Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Brent Gawey Date

Metabolomic Analysis of Microbial Small Molecules in Non-Alcoholic Fatty Liver Disease

By

Brent Gawey
Master of Science

Clinical Research

_____ [Advisor's signature]
Dean Jones, PhD
Advisor

_____ [Advisor's signature]
Andrew Neish, MD
Advisor

_____ [Advisor's signature]
Matthew Ryan Smith, PhD
Advisor

_____ [Advisor's signature]
Thomas Ziegler, MD MS
Advisor

_____ [Member's signature]
Jeffrey Collins, MD MS
Committee Member

_____ [Member's signature]
Amita Manatunga, PhD
Committee Member

Accepted:

_____
Kimberly Jacob Arriola, PhD, MPH
Dean of the James T. Laney School of Graduate Studies

_____
Date

Metabolomic Analysis of Microbial Small Molecules in Non-Alcoholic Fatty Liver Disease

By

Brent Gawey
B.S., University of Georgia, 2017

Advisors:
Dean Jones, PhD
Andrew Neish, MD
Thomas Ziegler, MD MS

An abstract of
A thesis submitted to the Faculty of the James T. Laney School of Graduate Studies of Emory University in partial fulfillment of the requirements for the degree of Master of Science in Clinical Research
2022

Abstract

Metabolomic Analysis of Microbial Small Molecules in Non-Alcoholic Fatty Liver Disease
By Brent Gawey

Introduction: High-resolution metabolomics (HRM) is an innovative platform that can quantify products of microbial metabolism in systemic circulation. δ-Valerobetaine (VB) and 5-methoxyindole-acetic acid (5-MIAA) are two recently discovered microbial metabolites with known systemic bioactivity. VB has been implicated in carnitine and fatty acid metabolism and 5-MIAA has been shown to activate hepatic detoxification pathways (Nrf2/ARE). These pathways are highly expressed in the liver and play a role in hepatic inflammatory diseases, such as non-alcoholic fatty liver disease (NAFLD). Our study aimed to investigate the associations of these novel metabolites with validated measures of liver fibrosis, hypothesizing that each metabolite would be associated with the underlying degree of liver fibrosis. We also performed metabolome wide association studies (MWAS) and pathway analysis to further characterize these metabolites.
Methods: This cross-sectional study included 139 participants with pre-existing liver fibrosis (mean age 51.08 ± 12.44 years, 66% female). Fasting plasma was analyzed using HRM methodology. Liver fibrosis was quantified using liver stiffness measurement (LSM) scores. Regression analyses were used to investigate associations between level of VB and 5-MIAA with LSM. An MWAS was performed to identify metabolites significantly associated with VB and 5-MIAA. These results were further analyzed using pathway analysis.
Results: In regression analysis, there were no significant associations between plasma level of VB and 5-MIAA with LSM. MWAS for VB identified a variety of lipids and an NAD precursor to be most significantly associated. Pathway analysis for VB revealed significant associations with limonene degradation, lysine and glutathione metabolism, and vitamin B1 metabolism. MWAS for 5-MIAA identified a variety of dicarboxylic acids and fatty acids demonstrating the most significant associations. Pathway analysis for 5-MIAA revealed significant associations with lipid metabolism, chondroitin sulfate degradation, and biopterin metabolism.
Conclusions: While regression analysis did not show any significant associations between plasma level of VB and 5-MIAA with LSM in patients with NAFLD, network and pathway analysis found a variety of metabolites and metabolic pathways with known biological importance in liver fibrosis significantly associated with VB and 5-MIAA. These associations offer future direction for analysis of the gut microbiome's role in host inflammatory responses and human disease.

Metabolomic Analysis of Microbial Small Molecules in Non-Alcoholic Fatty Liver Disease

By

Brent Gawey
B.S., University of Georgia, 2017

Advisors:
Dean Jones, PhD
Andrew Neish, MD
Thomas Ziegler, MD MS

A thesis submitted to the Faculty of the James T. Laney School of Graduate Studies of Emory University in partial fulfillment of the requirements for the degree of Master of Science in Clinical Research
2022

Acknowledgements

It is with tremendous gratitude and indebtedness that I begin to think about the extensive list of individuals who contributed to the production of this thesis. First, I would like to thank my lead mentor, Dr. Andrew Neish. I came to Dr. Neish in my second year of medical school, with the general idea that I wanted to do research in the gut microbiome. He quickly guided me to the right people, introduced me to the world of metabolomics, and fueled my creativity at every step of the way. He became a great friend through the process, as well as an outstanding mentor.

Secondly, I would like to thank co-mentor, Dr. Thomas Ziegler. I first met Dr. Ziegler to discuss my interests in the MSCR program. He took the time to listen to my career goals, provide insightful commentary, and guide me in my next steps. He contributed tremendously to my personal and professional development since our initial meeting and was a strong pillar of support throughout the MSCR year and beyond. His gregarious personality made our frequent contact very enjoyable, and I am fortunate to know such an exemplary leader. I am grateful for the many academic opportunities he has given me, particularly the opportunity to design my own research project as I began to independently engage in the process of scientific inquiry. He has truly made a lasting impact on my formative development.

Thirdly, to co-mentor, Dr. Dean Jones. Dr. Jones graciously took me into his lab and introduced me to crucial members responsible for the completion of this thesis. He provided the dataset this project interrogated and graciously allowed me to use the laboratory's software and equipment. He included me in lab meetings and shared multiple educational resources to support my growing interest in metabolomics methodology.

Importantly, I also owe a tremendous amount of gratitude to my thesis reader, Dr. Jeffrey Collins. Dr. Collins was actively engaged in this project from the start, giving thoughtful feedback on my writing and interrogating my thought process every step of the way. His level of engagement throughout the process taught me many valuable lessons in effective mentorship. Through his efforts, I now believe myself to be a much more capable critical thinker and more equipped to carry out the scientific process in the future. I have much appreciation for his unconditional support through countless rounds of edits and manuscript improvements. His wisdom and insight are truly appreciated.

Crucially, this thesis would not have been possible without my formal metabolomics training done by Dr. Ryan Smith. Dr. Smith exemplified idealistic levels of patience, understanding, and flexibility as he taught me the lab's metabolomics methodology. I greatly appreciate his thoughtful responses to my endless questions, and for the side conversations regarding different barbeque recipes. I owe a tremendous amount to him for all of his efforts.

I would like to thank Dr. Ken Liu of the Dean Jones laboratory for his initial ideas in formulating my thesis along with technical support to move this project to completion.

Thank you to Dr. Amita Manatunga for chairing my thesis committee and for your excellent instruction during the MSCR coursework.

I would also like to thank all of the MSCR faculty for their investment in me throughout the year. Particularly, Dr. Matthew Magee, for his instruction in class and numerous responses to my emails

that greatly assisted in the timely completion of this thesis. Additionally, I would like to thank Dr. Jessica Alvarez, for her kindness and willingness to take the time to explain the conceptual approach of clinical metabolomics to me.
To the MSCR program, for giving me the opportunity to learn how to ask intelligent, researchable questions and for efforts to make the learning environment enriching despite the ongoing COVID-19 pandemic. Thank you for the support of my training and for providing the framework with which I could be formally introduced to the world of science. Also, to my fellow MSCR students, for providing a collaborative environment that made it fun to learn in, despite the virtual setting.

Finally, I want to thank my family members for their support during this process and forcing me to develop the ability to clearly explain what "metabolomics" is in plain English. Also, for never failing to ask in conversation, "so when is your thesis going to be done"?

I am grateful for each individual listed above and for those not specifically mentioned who contributed to this project. This collaboration taught me important lessons in team science and the power of the collective. I hope that in my current and future work, I will make them proud and show them that their investment was worthwhile.

Table of Contents

List of figures and tables

List of Abbreviations

5-MIAA – 5-methoxyindole-acetic acid
ALP – alkaline phosphatase
ALT - alanine aminotransferase
AST- aspartate aminotransferase
CI – confidence interval
DAG – Directed acyclic graph
db/m - decibels per meter
FDR – False discovery rate
HDL – high-density lipoprotein
HILIC – Hydrophilic interaction liquid chromatography
HIPAA – Health Insurance Portability and Accountability Act of 1996
HMDB – Human metabolome database
HMDB – Human Metabolome DataBase
HRM – High-resolution metabolomics
K/μl – kilo per microliter
kg/m^2– kilogram per square meter
kPa – Kilopascals
LC-MS – Liquid chromatography mass spectrometry
LDL - low-density lipoprotein
LSM – Liver stiffness measurement
m/z – Mass to charge
max – maximum
mg/dL – milligrams per deciliter
min – minimum
MS – Mass spectrometry
MS/MS – Tandem mass spectrometry
MWAS – Metabolome wide association study
NAD - nicotinamide adenine dinucleotide.
NAFLD – Non-alcoholic fatty liver disease
NAFLD – Nonalcoholic fatty liver disease
NASH – Nonalcoholic steatohepatitis
Nrf2/ARE - nuclear erythroid 2-related factor 2/antioxidant response element
OR – odds ratio
r – Spearman's correlation coefficient
RSD – Relative standard deviation
RT – Retention time
SD – standard deviation
SEM – Standard error of the mean
Std. – standard
VB – δ-Valerobetaine
VCTE - vibration-controlled transient elastography

1. Introduction

The human gastrointestinal tract is home to trillions of microorganisms known to play an important role in the metabolism of ingested compounds (xenobiotics), including food, pharmaceutical drugs, and industrial chemicals.[1-5] This interaction produces a metabolic profile unique to an individual's gut microbial composition.[6,7] These microbiota-derived, bioactive metabolites can be readily absorbed and distributed throughout systemic circulation, having a potential impact on whole organism homeostasis and metabolism. While gut-microbiota driven metabolism can provide numerous benefits to the host, abnormalities of microbial composition have been associated with metabolic and inflammatory diseases, including adult-onset diabetes, inflammatory bowel disease, obesity, and nonalcoholic fatty liver disease.[6-10] How microbiota-driven xenobiotic metabolism mechanistically influences these aspects of human physiology beyond direct effects on the gut epithelium and the associated enteric immune system remains largely unknown.

