Dipartimento di Medicina, Chirurgia e Odontoiatria
"SCUOLA MEDICA SALERNITANA"

Corso di Dottorato
Medicina Traslazione dello
Sviluppo e dell'Invecchiamento Attivo
Ciclo XXXIII
Codice DOT1328517

Coordinatore
Prof. Palmiero Monteleone

Tesi di Dottorato

"Untargeted metabolomics as a diagnostic tool in NAFLD: discrimination of steatosis, steatohepatitis and cirrhosis"

Relatore
Char.mo Prof.
Marcello Persico

Candidato
Dott. Valerio Rosato
Matr. 8800900026

Anno Accademico 2017/2020

Abbreviations: ALT: Alanine aminotransferase (ALT); ANOVA: analysis of variance; AST: Aspartate aminotransferase; BMI: body mass index; CAP: Continuous Attenuation Parameter; GC-MS: gas chromatography-mass spectrometry; GGT: Gamma-glutamyl transferase (GGT); GLP-1r: receptor of Glucagon Like Peptide-1; HCC: hepatocellular carcinoma; HCV: hepatitis C virus; HDL: High Density Lipoprotein (HDL); HOMA: homeostasis model of assessment; NAFL: non-alcoholic fatty liver; NAFLD: non-alcoholic fatty liver disease; NASH: non-alcoholic steato-hepatitis; PLS-DA: partial least square discriminantt analysis; ROS: Reactive Oxygen Species; TCA: tricarboxylic acid; VIP: variable importance in projection.

ABSTRACT

Introduction: Non-Alcoholic Fatty Liver Disease encompasses a spectrum of diseases ranging from simple steatosis to steatohepatitis (or NASH), up to cirrhosis and hepatocellular carcinoma(HCC). The challenge is to recognize the more severe and/or progressive pathology. A reliable non-invasive method does not exist. Untargeted metabolomics is a novel method to discover biomarkers and give insights on diseases pathophysiology. **Objectives:** we applied metabolomics to understand if simple steatosis, steatohepatitis and cirrhosis in NAFLD patients have peculiar metabolites profiles that can differentiate them among each-others and from controls. **Methods**: Metabolomics signatures were obtained from 307 subjects from two separated enrollments. The first collected samples from 69 controls and 144 patients (78 steatosis, 23 NASH, 15 NASH-cirrhosis, 8 HCV-cirrhosis, 20 cryptogenic cirrhosis). The second, used as validation-set, enrolled 44 controls and 50 patients (34 steatosis, 10 NASH and 6 NASH-cirrhosis). The "Partial-Least-Square Discriminant-Analysis" (PLS-DA) was used to reveal class separation in metabolomics profiles between patients and controls and among each class of patients, and to reveal the metabolites contributing to class differentiation. **Results:** Several metabolites were selected as relevant, in particular: Glycocholic acid, Taurocholic acid, Phenylalanine, branched-chain amino acids increased at the increase of the severity of the disease from steatosis to NASH, NASH-cirrhosis, while glutathione decreased ($p<0.001$ for each). Moreover, an ensemble machine learning (EML) model was built using 10 different classification models. EML showed accuracy>80% in NAFLD evolution steps prediction. **Conclusions**: Metabolomics profiles of NAFLD patients could be a useful tool to non-invasively diagnose NAFLD and discriminate among the various stages of the disease, giving insights into its pathophysiology.

INTRODUCTION

Non-Alcoholic Fatty Liver Disease (NAFLD) is becoming an increasingly frequent finding in Hepatology ambulatories and clinics in the last decade. This condition is defined by the presence of any significant (>5% of hepatocytes) amount of fat accumulation in the liver in the absence of an "unsafe" quantity of alcohol consumption and any other cause of liver diseases(1). This term includes at least two different clinical entities: a form that represents only the accumulation of fat in the liver (also named steatosis or Non-Alcoholic Fatty Liver - NAFL), and the Non-Alcoholic Steato-Hepatitis (NASH) which is characterized by steatosis along with necroinflammation and fibrosis. This latter entity is considered a "progressive" form that has histological features that make it hardly distinguishable from alcoholic liver disease. Accordingly, liver histology is characterized by the presence of hepatocytes ballooning, lobular inflammation, perisinusoidal and perivenular fibrosis(2, 3). Even if these two entities always go under the same definition of NAFLD, they don't share the same natural history: the first being a “benign” presentation with no (or very rare) progression, the second being responsible of liver cirrhosis, Hepatocellular Carcinoma (HCC) and liver-related deaths(4). Nevertheless, these two conditions seem to share the same risk factors(5). NAFLD is a unique "challenge" for the hepatologists who are called to discriminate between patients with a non-progressive disease (namely “good storers”) and potentially progressive ones (“bad storers”)(6). Moreover, due to the to changes in dietary habits and increased sedentary lifestyle, has seen a worldwide increment in the last years, making it one of the most frequent liver diseases in the world(7). It is generally considered a “benign disease” with low rates of progression to fibrosis, cirrhosis and HCC(4). Nevertheless, due to the high number of affected patients, the prevalence of related cirrhosis increased overtime, and actually it represents the third cause of liver transplantation in the USA(8). Finally, even if the incidence of HCC in NAFLD patients is lower than that in HCV/HBV cirrhotic patients, the absolute burden of NASH-related HCC is higher, due to the higher number of patients with NAFLD in respect to HCV infected ones(9). It is very likely that the importance of this disease will continue to increase in the future, when the new therapies and prevention programs for hepatitis C and B are further reducing the size of viral infections of the liver. For these reasons, it is very important to recognize the mechanisms underlying its onset and progression in the liver. Even if many insights on this topic were made in the last years, various aspects of the pathophysiological mechanisms underlying this disease remain to be explored.

In this setting, it would be very useful to individuate one, or more, specific biomarkers, to differentiate NAFLD patients from general population and, between NAFLD presentations, simple steatosis (NAFL) from steatohepatitis (NASH) and, NASH related cirrhosis (NASH-Cirrhosis).

Metabolomics technique has the advantage to evaluate more accurately the "phenotype" of a disease, in respect to genes, transcripts and proteins, which very likely undergo to epigenetic, transcriptional and pre- and post-translational modifications (10).

In the field of NAFLD some studies on humans, have recently sketched partial metabolomics profiles of such disease mostly on lipidomics, or targeted metabolomics profiles(11-16). For this reason, it is of high interest to further evaluate the plasma metabolomics profile of a cohort of biopsy proven NAFLD patients in order to find one, or more, specific biomarkers capable to differentiate them from healthy controls and, between NAFLD patients, to discriminate progressive forms (NASH), and eventually NAFLD-cirrhosis, from non-progressive liver steatosis (NAFL), with the mean of untargeted metabolomics. Finally, in this study we have compared the untargeted metabolomics profiles of three different kind of cirrhosis (cryptogenic, NASH-related and HCV-related), in order to find any measurable difference in the metabolomic profile of such diseases.

