UNIVERSITY OF CALIFORNIA SAN DIEGO

Heritability and Shared Gene Effects of Serum Metabolites Associated with Hepatic Steatosis and Fibrosis

A Thesis submitted in partial satisfaction of the

requirements for the degree

Master of Science

in

Biology

by

Amy Liu

Committee in Charge:

Professor Rohit Loomba, Chair
Professor Gen-Sheng Feng, Co-Chair
Professor James T Kadonaga

2020

The Thesis of Amy Liu is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

Co-Chair

_____

Chair

University of California San Diego

2020

TABLE OF CONTENTS

Signature Page……..… iii

Table of Contents…..……...…… iv

List of Abbreviations……..…… vii

List of Figures…… ix

List of Tables……. x

Acknowledgements…… xi

Abstract of the Thesis…..……...…….……... xii

Chapter 1 Introduction……...…… 1

1.1 The Anatomy and Function of the Liver……. 3

1.2 What is Non-Alcoholic Fatty Liver Disease (NAFLD)?..... 4

1.2.1 The Difference Between Non-Alcoholic Fatty Liver Disease and Alcoholic Fatty Liver Disease……..…….. 6

1.3 Non-Alcoholic Fatty Liver Disease Progression……. 7

1.4 Current Diagnostic Tools for Liver-Related Disease…….. 9

1.4.1 Invasive Diagnostic Tools…… 10

1.4.2 Noninvasive Diagnostic Tools……. 10

1.5 Non-Alcoholic Fatty Liver Disease is a Heritable Trait…….. 13

1.5.1 Heritability of Hepatic Steatosis and Hepatic Fibrosis……. 14

1.5.2 Shared Gene Effects Between Hepatic Steatosis and Fibrosis and with Metabolic Risk Factors…….. 14

1.6 Genetic Variants Associated with NAFLD……. 15

1.7 Metabolome Profiling of Non-Alcoholic Fatty Liver Disease (NAFLD)…...… 16

1.8 Serum Metabolites of Non-Alcoholic Fatty Liver Disease (NAFLD)…… 16

1.9 Aim and Objectives……. 17

Chapter 2 Methodology

2.1 Participants and Design of the Twin and Family Cohort…… 18

2.1.1 Inclusion Criteria……. 19

2.1.2 Exclusion Criteria…… 19

2.2 Clinical Assessment and Laboratory Test……..... 20

2.3 Magnetic Resonance Imaging Assessment……..... 21

2.3.1 Magnetic Resonance Imaging- Proton Density Fat Fraction (MRI-PDFF) 21

2.3.2 Magnetic Resonance Elastography (MRE) Protocol…… 22

2.4 Untargeted Metabolome Profiling…….…….... 23

2.4.1 Sample Accessioning…….. 24

2.4.2 Sample Preparation …… 24

2.4.3 Quality Assessment (QA) and Quality Control (QC)…….. 25

2.4.4 Ultrahigh Performance Liquid Chromatography-Tandem Mass Spectroscopy (UPLC-MS/MS)……..……. 26

2.4.5 Gas Chromatography/Mass Spectrometry (GC/MS)…….. 27

2.4.6 Bioinformatics…… 27

2.4.7 Lipidomic Platform…… 28

2.5 Statistical Analysis

2.5.1 Data Preparation……. 28

2.5.2 Data Analysis……. 29

2.5.3 Heritability Estimates and Shared Gene Effect……. 30

2.5.4 Sample Size Estimation……. 31

Chapter 3 Results

3.1 Characteristics of the Twin and Family Cohort……..... 33

3.2 Serum Metabolites Associated with Hepatic Steatosis and Fibrosis in the Twin and Family Cohort……. 36

3.3 Heritability of Serum Metabolites in the Twin and Family Cohort…….. 36

3.4 Shared Gene Effects between Serum Metabolites and Hepatic Steatosis and Fibrosis 37

3.5 Pathways Associated with the Heritable Serum Metabolites……..... 42

3.6 Subpathways of the 57 Serum Metabolites with Shared Gene Effect with Hepatic Steatosis and Fibrosis……..…… 44

Chapter 4 Discussion……... 48

4.1 Strengths and Limitations……..……..... 49

4.2 Implications for Future Study……..... 50

References…….….. 52

LIST OF ABBREVIATIONS

NAFLD	Non-alcoholic fatty liver disease
NAFL	Non-alcoholic fatty liver
NASH	Non-alcoholic steatohepatitis
HCC	Hepatocellular carcinoma
MRI-PDFF	Magnetic Resonance Imaging-Proton Density Fat Fraction
MRE	Magnetic Resonance Elastography
kPa	Kilopascals
kHz	Kilohertz
AFLD	Alcoholic Fatty Liver Disease
SBP	Systolic Blood Pressure
DBP	Diastolic Blood Pressure
HDL	High Density Lipoprotein
LDL	Low-Density Lipoprotein
HOMA-IR	Homeostatic Model Assessment-Insulin Resistance
HbA1C	Hemoglobin A1c
BMI	Body Mass Index
ALT	Alanine Aminotransferase
AST	Aspartate Aminotransferase
GGT	Gamma-Glutamyl Transferase
INR	International Normalized Ratio
GWAS	Genome-Wide Association Studies
PNPLA-3	Patatin-like Phospholipase Domain-Containing Protein 3
TM6SF2	Transmembrane 6 Superfamily Member 2
SNP	Single Nucleotide Polymorphism
HIPPA	Health Insurance Portability and Accountability Act
MZ	Monozygotic

DZ	Dizygotic
HIV	Human Immunodeficiency Virus
GC/MS	Gas Chromatography/ Mass Spectrometry
LC	Liquid Chromatography
LC/MS	Liquid Chromatography-Mass Spectrometry
LC/MS/MS	Liquid Chromatography with Tandem Mass Spectrometry
UPLC-MS/MS	Ultrahigh Performance Liquid Chromatography-Tandem Mass Spectroscopy
LIMS	Laboratory Information Management System
QC	Quality Control
QA	Quality Assessment
HILIC	Hydrophilic Interaction Liquid Chromatography
rG	Shared Genetic Determination
rE	Shared Environmental Determination
SAM	S-Adenosyl Methionine
TCA Cycle	Tricarboxylic Acid Cycle

LIST OF FIGURES

Figure 1: The Progression of Nonalcoholic Fatty Liver Disease……..…… 7

Figure 2: Fibrosis Progression……..…… 8

Figure 3: Shared gene-effect between serum metabolites and (a) hepatic steatosis (b) hepatic fibrosis, and (c) hepatic steatosis and fibrosis. Shared gene-effect of the serum metabolites is presented in bar graphs as genetic determination estimates (rG) adjusted for age, sex and ethnicity colored based on eight metabolic superpathways…… 40

Figure 4: Summary of Serum Metabolite Heritability and Their Shared Gene Effect with Hepatic Steatosis and Fibrosis……... 43

Figure 5: 57 Serum Metabolites have a Shared Gene Effect with Hepatic Steatosis and Liver Fibrosis……..... 44

Figure 6: Amount of Serum Metabolites that have a Shared Gene Effect with Hepatic Steatosis (left), Hepatic Fibrosis (middle), and Hepatic Steatosis and Fibrosis (right)……... 46

LIST OF TABLES

Table 1: Baseline Characteristics between Non-NAFLD and NAFLD Individuals in the Twin and Family Cohort……..……..... 35

Table 2: Superpathways and Subpathways of the 57 Serum Metabolites that have a Shared Gene Effect with Hepatic Steatosis and Fibrosis……. 49

ACKNOWLEDGMENTS

I would like to acknowledge Professor Rohit Loomba for his support as the chair of my committee. Through multiple drafts and many long nights, his guidance has proved to be invaluable.

I would also like to acknowledge Dr. Cyrielle Caussy, a visiting scholar at our lab, without whom my research would have no doubt taken five times as long. It is their support that helped me in an immeasurable way.

This thesis contains material as it appears in "Link between gut microbiome derived metabolite and shared gene-effects with hepatic steatosis and fibrosis in NAFLD", Hepatology 68(3), 2018. Cyrielle Caussy, Cynthia Hsu, Min-Tzu Lo, Amy Liu, Ricki Bettencourt, Veeral H. Ajmera, Shirin Bassirian, Jonathan Hooker, Ethan Sy, Lisa Richards, Nicholas Schork, Bernd Schnabl, David A. Brenner, Claude B. Sirlin, Chi-Hua Chen, Rohit Loomba, Genetics of NAFLD in Twins Consortium. I would like to thank Dr. Cyrielle Caussy and Dr. Rohit Loomba for giving me the opportunity to be a co-author of this paper and to be involved with this project.

ABSTRACT OF THE THESIS

Heritability and Shared Gene Effects of Serum Metabolites Associated with Hepatic Steatosis and Fibrosis

by

Amy Liu

Master of Science in Biology

University of California San Diego, 2020

Professor Rohit Loomba, Chair
Professor Gen-Sheng Feng, Co-Chair

Studies have shown that hepatic steatosis and fibrosis are heritable and have a shared gene effect with metabolic risk factors, but the heritability of the serum metabolites associated with

NAFLD has not been assessed. The aim of this study was to investigate whether serum metabolites are heritable and if there are any shared gene effect with NAFLD. This study was a cross-sectional analysis of a prospective cohort of community-dwelling Twins and Family from southern California (156 subjects: 100 twins; 37 monozygotic, 13 dizygotic twin pairs; 56 sibling-sibling or parents-offspring). Hepatic steatosis was assessed using MRI-PDFF. Serum metabolite analysis was performed by Metabolon platform using UPLC-MS/MS and GC/MS. Serum metabolites associated with NAFLD were assessed using a Welch's t-test and a sampling permutation by running 1,000 random selections of the dataset for each serum metabolite was performed to assess the familial effect. The association of serum metabolites with NAFLD was confirmed using generalized estimating equations (PROC GENMOD) to account for intrapair correlation within twinship. An ACE model was used to estimate the heritability of the serum metabolites and the shared gene effect (R_G) between serum metabolites and NAFLD. Among 713 serum metabolites analyzed, 440 were heritable and of those, 57 had a shared gene effect with both hepatic steatosis and fibrosis; around 40% belonging to the amino acid superpathways. This study provides evidence that several serum metabolites associated with NAFLD are heritable and have shared gene effects suggesting that metabolites may be associated with the genetic susceptibility and pathogenesis of NAFLD.

Chapter 1

Introduction

Non-alcoholic fatty liver disease (NAFLD) is considered the leading cause of liver disease in the United States and worldwide.[12] Because NAFLD is currently a major cause of liver-related morbidity and mortality, there is a great need to characterize the heritability of NAFLD in order to identify patients who may be at risk for the disease, to improve the understanding of NAFLD pathogenesis, and to find potential noninvasive biomarkers that can diagnose and monitor NAFLD and NAFLD-related disease progression.

Previous seminal studies have shown that NAFLD is a heritable trait.[2] In addition, studies have demonstrated that hepatic steatosis is heritable and has a shared gene-effect with metabolic risk factors and liver fibrosis using a twin study design.[2] Although genetic variants associated with NAFLD located on PNPLA-3 and TM6SF2 genes, these common risk alleles do not account for all of the variance observed in NAFLD.[2]

Metabolome profiling has provided new insights into the molecular mechanisms of diseases including NAFLD. Several studies have identified specific metabolomics signatures associated with different stage of the disease in NAFLD. In addition, previous studies have shown a familial resemblance of the concentration of serum metabolites using twin and familial cohorts.[2] Furthermore, when combined within a twin study design, the heritability and shared gene-effect between serum metabolites and features of NAFLD allow further insight into the pathogenesis of NAFLD. If the gene regulation of a serum metabolite and features of NAFLD (e.g., hepatic

steatosis and fibrosis) significantly overlap, then targeting specific nodal points in the metabolite pathway could have a major therapeutic impact in NAFLD. In addition, the metabolite would also be a useful biomarker for the screening of patients at risk of advanced stage of NAFLD.[2]

The heritability of the serum metabolites associated with NAFLD has not been systematically assessed yet. Utilizing a well-characterized prospective cohort of community-dwelling twins and families, this project aimed to investigate whether serum metabolites are heritable and whether serum metabolites have a shared gene-effect with hepatic steatosis and fibrosis.

1.1. The Anatomy and Function of the Liver

The liver is an organ found on the upper right quadrant of the abdomen, beneath the diaphragm and on top of the stomach, right kidney, and intestines. The liver is unique due to the two distinct sources that supply blood to it: the hepatic artery which supplies oxygen rich blood and the hepatic portal vein which supplies nutrient rich blood.[3] The liver also has two main sections called lobes. Both lobes are made up of eight segments and these segments are made up of thousands of small lobes called lobules.[3] The lobules are connected to small ducts that connect with larger ducts to form the common hepatic duct. The hepatic duct transports material made by the liver cells called bile which helps to digest food to the gallbladder and duodenum, the first part of the small intestine.[3]

The liver is a vital organ in the human body and performs a variety of functions that help support metabolism, immunity, digestion, detoxification, and vitamin storage (to name a few).[3] One of the primary functions of the liver is to control the chemical levels in the blood. All the blood that leaves the stomach and intestine is processed (broken down) in the liver such that it creates nutrients for the body to use.[3] In addition, the liver metabolizes medicines in the blood into forms that are easier for the body to use as well as detoxification of xenobiotics in the blood. Additional important functions of the liver include:

- Creates bile, which carries away waste and break down fats in the small intestine during digestion
- Make cholesterol and proteins that help carry fats through the body
- Stores and release glucose as needed
- Iron and copper storage
- Make protein for blood plasma

- Regulates blood clotting
- Clears bilirubin to prevent jaundice, when the skin and eyes turn yellow
- Changes ammonia, a harmful substance, to urea, one of the end products of protein metabolism that is excreted in the urine[3]

The liver plays a plethora of functions. When it has broken down harmful substances, they are excreted into either bile or blood.[3] With bile by-products, they enter the intestine and eventually leave the body in bowel movements while blood by-products are filtered out by the kidneys and eventually leave the body in the form of urine.[3] Because of its role in multiple organ systems, it is prone to a variety of pathologies.

