AFP (dot-plot in Fig. 1), Child-Pugh score, and etiologic factors comprised of alcohol, viral infection (HBV, HCV), non-alcoholic fatty liver disease (NAFLD), and other less frequent etiologies including autoimmune, primary biliary cirrhosis (PBC), primary sclerosing cholangitis (PSC), and

**Figure 2.**

Box plots of AUC values obtained based on the training set during leave-one-out cross-validation of logistic regression (**A**) and SVM (**B**) models, and the corresponding 95% CI (**C**). ROC curves obtained while testing the logistic regression (**D**) and SVM (**E**) models during the leave-one-out cross-validation using four sets of predictors: AFP (dot dashed line), clinical covariates (dotted line), metabolites only (dashed line), combined metabolites and clinical covariates (solid line).

cryptogenic. The results of multivariable analysis and univariate logistic regression of these clinical covariates are shown in Table 2. Although we did not anticipate the clinical covariates age, sex, and ethnicity to be selected by the LASSO model, since in our study they were matched between HCC cases and cirrhotic controls, we included them anyway into the analysis and, as expected, they did not show up to be significant.

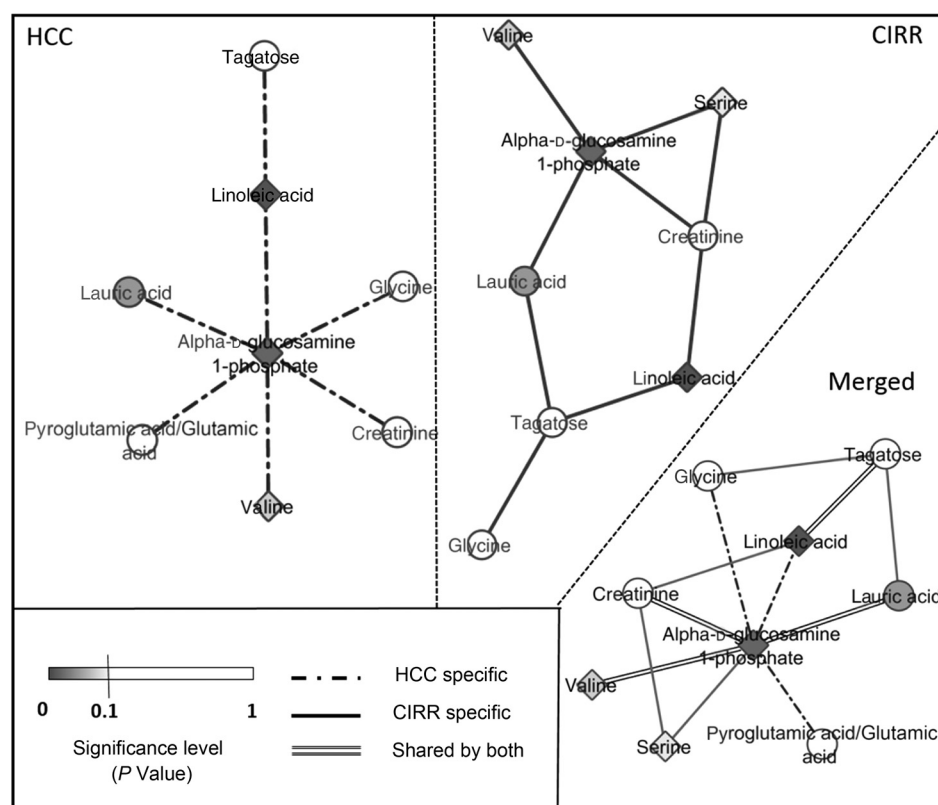
Performance evaluation of predictors

Through a leave-one-out cross-validation, we evaluated the performances of the predictors selected by LASSO in terms of their ability to distinguish cirrhotic controls from early-stage

HCC by excluding the group of patients with HCC stage III ($n = 3$). Figures 2A–C present box plots for AUC values and the corresponding 95% CI calculated on the basis of the training set during leave-one-out cross-validation of logistic regression and SVM models. Figures 2D and E depict ROC curves obtained while testing the logistic regression and SVM models during the leave-one-out cross-validation. As shown in these figures, the logistic regression model built by the LASSO-selected metabolites (AUC = 0.808) and clinical covariates (AUC = 0.788) led to improved performance compared with AFP (AUC = 0.723). Although the logistic regression model built by combining the LASSO-selected metabolites with clinical covariates in a panel

Figure 3.