Materials and Methods

Population and study design

Two separate enrollments were performed, one to build the classification models and one to test them. The first was a cohort of biopsy proven NAFLD patients and sex- and age-matched controls. One hundred and forty-four patients, and 69 controls, were enrolled and stratified for age, sex and BMI. NAFLD patients were divided on the basis of the histological presence of steatosis or steatohepatitis diagnosed by the Kleiner and Brunt criteria(3, 17). Patients were divided in three groups: Non-alcoholic fatty liver (NAFL, or Simple steatosis) n: 78; Non-alcoholic Steatohepatitis (NASH) n:23; Non-alcoholic steatohepatitis associated Cirrhosis (NASH-Cirrhosis) n:43. Controls were enrolled from a cohort of age and sex-matched healthy subjects (without any evidence of Metabolic Syndrome, diabetes and/or NAFLD) afferent to the transfusion center of the University of Salerno. Among the NASH-Cirrhosis patients, 8 had advanced hepatic disease related to hepatitis C virus (HCV) and 20 had "cryptogenic" cirrhosis (defined as patients without any medical history of the NAFLD associated presentations such as Metabolic Syndrome, Type 2 diabetes mellitus and

obesity)(18). These patients were also enrolled and used to compare the several cirrhosis types. Second cohort was composed from 94 subjects (44 CTRL, 34 NAFL, 10 NASH, 6 NASH-cirrhosis). This second, independent enrollment followed the same criteria and inclusion/exclusion criteria of the first enrollment. Supplementary Figure S1 reports a schematic representation of the study flow-chart. Data analyst researchers were single blinded about the NAFLD related diagnosis of the second enrollment subjects.

Of every patient were recorded: clinical history with alcohol consumption and smoking habits registration, physical examination with waist circumference and body mass index (BMI) evaluation, blood glucose and insulin, HOMA score, total and fractioned cholesterol, triglycerides, alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma-glutamyl transpeptidase (GGT), ferritin, blood count, metabolic syndrome evaluation by NCEP-ATPIII criteria. Moreover, in patients, abdomen ultrasonography with the evaluation of the bright liver echo pattern presence and grade(19) and liver Stiffness measurement with transient elastography (Echosense Fibroscan® device, model 502; EchoSense, Paris, France; equipped with an M probe) by a skilled hepatologist (MP and MM) were performed. Liver tissue samples were collected by performing a hepatic percutaneous biopsy with Surecut 17G needles, via the intercostal route, using an echo-guided or echo-assisted method. Liver specimens were used for histological examination if they were at least 1.5-cm long and contained >5 portal spaces. Biopsies will be evaluated with the Kleiner score(17) for necroinflammation grading and fibrosis staging, and by the Brunt score(2) for the presence and extent of steatosis by a skilled pathologist. Each patient was included in the study after giving an informed consent. Patients with clinical/ultrasonographical signs of cirrhosis (esophageal varices, spleen enlargement, low platelet count, caudate lobe hypertrophy at ultrasound examination) did not undergo to liver biopsy and were classified on the basis of the abovementioned diagnostic criteria in NASH-associated, cryptogenic or, when HCV-Ab positive, in HCV-related cirrhosis. The study was approved by the local ethical committee (CEI Campania Sud IRB n.8/2018).

Samples collection

Human tissue collection strictly adhered to the guidelines outlined in the Declaration of Helsinki IV edition(20). Blood samples were collected at enrollment from controls and at enrollment and three months after the treatment for cases, using a BD vacutainer (Becton Dickinson, Oxfordshire, UK) blood collection red

tube (with no additives). After centrifugation, the sample was immediately frozen to -80 °C until the time of analysis. All patients were asked to respect a 12-hour fast before blood collection.

Metabolomics analysis

Metabolite extraction and derivatization

Metabolome extraction, purification and derivatization was carried out with the MetaboPrep GC kit (Theoreo srl, Montecorvino Pugliano [SA], Italy) according to the manufacturer's instructions.

GC-MS analysis

GC-MS analysis was conducted according to Troisi et al.. Briefly, two µL samples of the derivatized solution were injected into the GC-MS system (GC-2010 Plus gas chromatograph coupled to a 2010 Plus single quadrupole mass spectrometer; Shimadzu Corp., Kyoto, Japan). Chromatographic separation was achieved with a 30 m 0.25 mm CP-Sil 8 CB fused silica capillary GC column with 1.00 µm film thickness (Agilent, J&W Scientific, Folsom, CA, USA), with helium as carrier gas. Untargeted metabolites were identified by comparing the mass spectrum of each chromatographic peak with the NIST library collection (NIST, Gaithersburg, MD, USA) (minimum overlap 85%), also using the Kovats' index (max tolerance 10%).. Metabolites emerged as relevant (see below) were further structural confirmed by means of analytical external standard according to the level 1 MSI (21).

Statistical analysis

Anthropometric parameters

Data are reported as mean±standard deviation for continuous variables and number (percentage) for categorical variables. Statistical analysis was performed using Statistica software (StatSoft, Oklahoma, USA) and Minitab (Minitab Inc, Pennsylvania, USA). Normal distribution of data was verified using the Shapiro-Wilks test. Since the data were normally distributed, we used one-way ANOVA with the Tukey post hoc test for inter-group comparisons. The alpha (α) value was set to 0.05. Pearson's chi-squared test was used to determine differences among groups for the categorical variables.

Classification models building

The chromatographic data were tabulated with one sample per row and one variable (metabolite) per column. Data pre-treatment consisted of normalizing each metabolite peak area to that of the internal standard followed by generalized log transformation and data scaling by autoscaling (mean-centered and divided by standard deviation of each variable). Three different ensemble machine learning (EML) models, based on a voting scheme statistically weighted by the individual classification accuracy of ten different classification models (decision tree [DT], partial least square discriminant analysis [PLS-DA], naïve Bayes [NB], random forest [RF], k-nearest neighbor [k-NN], artificial neuronal network [a-NN], support-vector machines [SVM], linear discriminant analysis [LDA], logistic regression [LR] and deep learning [DL]) was built according to Troisi et al.(22-24), using R and RapidMiner Studio version 9.5 (RapidMiner, Boston, MA, USA). Samples collected in the first enrollment were used for the purpose of training the ten individual models and using those results to generate the EML model. Application of the EML model using the metabolomic profile of blood samples from the subjects of the second enrollment constitutes the independent validation of the proposed models.

A genetic algorithm (GA) was also built to selects a subset of relevant metabolites. This was selected as the subset able to train a classification model with classification performance close to the one archivable from a model built on the whole metabolites set. GA mimics Darwinian forces of natural selection to optimize values of a function(25). An initial set of potential solutions were built, and their corresponding "fitness" values was calculated. Using the evolutive analogy, each solution represents an individual and the whole set were considered as a population.

The individuals with the best fitness values were randomly combined to produce offspring which will structure the next population. To do so, individuals were selected and undergo cross-over (mimicking genetic reproduction) and were also subjected to random mutations. This process was repeated several times producing many generations that created even better solutions.

For metabolites selection, the individuals were subsets of predictors that are encoded as binary; metabolites were either included or not in the subset. The fitness values were the measure of model performance (classification accuracy). The genetic algorithm was built using RapidMiner studio version 9.5 (RapidMiner, Boston, MA). Model's details were reported in Troisi et al.(22) The metabolites selected by means of GA were used to build ah heatmap representation of the metabolites' concentration change trough the disease evolution.

The metabolic pathway was constructed using MetScape application(26) of the software Cytoscape(27).

NAFLD evolution classification model building

To build the EML models able to discriminate among the several NAFLD stage we first divide the multi class problem in several dichotomic problems. We built 3 different EML models, one able to discriminate healthy subject from NAFLD affected ones (model 1), one to discriminate subjects with simple steatosis (NAFL) from patients with steatohepatitis or cirrhosis (model 2), one to discriminate NASH patients NASH-cirrhosis (model 3).

NAFLD Metabolomic Score (NAFLD-EML-Score)

For each classification model the cross-validation accuracy was evaluated. For each sample and for each classification the classification confidence was also evaluated. From these parameters a model score was evaluated multiplying the classification accuracy and the classification confidence. For the subjects classified as NAFLD (or as a next-stage disease), these scores were considered as is, while for each CTRL (or for subject considered in the less evolved NAFLD stage), the scores were multiplied by -1. A NAFLD-EML-score was calculated for each sample summing all the classification scores.