1.2. What is Non-Alcoholic Fatty Liver Disease (NAFLD)?

Non-alcoholic fatty liver (NAFLD) is recognized as one of the most common liver disease worldwide and is currently a major cause of liver-related morbidity and mortality.[2,4-7] To date, NAFLD has a global prevalence of 25%, and is increasing due to the increasing epidemic of obesity, Type 2 Diabetes, and metabolic syndrome which NAFLD has been shown to be closely associated with.[4] In the United States, 80-100 million adults are thought to have NAFLD.[2,6] NAFLD is classified into two major categories: nonalcoholic fatty liver (NAFL) and nonalcoholic steatohepatitis (NASH).[4] NAFL, also called simple steatosis, is considered the nonprogressive form of NAFLD. This is because in this form of the disease, the fat deposits in the liver rarely become inflamed and as a result, the chances of developing later stages of liver disease such as fibrosis or cirrhosis remain unlikely.[4,5] On the other hand, patients with NASH which affects 20%

of those who have NAFLD, the lipid deposits in the liver do become inflamed and as a result, can lead to the development of steatohepatitis (inflammation of fat in liver), fibrosis, cirrhosis, hepatocellular carcinoma (HCC) and liver related mortality.[4,5]

NAFLD is increasing recognized as the liver disease components of metabolic syndrome since clinically, NAFLD patients tend to be obese, with insulin resistance and/or type 2 diabetes, dyslipidemia, hypertriglycerides, and hypertension.[4] In fact, the prevalence of NAFLD in individuals with metabolic syndrome has been quite high such that NAFLD has been reported in over 76% of type 2 diabetics and over 90% in severely obese patients undergoing bariatric surgery.[4] Recent studies have also indicated that patients with obesity, diabetes mellitus, metabolic syndrome with hyperlipidemia/dyslipidemia make up the majority of patients with NAFLD and an even higher percentage of patients with these comorbidities have NASH.[4] In addition, diabetes in particular is a risk factor for progression to NASH, cirrhosis and mortality, and having poor glycemic control increases the risk of fibrosis in NASH.[4] The high burden of metabolic comorbidities associated with NAFLD and NASH in these recent studies have raised concern about the growing prevalence of NAFLD and higher risk the development of NASH.

Clinically, NAFLD is defined as a Magnetic Resonance Imaging- Proton Density Fat Fraction (MRI-PDFF) of $\geq 5\%$, which means that the measured average hepatic steatosis of the liver is $\geq 5\%$.[2] The diagnosis of NAFLD is considered in the absence of competing liver etiologies such as chronic viral hepatitis, steatohepatitis caused by known medications such as amiodarone or tamoxifen, and other chronic liver diseases such as autoimmune hepatitis, hemochromatosis, and Wilson's disease, or significant alcohol consumption.[4] Like other diseases such as viral hepatitis, hypertension, diabetes, NAFLD is considered a "silent killer" meaning that in most cases, people's symptoms are asymptomatic.[8] Many of those with NAFLD are unaware of the presence

or severity of their disease until symptoms develop as a result of progression of disease to later liver stages. It is no surprise because of this that NASH is now recognized as one of the leading causes of cirrhosis in adults in the United States and NASH-related cirrhosis is currently the second indication for liver transplantation in the US.[4]

1.2.1. The Difference Between Non-Alcoholic Fatty Liver Disease (NAFLD) and Alcoholic Fatty Liver Disease (AFLD)

Histologically, NAFLD is very similar to alcoholic fatty liver disease (AFLD) such as the main histologically features of NAFLD are similar to those of alcohol-induced liver disease and include steatosis, steatohepatitis, and varying degrees of fibrosis.[8] But one of the defining differences between a diagnosis of NAFLD and AFLD is the amount of alcohol consumed by that individual. In someone with NAFLD, there is no significant alcohol consumption. According to the United States Guidelines for NAFLD (endorsed as the American Association for the Study of Liver Diseases, American College of Gastroenterology, and American Gastroenterological Association NAFLD Guideline), significant alcohol use is defined as current or recent alcohol consumption of >21 drinks/week in men and >14 drinks/week in women[3] where one drink is approximately 10 grams of alcohol per one drink unit.[4]

1.3. Non-Alcoholic Fatty Liver Disease Progression

NAFLD is defined by an MRI-PDFF measuring ≥ 5% meaning that there is a ≥ 5% of fat buildup in cells of the liver. When there is just fat buildup in the liver, this is considered simple steatosis (NAFL) and for some this disease is nonprogressive and these individuals would rarely develop into NASH or later stage liver disease. However, 20% of these individuals can develop inflammation resulting in NASH and progression of this can lead to late stage liver disease such as cirrhosis as seen in Figure 1.1.

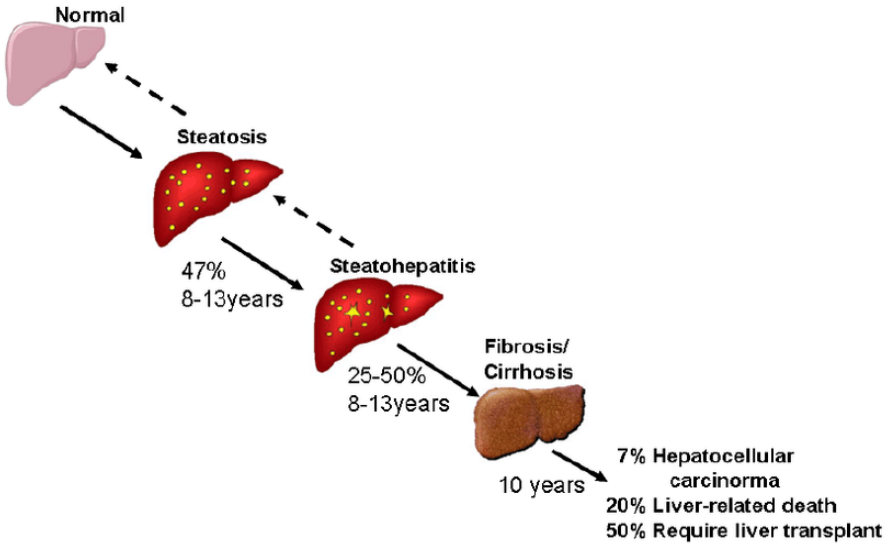


Figure 1: The Progression of Nonalcoholic Fatty Liver Disease.[9]

Studies have shown that the progression of NAFLD to NASH to fibrosis/cirrhosis takes an average of 24-30 years where to progress to each successive stage takes approximately 8-10 years.[9] Despite the slow progression of NAFLD, NAFLD continues to be an epidemic because it is asymptomatic. Many patients are unaware that they have the disease until complication from later stage liver disease such as cirrhosis develop. Later stage liver disease complications include ascites (fluid buildup in abdomen), portal hypertension gastropathy, esophageal varices, and hepatic encephalopathy (confusion). Once patients reach the cirrhotic stage of liver disease, there is an increased risk of developing HCC, having liver-related death due to complications, or requiring a liver transplantation.[9] The diagnosis of asymptomatic cirrhosis is usually made when incidental screening tests such as abnormal liver enzymes or radiological findings suggest liver disease.

Liver fibrosis (defined as MRE ≥ 3kPa) development and progression is the most important predictor of mortality in NAFLD.[5-6] It was found that compared to NAFLD patients with no fibrosis (stage 0), NAFLD patients with fibrosis were at an increased risk for all-cause mortality, and this risk increases as the fibrosis stages increases from stage 0 to stage 4.[4] The risk of all-cause mortality is higher with increasing fibrosis stage that even NAFLD patients with stage 1 fibrosis are at increased risk of mortality. The risk of liver related mortality increases on an exponential scale rather than on a linear scale with increase in the fibrosis stage. The risk of liver-related death is statistically higher only after progression to stage 2 fibrosis or higher.[4] Because of this increased risk of all-cause mortality with higher fibrotic stages, it is important to develop methods and tools that recognize the existence of fibrosis and prevent its progression.

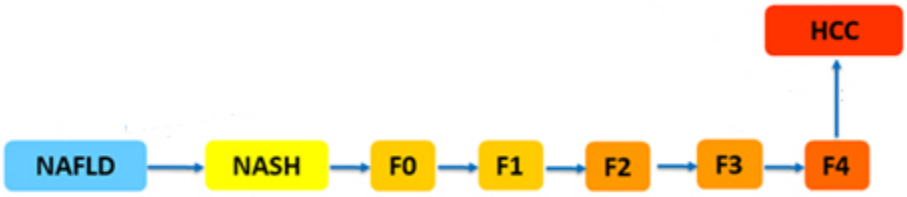


Figure 2: Fibrosis Progression[10]

Once NASH develops, there is a high risk of fibrosis (scar) developing. The development of fibrosis is the result of encapsulation or replacement of injured tissue by a collagenous scar.[10] Liver fibrosis results from the perpetuation of the normal wound healing response resulting in an abnormal continuation of fibrogenesis (connective tissue production and deposition).[10] Measuring the amount of fibrosis is called staging.[10] There are five stages: F0 indicates no scarring (no fibrosis); F1 indicates minimal scarring; F2 suggests that scarring has occurred and extended to multiple areas of the liver (significant fibrosis); F3 indicates fibrosis spreading and forming bridges with other fibrotic liver areas (severe fibrosis); F4 indicates cirrhosis.[10] The major clinical consequence of cirrhosis is decompensation of the liver such impaired hepatocyte (liver function), an increased intrahepatic persistence (portal hypertension), and the development of hepatocellular carcinoma (HCC).[10]

1.4. Current Diagnostic Tools for Liver-Related Disease

Correctly diagnosing and staging NAFLD and distinguishing the subset of patients that are likely to progress to NASH is not only critical for disease monitoring but also for potential implications for therapies.

As previously mentioned, NAFLD is a term that encompasses two types of patients: individuals with NAFL which is defined by an MRI-PDFF of ≥5% hepatic steatosis without incidence of hepatocellular injury and individuals with NASH, which is defined by an MRI-PDFF ≥5% hepatic steatosis and inflammation with hepatocellular injury (ballooning), with or without fibrosis.[1,11] Although the progression of NAFLD involves going from NAFL to NASH, disease progression likely involves a continuum with intermediate stages rather than a clear, distinct line that separates NAFL from NASH.[11] Furthermore, disease progression may not be linear and may take on stages of progression and regression.[11] Further disease progression among NASH patients involves development of fibrosis, cirrhosis, and cirrhosis-related complications such as hepatocellular carcinoma and end-stage liver.[9,11] Although accurately identifying NASH is important to disease monitoring, prognostication, and therapeutic considerations, no consistent biomarkers exist. In addition, despites advances in noninvasive diagnosis tools developed such as the MRI-PDFF, Fibroscan, and MRE (to name a few) to diagnose and monitor hepatic steatosis and fibrosis, these methods are limited and the liver biopsy continues to remain the gold standard for histologic diagnosis.[11]

1.4.1. Invasive Diagnostic Tools

The liver biopsy remains the gold standard for histologic diagnosis for NAFLD.[11] Although liver biopsy is the traditional and most widely accepted method of diagnosing NASH and staging fibrosis, its potential complications, invasiveness, sampling error, the vast cohort of patients who require disease staging, and the increasing -availability and accuracy of noninvasive methods have made liver biopsy less common than before. However, the primary reason why the liver biopsy still remains the gold standard for diagnosing and staging liver disease is due the limitations of noninvasive methods to assess different fibrotic stages.[11] The risk of liver-related mortality increase by a factor of 50 to 80 in patients going from stage three to stage four fibrosis and therefore, it is important to diagnose and stage the disease effectively.[11] Having limitations to distinguish between different fibrotic stages is unhelpful in advanced stage liver disease and therefore, practitioner turn to liver biopsy to confirm their diagnosis, evaluate the level of hepatocyte damage, and stage the fibrosis.[11]

1.4.2. Noninvasive Diagnostic Tools

The application of noninvasive approaches such as serum biomarkers and imaging modalities has come to the clinical forefront in helping with the diagnosis and monitoring of NAFLD and NASH-related fibrosis/cirrhosis. There is an increased risk of NAFLD in patients with insulin resistance, obesity, and metabolic syndrome, but it is not entirely clear which patient populations would benefit the most from screening for NAFLD and which diagnostic modality would be the most cost-effective.[1,11] According to the American Association for the Study of Liver Diseases, routine screening for NAFLD, even in high-risk patients, is not currently recommended

given the uncertainty of diagnostic tests, availability of treatments, and lack of clarity on long-term benefits and cost-effectiveness of screening.[11] Usually the diagnosis of NAFLD is due to incidental findings of elevations in laboratory enzymes or liver function tests or incidentally observations on imaging tests performed for other reasons.