Network analysis. Each node is shaded in proportion to its significance level – the darker the node, the smaller the adjusted *P* value; the node shape represents the fold change between HCC and CIRR (diamond nodes for fold change > 1, and circular nodes for fold change < 1); the edges represent whether the association between the metabolites was based on the data from the HCC group (dotted line), cirrhotic group (solid line), or shared by both (double line).



(AUC = 0.733) did not perform well, the SVM built using these predictors outperformed (AUC = 0.857) the remaining three sets of predictors: LASSO-selected metabolites (AUC = 0.805), LASSO-selected clinical covariates (AUC = 0.786), and AFP (AUC = 0.712).

Correlation analysis

Pearson correlation among the panel of metabolites selected from LASSO showed a strong relation between amino acids, fatty acids, alpha-D-glucosamine-1-phosphate and pyroglutamic acid/glutamic acid in HCC patients (Supplementary Fig. S1). In particular, creatinine is strongly correlated with pyroglutamic acid/glutamic acid, alpha-D-glucosamine 1-phosphate, and lauric acid. Lauric acid is also positively correlated with pyroglutamic acid/glutamic acid, alpha-D-glucosamine 1-phosphate, and linoleic acid. On the other hand, in cirrhotic controls (Supplementary Fig. S2), phosphoric acid shows positive correlation with linoleic acid which presents also a moderate and strong negative correlation with a furanose sugar. Valine shows a common moderate and strong positive correlation with isoleucine in both HCC cases and cirrhotic controls. In addition to the correlation among the LASSO-selected metabolites, we investigated their relationship with two of the three clinical covariates (AFP, Child-Pugh score) selected by LASSO. In this panel, AFP was negatively correlated with glycine, creatinine, and Child-Pugh score in cirrhotic controls. The Child-Pugh score presented a negative correlation with isoleucine in both groups of patients and with creatinine in HCC cases only.

Network and pathway analysis

Figure 3 shows GGM networks built for HCC and cirrhotic controls separately, along with the merged one. As shown in the

merged graph (Fig. 3, merged panel), there are four connected pairs, composed of five metabolites with statistically significant differences between HCC and cirrhotic groups (darker nodes). The five metabolites are alpha-D-glucosamine 1-phosphate, valine, serine, lauric acid, and linoleic acid. Among them, alpha-D-glucosamine 1-phosphate serves as a hub metabolite connected to all the other four. Further evaluation of these four metabolites by searching for the shortest path between each metabolite pair against the KEGG database (15) revealed 22 metabolites connecting the four metabolites as shown in Supplementary Fig. S3. Metabolite names and KEGG IDs for the original and recovered analytes are listed in Supplementary Table S2. Pathway enrichment analysis, using MetaboAnalyst 3.0, based on the original and recovered metabolites, derived from the GGM network analysis (Supplementary Fig. S3), showed the involvement of the selected metabolites in nine specific pathways common to both HCC and cirrhotic groups (Supplementary Table S3).

Discussion

In this study, we conducted targeted analysis of metabolites in plasma samples from HCC cases and patients with liver cirrhosis. LASSO regression analysis of the metabolomic data selected eleven metabolites and three clinical covariates including AFP. Combined by SVM in a panel, these predictors showed improved performance in disease classification, compared with AFP only (Fig. 2B). If successfully validated, the panel can potentially improve the ability to detect and monitor HCC in high risk population of cirrhotic patients.