The area under receiver operating characteristic (ROC) curves was calculated to evaluate the ability of the NAFLD-EML-score to predict the NAFLD presence or stage. Cutoff points were proposed after calculation for each NAFLD-EML-score (from minimum to maximum estimated values) the Youden's Index (sensitivity+specificity-1). DeLong et al.(28) non-parametric approach was used to compare the areas under the ROC curves.

Classification performance evaluation

Subjects from second enrollment were tested by means of the dichotomic models to predict the disease stage. The classification performances of the proposed dichotomic EML models were investigated in terms of

Sensitivity (SN), Specificity (SP), Positive Predictive Value (PPV), Negative Predictive Value (NPV), Positive Likelihood Ratio (PLR), Negative Likelihood Ratio (NLR) and Accuracy (A). Cumulative accuracies (A_c) were also evaluated as the global accuracy considering the accuracies of the previous steps. Table S1 (Supplementary information) reports the formulas described above.

RESULTS

The demographic and clinical–biochemical characteristics of NAFLD patients are reported in **Table 1.**

Table 1. Demographical and laboratory characteristics of the study population by liver Histology presentation (HCV and cryptogenic cirrhosis were not presented separately).

	Overall	**NAFL**	p NAFL vs NASH	**NASH**	p NASH vs NASH-Cirrhosis	**NASH-Cirrhosis**
Total Number	194	112	-	33	-	21
Age (years±SD)	60.85±16.0	57.7±17.8	0.95	57.5±14.7	**0.01**	65.8±10.1
Male Sex (%)	111 (57.2%)	66 (58.9%)	0.67	18 (54.5%)	0.66	11 (52.4%)
BMI (Kg/m^2)	29.6±4.3	29.6±4.0	0.17	30.8±4.5	0.8	31.1±5.1
AST (U/L)	37.12±28.8	33.0±32.7	0.32	43.5±56.1	0.55	40.6±19.8
ALT (U/L)	44.9±45.86	43.5±56.1	0.25	53.6±39.9	0.62	49.2±29.7
Diabetes (%)	50 (25.8%)	27 (24.1%)	**0.008**	16 (48.8%)	0.148	12 (57.1%)
Glycaemia	111.5±38.2	108.0±39.9	0.56	111.4±24.5	0.44	118.0±36.4
Serum Insulin	19.8±15.6	15.3±12.3	**0.005**	28.6±18.8	0.85	30.4±20.1
HOMA	5.3±5.1	3.9±3.4	**0.001**	8.0±7.3	0.84	7.3±6.7
Hypertension (%)	75 (38.6%)	58 (51.8%)	**0.011**	25 (75.8%)	0.964	21 (58.3%)
HDL Cholesterol (mg/dL)	44.7±12.9	45.1±12.7	0.88	45.4±11.7	0.39	45.4±11.7
Total Cholesterol (mg/dL)	177.8±42.9	183.2±39.9	0.16	194.5±40.7	**0.015**	166.4±39.3
Triglycerides (mg/dL)	140.2±72.4	144.5±75.3	0.20	163.1±66.7	0.22	137.3±79.4
Metabolic Syndrome (%)	103 (53.1%)	62 (55.3%)	**0.00001**	29 (87.9%)	0.81	17 (80.9%)
GGT (U/L)	71.2±81.6	52.1±72.1	0.10	77.5.1±82.4	0.33	103.2±108.3
Ferritin (µg/L)	178.7±208.2	221.9±214.0	0.53	185.0±225.5	0.57	137.3±235.6
Liver Stiffness	-	5.73±2.8	**0.006**	16.1±15.4	-	-

No differences in term of age, sex ratio, body mass index, serum cholesterol concentration (both HDL an LDL), triglycerides, ferritin and GGT resulted among the patients. Blood insulin concentration, homeostatic model assessment (HOMA), prevalence of hypertension, type 2 diabetes mellitus, metabolic syndrome and mean liver stiffness were significantly higher in NASH compared with NAFL patients.

Within one TIC chromatogram, over 250 signals were detected in a single specimen and some of these peaks were not investigated further as they were not consistently found in other sets of samples, too low in concentration, or of poor spectral quality to be confirmed as metabolites. A total of 228 endogenous metabolites were detected consistently. As shown in **Figure 1**, the PLS-DA score plots clearly differentiated controls and NAFLD patients (Figure 1A) and among NAFLD: NAFL, NASH (Figure 1B) and NASH-Cirrhosis ones (Figure 1C). The 15 highest scoring VIP variables (VIP score > 1.5) identified by PLS-DA are shown in Figure 1D as box and whisker plot.

The metabolite concentration variations induced by the liver disease progression allowed to divide the metabolites into three classes: those with lower concentrations in the controls and which increased in disease progression (isocitric acid, isoleucine and a not structural identified metabolite), those which have higher concentrations in the controls and whose concentration decrease in disease progression (xanthine, glutathione and glycolic acid) and those which concentration distribution is not strictly related to the disease progression (valine, asparagine, 4-deoxy erythronic acid, propanoic acid, palmitic acid, butanoic acid, stearic acid, phenylalanine and taurocholic acid).

