EUR Research Information Portal

Unravelling Obesity and Fatty Liver Disease Mechanisms

Publication status and date:
Published: 09/01/2024

Document Version
Publisher's PDF, also known as Version of record

Citation for the published version (APA):
Abozaid, Y. (2024). *Unravelling Obesity and Fatty Liver Disease Mechanisms: Insights from Population-Based Omics Studies*. [Doctoral Thesis, Erasmus University Rotterdam].

Link to publication on the EUR Research Information Portal

Terms and Conditions of Use
Except as permitted by the applicable copyright law, you may not reproduce or make this material available to any third party without the prior written permission from the copyright holder(s). Copyright law allows the following uses of this material without prior permission:

- you may download, save and print a copy of this material for your personal use only;
- you may share the EUR portal link to this material.

In case the material is published with an open access license (e.g. a Creative Commons (CC) license), other uses may be allowed. Please check the terms and conditions of the specific license.

Take-down policy
If you believe that this material infringes your copyright and/or any other intellectual property rights, you may request its removal by contacting us at the following email address: openaccess.library@eur.nl. Please provide us with all the relevant information, including the reasons why you believe any of your rights have been infringed. In case of a legitimate complaint, we will make the material inaccessible and/or remove it from the website.

Erasmus University Rotterdam
Making Minds Matter

Unravelling Obesity and Fatty Liver Disease Mechanisms;

Insights from Population-Based Omics Studies

Yasir Jameel Abozaid

Unravelling Obesity and Fatty Liver Disease Mechanisms; Insights from Population-Based Omics Studies

Yasir Jameel Abozaid

Acknowledgments
The work presented in this thesis was performed at the Department of Epidemiology, Erasmus MC, Rotterdam, the Netherlands. The studies presented herein were conducted primarily within the Rotterdam Study. We gratefully acknowledge the contributions provided by the study participants, staff, data management, participating general practitioners and health professionals involved.

The work presented in this thesis was conducted within the Molecular Systems Epidemiology group at the Department of Epidemiology Erasmus Medical Center Rotterdam, the Netherlands. The research was supported by a scholarship from the Ministry of higher education and scientific research, Kurdistan region of Iraq under the human capacity development program (HCDP). The Rotterdam Study is supported by the Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam.

Publication of this thesis was kindly supported by the Department of Epidemiology of the Erasmus Medical Center and by the Erasmus University Rotterdam.

Layout and printed by: Ridderprint BV | www.ridderprint.com
Cover design: Muhannad Jamal and Yasir Jameel Abozaid
ISBN: 978-94-6483-696-7

© 2024 Yasir Jameel Abozaid, Rotterdam, The Netherlands
The copyright is transferred to the respective publisher upon publication of the manuscript. No part of this thesis may be reproduced, stored in a retrieval system, or transmitted in any form or by any means without prior permission of the author or the publisher of the manuscript.

Unravelling Obesity and Fatty Liver Disease Mechanisms; Insights from Population-Based Omics Studies

Ontrafelen van mechanismen onderliggend aan obesitas en leververvetting; inzichten vanuit omics studies op populatie niveau

Thesis

To obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. A.L. Bredenoord

and in accordance with the decision of the Doctorate Board.
The public defence shall be held on
Tuesday, 9th of January 2024 at 10:30 hrs

by

Yasir Jameel Abozaid

born in Aqrah, Iraq

Erasmus University Rotterdam

DOCTORAL COMMITTEE

Promotor: Prof.dr. M. A. Ikram

Other members: Prof.dr. P.E. Slagboom
Prof.dr. L.J.W. van der Laan
Prof.dr. M. Schram

Copromotor: Dr. M. Ghanbari

Paranymphs: Ibrahim Ayada
Samantha Leonard

بسم الله الرحمن الرحيم

وقل رب زدني علما

صدق الله العظيم

Dedicated To my parents,

to my wife Narin

to my son, Laween and my daughter, Lana

Table of Contents

Chapter 1 General introduction **11**

Chapter 2 Epigenetic regulation of obesity traits and fatty liver disease **33**

2.1 Plasma circulating microRNAs associated with obesity, body fat distribution, and fat mass 35

2.2 Circulating microRNAs as potential biomarkers for fatty liver disease 55

2.3 Plasma microRNAs signature of alcohol consumption and their links to fatty liver disease 77

Chapter 3 Proteomics of fatty liver disease **105**

3.1 Plasma proteomic signature of fatty liver disease: the Rotterdam Study 107

Chapter 4 Metabolomics of fatty liver disease **127**

4.1 Circulating metabolites associated with fatty liver disease and liver enzymes: a comprehensive multi-platform populations-based study 129

Chapter 5 General discussion **179**

Chapter 6 Summary/ Samenvatting **199**

Chapter 7 Appendices **207**

Acknowledgments 209

List of manuscripts 213

PhD portfolio 215

About the author 217

Manuscripts that form the basis for this thesis

Chapter 2. Epigenetic regulation of obesity traits and fatty liver disease

Abozaid, Y.J., Zhang, X.*, Mens, M.M.*, Ahmadizar, F., Limpens, M., Ikram, M.A., Rivadeneira, F., Voortman, T., Kavousi, M. and Ghanbari, M., 2022. Plasma circulating microRNAs associated with obesity, body fat distribution, and fat mass: the Rotterdam Study. *International Journal of Obesity*, *46*(12), pp.2137-2144.

Zhang, X., Mens, M.M.*, **Abozaid, Y.J.***, Bos, D., Darwish Murad, S., de Knegt, R.J., Ikram, M.A., Pan, Q. and Ghanbari, M., 2021. Circulatory microRNAs as potential biomarkers for fatty liver disease: the Rotterdam study. *Alimentary Pharmacology & Therapeutics*, *53*(3), pp.432-442.

Karabegović, I., **Abozaid, Y.**, Maas, S.C., Labrecque, J., Bos, D., De Knegt, R.J., Ikram, M.A., Voortman, T. and Ghanbari, M., 2022. Plasma microRNA signature of alcohol consumption: the Rotterdam Study. *The Journal of nutrition*, *152*(12), pp.2677-2688.

Chapter 3. Proteomics of fatty liver disease

Abozaid, Y.J., Ayada, I.*, van Kleef, L.A.*, Vallerga, C.L., Pan, Q., Brouwer, W.P., Ikram, M.A., Van Meurs, J., de Knegt, R.J. and Ghanbari, M., 2023. Plasma proteomic signature of fatty liver disease: the Rotterdam Study. *Hepatology*, *78*(1), pp.284-294.

Chapter 4. Metabolomics of fatty liver disease

Yasir J. Abozaid, Ibrahim Ayada, Laurens A. van Kleef, Neil J Goulding, Jessica S. Williams-Nguyen, Robert C Kaplan, Robert J. de Knegt, Lynne E. Wagenknecht, Nicholette D. Palmer, Deborah A. Lawlor, Nicholas J Timpson, Jill M. Norris, Yii-Der Ida Chen, M. Arfan Ikram, Willem Pieter Brouwer, Mohsen Ghanbari. Circulating metabolites associated with fatty liver disease and liver enzymes: a comprehensive multi-platform populations-based study. (In submission)

*Authors denotes equal contribution

Chapter 1

General introduction

Introduction

The liver is a large organ found in the upper right quadrant of the abdomen.[1] It is responsible for an array of functions that are involved in metabolism, immunity, digestion, detoxification, vitamin storage, blood clotting, and many other functions.[2] Underlying etiologies in liver disease comprise viruses (Hepatitis A, B, C, D, and E), drugs, toxins, excessive alcohol drinking, metabolic and autoimmune diseases, and genetics [3, 4]. Most of these etiologies will eventually lead to liver fibrosis and cirrhosis (end-stage liver disease), which is when the liver fails to maintain the above mention functions. Obesity and alcohol consumption are two main risk factors for fatty liver disease (FLD), and they frequently coexist.[5] Moreover, there are considerable synergistic interaction effects between hazardous obesity-associated metabolic abnormalities and alcohol use in the development and progression of fatty liver disease.

Fatty liver disease includes two forms of alcoholic and non-alcoholic fatty liver disease (NAFLD). The latter form is closely linked to metabolic dysfunction and is the most common cause of chronic liver disease worldwide with a prevalence of over 33%.[6-8] This indicates the ongoing extent of FLD as a global health care burden, especially in developing countries. FLD is a term used to describe a range of related disorders **(Figure 1).** The earliest stage is hepatic steatosis which can progress into steatohepatitis (lobular inflammation and hepatocyte ballooning), a subpopulation of 10-20% of FLD individuals will further progress to fibrosis (scar tissue accumulation), and among them, 20% will develop to liver cirrhosis which in turn will cause hepatocellular carcinoma (HCC).[9-11] However, of special worry is the fact that non-cirrhotic individuals with steatohepatitis and fibrosis are also at an increased risk of developing HCC.[12]

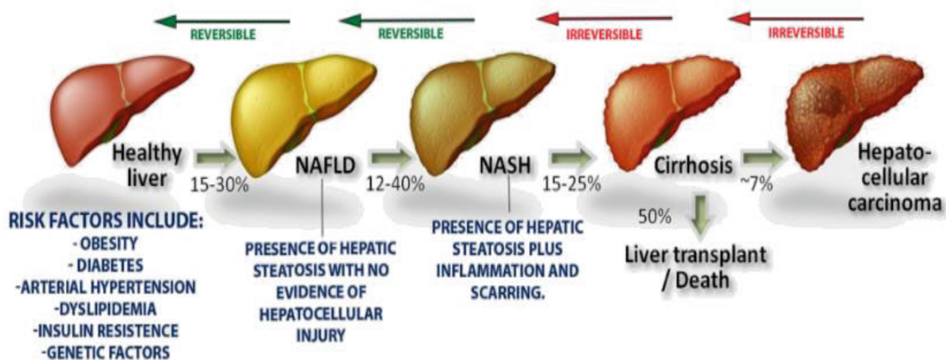


Figure 1. The spectrum of fatty liver disease (Figure created with https://txliver.com/patient-education/fatty-liver-2) Abbreviation: NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis.

Further complicating matter is the fact that NAFLD rarely cause the symptoms which are known to be liver related such as jaundice, a yellowing of the skin and the whites of the eyes, abdominal pain (especially on the right side), changes in the color of urine or stool, fatigue, nausea or vomiting, and finally easily bruising caused by impaired clotting, and swelling in the arms or legs (edema). Hence many

techniques have been explored in the last decades to allow the diagnosis of NAFLD including blood liver function tests, imaging tests (including ultrasound, MRI, or CT scan), and liver biopsy. At present, no drugs or pharmacological agents have been approved for the long-term treatment of NAFLD, but Resmetirom which is a thyroid hormone receptor agonist might be the first one to be approved.[13] Therefore, adopting a healthy lifestyle such as weight loss remains the cornerstone of treatment and has been shown in controlled trials to improve hepatic steatosis, hepatic inflammation, and fibrosis.[14]

Nomenclature of fatty liver disaese

In 1836, The term 'fatty liver' was first described by Thomas Addison from Newcastle upon Tyne, England.[15] Subsequently, in 1838 the pathologist Karl Rokitansky from Vienna, Austria documented hepatic fat accumulation that might be causative of cirrhosis in the autopsy specimens.[16] In 1958, Westwater and Feiner reported the histological findings of fatty infiltration of the liver in obese patients. In 1980, the term NAFLD and non-alcoholic steatohepatitis (NASH) was coined by Jurgen Ludwig to describe the progressive form of fatty liver disease.[17]

The current guidelines and consensus recommendations of NAFLD definition requires the exclusion of other causes of liver diseases and of a daily significant amount of alcohol, but the exact cut-offs to define 'significant' remains hotly debated. The proposed thresholds of alcohol consumption have varied from ≤1 drink (14 g/day) to 2-3 drinks (<30 g/day) in men and <20 g/day for women.[7, 18, 19]

Over the last two decades, many criticisms have been voiced about the nomenclature and definition of NAFLD.[17] Therefore, a panel recently reached a consensus that the disease should be renamed metabolic-associated fatty liver disease (MAFLD) and that the disease should be diagnosed by positive criteria.[20] The new inclusion-based diagnosis requires the simultaneous presence of both steatosis and metabolic dysfunction, and it is crucial to note that individuals with secondary causes of steatosis are not excluded from this diagnosis. This would open the door for efforts from the research community to update the nomenclature and sub-phenotype of the disease to accelerate the translational path to new treatments. Metabolic dysfunction was defined as a cluster of conditions that occur together including overweight, diabetes, or a combination of at least two minor criteria, such as increased blood pressure, high blood sugar, excess body fat around the waist, and abnormal cholesterol or triglyceride levels. Despite the increasing and strong evidence supporting the superior usefulness of the term MAFLD in clinical and academic practice, controversy abounds.[21]

Although both MAFLD and NAFLD are highly prevalent in the general population, with considerable overlap between them,[22] compared to NAFLD, significantly more individuals were additionally identified by MAFLD than were missed. Particularly, by using the MAFLD criteria, more individuals with liver damage were identified.

More recently, the global community has further chosen steatotic liver disease (SLD) as a comprehensive term encompassing the various causes of steatosis.[23] The term steatohepatitis was recognized as an important concept in understanding the pathophysiology and is retained within the new framework. Also, NAFLD has referred to as metabolic dysfunction-associated steatotic liver disease (MASLD), which includes patients with hepatic steatosis who also have at least one of five cardiometabolic risk factors such as overweight, fasting serum glucose, blood pressure, plasma triglycerides, and plasma HDL-cholesterol.

To accommodate individuals with MASLD who consume higher amounts of alcohol, a new category called MetALD has been introduced. MetALD describes individuals with MASLD who consume 140 g/week and 210 g/week of alcohol for females and males, respectively. Individuals with no metabolic parameters and no known cause for liver disease are categorized as having cryptogenic SLD. Additionally, metabolic dysfunction-associated steatohepatitis (MASH) replaces the term NASH in the updated classification. These updates, may provide a platform from which can increase disease awareness, reduce stigma and accelerate drug and biomarker development for the benefit of patients with MASLD, MASH and MetALD.

The pathphysiolgy of fatty liver disease

For many years, the so-called "two-hit model" of FLD pathogenesis has been accepted, according to which the disease began with simple steatosis (the first hit) and progressed to NASH through risk factors such as oxidative stress (the second hit).[24] However, we now understand that FLD pathophysiology is heterogeneous, multifactorial, and unlikely to be the same in every patient.[25] To understand the pathogenesis of FLD the most accepted hypothesis, a "multi parallel hits hypothesis", suggesting more sophisticated mechanisms based on the most current research, and it comprises a combination of elements rather than in series.[26] This hypothesis proposes that lipotoxicity of adipose tissue and alterations in gut microbial functions contribute to the evolution of inflammation and fibrosis in NAFLD.[27]

Recent insights into the pathogenesis of FLD indicate a complex interaction among environmental factors like obesity, changes in microbiota, and predisposing genetic variants resulting in disturbed lipid homeostasis and an excessive accumulation of triglycerides and other lipid species in hepatocytes.[28] Insulin resistance also plays a role in the development of FLD by increasing hepatic de novo lipogenesis and inhibition of adipose tissue lipolysis, with a subsequently increased flow of fatty acids (FA) in the liver.[29] Fat accumulates within the hepatocytes mainly as triglycerides originating from the esterification of glycerol and free FAs[30] **(Figure 2).**

Previous studies have detected that genetic factors and polymorphisms of several genes contribute to FLD and its results.[31-33] Yet, clinical investigations exploring the epigenetic reprogramming in non-alcoholic steatohepatitis (NASH) are in their early stages, but among the epigenetic markers, DNA methylation and microRNAs (miRNAs) have been extensively studied in relation to the risk of FLD.[34]

Furthermore, gut microbiome symbiosis may play a significant role in the pathogenesis of FLD by dysregulating the gut–liver axis[35] and is linked to a variety of extra intestinal diseases, including metabolic diseases such as insulin resistance and obesity.[36] Alteration of the gut microbiota associated with obesity and insulin resistance has consequences on both the homeostasis of energy and the systemic inflammation secondary to endotoxemia. Excessive fat accumulation in the liver can lead to mitochondrial dysfunction and endoplasmic reticulum stress and activation of the unfolded protein response, which will subsequently lead to the activation of inflammatory responses.[37, 38]

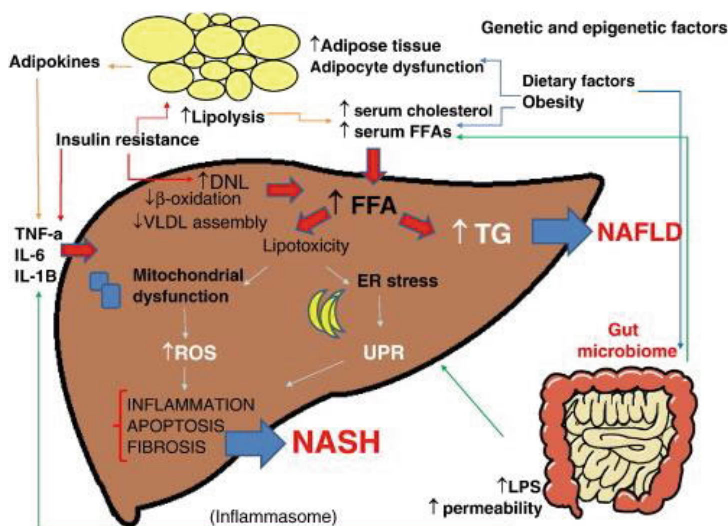


Figure 2. Multiple hit hypothesis for the development of fatty liver disease. (Figure created with ScienceDirect.com) Abbreviations: FFAs, free fatty acids; DNL, de novo lipogenesis; VLDL, very low-density lipoproteins; CH, cholesterol; TNF-α, tumor necrosis factor alpha; IL-6, interleukin 6; TG, triglycerides; ROS, reactive oxygen species; ER, endoplasmic reticulum; UPR, unfolded protein response; LPS, lipopolysaccharide; NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis.

Diagnosis of fatty liver disease

Currently, liver biopsy remains the gold standard for diagnosing FLD, but its widespread use is limited by the risk associated with an invasive procedure, cost, and sampling error.[39] Non-invasive imaging methods such as mainly conventional ultrasonography, computed tomography, magnetic resonance imaging, and other newer imaging technologies, are rapidly evolving and may replace biopsy in some circumstances. These imaging techniques are still relatively limited in the detection of inflammation (NASH), which is more important than steatosis in terms of its high risk for fibrosis, cirrhosis, and hepatocellular carcinoma.[40] Ultrasound and magnetic resonance imaging (MRI)-based markers have emerged as key noninvasive biomarkers in NAFLD with the ability of MRI to accurately detect hepatic steatosis and liver fibrosis.[41, 42] Emerging data support the use of MRI-derived proton density fat fraction (MRI-PDFF) as a non-invasive, quantitative, and accurate measure of liver fat content.[43] With regard to imaging of liver fibrosis, most clinical studies have been performed with transient elastography (FibroScan®, EchoSens).[44-46] The liver stiffness measurement (LSM) interquartile range/median LSM >0.3 kilopascals (kPa) and LSM ≥7.1 kPa were regarded as unreliable.[47]

In addition, biochemical assessment of FLD including alanine transaminase (ALT), aspartate aminotransferase (AST), and gamma-glutamyl transpeptidase (GGT) has been traditionally used to suggest FLD after the exclusion of secondary causes of steatosis. Nowadays, several algorithms incorporating multiple clinical and biochemical parameters have been evaluated for the diagnosis of simple steatosis. These surrogate markers include fatty liver index (FLI), an algorithm formula consisting of body mass index (BMI), waist circumference (WC), and serum levels of triglycerides and GGT to detect FLD with AUC (0.84).[48] While Fibrosis-4 (FIB-4) is widely used for the detection of liver fibrosis with AUC (0.76) and it is variables comprised of age, AST, ALT, and platelets and has been shown to have the best diagnostic accuracy for advanced fibrosis when compared with other noninvasive clinical scores.[49]

Prevalence and diagnosis of obesity

Obesity defined as a body mass index (BMI) of above 25 kg/m^2, is a global public health issue, with over 2.2 billion people meeting the definition of overweight or obese in 2015.[50] This condition is commonly associated with other metabolic disorders, such as type 2 diabetes, non-alcoholic fatty liver disease, cardiovascular diseases, and cancers.[51] The pathogenesis of obesity involves regulation of calorie utilization, appetite, and physical activity, but have complex interactions with availability of health-care systems, the role of socio-economic status, and underlying hereditary and environmental factors.[52] To manage and treat obesity, several approaches are commonly employed, including dietary modifications, exercise, supplementation with probiotics/prebiotics, medication, bariatric surgery, and behavioral interventions.[53] These strategies aim to address the complex nature of obesity and promote healthier lifestyles to improve overall well-being.

Fortunately, diagnosing obesity is less complicated than the diagnosis of FLD. This is primarily due to the availability of more simple and commonly used anthropometric methods for assessing obesity, which include measuring weight, stature, abdominal circumference, and skinfold measurements.[54] While more complex methods include bioelectrical impedance, dual-energy X-ray absorptiometry (DEXA), body density, and total body water estimates. BMI was calculated as weight (Kg) divided by height (m^2). Obesity by BMI was defined as BMI ≥ 30 kg/m^2.[55] WC was measured at the mid-point between the lower border of the ribs and the iliac crest. Women with WC ≥ 88 cm were classified as obese.[55] Waist-hip ratio (WHR) was calculated by dividing the average WC by the average hip circumference. WHR ≥ 0.85 were classified as obese in women.[55] In recent years, DEXA has become the reference tool in clinical routine to measure body composition and fat mass (FM) distribution.[56, 57] Furthermore, DEXA measurements are considered a strong and independent predictor of type 2 diabetes (T2D) and cardiovascular disease.[58] To determine the fat mass index (FMI) was calculated by dividing the total fat mass (in kilograms) by height squared in meters (m^2). FMI classified obesity if higher than 9 kg/m^2 in men and higher than 13 kg/m^2 in women.[59] Similar to FMI, the fat-free mass index (FFMI) was calculated by dividing fat-free mass (kg) by the height squared in meters (m^2). The values were 8.3 and 11.8 kg/m^2 in men and women with obese BMI.[60] While Android to Gynoid ratio (AGR) was defined as android fat divided by gynoid fat. An android/gynoid ratio greater than 1 would determine this and you may be at more risk of having a high visceral fat.