Nonalcoholic fatty liver disease (NAFLD) is the most common chronic liver disease in the United States.[11] NAFLD results from increased deposition of fat in liver cells and is associated with altered lipid and glucose metabolism.[12] Excess fat accumulation is toxic to liver cells, leading to inflammation, which can progress to nonalcoholic steatohepatitis (NASH). In states of prolonged hepatic inflammation, as seen in NASH, scarring and fibrosis develop, leading to liver cirrhosis. Liver cirrhosis is associated with increased rates of hepatocellular carcinoma and mortality.[13] Accordingly, liver fibrosis is the strongest predictor of liver and non-liver related outcomes in patients with NAFLD.

High-resolution metabolomics (HRM) is an innovative technique that allows for the study of host-microbial metabolism through the analysis of a large number of metabolites and biological

compounds. This provides mechanistic insight to the contribution and function of microbial-derived metabolites in human physiology, including metabolism. A recent study using HRM in concert with germ-free and conventional mice identified a variety of novel metabolites synthesized directly from gut microbiota that influence host physiology.[14] Specifically, a novel microbial-derived metabolite, δ-Valerobetaine (VB) was found to have significant biochemical activity toward carnitine and fatty acid metabolism, and is linked with dietary fat intake in animal models, with a hypothesized role in metabolic hepatic diseases, such as NAFLD (Figure 1).[15] Additionally, microbial-derived 5-methoxyindole-acetic acid (5-MIAA) was found to be a potent activator of host hepatic detoxification pathways (NRF-2/ARE), serving as a potential biomarker for hepatic metabolism and severity of hepatic steatosis (Figure 2).[14]

In this study, we utilized HRM to investigate the associations between the novel, gut-derived metabolites VB and 5-MIAA with NALFD using data from a participant cohort with measured scores of liver fibrosis as a proxy for underlying severity of NAFLD. Understanding that VB decreases fatty acid oxidation in the liver, we hypothesized that plasma δ-Valerobetaine (VB) will be associated with the degree of liver fibrosis in individuals with NAFLD. Further, because 5-MIAA drives antioxidant responses in the liver, we hypothesized that plasma 5-methoxyindole-3-acetic acid (5-MIAA) will be associated with the degree of liver fibrosis in individuals with NAFLD. Finally, we conducted an untargeted metabolome wide association study (MWAS) for both VB and 5-MIAA separately, to assess their association with concentrations of other detected metabolites in this same cohort, based on the hypothesis that plasma concentrations of VB and 5-MIAA are associated with differences in the broader metabolome.

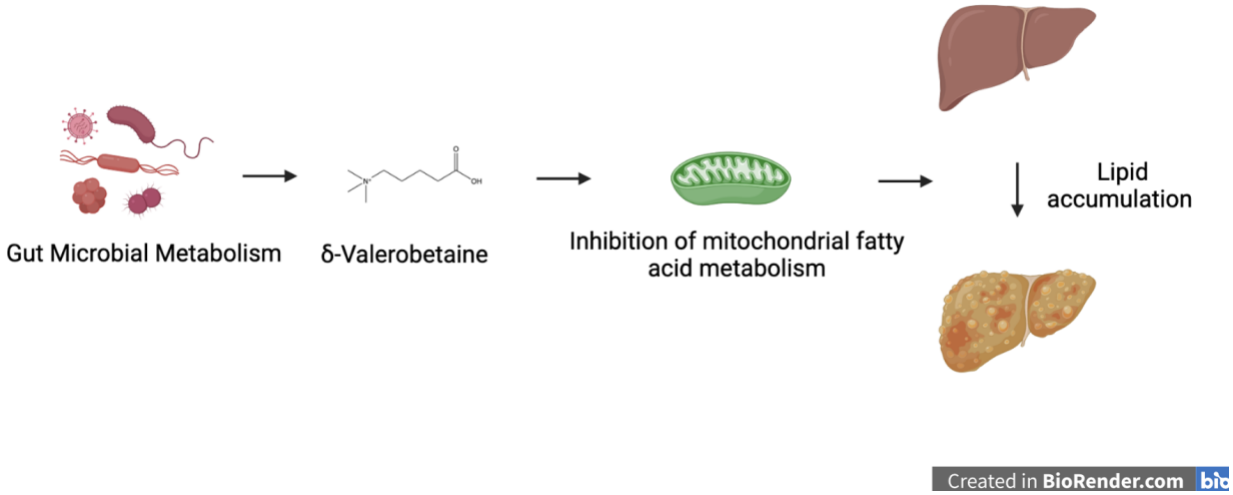

Figure 1 Molecular structure and biological function of gut microbial metabolite δ-Valerobetaine (VB), demonstrating its inhibitory effect on mitochondrial fatty acid oxidation through disrupted uptake of L-carnitine. These metabolic changes are associated with increased hepatic lipid accumulation, predisposing to the development and progression of NAFLD.

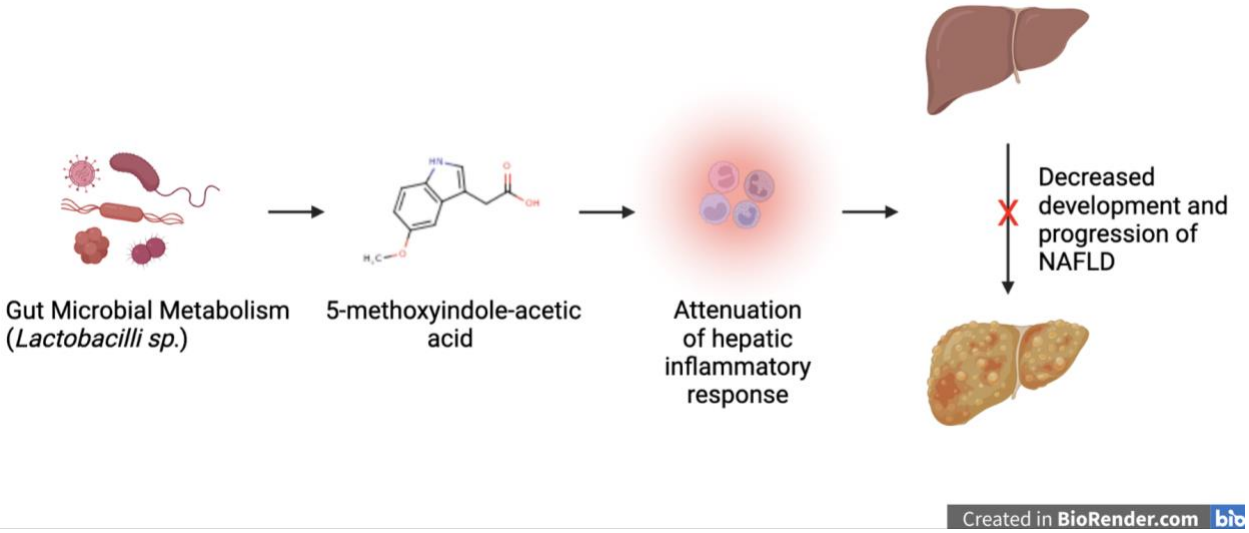

Figure 2 Molecular structure and biological function of gut microbial metabolite 5-methoxyindole-acetic acid (5-MIAA), demonstrating its ability to activate the nuclear erythroid 2-related factor 2/antioxidant response element (Nrf2/ARE) detoxification pathway in the liver and attenuate the hepatic inflammatory response. Reduced hepatic inflammation is associated with decreased development and progression of NAFLD.

2. Methods

2.1 Study Design, Setting, Participants

This cross-sectional study utilized archived, de-identified data and plasma samples from 139 adult patients with nonalcoholic fatty liver disease (NAFLD). Patients were recruited in the Indiana University Health network as part of a previously published, placebo-controlled clinical trial of obeticholic acid for adult patients with nonalcoholic steatohepatitis (NASH).[16] All participants were 18 years or older, able to read and understand English, and able to provide informed consent. Exclusion criteria were fasting for < 3 hours, current significant alcohol consumption, history of significant alcohol consumption for > 3 consecutive months within one year prior to screening, active substance abuse, participants with cardiac or gastrointestinal pacemakers, pregnant women or nursing mothers, patients unable or unwilling to sign the informed consent statement and HIPAA Authorization form, and participants who met criteria for prisoners or those under the care of the Department of Corrections. Significant alcohol consumption was defined as more than 20 grams per day in females and more than 30 grams per day in males. All measurements used for analyses were taken from baseline visits. All participants provided written informed consent and the study was approved by the affiliated Institutional Review Board.

2.2 Variables

Liver fibrosis was quantified by liver stiffness measurement (LSM) scores via Fibroscan® (ultrasound). LSM score was used to distinguish patients with advanced liver fibrosis and those without advanced liver fibrosis. Advanced fibrosis was defined as LSM ≥ 12.1 of 12.1 kilopascals (kPa) and non-advanced fibrosis as LSM < 12.1 kPa.[17]

2.3 Plasma High-Resolution Metabolomics (HRM)

Fasting plasma samples were collected for participants (n=139) using an established protocol in the Emory Clinical Biomarkers Laboratory.[18] In summary, 65 microliters of plasma previously-stored at −80°C was added to 130 mL acetonitrile along with a mixture of internal standard stable isotopes.[18] Samples were analyzed with three technical replicates using liquid chromatography-Fourier transform mass spectrometry (Dionex Ultimate 3000, Q-Exactive, Thermo Scientific, Waltham, MA, USA) with both a C18 column under negative-ionization and a hydrophilic interaction liquid chromatography (HILIC) [Accucore HILIC 100×2.1mm columns] column under positive ionization to maximize the detection of low-molecular-weight chemicals. Batches of 20 randomized samples were run with quality control samples analyzed at the beginning and end of each batch. Following analysis of all participant and quality control samples, the metabolic profiles from each analytical run were extracted by R-based packages apLCMS and xMSanalyzer with batch correction done by ComBat.[19-21] Output was displayed in mass to charge (*m/z*) feature tables with detected ions and their associated relative retention times and mass. Metabolic features without matches to common adducts within a 10 ppm window were labeled as "unknown."