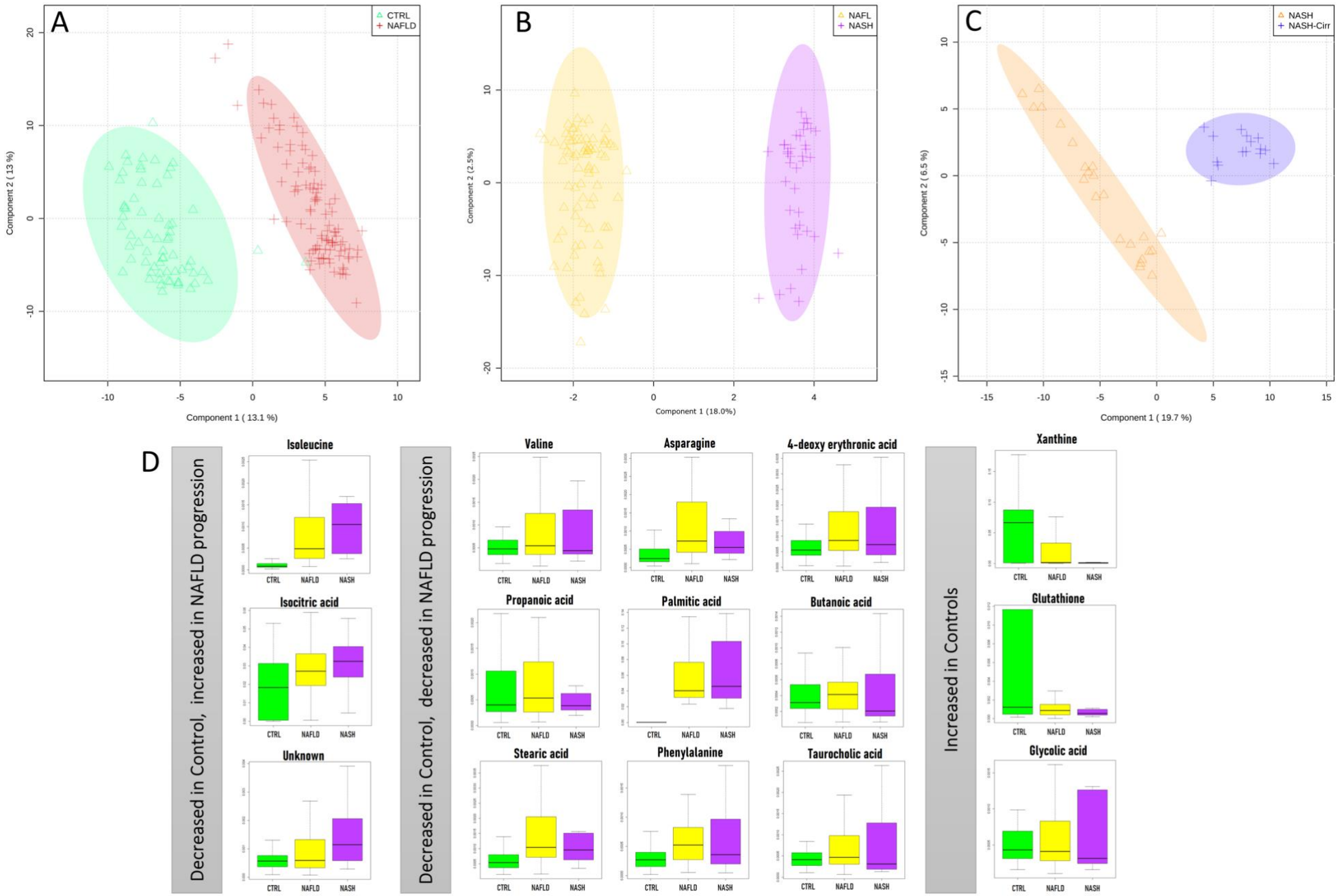
A
CTRL
NAFLD
Component 2 (13 %)
Component 1 (13.1 %)
B
NAFL
NASH
Component 2 (2.5%)
Component 1 (18.0%)
C
NASH
NASH-Cirr
Component 2 (6.5 %)
Component 1 (19.7 %)
D
Decreased in Control, increased in NAFLD progression
Isoleucine
Isocitric acid
Unknown
Decreased in Control, decreased in NAFLD progression
Valine
Asparagine
4-deoxy erythronic acid
Propanoic acid
Palmitic acid
Butanoic acid
Stearic acid
Phenylalanine
Taurocholic acid
Increased in Controls
Xanthine
Glutathione
Glycolic acid
CTRL
NAFLD
NASH

Figure 1.

Figure 1: (A) Partial least square discriminant analysis (PLS-DA) models to discriminate Controls (CTRL, triangles) and Non-alcoholic Fatty Liver Disease (NAFLD, crosses); (B) Simple Steatosis (NAFL, triangles) and Non-alcoholic Steatohepatitis (NASH, crosses); (C) NASH (triangles) and NASH related cirrhosis (crosses). The explained variance of each component is shown in parentheses on the corresponding axis. (C) Box and Whisker plot representation of the 15 top-scoring VIP metabolites (VIP-score ≥1.5)

Figure 2 reports a model similar to the one reported in Figure 1, in which cirrhosis patients were stratified according to their cirrhosis etiology as cirrhosis due to hepatic C virus (HCV) infection, cirrhosis on NASH and cryptogenic cirrhosis. Fifteen metabolites showed a VIP score higher that 1.5 in this PLS-DA model (Figure 2B): Galactose, Uric acid, Ribitol, Glyceric acid, Butanoic acid, Histidine, Phenylalanine, Stearic acid, Threonine, Palmitic acid which concentration was higher in NASH related cirrhosis patients. N-acetyl glucosamine and isoleucine with a higher concentration in HCV related cirrhosis and Eicosanoic, 5-hydroxyindolacetic and Aspartic acid that showed a high concentration in cryptogenic cirrhosis.

Figure 2.

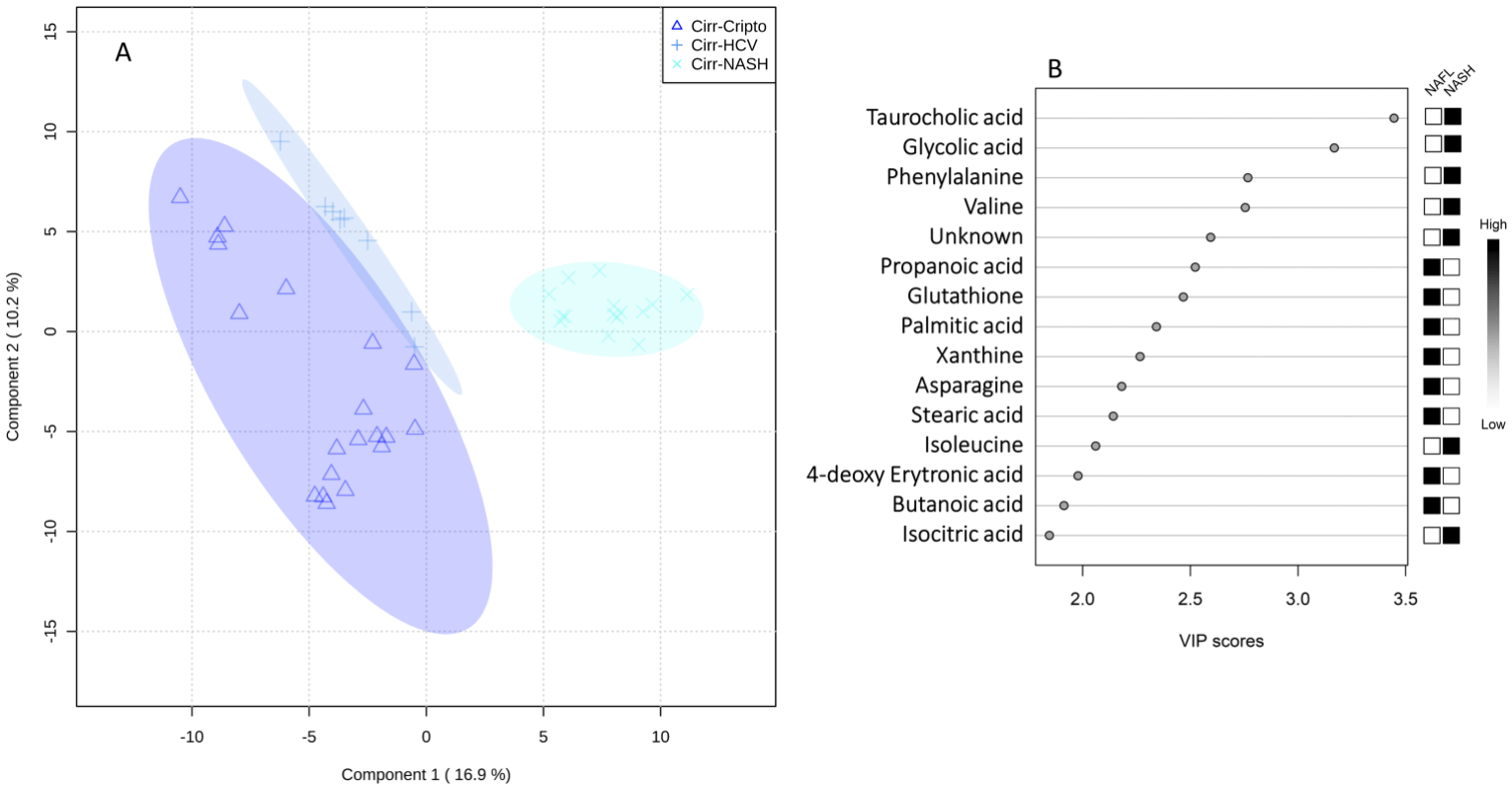

Figure 2: (A) Two dimensional representations of the Partial least square discriminant analysis (PLS-DA) models to discriminate the three forms of Cirrhosis: Cryptogenic (triangles), HCV-related, (crosses) and NASH-related (X). The explained variance of each component is shown in parentheses on the corresponding axis. (B) The 15 top-scoring VIP metabolites (VIP-score≥1.5) are shown. The boxes on the right indicate the relative amount of the corresponding metabolite in each group under study.