Omics data

The suffix -omics (Greek word) refers to a field of study in life sciences that focuses on large-scale data/information to understand life summed up in "omes" and "omics" such as genomics, epigenomics, transcriptomics, proteomics, and metabolomics.[61] These terms represent a comprehensive study of a genome, epigenome, transcriptome, proteome, and metabolome respectively. Integrating different omics layers and their effect on obesity and fatty liver disease can lead to a deeper comprehension of the underlying mechanisms of these diseases. Consequently, this knowledge can enhance the effectiveness of identifying novel biomarkers for early diagnosis of obesity and FLD, and ultimately may open up opportunities for the development of improved diagnostic methods and potential therapeutic targets **(Figure 3)**.

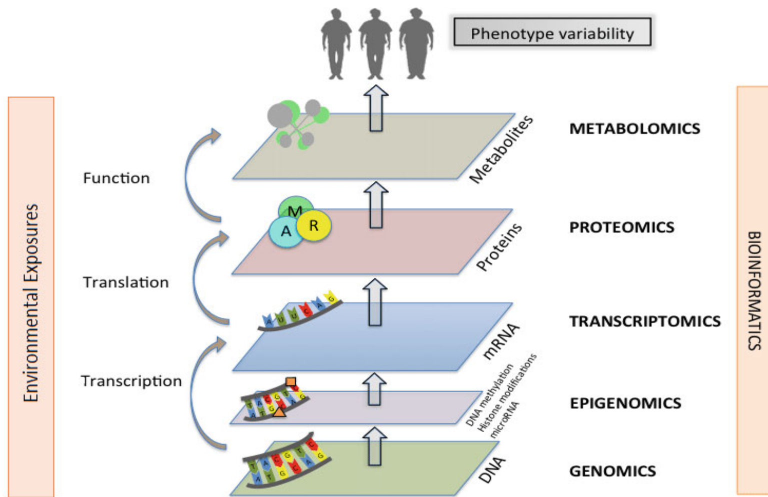


Figure 3. Multi-omics layers and the interactions with environmental factors in biological systems (Figure created with Br Med Bull, Volume 123, Issue 1, September 2017, Pages 159–173, https://doi.org/10.1093/bmb/ldx022).

Genetics

In the last two decades, substantial advances in the discovery of genetic determinants of both obesity and FLD have been made. The genetic variants in lipid metabolism genes are currently being studied to decipher their role in the predisposition toward liver disease. Both genome-wide association studies (GWAS) and candidate gene approaches have identified a number of variants, which have enriched our understanding of the genetic basis of NAFLD.[62] One of the most studied genes is patatin-like phospholipase domain-containing protein 3 (PNPLA3*)* that predisposes to a higher risk of fatty liver diseases, in particular, non-synonymous variant, rs738409 C>G (p.Ile148Met),[63] in addition to the fat mass and obesity-associated protein (FTO), the b-Klotho (KLB) and carboxylesterase (CES1).[64] The candidate gene approach was the predominant method used to study susceptibility to NAFLD until the first GWAS was published in 2008. This approach primarily focused on genes associated with various aspects of hepatic lipid metabolism (such as synthesis, storage, export, and oxidation), glycaemic metabolism, insulin signaling (including insulin resistance), as well as genes involved in immune responses, oxidative stress, inflammation, and fibrosis.[65-69] However, the majority of this research used relatively small sample sizes, and due to their limitations, the reported results were frequently inconsistent. In recent years, GWAS is used to screen the entire genome of large numbers of individuals to look for associations between millions of genetic variants within those individuals and their disease outcomes. It is based on the common disease-common variant hypothesis. These studies have helped to define common gene variants (minor allele frequency >1-5 %) that could potentially contribute to NAFLD. Additionally,

several large-scale GWAS on liver enzymes have been conducted, which help in our understanding of the mechanisms behind FLD.[70-73]

Epigenomics

Epigenome is the complete description of all the chemical modifications to DNA and histone proteins that regulate the expression of genes within the genome, without changing the DNA sequencing.[74, 75] Epigenetics is a reversible system that can be affected by genetic (heritable) and various environmental factors such as drugs and nutrition. Epigenetic marks, including DNA methylation, histone modification, and non-coding RNAs, are thought to influence gene expression primarily at the level of transcription; however, other steps in the process (for example, translation) may also be regulated epigenetically.[76] A recent study of DNA and its alterations has considerably increased our understanding of the function of epigenetics in regulating energy metabolism and expenditure in obesity and metabolic diseases.[77, 78] Moreover, recent advances have been made in understanding the potential roles of miRNAs in epigenetic regulation of obesity and FLD.[79, 80]

DNA methylation is the most widely studied epigenetic mechanism and refers to the addition of a methyl (-CH3) group to the DNA strand itself, often to the 5' carbon atom of a cytosine ring. This conversion of cytosine bases to 5-methylcytosine is catalysed by DNA methyltransferases (DNMTs).[81] These modified cytosine residues usually lie next to a guanine base (CpG methylation) and the result is two methylated cytosines positioned diagonally to each other on opposite strands of DNA. Of the roughly 28 million CpGs in the human genome, 60%–80% are generally methylated.[82] Over the past two decades, there has been a substantial increase in the number of publications focusing on DNA methylation in liver disease. Through omics analysis of the liver, researchers have found aberrant pathways linked to obesity and diabetes. This is accompanied by DNA hypomethylation of genes at the ATF-motif regulatory site. The activation of these pathways, coupled with *PRKCE* activation and hypomethylation, may contribute to the development of hepatic insulin resistance and steatosis.[83] A previous epigenome-wide association study (EWAS) reported associations between DNA methylation sites in *SLC7A11, SLC1A5, SLC43A1,* and phosphoglycerate dehydrogenase (*PHGDH*) genes with FLD.[84] Experimental evidence suggests that *SLC7A11* plays a role in lipid metabolism.

Together with DNA methylation, histone modifications (including histone acetylation, methylation, phosphorylation, ubiquitination, etc.) and non-coding RNAs represent the most well-known epigenetic mechanisms. It was demonstrated that histone modifications (acetylation) in fatty acid synthase modulated by carbohydrate-responsive element binding protein are associated with NAFLD.[85] In addition, changes in histone modification are a key component of an epigenetic network controlling adipogenesis, energy homeostasis, and obesity.[86]

There is a growing body of evidence that non-coding RNAs play an important epigenetic role in the pathogenesis of complex diseases. MicroRNAs are short non-coding RNA molecules that post-transcriptionally repress the expression of genes by binding mainly to 3' untranslated regions of their target mRNAs.[87, 88] Emerging evidence shows that miRNAs help to distinguish NAFLD and NASH severity, particularly, miR-34a, miR-192, and miR-122 (Ref). In addition, miR-122 showed moderate accuracy (AUC 0.82) to distinguish NAFLD from healthy controls.[89] Notably, miR-122 is the most abundant and specific miRNA in the liver.[90]

Epigenomics, in particular DNA methylation, is also widely studied to understand mechanism of obesity. Previous EWAS on obesity traits were conducted in population-based studies, including Rotterdam Study and Atherosclerosis Risk in Communities (ARIC),[91] the investigators reported a novel association between increased methylation in the *MSI2* and *LARS2* genes and higher BMI and WC in older adults. Moreover, CpG sites at *BRDT* and *MAP1A* were associated with BMI, and CpG sites at *TMEM49* and *LGALS3BP* were associated with WC. In addition, this study also confirmed 3 previously identified methylation loci (*CPT1A*, *ABCG1*, and *SREBF1*) to be associated with obesity-related traits. Lastly, several studies have demonstrated the association of miRNAs levels with obesity and fat distribution such as miR-196a2 and miR-196a, by using a genetic approach.[92, 93] Likewise, miR-146, miR-378, miR-143, miR-145, and miR-194 are shown to be involved in inflammatory processes that occur during obesity, while other miRNAs, including miR-196, found to be modulated during inflammatory processes in cancer, cardiovascular diseases or type 2 diabetes.[94] In both diseases, FLD as well as obesity, DNA methylation is being studied the most.

Metabolomics

Metabolome is the study of endogenous and exogenous metabolites in biological specimens (e.g. blood, urine, saliva, cerebral spinal fluid) and aims to provide semi-quantitative information on metabolite abundances in a biological system.[95] This includes fatty acids, amino acids, peptides, and carbohydrates. Recent improvements in metabolomics technologies reveal the unequivocal value of metabolomic tools in biomarker discovery, gene-function analysis, systems biology, and diagnostic platforms.[96] Metabolites represent the downstream expression of the genome, transcriptome, and proteome, thus helping to provide deep insight into disease pathophysiology.[97] In recent years, the development of instrumental systems, such as high-resolution nuclear magnetic resonance spectroscopy (NMR) and mass spectrometry (MS), ultra-performance liquid chromatography (ULC-MS), and more sophisticated bioinformatics and analytical techniques, have enabled more comprehensive coverage of the metabolome. Over the past two decades, NMR has emerged as one of the three principal analytical techniques used in metabolomics (the other two being gas chromatography coupled to mass spectrometry (GC-MS) and liquid chromatography coupled with single-stage mass spectrometry (LC-MS)).[98] NMR spectroscopy plays important and multifaceted roles that have benefited and continue to benefit the field of metabolomics. The unique

characteristics of NMR are valued for being non-invasive, nondestructive, fast, and for providing highly reproducible results.[99] MS has witnessed a very rapid growth in clinical metabolomics with targeted and non-targeted approaches over the past decade.[100] MS has emerged as a key technology for the selective and sensitive analysis of metabolites in biological samples, providing the ability to quantify and identify metabolites.[101]

NAFLD is strongly associated with well-known metabolic risk factors such as obesity and insulin resistance. Atherogenic dyslipidemia, characterized by plasma hypertriglyceridemia, increased small dense low-density lipoprotein (LDL) particles and decreased high-density lipoprotein cholesterol (HDL-C) levels, is often observed in NAFLD patients.[102] It has been shown that an increase in intrahepatic triglycerides (IHTG) is related to the development of steatosis in individuals with NAFLD.[103-105] The accumulation of triglycerides (TG) in liver tissue happens when the rate of hepatic TG production is greater than the combined rates of TG export in very-low-density lipoprotein (VLDL) particles and intrahepatic oxidation of TG-derived fatty acids. Lipidomics is a new rapidly growing field that allows the overall and detailed investigation of the whole lipid composition in a given biological matrix. Lipid profiling of liver biopsies of patients with NAFLD has previously revealed several changes in glycerophospholipids and sphingolipids concentrations and alterations in fatty acid pattern compared to healthy control, which correlates with disease progression.[106] New promising non-invasive biomarkers and techniques have been developed, evaluated, and assessed, including biochemical markers, imaging modalities, and the most recent multi-omics approaches.[107] Nevertheless, among the “omics”, metabolomics/lipidomics have the best potential for leading to the development of important tools for the diagnosis and staging of FLD.[108] Integration of metabolome and lipidome offers a complete atlas of the metabolic landscape,[109] enabling comprehensive network analysis to identify critical metabolic drivers in disease pathology, facilitating the study of interconnection between lipids and other metabolites in disease progression. Moreover, previous studies provide evidence for changes in phospholipid and amino acid metabolism that may be linked to obesity based on *FTO* genotype.[110] This may contribute to a better understanding of the biochemical networks underlying the development of obesity.

Proteomics

Proteomics is a powerful tool in the study of changes in the protein expression of proteomes of different populations of patients.[111] Unlike genes, a variety of factors, such as the environment, the stage of the disease, the use of medications, and dietary habits, can affect protein levels. Since disease-related molecular changes are reflected at the transcriptome and proteome level, the identification of these proteins may have a huge impact by increasing the availability of molecular markers for early diagnosis and therapy and providing a better understanding of the underlying pathophysiology.[112] Recent technological advances now make it possible to measure a large number of proteins in a large number of individuals.[113] Two main techniques can measure the concentration of thousands human plasma proteins,

such as SomaScan and Olink technologies. Both techniques have been shown to be successful in identifying novel biomarkers, while a comparison of the techniques showed the synergistic nature of these technologies to better identify disease mechanisms. The Olink PEA technology uses a dual recognition DNA-coupled immunoassay that rapidly allows for protein identification and relative quantification with high sensitivity and specificity to provide unique, enabling tools for protein biomarker discovery and development.[114] A current meta-analysis study demonstrated that fibroblast growth factor 21 (FGF-21) and cytokeratin 18 (CK-18) could be used as biomarkers to diagnose NAFLD, especially NASH.[115] Furthermore, other studies found the potential role of FGF-21 as a biomarker for NAFLD,[116, 117] which is secreted in the liver in response to peroxisome proliferator-activated receptor (PPAR)-α activation. The identification of these proteins may have a huge impact by increasing the availability of molecular markers for early diagnosis and therapy monitoring.

Objective of this thesis and outline

The overall aim of this thesis is to investigate the molecular determinants and underlying pathways of obesity and FLD and to identify potential biomarkers for their early diagnosis. This aim was studied by integrating various population-based omics data as well as clinical data available in the Rotterdam Study and through conducting different advanced molecular epidemiological studies. In **Chapter 2** I mainly focused on the epigenetic regulation of obesity and FLD. In **Chapter 2.1**, I investigated the association between plasma circulating miRNAs and obesity-related traits in a population-based setting; in **Chapter 2.2**, I aimed to study the potential circulatory miRNAs as regulators and biomarkers for FLD; and in **Chapter 2.3**, I study the association between alcohol consumption and circulatory miRNAs, then whether these miRNA could mediate the association between alcohol consumption and FLD. In **Chapter 3**, I focused on the proteomic determinants of FLD. In **this chapter**, I aimed to identify specific plasma proteins associated with FLD and liver fibrosis using population-based data. Finally, in **Chapter 4**, I assessed the relationships between circulating metabolites and NAFLD. In **this chapter**, I used data on circulating metabolites and NAFLD in Rotterdam study (RS), Avon Longitudinal Study of Parents and Children (ALSPAC), The Insulin Resistance Atherosclerosis Family Study (IRASFS) and Study of Latinos (SOL) to conduct a multi-platform based meta-analysis.

Study population

The Rotterdam Study

The majority of studies described in this thesis were conducted within the Rotterdam Study, a large population-based cohort study, also known as "Erasmus Rotterdam Gezondheid Onderzoek (ERGO)".[118] The Rotterdam study is an ongoing prospective cohort study the first subcohort (RS-I) was initiated in 1990 with individuals ≥55 y of age (N = 7983). The study was extended by including a second subcohort (RS-II) in 2000 (N = 3011, ≥55 y of age), and a third subcohort (RS-III) in 2006 (N = 3932, ≥45 y of age). In June 2016, the recruitment of another extension (RS-IV) began to enroll participants (N= 3005, ≥40 y age). All participants in the study provided written informed consent to participate and to obtain information from their treating physicians. The participants were examined at the baseline of the respective subcohort and re-examined every 3-5 years. Shortly, a home interview was conducted and the participants had an extensive set of examinations in the research center in Ommoord. I used miRNA profiling data which has been conducted in 2,000 randomly selected individuals from the fourth round of Rotterdam Study I (RS-I-4) and the second round of Rotterdam Study II (RS-II-2) between January 2002 and December 2005. In addition, proteomics data was used from 3,596 individuals of the first visit of the third cohort (RS-III-1). Nightingale-based metabolomics data were used from the fifth visit of the first cohort (RS-I-5), a third visit of the second cohort (RS-II-3), and a second visit of the third cohort (RS-III-2). While metabolon-based metabolomics data were used from the fourth visit of the first cohort (RS-I-4) and the second visit of the third cohort (RS-III-2). **Figure 4** illustrates molecular layers across different cohorts of the Rotterdam Study used in this thesis.

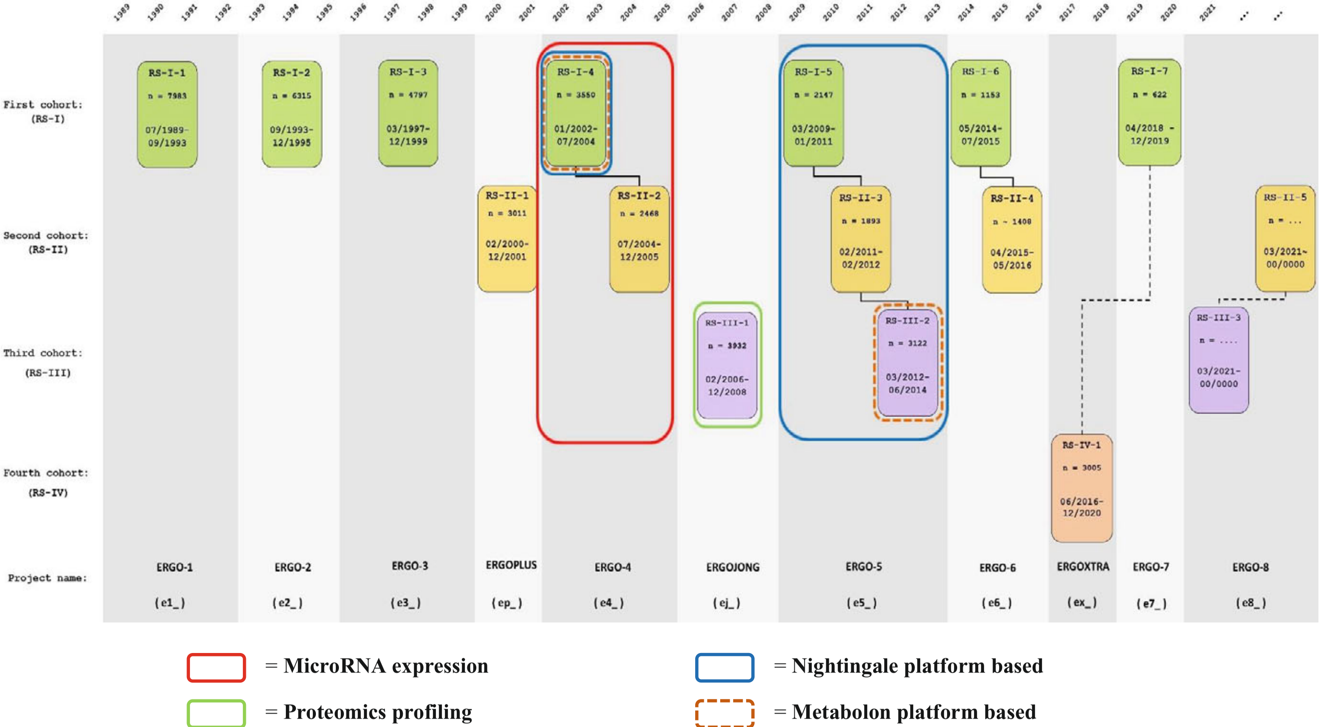


Figure 4. Different omics dataset of the Rotterdam Study. Red indicates cohorts in which miRNA expression was conducted (RS-1-4 and RS-II-2). Green color indicates cohorts in which proteomics profiling was conducted (RS-III-2). Blue indicates cohorts in which Nightingale platform-based metabolomics was conducted (RS-1-4, RS-I-5, RS-II-3, and RS-III-2). Orange indicates cohorts in which metabolon platform-based metabolomics was conducted (RS-1-4 and RS-III-2). Abbreviation: RS, Rotterdam Study.

References

1. Abdel-Misih SRZ, Bloomston M. Liver anatomy. Surgical Clinics. 2010;90(4):643-53.
2. Kalra A, Yetiskul E, Wehrle CJ, Tuma F. Physiology, Liver. 2022.
3. Zhang X. Molecular Epidemiology towards Deciphering Underlying Mechanisms of Fatty Liver Disease. 2022.
4. Tilg H, Adolph TE, Dudek M, Knolle P. Non-alcoholic fatty liver disease: the interplay between metabolism, microbes and immunity. Nat Metab. 2021;3(12):1596-607.
5. Åberg F, Färkkilä M, editors. Drinking and obesity: alcoholic liver disease/nonalcoholic fatty liver disease interactions. Seminars in liver disease; 2020: Thieme Medical Publishers.
6. Younossi ZM, Koenig AB, 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):73-84.
7. Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, et al. The diagnosis and management of nonalcoholic fatty liver disease: practice guidance from the American Association for the Study of Liver Diseases. Hepatology. 2018;67(1):328-57.
8. Le MH, Yeo YH, Li X, Li J, Zou B, Wu Y, et al. 2019 Global NAFLD prevalence: a systematic review and meta-analysis. Clinical Gastroenterology and Hepatology. 2022;20(12):2809-17. e28.
9. Liu J, Ayada I, Zhang X, Wang L, Li Y, Wen T, et al. Estimating global prevalence of metabolic dysfunction-associated fatty liver disease in overweight or obese adults. Clinical Gastroenterology and Hepatology. 2022;20(3):e573-e82.
10. Sanyal AJ, Harrison SA, Ratziu V, Abdelmalek MF, Diehl AM, Caldwell S, et al. The natural history of advanced fibrosis due to nonalcoholic steatohepatitis: data from the simtuzumab trials. Hepatology. 2019;70(6):1913-27.
11. Kanwal F, Kramer JR, Mapakshi S, Natarajan Y, Chayanupatkul M, Richardson PA, et al. Risk of hepatocellular cancer in patients with non-alcoholic fatty liver disease. Gastroenterology. 2018;155(6):1828-37. e2.
12. Huang DQ, El-Serag HB, Loomba R. Global epidemiology of NAFLD-related HCC: trends, predictions, risk factors and prevention. Nature reviews gastroenterology & hepatology. 2021;18(4):223-38.
13. Harrison SA, Bashir MR, Guy CD, Zhou R, Moylan CA, Frias JP, et al. Resmetirom (MGL-3196) for the treatment of non-alcoholic steatohepatitis: a multicentre, randomised, double-blind, placebo-controlled, phase 2 trial. The Lancet. 2019;394(10213):2012-24.
14. Esteban JP, Dinani A. Lifestyle interventions beyond diet and exercise for patients with nonalcoholic fatty liver disease. Gastroenterology & Hepatology. 2020;16(3):119.
15. Ayonrinde OT. Historical narrative from fatty liver in the nineteenth century to contemporary NAFLD–Reconciling the present with the past. JHEP Reports. 2021;3(3):100261.
16. Lonardo A, Leoni S, Alswat KA, Fouad Y. History of nonalcoholic fatty liver disease. International Journal of Molecular Sciences. 2020;21(16):5888.
17. Fouad Y, Waked I, Bollipo S, Gomaa A, Ajlouni Y, Attia D. What's in a name? Renaming 'NAFLD'to 'MAFLD'. Liver international. 2020;40(6):1254-61.
18. European Association for the Study of the L, European Association for the Study of D, European Association for the Study of O. EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. J Hepatol. 2016;64(6):1388-402.
19. Siddiqui MS, Harrison SA, Abdelmalek MF, Anstee QM, Bedossa P, Castera L, et al. Case definitions for inclusion and analysis of endpoints in clinical trials for nonalcoholic steatohepatitis through the lens of regulatory science. Hepatology. 2018;67(5):2001-12.
20. Shiha G, Alswat K, Al Khatry M, Sharara AI, Örmeci N, Waked I, et al. Nomenclature and definition of metabolic-associated fatty liver disease: a consensus from the Middle East and north Africa. The lancet Gastroenterology & hepatology. 2021;6(1):57-64.
21. Fouad Y, Dufour JF, Zheng MH, Bollipo S, Desalegn H, Grønbaek H, et al. The NAFLD-MAFLD debate: Is there a Consensus-on-Consensus methodology? Liver International. 2022;42(4):742-8.