Each metabolic feature was characterized by its mass to charge (*m/z*) ratio, retention time (RT), and peak intensity. Metabolic features were annotated using xMSannotator.[22] xMSannotator is an R-based package that uses multiple criteria to provide score-based metabolite annotation using the Human Metabolome DataBase (HMDB).[23] Metabolite identities were confirmed by comparing coelution with standard curves from authentic chemical standards. Chemical identity level was based on Schymanski et al. (2014), where 1 represents the highest level of identification and 5 represents the lowest.[24] Results were only reported for columns on which the metabolites

could be annotated. The authentic chemical standard for VB was from an in-house validated library of over 920 chemicals. The authentic chemical standard for 5-MIAA was individually purchased with stated purity >95%.

2.4 Data Analysis

Descriptive statistics were performed for clinical and demographic variables using JMP Pro (Version 15, SAS Institute Inc, Cary, NC). Summary statistics are reported as mean ± standard deviation (SD) or count and percentage. LSM was analyzed as a dichotomous variable with logistic regression and as a continuous variable with linear regression. To dichotomize LSM, advanced fibrosis was denoted as LSM ≥ 12.1 kilopascals (kPa) and non-advanced fibrosis as LSM < 12.1 kPa.[17] Plasma concentrations of metabolites were log2 transformed and quantile normalized for all analyses. Model building was guided by directed acyclic graph (DAG) theory. DAGs allow for the assessment of potential relationships amongst variables, providing discernment for which variables should be included or excluded from analyses to minimize bias.[25] Variables existing on open back door pathways were considered to be potentially confounding of the primary exposure outcome relationship, and were included in the model. Variables existing on closed directed pathways were considered mediators and were not included in the analysis to prevent removing part of the variable's effect. Regression analyses for Aim 1 and 2 were performed using xMSPanda (https://github.com/kuppal2/xmsPANDA). xMSPANDA is an R-based package that can perform classification, regression-based feature selection, and statistical analyses for metabolites of interest.

The metabolome-wide association studies (MWAS) for VB and 5-MIAA were performed using the xmsPANDA package in R studio.[26] xmsPANDA uses multiple linear regression analyses

to identify metabolites significantly associated with 5-MIAA and VB. False discovery rate (FDR) in regression analyses was controlled for using the Benjamini–Hochberg method (q = 0.2) or a raw p-value of 0.05 if the Benjamini-Hochberg method failed to identify any significant associations.[27] Metabolites not present in at least 10% of the samples and at least 80% of samples for a group were excluded from analyses. For association analysis, Spearman's correlation threshold was set to 0.4. Metabolites significantly associated with VB and 5-MIAA were selected for pathway enrichment analysis using mummichog.[28] Mummichog is a Python-based program that maps statistically significant metabolic features with known human biological pathways. A raw p-value significance threshold of 0.05 was used as a cutoff for metabolites associated with VB and 5-MIAA. Enriched pathways were included if they contained four or more overlapping metabolic features with at least one metabolite with a confirmed identity.

3. Results

3.1 Descriptive Data

A total of 139 patients met the inclusion criteria. Table 1 shows baseline demographic, clinical, and laboratory characteristics of the 139 participants analyzed in this study. Overall, participants were mostly female (n=92, 66.2%) and had non-advanced liver fibrosis (n=97, 70%), with an average age of 51.1 years (± 12.4 years). Type 2 diabetes was prevalent in 49% (n=68) of participants with an average BMI of 35.9 (± 7.52). Of patients with non-advanced liver fibrosis, LSM was 6.55 kPa (± 2.19 kPa) compared to 24.82 kPa (± 13.11 kPa) in patients with advanced liver fibrosis. A bivariate analysis was also performed to determine factors associated with LSM score (Table 1).

Table 1. Baseline characteristics and bivariate analysis for patients (n=139) with non-advanced liver fibrosis versus those with advanced liver fibrosis.

	Non-Advanced Fibrosis (LSM < 12.1 kPa) N = 97	**Advanced Fibrosis (LSM ≥ 12.1 kPa) N = 42**	**OR (95% CI)**
Age (years)	50.90 ± 12.50[1]	52.70 ± 12.44	1.02 (0.99, 1.05)
Sex			
Male		13 (0.31%)	REF
Female	69 (64.49%)[2]	29 (69%)	1.20 (0.56, 2.67)
Missing	0 (0%)	0 (0%)	--
BMI (kg/m^2)	34.81 ± 7.30	38.93 ± 6.96	1.08 (1.02, 1.14)
Type 2 Diabetes Mellitus			
No	48 (49.5%)	21 (50%)	REF
Yes	49 (50.5%)	21 (50%)	0.98 (0.48, 2.02)
Missing	0 (0%)	0 (0%)	--
FibroScan VCTE Data			
Liver Stiffness Measurement (kPa)	6.55 ± 2.19	24.82 ± 13.11	--
Controlled Attenuation Parameter (db/m)	316.53 ± 52.47	346.84 ± 50.38	1.01 (1.00, 1.02)
Laboratory Data			
Platelets (K/μL)	252.71 ± 78.81	206.28 ± 107.69	1.00 (0.99, 1.01)
ALT (U/L)	62.60 ± 50.37	56.10 ± 40.0	1.00 (0.99, 1.01)
AST (U/L)	51.10 ± 66.85	51.03 ± 39.84	1.00 (0.99, 1.01)
ALP (U/L)	92.04 ± 53.18	89.36 ± 39.18	1.00 (0.99, 1.01)
Total Bilirubin (mg/dL)	0.64 ± 0.44	0.70 ± 0.42	1.67 (0.72, 3.88)
Cholesterol, Total (mg/dL)	169.82 ± 44.30	167.0 ± 37.0	1.00 (0.99, 1.01)
LDL (mg/dL)	106.21 ± 62.65	94.75 ± 32.31	1.00 (0.99, 1.01)
HDL (mg/dL)	40.46 ± 10.14	42.47 ± 10.26	1.03 (0.99, 1.08)
Triglycerides (mg/dL)	169.01 ± 97.17	151.81 ± 54.68	0.99 (0.98, 1.00)

Abbreviations: OR – odds ratio, CI – confidence interval, kg/m^2– kilogram per square meter, LSM – liver stiffness measurement, kPa – kilopascal, VCTE - vibration-controlled transient elastography, db/m - decibels per meter, ALT - alanine aminotransferase, AST- aspartate aminotransferase, ALP – alkaline phosphatase, LDL - low-density lipoprotein, HDL – high-density lipoprotein, K/µl – kilo per microliter, mg/dL – milligrams per deciliter

1. Value represents mean ± standard deviation.
2. Value represents column percentage, n (%).
3. Missing values were excluded from bivariate analyses for the outcome (level of liver fibrosis) based on analysis of effect of missing values on measures of association using sensitivity analyses.

3.2 Main Results - Metabolite Identification and Quantification

VB was detected as an $[M+H]^+$ adduct (*m/z,* 160.13322) eluting at 82.57 seconds in HILIC positive mode, however VB was not detected as an annotatable ion (*m/z* with retention time and intensity; hereafter termed "metabolite") in C18 negative mode. 5-MIAA was detected as an $[M+H]^+$ adduct (*m/z,* 170.06092) eluting at 95.37 seconds and an $[M-H]^-$ adduct (*m/z*, 204.0667) eluting at 40.4 seconds (Table 3). VB was annotated at a level 1 on the HILIC positive column. 5-MIAA was annotated at a level 2 on the C18 negative and HILIC positive columns.[24]

Table 2 Detection of δ-valerobetaine and 5-methoxyindole-3-acetic acid primary adducts based on mass-to-charge (*m/z*) ratio and retention time (RT) with associated chemical identity confidence score.

Metabolite	Adduct	Detected m/z	RT sec	Chemical Identity Level[A]
δ-Valerobetaine	$[M+H]^+$	160.13322	82.57	1
5-methoxyindole-3-acetic acid	$[M-H]^-$	204.06665	40.4	2
	$[M+H]^+$	170.06092	95.37	2

Abbreviations: *m/z*, mass to charge; RT, retention time

A. Chemical identity according to Schymanski et al. (2014), where 1 represents the highest level of identification and 5 represents the lowest.[24]

3.3 Main Results – δ-Valerobetaine Targeted Analysis

In the crude logistic regression model (Table 4), there was no significant association between plasma level of *δ*-valerobetaine and level of liver fibrosis (OR 1.03; 95% CI 0.89, 1.20). There was no significant difference in plasma level of VB between patients with non-advanced fibrosis and advanced fibrosis (21.11 ± 2.31 and 21.31 ± 2.85, respectively; p = 0.96) (Figure 3). Further considering the relationship between plasma VB (primary exposure) and NAFLD (primary outcome), an assessment of covariates was done by literature review and directed acyclic graph theory to construct a DAG, which demonstrated the hypothesized causal relationship between selected variables included in Table 1. The age variable existed on an open backdoor path and thus was considered as a potential confounder by DAG theory and included in the model based on these assumptions (Appendix, Figure A.1). Multivariable regression analysis found no significant association between plasma level of VB and level of liver fibrosis (adjusted OR 1.02; 95% CI 0.88, 1.19) (Table 4).

Linear regression analyses were run to assess the relationship between LSM and quantile normalized level of plasma VB (Table 5). LSM was log transformed for analyses. In the simple linear model, VB was not significantly associated with LSM (regression coefficient 0.01; SE 0.01; p-value 0.38). The model had poor discrimination, explaining only 0.6% of the variation in LSM (R-square = 0.006; Figure 4). In the adjusted model including age, VB was not significantly associated with LSM (regression coefficient 0.007; SE 0.01; p-value 0.52).

Table 3 Multivariable logistic model for the association between plasma level of δ-Valerobetaine and level of liver fibrosis.

Participant Characteristics	Crude OR (95% CI)	Adjusted* OR (95% CI)
δ-Valerobetaine, $[M+H]^+$	1.03 (0.89, 1.20)	1.02 (0.88, 1.19)

Abbreviations: OR – odds ratio, CI – confidence interval
*Model adjusted for age at consent.