Metabolites selected as the most relevant by means of GA, were used to build the heatmap reported in **figure 3**. Cluster analysis performed on these data highlight the presence of three groups of metabolites: one which concentration constantly decrease from controls versus disease progression (fig. 3A), one which concentration constantly increase (fig. 3B) and one which concentration follow a more complex route (decreased in early disease stage and increased in advanced ones compared to controls), (see fig. 3C).

Figure 3.

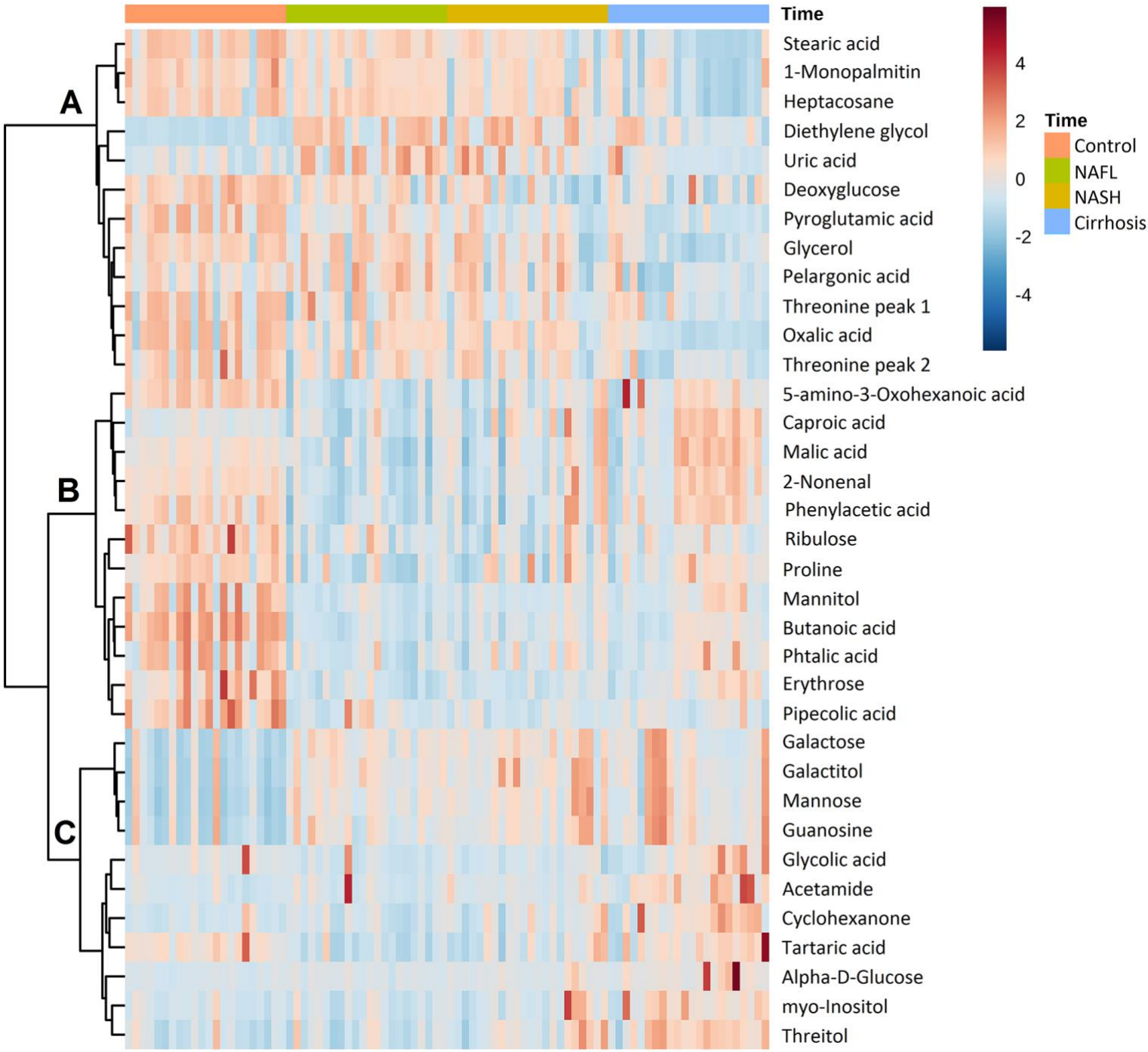

Figure 3. Heatmap representation of the genetic algorithm selected metabolites trough the NAFLD evolution stages. Metabolites shown 3 different behaviors:
(A) especially composed of fatty acids related metabolites decreased from the healthy subjects to the NASH-Cirrhosis ones.
(B) accounted from amino acid and amino acid related metabolites and gut microbiota related metabolites decrease in the first steps of the NAFLD (NAFL and NASH) while increase in the lasts (NASH-related Cirrhosis).
(C) composed especially from sugars constantly increase during the NAFLD evolution.

Selected metabolites ontology was reported in **table 2**.

Table 2		
Metabolite	Selection Criteria	HMDB ID
Isoleucine	VIP-score >1.5 NAFLD Vs CTRL	HMDB0000172
Isocitric acid		HMDB0000193
Valine		HMDB0000883
Asparagine		HMDB0000168
4-Deoxyerythronic acid		HMDB0000498
Propanoic acid		HMDB0000237
Palmitic acid		HMDB0000220
Butanoic acid		HMDB0000039
Stearic acid		HMDB0000827
Phenylalanine		HMDB0000159
Taurocholic acid		HMDB0000036
Xanthine		HMDB0000292
Glutathione		HMDB0000125
Glycolic acid		HMDB0000115
Galactose	VIP-score >1.5 Cirrhosis etiology	HMDB0000143
Acetyl glucosamine		HMDB0000803
Eicosanoic acid		HMDB0002212
Uric acid		HMDB0000289
Ribitol		HMDB0000508
Isoleucine		HMDB0000172
Glyceric acid		HMDB0000139
Butanoic acid		HMDB0000039
Histidine		HMDB0000177
Indolacetic acid		HMDB0000197
Stearic acid		HMDB0000827
Threonine		HMDB0000167
Aspartic acid		HMDB0000191

The metabolic pathway analysis of the metabolites selected (VIP-score > 1.5 in the PLS-DA model built on Control, NAFLD, NASH and NASH-Cirrhosis patient) is summarized in the metabolic systems map shown in **Figure 4.**