When performing a lab panel, an elevation in the liver enzyme, aminotransferase (ALT), levels is the first indication suggesting the presence of hepatic steatosis, but advanced liver disease can present itself with normal aminotransferase levels and therefore, indicates the poor sensitivity and specificity of using ALT alone as a biomarker in diagnosing and staging NAFLD-related disease.[12,13] In addition, previous studies have suggested that normal ALT levels are found in 30% to 60% of patients with biopsy-confirmed hepatic steatosis.[11] Overall, elevation of ALT itself may also not be a strong predictor of risk of disease progression and therefore, there is a need to incorporate multiple tools to accurately assess disease severity.[11]

In addition, multiple radiologic techniques are available. Ultrasonography (US) is often the initial radiologic assessment performed to detect hepatic steatosis given its low cost and ease of accessibility.[11] However, US has been found to be suboptimal in patients with lower levels of hepatic steatosis or early-stage NAFL and therefore, this modality is better suited for those with moderate to severe hepatic steatosis.[11] In addition, computed tomography (CT) scans have not been shown to have improved sensitivities in mild steatosis, but carries additional radiation exposure and increased cost.[11] Thus, it might not be the most ideal modality for diagnosing NAFLD. The emerging new imaging tool used to diagnose NAFLD is MRI-PDFF which several studies have shown has been able to distinguish longitudinal changes in steatosis and fibrosis in the liver making MRI-PDFF a widely acceptable tool to diagnose NAFLD.[11] However, MRI machines and scans

can be expensive and because NAFLD is an asymptomatic disease, patients might not feel inclined to get a scan.

Furthermore, emerging radiologic technologies with high accuracy for assessing liver fibrosis continue to play increasingly important clinical roles in NASH patients. Vibration-controlled transient elastography (Fibroscan) is the most-studied and commonly accessible outpatient technology for identifying the severity of hepatic fibrosis, as well as one of the most reliable and accurate modalities in the noninvasive assessment of hepatic fibrosis.[11] However, the accuracy of this tool is limited by increasing abdominal and visceral adiposity as well as other factors that may affect potential assessment of liver stiffness (eg, abdominal ascites and severe active hepatitis).[11] Combining that limitation and that the majority of NAFLD patients are obese, the accuracy of this tool to access fibrosis in this population is further diminished. Another radiologic technology that has been developed to access liver fibrosis is the Magnetic resonance elastography (MRE). MRE is a highly accurate noninvasive diagnostic tool for quantifying fibrosis that is capable of being used in clinical trial outcomes and in disease monitoring in clinical practice.[15] Some studies have suggested that MRE is superior to Fibroscan, but is limited by cost and availability, and is typically offered only at major academic medical centers.[11]

The application of noninvasive approaches, including serum biomarkers and imaging modalities, has come to the clinical forefront. There are currently several noninvasive methods used to assess hepatic steatosis and fibrosis, and despite their limitations, they provide valuable information when evaluating patients with NAFLD.

1.5. Non-Alcoholic Fatty Liver Disease is a Heritable Trait

Previous seminal studies have shown that NAFLD is a heritable trait.[7,16]

As previously mentioned, obesity is a strong risk factor for the development of NAFLD. However, despite obesity being a risk factor, individuals with who are obese or even morbidly obese do not always develop NAFLD. There are even instances of NAFLD occurring in individuals of normal weight.[16] One explanation for these variations is that the development of NAFLD is possibly heritable. The fact that there have been many clinical cases and retrospective case studies of familial clustering of NAFLD in that patients with NAFLD were from families with multiple family members also showing NAFLD highly suggest heritability of NAFLD. In addition, familial aggregation studies of overweight children with and without NAFLD showed that liver fat fraction, and the condition of fatty liver are heritable traits.[16] In addition to familial aggregation of NAFLD, the phenotypic features associated with NAFLD may also be more prevalent in families with NAFLD.[16] Furthermore, there are racial and ethnic differences in the prevalence of NAFLD.[4,16] It was seen that patients of comparable body mass index (BMI) and insulin resistance, NAFLD was more likely to occur in Hispanic Americans than among African Americans.[16] And although environmental risk factors are likely to influence the development of NAFLD, aspects of the observed variation in NAFLD phenotype expression in persons with similar metabolic risk factors strongly implicate a genetic contribution.[16]

1.5.1. Heritability of Hepatic Steatosis and Hepatic Fibrosis

Prospective twin studies have established that hepatic steatosis and fibrosis are heritable.[17]

1.5.2. Shared Gene Effects Between Hepatic Steatosis and Fibrosis and with Metabolic Risk Factors

Prospective twin studies have established that there is a shared gene effect between hepatic steatosis and fibrosis.[1]

Prospective twin studies have shown that there are shared gene effect between hepatic steatosis and metabolic risk factors such as systolic blood pressure (SBP), diastolic blood pressure (DBP), high-density-lipoproteins (HDL) cholesterol, triglycerides, glucose, homeostatic model assessment- insulin resistance (HOMA-IR), insulin, and HbA1c.[1] Twin studies have also indicated that there is shared gene effect between hepatic fibrosis with metabolic risk factors.[17] Overall indicating that there is a genetic association between hepatic steatosis and metabolic risk factors and hepatic fibrosis and metabolic risk factors.

Shared environmental effects between hepatic steatosis and metabolic risk factors were also previously examined and it was found that there is no significant environmental effect between hepatic steatosis and other metabolic risk factors.[17] The same case was also found when examining the shared environmental effect between hepatic fibrosis and metabolic risk factors.[17]

Overall, these results suggest that genetic may potentially play a larger role in contributing to the development of hepatic steatosis and fibrosis than the environment.

1.6. Genetic Variants Associated with NAFLD

Genome-wide association studies (GWAS) have identified genetic variants associated with NAFLD located on patatin-like phospholipase domain containing-3 (PNPLA-3) and TM6SF2.[18-20] PNPLA-3 gene has been shown to be strongly influence the pathogenesis of NAFLD and affect the severity of the disease.[19] In a meta-analysis study of the influence of a I148M (rs738409 C/G) single nucleotide polymorphism (SNP) of PNPLA-3 on NAFLD, it was found that the variant of this gene exerted a strong influence on not only the liver fat accumulation but also on the susceptibility of a more aggressive disease.[19] Previous studies have also shown that PNPLA3 variants have been associated with obesity, ALT levels, and insulin sensitivity and secretion.[18]

In addition, a region on chromosome 19 (19p13) flanking, TM6SF2 (rs58542926 c.449 C>T, p.Glu167Lys) has been reported to be associated with NAFLD as well as plasma cholesterol, triglyceride, and low-density lipoprotein levels in several previous studies.[20] A recent study has also confirmed that TM6SF2 minor allele changes is higher in those with NAFLD and that carriage of each copy of the TM6SF2 rs58542926 C>T minor allele was associated with increased risk of steatosis.[20] Furthermore, TM6SF2 gene has a strong association between TM6SF2 rs58542926 and fibrosis stage.[20] Previous studies showed that those with mild fibrosis (F0-1) and advanced fibrosis (F2-4), carriage of each copy of the TM6SF2 rs58542926 C>T minor allele was associated consistently with a significant increased risk of advanced fibrosis.[20]

Hepatic steatosis is the initial stage in the development of NAFLD, whereas hepatic fibrosis is an important indicator of the morbidity and mortality of individuals with later stage liver disease.

The genes PNPLA3 and TM6SF2 have been shown to modify the risks of hepatic steatosis and fibrosis, and other genetic pathways associated with steatosis, fibrosis, and other metabolic traits.[18-20] Although, PNPLA3 and TM6SF2 explain genetic associations with the development of hepatic steatosis and fibrosis in NAFLD, the genes only explain a fraction of the variance and complexity that is seen in NAFLD. For instance, the PNPLA3 genotype only explains 10-12% of the variance of the trait.[17] Therefore, 90% of the variance in the trait remains to be elucidated.[17] Further associated factors must be analyzed to have a further understanding of the pathogenesis and development of NAFLD.

1.7. Metabolome Profiling of Non-Alcoholic Fatty Liver Disease (NAFLD)

Metabolome profiling has provided a greater understanding of the molecular mechanisms of diseases including NAFLD. Several studies have identified specific metabolomic signatures associated with different stages of the disease in NAFLD.[21-25]

1.8. Serum Metabolites of Non-Alcoholic Fatty Liver Disease (NAFLD)

The heritability of the concentration of circulating serum metabolites has been relevant to for determining whether they are suitable as biomarkers for disease. For example, if a metabolite was found to be 100% heritable then it is not likely for this metabolite to be a biomarker for diseases that are chiefly driven by environmental factors.[26] Familial studies have been performed

in order to assess the heritability of serum metabolites and help explain the heritability of metabolic diseases when levels of metabolites differ. The determination of the heritability of serum metabolite levels is also important because it contains the contributions of all genetic variants that can influence an individual's metabolite concentration and knowing this can therefore set an upper limit to the contribution of genetic variants.[26]

Several studies have demonstrated a familial resemblance of the concentration of serum metabolites using twin and familial cohorts.[26-30] In addition, when combined with a twin study design, the heritability and shared gene effect between serum metabolites and features of NAFLD provide further understanding of the pathogenesis of NAFLD. If the gene regulation of a serum metabolite and features of NAFLD such as hepatic steatosis and fibrosis significantly overlap, then targeting specific points in the metabolite pathway could have a major therapeutic impact in NAFLD. Furthermore, the metabolites would also be a useful biomarker for the screening of patients at risk of advanced stage of NAFLD.

1.9. Aim and Objectives

The heritability of serum metabolites associated with NAFLD have not been systemically assessed. The aims of my project were to use a well-characterized prospective cohort of community-dwelling twins and families in order to

- To investigate whether serum metabolites are heritable.
- To determine whether serum metabolites have a shared genetic effect with hepatic steatosis and fibrosis
- To identify metabolic pathways associated with NAFLD related serum metabolites

Chapter 2

Methodology

2.1. Participants and Design of the Twin and Family Cohort

This was a cross sectional analysis of a prospective cohort study of patients from the Twin and Family Study (ClinicalTrials.gov:NCT01643512) residing in southern California. This study included a total of 156 participants, 100 twins (50 twin pairs) including 37 pairs of monozygotic (MZ), 13 pairs of dizygotic (DZ) and 56 participants either siblings or parents-offspring recruited the University of California at San Diego (UCSD) NAFLD Research Center[1,17, 31-33] between December 2011 and January 2014.

All participants underwent a standardized clinical research visit including detailed medical history, alcohol quantification using Skinner and Audit questionnaire, anthropometric exam, physical exam, and biochemical testing at the University of California at San Diego (UCSD) NAFLD Research Center.[3-6] NAFLD was assessed clinically and quantified by magnetic-resonance-imaging proton-density-fat-fraction (MRI-PDFF) and liver fibrosis was assessed by magnetic resonance elastography (MRE) at the MR3T Research Laboratory. Research visits and imaging procedures were performed the same day for each pair of twins, parent-offspring or siblings. This study was Health Insurance Portability and Accountability Act (HIPAA) compliant

and was approved by the UCSD Institutional Review Board approval number: 111282. Informed written consent was obtained from each participant before enrolling in the study.

2.1.1 Inclusion Criteria

Patients were included if they were twins, siblings or parent-offspring at least 18 years old, willing and able to complete all research procedures and observations. For each twin pair, a detailed assessment of twinship status (ie, monozygotic (MZ) or dizygotic (DZ)) was obtained. The majority of twin-pairs[33] were diagnosed by their physician as either MZ or DZ by genetic testing. Furthermore, twin-ship status was confirmed by using a previously published questionnaire.[16]

2.1.2 Exclusion Criteria

Participants were excluded from the study if they met any of the following criteria: significant alcohol intake (>10 g/day in females or >20 g/day in males) for at least 3 consecutive months over the previous 12 months or if the quantity of alcohol consumed could not be reliably ascertained; clinical or biochemical evidence of liver diseases other than NAFLD (eg, viral hepatitis, HIV, coeliac disease, cystic fibrosis, autoimmune hepatitis); metabolic and/or genetic liver disease (eg, Wilson's disease, haemochromatosis, polycystic liver disease, alpha-1-antitrypsin deficiency, dysbetalipoproteinaemia); clinical or laboratory evidence of systemic infection or any other clinical evidence of liver disease associated with hepatic steatosis; use of drugs known to cause hepatic steatosis (eg amiodarone, glucocorticoids, methotrexate, L-asparaginase and valproic acid) for at least 3 months in the last past 6 months; history of bariatric

surgery; presence of systemic infectious illnesses; females who were pregnant or nursing at the time of the study; contraindications to MRI (eg metal implants, severe claustrophobia, body circumference greater than the imaging chamber); any other condition(s) which, based on the principal investigator's opinion, may significantly affect the participant's compliance, competence, or ability to complete the study.

2.2. Clinical Assessment and Laboratory Test

All participants underwent a standardized clinical research visit at the UCSD NAFLD Research Center. A detailed history was obtained from all participants. A physical exam, which included vital signs, height, weight, and anthropometric measurements, was performed by a trained clinical investigator. Body mass index was defined as the body weight (in kilograms) divided by height (in meters) squared. Alcohol consumption was documented outside clinical visits and confirmed in the research clinic using the Alcohol Use Disorders Identifications Test and the Skinner questionnaire. A detailed history of medications was obtained and no patient took medications known or suspected to cause steatosis or steatohepatitis. Other causes of liver disease and secondary causes of hepatic steatosis were systemically ruled out using detailed history and laboratory data. After completion of the earlier described elements of the history and physical examination, participants had the following fasting laboratory work: complete blood count, screening etiologic tests (hepatitis B, surface antigen, hepatitis C antibody, and iron panel including serum ferritin), clinical chemistry (creatinine, total protein, blood urea nitrogen, uric acid), hepatic panel (total bilirubin, direct bilirubin, aspartate aminotransferase, alanine aminotransferase, alkaline phosphatase, g-glutamyltransferase, albumin, prothrombin time, and international normalized ratio), lipid profile, hemoglobin A1c, and glucose-insulin levels.