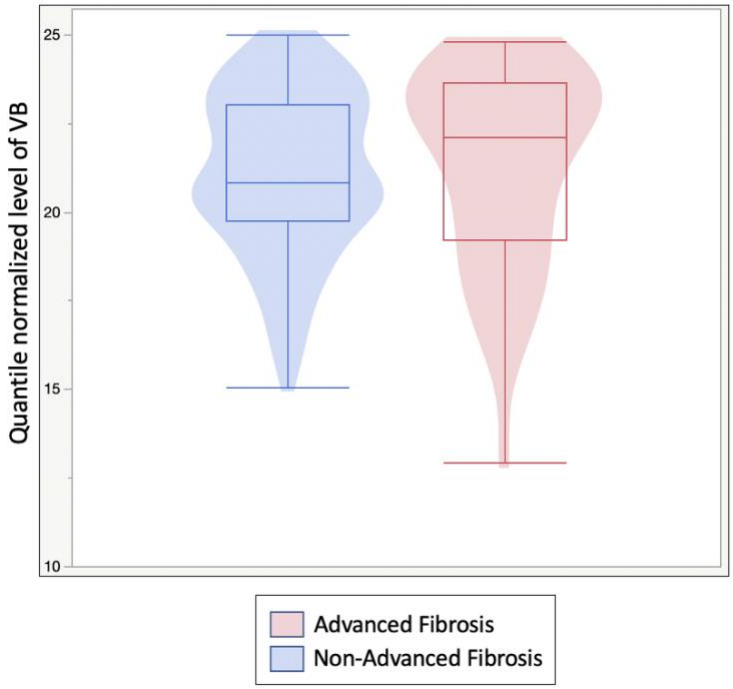

Figure 3 Box plots with violin plot overlay to demonstrate plasma level of VB in 97 patients with non-advanced fibrosis and 42 patients with advanced fibrosis. Boxplot 'boxes' indicate the first and third quartiles of the data. The middle lines indicate the medians. Boxplot 'whiskers' indicate the inner fences of the data. There was no significant difference in plasma level of VB between patients with non-advanced fibrosis and advanced fibrosis (21.11 ± 2.31 and 21.31 ± 2.85, respectively; p = 0.96).

Table 4 Univariate regression for the association between log transformed LSM and quantile normalized level of δ-valerobetaine.

Variable	Mean (SD) [Range]	Crude Regression Coefficient (Std. Error, p-value)	Adjusted[B] Regression Coefficient (Std. Error, p-value)
LSM[A]	0.96 (0.30) [0.32 - 1.86]	--	--
δ-Valerobetaine, $[M+H]^{+}$ [A]	21.17 (2.48) [12.93 - 25.01]	0.01 (0.01, 0.38)	0.007 (0.01, 0.52)

Abbreviations: SD – standard deviation, min – minimum, max – maximum, Std. – standard

A. Log normalized variable.
B. For the final model containing plasma level of δ-Valerobetaine and participant's age.

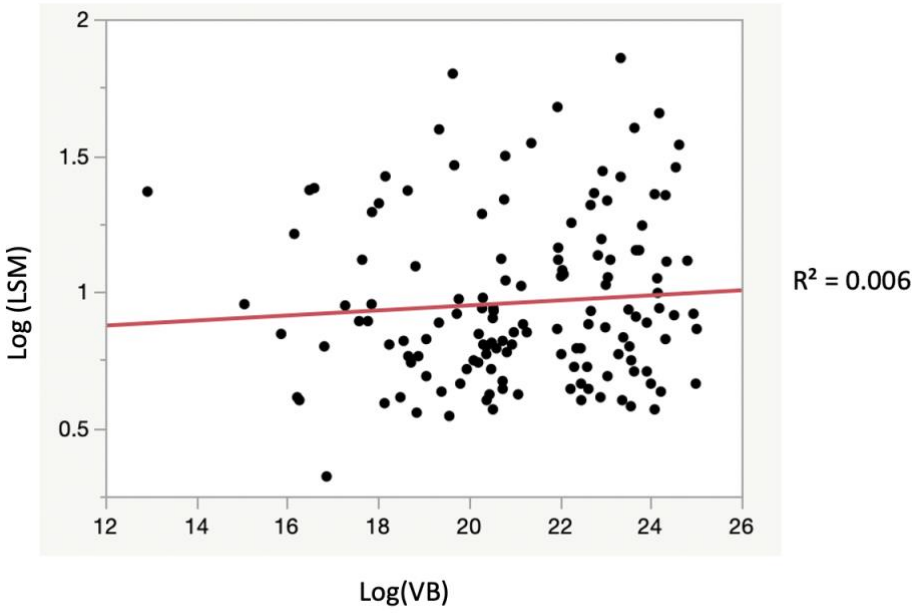

Figure 4 Intensity values for quantile normalized level of plasma VB (identified by *m/z* and retention time) for individual plasma samples plotted as a function of the associated log transformed LSM measurement for the sample.

3.4 Main Results – δ-Valerobetaine Untargeted Analysis

An untargeted metabolome wide association study (MWAS) detected 18,009 ions (*m/z* with retention time and intensity; hereafter termed "metabolites") in HILIC positive mode. Metabolites not present in at least 10% of the samples were excluded from analyses, leaving 16,958 metabolites. VB was not annotated in C18 negative mode, and therefore was not included in the MWAS. A raw p-value of 0.05 was used for significance in order to identify regulated metabolic pathways. Using this method, 5,560 metabolites were significantly associated with the plasma VB level. The Manhattan plot (Figure 5) showed a broad range of *m/z*, indicating that correlated metabolites were diverse in size and physical properties. An unknown metabolite (*m/z* 308.8586, RT 120s) had the highest –log P ($r = 0.37$, $p < 0.0001$). Correlated features were used to search the HMDB metabolomics database for matches to common ions (H+, H-) with 10 ppm tolerance. 13 of the top 25 metabolites were of unknown identity. Other metabolites significantly associated with VB with confirmed identification are shown in Table 6.

Analysis of pathways associated with VB demonstrated significant ($p<0.05$) enhancement of the following metabolic pathways related to nutrient metabolism (Figure 6): limonene/pinene degradation, Vitamin B1 (thiamine) metabolism), amino acid metabolism (aspartate and asparagine metabolism, glutathione metabolism), sialic acid metabolism, and lipid metabolism and polyunsaturated fatty acid biosynthesis.

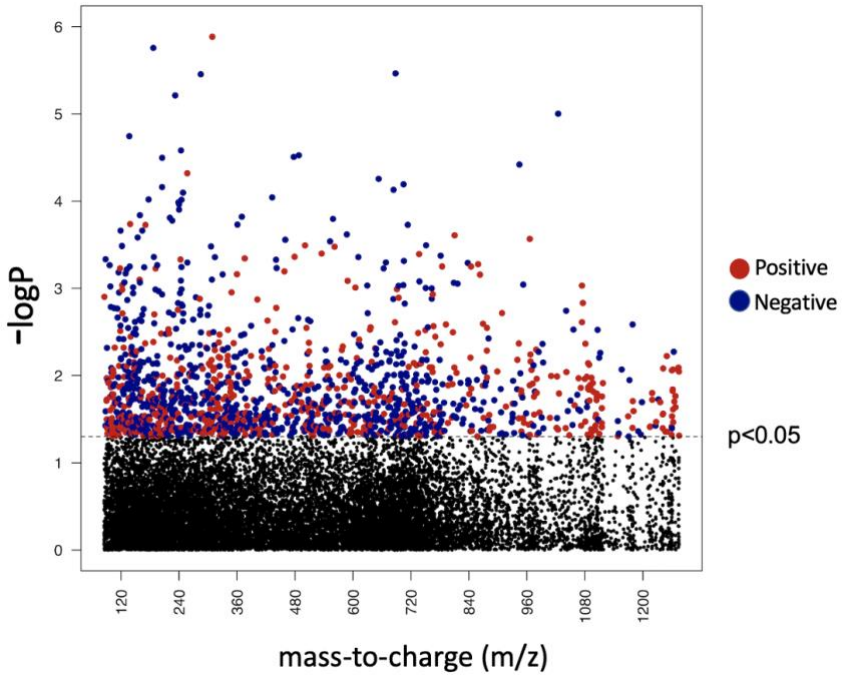

Figure 5 Type 1 Manhattan plot showing the negative log p (−log p) for correlation of 5,560 significant features at p = 0.05 as a function of the *m/z* with VB level. *m/z* 308.8586 had *r* = 0.37 and the highest -log10 P. Features positively associated with VB are represented in red, and those with negative associations are in blue. Most significant metabolites were of unknown origin.

Table 5 Selected metabolites (hydrogen adducts) significantly correlated with δ-valerobetaine in an untargeted analysis of human plasma in patients with varying degrees of liver fibrosis.

Metabolite	Adduct	Detected *m/z*	RT sec	Correlation *r*	P-value	P-value (FDR corrected)	Metabolite Class
3-Mercaptohexyl butyrate	$[M+H]^+$	205.1264	85	- 0.33	< 0.0001	< 0.0001	Fatty acid ester
Tiglylcarnitine	$[M+H]^+$	244.1546	29	- 0.32	< 0.0001	< 0.0001	Acylcarnitine
Hydroxyhexanoycarnitine	$[M+H]^+$	240.1614	125	- 0.29	< 0.0001	0.0001	Acylcarnitine
1-(beta-D-Ribofuranosyl)-14-dihydronicotinamide	$[M+H]^+$	257.1129	112	0.32	< 0.0001	< 0.0001	NAD precursor

Abbreviations: m/z - mass-to-charge ratio, RT - retention time, *r* – Spearman's correlation coefficient, FDR – false discovery rate, NAD - nicotinamide adenine dinucleotide.