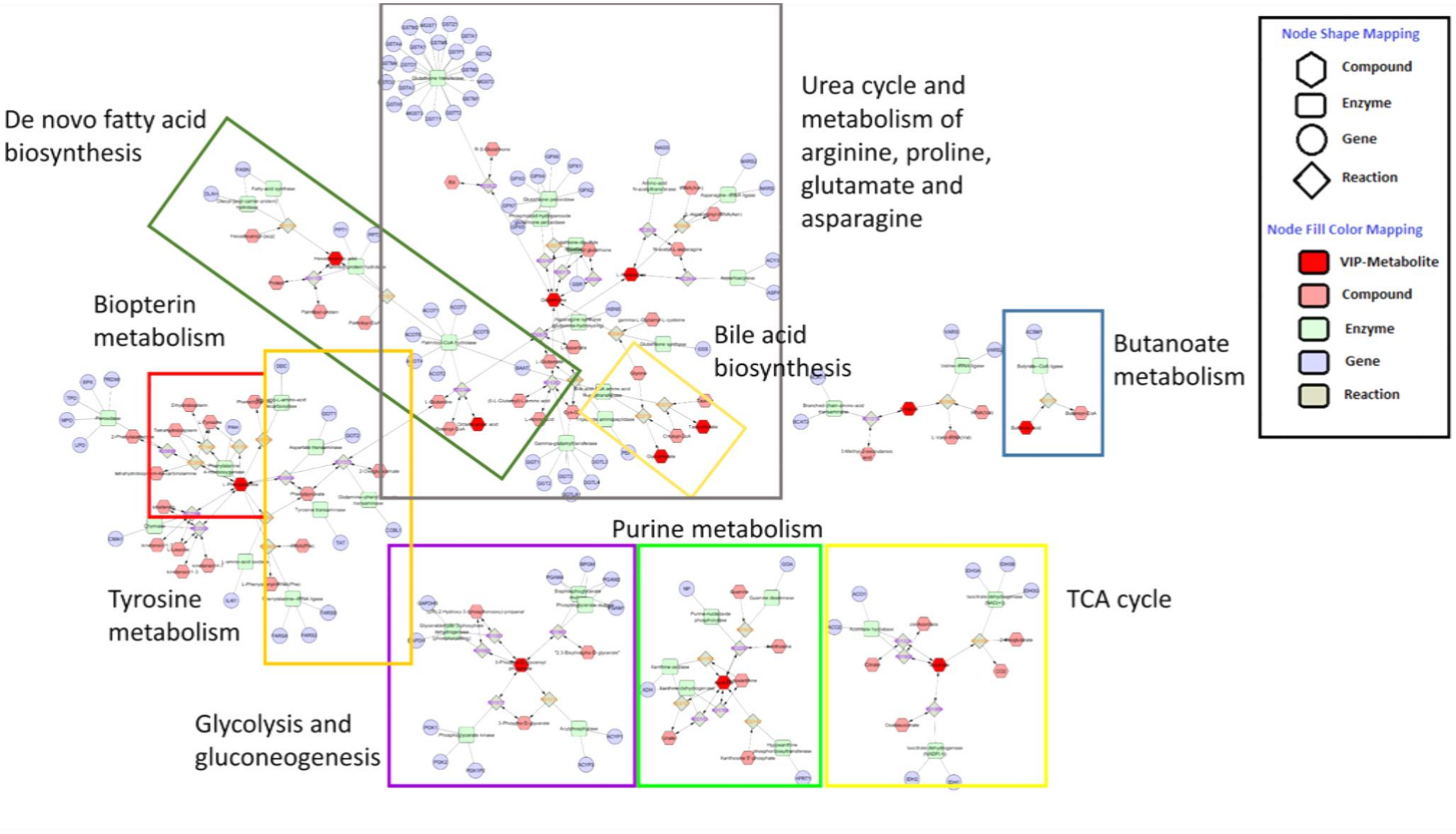
De novo fatty acid biosynthesis
Urea cycle and metabolism of arginine, proline, glutamate and asparagine
Biopterin metabolism
Bile acid biosynthesis
Butanoate metabolism
Purine metabolism
Tyrosine metabolism
TCA cycle
Glycolysis and gluconeogenesis
Node Shape Mapping
Compound
Enzyme
Gene
Reaction
Node Fill Color Mapping
VIP-Metabolite
Compound
Enzyme
Gene
Reaction

Figure 4. Metabolic systems map summarizing the shortest route that may explain the interactions among the 15 selected metabolites. There is a clear interplay of several pathways involving: De novo fatty acid biosynthesis; Urea cycle and metabolism of arginine, proline, glutamate, aspartate and asparagine; Biopterin metabolism; Bile acid biosynthesis; Butanoate metabolism Tyrosine metabolism; Glycolysis and gluconeogenesis; Purine metabolism; TCA cycle.

There is a definite interaction of several pathways involving de novo fatty acid biosynthesis; Urea cycle and metabolism of arginine, proline, glutamate, aspartate and asparagine; Biopterin metabolism; Bile acid biosynthesis; Butanoate metabolism Tyrosine metabolism; Glycolysis and gluconeogenesis; Purine metabolism; TCA cycle.

Classification performance evaluated on test set were summarized in **table 3**.

Table 3. Classification models performance. Values are reported as result ± standard error. Cumulative accuracies were reported in brackets.

	Sensitivity (%)	**Specificity (%)**	**PLR**	**NLR**	**PPV (%)**	**NPV (%)**	**Accuracy (%)**
Model 1 (Ctrl Vs NAFLD)	96.0±2.8	97.7±2.2	42.24	0.041	98.0±2.0	95.6±3.1	96.8±2.1 (---)
Model 2 (NAFL Vs NASH & NASH-Cirrhosis)	94.1±4.0	93.8.3±6.1	15.06	0.063	97.0±3.0	94.0±4.2	94.0±4.2 (91.0±2.1)
Model 3 (NASH Vs NASH-Cirrhosis)	80.0±12.6	83.3±15.2	4.80	0.240	88.9±10.5	71.4±17.1	81.3±12.2 (73.9±9.7)

PLR: positive likelihood ratio; NLR: negative likelihood ratio; PPV: positive prognostic value; NPV: negative prognostic value.

Blind analysis using the subjects of the second enrollment shown an accuracy for NAFLD identification (model 1) of 96.8% (91/94 correctly identified), 94.0% for NAFL identification (model 2, 47/50 correctly identified). Cumulative accuracy for model 2 was 91.0%. NASH identification (model 3) accuracy was 81.3 % (13/16 correctly identified) while cumulative accuracy was 73.9%.

DISCUSSION

Here we report the serum metabolomic analysis results to obtain a comprehensive view of changes in several metabolite and metabolic pathways in patients with NAFLD, in order to identify disease related patterns and to identify biochemical perturbations. The data revealed significant changes in certain key pathways,

specifically bile acids, lipid and amino acid metabolism. Metabolomic changes due to NAFLD were already reported both in adults(13) and in pediatric studies(29).

In our NAFLD metabolomic signature isoleucine, valine and asparagine have showed a great relevance. Amino acids involvement in NAFLD was already been reported (13, 29). The increase in essential amino acids suggests a higher rate of protein turnover. In addition, the incremental amino acids values in the disease progression suggests that changes in protein turnover may be a late event in the progression of steatosis to NASH, and may be modulated by other factors such as cytokines and inflammation, in addition to insulin resistance(30, 31). Moreover, branched-chain amino acids and, in particular, isoleucine have been associated with gluconeogenesis, insulin resistance and increased risk of type 2 diabetes development, as well as cardiometabolic risk(32-34). In our study we observed a significant increase in isoleucine in our patients according to the NAFLD presentation (from simple steatosis to cirrhosis), whereas valine and asparagine, even if increased in respect to healthy controls, showed an opposite trend, in the same way as it has been described by other authors(35).