Among the LASSO-selected 11 metabolites, four (valine, isoleucine, glutamic acid, and the furanose sugar) were also found statistically significant in a GC-MS-based metabolomic analysis, we conducted previously using plasma samples from HCC cases and patients with liver cirrhosis recruited in Tanta, Egypt (10). Of the three branched-chain amino acids (BCAA), valine and isoleucine were elevated in HCC versus cirrhosis in both the United States and Egyptian cohorts. Although it was not statistically significant, leucine too showed increased level in HCC versus cirrhosis, consistent with our previous findings (10). BCAAs have been reported to have a crucial role in cancer by regulating the anabolic process involving protein synthesis and degradation, needs that are shared by both tumor and normal cells (18). The severe muscle wasting syndrome experienced by many patients with cancer has motivated the use of BCAA supplements, as already extensively used in the athletic field for performance improvement and muscle mass. Therefore, the use of BCAAs as biomarkers is challenging due to the competing energetic and proliferative demands in both healthy and disease states (18–20). However, high levels of BCAAs in HCC samples could be due to their potential tumorigenic effect in liver and may be a significant component of diagnostic testing panels for monitoring the risk of cancer (19). While we found in this study reduced level of glutamic acid in HCC versus cirrhosis, we observed increased level in our previous study involving the Egyptian cohort (10). In another plasma metabolomics study conducted on patients with liver diseases (21), the level of glutamic acid was found decreased in all three types of liver disease (hepatitis, cirrhosis, HCC) when compared with healthy controls. According to the authors, this remarkable reduction can be explained by the altered activities and ratio across the three groups of patients of two monitored transaminases. Among the LASSO-selected sugars, tagatose appears to be downregulated in HCC similarly to sorbose, another furanose sugar, identified in our previous study using the Egyptian cohort (10). To investigate their nature and contribution in promoting hypoxia-inducible factors prevalent in low oxygen environments, as in solid tumors like HCC (17), the use of complementary techniques aimed at discriminating sugar isoforms will be necessary.

The selection by LASSO, of multiple clinical variables, in addition to AFP, seems to be in agreement with several epidemiologic and clinical studies that have shown the increased sensitivity of early detection of HCC in clinical practice when incorporating longitudinal data or adjusting for patient characteristics in addition to the conventional AFP assay (22).

The result of our correlation analysis shows that valine has a moderate and strong positive correlation with isoleucine in both HCC cases and cirrhotic controls, respectively, suggesting its connection not only to cancer but other liver diseases such as cirrhosis as reported previously (18). Also, the Child-Pugh score, a prognostic indicator of liver diseases and necessity of liver transplantation, presented a negative correlation with isoleucine in both groups of patients. This correlation has been previously reported in patients with liver diseases, where the muscle and blood amino acids metabolism were investigated (23).

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *Cancer J Clin* 2015;65:87–108.

The result of the pathways enrichment analysis (Supplementary Table S3; Supplementary Fig. S3) reveals the hepatic metabolome interchange between lipids and water-soluble metabolites crucial for liver energy production and consumption (24), therefore essential for aberrant metabolic reprogramming happening in cancer cells (25).

In summary, the combination of metabolites with clinical covariates, including AFP, has led to better area under the ROC curve in distinguishing early HCC cases in patients with liver cirrhosis when compared with the results obtained by using AFP. Previous HCC-related metabolomics studies, conducted using complementary metabolomics platforms and multivariate analysis, including a GC-MS study conducted on an Egyptian cohort, revealed similar metabolomics findings to the ones reported in this paper. Due to the small sample size used in this study, replication of these findings through a larger cohort including samples that represent diverse populations is desired. Following appropriate validation, the metabolites discovered in this study could contribute to better understanding of the development of HCC and allow early detection of HCC in patients with liver cirrhosis. Most of the clinical covariates selected by LASSO are commonly reported as HCC risk factors. Thus, following validation of the metabolites discovered in this study, their combination with the selected clinical covariates in a panel could contribute to better understanding of the development of HCC and to improving our ability to detect early-stage HCC in patients with liver cirrhosis.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: C. Di Poto, A. Ferrarini, R.S. Varghese, K. Shetty, H.W. Resson

Development of methodology: C. Di Poto, A. Ferrarini, Y. Luo
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): A. Ferrarini, R.S. Varghese, Y. Luo, C.S. Desai

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): C. Di Poto, A. Ferrarini, Y. Zhao, R.S. Varghese, C. Tu, Y. Zuo, M. Wang, M.R. Nezami Ranjbar, C. Zhang, M.G. Tadesse

Writing, review, and/or revision of the manuscript: C. Di Poto, A. Ferrarini, Y. Zhao, R.S. Varghese, Y. Zuo, C.S. Desai, K. Shetty, M.G. Tadesse, H.W. Resson
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): C. Di Poto, A. Ferrarini, K. Shetty

Study supervision: C. Di Poto, A. Ferrarini, H.W. Resson

Acknowledgments

The authors acknowledge Dr. Tsung-Heng Tsai for providing constructive comments.

Grant Support

This work was supported by U01CA185188 (awarded to H.W. Resson).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received May 4, 2016; revised November 1, 2016; accepted November 14, 2016; published OnlineFirst December 2, 2016.