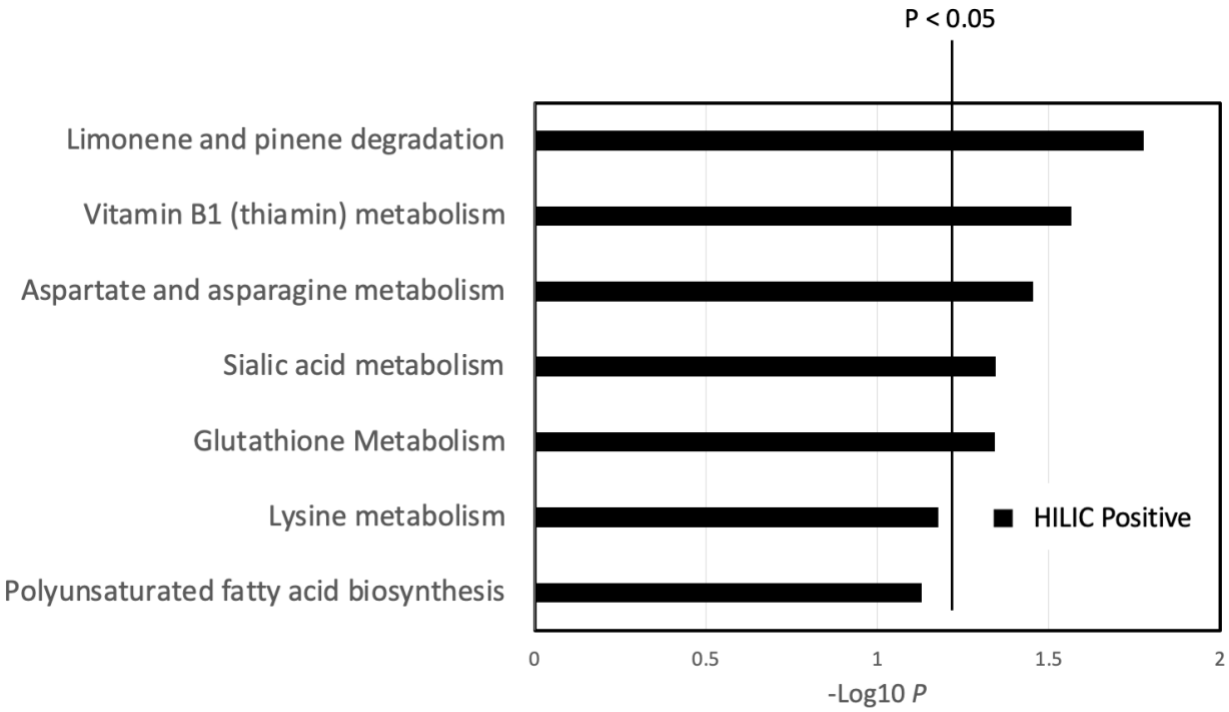

Figure 6. Pathway enrichment analysis ($p < 0.05$) using *mummichog*. The y-axis shows metabolic pathways significantly associated with metabolome wide alterations associated with VB. The x-axis shows the -log p-value for pathway enrichment. Results are only shown for HILIC positive ionization mode, as VB was unable to be annotated using C18 negative chromatography. Limonene and pinene degradation was the most significantly enhanced pathway.

3.5 Main Results – 5-methoxyindole-acetic acid Targeted Analysis

In the crude logistic regression model (Table 4), there was no significant association between plasma level of 5-MIAA and the level of liver fibrosis (OR 1.46; 95% CI 0.86, 2.47). There was no significant difference in plasma level of 5-MIAA between patients with non-advanced fibrosis and advanced fibrosis (21.06 ± 0.77 and 21.23 ± 0.78, respectively; p = 0.2143) (Figure 7). Considering the relationship between plasma 5MIAA (primary exposure) and NAFLD (primary outcome), an assessment of covariates was done by literature review and directed acyclic graph (DAG) theory, which demonstrated the hypothesized causal relationship between all variables included in Table 1. There were no covariates on open backdoor paths in the DAG and thus no covariates were considered confounders by DAG theory and included in the model based on these assumptions (Appendix, Figure A.2).

Linear regression analysis was performed to assess the relationship between LSM and quantile normalized level of plasma 5-MIAA (Table 7). LSM was log transformed for analyses. In the linear model, 5-MIAA was not significantly associated with LSM (regression coefficient 0.05; SE 0.04; p-value 0.16). The model had poor discrimination, explaining only 1.0% of the variation in LSM (R-square = 0.01; Figure 8).

Table 6. Regression model for the association between plasma level of 5-MIAA with C18 negative separation and level of liver fibrosis.

Participant Characteristics	Crude OR (95% CI)
5-methoxyindole-3-acetic acid, $[M-H]^-$	1.46 (0.86, 2.47)

Abbreviations: OR – odds ratio, CI – confidence interval

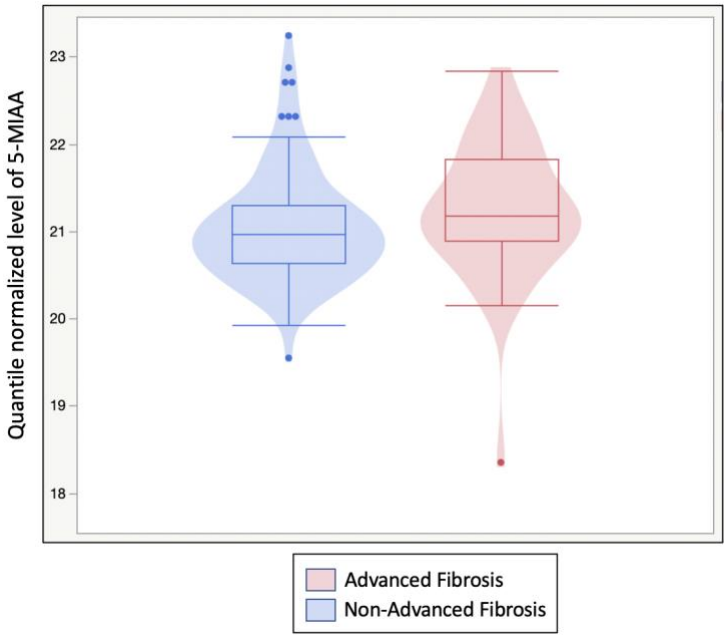

Figure 7 Box plots with violin plot overlay to demonstrate 5-MIAA level with C18 separation in 97 patients with non-advanced fibrosis and 42 patients with advanced fibrosis. Boxplot 'boxes' indicate the first and third quartiles of the data. The middle lines indicate the medians. Boxplot 'whiskers' indicate the inner fences of the data. Outliers are represented as individual points. There was no significant difference in plasma level of 5-MIAA between patients with non-advanced fibrosis and advanced fibrosis (21.06 ± 0.77 and 21.23 ± 0.78, respectively; p = 0.2143).

Table 7: Univariate regression for the association between LSM (kPa) and plasma level of 5-methoxyindole-3-acetic acid.

Variable	Mean (SD)	Min, Max	Regression Coefficient Estimate (Std. Error)	P-value
LSM[A]	0.96 (0.30)	0.32, 1.86	--	--
5-methoxyindole-3-acetic acid, $[M-H]^{-A}$	21.11 (0.7)	18.35, 23.24	0.05 (0.04)	0.16

Abbreviations: SD – standard deviation, min – minimum, max – maximum, Std. – standard
A. Log normalized variable.

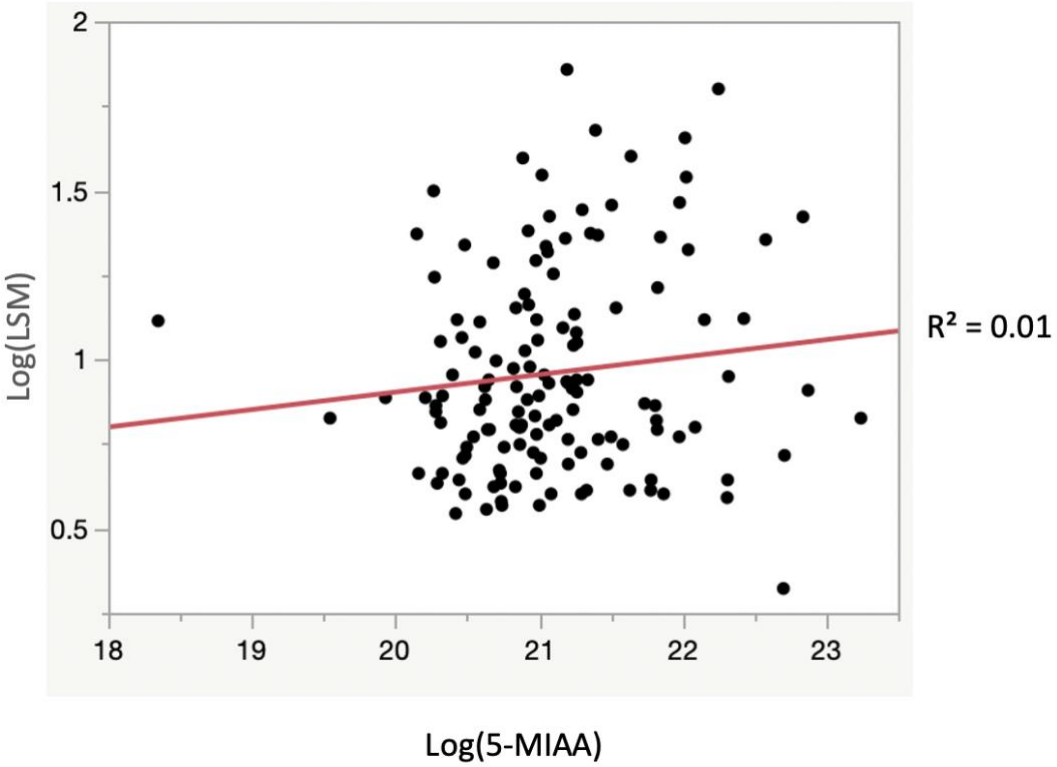

Figure 8 Intensity values for quantile normalized level of plasma 5-MIAA (identified by *m/z* and retention time) for individual plasma samples plotted as a function of the associated log-transformed LSM measurement for the sample.

3.6 Main Results – 5-methoxyindole-Acetic Acid Untargeted Analysis

An untargeted MWAS detected 11,993 metabolites using C18 negative electrospray ionization chromatography. After filtering for metabolites not present in at least 10% of the samples, 11,497 were included in analyses, of which 492 were found to be significantly associated with plasma 5-MIAA level with a rigorous FDR cutoff of q = 0.2. With HILIC positive ionization mode, 18,009 metabolites were detected. After filtering metabolites not present in at least 10% of the samples, 16,968 features underwent downstream analysis, with 788 features significantly associated with plasma 5-MIAA level at a raw p-value of 0.05. A raw p-value of 0.05 was used for significance for metabolites separated using HILIC positive chromatography in the mummichog pathway analytical program.