Kalhan et al.(13) reported that the changes in non-essential amino acids, aspartate and glutamate, may be due to increased anaplerosis of amino acids into the TCA cycle, resulting in an increased cataplerosis to insure the required removal of the resulting carbon skeletons of these amino acids from the cycle. Moreover, it has been demonstrated in a murine model of glutathione deficiency, that aspartate and glutamate increase in response of GSH deficiency as a metabolic adaptation for the maintenance of the redox and metabolic homeostasis of the liver(36). In this way, the higher levels of asparagine in our NAFLD cohort compared to controls subjects, in addition to higher glutathione turnover, could also be due to increased transamination of amino acids being degraded in the liver and skeletal muscle. Moreover, during the evolution of the inflammation/fibrosis process, intermediate metabolites of the tricarboxylic acid cycle, branched chain amino acids and fatty acids are accumulated in the sera of affected individuals. These results were recently confirmed also on a murine model of fibrosis induced by CCl_4 (37). BCAA were also correlated with insulin resistance and cardiometabolic risk in NAFLD patients(34). Finally, Isoleucine and citric acid were also reported as potential biomarker for NAFLD in an untargeted metabolomic study on an Egyptian cohort (38).

Our results showed a reduction of xanthine amount in patients' serums during the NAFLD progression. Xanthine oxidation is a rate-limiting reaction in the uric acid production and is already been reported as a

candidate link between NAFLD and hyperuricemia(39). Xanthine is physiologically oxidated to urate and hydrogen peroxide (40), allowing the production of Reactive Oxygen Species (ROS). Low xanthine serum level can be reasonably associated with high ROS levels (39). This result is also in line with the coexisting reduction in the glutathione levels in the disease progression.

Glycolic acid showed a decrease trend during the disease progression. It has been reported to be higher in small intestine bacterial overgrowth children (29). At the same time our results showed a significant role (even if the serum concentration is not directly linked to the disease progression) of two important short chain fatty acids that are mainly produced from the gut microbiota (propanoic and butyric acid). Short chain fatty acids, and in particular propionate and butyrate, have been demonstrated to exert an important protective role in NAFLD onset and progression, through the modulation of the inflammation and the insulin resistance via the activation of the receptor of Glucagon Like Peptide-1 (GLP1r) in the liver (41). Several studies in which the gut microbiota was manipulated, and observational studies in patients with NAFLD, have provided evidence that dysbiosis contributes to the pathogenesis of NAFLD (41, 42). Dysbiosis increases gut permeability to bacterial products and increases hepatic exposure to injurious substances that increase hepatic inflammation and fibrosis (43). Taurocholic and fatty acids as palmitic or stearic were also reported altered in NAFLD (31, 44, 45) coherently with our finding. The increase in oxidative stress in the progression of NAFLD is demonstrated by the reduction of serum Glutathione levels. This evidence has been repeatedly reported in the literature (34, 46).

Nowadays to differentiate the presence of a simple, non-evolutive, liver steatosis (NAFL) from a potentially worsening steatohepatitis still represents a diagnostic issue. In fact, the definitive differential diagnosis between these two entities still relay on a such invasive technique as liver biopsy, which, in the clinical practice, is difficult to propose to a large number of patients, of whom only a minority is potentially affected by the evolutionary form of the disease (NASH) (7). To address this issue, various clinical scores have been proposed for the use in NAFLD patients, such as “Fatty Liver Index”, “NAFLD Fibrosis Score”, and analogues(47, 48). Even if an approach of this type has the advantage of being a non-invasive method for discriminating between potentially benign and evolutionary diseases, it has various limitations, based primarily on the lack of validation on large cohorts and different populations, and on the scarce power to evaluate “intermediate presentations”. These methods have, in fact, a large “grey zone” in which they fail to address the real risk of the patients. Other non-invasive methods relay on imaging techniques (such as Fibroscan, Continuous

Attenuation Parameter – CAP, Ultrasonography and Magnetic Resonance) that can, at their best, quantify the liver fat content rather than differentiate between simple steatosis and steatohepatitis. In this way, other more precise non-invasive approaches to differentiate the various clinical expressions of NAFLD are desirable. Our results indicated the serum metabolomics signature can aid the NAFLD identification and its evolution stage. This approach showed, even on a small validation size cohort, good performance of each classification models, although the cumulative accuracy rapidly decreases with the increase of classification steps. This is a weakness inherent in our approach, but it can be minimized with larger validation chords.

In summary, our analysis of the metabolomic profiles of NAFL, NASH and NASH-cirrhosis patients showed a significant separation between these groups at the PLS-DA analysis, suggesting that metabolomic profiles, with the above mentioned metabolites identified as "VIP", can differentiate the patients on the basis of the severity of their liver disease (Figure 3). Finally, the analysis of metabolomic profiles of cirrhotic patients of different etiologies (HCV-associated, cryptogenic and NASH-associated cirrhosis) also showed a peculiar and significantly different metabolomic profile of 15 VIP metabolites, suggesting that these conditions can be discriminated by performing a targeted metabolomic profile on these patients. This is of particular interest for the differential diagnosis of cryptogenic- and NASH-related cirrhosis that often represents a challenge in the clinical practice.

A limitation of our study is the not prospective study design. The disease progression was evaluated among different subject at different disease stages. To follow the same patients during the disease progression (or not progression) can highlight clearly the mechanism involved in these events. NAFLD disease progression is a very slow process and can be completed in several decades. So, this kind of study will take a very long time

CONCLUSION

Our results indicated, even if in a preliminary way, that the untargeted metabolomic can represent an appropriate and useful tool for the diagnosis and the assessment of the different forms of NAFLD and to understand the mechanisms involved in its progression.

References:

1. Angulo P. Nonalcoholic fatty liver disease. N Engl J Med. 2002;346(16):1221-31.

2. Brunt EM, Janney CG, Di Bisceglie AM, Neuschwander-Tetri BA, Bacon BR. Nonalcoholic steatohepatitis: a proposal for grading and staging the histological lesions. Am J Gastroenterol. 1999;94(9):2467-74.

3. Brunt EM. Pathology of nonalcoholic fatty liver disease. Nat Rev Gastroenterol Hepatol. 2010;7(4):195-203.

4. Matteoni CA, Younossi ZM, Gramlich T, Boparai N, Liu YC, McCullough AJ. Nonalcoholic fatty liver disease: a spectrum of clinical and pathological severity. Gastroenterology. 1999;116(6):1413-9.

5. Younossi Z, Anstee QM, Marietti M, Hardy T, Henry L, Eslam M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. Nat Rev Gastroenterol Hepatol. 2018;15(1):11-20.

6. Marra F, Lotersztajn S. Pathophysiology of NASH: perspectives for a targeted treatment. Curr Pharm Des. 2013;19(29):5250-69.

7. Chalasani N, Younossi Z, Lavine JE, Diehl AM, Brunt EM, Cusi K, et al. The diagnosis and management of non-alcoholic fatty liver disease: practice Guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association. Hepatology. 2012;55(6):2005-23.

8. White DL, Kanwal F, El-Serag HB. Association between nonalcoholic fatty liver disease and risk for hepatocellular cancer, based on systematic review. Clin Gastroenterol Hepatol. 2012;10(12):1342-59 e2.

9. Masarone M, Federico A, Abenavoli L, Loguercio C, Persico M. Non alcoholic fatty liver: epidemiology and natural history. Rev Recent Clin Trials. 2014;9(3):126-33.

10. Patti GJ, Yanes O, Siuzdak G. Innovation: Metabolomics: the apogee of the omics trilogy. Nat Rev Mol Cell Biol. 2012;13(4):263-9.

11. Bril F, Millan L, Kalavalapalli S, McPhaul MJ, Caulfield MP, Martinez-Arranz I, et al. Use of a metabolomic approach to non-invasively diagnose non-alcoholic fatty liver disease in patients with type 2 diabetes mellitus. Diabetes Obes Metab. 2018;20(7):1702-9.