22. Ayada I, van Kleef LA, Alferink LJM, Li P, de Knegt RJ, Pan Q. Systematically comparing epidemiological and clinical features of MAFLD and NAFLD by meta-analysis: focusing on the non-overlap groups. Liver International. 2022;42(2):277-87.
23. Rinella ME, Lazarus JV, Ratziu V, Francque SM, Sanyal AJ, Kanwal F, et al. A multi-society Delphi consensus statement on new fatty liver disease nomenclature. Hepatology (Baltimore, Md). 2023.
24. Day CP, James OFW. Steatohepatitis: a tale of two "hits"? : Elsevier; 1998. p. 842-5.
25. Friedman SL, Neuschwander-Tetri BA, Rinella M, Sanyal AJ. Mechanisms of NAFLD development and therapeutic strategies. Nature medicine. 2018;24(7):908-22.
26. Lonardo A, Nascimbeni F, Maurantonio M, Marrazzo A, Rinaldi L, Adinolfi LE. Nonalcoholic fatty liver disease: Evolving paradigms. World journal of gastroenterology. 2017;23(36):6571.
27. Tilg H, Adolph TE, Moschen AR. Multiple parallel hits hypothesis in nonalcoholic fatty liver disease: revisited after a decade. Hepatology (Baltimore, Md). 2021;73(2):833.
28. Arab JP, Arrese M, Trauner M. Recent insights into the pathogenesis of nonalcoholic fatty liver disease. Annual Review of Pathology: Mechanisms of Disease. 2018;13:321-50.
29. Caturano A, Acierno C, Nevola R, Pafundi PC, Galiero R, Rinaldi L, et al. Non-alcoholic fatty liver disease: From pathogenesis to clinical impact. Processes. 2021;9(1):135.
30. Jacome-Sosa MM, Parks EJ. Fatty acid sources and their fluxes as they contribute to plasma triglyceride concentrations and fatty liver in humans. Current opinion in lipidology. 2014;25(3):213-20.
31. Severson TJ, Besur S, Bonkovsky HL. Genetic factors that affect nonalcoholic fatty liver disease: A systematic clinical review. World journal of gastroenterology. 2016;22(29):6742.
32. BasuRay S, Wang Y, Smagris E, Cohen JC, Hobbs HH. Accumulation of PNPLA3 on lipid droplets is the basis of associated hepatic steatosis. Proceedings of the National Academy of Sciences. 2019;116(19):9521-6.
33. Cohen JC, Horton JD, Hobbs HH. Human fatty liver disease: old questions and new insights. Science. 2011;332(6037):1519-23.
34. Zhang X, Asllanaj E, Amiri M, Portilla-Fernandez E, Bramer WM, Nano J, et al. Deciphering the role of epigenetic modifications in fatty liver disease: A systematic review. European journal of clinical investigation. 2021;51(5):e13479.
35. Pezzino S, Sofia M, Faletra G, Mazzone C, Litrico G, La Greca G, et al. Gut–Liver Axis and Non-Alcoholic Fatty Liver Disease: A Vicious Circle of Dysfunctions Orchestrated by the Gut Microbiome. Biology. 2022;11(11):1622.
36. Muscogiuri G, Cantone E, Cassarano S, Tuccinardi D, Barrea L, Savastano S, et al. Gut microbiota: a new path to treat obesity. International journal of obesity supplements. 2019;9(1):10-9.
37. Liu X, Green RM. Endoplasmic reticulum stress and liver diseases. Liver research. 2019;3(1):55-64.
38. Buzzetti E, Pinzani M, Tsochatzis EA. The multiple-hit pathogenesis of non-alcoholic fatty liver disease (NAFLD). Metabolism. 2016;65(8):1038-48.
39. Boyum JH, Atwell TD, Schmit GD, Poterucha JJ, Schleck CD, Harmsen WS, et al., editors. Incidence and risk factors for adverse events related to image-guided liver biopsy. Mayo Clinic Proceedings; 2016: Elsevier.
40. Li Q, Dhyani M, Grajo JR, Sirlin C, Samir AE. Current status of imaging in nonalcoholic fatty liver disease. World journal of hepatology. 2018;10(8):530.
41. Caussy C, Johansson L. Magnetic resonance-based biomarkers in nonalcoholic fatty liver disease and nonalcoholic steatohepatitis. Endocrinology, Diabetes & Metabolism. 2020;3(4):e00134.
42. Ajmera V, Loomba R. Imaging biomarkers of NAFLD, NASH, and fibrosis. Molecular Metabolism. 2021;50:101167.
43. Caussy C, Reeder SB, Sirlin CB, Loomba R. Noninvasive, quantitative assessment of liver fat by MRI-PDFF as an endpoint in NASH trials. Hepatology. 2018;68(2):763-72.
44. Fraquelli M, Rigamonti C. Diagnosis of cirrhosis by transient elastography: what is hidden behind misleading results. Hepatology. 2007;46(1):282.
45. Gomez-Dominguez E, Mendoza J, Garcia-Buey L, Trapero M, Gisbert JP, Jones EA, et al. Transient elastography to assess hepatic fibrosis in primary biliary cirrhosis. Alimentary pharmacology & therapeutics. 2008;27(5):441-7.

46. Friedrich-Rust M, Wunder K, Kriener S, Sotoudeh F, Richter S, Bojunga J, et al. Liver fibrosis in viral hepatitis: noninvasive assessment with acoustic radiation force impulse imaging versus transient elastography. Radiology. 2009;252(2):595-604.
47. El-Kamary SS, Mohamed MM, El-Raziky M, Shardell MD, Shaker OG, ElAkel WA, et al. Liver fibrosis staging through a stepwise analysis of non-invasive markers (FibroSteps) in patients with chronic hepatitis C infection. Liver International. 2013;33(7):982-90.
48. Bedogni G, Bellentani S, Miglioli L, Masutti F, Passalacqua M, Castiglione A, et al. The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population. BMC gastroenterology. 2006;6(1):1-7.
49. McPherson S, Stewart SF, Henderson E, Burt AD, Day CP. Simple non-invasive fibrosis scoring systems can reliably exclude advanced fibrosis in patients with non-alcoholic fatty liver disease. Gut. 2010;59(9):1265-9.
50. Seidell JC, Halberstadt J. The global burden of obesity and the challenges of prevention. Annals of Nutrition and Metabolism. 2015;66(Suppl. 2):7-12.
51. Yang M, Liu S, Zhang C, editors. The related metabolic diseases and treatments of obesity. Healthcare; 2022: MDPI.
52. Lin X, Li H. Obesity: Epidemiology, Pathophysiology, and Therapeutics. Frontiers in Endocrinology. 2021:1070.
53. Baker JS, Supriya R, Dutheil F, Gao Y. Obesity: treatments, conceptualizations, and future directions for a growing problem. Biology. 2022;11(2):160.
54. Duren DL, Sherwood RJ, Czerwinski SA, Lee M, Choh AC, Siervogel RM, et al. Body composition methods: comparisons and interpretation. Journal of diabetes science and technology. 2008;2(6):1139-46.
55. World Health O. Obesity: preventing and managing the global epidemic. 2000.
56. Kaul S, Rothney MP, Peters DM, Wacker WK, Davis CE, Shapiro MD, et al. Dual-energy X-ray absorptiometry for quantification of visceral fat. Obesity. 2012;20(6):1313-8.
57. Andreoli A, Scalzo G, Masala S, Tarantino U, Guglielmi G. Body composition assessment by dual-energy X-ray absorptiometry (DXA). La radiologia medica. 2009;114(2):286-300.
58. Vasan SK, Osmond C, Canoy D, Christodoulides C, Neville MJ, Di Gravio C, et al. Comparison of regional fat measurements by dual-energy X-ray absorptiometry and conventional anthropometry and their association with markers of diabetes and cardiovascular disease risk. International journal of obesity. 2018;42(4):850-7.
59. Kelly TL, Wilson KE, Heymsfield SB. Dual energy X-Ray absorptiometry body composition reference values from NHANES. PloS one. 2009;4(9):e7038.
60. Kyle UG, Schutz Y, Dupertuis YM, Pichard C. Body composition interpretation: contributions of the fat-free mass index and the body fat mass index. Nutrition. 2003;19(7-8):597-604.
61. Yadav SP. The wholeness in suffix -omics, -omes, and the word om. J Biomol Tech. 2007;18(5):277.
62. Eslam M, George J. Genetic and epigenetic mechanisms of NASH. Hepatology international. 2016;10(3):394-406.
63. Trépo E, Romeo S, Zucman-Rossi J, Nahon P. PNPLA3 gene in liver diseases. Journal of hepatology. 2016;65(2):399-412.
64. Rivera-Iñiguez I, Panduro A, Roman S, González-Aldaco K. What do we know about nutrient-based strategies targeting molecular mechanisms associated with obesity-related fatty liver disease? Annals of Hepatology. 2023;28(1):100874.
65. Gambino R, Cassader M, Pagano G, Durazzo M, Musso G. Polymorphism in microsomal triglyceride transfer protein: a link between liver disease and atherogenic postprandial lipid profile in NASH? Hepatology. 2007;45(5):1097-107.
66. Dong H, Wang J, Li C, Hirose A, Nozaki Y, Takahashi M, et al. The phosphatidylethanolamine N-methyltransferase gene V175M single nucleotide polymorphism confers the susceptibility to NASH in Japanese population. Journal of hepatology. 2007;46(5):915-20.
67. Namikawa C, Shu-Ping Z, Vyselaar JR, Nozaki Y, Nemoto Y, Ono M, et al. Polymorphisms of microsomal triglyceride transfer protein gene and manganese superoxide dismutase gene in non-alcoholic steatohepatitis. Journal of hepatology. 2004;40(5):781-6.

68. Wong VWS, Wong GLH, Tsang SWC, Hui AY, Chan AWH, Choi PCL, et al. Genetic polymorphisms of adiponectin and tumor necrosis factor-alpha and nonalcoholic fatty liver disease in Chinese people. Journal of gastroenterology and hepatology. 2008;23(6):914-21.
69. Dixon JB, Bhathal PS, Jonsson JR, Dixon AF, Powell EE, O'Brien PE. Pro-fibrotic polymorphisms predictive of advanced liver fibrosis in the severely obese. Journal of hepatology. 2003;39(6):967-71.
70. Pazoki R, Vujkovic M, Elliott J, Evangelou E, Gill D, Ghanbari M, et al. Genetic analysis in European ancestry individuals identifies 517 loci associated with liver enzymes. Nature communications. 2021;12(1):1-12.
71. Chen VL, Du X, Chen Y, Kuppa A, Handelman SK, Vohnoutka RB, et al. Genome-wide association study of serum liver enzymes implicates diverse metabolic and liver pathology. Nature communications. 2021;12(1):1-13.
72. Chambers JC, Zhang W, Sehmi J, Li X, Wass MN, Van der Harst P, et al. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. Nature genetics. 2011;43(11):1131-8.
73. Seo JY, Lee J-E, Chung GE, Shin E, Kwak M-S, Yang JI, et al. A genome-wide association study on liver enzymes in Korean population. Plos one. 2020;15(2):e0229374.
74. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. Genome biology. 2017;18(1):1-15.
75. Lee J, Kim Y, Friso S, Choi S-W. Epigenetics in non-alcoholic fatty liver disease. Molecular aspects of medicine. 2017;54:78-88.
76. Gibney ER, Nolan CM. Epigenetics and gene expression. Heredity. 2010;105(1):4-13.
77. Modi A, Khokhar M, Sharma P, Joshi R, Mishra SS, Bharshankar RN, et al. Leptin DNA Methylation and Its Association with Metabolic Risk Factors in a Northwest Indian Obese Population. Journal of Obesity & Metabolic Syndrome. 2021;30(3):304.
78. Mahmoud AM. An overview of epigenetics in obesity: The role of lifestyle and therapeutic interventions. International Journal of Molecular Sciences. 2022;23(3):1341.
79. Ramzan F, Vickers MH, Mithen RF. Epigenetics, microRNA and metabolic syndrome: a comprehensive review. International Journal of Molecular Sciences. 2021;22(9):5047.
80. Lin H-Y, Yang Y-L, Wang P-W, Wang F-S, Huang Y-H. The emerging role of microRNAs in NAFLD: highlight of microRNA-29a in modulating oxidative stress, inflammation, and beyond. Cells. 2020;9(4):1041.
81. Fernandez A, O'Leary C, O'Byrne KJ, Burgess J, Richard DJ, Suraweera A. Epigenetic mechanisms in DNA double strand break repair: A clinical review. Frontiers in Molecular Biosciences. 2021;8:653.
82. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. Nature Reviews Genetics. 2013;14(3):204-20.
83. Kirchner H, Sinha I, Gao H, Ruby MA, Schönke M, Lindvall JM, et al. Altered DNA methylation of glycolytic and lipogenic genes in liver from obese and type 2 diabetic patients. Molecular metabolism. 2016;5(3):171-83.
84. Nano J, Ghanbari M, Wang W, de Vries PS, Dhana K, Muka T, et al. Epigenome-wide association study identifies methylation sites associated with liver enzymes and hepatic steatosis. Gastroenterology. 2017;153(4):1096-106. e2.
85. Cai C, Yu H, Huang G, Du X, Yu X, Zhou Y, et al. Histone modifications in fatty acid synthase modulated by carbohydrate responsive element binding protein are associated with non-alcoholic fatty liver disease. International Journal of Molecular Medicine. 2018;42(3):1215-28.
86. Okamura M, Inagaki T, Tanaka T, Sakai J. Role of histone methylation and demethylation in adipogenesis and obesity. Organogenesis. 2010;6(1):24-32.
87. Landrier J-F, Derghal A, Mounien L. MicroRNAs in obesity and related metabolic disorders. Cells. 2019;8(8):859.
88. Pirola CJ, Sookoian S. Metabolic dysfunction-associated fatty liver disease: advances in genetic and epigenetic implications. Current Opinion in Lipidology. 2022;33(2):95-102.
89. Liu C-H, Ampuero J, Gil-Gómez A, Montero-Vallejo R, Rojas Á, Muñoz-Hernández R, et al. miRNAs in patients with non-alcoholic fatty liver disease: A systematic review and meta-analysis. Journal of hepatology. 2018;69(6):1335-48.

90. Chun K-H. Molecular targets and signaling pathways of microRNA-122 in hepatocellular carcinoma. Pharmaceutics. 2022;14(7):1380.
91. Dhana K, Braun KVE, Nano J, Voortman T, Demerath EW, Guan W, et al. An epigenome-wide association study of obesity-related traits. American journal of epidemiology. 2018;187(8):1662-9.
92. Ghanbari M, Sedaghat S, De Looper HWJ, Hofman A, Erkeland SJ, Franco OH, et al. The association of common polymorphisms in mi R-196a2 with waist to hip ratio and mi R-1908 with serum lipid and glucose. Obesity. 2015;23(2):495-503.
93. Hilton C, Neville MJ, Wittemans LBL, Todorcevic M, Pinnick KE, Pulit SL, et al. MicroRNA-196a links human body fat distribution to adipose tissue extracellular matrix composition. EBioMedicine. 2019;44:467-75.
94. Brettfeld C, Maver A, Aumuller E, Peterlin B, Haslberger AG. MicroRNAs responsible for inflammation in obesity. Journal of Endocrinology and Metabolism. 2017;7(3):77-85.
95. Tan SZ, Begley P, Mullard G, Hollywood KA, Bishop PN. Introduction to metabolomics and its applications in ophthalmology. Eye. 2016;30(6):773-83.
96. Zhang A, Sun H, Yan G, Wang P, Wang X. Metabolomics for biomarker discovery: moving to the clinic. Biomed research international. 2015;2015.
97. Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. Nature reviews Molecular cell biology. 2016;17(7):451-9.
98. Emwas A-H, Roy R, McKay RT, Tenori L, Saccenti E, Gowda GAN, et al. NMR spectroscopy for metabolomics research. Metabolites. 2019;9(7):123.
99. Pan Z, Raftery D. Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics. Analytical and bioanalytical chemistry. 2007;387(2):525-7.
100. Zang X, Monge ME, Fernandez FM. Mass spectrometry-based non-targeted metabolic profiling for disease detection: Recent developments. TrAC Trends in Analytical Chemistry. 2019;118:158-69.
101. Ramautar R. Capillary electrophoresis–mass spectrometry for metabolomics–from metabolite analysis to metabolic profiling. 2018.
102. Deprince A, Haas JT, Staels B. Dysregulated lipid metabolism links NAFLD to cardiovascular disease. Molecular metabolism. 2020;42:101092.
103. Smith GI, Shankaran M, Yoshino M, Schweitzer GG, Chondronikola M, Beals JW, et al. Insulin resistance drives hepatic de novo lipogenesis in nonalcoholic fatty liver disease. The Journal of clinical investigation. 2020;130(3):1453-60.
104. Fabbrini E, Mohammed BS, Magkos F, Korenblat KM, Patterson BW, Klein S. Alterations in adipose tissue and hepatic lipid kinetics in obese men and women with nonalcoholic fatty liver disease. Gastroenterology. 2008;134(2):424-31.
105. Fabbrini E, Magkos F, Mohammed BS, Pietka T, Abumrad NA, Patterson BW, et al. Intrahepatic fat, not visceral fat, is linked with metabolic complications of obesity. Proceedings of the National Academy of Sciences. 2009;106(36):15430-5.
106. Kartsoli S, Kostara CE, Tsimihodimos V, Bairaktari ET, Christodoulou DK. Lipidomics in non-alcoholic fatty liver disease. World Journal of Hepatology. 2020;12(8):436.
107. Martinou E, Pericleous M, Stefanova I, Kaur V, Angelidi AM. Diagnostic modalities of non-alcoholic fatty liver disease: from biochemical biomarkers to multi-Omics non-Invasive approaches. Diagnostics. 2022;12(2):407.
108. Perakakis N, Stefanakis K, Mantzoros CS. The role of omics in the pathophysiology, diagnosis and treatment of non-alcoholic fatty liver disease. Metabolism. 2020;111:154320.
109. Wang R, Li B, Lam SM, Shui G. Integration of lipidomics and metabolomics for in-depth understanding of cellular mechanism and disease progression. Journal of genetics and genomics. 2020;47(2):69-83.
110. Kim Y-J, Lee H-S, Kim YK, Park S, Kim J-M, Yun JH, et al. Association of metabolites with obesity and type 2 diabetes based on FTO genotype. PloS one. 2016;11(6):e0156612.
111. Rodríguez-Suárez E, Duce AM, Caballería J, Arrieta FM, Fernández E, Gómara C, et al. Non-alcoholic fatty liver disease proteomics. PROTEOMICS–Clinical Applications. 2010;4(4):362-71.
112. Lovestone S, Güntert A, Hye A, Lynham S, Thambisetty M, Ward M. Proteomics of Alzheimer's disease: understanding mechanisms and seeking biomarkers. Expert Review of Proteomics. 2007;4(2):227-38.

113. Pietzner M, Wheeler E, Carrasco-Zanini J, Kerrison ND, Oerton E, Koprulu M, et al. Synergistic insights into human health from aptamer-and antibody-based proteomic profiling. Nature communications. 2021;12(1):1-13.
114. Assarsson E, Lundberg M, Holmquist G, Björkesten J, Bucht Thorsen S, Ekman D, et al. Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. PloS one. 2014;9(4):e95192.
115. He L, Deng L, Zhang Q, Guo J, Zhou J, Song W, et al. Diagnostic value of CK-18, FGF-21, and related biomarker panel in nonalcoholic fatty liver disease: a systematic review and meta-analysis. BioMed research international. 2017;2017.
116. Li H, Dong K, Fang Q, Hou X, Zhou M, Bao Y, et al. High serum level of fibroblast growth factor 21 is an independent predictor of non-alcoholic fatty liver disease: a 3-year prospective study in China. Journal of hepatology. 2013;58(3):557-63.
117. Cantero I, Abete I, Bullón-Vela V, Crujeiras AB, Casanueva FF, Zulet MA, et al. Fibroblast growth factor 21 levels and liver inflammatory biomarkers in obese subjects after weight loss. Archives of Medical Science: AMS. 2022;18(1):36.
118. Ikram MA, Brusselle G, Ghanbari M, Goedegebure A, Ikram MK, Kavousi M, et al. Objectives, design and main findings until 2020 from the Rotterdam Study. European journal of epidemiology. 2020;35(5):483-517.

Chapter 2

Epigenetic regulation of obesity traits and fatty liver disease

Chapter 2.1

Plasma circulating microRNAs associated with obesity, body fat distribution, and fat mass

Yasir J. Abozaid, Xiaofang Zhang*, Michelle M.J. Mens*, Fariba Ahmadizar, Marlou Limpens, M. Arfan Ikram, Fernando Rivadeneira, Trudy Voortman, Maryam Kavousi, Mohsen Ghanbari

International Journal of Obesity 46, no. 12 (2022): 2137-2144.