Manhattan plots (Figures 9, 10) showed a broad range of *m/z* correlated with 5-MIAA, indicating that correlated metabolites were diverse in size and physical properties. Correlated features were used to search the HMDB metabolomics database for matches to common ions (H+, H-) with 10 ppm tolerance. Caffeic Acid (*m/z* 181.0507, RT 37.4 s) had the highest –log P (r = 0.56, $p < 0.0001$). Other metabolites significantly associated with 5-MIAA with confirmed identification are shown in Table 9.

Pathway analysis revealed significant ($p<0.05$) enhancement of the following pathways associated with 5-MIAA: lipid metabolism (de novo fatty acid biosynthesis, glycerophospholipid metabolism, fatty acid metabolism/activation, glycosphingolipid metabolism, omega-3 fatty acid metabolism), saccharide metabolism (chondroitin sulfate degradation, heparan sulfate degradation, prostaglandin formation, galactose metabolism), amino acid metabolism (histidine metabolism, methionine and cysteine metabolism, lysine metabolism, urea cycle/amino group metabolism, aspartate and asparagine metabolism), mitochondrial metabolism (CoA catabolism,

glyoxylate and dicarboxylate metabolism), antioxidant response (biopterin metabolism), and sialic acid metabolism (Figure 11).

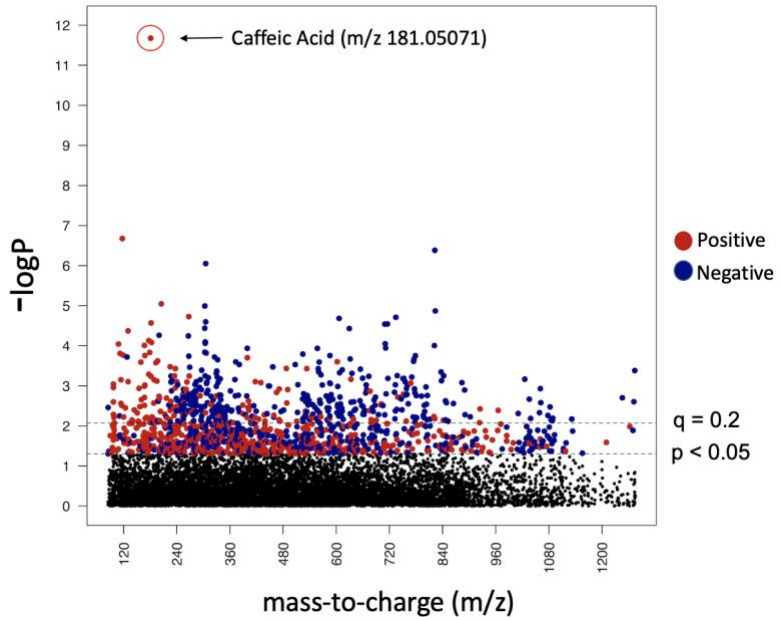

Figure 9 Type 1 Manhattan plot showing the negative log p (−log p) for correlation of 492 significant features (q = 0.2) as a function of the *m/z* with 5-MIAA level with C18 negative separation. *m/z* 181.0507 (Caffeic Acid) had *r* = 0.56 and the highest -log10 P. Each dot represents one metabolic feature. Features positively associated with 5-MIAA are represented in red, and those with negative associations are in blue. Black dots (below p < 0.05) represent metabolites not significantly related to 5-MIAA. False discovery rate (FDR) threshold is labeled as q = 0.2.

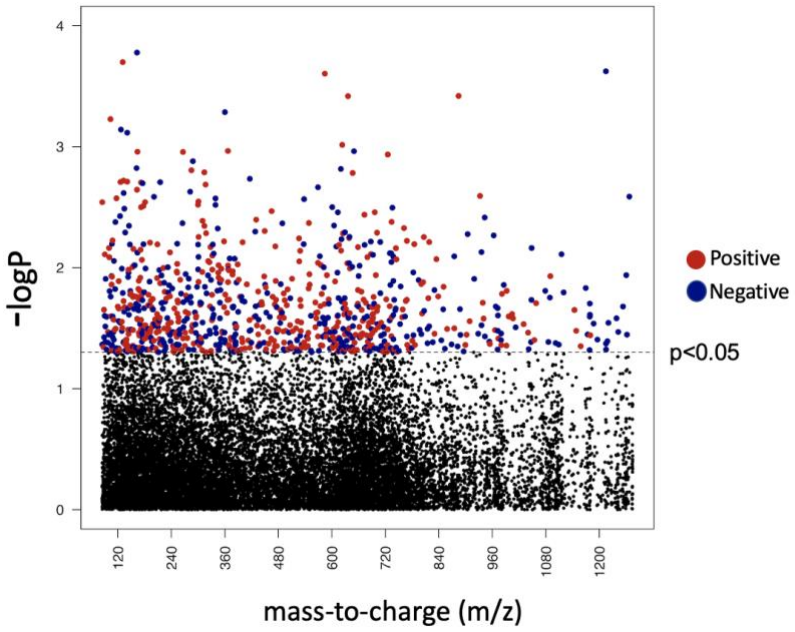

Figure 10 Type 1 Manhattan plot showing the negative log p (−log p) for correlation of 788 significant features (p-value < 0.05) as a function of the *m/z* with 5-MIAA level with HILIC positive separation. *m/z* 162.9739 had p = - 0.31 and the highest -log10 P. Features positively associated with 5-MIAA are represented in red, and those with negative associations are in blue.

Table 8 Selected metabolites (hydrogen adducts) significantly correlated with 5-MIAA in an untargeted analysis of human plasma in patients with varying degrees of liver fibrosis.

Metabolite	Adduct	Detected *m/z*	RT sec	Correlation *r*	P-value	P-value (FDR corrected)	Metabolite Class
Caffeic Acid	$[M-H]^-$	181.05071	37.4	0.56	< 0.0001	< 0.0001	Dicarboxylic acid
1-eicosyl-glycero-3-phosphoserine	$[M-H]^-$	822.5664	230.9	-0.41	< 0.0001	0.002	Glycerophosplipid
Eicosatetraenoic acid	$[M-H]^-$	305.34645	246	-0.4	< 0.0001	0.003	Fatty Acyl (Long chain fatty acid)
Chlorophyllin	$[M-H]^-$	205.07018	40	0.38	< 0.0001	0.02	Tetrapyrrole
5-amino-5-oxopentanoic acid	$[M-H]^-$	130.05097	38.4	0.35	< 0.0001	0.027	Keto Acid
Methyl (2E,6Z)-dodecadienoate	$[M+H]^-$	266.2205	223	- 0.34	< 0.0001	0.034	Fatty Acyl (fatty acid ester)
1-hexadecyl-2-tetradecanoyl-glycero-3-phosphate	$[M-H]^-$	606.4583	234.7	-0.34	< 0.0001	0.024	Glycerophospholipid
N'-arachidonoyl-N''-diethylethylenediamine	$[M-H]^-$	200.17377	159.7	-0.34	< 0.0001	0.034	Keto Acid
N-(heptadecanoyl)-sphing-4-enine-1-phosphocholine	$[M-H]^-$	715.57409	288.4	-0.33	< 0.0001	0.024	Phosphosphingolipid
Methyl (2E,4Z)-hexadecadienoate	$[M-H]^-$	266.22052	223.3	-0.33	< 0.0001	0.034	Fatty Acyl (fatty acid ester)
33-Thiobispropanoic acid	$[M-H]^-$	177.0227	36.6	0.35	< 0.0001	0.042	Dicarboxylic acid

Abbreviations: *m/z* - mass-to-charge ratio, RT - retention time, *r* – Spearman's correlation coefficient, FDR – false discovery rate

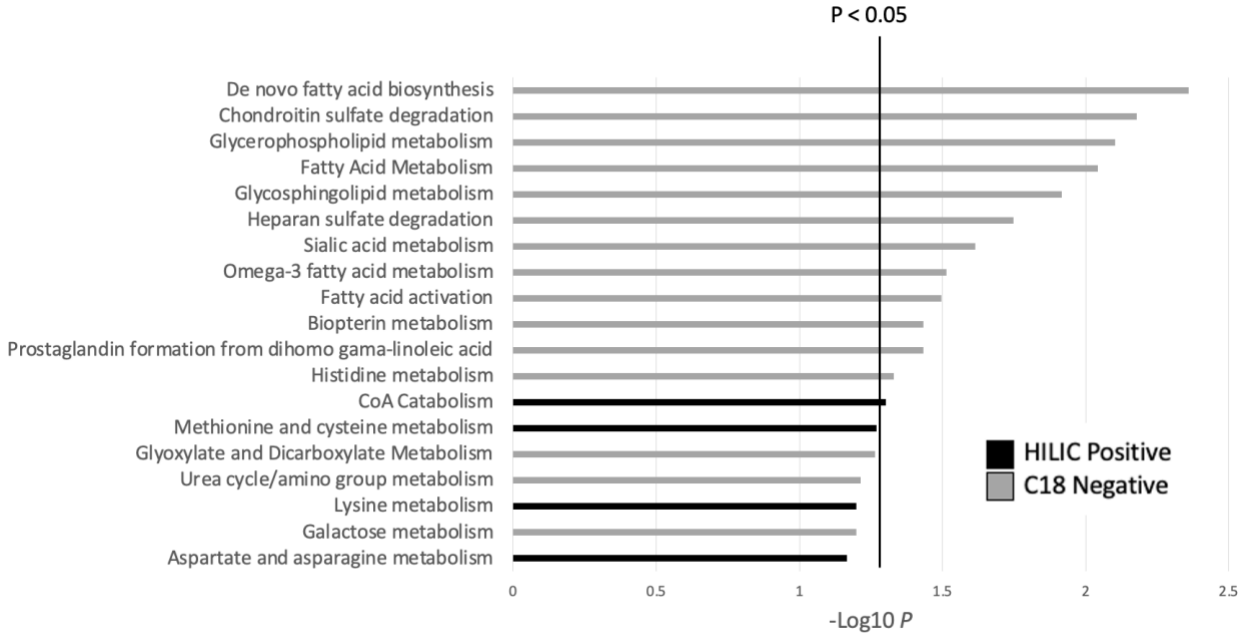

Figure 11 Pathway enrichment analysis ($p < 0.05$) using *mummichog*. The y-axis shows metabolic pathways significantly associated with metabolome wide alterations associated with 5-MIAA. The x-axis shows the -log p-value for pathway enrichment. Results are shown for both HILIC positive and C18 negative ionization modes. De novo fatty acid biosynthesis was the most significantly enhanced pathway.