12. Cano A, Alonso C. Deciphering non-alcoholic fatty liver disease through metabolomics. Biochem Soc Trans. 2014;42(5):1447-52.

13. Kalhan SC, Guo L, Edmison J, Dasarathy S, McCullough AJ, Hanson RW, et al. Plasma metabolomic profile in nonalcoholic fatty liver disease. Metabolism. 2011;60(3):404-13.

14. Mayo R, Crespo J, Martinez-Arranz I, Banales JM, Arias M, Minchole I, et al. Metabolomic-based noninvasive serum test to diagnose nonalcoholic steatohepatitis: Results from discovery and validation cohorts. Hepatol Commun. 2018;2(7):807-20.

15. Romero-Ibarguengoitia ME, Vadillo-Ortega F, Caballero AE, Ibarra-Gonzalez I, Herrera-Rosas A, Serratos-Canales MF, et al. Family history and obesity in youth, their effect on acylcarnitine/aminoacids metabolomics and non-alcoholic fatty liver disease (NAFLD). Structural equation modeling approach. PLoS One. 2018;13(2):e0193138.

16. Voiculescu M, Nanau RM, Neuman MG. Non-invasive biomarkers in non-alcoholic steatohepatitis-induced hepatocellular carcinoma. J Gastrointestin Liver Dis. 2014;23(4):425-9.

17. Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology. 2005;41(6):1313-21.

18. Thuluvath PJ, Kantsevoy S, Thuluvath AJ, Savva Y. Is cryptogenic cirrhosis different from NASH cirrhosis? J Hepatol. 2018;68(3):519-25.

19. Palmentieri B, de Sio I, La Mura V, Masarone M, Vecchione R, Bruno S, et al. The role of bright liver echo pattern on ultrasound B-mode examination in the diagnosis of liver steatosis. Dig Liver Dis. 2006;38(7):485-9.

20. World Medical A. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA. 2013;310(20):2191-4.

21. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics. 2007;3(3):211-21.

22. Troisi J, Landolfi A, Sarno L, Richards S, Symes S, Adair D, et al. A metabolomics-based approach for non-invasive screening of fetal central nervous system anomalies. Metabolomics. 2018;14(6):77.

23. Troisi J, Sarno L, Martinelli P, Di Carlo C, Landolfi A, Scala G, et al. A metabolomics-based approach for non-invasive diagnosis of chromosomal anomalies. Metabolomics. 2017;13(11).

24. Troisi J, Sarno L, Landolfi A, Scala G, Martinelli P, Venturella R, et al. Metabolomic Signature of Endometrial Cancer. J Proteome Res. 2018;17(2):804-12.

25. Mitchell M. An introduction to genetic algorithms. Cambridge, Mass.: MIT Press; 1996. viii, 205 p. p.

26. Karnovsky A, Weymouth T, Hull T, Tarcea VG, Scardoni G, Laudanna C, et al. Metscape 2 bioinformatics tool for the analysis and visualization of metabolomics and gene expression data. Bioinformatics. 2012;28(3):373-80.

27. Nishida K, Ono K, Kanaya S, Takahashi K. KEGGscape: a Cytoscape app for pathway data integration. F1000Res. 2014;3:144.

28. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44(3):837-45.

29. Troisi J, Pierri L, Landolfi A, Marciano F, Bisogno A, Belmonte F, et al. Urinary Metabolomics in Pediatric Obesity and NAFLD Identifies Metabolic Pathways/Metabolites Related to Dietary Habits and Gut-Liver Axis Perturbations. Nutrients. 2017;9(5).

30. Gougeon R, Morais JA, Chevalier S, Pereira S, Lamarche M, Marliss EB. Determinants of whole-body protein metabolism in subjects with and without type 2 diabetes. Diabetes Care. 2008;31(1):128-33.

31. Kalhan SC. Fatty acids, insulin resistance, and protein metabolism. J Clin Endocrinol Metab. 2009;94(8):2725-7.

32. Zhao X, Han Q, Liu Y, Sun C, Gang X, Wang G. The Relationship between Branched-Chain Amino Acid Related Metabolomic Signature and Insulin Resistance: A Systematic Review. J Diabetes Res. 2016;2016:2794591.

33. Asghari G, Farhadnejad H, Teymoori F, Mirmiran P, Tohidi M, Azizi F. High dietary intake of branched-chain amino acids is associated with an increased risk of insulin resistance in adults. J Diabetes. 2018;10(5):357-64.

34. Gaggini M, Carli F, Rosso C, Buzzigoli E, Marietti M, Della Latta V, et al. Altered amino acid concentrations in NAFLD: Impact of obesity and insulin resistance. Hepatology. 2018;67(1):145-58.

35. Lake AD, Novak P, Shipkova P, Aranibar N, Robertson DG, Reily MD, et al. Branched chain amino acid metabolism profiles in progressive human nonalcoholic fatty liver disease. Amino Acids. 2015;47(3):603-15.

36. Chen Y, Golla S, Garcia-Milian R, Thompson DC, Gonzalez FJ, Vasiliou V. Hepatic metabolic adaptation in a murine model of glutathione deficiency. Chem Biol Interact. 2019;303:1-6.

37. Chang H, Meng HY, Liu SM, Wang Y, Yang XX, Lu F, et al. Identification of key metabolic changes during liver fibrosis progression in rats using a urine and serum metabolomics approach. Sci Rep. 2017;7(1):11433.

38. Nezami Ranjbar MR, 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(6):e0127299.

39. Xu C, Wan X, Xu L, Weng H, Yan M, Miao M, et al. Xanthine oxidase in non-alcoholic fatty liver disease and hyperuricemia: One stone hits two birds. J Hepatol. 2015;62(6):1412-9.

40. Harrison R. Physiological roles of xanthine oxidoreductase. Drug Metab Rev. 2004;36(2):363-75.

41. Ding Y, Yanagi K, Cheng C, Alaniz RC, Lee K, Jayaraman A. Interactions between gut microbiota and non-alcoholic liver disease: The role of microbiota-derived metabolites. Pharmacol Res. 2019;141:521-9.

42. Leung C, Rivera L, Furness JB, Angus PW. The role of the gut microbiota in NAFLD. Nat Rev Gastroenterol Hepatol. 2016;13(7):412-25.

43. Ma J, Zhou Q, Li H. Gut Microbiota and Nonalcoholic Fatty Liver Disease: Insights on Mechanisms and Therapy. Nutrients. 2017;9(10).

44. Calvo N, Beltran-Debon R, Rodriguez-Gallego E, Hernandez-Aguilera A, Guirro M, Marine-Casado R, et al. Liver fat deposition and mitochondrial dysfunction in morbid obesity: An approach combining metabolomics with liver imaging and histology. World J Gastroenterol. 2015;21(24):7529-44.

45. Stiuso P, Scognamiglio I, Murolo M, Ferranti P, De Simone C, Rizzo MR, et al. Serum oxidative stress markers and lipidomic profile to detect NASH patients responsive to an antioxidant treatment: a pilot study. Oxid Med Cell Longev. 2014;2014:169216.

46. Liu W, Baker SS, Baker RD, Zhu L. Antioxidant Mechanisms in Nonalcoholic Fatty Liver Disease. Curr Drug Targets. 2015;16(12):1301-14.

47. Angulo P, Hui JM, Marchesini G, Bugianesi E, George J, Farrell GC, et al. The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. Hepatology. 2007;45(4):846-54.

48. Bedogni G, Bellentani S, Miglioli L, Masutti F, Passalacqua M, Castiglione A, et al. The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population. BMC Gastroenterol. 2006;6:33.