***Authors denotes equal contribution**

Abstract

Background: MicroRNAs (miRNAs) represent a class of small non-coding RNAs that regulate gene expression post-transcriptionally and are implicated in the pathogenesis of different diseases. Limited studies have investigated the association of circulating miRNAs with obesity and body fat distribution and their link to obesity-related diseases using population-based data.

Methods: We conducted a genome-wide profile of circulating miRNAs in plasma, collected between 2002 and 2005, in 1208 participants from the population-based Rotterdam Study cohort. Obesity and body fat distribution were measured as body mass index (BMI), waist-to-hip ratio (WHR), android-fat to gynoid-fat ratio (AGR) and fat mass index (FMI) measured by anthropometrics and Dual X-ray Absorptiometry. Multivariable linear regression models were used to assess the association of 591 miRNAs well-expressed in plasma with these traits adjusted for potential covariates. We further sought for the association of identified miRNAs with cardiovascular and metabolic diseases in the Rotterdam study and previous publications.

Results: Plasma levels of 65 miRNAs were associated with BMI, 40 miRNAs with WHR, 65 miRNAs with FMI and 15 miRNAs with AGR surpassing the Bonferroni-corrected $P < 8.46\times10^{-5}$. Of these, 12 miRNAs were significantly associated with all traits, while four miRNAs were associated only with WHR, three miRNAs only with FMI, and miR-378i was associated only with AGR . The most significant association among the overlapping miRNAs was with miR-193a-5p, which was shown to be associated with type 2 diabetes and hepatic steatosis in the Rotterdam Study. Moreover, five of the obesity-associated miRNAs and two of the body fat distribution miRNAs have been correlated previously to cardiovascular disease.

Conclusions: This study indicates that plasma levels of several miRNAs are associated with obesity and body fat distribution that may help us to better understand the underlying mechanisms and could have the biomarker potential for obesity-related metabolic diseases.

Introduction

Obesity is a growing health problem worldwide; its prevalence is increasing in both developed and developing countries.[1] According to WHO data, 39% of the global population of adults are overweight and of these, 13% are obese. Obesity can contribute to the development of metabolic disorders, such as cardiovascular disease (CVD), type 2 diabetes (T2D) and fatty liver disease.[2, 3] The pathogenesis of obesity involves regulation of calorie utilization, appetite, and physical activity, but has complex interactions with the availability of healthcare systems, and underlying genetic and environmental factors.[4-8] Obesity, particularly abdominal obesity due to the presence of large amounts of visceral adipose tissue (VAT), which is highly metabolically active,[9] predisposes a person to a number of other cardiovascular risk factors, and is an independent predictor of clinical CVD including coronary heart disease, heart failure and stroke.[10]

Many studies have focused on the role of body fat distribution and the metabolic complication of obesity, one of these studies has shown that lower and upper body subcutaneous fat, and visceral fat depots have unique characteristics with regards to fatty acid metabolism.[11] Selective dysregulation of these depots probably plays an important role in the metabolic complications of obesity. Anthropometric indicators have also been proposed to diagnose the health risks considering the increased body fat.[12] The most widely used is still body mass index (BMI), but it has limitations.[13] For example, BMI does not measure body fat or fluid retention,[14] so the anthropometric data are relatively poor indicators for older people and athletes. The metabolic consequences of adiposity are dictated not only by absolute adipose tissue (AT) mass, but also by its distribution,[15-17] to the extent that waist-to-hip ratio (WHR) is a stronger predictor of myocardial infarction than BMI in women.[18] Besides the anthropometric measurements, Dual X-ray absorptiometry (DXA) parameters could indicate fat mass and adiposity.

MicroRNAs (miRNAs) represent a class of small non-coding RNAs with the main role of regulating gene expression post-transcriptionally through degradation or repression of messenger RNAs.[19] In recent years, remarkable progress has been made toward identifying the origin and function of miRNAs, focusing on their potential use in both the research to understand disease mechanisms and the clinic to identify potential biomarkers.[20-22] Notably, miRNAs have been shown to regulate adipose tissue metabolism, insulin secretion and action. The aberrant expression of miRNAs play a role in the development of obesity,[23] though the role of miRNAs in regulating fat distribution remains poorly understood.[24] Although extensive research has explored the role of miRNAs in the pathogenesis of metabolic diseases, limited population-based studies have investigated the association of circulating miRNAs in plasma with obesity and body fat distribution. In the current study, we conducted a genome-wide screening to investigate the relationship between plasma levels of miRNAs with obesity, body fat distribution and fat mass in the population-based Rotterdam Study cohort. We further looked up for the correlation of obesity-related miRNAs with cardiovascular and metabolic diseases.

Methods

Study population

This study was embedded within the Rotterdam Study, a prospective cohort of individuals aged ≥ 45 years living in the Ommoord district of Rotterdam, the Netherlands. The design of the Rotterdam Study has extensively been described elsewhere.[25] In 1990, 7983 persons aged 55 years or older were recruited to participate in the first cohort of the RS-I. In 2000, the second cohort RS-II was extended by 3011 participants who moved to Ommoord or had become 55 years of age. In the current study, we used the expression profiles of circulating miRNAs in plasma, collected between 2002 and 2005, from a random subset (n = 1000) of the fourth visit of the first cohort (RS-I-4) and a random subset (n = 999) of the second visit of the second cohort (RS-II-2). Among them, 1331 individuals with data on anthropometric measurements (BMI and WHR) and Dual X-ray absorptiometry (DXA) measurements (fat mass index (FMI) and android-fat to the gynoid-fat ratio (AGR) were included in our analyses. Furthermore, 123 participants that were excluded because of missing data in covariates resulting in 1208 participants that were included for the statistical analysis **(Figure 1**).

The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC University Medical Center (registration number MEC 02.1015) and the Dutch Ministry of Health, Welfare and Sport (Population Screening Act WBO, license number 1071272-159521-PG).The participants included in the current study provided written informed consent.

miRNA expression profiling

Blood samples were collected in EDTA-treated containers and centrifuged. Plasma was then aliquoted and frozen at −80 °C according to standard procedures. Subsequently, Plasma levels of cell-free miRNAs were determined using the HTG EdgeSeq miRNA Whole Transcriptome Assay (WTA), which measures the expression of 2083 mature human miRNAs (HTG Molecular Diagnostics, Tuscon, AZ). Each sample was sequenced on an Illumina NextSeq 500 sequencer (Illumina, San Diego, CA). The whole transcriptome assay characterizes miRNA expression patterns and measures the expression of 13 housekeeping genes, allowing flexibility in data normalization and analysis.

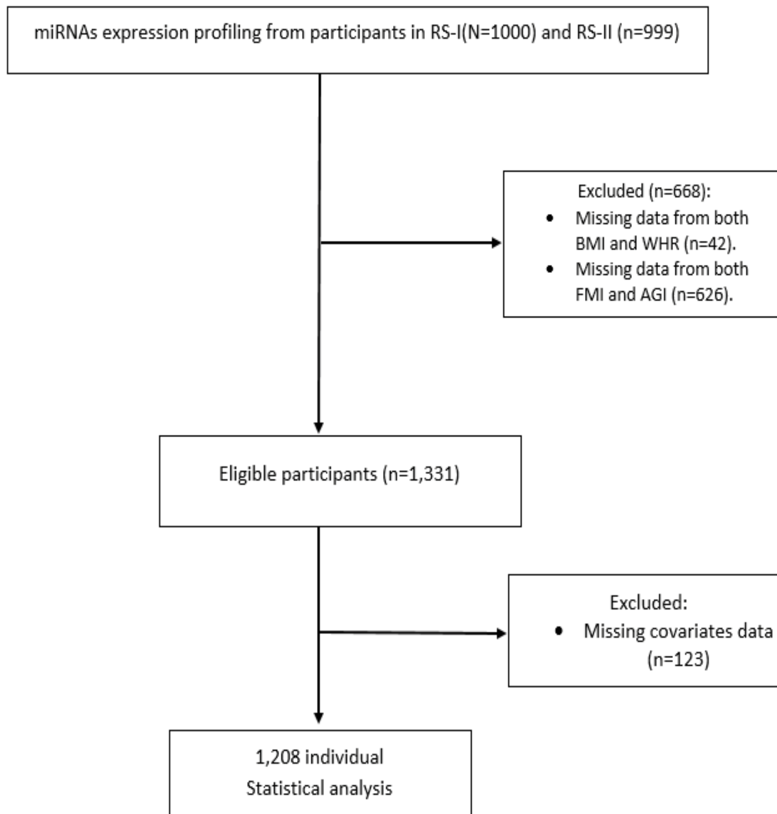


Figure 1. Flowchart of the study participants. Abbreviations: RS, Rotterdam study; BMI, Body mass index; WHR, Waist-hip ratio; FMI, Fat mass index; AGR, Android to gynoid-fat ratio; miRNAs, microRNAs.

Quantification of miRNA expression was based on counts per million (CPM). Log2 transformation of counts per million was used as standardization and adjusted for total reads within each sample. The lower limit of quantification was used to select well-expressed miRNAs. The lower limit of quantification level was based on a monotonic decreasing spline curve fit between the means and standard deviations of all miRNAs in the dataset with 1999 subjects. In our definition, well-expressed miRNA levels in plasma were those with > 50% values above the lower limit of quantification. This includes a set of 591 miRNAs that were used for association analysis.

Anthropometric and body composition measurements

Height and weight were measured while the participants were standing without shoes and heavy outerwear. BMI was calculated as weight divided by height squared (kg/m^2). Waist circumference was

measured at the level midway between the lower rib margin and the iliac crest with participants in standing position without heavy outerwear and with emptied pockets, breathing out gently. The waist-hip ratio (WHR) was calculated as waist measurement dividing waist by hip.

Total body fat was assessed by dual-energy x-ray absorptiometry (DXA) scan with Idxa total body-fan beam densitometer (GE Lunar Corp, Madison, WI, USA) following manufacturer protocols, with scans analyzed with enCORE software V13.6 using pre-determined regions of interest. The method of body components measurement by DXA is described in detail elsewhere.[26] We calculated FMI by dividing total fat mass by height squared in meters (kg/m^2). Additionally, we calculated the AGR by dividing android-fat by gynoid-fat.

Metabolic syndrome

To define MetS, we used the joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; the National Heart, Lung, and Blood Institute; the American Heart Association; the World Heart Federation; the International Atherosclerosis Society; and the International Association for the Study of Obesity.[27] According to this definition, an individual had MetS if three out of five of the following conditions are met: (1) waist circumference ≥94 cm for men and ≥80 cm for women; (2) triglycerides (TG) (≥150 mg/dl (1.7 mmol/l) or lipid-lowering medication use; (3) high-density lipoprotein cholesterol (HDL-C) (< 40 mg/dl (1.03 mmol/l)) in males and < 50 mg/dl (1.29 mmol/l) in females) or, or lipid-lowering medication use; (4) systolic blood pressure ≥130 mmHg and/or diastolic blood pressure ≥85 mmHg) or, blood-pressure lowering medication use; (5) fasting serum glucose ≥ 100 mg/dl (5.6 mmol/l) or previously diagnosed type 2 diabetes.

Serum glucose was measured by the hexokinase enzymatic method after overnight fasting. TG was measured in fasting serum using the enzymatic method. HDL-C was measured by automatic enzymatic method from fasting serum after precipitation of non-HDL fraction. Blood pressure was measured at the right brachial artery using a random-zero sphygmomanometer after 5 minutes of rest with the participants in a sitting position. The mean of the two consecutive measurements was used. MetS was obtained from the fifth visit of the first cohort (RS-I-5) and the third visit of the second cohort (RS-II-3).

Covariates

Covariates information on age, sex, smoking status (never, former, current) and physical activity (METhours per week) were obtained through questionnaires.[25] Measures of physical activity were obtained in the third visit of the original cohort (RS-I-3) and the first visit of the extended cohort (RS-II-1). Dietary intake was assessed in the first visit of the original cohort (RS-I-1) and the first visit of the extended cohort (RS-II-1), Dietary intake was assessed using validated Food-Frequency Questionnaires (FFQ), from which we calculated diet quality scores and energy intake, as described elsewhere.[28] Measures of total dietary scores and energy intake were gathered from the FFQ baseline. At baseline and during follow-up, T2D cases were ascertained using general practitioners' records, hospital discharge letters, and serum glucose measurements collected from center visits. T2D was defined according to the World Health Organization definition as fasting glucose levels of ≥7.0 mmol/L, non-fasting glucose levels ≥11.1 mmol/L, or the use of glucose-lowering medications. Glucose measurements were obtained during visits to the research center.

Statistical analysis

Population characteristics were represented as mean ± SD for continuous variables, and categorical variables were expressed as numbers and percentages. Participants with missing values on multiple covariates were excluded from the analysis. Analyses were performed in the total population and stratified by sex. We used multivariable linear regression models to investigate the associations between 591 well-expressed miRNA (exposure) and obesity related parameters (outcome). The Bonferroni-corrected p-value threshold was calculated based on the number of tested miRNAs ($0.05/591 = 8.46 \times 10^{-5}$). The basic model (M1) was adjusted for age, sex, and cohort. In the second model (M2), we adjusted for smoking status and physical activity, which is considered the main model. Sensitivity analysis was performed by adjusting for more variables.

Furthermore, the association between miRNAs and FMI, AGR and WHR was further evaluated by using linear regression models adjusting for BMI to determine associated miRNAs independent from BMI. As sensitivity analysis, we also adjusted the second model for more covariates including energy intake and diet quality score. Next, multivariable logistic regression models were used to investigate longitudinally the association of the plasma levels of the miRNA with prevalence of MetS, in model 1 and model 2.

All analyses were performed using SPSS statistical software (SPSS, version25; IBM Crop) and R software version 3.5.2 (The R Foundation for Statistical Computing).

Results

The descriptive of the participants included in the cross-sectional study are presented in **Table 1** and **Table S1**. The mean value (SD) of age was 72.17±6.9 years, and 56.6% were females. Compared to women, men had a higher mean value of WHR and AGR, while the mean value for FMI and physical activity were higher among women. In total 649 participants had a prevalence of metabolic syndrome (MetS) representing 53.4% of the population in our analysis. Hypertension was defined in 944 (78.1%) participants. The mean standard deviation (SD) value of the total serum cholesterol of the study population was 5.6±1.0. T2D was detected in 11.4% of participants according to the World Health Organization definition [29].

Table 1. Characteristics of the Rotterdam Study participants

Characteristics	Total population	Men	Women
Number	1208	524	684
Age	72.2±6.9	72.2±6.68	72.14±7.1
Gender, F	684 (56.6)	524 (43.4)	684 (56.6)
Obesity traits			
Body mass index, (kg/m²)	27.5±4.1	27.4 ±3.28	27.6 ±4.61
Waist to hip ratio	0.90 ±0.09	0.97±0.07	0.85±0.07
Fat mass index, (kg/m²)	0.96 ±0.3	0.79±0.24	1.09±0.34
Android to gynoid-fat ratio	0.64±0.2	0.80±0.16	0.52±0.13
Incidence diabetes mellitus, n (%)	138 (11.4)	55 (10.4)	83 (12.1)
Smoking status, n (%)			
Never smoking	178 (14.7%)	85 (16.3)	93 (13.5)
Ex. Smoking	687 (56.9)	376 (71.7)	311(45.5)
Current smoking	343 (28.4)	63 (12.0)	280 (41.0)
Physical activity (METh/wk)	87.8±43.26	74.83±41.16	97.81±42.02
Rotterdam study cohort, n (%)			
Rs_I-4	808 (66.9)	355 (67.7)	453 (66.3)
Rs_II-2	400 (33.1)	169 (32.3)	231 (33.7)

The table shows characteristics of study participants (1208); Variables are represented as mean (± standard deviation), or number (%). Missing values were excluded from the baseline.

Using linear regression analysis and in the multivariable model 2, we found 65 miRNAs for BMI, 40 for WHR, 65 for FMI, and 15 for AGR to be significantly associated at the Bonferroni-corrected $P< 8.46\times10^{-5}$ **(Table S2)**. The overlapping miRNAs among these traits are shown in **Figure 2**, including 12 miRNAs associated with all four traits that are shown in **Table 2**. The miRNAs exclusively associated with each of the traits are highlighted in **Table S2**. For example, miR-378i ($\beta= 0.07$, $P= 4.43\times10^{-6}$) was associated only with AGR, while miR-1304-3p ($\beta= 0.11$, $P= 3.0\times10^{-6}$), miR-566 ($\beta= 0.07$, $P= 1.73\times10^{-5}$), and miR-324-

3p (β= -0.10, P= 7.37×10^{-7}) were associated only with FMI. In the sensitivity analysis adjusting for energy intake and diet quality score, the results were less significant, but 5 out of the 12 common miRNAs remained statistically significant (**Table S3**). In addition, we observed variation in the plasma levels of 12 miRNAs within three BMI categories (normal weight, overweight and obesity) as shown in **Figure 3**.

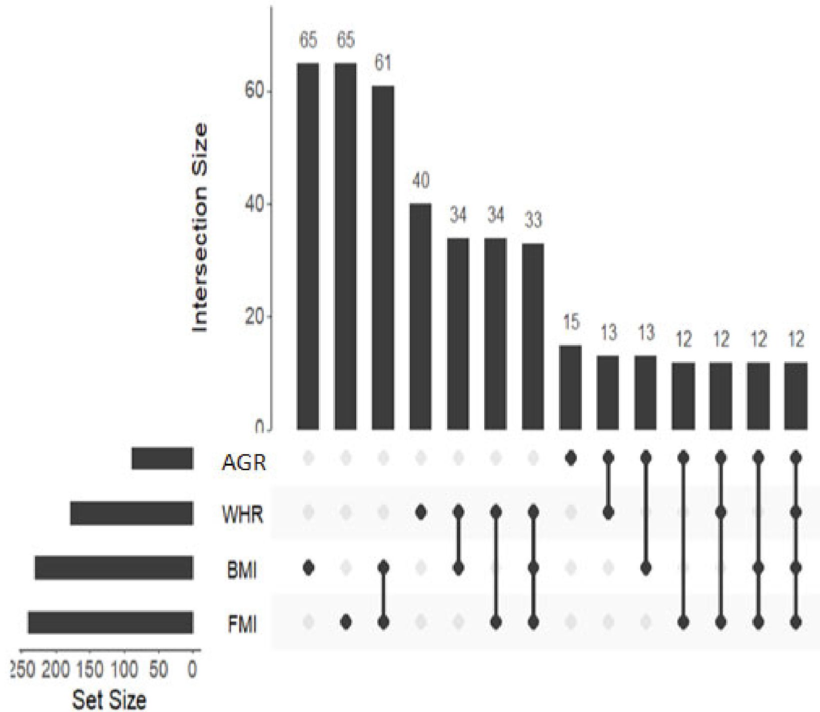


Figure 2. UpSet plots showing the intersections of miRNAs associated with the four traits**.** The bars chart visualized the numbers of miRNAs associated with obesity-related traits in model 2 adjusted for age, sex, rs_cohort, smoking status and physical activity. Dark circles in the matrix indicate sets that are part of the intersection. Abbreviations: AGR, Android to gynoid-fat ratio; WHR, Waist-hip ratio; BMI, Body mass index; FMI, Fat mass index.

Table 2. Circulatory miRNAs significantly associated with all four obesity related-traits

	BMI		WHR		AGR		FMI	
miRNA ID	**Beta**	***P*-value**	**Beta**	***P*-value**	**Beta**	***P*-value**	**Beta**	***P*-value**
miR-145-5p	-1.41	2.28E-11	-0.02	2.74E-06	-0.03	3.36E-05	-0.11	2.07E-12
miR-149-3p	1.38	1.23E-07	0.03	1.36E-08	0.04	2.43E-05	0.11	4.50E-08
miR-193a-5p	1.85	2.94E-13	0.03	2.72E-12	0.06	2.18E-11	0.13	1.83E-11
miR-345-5p	-2.87	2.23E-11	-0.04	1.82E-08	-0.07	1.32E-05	-0.24	5.44E-14
miR-3937	-1.01	2.18E-13	-0.02	2.97E-10	-0.02	8.40E-05	-0.08	1.48E-13
miR-4433b-5p	-1.65	3.06E-07	-0.02	1.88E-05	-0.05	1.57E-05	-0.11	4.92E-06
miR-4478	1.38	5.63E-07	0.03	1.10E-07	0.04	2.13E-05	0.10	8.91E-07
miR-6088	1.58	2.66E-11	0.03	9.70E-10	0.03	5.97E-05	0.12	2.76E-11
miR-6799-5p	1.29	6.27E-07	0.02	2.85E-06	0.04	7.23E-05	0.09	2.67E-06
miR-6803-5p	1.80	4.29E-13	0.03	1.06E-12	0.04	1.44E-05	0.13	4.91E-13
miR-6821-5p	1.42	3.63E-11	0.03	1.00E-12	0.04	1.55E-07	0.11	4.29E-12
miR-7107-5p	1.42	6.31E-12	0.02	2.74E-09	0.03	3.50E-05	0.11	2.21E-12

The table shows 12 miRNAs that were overlapped and significantly associated with all four studied traits in model 2. Model 2 is adjusted for age, sex, cohort, physical activity and smoking status. The *P*-value threshold is 8.46×10^{-5} (after Bonferroni correction 0.05/591 miRNAs). Abbreviations: miRNA, microRNA; BMI, body mass index; WHR, waist to hip ratio, AGR, android to gynoid-fat ratio; FMI, fat mass index.

The association of plasma miRNAs with AGR, FMI, and WHR adjusted for BMI and other covariates in multilinear regression model 2, is shown in **Table S4**. This analysis showed two miRNAs (miR-193a-5p and miR-378i) to be significantly associated with AGR and FMI. **Figure 4** shows the association of miRNAs with AGR and FMI adjusted for BMI. Three miRNAs (miR-193a-5p, miR-6821-5p and miR-6803-5p) were associated with WHR adjusted for BMI.