4. Discussion

4.1 *Key results (Targeted and untargeted analysis of VB)*

δ-Valerobetaine (VB), a gut microbial-derived diet dependent obesogen, has been shown to disrupt mitochondrial fatty acid oxidation via inhibition of the carnitine shuttle in the liver and positively correlate with visceral adipose tissue mass and body mass index in humans.[15] Therefore, we suspected higher levels of VB in patients with higher scores of liver fibrosis (LSM). In this study, there were no significant differences in plasma levels of VB between patients with advanced and non-advanced fibrosis. Additionally, VB was not associated with liver fibrosis in the regression analyses. However, the MWAS linked VB to several metabolites and pathways previously associated with the pathogenesis of NAFLD. Results from pathway analysis revealed significant associations between plasma levels of VB and limonene degradation, lysine and glutathione metabolism, and vitamin B1 (thiamine) metabolism. Limonene degradation has been associated with gut microbial metabolism and administration of D-limonene has been shown to precipitate liver fibrosis in the murine model.[29-31] Lysine and glutathione metabolism are notably altered in patients with NAFLD.[32,33] Further, multiple studies have linked alterations in amino acid metabolism directly to imbalances in human gut microbiota.[34,35] Alterations in vitamin B1 (thiamine) metabolism have been implicated in hepatic fat accumulation, a key pathogenic contributor to progression of liver fibrosis in patients with NAFLD.[36] These findings support our hypothesized importance of the microbiome-gut-liver axis in NAFLD and liver fibrosis.

Metabolites significantly associated with VB in the MWAS included a variety of lipids (two acylcarnitines, one fatty acid ester) and an NAD precursor. Metabolic disturbance of lipids is a well-documented hallmark of NAFLD and a known physiologic effect of VB.[37,38] Additionally,

prior studies have shown that patients with high levels of hepatic steatosis have an altered demand for NAD, which could explain the observed association with VB in our study.[33]

These findings suggest VB is metabolically linked to various pathways and metabolites implicated in liver fibrosis and NAFLD. In this cohort of patients with pre-existing liver fibrosis, it is unclear whether changes in VB level are due to metabolome wide alterations from liver fibrosis, or if VB precedes these underlying changes. While VB did not demonstrate any significant associations with severity of liver fibrosis in those who already have NAFLD, future studies should aim to explore the temporality of this relationship.

4.2 *Key results (Targeted and untargeted analysis of 5-MIAA)*

Previous studies have demonstrated that 5-MIAA is functionally protective against oxidative liver injury, decreases insulin resistance, and improves lipid metabolism.[14,39] Therefore, we expected increased plasma levels of 5-MIAA in patients with lower LSM scores. In this study, there were no significant differences in level of 5-MIAA between patients with advanced and non-advanced fibrosis and 5-MIAA was not a strong predictor of liver fibrosis in regression analyses. However, the MWAS linked 5-MIAA to metabolites and pathways previously documented to play important roles in NAFLD.

Results from pathway analysis revealed significant associations between plasma levels of 5-MIAA and lipid metabolism, chondroitin sulfate degradation, and biopterin metabolism. Together, the widespread changes in numerous pathways of lipid metabolism is consistent with the known metabolic disturbance of lipids as a hallmark of NAFLD.[32] Chondroitin sulfate degradation has been linked to gut microbial metabolism, lipogenesis, and systemic inflammation. These metabolic processes are key factors in the pathogenesis of NAFLD and further highlight the underlying implications of gut microbial metabolism in hepatic disease.[40] Biopterin metabolism is

associated with the hepato-protective antioxidant response (Nrf2/ARE pathway).[41] Nrf2/ARE plays a key role in the regulation of hepatic inflammation, detoxification, and development of fibrosis.[42,43] Thus, our data show 5-MIAA is involved in a robust metabolic network of lipid metabolism, inflammatory regulation, and antioxidant response, all of which have contributory roles in the underlying pathogenesis of NAFLD and its progression.

Metabolites significantly associated with 5-MIAA in the MWAS primarily included dicarboxylic acids (caffeic acid, 33-thiobispropanoic acid) and a variety of fatty acids (glycerophospholipids, fatty acylcarnitines, phosphosphingolipids, keto acids). Dicarboxylic acids have been associated with hepatic lipid deposition and mitochondrial fatty acid oxidation.[44] Specifically, caffeic acid, a metabolite derived from plant digestion, has been found to regulate lipid accumulation in the murine model, attenuating gut microbiota dysbiosis and protecting against hepatic steatosis in high fat diets.[45,46] Studies using lipidomic approaches have demonstrated that patients with NAFLD have altered levels of glycerophospholipids and phosphosphingolipids, two major membrane lipids with intracellular signaling capacity, when compared to healthy controls.[47,48] Fatty acids and their metabolites are directly implicated in NAFLD pathogenesis, with previous findings showing altered ratios of fatty acid composition and disrupted fatty acid metabolism in patients with NASH.[47,49-51] Also of interest to this study, research has shown that differences in free fatty acid profiles are directly linked to changes in gut microbiota composition.[52,53]

Taken together, these findings suggest 5-MIAA is metabolically linked to many important pathways implicated in NAFLD and that further study of the role of gut microbial metabolism in the pathogenesis and progression of NAFLD is warranted. As stated previously, all patients in this cohort had pre-existing liver fibrosis, and therefore it remains unknown whether differences in

plasma 5-MIAA may precede progression to cirrhosis. This should be a target for future studies, along with assessment of 5-MIAA as a reliable biomarker of hepatic inflammatory response and to further discern its role in host-microbiome cross talk.

4.3 *Limitations*

Limitations of this study include the cross-sectional design, which limits the capacity for causal inference and relatively small sample size, which limits generalizability. Future, larger studies interrogating similar hypotheses should include longitudinal sampling to allow for investigation of temporal associations between VB and 5-MIAA and the severity of liver fibrosis. Specifically, this will allow for measurement of these metabolites before the development of liver fibrosis, as our study only included patients with pre-existing NAFLD. Another limitation derives from the collection of samples from plasma, which places measured metabolites at risk of biotransformation in liver. This potentially makes associations harder to detect in humans when compared to murine models, where samples can be obtained from the portal vein, which feeds gut-derived metabolites into the liver. Another limitation is that our study lacked a healthy control group. Importantly, future studies should include healthy participants without NAFLD, as no studies of plasma 5-MIAA and VB have been previously documented in patients with NAFLD versus those without this metabolic disorder.

4.4 *Interpretation*

While targeted metabolites did not significantly differ between levels of fibrosis, the MWAS for each metabolite suggests these metabolites are associated with metabolic networks previously shown to have biologic importance in NAFLD. Further, many metabolites detected in the MWAS were of unknown origin, which could represent a broad network of previously

unexplored metabolic pathways associated in NAFLD. This supports the notion of gut microbial cross talk at the host-microbe interface, and the role of the gut microbiome in human disease. This study has raised important questions regarding mechanistic relationships between gut microbial metabolism and human metabolic activity that should be further explored in future studies.

4.5 *Generalizability*

Without a healthy control group, the findings in this study are limited in their clinical applicability and external validity. However, the results of this study do provide a general strategy for the use of high-resolution metabolomics for the interrogation of gut-microbiome metabolism in various disease states. Future studies including a healthy control group will allow for more developed clinical and translational insights using the conceptual framework demonstrated in this study.

5. Appendix

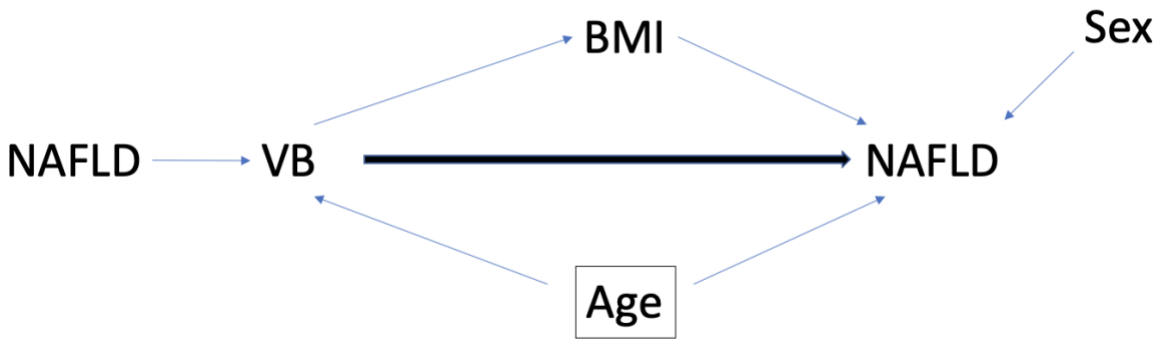

Figure A.1 Directed Acyclic Graph (DAG) for the assessment of covariate relationships with plasma VB (primary exposure) and NAFLD (primary outcome). Age was on an open backdoor path and thus was considered as a potential confounder by DAG theory and included in the model based on these assumptions.

A. Includes total cholesterol, LDL-C, and triglycerides.

Figure A.2 Directed Acyclic Graph (DAG) for the assessment of covariate relationships with plasma 5-MIAA (primary exposure) and NAFLD (primary outcome). There were no covariates on open backdoor paths in the DAG and thus no covariates were considered confounders by DAG theory and included in the model based on these assumptions.