2. American Cancer Society. Breast cancer in situ. *Cancer Facts and Figures* 2015. Atlanta, GA: American Cancer Society; 2015

3. Forner A, Llovet JM, Bruix J. Hepatocellular carcinoma. *Lancet* 2012;379:1245–55.
4. Bruix J, Sherman M. Management of hepatocellular carcinoma: an update. *Hepatology* 2011;53:1020–2.
5. Kim JU, Shariff MI, Crossey MM, Gomez-Romero M, Holmes E, Cox IJ, et al. Hepatocellular carcinoma: review of disease and tumor biomarkers. *World J Hepatol* 2016;8:471.
6. Li D, Mallory T, Satomura S. AFP-L3: a new generation of tumor marker for hepatocellular carcinoma. *Clin Chim Acta* 2001;313:15–9.
7. Lok AS, Sterling RK, Everhart JE, Wright EC, Hoefs JC, Di Bisceglie AM, et al. Des- γ -carboxy prothrombin and α -fetoprotein as biomarkers for the early detection of hepatocellular carcinoma. *Gastroenterology* 2010;138:493–502.
8. Liesenfeld DB, Habermann N, Owen RW, Scalbert A, Ulrich CM. Review of mass spectrometry-based metabolomics in cancer research. *Cancer Epidemiol Biomarkers Prev* 2013;22:2182–201.
9. Tibshirani R. Regression shrinkage and selection via the lasso. *J Royal Stat Soc. Series B (Methodological)* 1996;267–88.
10. Ranjbar MRN, Luo Y, Di Poto C, Varghese RS, Ferrarini A, Zhang C, et al. GC-MS based plasma metabolomics for identification of candidate biomarkers for hepatocellular carcinoma in Egyptian cohort. *PLoS ONE* 2015;10:e0127299.
11. Stein SE. An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *J Am Soc Mass Spectrometry* 1999;10:770–81.
12. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc. Series B (Methodological)* 1995;57:289–300.
13. Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 2008;9:432–41.
14. Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and regression trees. Boca Raton, FL: CRC Press; 1984.
15. Posma JM, Robinette SL, Holmes E, Nicholson JK. MetaboNetworks, an interactive Matlab-based toolbox for creating, customizing and exploring sub-networks from KEGG. *Bioinformatics* 2014;30:893–5.
16. Xia J, Sinelnikov IV, Han B, Wishart DS. MetaboAnalyst 3.0—making metabolomics more meaningful. *Nucleic Acids Res* 2015;43:W251–7.
17. Armitage EG, Kotze HL, Allwood JW, Dunn WB, Goodacre R, Williams KJ. Metabolic profiling reveals potential metabolic markers associated with Hypoxia Inducible Factor-mediated signalling in hypoxic cancer cells. *Sci Rep* 2015;5:15649.
18. O'Connell TM. The complex role of branched chain amino acids in diabetes and cancer. *Metabolites* 2013;3:931–45.
19. Tom A, Nair KS. Assessment of branched-chain amino acid status and potential for biomarkers. *J Nutr* 2006;136:324S–30S.
20. Liu KA, Lashinger LM, Rasmussen AJ, Hursting SD. Leucine supplementation differentially enhances pancreatic cancer growth in lean and overweight mice. *Cancer Metab* 2014;2:6.
21. Lin X, Zhang Y, Ye G, Li X, Yin P, Ruan Q, et al. Classification and differential metabolite discovery of liver diseases based on plasma metabolic profiling and support vector machines. *J Separation Sci* 2011;34:3029–36.
22. Singal AG, El-Serag HB. Hepatocellular carcinoma from epidemiology to prevention: translating knowledge into practice. *Clin Gastroenterol Hepatol* 2015;13:2140–51.
23. Dam G, Sørensen M, Buhl M, Sandahl TD, Møller N, Ott P, et al. Muscle metabolism and whole blood amino acid profile in patients with liver disease. *Scand J Clin Lab Invest* 2015;75:674–80.
24. Beyoğlu D, Idle JR. The metabolomic window into hepatobiliary disease. *J Hepatol* 2013;59:842–58.
25. Zhou S, Huang C, Wei Y. The metabolic switch and its regulation in cancer cells. *Sci China Life Sci* 2010;53:942–58.