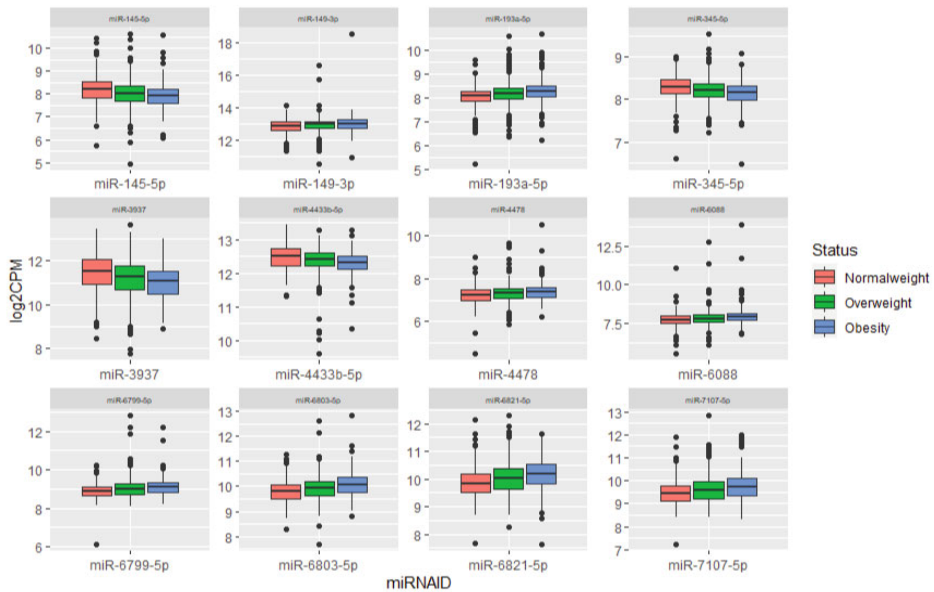


Figure 3. Comparison between expression levels of the identified miRNAs in 3 body mass index subgroups. The figure shows the comparison between expression of 12 miRNAs and BMI subgroups including normal weight (18.5-24.9), overweight (25-29.9) and obesity (30.0 or above) were significantly associated with obesity-related traits. Abbreviations: CPM, counts per million; miRNAs, microRNAs.

Sensitivity analyses were also conducted to assess the association of plasma levels of miRNAs with the studied traits stratified by sex, using multilinear regression models. In model 2 in females, the vast majority of the 12 common miRNAs were strongly associated and overlapped with all traits (**Table S5A**), while in males, the 12 miRNAs were less significant (**Table S5B**). Moreover, we found 5 new miRNAs (miR-185-5p, miR-19a-3p, miR-19b-3p, miR-7150 and miR-93-5p) to be associated with the four traits only in females surpassing the Bonferroni-corrected p-value threshold **(Table S5C)**.

In the longitudinal analysis, we found 24 miRNAs to be significantly associated with the prevalence of metabolic syndrome using logistic regression in the multivariable model 2. Of these, 7 miRNAs (miR-193a-5p, miR-149-3p, miR-3937, miR-6088, miR-6821-5p, miR-6803-5p and miR-4433b-5p) were overlapped with the 12 common miRNAs associated with all four traits (**Table S6)**. We further searched in the previous literature for the association with cardiovascular and metabolic diseases of the 12 common miRNAs. We found evidence for 7 of the 12 miRNAs to be correlated with hepatic steatosis, T2D, hypertension, heart failure, stroke and myocardial infarction in previous studies (**Table S7**).

Discussion

In this population-based study, we investigated the association of circulating miRNAs in plasma with obesity, body fat distribution and fat mass. We found several miRNAs associated with these traits. These include 12 miRNAs that were overlapped among all traits, consisting of 8 miRNAs (miR-193a-5p, miR-149-3p, miR-4478, miR-6088, miR-6803-5p, miR-6799-5p, miR-6821-5p and miR-7107-5p) that were positively and 4 miRNAs (miR-145-5p, miR-345-5p, miR-3937 and miR-4433b-5p) that were negatively associated with the studied traits. Some of the identified miRNAs were associated with cardio-metabolic diseases that help us to better understand the underlying mechanisms and may serve as potential biomarkers for obesity-related diseases.

Body fat distribution is an important predictor of metabolic abnormalities in obese humans, dysregulation of free fatty acid (FFA) release, especially from upper body subcutaneous adipose tissue, appears to contribute substantially to these metabolic disturbances.[30] In this study, we used a new genome-wide RNA-seq based assay to investigate the association of plasma circulating miRNAs with both obesity and body fat distribution. In general, RNA-seq analysis of cell-free miRNAs can give a higher sensitivity than the other miRNA profiling approaches (e.g., qPCR). This measures the expression levels of miRNAs over a wide dynamic range with the ability to identify novel miRNAs.[31] Furthermore, the stability of cell-free miRNAs is high in body fluids. The accessibility of blood samples in cohort studies, which is easier to collect than biopsy of diseased tissues, indicates the potential of circulating miRNAs as biomarkers in clinical traits.[32] In this line, aberrant expression of miRNAs in different tissues (e.g., pancreas, liver and adipose tissue) have been linked with obesity [33] and metabolic diseases in previous studies.[34, 35] Several studies have also demonstrated the association of miRNAs levels with obesity and fat distribution by using genetic approach.[24, 36, 37] For example, the role of miR-196a in regulating human body fat distribution has been found by genetic association study of miRNA-related SNPs and using tissue-specific miRNAs from 15 men (8 lean (BMI < 25) and 7 overweight (BMI > 25)). However, previous miRNA profiling studies have mainly used qPCR-based methods. They are also limited to a smaller number of miRNAs and sample sizes, while our study is conducted in a population-based setting with a much larger sample size and investigates hundreds of circulatory miRNAs.

The most significant miRNA associated with all four traits in our analyses was miR-193a-5p. Previous study on human has reported that surgery-induced weight loss led to a noticeable decrease of circulating miR-193a-5p.[38] Another study on animal model has shown that miR-193a-5p could be used as a biomarker for adiposity, and the authors have shown by functional analysis that this miRNA could regulate 33 pathways with two novel pathways included Glucagon signaling pathway and glycolysis/gluconeogenesis.[39] In the present study, we observed a significant increase in circulating levels of miR-193a-5p in all four studies traits. Furthermore, the circulating miR-193a-5p was positively associated with metabolic syndrome in our cohort. This miRNA has been reported by our group recently

to be implicated in fatty liver disease.[40] At the same time, in the most recent publication with the Rotterdam Study data, we found the link between miR-193a-5p and the incident of T2D.[41] These findings may indicate a strong role of miR-193a-5p in adiposity and metabolic pathways that warrant further experimental validation studies.

Our results moreover showed that circulating miR-145-5p is negatively associated with obesity traits. In this line, a previous study has reported that plasma levels of miR-145-5p were significantly altered following long-term fasting; the expression levels of miR-145-5p have changed in the opposite direction and negatively associated with obesity.[42] Lin et al. also found that miR-145 directly targets and represses FoxO1 and abhydrolase domain containing 5 (ABHD5 or Cgi58) activators of lipolytic activity, and forced expression of miR-145 attenuates lipolysis in white adipose tissue.[22, 43] Another study has shown that miR-145 expression is downregulated in human adipose tissue in insulin-resistant versus insulin-sensitive patients.[44] Viesti et al., have further found that miR-145 had a reduced tendency in the subcutaneous adipose tissue in patients with obesity.[45] All these findings support the important regulatory role of miR-145 in the underlying mechanisms of obesity.

In line with our observations, Pascut et al. have found that increased circulating levels of miR-7107-5p was prevalently expressed in a population with obesity and steatosis.[46] In addition, Liu, X et al, have shown that the expression of miR-345-5p decreased during adipogenic differentiation via targeting vascular endothelial growth factor B. They have demonstrated that overexpression of miR-345-5p reduced lipid accumulation in adipocytes, and the expression of adipocyte related genes is essential to lipogenic transcription, fatty acid synthesis and fatty acid transport.[47] In addition, a previous study on the mouse model has revealed an important mechanism of miR-149-3p and Prdm16-dependent regulation of energy expenditure upon chronic high-fat feeding, indicating that SAT miR-149-3p can serve as a therapeutic target to defend against diet-induced obesity and metabolic dysfunctions.[48] Here we also found that plasma levels of miR-149-3p are inversely associated with obesity and fat distribution.

On the other hand, we found member of miR-378 family, in particular miR-378i, to be associated only with fat distribution. A previous study on human has demonstrated particularly miR-378 as potential biomarker for predicting the risk of complications, especially insulin resistance in obesity, and reported that miR-378 expression is influenced by free fatty acids, adiponectin, and dexamethasone, the interaction of which may be involved in the pathogenesis of obesity-induced insensitivity to insulin.[49] Carrer et al. observed that mice genetically lacking miR-378 are resistant to obesity induced by a high fat diet. Therefore, a lowering in fat mass deposits was observed as well as a diminution of adipocyte size, raising the possibility that this miRNA might be necessary for efficient hypertrophy and lipid absorption in adipocytes.[50] MiRNA-378 is known to be highly stimulated during adipogenesis. Moreover, Gerin et al. found that miRNA378/378* overexpression during adipogenesis increases triacylglycerol accumulation in adipocytes due to the increase of de novo lipogenesis.[51]

Furthermore, we found that miR-342-5p to be inversely associated with total body fat, so we are in line with two previous studies that demonstrated miR-342-5p to be negatively associated with obesity.[52] Finally, the study by Yang, Zheng, et al. has proved evidence showing that derived expression miRNAs including miR-566 is associated with visceral adipose tissue in patients with obesity via regulation of fibroblast growth factor 2 (FGF2), FOS like 2 (FOSL2) and its subunit, AP-1 transcription factor, and adenosine monophosphate deaminase AMPD3, respectively.[53]

The other seven identified miRNAs for obesity traits, including miR-3937, miR-4433b-5p, miR-4478, miR-6088, miR-6799-5p, miR-6803-5p, and miR-6821-5p, and miR-1304-3p associated only with fat mass are new, which means we are the first to report their associations with obesity and fat mass. Yet, the majority of these miRNAs are shown to be associated with cancer. As many studies revealed that approximately 20% of many common cancers are caused by excess body fat accumulation,[54-56] there might be a role for these miRNAs in developing cancer. Future experimental studies are needed to investigate their potential roles in common molecular mechanisms underlying obesity and cancer.

In our search in previous studies on the association between the obesity-related miRNAs and cardiovascular and metabolic diseases, we found evidence for five miRNAs. Particularly, the top obesity-associated miRNA (miR-193a-5p), which has been shown by our group and other to be associated with hepatic steatosis and T2D [40, 41]. In addition, the high expression of miR-145-5p has been a risk factor for essential hypertension.[57] Another study has found the role of miR-345-5p (as part of lncRNA-mediated ceRNA networks) in the pathophysiological process of heart failure and its potential regulatory functions on programmed cell death.[58] Serum miR-6803-5p has been predicted the risk of cerebrovascular disorders before the onset of, for instance, stroke.[59] Finally, miR-4478 and soluble leptin receptor has been suggested to be used as predictors of non-ST-segment elevation myocardial infarction [NSTEMI].[60] In addition, miR-378 was associated with coronary heart disease,[61] while miR-566 was associated with acute myocardial infarction.[62]

The key strengths of this study include the large sample size compared to previous studies, the availability of various clinical data and DXA measures from a population-based prospective cohort study, and the use of a new RNA-seq based assay covering hundreds of known human miRNAs. However, the results reported herein should be considered in the light of some limitations. First, although our findings are from two sub-cohorts of the Rotterdam Study, additional replication in an independent cohort with more stratification analyses could confirm our findings. Second, during the baseline examination of this study, traditional parameters such as physical activity, energy intake and total dietary score were not measured. Therefore, these parameters were used as proxies from previous visits and could not calculate the differences in disease predicted capacity between the identified miRNAs and parameters. Thirdly, as miRNAs are tissue-specific, it would be interesting to additionally study the expression and gene targeting

of the identified miRNAs in adipose tissue or relevant cell lines to identify their regulatory roles in obesity pathways.

In conclusion, this study demonstrates plasma levels of several miRNAs to be associated with obesity and body fat distribution in a population-based setting. As circulatory miRNAs have opened a promising research avenue for the detection of non-invasive biomarkers for the identification of subjects at risk of complex diseases, the identified miRNAs may give insights into the underlying molecular pathways and have the biomarker potential for obesity-related diseases. Future epidemiologic studies with larger sample sizes and long follow-up time and functional studies are needed to confirm the role of identified miRNAs in the molecular pathways of obesity.

Supplementary material

Supplementary material is a viable on

References

1. Csige I, Ujvárosy D, Szabó Z, Lőrincz I, Paragh G, Harangi M, et al. The impact of obesity on the cardiovascular system. Journal of diabetes research. 2018;2018.
2. Ghaben AL, Scherer PE. Adipogenesis and metabolic health. Nature reviews Molecular cell biology. 2019;20(4):242-58.
3. Bovet P, Chiolero A, Gedeon J. Health effects of overweight and obesity in 195 Countries. The New England journal of medicine. 2017;377(15):1495-6.
4. Lin X, Li H. Obesity: Epidemiology, Pathophysiology, and Therapeutics. Frontiers in Endocrinology. 2021:1070.
5. Lovejoy JC. The menopause and obesity. Primary Care: Clinics in Office Practice. 2003;30(2):317-25.
6. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. New England journal of medicine. 2007;357(4):370-9.
7. Wardle J, Haase AM, Steptoe A, Nillapun M, Jonwutiwes K, Bellisie F. Gender differences in food choice: the contribution of health beliefs and dieting. Annals of behavioral medicine. 2004;27(2):107-16.
8. Yang W, Kelly T. J He. Genetic Epidemiology of Obesity Epidemiol Rev. 2007;29:49-61.
9. Després J-P, Lemieux I. Abdominal obesity and metabolic syndrome. Nature. 2006;444(7121):881-7.
10. Ortega FB, Lavie CJ, Blair SN. Obesity and cardiovascular disease. Circulation research. 2016;118(11):1752-70.
11. Jensen MD. Role of body fat distribution and the metabolic complications of obesity. The Journal of Clinical Endocrinology & Metabolism. 2008;93(11_supplement_1):s57-s63.
12. Sigulem DM, Devincenzi MU, Lessa AC. Diagnosis of child and adolescent nutritional status. J Pediatr (Rio J). 2000;76(Suppl 3):S275-84.
13. Neovius M, Linné Y, Barkeling B, Rössner S. Discrepancies between classification systems of childhood obesity. Obesity Reviews. 2004;5(2):105-14.
14. Oreopoulos A, Kalantar-Zadeh K, Sharma AM, Fonarow GC. The obesity paradox in the elderly: potential mechanisms and clinical implications. Clinics in geriatric medicine. 2009;25(4):643-59.
15. Liu DJ, Peloso GM, Yu H, Butterworth AS, Wang X, Mahajan A, et al. Exome-wide association study of plasma lipids in> 300,000 individuals. Nature genetics. 2017;49(12):1758-66.
16. Emdin CA, Khera AV, Natarajan P, Klarin D, Zekavat SM, Hsiao AJ, et al. Genetic association of waist-to-hip ratio with cardiometabolic traits, type 2 diabetes, and coronary heart disease. Jama. 2017;317(6):626-34.
17. Lotta LA, Gulati P, Day FR, Payne F, Ongen H, Van De Bunt M, et al. Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. Nature genetics. 2017;49(1):17-26.
18. Cao Q, Yu S, Xiong W, Li Y, Li H, Li J, et al. Waist-hip ratio as a predictor of myocardial infarction risk: A systematic review and meta-analysis. Medicine. 2018;97(30).
19. Otsuka M, Kishikawa T, Yoshikawa T, Yamagami M, Ohno M, Takata A, et al. MicroRNAs and liver disease. Journal of human genetics. 2017;62(1):75-80.
20. Schueller F, Roy S, Vucur M, Trautwein C, Luedde T, Roderburg C. The role of miRNAs in the pathophysiology of liver diseases and toxicity. International journal of molecular sciences. 2018;19(1):261.
21. Pirola CJ, Gianotti TF, Castaño GO, Mallardi P, San Martino J, Ledesma MMGL, et al. Circulating microRNA signature in non-alcoholic fatty liver disease: from serum non-coding RNAs to liver histology and disease pathogenesis. Gut. 2015;64(5):800-12.
22. Condrat CE, Thompson DC, Barbu MG, Bugnar OL, Boboc A, Cretoiu D, et al. miRNAs as biomarkers in disease: latest findings regarding their role in diagnosis and prognosis. Cells. 2020;9(2):276.
23. Landrier J-F, Derghal A, Mounien L. MicroRNAs in obesity and related metabolic disorders. Cells. 2019;8(8):859.
24. Hilton C, Neville MJ, Wittemans LBL, Todorcevic M, Pinnick KE, Pulit SL, et al. MicroRNA-196a links human body fat distribution to adipose tissue extracellular matrix composition. EBioMedicine. 2019;44:467-75.
25. Ikram MA, Brusselle G, Ghanbari M, Goedegebure A, Ikram MK, Kavousi M, et al. Objectives, design and main findings until 2020 from the Rotterdam Study. European journal of epidemiology. 2020;35(5):483-517.

26. Alferink LJM, Trajanoska K, Erler NS, Schoufour JD, de Knegt RJ, Ikram MA, et al. Nonalcoholic fatty liver disease in the Rotterdam study: about muscle mass, sarcopenia, fat mass, and fat distribution. Journal of bone and mineral research. 2019;34(7):1254-63.
27. Alberti KGMM, Eckel RH, Grundy SM, Zimmet PZ, Cleeman JI, Donato KA, et al. Harmonizing the metabolic syndrome: a joint interim statement of the international diabetes federation task force on epidemiology and prevention; national heart, lung, and blood institute; American heart association; world heart federation; international atherosclerosis society; and international association for the study of obesity. Circulation. 2009;120(16):1640-5.
28. Voortman T, Kiefte-de Jong JC, Ikram MA, Stricker BH, van Rooij FJA, Lahousse L, et al. Adherence to the 2015 Dutch dietary guidelines and risk of non-communicable diseases and mortality in the Rotterdam Study. European journal of epidemiology. 2017;32(11):993-1005.
29. World Health O. Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: report of a WHO/IDF consultation. 2006.
30. Santosa S, Jensen MD. Why are we shaped differently, and why does it matter? American Journal of Physiology-Endocrinology and Metabolism. 2008;295(3):E531-E5.
31. Eminaga S, Christodoulou DC, Vigneault F, Church GM, Seidman JG. Quantification of microRNA expression with next-generation sequencing. Current protocols in molecular biology. 2013;103(1):4.17. 1-4.. 4.
32. Szelenberger R, Kacprzak M, Saluk-Bijak J, Zielinska M, Bijak M. Plasma MicroRNA as a novel diagnostic. Clinica Chimica Acta. 2019;499:98-107.
33. Kunej T, Skok DJ, Zorc M, Ogrinc A, Michal JJ, Kovac M, et al. Obesity gene atlas in mammals. Journal of Genomics. 2013;1:45.
34. Jiang Q, Wang Y, Hao Y, Juan L, Teng M, Zhang X, et al. miR2Disease: a manually curated database for microRNA deregulation in human disease. Nucleic acids research. 2009;37(suppl_1):D98-D104.
35. Dumortier O, Hinault C, Van Obberghen E. MicroRNAs and metabolism crosstalk in energy homeostasis. Cell metabolism. 2013;18(3):312-24.
36. Ghanbari M, Sedaghat S, De Looper HWJ, Hofman A, Erkeland SJ, Franco OH, et al. The association of common polymorphisms in mi R-196a2 with waist to hip ratio and mi R-1908 with serum lipid and glucose. Obesity. 2015;23(2):495-503.
37. Cheng M, Mei B, Zhou Q, Zhang M, Huang H, Han L, et al. Computational analyses of obesity associated loci generated by genome-wide association studies. PloS one. 2018;13(7):e0199987.
38. Ortega FJ, Mercader JM, Catalan V, Moreno-Navarrete JM, Pueyo N, Sabater M, et al. Targeting the circulating microRNA signature of obesity. Clinical chemistry. 2013;59(5):781-92.
39. Youssef EM, Elfiky AM, Abu-Shahba N, Elhefnawi MM. Expression profiling and analysis of some miRNAs in subcutaneous white adipose tissue during development of obesity. Genes & nutrition. 2020;15(1):1-14.
40. Zhang X, Mens MMJ, Abozaid YJ, Bos D, Darwish Murad S, de Knegt RJ, et al. Circulatory microRNAs as potential biomarkers for fatty liver disease: the Rotterdam study. Alimentary Pharmacology & Therapeutics. 2021;53(3):432-42.
41. Mens MMJ, Mustafa R, Ahmadizar F, Ikram MA, Evangelou M, Kavousi M, et al. MiR-139-5p is a causal biomarker for type 2 diabetes; Results from genome-wide microRNA profiling and Mendelian randomization analysis in a population-based study. medRxiv. 2021.
42. Ravanidis S, Grundler F, de Toledo FW, Dimitriou E, Tekos F, Skaperda Z, et al. Fasting-mediated metabolic and toxicity reprogramming impacts circulating microRNA levels in humans. Food and Chemical Toxicology. 2021;152:112187.
43. Lin Y-Y, Chou C-F, Giovarelli M, Briata P, Gherzi R, Chen C-Y. KSRP and MicroRNA 145 are negative regulators of lipolysis in white adipose tissue. Molecular and cellular biology. 2014;34(12):2339-49.
44. Kirby TJ, Walton RG, Finlin B, Zhu B, Unal R, Rasouli N, et al. Integrative mRNA-microRNA analyses reveal novel interactions related to insulin sensitivity in human adipose tissue. Physiological genomics. 2016;48(2):145-53.
45. Viesti A. Collares R, Salgado Jr W, Pretti da Cunha Tirapelli D, dos Santos JS. The expression of LEP, LEPR, IGF1 and IL10 in obesity and the relationship with microRNAs. PLoS One. 2014;9(4):e93512.