6. References

1. Chen F, Stappenbeck TS. Microbiome control of innate reactivity. *Curr Opin Immunol.* 2019;56:107-113.
2. Curro D. The role of gut microbiota in the modulation of drug action: a focus on some clinically significant issues. *Expert Rev Clin Pharmacol.* 2018;11(2):171-183.
3. Klaassen CD, Cui JY. Review: Mechanisms of How the Intestinal Microbiota Alters the Effects of Drugs and Bile Acids. *Drug Metab Dispos.* 2015;43(10):1505-1521.
4. Levy M, Thaiss CA, Zeevi D, et al. Microbiota-Modulated Metabolites Shape the Intestinal Microenvironment by Regulating NLRP6 Inflammasome Signaling. *Cell.* 2015;163(6):1428-1443.
5. Maurice CF, Haiser HJ, Turnbaugh PJ. Xenobiotics shape the physiology and gene expression of the active human gut microbiome. *Cell.* 2013;152(1-2):39-50.
6. Schroeder BO, Backhed F. Signals from the gut microbiota to distant organs in physiology and disease. *Nat Med.* 2016;22(10):1079-1089.
7. Sharon G, Garg N, Debelius J, Knight R, Dorrestein PC, Mazmanian SK. Specialized metabolites from the microbiome in health and disease. *Cell Metab.* 2014;20(5):719-730.
8. Liu R, Hong J, Xu X, et al. Gut microbiome and serum metabolome alterations in obesity and after weight-loss intervention. *Nat Med.* 2017;23(7):859-868.
9. Ning G. Decade in review-type 2 diabetes mellitus: At the centre of things. *Nat Rev Endocrinol.* 2015;11(11):636-638.
10. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature.* 2006;444(7122):1027-1031.
11. Byrne CD, Targher G. NAFLD: a multisystem disease. *J Hepatol.* 2015;62(1 Suppl):S47-64.
12. Fabbrini E, Sullivan S, Klein S. Obesity and nonalcoholic fatty liver disease: biochemical, metabolic, and clinical implications. *Hepatology.* 2010;51(2):679-689.
13. Tarao K, Nozaki A, Ikeda T, et al. Real impact of liver cirrhosis on the development of hepatocellular carcinoma in various liver diseases-meta-analytic assessment. *Cancer Med.* 2019;8(3):1054-1065.
14. Saeedi BJ, Liu KH, Owens JA, et al. Gut-Resident Lactobacilli Activate Hepatic Nrf2 and Protect Against Oxidative Liver Injury. *Cell Metab.* 2020;31(5):956-968 e955.
15. Liu KH, Owens JA, Saeedi B, et al. Microbial metabolite delta-valerobetaine is a diet-dependent obesogen. *Nat Metab.* 2021;3(12):1694-1705.
16. Ratziu V, Sanyal AJ, Loomba R, et al. REGENERATE: Design of a pivotal, randomised, phase 3 study evaluating the safety and efficacy of obeticholic acid in patients with fibrosis due to nonalcoholic steatohepatitis. *Contemp Clin Trials.* 2019;84:105803.
17. Siddiqui MS, Vuppalanchi R, Van Natta ML, et al. Vibration-Controlled Transient Elastography to Assess Fibrosis and Steatosis in Patients With Nonalcoholic Fatty Liver Disease. *Clin Gastroenterol Hepatol.* 2019;17(1):156-163 e152.
18. Soltow QA, Strobel FH, Mansfield KG, Wachtman L, Park Y, Jones DP. High-performance metabolic profiling with dual chromatography-Fourier-transform mass spectrometry (DC-FTMS) for study of the exposome. *Metabolomics.* 2013;9(1 Suppl):S132-S143.

19. Yu T, Park Y, Johnson JM, Jones DP. apLCMS--adaptive processing of high-resolution LC/MS data. *Bioinformatics.* 2009;25(15):1930-1936.
20. Uppal K, Soltow QA, Strobel FH, et al. xMSanalyzer: automated pipeline for improved feature detection and downstream analysis of large-scale, non-targeted metabolomics data. *BMC Bioinformatics.* 2013;14:15.
21. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.* 2007;8(1):118-127.
22. Uppal K, Walker DI, Jones DP. xMSannotator: An R Package for Network-Based Annotation of High-Resolution Metabolomics Data. *Anal Chem.* 2017;89(2):1063-1067.
23. Wishart DS, Guo A, Oler E, et al. HMDB 5.0: the Human Metabolome Database for 2022. *Nucleic Acids Res.* 2022;50(D1):D622-D631.
24. Schymanski EL, Jeon J, Gulde R, et al. Identifying small molecules via high resolution mass spectrometry: communicating confidence. *Environ Sci Technol.* 2014;48(4):2097-2098.
25. Lipsky AM, Greenland S. Causal Directed Acyclic Graphs. *JAMA.* 2022.
26. Uppal K, Walker DI, Liu K, Li S, Go YM, Jones DP. Computational Metabolomics: A Framework for the Million Metabolome. *Chem Res Toxicol.* 2016;29(12):1956-1975.
27. Benjamini Y, Cohen R. Weighted false discovery rate controlling procedures for clinical trials. *Biostatistics.* 2017;18(1):91-104.
28. Li S, Park Y, Duraisingham S, et al. Predicting network activity from high throughput metabolomics. *PLoS Comput Biol.* 2013;9(7):e1003123.
29. Jing L, Zhang Y, Fan S, et al. Preventive and ameliorating effects of citrus D-limonene on dyslipidemia and hyperglycemia in mice with high-fat diet-induced obesity. *Eur J Pharmacol.* 2013;715(1-3):46-55.
30. Marmulla R, Harder J. Microbial monoterpene transformations-a review. *Front Microbiol.* 2014;5:346.
31. Ramos CAF, Sá RdCdS, Alves MF, et al. Histopathological and biochemical assessment of d-limonene-induced liver injury in rats. *Toxicol Rep.* 2015;2:482-488.
32. Goffredo M, Santoro N, Trico D, et al. A Branched-Chain Amino Acid-Related Metabolic Signature Characterizes Obese Adolescents with Non-Alcoholic Fatty Liver Disease. *Nutrients.* 2017;9(7).
33. Mardinoglu A, Bjornson E, Zhang C, et al. Personal model-assisted identification of NAD(+) and glutathione metabolism as intervention target in NAFLD. *Mol Syst Biol.* 2017;13(3):916.
34. Henao-Mejia J, Elinav E, Jin C, et al. Inflammasome-mediated dysbiosis regulates progression of NAFLD and obesity. *Nature.* 2012;482(7384):179-185.
35. Mardinoglu A, Shoaie S, Bergentall M, et al. The gut microbiota modulates host amino acid and glutathione metabolism in mice. *Mol Syst Biol.* 2015;11(10):834.
36. Kalyesubula M, Mopuri R, Asiku J, et al. High-dose vitamin B1 therapy prevents the development of experimental fatty liver driven by overnutrition. *Dis Model Mech.* 2021;14(3).
37. Enooku K, Nakagawa H, Fujiwara N, et al. Altered serum acylcarnitine profile is associated with the status of nonalcoholic fatty liver disease (NAFLD) and NAFLD-related hepatocellular carcinoma. *Sci Rep.* 2019;9(1):10663.

38. Fabbrini E, Mohammed BS, Magkos F, Korenblat KM, Patterson BW, Klein S. Alterations in adipose tissue and hepatic lipid kinetics in obese men and women with nonalcoholic fatty liver disease. *Gastroenterology.* 2008;134(2):424-431.
39. Ji Y, Gao Y, Chen H, Yin Y, Zhang W. Indole-3-Acetic Acid Alleviates Nonalcoholic Fatty Liver Disease in Mice via Attenuation of Hepatic Lipogenesis, and Oxidative and Inflammatory Stress. *Nutrients.* 2019;11(9).
40. Pessentheiner AR, Ducasa GM, Gordts P. Proteoglycans in Obesity-Associated Metabolic Dysfunction and Meta-Inflammation. *Front Immunol.* 2020;11:769.
41. McNeill E, Crabtree MJ, Sahgal N, et al. Regulation of iNOS function and cellular redox state by macrophage Gch1 reveals specific requirements for tetrahydrobiopterin in NRF2 activation. *Free Radic Biol Med.* 2015;79:206-216.
42. Xu D, Xu M, Jeong S, et al. The Role of Nrf2 in Liver Disease: Novel Molecular Mechanisms and Therapeutic Approaches. *Front Pharmacol.* 2019;9:1428-1428.
43. Tang W, Jiang Y-F, Ponnusamy M, Diallo M. Role of Nrf2 in chronic liver disease. *World J Gastroenterol.* 2014;20(36):13079-13087.
44. Zhang X, Gao T, Deng S, et al. Fasting induces hepatic lipid accumulation by stimulating peroxisomal dicarboxylic acid oxidation. *J Biol Chem.* 2021;296:100622.
45. Kim HM, Kim Y, Lee ES, Huh JH, Chung CH. Caffeic acid ameliorates hepatic steatosis and reduces ER stress in high fat diet–induced obese mice by regulating autophagy. *Nutrition.* 2018;55-56:63-70.
46. Mu H-N, Zhou Q, Yang R-Y, et al. Caffeic acid prevents non-alcoholic fatty liver disease induced by a high-fat diet through gut microbiota modulation in mice. *Food Research International.* 2021;143:110240.
47. Chiappini F, Coilly A, Kadar H, et al. Metabolism dysregulation induces a specific lipid signature of nonalcoholic steatohepatitis in patients. *Sci Rep.* 2017;7:46658.
48. Gorden DL, Myers DS, Ivanova PT, et al. Biomarkers of NAFLD progression: a lipidomics approach to an epidemic. *J Lipid Res.* 2015;56(3):722-736.
49. Araya J, Rodrigo R, Videla LA, et al. Increase in long-chain polyunsaturated fatty acid n - 6/n - 3 ratio in relation to hepatic steatosis in patients with non-alcoholic fatty liver disease. *Clin Sci (Lond).* 2004;106(6):635-643.
50. Koo SH. Nonalcoholic fatty liver disease: molecular mechanisms for the hepatic steatosis. *Clin Mol Hepatol.* 2013;19(3):210-215.
51. Puri P, Baillie RA, Wiest MM, et al. A lipidomic analysis of nonalcoholic fatty liver disease. *Hepatology.* 2007;46(4):1081-1090.
52. Rodríguez-Carrio J, Salazar N, Margolles A, et al. Free Fatty Acids Profiles Are Related to Gut Microbiota Signatures and Short-Chain Fatty Acids. *Frontiers in Immunology.* 2017;8.
53. Kindt A, Liebisch G, Clavel T, et al. The gut microbiota promotes hepatic fatty acid desaturation and elongation in mice. *Nature Communications.* 2018;9(1):3760.