2.3. Magnetic Resonance Imaging Assessment

MRI was performed at the UCSD MR3T Research Laboratory using the 3T research scanner (GE Signa EXCITE HDxt; GE Healthcare, Waukesha, WI). MRI-PDFF was used to measure hepatic steatosis and MRE was used to measure hepatic fibrosis.

2.3.1. Magnetic Resonance Imaging- Proton Density Fat Fraction (MRI-PDFF)

Protocols The MRI proton-density fat fraction protocol has been previously described and validated as a standardized and objective measure of liver fat content.[17, 19] Patients were scanned in a supine position using a 3T MR scanner (SIGNA Excite HDxt; GE Medical Systems, Milwaukee, WI, USA), with an 8-channel torso-phased array surface coil centered over the liver. Images were obtained once at baseline, and again at post-treatment. Noncontrast axial-magnitude MR images were obtained of the whole liver using a 2-dimensional spoiled gradient-recalled-echo sequence. A low flip-angle (10°) was used at a repetition time of more than 100 miliseconds (ms) to minimize T1 effects. Six fractional echo magnitude images were obtained at serial opposed-phase and in-phase echo times 1.15, 2.3, 3.45, 4.6, 5.75 and 6.9 ms in a single breathhold (12–24 s). Other imaging parameters included: 8–10-mm slice thickness, 14–26 slices covering the whole liver, 0-mm-slice gaps, 192 × 192 base matrix, 1 signal average and rectangular field of view adjusted to the body habitus and breath-hold capacity. By using a custom open-source software plug-in for Osirix (Pixmeo Co., Geneva, Switzerland) that corrects for exponential T2* decay and that incorporates a multipeak fat spectral model, MRI proton-density fat fraction parametric maps

were reconstructed offline from the source MR images. Circular regions of interest with a 1-cm radius were placed in each of the four right liver lobe segments (segments 5–8) on the proton-density fat fraction maps. Proton-density fat fraction values were recorded for each region of interest/segment, and a final right-lobe MRI proton-density fat fraction value for each participant was obtained by averaging the values of the four corresponding regions of interest.

2.3.2. Magnetic Resonance Elastography (MRE) Protocol

MRE was performed using previously published methods[18,20-22] using software and hardware available commercially from Resoundant Inc., Rochester, MN. Briefly, an acoustic active driver placed outside of the MRI room delivers continuous vibrations at 60 Hz to an acoustic passive driver attached with an elastic band to the body wall anterior to the liver. During the transmission of the vibrations, a 2D gradient-recalled-echo MRE pulse sequence is performed, leading to the acquisition of four non-contiguous axial slices, 10-mm thick with 10-mm interslice gaps, that were acquired in a 16-second breathhold through the widest transverse section of the liver. The following acquisition parameters were used: repetition time 50 ms, echo time 20.2 ms, flip angle 30 degrees, matrix 256 x 64, field of view 48 x 48 cm, one signal average, receiver bandwidth ± 33 kHz (confirm), and parallel imaging acceleration factor 2. The pulse sequence utilizes oscillating motion-sensitizing gradients which encode tissue motion into MR signal phases, thus generating wave images depicting the shear waves within the liver. The sequence is repeated four times, with the phase relationship (phase offset) between the vibrations and the oscillating motion-sensitizing gradients adjusted each time, leading to the production at each slice location of

wave images located at four evenly spaced time points over the wave cycle. The total acquisition time for a patient is approximately two minutes (with four 16-second long breathholds with short recovery time in between).

At each slice location, wave images are processed automatically on a scanner computer using specialized software that utilizes an inversion algorithm to produce quantitative, cross-sectional maps called elastograms depicting tissue stiffness. Four elastograms, one at each of the four slice locations, are generated. The elastograms are color maps that depict stiffness in different regions of the liver with a color scale in units of kilopascals (kPa). The elastograms are transferred offline for analysis by a single experienced image analyst in the MR3T research laboratory with at least six months of experience working with MRE. The image analyst uses a custom software package to manually draw regions of interest (ROI) on the elastograms. ROIs are drawn at each of four slice locations in areas of the liver where the corresponding wave images depict clearly observable wave propagation, and avoiding artifacts, large blood vessels, and liver edges. The mean liver stiffness was calculated by averaging the per-pixel stiffness values across ROIs at four slice locations. The final results are automatically outputted to an electronic spreadsheet.

2.4. Untargeted Metabolome Profiling

Serum metabolite assessment was performed by Metabolon, Inc (Durham, NC, USA). Samples were extracted and split into equal parts for analysis on the GC/MS and LC/MS/MS platforms.[23] Software was used to match ions to an in-house library of standards for metabolite identification and for metabolite quantitation by peak area integration.[24] A number of internal standards were added to each experimental and process standard sample just prior to injection into the mass spectrometers. A measure of the platform variability was determined by calculating the

median relative standard deviation (RSD) for these internal standards. Profiling of samples on the lipidomics platform was also carried out and that data was merged with results from the global profiling platform.

2.4.1. Sample Accessioning

Following receipt, samples were inventoried and immediately stored at -80oC. Each sample received was accessioned into the Metabolon LIMS system and was assigned by the LIMS a unique identifier that was associated with the original source identifier only. This identifier was used to track all sample handling, tasks, results, etc. The samples (and all derived aliquots) were tracked by the LIMS system. All portions of any sample were automatically assigned their own unique identifiers by the LIMS when a new task was created; the relationship of these samples was also tracked. All samples were maintained at -80oC until processed.

2.4.2. Sample Preparation

Samples were prepared using the automated MicroLab STAR® system from Hamilton Company. Recovery standards were added prior to the first step in the extraction process for QC purposes. Sample preparation was conducted using a methanol extraction to remove the protein fraction while allowing maximum recovery of small molecules. The resulting extract was divided into five fractions: one for analysis by UPLC-MS/MS with positive ion mode electrospray ionization, one for analysis by UPLC-MS/MS with negative ion mode electrospray ionization, one for LC polar platform, one for analysis by GC-MS, and one sample was reserved for backup. Samples were placed briefly on a TurboVap® (Zymark) under nitrogen to remove the organic

solvent. For LC, the samples were stored under nitrogen overnight. For GC, the samples were dried under vacuum overnight. Samples were then prepared for the appropriate instrument, either LC/MS or GC/MS.

2.4.3. Quality Assessment (QA) and Quality Control (QC)

Several types of controls were analyzed in concert with the experimental samples: a pooled matrix sample generated by taking a small volume of each experimental sample (or alternatively, use of a pool of well-characterized human plasma) served as a technical replicate throughout the data set; extracted water samples served as process blanks; and a cocktail of QC standards that were carefully chosen not to interfere with the measurement of endogenous compounds were spiked into every analyzed sample, allowed instrument performance monitoring and aided chromatographic alignment. Instrument variability was determined by calculating the median relative standard deviation (RSD) for the standards that were added to each sample prior to injection into the mass spectrometers. Overall process variability was determined by calculating the median RSD for all endogenous metabolites (i.e., non-instrument standards) present in 100% of the pooled matrix samples. Experimental samples were randomized across the platform run with QC samples spaced evenly among the injections.

2.4.4 Ultrahigh Performance Liquid Chromatography-Tandem Mass Spectroscopy (UPLC-MS/MS)

The LC/MS portion of the platform was based on a Waters ACQUITY ultra-performance liquid chromatography (UPLC) and a Thermo Scientific Q-Exactive high resolution/accurate mass spectrometer interfaced with a heated electrospray ionization (HESI-II) source and Orbitrap mass analyzer operated at 35,000 mass resolution. The sample extract was dried then reconstituted in acidic or basic LC-compatible solvents, each of which contained 8 or more injection standards at fixed concentrations to ensure injection and chromatographic consistency. One aliquot was analyzed using acidic positive ion optimized conditions and the other using basic negative ion optimized conditions in two independent injections using separate dedicated columns (Waters UPLC BEH C18-2.1x100 mm, 1.7 μm). Extracts reconstituted in acidic conditions were gradient eluted from a C18 column using water and methanol containing 0.1% formic acid. The basic extracts were similarly eluted from C18 using methanol and water, however with 6.5mM Ammonium Bicarbonate. The third aliquot was analyzed via negative ionization following elution from a HILIC column (Waters UPLC BEH Amide 2.1x150 mm, 1.7 μm) using a gradient consisting of water and acetonitrile with 10mM Ammonium Formate. The MS analysis alternated between MS and data-dependent MS2 scans using dynamic exclusion, and the scan range was from 80-1000 m/z. Raw data files are archived and extracted as described below.

2.4.5 Gas Chromatography/Mass Spectrometry (GC/MS)

The samples destined for GC/MS analysis were re-dried under vacuum desiccation for a minimum of 24 hours prior to being derivatized under dried nitrogen using bistrimethyl-silyl-trifluoroacetamide (BSTFA). The GC column was a 20m X 0.18 mm ID, with 5% phenyl; 95% dimethylsilicone phase. The temperature ramp was from 60° to 340° C in a 18 minute period. Samples were analyzed on a Thermo-Finnigan Trace DSQ fast-scanning single-quadrupole mass spectrometer using electron impact ionization at unit mass resolution. The instrument was tuned and calibrated for mass resolution and mass accuracy on a daily basis. The information output from the raw data files was automatically extracted as discussed below.

2.4.6. Bioinformatics

The informatics system consisted of four major components, the Laboratory Information Management System (LIMS), the data extraction and peak-identification software, data processing tools for QC and compound identification, and a collection of information interpretation and visualization tools for use by data analysts. The hardware and software foundations for these informatics components were the LAN backbone, and a database server running Oracle 10.2.0.1 Enterprise Edition.

2.4.7. Lipidomic Platform

TrueMass® Lipomic Panel: Lipids were extracted in the presence of authentic internal standards by the method of Folch et al. (J Biol Chem 226:497-509) using chloroform:methanol (2:1 v/v). Lipids were transesterified in 1% sulfuric acid in methanol in a sealed vial under a nitrogen atmosphere at 100°C for 45 minutes. The resulting fatty acid methyl esters were extracted from the mixture with hexane containing 0.05% butylated hydroxytoluene and prepared for GC by sealing the hexane extracts under nitrogen. Fatty acid methyl esters were separated and quantified by capillary GC (Agilent Technologies 6890 Series GC) equipped with a 30 m DB 88 capillary column (Agilent Technologies) and a flame ionization detector.

2.5. Statistical Analysis

2.5.1. Data Preparation

Among 1,181 serum metabolites detected, 411 unnamed serum metabolites were excluded from the analysis. Metabolites for which there were more than 50% missing values, most definitely due to levels below the limits of detection (57 out of 770 and 69 out of 721 identified metabolites in the discovery and validation cohort, respectively), were also excluded from the analysis. For the remaining serum metabolites with fewer than 50% missing values, values were imputed to half of the observed minimum value for each metabolite assuming that the metabolites were under the limit of detection. In the twin and family cohort, if both twins in a pair had a missing value for the same metabolites, the missing values were left blank to avoid bias heritability estimates and the

twin pairs were excluded from the heritability analysis. All the data were log-transformed prior to statistical analysis.

2.5.2. Data Analysis

Patients' demographic, anthropometric, clinical, and biochemical characteristics were summarized. Categorical variables were shown as counts and percentages, and associations were tested using a chi-squared test or Fisher's exact test. Normally distributed continuous variables were shown as mean (± standard deviation), and differences between groups were analyzed using a two-independent samples *t* test or Wilcoxon-Mann-Whitney test.

Statistical comparison of serum metabolites between NAFLD and non-NAFLD groups and between individuals with the highest MRE measurements >4.17 kPa (corresponding to the ninety-fifth percentile) and the individuals with the lowest measurements <1.67 kPa (corresponding to the ninety-fifth percentile) were assessed using Welch's *t* tests, and Pearson correlation between serum metabolites and hepatic steatosis assessed by MRI-PDFF and hepatic fibrosis assessed by MRE or liver biopsy were performed followed by pairwise partial correlation conditional on age, sex, Hispanic ethnicity and obesity. To further explore familial effects within the dataset, a sampling permutation median *P* value was calculated for each metabolite by running 1,000 random selections of the dataset and calculating the *P* value for all compounds, followed by calculating the median *P* value for each metabolite from the 1,000 simulations. Statistical analyses were performed using ArrayStudio and the programs R. The False Discovery Rate using the Benjamini-Hochberg method was used to account for multiple testing.