46. Pascut D, Tamini S, Bresolin S, Giraudi P, Basso G, Minocci A, et al. Differences in circulating microRNA signature in Prader–Willi syndrome and non-syndromic obesity. Endocrine connections. 2018;7(12):1262-74.
47. Liu X, He Y, Feng Z, Sheng J, Dong A, Zhang M, et al. miR-345-5p regulates adipogenesis via targeting VEGF-B. Aging (Albany NY). 2020;12(17):17114.
48. Zheng S, Guo S, Sun G, Shi Y, Wei Z, Tang Y, et al. Gain of Metabolic Benefit with Ablation of miR-149-3p from Subcutaneous Adipose Tissue in Diet-Induced Obese Mice. Molecular Therapy-Nucleic Acids. 2019;18:194-203.
49. Xu L, Jiang L, Gu K, Liu Z, Xu X. Regulation of MicroRNA-378 expression in mature human adipose tissue cells by adiponectin, free fatty acids and dexamethasone. Tropical Journal of Pharmaceutical Research. 2018;17(1):29-34.
50. Carrer M, Liu N, Grueter CE, Williams AH, Frisard MI, Hulver MW, et al. Control of mitochondrial metabolism and systemic energy homeostasis by microRNAs 378 and 378. Proceedings of the National Academy of Sciences. 2012;109(38):15330-5.
51. Gerin I, Bommer GT, McCoin CS, Sousa KM, Krishnan V, MacDougald OA. Roles for miRNA-378/378* in adipocyte gene expression and lipogenesis. American Journal of Physiology-Endocrinology and Metabolism. 2010;299(2):E198-E206.
52. Choi WH, Ahn J, Um MY, Jung CH, Jung SE, Ha TY. Circulating microRNA expression profiling in young obese Korean women. Nutrition Research and Practice. 2020;14(4):412-22.
53. Yang Z, Wei Z, Wu X, Yang H. Screening of exosomal miRNAs derived from subcutaneous and visceral adipose tissues: determination of targets for the treatment of obesity and associated metabolic disorders. Molecular medicine reports. 2018;18(3):3314-24.
54. Lancet T. The link between cancer and obesity. 2017. p. 1716.
55. De Pergola G, Silvestris F. Obesity as a major risk factor for cancer. Journal of obesity. 2013;2013.
56. Hopkins BD, Goncalves MD, Cantley LC. Obesity and cancer mechanisms: cancer metabolism. Journal of clinical oncology. 2016;34(35):4277.
57. Tao Z, Zheng S, He X, Sun J, He C, Zhang L. Hsa_circ_0037897 may be a risk factor for essential hypertension via hsa-miR-145-5p. Clinical and Experimental Hypertension. 2021;43(3):281-6.
58. Zhou Q, Li D, Zheng H, He Z, Qian F, Wu X, et al. A novel lncRNA-miRNA-mRNA competing endogenous RNA regulatory network in lung adenocarcinoma and kidney renal papillary cell carcinoma. Thoracic cancer. 2021;12(19):2526-36.
59. Sonoda T, Matsuzaki J, Yamamoto Y, Sakurai T, Aoki Y, Takizawa S, et al. Serum microRNA-based risk prediction for stroke. Stroke. 2019;50(6):1510-8.
60. Gholikhani-Darbroud R, Khaki-Khatibi F, Mansouri F, Hajahmadipoorrafsanjani M, Ghojazadeh M. Decreased circulatory microRNA-4478 as a specific biomarker for diagnosing non-ST-segment elevation myocardial infarction (NSTEMI) and its association with soluble leptin receptor. Bratislavske lekarske listy. 2017;118(11):684-90.
61. Zhang H, Hao J, Sun X, Zhang Y, Wei Q. Circulating pro-angiogenic micro-ribonucleic acid in patients with coronary heart disease. Interactive cardiovascular and thoracic surgery. 2018;27(3):336-42.
62. Chen L, Bai J, Liu J, Lu H, Zheng K. A four-microRNA panel in peripheral blood identified as an early biomarker to diagnose acute myocardial infarction. Frontiers in Physiology. 2021;12.

Chapter 2.2

Circulating microRNAs as potential biomarkers for fatty liver disease

Xiaofang Zhang, Michelle M.J. Mens*, **Yasir J. Abozaid***, Daniel Bos, Sarwa Darwish Murad, Robert. J. de Knegt, M. Arfan Ikram, Qiuwei Pan, Mohsen Ghanbari

Aliment Pharmacol Ther. 2021 Feb;53(3):432-442.

***Authors denotes equal contribution**

Abstract

Background: Fatty liver disease (FLD) is the most common cause of liver dysfunction in developed countries. There is a great interest in developing clinically valid and minimally-invasive biomarkers to enhance early diagnosis of FLD.

Aim: To investigate the potential of circulatory microRNAs (miRNAs) as biomarkers of FLD at the population level.

Methods: Plasma levels of 2083 miRNAs were measured by RNA-sequencing in 1,999 participants from the prospective population-based Rotterdam Study cohort. The Hounsfield Unit (HU) attenuation of liver was measured using non-enhanced computed tomography (CT-scan). Logistic and linear regression models adjusting for potential confounders were used to examine the association of circulatory miRNAs with liver enzymes (n=1,991) and CT-based FLD (n=954). Moreover, the association of miRNAs with hepatic steatosis and liver fibrosis were assessed longitudinally in individuals who underwent abdominal ultrasound (n=1,211) and transient elastography (n=777) after a median follow-up of >6 years.

Results: Cross-sectional analysis showed 61 miRNAs significantly associated with serum Gamma-glutamyl transferase and/or Alkaline phosphatase levels (Bonferroni-corrected p-value$<8.46\times10^{-5}$). Moreover, 17 miRNAs were significantly associated with CT-based FLD (p-value$<8.46\times10^{-5}$), 14 of them were associated with liver enzymes. Longitudinal analysis showed that four of the 14 miRNAs (miR-193a-5p, miR-122-5p, miR-378d, and miR-187-3p) were significantly associated with hepatic steatosis (p-value$<3.57\times10^{-3}$) and three (miR-193a-5p, miR-122-5p and miR-193b-3p) were nominally associated with liver fibrosis (p-value<0.05). Nine of the 14 identified miRNAs were involved in pathways underlying liver diseases.

Conclusions: This study indicates that plasma levels of several miRNAs can be used as biomarkers of FLD, laying the groundwork for future clinical applications.

Introduction

Fatty liver disease (FLD), is the most common cause of liver dysfunction in developed countries, that is also increasing in developing countries,[1] is defined as an excess accumulation of fat in hepatocytes.[2] Specifically, non-alcoholic fatty liver disease (NAFLD) is characterized by fat accumulation in hepatocytes not due to excess alcohol consumption.[3] The disorder covers a broad spectrum of underlying conditions, ranging from simple fatty liver to inflammation, which can progress to fibrosis, cirrhosis and even liver cancer.[4] FLD is strongly associated with obesity, hypertension, dyslipidemia and insulin resistance, regarded as hepatic manifestation of the metabolic syndrome.[5] Currently, liver biopsy is the gold standard for diagnosing and staging of FLD, but its application is limited by the invasive nature, risk of complications and high cost.[6] Various imaging modalities, such as computed tomography (CT) scan and ultrasound, have also been used for detecting the presence or quantifying the severity of liver fat noninvasively.[7] However, the limited diagnostic accuracy of detecting mild degree hepatic steatosis with CT and ultrasound is an issue that should be taken into consideration.[7] Transient elastography is non-invasive technique that uses both ultrasound and low-frequency elastic waves to qualify liver fibrosis. However, recent research suggests that steatosis may influence its diagnostic performance.[8] Controlled attenuation parameter (CAP) is an ultrasound-based diagnostic method and added to transient elastography enables simultaneous assessment of steatosis and fibrosis,[9] but the clinical application of CAP is limited by influences of covariates.[10] Therefore, the development of clinically valid and minimally-invasive methods are required to enhance early diagnosis of FLD.

MicroRNAs (miRNAs) are small non-coding RNA molecules of 20–25 nucleotides in length that regulate gene expression at the post-transcriptional level.[11] Recently, the interest in miRNAs has increased tremendously because they offer new insights into disease mechanisms and have a great potential to be used in the clinic as diagnostic biomarkers and/or even therapeutic targets.[12, 13] In line with this, numerous studies have reported increased levels of circulating miR-122 in liver diseases with different etiologies and suggested this miRNA as a potential biomarker and target of therapy in liver dysfunction.[14, 15] Although extensive research has explored the role of miRNAs in the pathophysiology of liver diseases, little is known about the potential of circulatory miRNAs as FLD biomarker in the population level. Moreover, the available studies have mainly used qPCR-based methods and limited to small number of miRNAs and sample sizes.[16, 17]

The aim of this study was to systematically investigate the association of circulating miRNAs in plasma with FLD in a population-based setting. To achieve this aim, we conducted regression models to identify miRNAs that are associated with FLD and liver enzymes at the baseline in the Rotterdam Study cohort. Moreover, we performed subsequent analyses to check whether the identified miRNAs are linked to the risk of hepatic steatosis or liver fibrosis after follow-up and are involved in the known pathways underlying liver diseases.

Material and methods

Study population

This study was embedded within the framework of the Rotterdam Study (RS), a prospective cohort study of individuals aged ≥45 years living in the Ommoord district of Rotterdam, the Netherlands. The objectives and design of the Rotterdam Study have been described in detail elsewhere.[18] In 1989, the first cohort of study participants (RS-I) comprised 7983 persons aged 55 years or over. In 2000, the second cohort (RS-II) was extended to include an additional 3011 participants who moved into the study district or had become 55 years of age. A further extension of the Rotterdam Study cohort (RS-III) formed in 2006 and include 3932 participants living in the research area and aged 45 years and older. Follow-up examinations were scheduled periodically, approximately every 3-5 years. All participants in the study provided written inform consent to participate and to obtain information from their treating physicians.

For this study, we used the expression profiles of circulating miRNA in plasma, collected between 2002 and 2005, from a random subset (n=1,000) of the fourth visit of the first cohort (RS-I-4) and a random subset (n=999) of the second visit of the second cohort (RS-II-2). Among them, 1,991 participants had serum Gamma-glutamyl transferase (GGT) and Alkaline phosphatase (ALP) levels available at baseline that were included to investigate the association of miRNAs with liver enzymes. Moreover, 954 participants who underwent CT-scan from June 2003 to February 2006 were included for investigating the associations of miRNAs with FLD in a cross-sectional setting (**Figure 1**).

For the longitudinal analysis, 1,999 participants at the baseline were followed-up >6 years, until the fifth visit of the first cohort (RS-I-5) and the third visit of the second cohort (RS-II-3). Among these, 1,211 participants who underwent abdominal ultrasound between January 2009 and June 2014 were included to investigate the association of miRNAs with hepatic steatosis (424 cases). Of these, 1,147 participants were included to investigate the association of miRNAs with NAFLD (321 cases) and alcoholic FLD (76 cases). Moreover, out of the 1,999 individuals, 777 participants who underwent transient elatography were included to investigate the association of miRNAs with liver fibrosis (33 cases). A more detailed flow chart for the selection of study participants is shown in **Figure 1**.

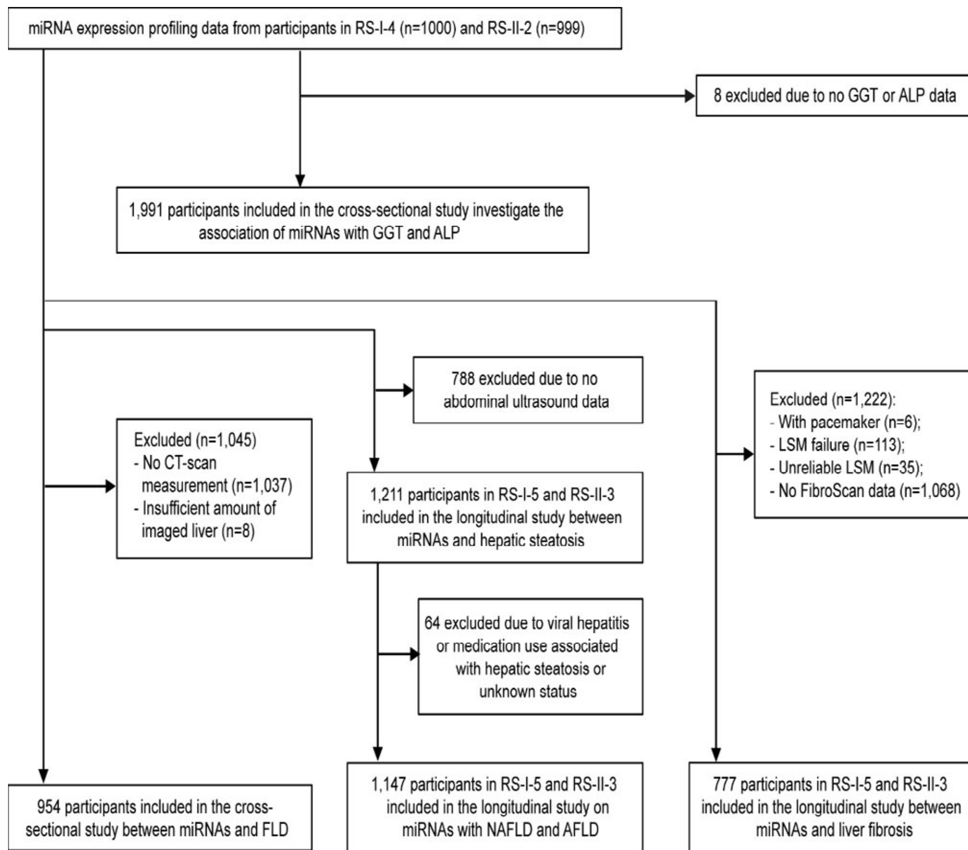


Figure 1. An overview of the study to identify circulatory miRNAs associated with FLD. Cross-sectional studies at baseline were performed in participants from RS-I-4 and RS-II-2 with liver enzymes and CT-scan data available. Longitudinal studies were performed in participants from RS-I-5 and RS-II-3 who underwent abdominal ultrasound or transient elastography. Abbreviations: AFLD, alcoholic fatty liver disease; ALP, Alkaline phosphatase; CT, computed tomography; FLD, fatty liver disease; GGT, gamma-glutamyl transferase; LSM, liver stiffness measurement.; miRNAs, microRNAs; NAFLD, non-alcoholic fatty liver disease; RS, Rotterdam Study; RS-I-4, the fourth visit of the first cohort; RS-I-5, the fifth visit of the first cohort; RS-II-2, the second visit of the second cohort; RS-II-3, the third visit of the second cohort

MiRNA expression profiling

Plasma levels of cell-free miRNAs were determined using the HTG EdgeSeq miRNA Whole Transcriptome Assay (WTA), which measures the expression of 2083 human mature miRNAs. The WTA characterizes miRNA expression patterns, and measures the expression of 13 housekeeping genes, that allows flexibility in data normalization and analysis. Plasma samples, for two re-measurements that generally is sufficient to obtain a valid result for all samples, were sent to HTG Molecular Diagnostics (AZ, USA) for sequencing. Each sample was tagged individually with molecular barcodes, tagged samples were pooled and sequenced on an Illumina NextSeq 500 sequencer (Illumina, San Diego, CA, USA). Quantification of miRNA expression was based on counts per million (CPM). The log2 transformation of CPM was used as standardization and adjustment for total reads within each sample. The miRNAs with log2 CPM< 1.0 were considered as not expressed in the samples. Of the 2083 miRNAs, 591 miRNAs were expressed at good levels in plasma. These 591 well-expressed miRNAs are those with >50% values above Lower Limit of Quantification (LLOQ). The LLOQ level is based on a monotonic decreasing spline curve fit between the means and standard deviations of all miRNAs.

Assessment of liver fat with CT scan

As part of a larger project on the assessment of vascular calcification, ECG-gated, cardiac, non-enhanced CT scanning on a 16-slice (n=251) or 64-slice (n=703) multi-detector CT scanner (Somatom Sensation 16 or 64, Siemens, Forchheim, Germany). Imaging parameters of the scans are described in detail elsewhere.[19]

Using this cardiac scans, we evaluated the liver density (attenuation) using a standardized procedure. First, we placed three circular regions of interest (ROIs) in the liver and calculated the mean liver attenuation (LA) within these regions.[20] These ROIs are delineated throughout the imaged liver tissue (including both the left and right liver lobes) are carefully chosen to include only liver tissue, and avoiding the large blood vessels, cysts, or focal lesions. Next, we calculated the mean Hounsfield unit (HU) value from these three measurements as a marker of total amount of liver fat, which is a reliable proxy for the mean LA value of the whole liver.[20] All measurements were done using Philips iSite Enterprise software (Royal Philips Electronics N.V. 2006) and described in detail elsewhere.[21]

The CT diagnosis of liver fat is made by measuring mean LA in HU or the difference between the liver and spleen.[20] As the amount of liver fat increases, the measured LA decreases, that means low LA was equal to high risk of fatty liver. However, in the present study fatty liver disease was defined as mean LA< 40 HU.[22]

Assessment of hepatic steatosis and liver fibrosis

Hepatic steatosis was assessed by using abdominal ultrasound, which was carried out by a certified and skilled technician (Pavel Taimr) on Hitachi HI VISION 900.[23] Images were stored digitally and re-assessed by a single hepatologist with more than ten years of experience in ultrasonography. Diagnosis of steatosis was determined dichotomously as presence of a hyperechogenic liver parenchyma according to the protocol by Hamaguchi et al.[24]

Moreover, liver fibrosis was assessed using transient elastography (FibroScan®, EchoSens, Paris, France). Applied implementation of this examination has been described in detail previously 25. Liver stiffness measurement (LSM) was performed by a single certified and experienced operator, who obtained 10 serial measurements using either the M or XL-probe dependent on the thickness of the subcutaneous fat layer 23. Moreover, LSM interquartile range/median LSM >0.3 kilopascals (kPa) and LSM ≥ 7.1 kPa were regarded as poorly reliable 26. In the present study, LSM ≥ 9.0 kPa was used as a cutoff suggesting clinically relevant liver fibrosis.[8]

Assessment of covariates and liver enzymes

Information on smoking behavior, medication use and blood sampling, was obtained during home interviews.[18] Height and weight were measured, and the body mass index (BMI) [(weight in kg)/(height in m)^2] was calculated. Waist circumferences was measured at the level midway between the lower rib margin and the iliac crest with the participant in a standing position. Smoking status was categorized into never, current or former and were classified (yes/no), for ever-smokers were regarded as current and former smokers combined. Alcohol consumption was assessed in grams of ethanol per day and were classified (yes/no). Excessive alcohol consumption was defined as alcohol intake >30 g/day for men and >20 g/day for women.[23] Hypertension was defined as a systolic blood pressure (BP) ≥140mmHg or a diastolic BP ≥90 mmHg or the use of BP-lowering drugs prescribed for hypertension.[27] Diabetes mellitus was defined according to recent WHO guidelines[28] as fasting blood glucose ≥7.0 mmol/L or non-fasting blood glucose between ≥11.1 mmol/L or the use of antidiabetic medication. From the blood samples, concentrations high-density lipoprotein (HDL) cholesterol were determined using enzymatic procedures.[29] Serum GGT and ALP levels were determined within 2 weeks using a Merck Diagnostica kit on an Elan Autoanalyzer (Merc, Darmstadt, Germany). According to local cutoffs, elevation of GGT was defined as >34 U/L for women and >49 U/L for men, and elevation of ALP was defined as >97 U/L for women and >114 U/L for men.[30]

Statistical analysis

Continuous variables are reported as mean ± standard deviation (SD) unless stated otherwise and categorical variables were presented as sample sizes and percentages. To obtain a normal distribution, skewed variables (serum HDL cholesterol, GGT and ALP) were log transformed. In addition, the amount of liver fat (A) had a left skewed and we used exponential transformed values (B) using the formula [$B=A^{3.5}/10000$].[21]

The multivariable linear regression models were used to check the association of miRNA levels with serum GGT and ALP levels. Beta, standard error (SE), p-value were reported. The Bonferroni-corrected p-value threshold was calculated based on the number of tested miRNAs ($0.05/591=8.46\times10^{-5}$). In basic model (model 1), we adjusted the analysis for age and sex. The multivariable model (model 2), was additionally adjusted for waist circumference, ever smoking, alcohol consumption, hypertension, diabetes mellitus and serum HDL cholesterol. Because the missing values were likely to be missing at random and for avoidance of loss in efficiency, missing values on covariates (ranging from 0.1% to 1.7%) were imputed using a multiple imputation technique (N=5 imputations). All analyses were done using SPSS statistical software (SPSS, version 25; IBM Corp, Armonk) and R software version 3.5.2 (The R Foundation for Statistical Computing, Vienna, Austria). Sensitivity analyses were performed by adjusting for more variables. Model 3 was built by adding GGT and ALP to model 2. In model 4, we further adjusted for potential intermediator factors including in the model 2, use of lipid-lowering medication, use of bile and liver medications. In model 5, we adjusted for all potential intermediator factors (including model 2, GGT, ALP, use of lipid-lowering medication, use of bile and liver medications).

Furthermore, multivariable logistic regression models were used to investigate longitudinally the association of the plasma levels of the identified miRNAs with prevalence of hepatic steatosis and liver fibrosis after a median follow-up of 6.4 years [interquartile range (IQR): 5.9-7.0 years]. The Bonferroni correction was used to set the significance threshold.

Two databases, the Human miRNA tissue atlas (https://ccb-web.cs.uni-saarland.de/tissueatlas)[31] and Human miRNA expression profiles (https://guanfiles.dcmb.med.umich.edu/mirmine/ index.html), were used to check whether the identified miRNAs are expressed in the liver. We also searched the literature[32-35] and several web tools (e.g., miR2Diease and GWAS catalog) to see whether the identified miRNAs are associated with liver function and diseases.

Results

At baseline, 954 participants who had miRNA expression data and CT-based liver fat measurement were included to test the association of miRNAs and FLD. The mean age of the study population was 68.8±6.7 years, and 46.6% were male. The mean LA in the population was 61.6 HU (IQR: 55.4-65.6 HU). Among the study participants, 14.8% were diagnosed with cancer, but none of them was diagnosed with liver cancer. At follow-up, 1,211 participants who had underwent abdominal ultrasound were included to test the association of miRNAs and hepatic steatosis. The man age of the study population was 76.3±6.5 years, and 42.4% were male. Lifestyle, clinical and biochemical characteristics of all study participants are presented in **Table 1**. Comparison of characteristics between healthy controls and FLD patients based on CT-scan and ultrasound data (in baseline and follow-up study) are shown in **Table S1**. The participants with FLD have significantly higher BMI, waist circumference and alcohol consumption than healthy controls. In addition, compare to healthy controls, individuals with FLD have significantly lower serum HDL cholesterol.

Table 1. Characteristics of the study population

Characteristic	Baseline (n=954) With CT-scan data	Follow-up (n=1,211) With ultrasound data
Age, years	68.8±6.7	76.3±6.5
Male, n (%)	445 (46.6)	513 (42.4)
Body mass index , kg/m^2	27.9±4.0	27.5±4.1
Waist circumference, cm	94.4±11.7	93.2±12.0
Hypertension, n (%)	709 (74.3)	1,047 (86.5)
Blood-pressure-lowering medication, n (%)	385 (40.4)	644 (53.2)
Smoking status, n (%)		
Ever	677 (71.0)	791 (65.3)
Current	130 (13.6)	116 (9.6)
Former	547 (57.4)	675 (55.7)
Diabetes mellitus, n (%)	123 (12.9)	156 (12.9)
Use of lipid-lowering medication, n (%)	246 (25.8)	372 (30.7)
Alcohol intake, grams/day	8.6 (1.4-20.0)	8.6 (1.6-8.6)
Mean liver attenuation, HU	61.6 (55.4-65.6)	–
Serum hdl cholesterol, mmol/l	1.4 (1.2-1.7)	1.4 (1.2-1.7)
GGT level, u/l	26.0 (18.0-39.0)	24.0 (17.0-34.2)
ALP level, u/l	77.0 (66.0-91.0)	68.0 (57.0-80.0)
Cancer, n (%)	141 (14.8)	–
Liver cancer, n (%)	0 (0)	–
Fatty liver, n (%)	47 (4.93)	424 (35)

Note: The table shows characteristics of 954 participants with CT-scan data at baseline and 1211 participants with ultrasound data at follow-up. Values are represented as mean (±standard deviation), sample sizes (%), or median (inter-quartile range) for characteristics with skewed distributions. Abbreviations: ALP, Alkaline phosphatase; BP, blood pressure; GGT, Gamma-glutamyl transferase; HDL, high-density lipoprotein; HU, Hounsfield unit.

In the linear regression analysis with liver enzymes, 37 miRNAs were significantly associated with serum GGT levels and 29 miRNAs with serum ALP levels, at the Bonferroni-corrected p-value< 8.46×10^{-5} (0.05/591 well-expressed miRNAs) (**Table S2** and **Table S3**, respectively). Volcano plot showing differently expressed miRNAs in relation to GGT and ALP levels is depicted in **Figure 2A-B**.

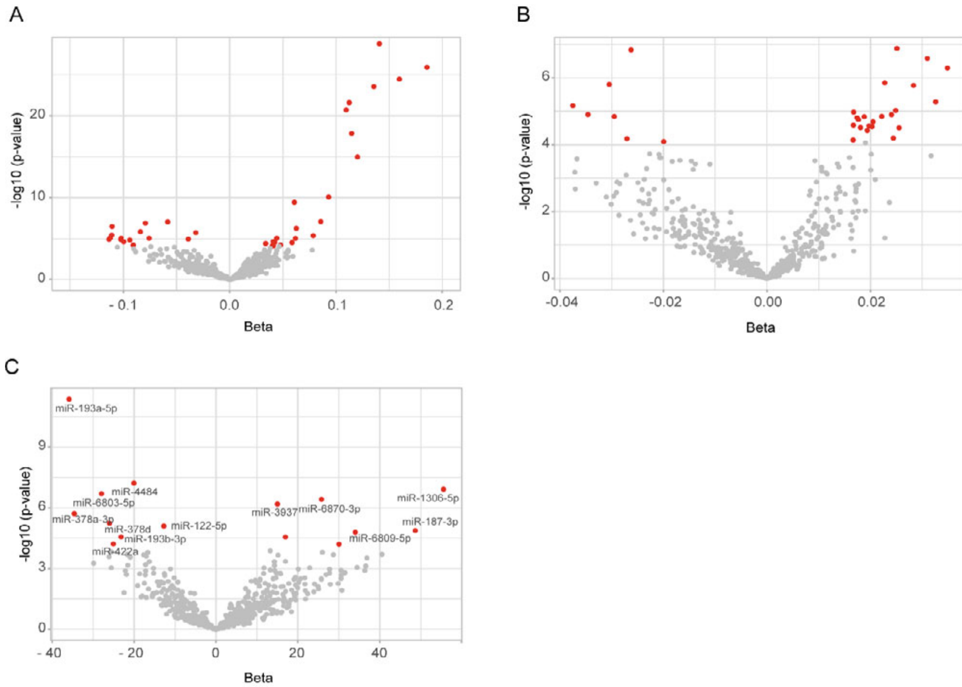


Figure 2. Volcano plots showing correlation between plasma levels of miRNAs and GGT (A), ALP (B), and continuous Hounsfield Unit values (C). The red dots indicate miRNAs significantly associated at Bonferroni-corrected P < 8.46 × 10−5. The grey dots indicate miRNAs with no significant association. The name of miRNAs that were significantly associated with liver enzymes and the continuous Hounsfield Unit values are mentioned in (C). Abbreviations: ALP, Alkaline phosphatase; GGT, Gamma-glutamyl transferase; miRNAs, microRNAs.

Using a linear regression analysis of the continuous HU values, in the multivariable model 2, we found 15 miRNAs to be significantly associated at Bonferroni-corrected p-value< 8.46×10^{-5} (**Table S4**). Volcano plot showing differently expressed miRNAs in relation to the continuous HU values is depicted in **Figure 2C**. In addition, in a logistic regression model testing the association of miRNA levels with the dichotomous HU values assessing FLD (mean LA ≤ or >40), 6 miRNAs were significantly associated (p-value< 8.46×10^{-5}) (**Figure 3**). In model 3, further adjusting for GGT and ALP changed slightly the associations and with less miRNAs significant associated, but 2 out of the 17 and 6 out of 17 miRNAs remained significant using dichotomous and continuous Hounsfield Unit values, respectively (p-value< 8.46×10^{-5}) (**Table S5** and **Table S6**). Additional adjustment for use of lipid-lowering medication, use of bile and liver medications (model 4) did not change the observed associations between miRNAs and FLD (**Table S5** and **Table S6**). In model 5, we added GGT, ALP, use of serum lipid reducing agents, use of bile and liver medications to model 2, the associations of miRNAs with FLD were similar to model 3, the same miRNAs remained significant (**Table S5** and **Table S6**).

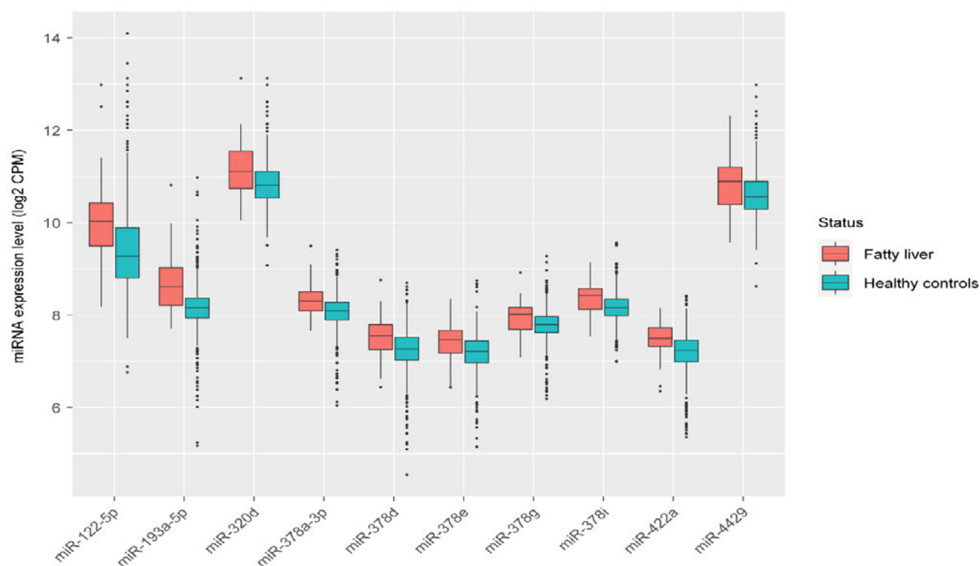


Figure 3. Comparison between expression levels of the top 10 miRNAs in patients with fatty liver disease and healthy controls.

Of these, 6 miRNAs (miR-193a-5p, miR-378a-3p, miR-422a, miR-378d, miR-320d and miR-378e) were significantly associated with fatty liver disease at Bonferroni-corrected $P < 8.46 \times 10^{-5}$. Abbreviations: CPM, counts per million; miRNAs, microRNAs.

Collectively, our cross-sectional studies with the baseline data revealed 61 unique miRNAs associated with liver enzymes (GGT and/or ALP) and 17 unique miRNAs associated with CT-based FLD (continuous or dichotomous HU values). Of these, 14 miRNAs were common in both lists that were selected for further analyses (**Table 2**).

The longitudinal analysis was performed for the 14 miRNAs by using ultrasound and FibroScan data. Using a logistic regression in the multivariable model 2, we found significant association between four of the 14 miRNAs (miR-193a-5p, miR-122-5p, miR-378d, and miR-187-3p) with hepatic steatosis at Bonferroni-corrected p-value< 3.57×10^{-3} (0.05/14 miRNAs) (**Table 3**). Moreover, we found significant association of miR-122-5p and miR-187-3p with NAFLD (p-value< 3.57×10^{-3}), and miR-3937 was nominally (p-value< 0.05) associated with alcoholic FLD (**Table S7**). Using FibroScan data and in a multivariable logistic regression model, we found miR-193a-5p (p-value= 5.58×10^{-3}, β= 1.11), miR-122-5p (p-value=0.0147, β= 0.45) and miR-193b-3p (p-value= 0.0102, β= 1.19) were to be nominally associated with liver fibrosis (**Table 3**).

Table 2. Circulatory miRNAs significantly associated with CT-scan based fatty liver disease and serum liver enzymes

miRNA ID	CT-based FLD				Liver enzymes			
	Dichotomous HU values		Continues HU values		GGT		ALP	
	Beta	p-value	Beta	p-value	Beta	p-value	Beta	p-value
miR-193a-5p	1.86	2.02×10^{-10}	-35.88	4.26×10^{-12}	0.14	6.15×10^{-29}	0.001	7.99×10^{-01}
miR-4484	0.71	8.33×10^{-4}	-20.04	6.02×10^{-08}	0.03	4.02×10^{-05}	0.01	3.68×10^{-04}
miR-1306-5p	-0.60	3.15×10^{-01}	55.59	1.21×10^{-07}	-0.09	6.45×10^{-05}	-0.02	1.12×10^{-01}
miR-378a-3p	2.36	1.90×10^{-07}	-34.61	1.95×10^{-06}	0.09	8.45×10^{-08}	-0.01	3.98×10^{-1}
miR-6803-5p	0.70	2.68×10^{-02}	-27.96	2.01×10^{-07}	0.05	6.31×10^{-05}	0.01	1.03×10^{-02}
miR-6870-3p	-0.59	1.60×10^{-02}	25.79	3.87×10^{-07}	-0.06	9.04×10^{-08}	-0.02	8.00×10^{-05}
miR-3937	-0.54	1.83×10^{-03}	15.02	6.59×10^{-07}	-0.03	1.90×10^{-06}	-0.01	3.84×10^{-04}
miR-122-5p	0.54	1.72×10^{-04}	-12.74	8.05×10^{-06}	0.14	2.39×10^{-116}	0.01	4.73×10^{-04}
miR-422a	2.01	9.87×10^{-06}	-25.07	6.09×10^{-05}	0.06	9.71×10^{-06}	0.002	7.37×10^{-01}
miR-378d	1.75	1.11×10^{-05}	-26.00	5.87×10^{-06}	0.06	5.55×10^{-07}	-0.004	4.92×10^{-01}
miR-187-3p	-1.31	2.52×10^{-02}	48.67	1.33×10^{-05}	-0.11	3.78×10^{-06}	-0.03	1.80×10^{-02}
miR-6809-5p	-0.73	5.51×10^{-02}	34.03	1.61×10^{-05}	-0.08	1.45×10^{-06}	-0.01	1.11×10^{-01}
miR-193b-3p	1.01	2.70×10^{-03}	-23.17	2.77×10^{-05}	0.11	2.01×10^{-21}	-0.001	7.85×10^{-01}
miR-4713-3p	-0.55	1.02×10^{-02}	16.96	2.80×10^{-05}	-0.03	8.53×10^{-04}	-0.01	1.13×10^{-02}
miR-320d	1.16	2.87×10^{-05}	-14.30	4.54×10^{-03}	0.02	1.72×10^{-01}	0.004	4.10×10^{-01}
miR-34b-3p	-0.76	8.88×10^{-02}	30.06	6.27×10^{-05}	-0.06	1.79×10^{-04}	-0.01	3.41×10^{-01}
miR-378e	1.64	6.88×10^{-05}	-21.54	6.86×10^{-04}	0.06	3.13×10^{-05}	-0.01	4.42×10^{-01}

Note: Model 2: adjusted for age, sex, waist circumference, ever smoking, alcohol consumption, hypertension, diabetes mellitus and serum HDL cholesterol. The table is sorted based on Bonferroni-corrected P value association of miRNAs with dichotomous or continues Hounsfield Unit values in model 2. The Bonferroni-corrected significance threshold is $P < 8.46 \times 10^{-5}$ (0.05/591 miRNAs). The P values surpassing the significance threshold are "a". Abbreviations: ALP, Alkaline phosphatase; FLD, fatty liver disease; GGT, Gamma-glutamyl transferase; HU, Hounsfield Unit; miRNA, microRNA.

Additionally, we searched the Human miRNA tissue atlas and the miRmine database to see whether the 14 miRNAs associated with FLD in plasma are also expressed in the liver that are shown in **Table S8**. Among them, miR-122-5p is a specifically expressed miRNA with the tissue specificity index (TSI) of 0.97 and highly expressed in the liver. Then, we sought to find whether the 14 identified miRNAs are reported in previous studies to be associated with liver function or/and diseases. A summary of evidence for associations between 9 of these miRNAs and liver diseases are shown in **Table S9**. Finally, we extracted SNPs annotated to the 16 identified miRNAs and checked their associations with FLD and liver enzymes using summary statistics data from pervious GWAS.[34, 35] There were 63 SNPs related to miR-193a-5p, miR-378d and miR-193b-3p, none of them showed significant association after correcting the p-value for multiple testing based on the number of tested SNPs.

Table 3. Longitudinal study of the 14 identified miRNAs with hepatic steatosis and liver fibrosis

miRNA ID	Hepatic Steatosis			Liver Fibrosis		
	Beta	SE	p-value	Beta	SE	p-value
miR-193a-5p	0.54	0.16	2.02×10^{-10}	1.11	0.40	5.58×10^{-03}
miR-4484	0.21	0.10	4.65×10^{-02}	0.41	0.28	1.37×10^{-01}
miR-1306-5p	-0.33	0.28	2.36×10^{-01}	-1.16	-0.73	1.14×10^{-01}
miR-378a-3p	0.37	0.21	8.19×10^{-02}	1.23	0.64	5.67×10^{-02}
miR-6803-5p	0.13	0.15	3.79×10^{-01}	0.70	0.41	8.42×10^{-02}
miR-6870-3p	-0.22	0.14	1.18×10^{-01}	0.07	0.45	8.80×10^{-01}
miR-3937	-0.13	0.09	1.44×10^{-01}	-0.18	0.26	4.87×10^{-01}
miR-122-5p	0.33	0.08	6.06×10^{-05}	0.45	0.18	1.47×10^{-02}
miR-422a	0.37	0.18	4.12×10^{-02}	0.96	0.54	7.61×10^{-02}
miR-378d	0.54	0.17	1.65×10^{-03}	0.73	0.53	1.69×10^{-01}
miR-187-3p	-1.01	0.34	2.76×10^{-03}	-0.42	0.98	6.68×10^{-01}
miR-6809-5p	-0.55	0.24	2.49×10^{-02}	-0.66	-0.76	3.83×10^{-01}
miR-193b-3p	0.22	0.15	1.42×10^{-01}	1.19	0.46	1.02×10^{-02}
miR-378e	0.50	0.18	7.25×10^{-03}	0.94	0.60	1.14×10^{-01}

Note: Model 2: adjusted for age, sex, waist circumference, ever smoking, alcohol consumption, hypertension, diabetes mellitus and serum HDL cholesterol. The table is sorted based on the association of 14 miRNAs with the continues Hounsfield Unit values in the cross-sectional study. The p-values surpassing the Bonferroni-corrected threshold of p-value< 3.57×10^{-3} (0.05/14 miRNAs) are bold and nominal associations with p-value <0.05 are underlined. Abbreviations: miRNA, microRNA; SE, standard error.

Discussion

In this study, we investigated the association between circulating miRNAs and liver enzymes in a population-based setting and found 61 unique miRNAs to be associated with serum GGT or ALP levels. Moreover, we found plasma levels of 17 miRNAs to be associated with CT-based FLD, 14 of these were also associated with the liver enzymes. Higher plasma levels of three and lower plasma level of one of these 14 miRNAs were significantly associated with hepatic steatosis after >6 years follow-up. These findings indicate that plasma levels of miRNAs can be considered as potential biomarkers of FLD and hepatic steatosis in the general population.

Several studies have demonstrated the potential of miRNAs to be used as biomarkers for liver diseases.[36-38] However, previous studies have conducted for subset of miRNAs, using qPCR-based methods, or on the modest sample sizes. While our study is embedded within the Rotterdam Study with much larger sample size, based on RNA-sequencing method, conducted genome-wide profiling of almost all important cell-free miRNAs, and adjusted for a broad range of potential confounders, such as waist circumference, smoking status, alcohol consumption, hypertension and diabetes mellitus, which have been overlooked in most of previous studies.[36, 39] Such a large-scale population-based study with long term follow-up data provided a more statistical power to detect multiple significant associations. Compared to microarray or qPCR-based profiling techniques, the cell-free RNA-seq analysis can provide higher sensitivity to measure miRNAs expression levels over a wide dynamic range and with ability to identify novel miRNAs.[40] Additionally, due to the high stability of cell-free miRNAs in body fluids and accessibility of plasma compared with the target tissue, the identified miRNAs can be considered as potential easy-to-use biomarkers in clinical routine.[41]

Previous studies on human or mouse model have demonstrated particularly miR-122-5p as potential biomarkers and therapeutic target for liver diseases.[39, 42, 43] In line with previous studies, we found that the higher plasma miR-122-5p level is significantly associated with FLD and liver enzymes also in a population-based setting. In addition to the well-established liver-associated miR-122, we found evidence in previous studies for 8 of the other identified miRNAs in our study to be associated with liver diseases, indicating the importance of these miRNAs in pathways underlying liver function and diseases. In particular, the expression of miR-193a-5p, which is one of our top miRNAs associated with FLD and hepatic steatosis, is reported to be upregulated in HCC tissues,[44] whereas miR-193a-5p can distinguish HCC from other non-HCC individuals[43] and inhibited HCC development through targeting *SPOCK1*.[45] Similarly, miR-422a was related to NAFLD,[46] miR-378d[47] miR-187-3p,[48] miR-6809-5p[49] and miR-4484[50] were associated with HCC. Moreover, miR-193b-3p which were significantly associated with FLD and nominally associated with liver fibrosis in our study, have been verified previously to be involved in the pathogenesis of liver fibrosis in vitro.[51] Finally, the members of miR-378 family, in particular miR-378a-3p that has been also proposed to have a therapeutic potential for liver fibrosis.[52]

Our results showed a minimal of the use of serum lipid reducing agents, use of bile and liver medications on the observed associations between miRNAs and FLD. We observed slightly change in the associations between miRNAs and FLD by adding GGT and ALP to the model. This difference may indicate that liver enzymes have more stronger links to FLD compared to medication use. Also, we did not find significant association between SNPs related to the identified miRNAs and FLD in the summary statistics from previous GWAS, but we need to take into consideration the sample sizes of available GWAS of liver diseases. To date, the GWAS on liver diseases are mainly from two studies,[34, 35] Nakamura et al. conducted a study to identify susceptibility loci for primary biliary cirrhosis, a GWAS in 963 Japanese individuals and in a subsequent replication study including 1,402 other Japanese individuals. The sample sizes of this study is limited and the analysis conducted in Asia population, while our study was conducted in European population, and miRNAs expression might exhibits population differences.[53] In addition, Namjou et al. conducted a GWAS using both adult and pediatric participants (1,106 NAFLD cases and 8,571 controls) from electronic medical records to identify genetic contributions to NAFLD. As the cohorts in that GWAS study represent many geographic area in USA, other ancestry groups are under-represented in the electronic medical records. Thus, it is possible that future trans-ethnic GWAS with larger samples sizes find association between some of the identified miRNAs and FLD.

Our study has some limitations that should be considered. First, in the cross-sectional observational study the ability to assess causality or temporality is limited. We therefore assessed additionally the associations of the identified miRNAs as biomarkers for diagnosis hepatic steatosis after follow-up. Future studies are still needed to confirm our findings in longitudinal settings considering the incidence date, longer follow-up time and in different age groups. Second, we defined the mean LA< 40 HU as FLD and found 47 cases out of 954 individuals (5%), which is lower than the excepted prevalence of FLD in the general population. A liver-to-spleen ratio< 1.0 is comparable to using a mean LA cut off ≤ 51 HU for diagnosis of mild liver fat.[54] However, previous studies have demonstrated different cut-offs, mainly a cut-off value of 40 HU on non-enhanced CT as the most clinically indicator for moderate-to severe steatosis.[20, 22] In our study, the cut-off value is 40 HU for FLD, it is relatively strict than using the mean LA≤ 51 HU, which increases the certainty of identifying participants with true FLD and also results in a lower prevalence. Therefore, we performed cross-sectional analysis at baseline with the continuous HU values and liver enzymes as well. The majority (14 out of 17) of the identified miRNAs with CT-scan data showed significant association with liver enzyme, indicating the robustness of our results. Yet, compare to the liver biopsy as the gold standard, but invasive method, to measure FLD, CT-based liver fat has limited diagnostic accuracy of detecting mild degree hepatic steatosis. Therefore, there might be some known or unknown causes for low density of the liver on CT scan. Also, there might be some inconsistency between CT-scans and ultrasound data for diagnosing FLD and hepatic steatosis. In an optimal setting, one should use the repeated measurement of liver fat by a similar diagnostic method for longitudinal analysis.