2.5.3. Heritability Estimates and Shared Gene-Effect

AE models were used to estimate the shared genetic determination (rG) and shared environmental determination (rE) between twin pairs. In the classical twin study of sets of monozygotic and dizygotic twins, four latent factors can account for the variance of any phenotype: additive genetic effects (A); nonadditive genetic effects, including dominance (D); common or shared environmental effects (C); and nonshared or individual-specific environmental effects (E).[25] Because monozygotic twins are presumed to be genetically identical, they correlate perfectly (r = 1.0) with respect to both additive and nonadditive genetic effects. Dizygotic twins share, on average, 50% of their genes, resulting in correlations of 0.50 for additive genetic effects and 0.25 for nonadditive genetic effects. The C term is defined as environmental factors that make twins similar; hence, common environmental factors correlate 1.0 across twin pairs, regardless of zygosity. The E term represents environmental factors that lead to differences between twins. Because these are individual-specific factors, they are assumed to be uncorrelated across twins. Error is assumed to be random across individuals, so measurement error forms part of the estimate of E in these analyses. These latent factors comprise what are referred to as the univariate ACE or ADE models; due to model underidentification, an ACDE model cannot be tested in the classical twin design.[25]

The ACE and ADE models are easily extended to the multivariate case.[25] In addition to genetic and environmental sources of variance, sources of covariance can be examined in the bivariate model. In the present study, we used bivariate models to compute genetic correlations between two phenotypes. A phenotypic correlation measures shared variance; a genetic correlation

measures shared genetic variance. More specifically, a phenotypic correlation is defined as the total covariance (genetic plus environmental) of two variables divided by the square root of the product of the total variance of variable 1 and the total variance of variable 2. After decomposing the sources of variance in the bivariate model, we computed genetic correlations. These are defined as the genetic covariance divided by the square root of the product of the genetic variance of variable 1 and the genetic variance of variable 2. The analyses were performed using OpenMx, a structural equation modeling software package for genetically informative data (http://openmx.psyc.virginia.edu). Prior to the model fitting, the measures were adjusted for controlling age, gender, and ethnicity. Overall, AE models tended to provide the best fits to the data. Consequently, the genetic effects estimated in these AE models refer to broad-sense heritability, reflecting the proportion of phenotypic variance accounted for by the combined effect of all genetic influences (A + D).

2.5.4. Sample Size Estimation

Previous studies have reported that the mean heritability estimates of circulating metabolites was 0.53, ranging from 0.21 to 0.77 in a Netherlands twin cohort[26] and the median heritability estimates of serum metabolites was 0.49, ranging from 0.23 to 0.76 in a UK twins cohort.[28] We have also previously estimated the heritability of hepatic steatosis to be approximately 0.5.[17] Therefore, we anticipated that the heritability of serum metabolite and hepatic steatosis with one another would be approximately 0.50. It has been shown that to detect an additive genetic component of 0.4-0.8 in an ACE model, approximately 36-74 twin pairs are needed to produce a power of 0.95 with an alpha value of 0.05.[34] Therefore, the 50 twin pairs

included in this study would be adequate to assess the heritability and shared gene-effect of serum metabolites and hepatic steatosis in this cohort.

Chapter 3

Results

3.1. Characteristics of the Twin and Family Cohort

This study included a total of 156 participants, 100 twins (50 twin pairs) including 37 pairs of MZ, 13 pairs of DZ, and 56 participants either siblings or parents-offspring who underwent serum metabolites assessment, clinical evaluation and advanced MRI assessment. The mean (± standard deviation) age and body mass index (BMI) was 46.3 years (± 19.8) and 26.6 kg/m^2 (± 6.0), respectively. The prevalence of NAFLD as defined by MRI-PDFF ≥5% was 23% (36/156) and the prevalence of hepatic fibrosis as defined by MRE ≥3 kPa was 18% (28/156). Detailed demographic, biochemical, and imaging data of subjects with NAFLD compared to subjects without NAFLD is provided in Table 1.

Table 1: Baseline Characteristics between Non-NAFLD and NAFLD Individuals in the Twin and Family Cohort[2]

Mean values are provided with the standard deviation in parenthesis, unless otherwise noted as n (%). Differences between individuals with and without NAFLD were evaluated with t tests or the Wilcoxon-Mann-Whitney test for continuous variables and the chi squared or Fisher exact test for categorical variables. Bold indicates significant *P*-values <0.05.[2]

Characteristics	**Overall** (n=156)	**NAFLD MRI-PDFF ≥ 5%** (n=36)	**Non-NAFLD MRI-PDFF < 5%** (n=120)	***P***
Relationship		22 Twins; 14 Other relatives (parent, offspring, sibling)	78 Twins; 42 Other relatives (parent, offspring, sibling)	
Age years	46.3 (19.8)	55.2 (15.7)	43.6 (20.2)	**0.001**
Female, n (%)	114 (73.1)	20 (55.6)	94 (78.3)	**0.007**
Race				0.477
White, n (%)	115 (73.7)	25 (69.4)	90 (75.0)	
Hispanic, n (%)	26 (16.7)	7 (19.4)	19 (15.8)	
BMI (kg/m^2)	26.6 (6.0)	31.8 (5.9)	25.0 (5.2)	**<0.0001**
SBP (mm Hg)	123.7 (21.4)	137.4 (22.6)	119.5 (19.3)	**<0.001**
DBP (mmHg)	76.6 (12.3)	81.0 (12.4)	75.2 (12.0)	0.13
Waist circumference (cm)	90.5 (13.4)	102.4 (13.2)	86.9 (11.2)	**<0.0001**
Hip Circumference (cm)	101.1 (11.9)	108.8 (13.5)	98.7 (10.34)	**<0.0001**
Glucose (mg/dL)	91.7 (16.7)	102.5 (27.5)	88.4 (9.6)	**0.005**
Insulin (U/L)	11.8 (21.8)	25.5 (42.7)	7.67 (4.2)	**0.019**
HbA1c	5.8 (0.5)	6.1 (0.7)	5.7 (0.4)	**0.001**
HOMA-IR	2.8 (5.9)	6.6 (11.4)	1.7 (1.0)	**0.016**
AST (U/L)	23.8 (9.8)	28.4 (15.8)	22.4 (6.6)	**0.035**
ALT (U/L	22.7 (14.8)	31.5 (19.6)	20.1 (11.9)	**0.002**
Alkaline phosphatase (U/L)	69.4 (22.8)	72.4 (25.2)	68.5 (22.1)	0.372
Total bilirubin (mg/dL)	0.5 (0.2)	0.43 (0.20)	0.46 (0.23)	0.423

Table 1: Baseline Characteristics between Non-NAFLD and NAFLD Individuals in the Twin and Family Cohort[2]

Mean values are provided with the standard deviation in parenthesis, unless otherwise noted as n (%). Differences between individuals with and without NAFLD were evaluated with t tests or the Wilcoxon-Mann-Whitney test for continuous variables and the chi squared or Fisher exact test for categorical variables. Bold indicates significant *P*-values <0.05.[2] Continued.

Characteristics	**Overall** (n=156)	**NAFLD MRI-PDFF ≥ 5%** (n=36)	**Non-NAFLD MRI-PDFF < 5%** (n=120)	***P***
Direct bilirubin (mg/dL)	0.1 (0.04)	0.12 (0.04)	0.12 (0.04)	0.540
Albumin (g/dL)	4.5 (0.3)	4.5 (0.3)	4.6 (0.3)	0.170
GGT (U/L)	25.2 (29.1)	44.7 (53.0)	19.3 (10.8)	**0.007**
Total cholesterol (mg/dL)	189.4 (40.3)	189.2 (39.8)	189.5 (40.6)	0.970
HDL-cholesterol (mg/dL)	63.1 (19.3)	49.6 (11.5)	67.2 (19.3)	**<0.0001**
LDL-cholesterol (mg/dL)	107.1 (35.0)	111.1 (35.1)	105.9 (35.1)	0.441
Triglycerides (mg/dL)	98.3 (61.3)	151.1 (70.3)	82.2 (48.2)	**<0.0001**
White blood cell count (x 10^3/uL)	6.0 (1.6)	6.8 (1.5)	5.8 (1.6)	**0.002**
Hemoglobin (g/dL)	13.8 (2.1)	14.6 (3.8)	13.6 (1.1)	0.116
Hematocrit (%)	40.5 (3.4)	41.2 (4.2)	40.3 (3.1)	0.232
Platelet count (x10^3/uL)	255.4 (51.2)	249.8 (57.8)	257.1 (49.1)	0.454
INR	1.1 (0.3)	1.1 (0.3)	1.1 (0.2)	0.924
Ferritin (ng/mL)	91.2 (89.3)	121.1 (136.5)	82.4 (67.8)	0.113
MRI-PDFF (%)	4.7 (5.4)	12.7 (6.3)	2.3 (0.9)	**<0.0001**
MRE (kPa)	2.4 (0.8)	3.2 (1.3)	2.1 (0.4)	**<0.0001**

3.2. Serum Metabolites Associated with Hepatic Steatosis and Fibrosis in the Twin and Family Cohort

Among the 713 metabolites analyzed, 153 belonging to eight super pathways (amino-acid, peptide, energy, lipid, carbohydrate, nucleotide, cofactor and vitamins and xenobiotics) were significantly differentially expressed in individuals with NAFLD (MRI-PDFF ≥5%) compared to individuals without NAFLD (MRI-PDFF <5%) and 86 belonging to eight super pathways (amino-acid, peptide, energy, lipid, carbohydrate, nucleotide, cofactor and vitamins and xenobiotics) were significantly differentially expressed in individuals with the higher MRE measurements (>4.17 kPa, ninety-fifth percentile) compared to individual with lower MRE measurements (<1.61 kPa, fifth percentile), after assessment of the familial effect.

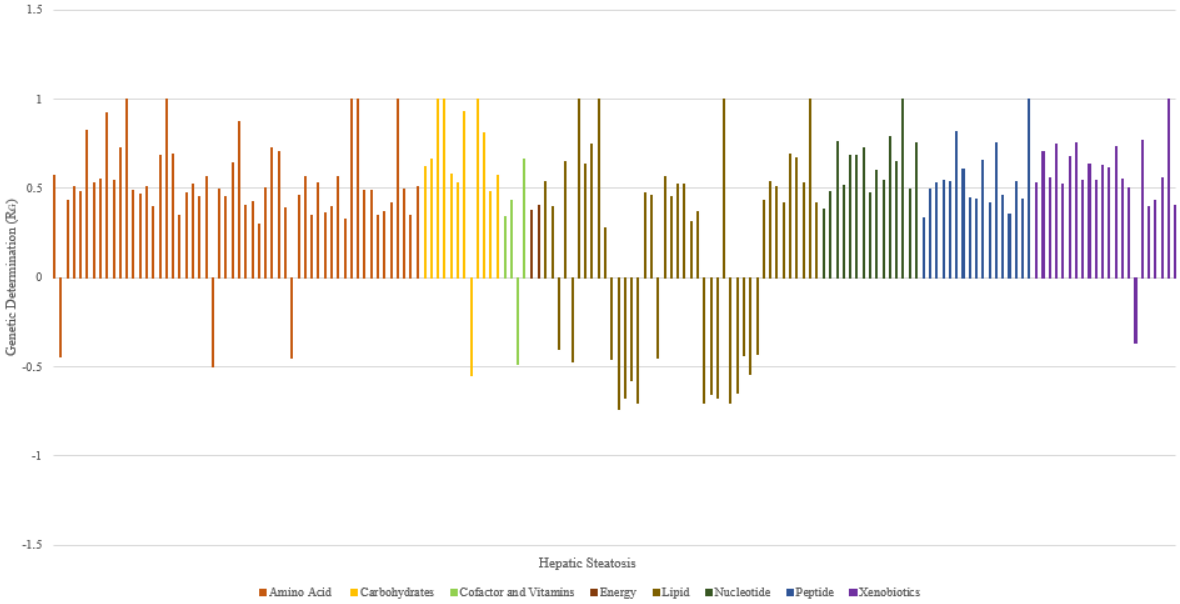
3.3. Heritability of Serum Metabolites in the Twin and Family Cohort

Among the 713 serum metabolites analyzed in the twins, 440 serum metabolites were heritable with significant heritability estimates (h^2) ranging from 0.28 to 0.91 after adjustment for age, sex and ethnicity.

3.4. Shared Gene Effects between Serum Metabolites and Hepatic Steatosis and Fibrosis

Among the 440 heritable serum metabolites identified, 170 serum metabolites had a significant shared gene-effect with hepatic steatosis, as measured by MRI-PDFF and 94 serum metabolites had a significant shared gene-effect with hepatic fibrosis as measured by MRE. Among them, 56 serum metabolites had a significant shared gene-effect with both hepatic steatosis and fibrosis (Figure 3-4). The fold-change of these 56 serum metabolites in individuals with NAFLD versus non-NAFLD and in individuals with higher MRE measurement versus lower MRE measurement (Figure 5).

(a)



(b)

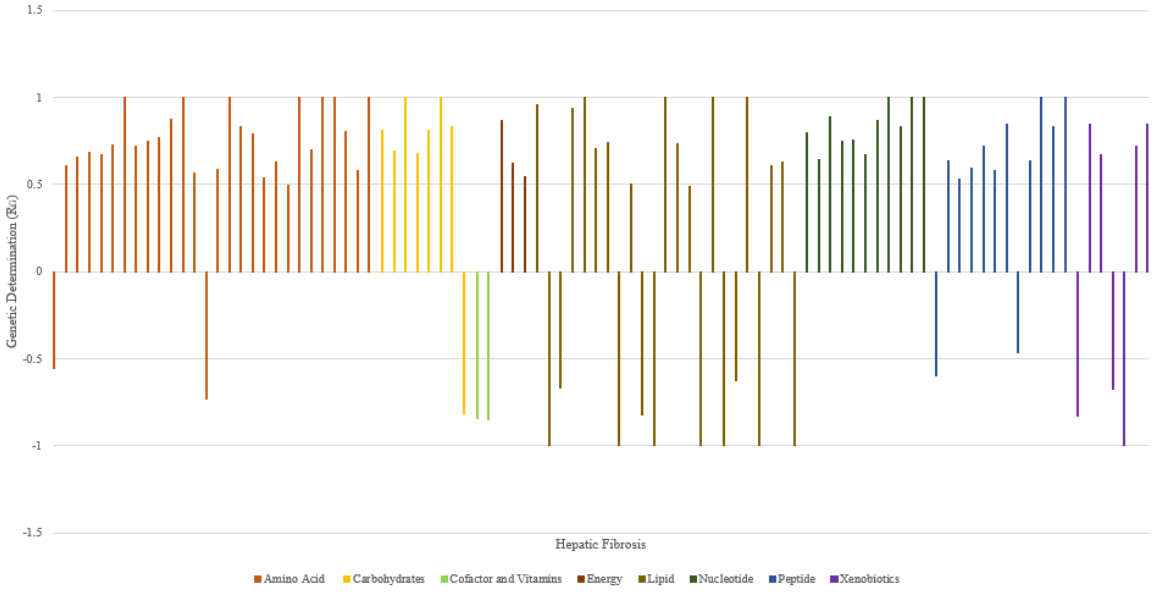


Figure 3: Shared gene-effect between serum metabolites and (a) hepatic steatosis (b) hepatic fibrosis, and (c) hepatic steatosis and fibrosis. Shared gene-effect of the serum metabolites is presented in bar graphs as genetic determination estimates (rG) adjusted for age, sex and ethnicity colored based on eight metabolic superpathways.[2]

(c)

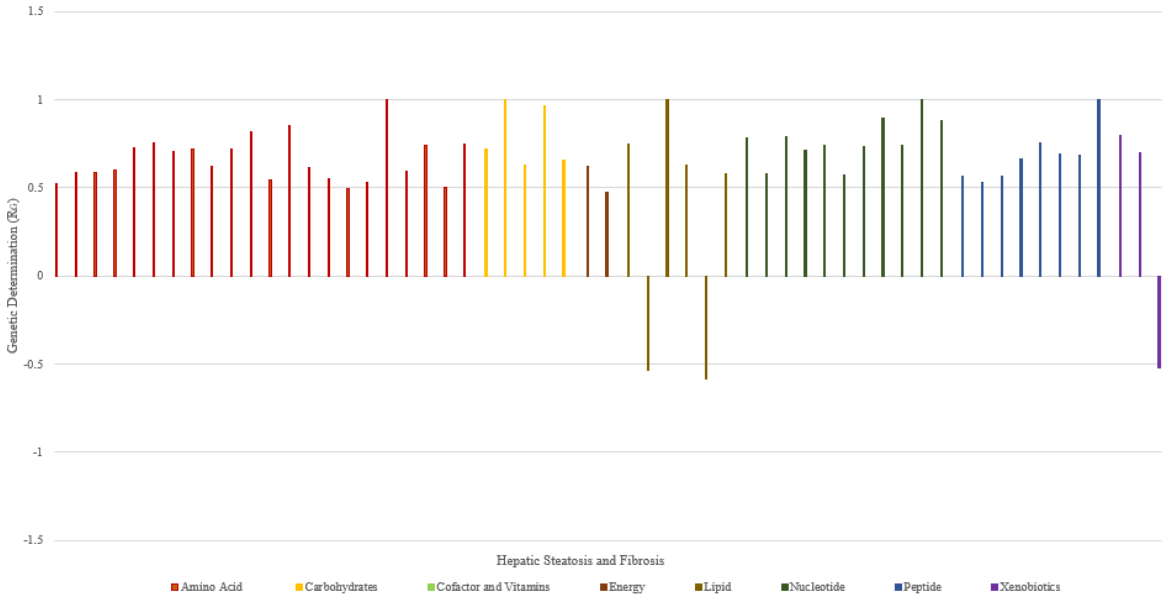


Figure 3: Shared gene-effect between serum metabolites and (a) hepatic steatosis (b) hepatic fibrosis, and (c) hepatic steatosis and fibrosis. Shared gene-effect of the serum metabolites is presented in bar graphs as genetic determination estimates (rG) adjusted for age, sex and ethnicity colored based on eight metabolic superpathways.[2] Continued.

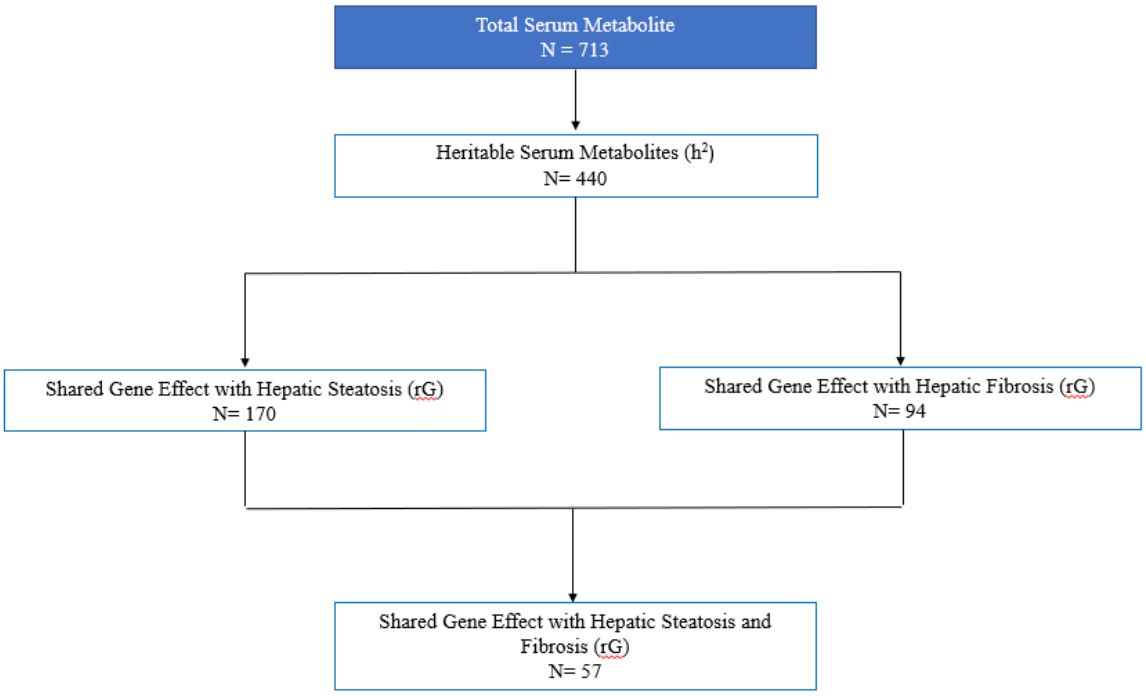


Figure 4: Summary of Serum Metabolite Heritability and Their Shared Gene Effect with Hepatic Steatosis and Fibrosis.

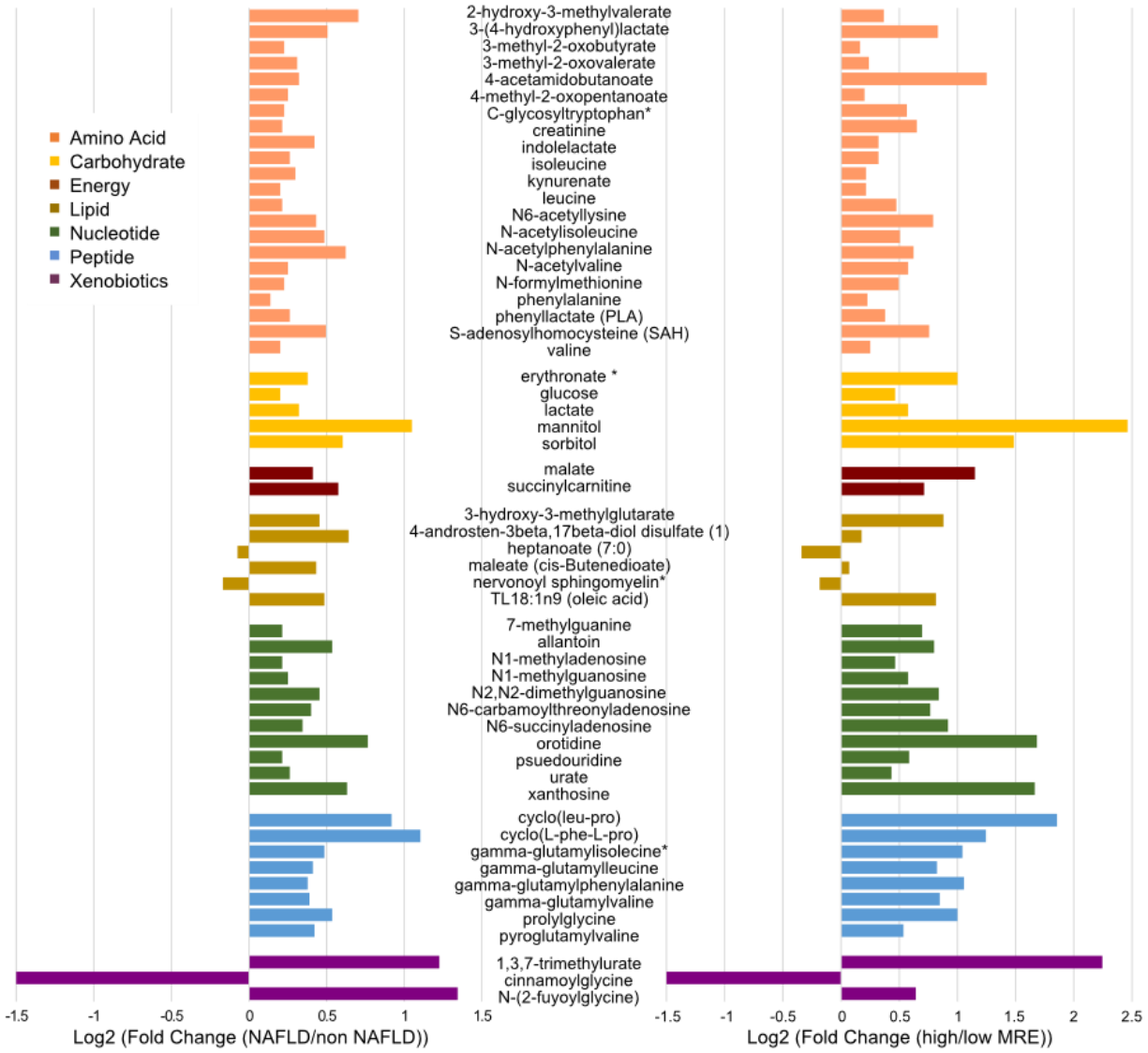


Figure 5: 57 Serum Metabolites have a Shared Gene Effect with Hepatic Steatosis and Liver Fibrosis[2]

Fifty-seven serum metabolites shared gene-effect with hepatic steatosis and fibrosis. The variation of the 57 serum metabolites, colored based on eight superpathways, with shared gene-effect with hepatic steatosis and fibrosis is depicts as $\log_2$ fold-change in participants with NALFD versus non NAFLD as defined by MRI-PDFF ≥5% (left panel) and in participants with higher measurement MRE (>4.17 kPa) versus lower measurement of MRE (<1.67 kPa) (right panel).[2]

3.5. Pathways Associated with the Heritable Serum Metabolites

The 170 serum metabolites that had a significant shared gene effect with hepatic steatosis were categorized into eight metabolomic superpathways (amino acid, carbohydrate, cofactor & vitamins, energy, lipid, nucleotide, peptide, and xenobiotics). Out of the 170 heritable serum metabolites, 56 were found in the amino acid superpathway, 12 were found in the carbohydrates superpathway, 4 were found in the cofactors & vitamins superpathway, 42 were found in the lipid superpathway, 15 were found in the nucleotide superpathway, 17 were found in the peptide superpathway, and 22 were found in the xenobiotics superpathway (Figure 6). Overall, the majority of serum metabolites that had a significant shared gene effect with hepatic steatosis were from the amino acid (33%) and lipid (25%) superpathways.

The 94 serum metabolites that had a significant shared gene effect with hepatic fibrosis were also categorized. Out of the 94 heritable serum metabolites that have a shared gene effect with hepatic fibrosis, 28 were found in the amino acid superpathway, 8 were found in the carbohydrates superpathway, 2 were found in the cofactors & vitamins superpathway, 23 were found in the lipid superpathway, 11 were found in the nucleotide superpathway, 12 were found in the peptide superpathway, and 7 were found in the xenobiotics superpathway (Figure 6). A majority of the serum metabolites that have a significant shared gene effect with hepatic fibrosis were also from the amino acid (30%) and lipid (24%) superpathways.

Furthermore, the 57 serum metabolites that had a significant shared gene effect with hepatic steatosis and fibrosis were also categorized. Out of the 57 heritable serum metabolites that have a shared gene effect with hepatic steatosis and fibrosis, 22 were found in the amino acid

superpathway, 5 were found in the carbohydrates superpathway, 2 were found in energy superpathway, 6 were found in the lipid superpathway, 11 were found in the nucleotide superpathway, 8 were found in the peptide superpathway, 3 were found in the xenobiotics superpathway, and none were found in the the cofactors & vitamins superpathway (Figure 6). Around 39% of the serum metabolites that had a shared gene effect with hepatic steatosis and fibrosis were from the amino acid superpathway. Interestingly, there were fewer lipid serum metabolites that had a shared gene effect with hepatic steatosis and fibrosis (around 11%). Instead, serum metabolites from the nucleotide superpathway had a greater shared gene effect with both hepatic steatosis and fibrosis (around 19%).

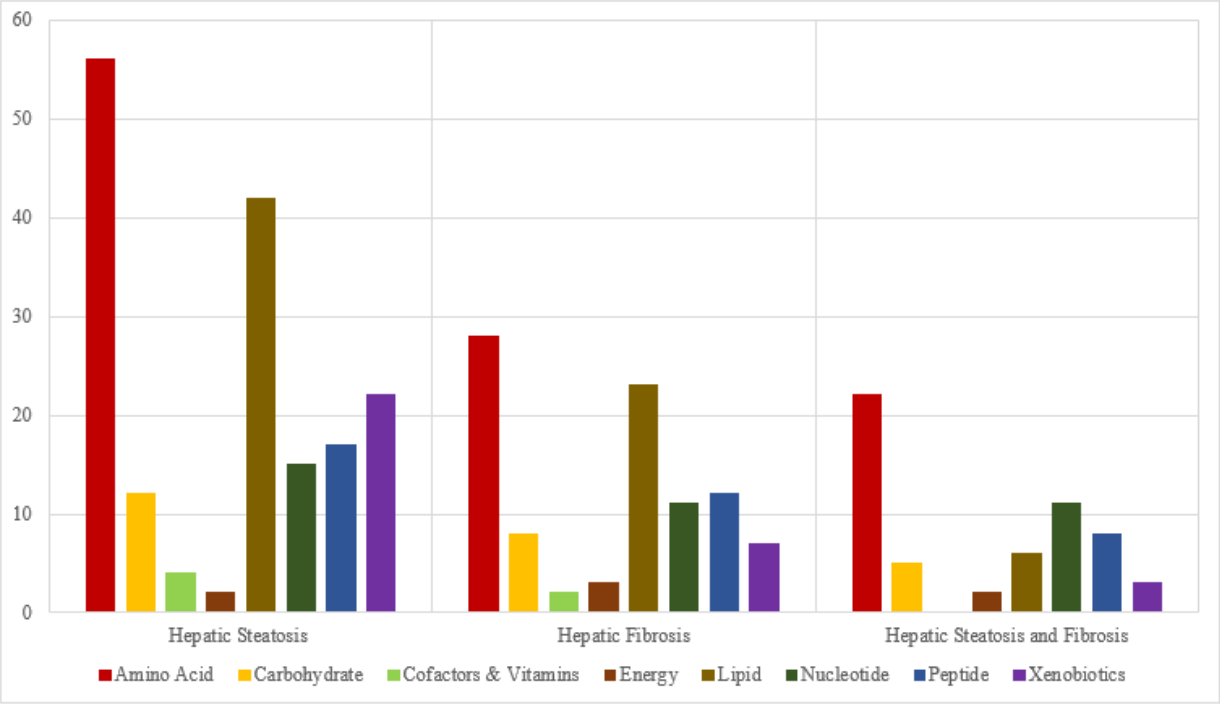


Figure 6: Amount of Serum Metabolites that have a Shared Gene Effect with Hepatic Steatosis (left), Hepatic Fibrosis (middle), and Hepatic Steatosis and Fibrosis (right).

3.6. Subpathways of the 57 Serum Metabolites With Shared Gene Effect with Hepatic Steatosis and Fibrosis

The 57 serum metabolites with a shared gene effect with hepatic steatosis and fibrosis were categorized into their subpathways. There were 22 serum metabolites found in the amino acid superpathway. These were sub-divided into subpathways. The majority of amino acid serum metabolites (41%) were found to be from the leucine, isoleucine, and valine metabolism. Around 18% and 14% were found in the phenylalanine and tyrosine metabolism and tryptophan metabolism subpathways, respectively. Around 10% were found in both the methionine, cysteine, SAM and taurine metabolism and lysine metabolism. Around 5% were involved in the creatine and polyamine metabolism.

There were 5 serum metabolites associated with the carbohydrate superpathway and around 40% of the serum metabolites were found in the glycolysis, gluconeogenesis, and pyruvate metabolism subpathway, 40% were found in the fructose, mannose, and galactose metabolism subpathway, and 20% were found in the aminosugar metabolism subpathway.

In addition, the two serum metabolites associated with the energy superpathway were part of the TCA cycle.

The 6 serum metabolites that were associated with the lipid superpathway were all even divided between the mevalonate metabolism, medium chain fatty acid, fatty acid, dicarboxylate, steroid, sphingolipid metabolism, and total fatty acid subpathways.

There were 11 serum metabolites associated with the nucleotide superpathway where the purine metabolism, (hypo) xanthine/inosine containing subpathway, purine metabolism, adenine containing subpathway, purine metabolism, guanine containing subpathway each contained 27% of the total serum metabolites. The remaining 19% were found in the pyrimidine metabolism, uracil containing and orotate containing subpathway .

In addition, there were 8 serum metabolites associated with the peptide superpathway where 50% of the serum metabolites were found in the gamma-glutamyl amino acid subpathway and 50% were found in the dipeptide subpathway.

Lastly, there were 3 serum metabolites found in the xenobiotics and 2 were found in the food component/plant subpathway and 1 was found in the xanthine subpathway. There were no serum metabolites associated with the cofactors and vitamins superpathway (Table 2).

Table 2: Superpathways and Subpathways of the 57 Serum Metabolites that have a Shared Gene Effect with Hepatic Steatosis and Fibrosis.

Superpathway	**Number of Serum Metabolites in Superpathway**	**Subpathway**	**Number of Serum Metabolite in Subpathways**	**Number of Subpathway Serum Metabolites/ Number of Superpathway Serum Metabolites (%)**
Amino Acid	22	Leucine, Isoleucine, Valine Metabolism	9	41%
		Phenylalanine and Tyrosine Metabolism	4	18%
		Tryptophan Metabolism	3	14%
		Methionine, Cysteine, SAM, and Taurine Metabolism	2	10%
		Lysine Metabolism	2	10%
		Creatine and polyamine Metabolism	1	5%
Carbohydrate	5	Glycolysis, Gluconeogenesis, and Pyruvate Metabolism	2	40%
		Fructose, Mannose, and Galactose Metabolism	2	40%
		Aminosugar Metabolism	1	20%
Energy	2	TCA Cycle	2	100%
Lipid	6	Mevalonate Metabolism	1	17%
		Medium Chain Fatty Acid Metabolism	1	17%

Table 2: Superpathways and Subpathways of the 57 Serum Metabolites that have a Shared Gene Effect with Hepatic Steatosis and Fibrosis. Continued.

Superpathway	Number of Serum Metabolites in Superpathway	Subpathway	Number of Serum Metabolite in Subpathways	Number of Subpathway Serum Metabolites/ Number of Superpathway Serum Metabolites (%)
Lipid	6	Fatty Acid, Dicarboxylate Metabolism	1	17%
		Steroid Metabolism	1	17%
		Sphingolipid Metabolism	1	17%
		Totally Fatty Acid Metabolism	1	17%
Nucleotide	11	Purine Metabolism, (Hypo) Xanthine/Inosine Containing	3	27%
		Purine Metabolism, Adenine Containing	3	27%
		Purine Metabolism, Guanine Containing	3	27%
		Pyrimidine Metabolism, Uracil Containing	1	19%
		Pyrimidine Metabolism, Orotate Containing	1	19%
Peptide	8	Gamma-Glutamyl Amino Acid	4	50%
		Dipeptide	4	50%
Xenobiotics	3	Food Component/Plant	2	67%
		Xanthaine	1	33%

Chapter 4

Discussion

Utilizing a uniquely, well characterized, discovery cohort of community-dwelling twins and families, out of the 713 metabolites analyzed, there was 153 serum metabolites that were significantly expressed in individuals with NAFLD compared to individuals without NAFLD and 86 serum metabolites were significantly expressed in individuals with the higher MRE measurements than individual with lower MRE measurements. In addition, the main findings of this study were that out of 713 serum metabolites analyzed, there were 440 serum metabolites that were heritable. And out of the 440 heritable serum metabolites, 170 had a significant shared gene effect with hepatic steatosis, 94 had a significant shared gene effect with hepatic fibrosis, and 57 serum metabolites had significant shared gene effect with hepatic steatosis and fibrosis.

These 57 serum metabolites that were found to have a significant shared gene effect with hepatic steatosis and fibrosis were categorized into eight metabolic superpathway (amino acid, carbohydrate, co-factors and vitamins, energy, lipid, nucleotide, peptide, and xenobiotics) where the majority of the serum metabolites were associated with the amino acid superpathway (39%) and nucleotide superpathway (19%).

Studies have shown that both NAFLD [17,35,36] and serum metabolites are heritable traits.[30-32,37,38] In addition, it has been demonstrated that there are significant shared gene-effects between hepatic steatosis and metabolic risk factors.[4,39] This project was built on the results of those studies to show additional heritability and shared gene effect between hepatic steatosis and fibrosis and serum metabolites.

In this study, we reported similar heritability estimates of serum metabolites as previous studies using a twin and family model.[26-30] Of the 57 serum metabolites that were found to have a shared gene effect with hepatic steatosis and fibrosis, around 40% of the serum metabolites were associated with the amino-acid superpathway suggesting that amino acids should be looked at more closely to further understand its role in the pathogenesis of NAFLD. In addition, these results align with previous studies showing differences in amino-acid pathway associated with features of NAFLD and underlie the potential role of amino-acids in the development and progression of NAFLD.[22-25] Overall, this indicates that the amino acid superpathway is a potentially good starting point for future study for the identification of biomarkers and understanding the pathogenesis of NAFLD.

4.1. Strengths and Limitations

There are several notable strengths in this study, including the use of an independent, well-characterized prospective cohorts in which conditions such as excessive alcohol use, steatogenic therapies, viral hepatitis, and secondary causes of steatosis were systematically excluded. The twin and family cohort allowed for the assessment of the heritability estimates and shared gene-effects using accurate and reproducible non-invasive imaging biomarker for the quantification of hepatic steatosis and fibrosis.[2]

However, this study does have the following limitations; this cross-sectional study design only allowed capturing a snapshot of metabolomics across the different features of NAFLD. Therefore, it is not possible to determine whether the serum metabolites that were obtained in the results were a causal factor in the pathogenesis of NAFLD. Finally, although the association between serum metabolites and hepatic steatosis and fibrosis were significant after adjustment for

age, sex, obesity and Hispanic ethnicity, potential other confounding factors such as diet could not be excluded.[2]

Further longitudinal studies are needed to determine the serum metabolites in the development of NAFLD and whether serum metabolites are a reliable biomarker associated with longitudinal changes in hepatic steatosis and fibrosis.

4.2. Implications for Future Study

In this study, we determined that serum metabolites were heritable and that they have a significant shared gene effect with both hepatic steatosis and fibrosis. Further studies with larger sample sizes are needed to determine the precise genetics mechanism underlying the relationship between serum metabolites and NAFLD. In addition, further investigation needs to be done to see if common mutations and polymorphism account for the heritability of serum metabolites related to both hepatic steatosis and fibrosis. Additionally, further analysis on the serum metabolites associated with the different super- and sub- pathways needs to be done to further understand their relationships to both hepatic steatosis and fibrosis. Furthermore, additional studies need to be done to validate, assess, and determine if serum metabolites could be a useful non-invasive biomarker as well as a marker for monitoring NAFLD and NAFLD related fibrosis.

This thesis contains material as it appears in “Link between gut microbiome derived metabolite and shared gene-effects with hepatic steatosis and fibrosis in NAFLD”, Hepatology 68(3), 2018. Cyrielle Caussy, Cynthia Hsu, Min-Tzu Lo, Amy Liu, Ricki Bettencourt, Veeral H. Ajmera, Shirin Bassirian, Jonathan Hooker, Ethan Sy, Lisa Richards, Nicholas Schork, Bernd Schnabl, David A. Brenner, Claude B. Sirlin, Chi-Hua Chen, Rohit Loomba, Genetics of NAFLD in Twins Consortium. I would like to thank Dr. Cyrielle Caussy and Dr. Rohit Loomba for giving me the opportunity to be a co-author of this paper and to be involved with this project.

REFERENCES

1. Cui, J., Chen, CH., Lo, MT., Schork, N., Bettencourt, R., Gonzalez, M.P., Bhatt, A., Hooker, J., Shaffer, K., Nelson, K.E., Long, M.T., Brenner, D.A., Sirlin, C.B., Loomba, R., *Shared genetic effects between hepatic steatosis and fibrosis: A prospective twin study*. Hepatology, 2016. **64**(5): p. 1547- 1558.

2. Caussy, C., Hsu, C., Lo, MT., Liu, A., Bettencourt, R., Ajmera, V.H., Bassirian, S., Hooker, J., Sy, E., Richards, L., Schork, N., Schnabl, B., Brenner, D.A., Sirlin, C.B., Chen, CH., Loomba, R., Genetics of NAFLD in Twins Consortium., *Link between gut-microbiome derived metabolites and shared gene-effects with hepatic steatosis and fibrosis in NAFLD.* Hepatology, 2018. **68**(3): p. 918-932.

3. "Anatomy and Function of the Liver." *Stanford Children's Health*, Stanford Medicine, https://www.stanfordchildrens.org/en/topic/default?id=anatomy-and-function-of-the-liver-90-P03069.

4. Younossi, Z.M., Koenig, A.B., Abdelatif, D., Fazel, Y., Henry, L., Wymer, M., *Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes*. Hepatology, 2016. **64**(1): p. 73- 84.

5. Loomba, R., Sanyal, A.J., *The global NAFLD epidemic*. Nature Reviews Gastroenterology & Hepatology, 2013. **10**(11): p. 686- 690.

6. Dulai, P.S., Singh, S., Patel, J., Soni, M., Prokop, L.J., Younossi, Z., Sebastiani, G., Mattias, E., Hagstrom, H., Nasr., Stal, P., Wong, V.W.S., Kechagias, S., Hultcrantz, R., Loomba, R., *Increased risk of mortality by fibrosis stage in nonalcoholic fatty liver disease: Systematic review and meta-analysis*. Hepatology, 2017. **65**(5): p. 1557- 1565.

7. Struben, V.M., Hespenheide, E.E., Caldwell, S.H., *Nonalcoholic steatohepatitis and cryptogenic cirrhosis within kindreds.* American Journal of Medicine, 2000. **108**(1): p. 9- 13.

8. El-Kader, S.M.A., El-Den Ashmawy, E.M.S., *Non-alcoholic fatty liver disease: The Diagnosis and Management*. World Journal of Hepatology, 2015. **7**(6): p. 846-858.

9. Moore, J.B., *Non-alcoholic fatty liver disease: The hepatic consequence of obesity and the metabolic syndrome.* The Proceedings of the Nutrition Society, 2010. **69**(2): p. 211-220.

10. Pavlov, C.S., Casazza, G., Nikolova, D., Tschatzis, E., Burroughs, A.K., Ivashkin, V., Gluud, C., *Transient elastography for measurement of liver fibrosis and cirrhosis in people with alcoholic liver disease.* Cochrane Systemic Review, 2015.

11. Wong, T., Wong, R.J., Gish, R.G., *Diagnostic and Treatment Implications of Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis*. Gastroenterology & Hepatology, 2019. **15**(2): p. 83-89.

12. Mofrad, P., Contos, M.J., Haque, M., Sargeant, C., Fisher, R.A., Luketic, V.A., Sterling, R.K., Shiffman, M.L., Stravitz, R.T., Sanyal, A.J., *Clinical and histologic spectrum of nonalcoholic fatty liver disease associated with normal ALT values*. Hepatology, 2003. **37**(6): p. 1286-1292.

13. Amarapurkar, D.N., Patel, N.D., *Clinical spectrum and natural history of non-alcoholic steatohepatitis with normal alanine aminotransferase values*. Tropical Gastroenterology, 2004. **25**(3): p. 130-134.

14. Maximos, M., Bril, F., Portillo Sanchez, P., Lomonaco, R., Orsak, B., Biernacki, D., Suman, A., Weber, M., Cusi, K., *The role of liver fat and insulin resistance as determinants of plasma aminotransferase elevation in nonalcoholic fatty liver disease*. Hepatology, 2014. **61**(1): p. 153-160.

15. Han, MA., Saouaf, R., Ayoub, W., Todo, T., Mena, E., Noureddin, M., *Magnetic resonance imaging and transient elastography in the management of nonalcoholic fatty liver disease (NAFLD)*. Expert Review of Clinical Pharmacology, 2017. **10**(4): p. 379-390.

16. Schwimmer, J.B., Celedon, M.A., Lavine, J.E., Salem, R., Campbell, N., Schork, N.J., Shiehmorteza, M., Yokoo, T., Chavez, A., Middleton, M.S., Sirlin, C.B., *Heritability of nonalcoholic fatty liver disease*. Gastroenterology, 2009. **136**(5): p. 1585- 1592.

17. Loomba, R., Schork, N., Chen, CH., Bettencourt, R., Bhatt, A., Ang, B., Nguyen, P., Hernandez, C., Richards, L., Salotti, J., Lin, S., Seki, E., Nelson, K.E., Sirlin, C.B., Brenner, D., Genetics of NALFD in Twins Consortium., *Heritability of hepatic fibrosis and steatosis based on a prospective twin study*. Gastroenterology, 2015. **149**(7): p. 1784- 1793.

18. Speliotes, E.K., Yerges-Armstrong, L.M., Wu, J., Hernaez, R., Kim, L.J., Palmer, C.D., Gudnason, V., Eiriksdottir, G., Garcia, M.E., Launer, L.J., Nalls, M.A., Clark, J.M., Mitchell, B.D., Shuldiner, A.R., Butler, J.L., Tomas, M., Hoffman, U., Hwang, SJ., Massaro, J.M., O'Donnell, C.J., Sahani, D.V., Salomaa, V., Schadt, E.E., Schwartz, S.M., Siscovick, D.S., NASH CRN., GIANT Consortium., MAGIC Investigators., Voight, B.F., Carr, J.J., Feitosa, M.F., Harris, T.B., Fox, C.S., Smith, A.V., Lao, W.H.L., Hirschhorn, J.N., Borecki, I.B., GOLD Consortium., *Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits.* PLOS Genetics, 2011. **7**(3): p. e1001324.

19. Sookoian, S., Pirola, C.J., *Meta-analysis of the influence of I148M variant of patatin-like phospholipase domain containing 3 gene (PNPLA3) on the susceptibility and histological severity of nonalcoholic fatty liver disease*. Hepatology, 2011. **53**(6): p. 1883- 1894.

20. Liu, YL., Reeves, H.L., Burt, A.D., Tiniakos, D., McPherson, S., Leathart, J.B., Allison, M.E.D., Alexander, G.J., Piguet, AC., Anty, R., Donaldson, P., Aithal, G.P., Francque, S., Gaal, L.V., Clement, K., Ratziu, V., Dufour, JF., Day, C.P., Daly, A.K., Anstee, Q.M.,

TM6SF2 rs58542926 influences hepatic fibrosis progression in patients with non-alcoholic fatty liver disease. Nature Communications, 2014. **5**: p. 4309.

21. Alonso, C., Fernandez-Ramos, D., Varela-Rey, M., Martinez-Arranz, I., Navasa, N., Van Liempd, S.M., Lavin Trueba, J.L., Mayo, R., Ilisso, C.P., de Juan, V.G., Iruarrizaga-Lejarreta, M., dela Cruz-Villar, L., Minchole, I., Robinson, A., Crespo, J., Martin-Duce, A., Romero-Gomez, M., Sann, H., Platon, J., Van Eyk, J., Aspichueta, P., Noureddin, M., Falcon-Perez, J.M., Anguita, J., Aransay, A.M., Martinez-Chantar, M.L., Lu, S.C., Mato, J.M., *Metabolomic Identification of Subtypes of Nonalcoholic Steatohepatitis.* Gastroenterology, 2017. **152**(6): p. 1449- 1461.

22. Jin, R., Banton, S., Tran, V.T., Konomi, J.V., Li, S., Jones, D.P., Vos, M.B., *Amino Acid Metabolism is Altered in Adolescents with Nonalcoholic Fatty Liver Disease-An Untargeted, High Resolution Metabolomics Study*. Journal of Pediatrics, 2016. **17**: p. 14- 19.e5

23. Mardinoglu, A., Agren, R., Kampf, C., Asplund, A., Uhlen, M., Nielsen, J., *Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease.* Nature Communications, 2014. **5**: p. 3083.

24. Kaikkonen, J.E., Wurtz, P., Suomela, E., Lehtovirta, M., Kangas, A.J., Jula, A., Mikkila, V., Viikari, J.S.A., Juonala, M., Ronnemaa, T., Hutri-Kahonen, N., Kahonen, M., Lehtimaki, T., Soininen, P., Ala-Korpela, M., Raitakari, O.T., *Metabolic profiling of fatty liver in young and middle-aged adults: Cross-sectional and prospective analyses of the Young Finns Study*. Hepatology, 2017. **65**(2): p. 491- 500.

25. Sookoian, S., Puri, P., Castano, G.O., Scian, R., Mirshahi, F., Sanyal, A.J., Pirola, C.J., *Nonalcoholic steatohepatitis is associated with a state of betaine-insufficiency*. Liver International, 2017. **37**(4): p. 611- 619.

26. Draisma, H.H., Beekman, M., Pool, R., van Ommen, G.J., Vaarhorst, A.A.M., de Craen, A.J.M., Willemsen, G., Slagnoom, P.E., Boomsma, D.I., *Familial resemblance for serum metabolite concentrations*. Twin Research and Human Genetics, 2013. **16**(5): p. 948- 961.

27. Kettunen, J., Tukiainen, T., Sarin, AP., Ortega-Alonso, A., Tikkanen, E., Lyytikainen, LP., Kangas, A.J., Soininen, P., Wurtz, P., Silander, K., Dick, D.M., Rose, R.J., Savolainen, M.J., Viikari, J., Kahonen, M., Lehtimaki, T., Pietilainen, K.H., Inouye, M., McCarthy, M.I., Jula, A., Eriksson, J., Raitakari, I.T., Salomaa, V., Kaprio, J., Jarvelin, MR., Peltonen, L., Perola, M., Freimer, N.B., Ala-Kopela, M., Palotie, A., Ripatti, S., *Genome-wide association study identifies multiple loci influencing human serum metabolite levels*. Nature Genetics, 2012. **44**(3): p. 269- 276.

28. Long, T., Hicks, M., Yu, HC., Biggs, W.H., Kirkness, E.F., Menni, C., Zierer, J., Small, K.S., Mangino, M., Messier, H., Brewerton, S., Turpaz, Y., Perkins, B.A., Evans, A.M., Miller, L.A.D., Guo, L., Caskey, C.T., Schork, N.J., Garner, C., Spector, T.D., Venter, J.C.,

Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. Nature Genetics, 2017. **49**(4): p. 568- 578.

29. Shin, SY., Fauman, E.B., Petersen, AK., Krumsiek, J.K., Santos, R., Huang, J., Arnold, M., Erte, I., Forgetta, V., Yang, TP., Walter, K., Menni, C., Chen, L., Vasquez, L., Valdes, A.M., Hyde, C.L., Wang, V., Ziemek, D., Roberts, P., Xi, L., Grundberg, E., The Multiple Tissues Human Expression Resource (MuTHER) Consortium, Waldenberger, M., Richards, J.B., Mohney, R.P., Milburn, M.V., John, S.L., Trimmer, J., Thesis, F.J., Overington, J.P., Suhre, K., Brosnan, J., Gieger, C., Katenmuller, G., Spector, T.D., Soranzo, N., *An atlas of genetic influences on human blood metabolites*. Nature Genetics, 2014. **46**(6): p. 543- 550.

30. Nicholson, G., Rantalainen, M., Maher, A.D., Li, J.V., Malmodin, D., Ahmadi, K.R., Faber, J.H., Hallgrimsdottir, I.B., Barrett, A., Toft, H., Krestyaniova, M., Viksna, J., Guha Neogi, S., Dumas, ME., Sarkans, U., The MoIPAGE Consortium, Silverman, B.W., Donnelly, P., Nicholson, J.K., Allen, M., Zondervan, K.T., Lindon, J.C., Spector, T.D., McCarthy, M.I., Holmes, E., Baunsgaard, D., Holmes, C.C., *Human metabolic profiles are stably controlled by genetic and environmental variation*. Molecular Systems Biology, 2011. **7**: p. 525.

31. Zarrinpar, A., Gupta, S., Maurya, M.R., Subramaniam, S., Loomba, R., *Serum microRNAs explain discordance of non-alcoholic fatty liver disease in monozygotic and dizygotic twins: a prospective study*. Gut, 2016. **65**(9): p. 1546- 1554.

32. Caussy, C., Soni, M., Cui, J., Bettencourt, R., Schork, N., Chen, CH., Al Ikhwan, M., Bassirian, S., Cepin, S., Gonzalez, M.P., Mendler, M., Kono, Y., Vodkin, I., Mekeel, K., Haldorson, J., Hemming, A., Andrews, B., Salotti, J., Richards, L., Brenner, D.A., Sirlin, C.B., Loomba, R., the Familial NAFLD Cirrhosis Research Consortium. *Nonalcoholic fatty liver disease with cirrhosis increases familial risk for advanced fibrosis*. Journal of Clinical Investigation, 2017. **127**(7): p. 2697- 2704.

33. Wu, H., Esteve, E., Tremaroli, V., Khan, MT., Caesar, R., Manneras-Holm, L., Stahlman, M., Olsson, L.M., Serino, M., Planas-Felix, M., Xifra, G., Mercader, J.M., Torrents, D., Burcelin, R., Ricart, W., Perkins, R., Fernandez-Real, J.M., Backhed, F., *Metformin alters the gut microbiome of individuals with treatment-naive type 2 diabetes, contributing to the therapeutic effects of the drug*. Nature Medicine, 2017. **23**(7): p. 850- 858.

34. Visscher, P.M., *Power of the classical twin design revisited*. Twin Research and Human Genetics, 2004. **7**(5): p. 505- 512.

35. Loomba, R., Abraham, M., Unalp, A., Wilson, L., Lavine, J., Doo, E., Bass, N.M., *Association between diabetes, family history of diabetes, and risk of nonalcoholic steatohepatitis and fibrosis*. Hepatology, 2012. **56**(3): p. 943- 951.

36. Gill, S.R., Pop, M., Deboy, R.T., Eckburg, P.B., Turnbaugh, P.J., Samuel, B.S., Gordon, J.I., Relman, D.A., Raser-Liggett, C.M., *Metagenomic analysis of the human distal gut microbiome*. Science, 2006. **312**(5778): p. 1355- 1359.

37. Permutt, Z., Le, T.A., Peterson, M.R., Seki, E., Brenner, D.A., Sirlin, C., Loomba, R., *Correlation between liver histology and novel magnetic resonance imaging in adult patients with non-alcoholic fatty liver disease - MRI accurately quantifies hepatic steatosis in NAFLD*. Alimentary Pharmacology and Therapeutics, 2012. **36**(1): p. 22- 29.

38. Patel, N.S., Peterson, M.R., Brenner, D.A., Heba, E., Sirlin, C., Loomba, R., *Association between novel MRI-estimated pancreatic fat and liver histology-determined steatosis and fibrosis in non-alcoholic fatty liver disease*. Alimentary Pharmacology and Therapeutics, 2013. **37**(6): p. 630- 639.

39. Zhu, L., Baker, S.S., Gill, C., Liu, W., Alkhouri, R., Baker, R.D., Gill, S.R., *Characterization of gut microbiomes in nonalcoholic steatohepatitis (NASH) patients: a connection between endogenous alcohol and NASH*. Hepatology, 2013. **57**(2): p. 601- 609.

.