In conclusion, we found that plasma levels of several miRNAs were significantly associated with FLD and hepatic steatosis that can be considered as plasma disease biomarkers in this population-based study. Future research need to be conducted even with more sample sizes and longer follow-up times in order to confirm the potential of the identified miRNAs as biomarkers for early diagnosis and progression of FLD and also to uncover underlying molecular mechanisms by which these miRNAs may control liver fat.

Supplementary material

Supplementary material is a viable on

References

1. Younossi Z, Tacke F, Arrese M, Chander Sharma B, Mostafa I, Bugianesi E, et al. Global Perspectives on Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis. Hepatology. 2019;69(6):2672-82.
2. Fatiha Nassir, R. Scott Rector, Ghassan M. Hammoud, Ibdah JA. Pathogenesis and Prevention of Hepatic Steatosis. Gastroenterology & Hepatology. 2015;11(3):167-75.
3. VanWagner LB, Ning H, Allen NB, Ajmera V, Lewis CE, Carr JJ, et al. Alcohol Use and Cardiovascular Disease Risk in Patients With Nonalcoholic Fatty Liver Disease. Gastroenterology. 2017;153(5):1260-72 e3.
4. DiStefano JK, Gerhard GS. Circulating microRNAs in nonalcoholic fatty liver disease. Expert Rev Gastroenterol Hepatol. 2016;10(2):161-3.
5. Eslam M, Sanyal AJ, George J, International Consensus P. MAFLD: A Consensus-Driven Proposed Nomenclature for Metabolic Associated Fatty Liver Disease. Gastroenterology. 2020;158(7):1999-2014 e1.
6. Musaddaq G, Shahzad N, Ashraf MA, Arshad MI. Circulating liver-specific microRNAs as noninvasive diagnostic biomarkers of hepatic diseases in human. Biomarkers. 2019;24(2):103-9.
7. Lee DH. Imaging evaluation of non-alcoholic fatty liver disease: focused on quantification. Clin Mol Hepatol. 2017;23(4):290-301.
8. Eddowes PJ, Sasso M, Allison M, Tsochatzis E, Anstee QM, Sheridan D, et al. Accuracy of FibroScan Controlled Attenuation Parameter and Liver Stiffness Measurement in Assessing Steatosis and Fibrosis in Patients With Nonalcoholic Fatty Liver Disease. Gastroenterology. 2019;156(6):1717-30.
9. Karlas T, Petroff D, Sasso M, Fan JG, Mi YQ, de Ledinghen V, et al. Impact of controlled attenuation parameter on detecting fibrosis using liver stiffness measurement. Aliment Pharmacol Ther. 2018;47(7):989-1000.
10. Oeda S, Takahashi H, Imajo K, Seko Y, Ogawa Y, Moriguchi M, et al. Accuracy of liver stiffness measurement and controlled attenuation parameter using FibroScan((R)) M/XL probes to diagnose liver fibrosis and steatosis in patients with nonalcoholic fatty liver disease: a multicenter prospective study. J Gastroenterol. 2020;55(4):428-40.
11. Otsuka M, Kishikawa T, Yoshikawa T, Yamagami M, Ohno M, Takata A, et al. MicroRNAs and liver disease. Journal Of Human Genetics. 2016;62:75.
12. Schueller F, Roy S, Vucur M, Trautwein C, Luedde T, Roderburg C. The Role of miRNAs in the Pathophysiology of Liver Diseases and Toxicity. Int J Mol Sci. 2018;19(1).
13. Pirola CJ, Fernandez Gianotti T, Castano GO, Mallardi P, San Martino J, Mora Gonzalez Lopez Ledesma M, et al. Circulating microRNA signature in non-alcoholic fatty liver disease: from serum non-coding RNAs to liver histology and disease pathogenesis. Gut. 2015;64(5):800-12.
14. Roderburg C, Luedde T. Circulating microRNAs as markers of liver inflammation, fibrosis and cancer. J Hepatol. 2014;61(6):1434-7.
15. Liu CH, Ampuero J, Gil-Gomez A, Montero-Vallejo R, Rojas A, Munoz-Hernandez R, et al. miRNAs in patients with non-alcoholic fatty liver disease: A systematic review and meta-analysis. J Hepatol. 2018;69(6):1335-48.
16. Mehta R, Otgonsuren M, Younoszai Z, Allawi H, Raybuck B, Younossi Z. Circulating miRNA in patients with non-alcoholic fatty liver disease and coronary artery disease. BMJ Open Gastroenterol. 2016;3(1):e000096.
17. Zarrinpar A, Gupta S, Maurya MR, 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):1546-54.
18. Ikram MA, Brusselle G, Ghanbari M, Goedegebure A, Ikram MK, Kavousi M, et al. Objectives, design and main findings until 2020 from the Rotterdam Study. Eur J Epidemiol. 2020.
19. Odink AE, van der Lugt A, Hofman A, Hunink MG, Breteler MM, Krestin GP, et al. Association between calcification in the coronary arteries, aortic arch and carotid arteries: the Rotterdam study. Atherosclerosis. 2007;193(2):408-13.
20. Kodama Y, Ng CS, Wu TT, Ayers GD, Curley SA, Abdalla EK, et al. Comparison of CT methods for determining the fat content of the liver. AJR Am J Roentgenol. 2007;188(5):1307-12.

21. Wolff L, Bos D, Murad SD, Franco OH, Krestin GP, Hofman A, et al. Liver fat is related to cardiovascular risk factors and subclinical vascular disease: the Rotterdam Study. Eur Heart J Cardiovasc Imaging. 2016;17(12):1361-7.
22. Seong Ho P, Pyo Nyun K, Kyoung Won K, Sang Won L, Seong Eon Y, Sung Won P, et al. Macrovesicular Hepatic Steatosis in Living Liver Donors: Use of CT for Quantitative and Qualitative Assessment. Radiology. 2006;239(1):105-12.
23. Alferink LJM, Trajanoska K, Erler NS, Schoufour JD, de Knegt RJ, Ikram MA, et al. Nonalcoholic Fatty Liver Disease in The Rotterdam Study: About Muscle Mass, Sarcopenia, Fat Mass, and Fat Distribution. J Bone Miner Res. 2019;34(7):1254-63.
24. Hamaguchi M, Kojima T, Itoh Y, Harano Y, Fujii K, Nakajima T, et al. The severity of ultrasonographic findings in nonalcoholic fatty liver disease reflects the metabolic syndrome and visceral fat accumulation. Am J Gastroenterol. 2007;102(12):2708-15.
25. Koehler EM, Plompen EP, Schouten JN, Hansen BE, Darwish Murad S, Taimr P, et al. Presence of diabetes mellitus and steatosis is associated with liver stiffness in a general population: The Rotterdam study. Hepatology. 2016;63(1):138-47.
26. Boursier J, Zarski JP, de Ledinghen V, Rousselet MC, Sturm N, Lebail B, et al. Determination of reliability criteria for liver stiffness evaluation by transient elastography. Hepatology. 2013;57(3):1182-91.
27. Williams B, Mancia G, Spiering W, Agabiti Rosei E, Azizi M, Burnier M, et al. 2018 ESC/ESH Guidelines for the management of arterial hypertension. Eur Heart J. 2018;39(33):3021-104.
28. Origanization.WH. guidelines - World Health Organization. 2016.
29. Braun KVE, Dhana K, de Vries PS, Voortman T, van Meurs JBJ, Uitterlinden AG, et al. Epigenome-wide association study (EWAS) on lipids: the Rotterdam Study. Clin Epigenetics. 2017;9:15.
30. Koehler EM, Sanna D, Hansen BE, van Rooij FJ, Heeringa J, Hofman A, et al. Serum liver enzymes are associated with all-cause mortality in an elderly population. Liver Int. 2014;34(2):296-304.
31. Ludwig N, Leidinger P, Becker K, Backes C, Fehlmann T, Pallasch C, et al. Distribution of miRNA expression across human tissues. Nucleic Acids Res. 2016;44(8):3865-77.
32. Huan T, Rong J, Liu C, Zhang X, Tanriverdi K, Joehanes R, et al. Genome-wide identification of microRNA expression quantitative trait loci. Nat Commun. 2015;6:6601.
33. Majid Nikpay, Kaitlyn Beehler, Armand Valsesia, Jorg Hager, Mary-Ellen Harper, Robert Dent, et al. Genome-wide identification of circulating-miRNA expression quantitative trait loci reveals the role of several miRNAs in the regulation ofcardiometabolic phenotypes. European Society of Cardiology. 2019;115(11):1629-45.
34. Nakamura M, Nishida N, Kawashima M, Aiba Y, Tanaka A, Yasunami M, et al. Genome-wide association study identifies TNFSF15 and POU2AF1 as susceptibility loci for primary biliary cirrhosis in the Japanese population. Am J Hum Genet. 2012;91(4):721-8.
35. Namjou B, Lingren T, Huang Y, Parameswaran S, Cobb BL, Stanaway IB, et al. GWAS and enrichment analyses of non-alcoholic fatty liver disease identify new trait-associated genes and pathways across eMERGE Network. BMC Med. 2019;17(1):135.
36. Salvoza NC, Klinzing DC, Gopez-Cervantes J, Baclig MO. Association of Circulating Serum miR-34a and miR-122 with Dyslipidemia among Patients with Non-Alcoholic Fatty Liver Disease. PLoS One. 2016;11(4):e0153497.
37. Auguet T, Aragones G, Berlanga A, Guiu-Jurado E, Marti A, Martinez S, et al. miR33a/miR33b* and miR122 as Possible Contributors to Hepatic Lipid Metabolism in Obese Women with Nonalcoholic Fatty Liver Disease. Int J Mol Sci. 2016;17(10).
38. Raitoharju E, Seppala I, Lyytikainen LP, Viikari J, Ala-Korpela M, Soininen P, et al. Blood hsa-miR-122-5p and hsa-miR-885-5p levels associate with fatty liver and related lipoprotein metabolism-The Young Finns Study. Sci Rep. 2016;6:38262.
39. Brandt S, Roos J, Inzaghi E, Kotnik P, Kovac J, Battelino T, et al. Circulating levels of miR-122 and nonalcoholic fatty liver disease in pre-pubertal obese children. Pediatr Obes. 2018;13(3):175-82.
40. Eminaga S, Christodoulou DC, Vigneault F, Church GM, Seidman JG. Quantification of microRNA expression with next-generation sequencing. Curr Protoc Mol Biol. 2013;Chapter 4:Unit 4 17.

41. Szelenberger R, Kacprzak M, Saluk-Bijak J, Zielinska M, Bijak M. Plasma MicroRNA as a novel diagnostic. Clin Chim Acta. 2019;499:98-107.
42. Cheng JL, Zhao H, Yang SG, Chen EM, Chen WQ, Li LJ. Plasma miRNA-122-5p and miRNA-151a-3p identified as potential biomarkers for liver injury among CHB patients with PNALT. Hepatol Int. 2018;12(3):277-87.
43. Jin Y, Wong YS, Goh BKP, Chan CY, Cheow PC, Chow PKH, et al. Circulating microRNAs as Potential Diagnostic and Prognostic Biomarkers in Hepatocellular Carcinoma. Sci Rep. 2019;9(1):10464.
44. Wang JT, Wang ZH. Role of miR-193a-5p in the proliferation and apoptosis of hepatocellular carcinoma. European Review for Medical and Pharmacological Sciences. 2018;22:7233-9.
45. Li P, Xiao Z, Luo J, Zhang Y, Lin L. MiR-139-5p, miR-940 and miR-193a-5p inhibit the growth of hepatocellular carcinoma by targeting SPOCK1. J Cell Mol Med. 2019;23(4):2475-88.
46. Latorre J, Moreno-Navarrete JM, Mercader JM, Sabater M, Rovira O, Girones J, et al. Decreased lipid metabolism but increased FA biosynthesis are coupled with changes in liver microRNAs in obese subjects with NAFLD. Int J Obes (Lond). 2017;41(4):620-30.
47. Gao L, Xiong D-D, He R-Q, Lai Z-F, Liu L-M, Huang Z-G, et al. Identifying TF-miRNA-mRNA regulatory modules in nitidine chloride treated HCC xenograft of nude mice. American journal of translational research. 2019;11:7503-22.
48. Dou C, Liu Z, Xu M, Jia Y, Wang Y, Li Q, et al. miR-187-3p inhibits the metastasis and epithelial-mesenchymal transition of hepatocellular carcinoma by targeting S100A4. Cancer Lett. 2016;381(2):380-90.
49. Yang PW, Lu ZY, Pan Q, Chen TT, Feng XJ, Wang SM, et al. MicroRNA-6809-5p mediates luteolin-induced anticancer effects against hepatoma by targeting flotillin 1. Phytomedicine. 2019;57:18-29.
50. Pascut D, Cavalletto L, Pratama MY, Bresolin S, Trentin L, Basso G, et al. Serum miRNA Are Promising Biomarkers for the Detection of Early Hepatocellular Carcinoma after Treatment with Direct-Acting Antivirals. Cancers (Basel). 2019;11(11).
51. Ju B, Nie Y, Yang X, Wang X, Li F, Wang M, et al. miR-193a/b-3p relieves hepatic fibrosis and restrains proliferation and activation of hepatic stellate cells. J Cell Mol Med. 2019;23(6):3824-32.
52. Hyun J, Wang S, Kim J, Rao KM, Park SY, Chung I, et al. MicroRNA-378 limits activation of hepatic stellate cells and liver fibrosis by suppressing Gli3 expression. Nat Commun. 2016;7:10993.
53. Huang RS, Gamazon ER, Ziliak D, Wen Y, Im HK, Zhang W, et al. Population differences in microRNA expression and biological implications. RNA Biol. 2011;8(4):692-701.
54. Zeb I, Li D, Nasir K, Katz R, Larijani VN, Budoff MJ. Computed tomography scans in the evaluation of fatty liver disease in a population based study: the multi-ethnic study of atherosclerosis. Acad Radiol. 2012;19(7):811-8.

Chapter 2.3

Plasma microRNAs signature of alcohol consumption and their links to fatty liver

Irma Karabegović, **Yasir Abozaid*,** Silvana C.E. Maas*, Jeremy Labrecque, Daniel Bos, Robert J. De Knegt, M. Arfan Ikram, Trudy Voortman, Mohsen Ghanbari

Journal of Nutrition 2023 Jan 14;152(12):2677-2688

***Authors denotes equal contribution**

Abstract

Background: MicroRNAs (miRNAs) represent a class of non-coding RNAs that regulate gene expression and are implicated in the pathogenesis of different diseases. Alcohol consumption might affect the expression of miRNAs, which in turn could play a role in risk of diseases.

Objective: We investigated whether plasma concentrations of miRNAs are altered by alcohol consumption. Given the existing evidence showing the link between alcohol and liver diseases, we further explored the extent to which these associations are mediated by miRNAs.

Design: Profiling of plasma miRNAs was conducted using HTG EdgeSeq miRNA Whole Transcriptome Assay in 1933 participants of the Rotterdam Study. Linear regression was implemented to explore the link between alcohol consumption (glasses/day) and miRNAs levels, adjusted for age, sex, cohort, BMI, and smoking. Sensitivity analysis for alcohol-categories (non-, light-, and heavy-drinkers) was performed, where light drinkers-corresponded to 0-2 glasses/day in men; 0-1 glasses/day in women and heavy->2 glasses/day in men; >1 glasses/day in women. Moreover, we utilized the alcohol-associated miRNAs to explore their potential mediatory role between alcohol consumption and liver-related traits. Finally, we retrieved putative target genes of identified miRNAs to gain an understanding of the molecular pathways concerning alcohol consumption.

Results: Plasma levels of miR-193b-3p, miR-122-5p, miR-3937, and miR-4507 were significantly associated with alcohol consumption surpassing the Bonferroni-corrected *P-value*$< 8.46\times10^{-5}$. The top significant association was observed for miR-193b-3p (β=0.087, *P-value*$= 2.90\times10^{-5}$). Furthermore, a potential mediatory role of miR-3937 and miR-122-5p was observed between alcohol consumption and liver traits. Pathway analysis of putative target genes revealed involvement in biological regulation and the cellular processes.

Conclusions: This study indicates that alcohol consumption is associated with plasma concentrations of four miRNAs. We outline a potential mediatory role of two alcohol-associated miRNAs (miR-3937 and miR-122-5p), laying the groundwork for further exploration of miRNAs as potential mediators between lifestyle factors and disease development.

Introduction

Alcohol consumption is a modifiable lifestyle factor and a leading risk factor for the global burden of many diseases. Given its widespread nature, alcohol has been estimated to contribute to 2.7 million deaths and 4% of the global disease burden annually.[1] High alcohol intake has been associated with an increased risk of stroke, peripheral artery disease,[2] liver diseases,[3-6] various cancers,[7-10] overall all-cause mortality,[11] and many other diseases.[12] Although numerous molecular mechanisms have been postulated to explain the link between alcohol consumption and the risk of various diseases, this complex etiology remains to be explored.[13-15] The liver is the primary organ for metabolizing and detoxification of alcohol,[16] while excessive alcohol consumption can have a severe impact on liver health, including fatty liver, alcoholic hepatitis, and cirrhosis.[17] In addition, only 10-20% of chronic alcohol consumers will progress to advanced alcoholic liver disease (ALD).[18] The exact molecular mechanisms involved in alcohol-related liver traits and diseases are still not fully elucidated.[19, 20] Behavioural factors, including alcohol consumption, have been linked with epigenetic markers,[21-23] while these epigenetic markers have also been linked to several diseases.[20] Epigenetic mechanisms include DNA methylation, histone protein modifications, and RNA-mediated regulation by non-coding RNAs.[20, 24]

MicroRNAs (miRNAs) are small non-coding RNA molecules (around 22 nucleotides in length) that regulate gene expression at the post-transcriptional level. As such, miRNAs are estimated to regulate the expression of more than half of the protein-coding genes in our genome.[25] They are considered as a type of epigenetic regulation whose mechanism of action relies on the degradation of messenger (m)RNAs and translational repression.[26] An extensive body of research has demonstrated that dysregulation of miRNAs is associated with disease risk.[27-32] Moreover, recent studies have indicated an influence of modifiable lifestyle factors (such as smoking and diet) on miRNA expression levels.[33] Two before-after studies with small sample sizes (n=16-18)[34, 35] showed differential expression of miRNAs following exposure to alcohol consumption, including miR-122-5p, a highly expressed liver miRNA. However, limited studies were conducted to explore the association between expression levels of miRNAs and alcohol consumption in larger sample sizes.[33] As identifying alcohol-associated changes in miRNAs expression might help to elucidate the mechanism of action between alcohol consumption and health outcomes, it is of crucial importance to explore this niche. In this study, we aimed to investigate the association of plasma miRNAs with alcohol consumption and to explore whether there is a mediating effect for the alcohol-associated miRNAs in the cross-sectional association of alcohol consumption with liver function and disease, using data from the large population-based prospective Rotterdam Study cohort.[36]

Methods

Study population

This study was conducted in the Rotterdam Study, which is an ongoing prospective population-based cohort study. In brief, the Rotterdam Study consists of four sub-cohorts. The first subcohort (RS-I) was initiated in 1990 with individuals ≥55 years of age (n=7983). The study was extended by including a second sub-cohort (RS-II) in 2000 (n=3011, ≥ 55 years of age), a third sub-cohort (RS-III) in 2006 (n=3932, ≥ 45 years of age), and the most recent fourth sub-cohort (RS-IV) in 2016 (n=3005, ≥ 40 years of age). In addition to these baseline examinations, the participants were re-examined during follow-up every 3 to 5 years. More in-depth details regarding the design of the Rotterdam Study can be found elsewhere.[36]

For the present study, 1000 participants were included from the fourth visit of Rotterdam Study-I (RS-I-4) and 1000 participants from the second visit of Rotterdam Study-II (RS-II-2), for whom we had miRNA expression data measured (in total n=2000). These visits of the Rotterdam Study occurred between 2002 and 2005. From the 2000 unique individuals, 1 participant was excluded due to missing profiling data for all miRNAs, while 66 were excluded due to missing data on alcohol consumption. In total, 1933 non-overlapping participants were included in our analysis. The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC and by the Dutch Ministry of Health, Welfare, and Sport.[36]

MiRNA expression profiling

Blood samples were collected in EDTA treated containers and centrifuged, followed by plasma aliquotation, and stored at -80°C, according to the standard procedures. Plasma samples were then used for miRNA expression profiling using the HTG EdgeSeq miRNA Whole Transcriptome Assay (WTA) (HTG Molecular Diagnostics, Tuscon, AZ, USA). The WTA measured the expression of 2083 human miRNAs using the Illumina NextSeq sequencer (Illumina, San Diego, CA, USA). The assay characterizes miRNA expression patterns and hereby measured the expression of 13 housekeeping genes- providing flexibility in data analysis and normalization. The miRNA expression quantification was based on counts per million (CPM), which were log2 transformed and used as standardization, adjusting for total reads within each sample. Furthermore, the miRNAs showing log2 CPM <1.0 were referred to as low expressed and the well-expressed miRNAs were defined as those with >50% of values above the lower limit of quantification (LLOQ), resulting in a total of 591 miRNAs, which were used in our analysis. The LLOQ was used for the selection of well-expressed miRNAs (n=591), which was based on a monotonic decreasing spline curve fit between the means and standard deviations of all miRNAs in the whole study participants.

Assessment of alcohol consumption

Participants were administered interviews at home by research assistants, where they were asked about their alcohol consumption. The first question asked whether participants ever drank alcohol. If the answer was confirmative, it was later followed by more extensive questions on the type of alcohol (e.g., beer, red wine, white wine, moderately strong spirits such as Campari, Martini, sherry, and strong spirits such as rum, brandy, whisky) and frequency of consumption per week. This information was collected and used to calculate the average alcohol consumption in glasses/day. The glasses/day information could be used to estimate grams of alcohol, assuming that one glass of alcohol would roughly correspond to 10 grams of alcohol.[37] As our study population also included a percentage of current non-drinkers of alcohol (n=307, 15.88%), the alcohol consumption variable was right-skewed. To satisfy the assumption of normality of residuals in linear regression, we applied transformation of (log(glasses/day +1)), according to the approach reported by Liu et al.[38] Furthermore, alcohol consumption was categorized in non-drinkers (glasses/day=0), light drinkers (0<-≤2 glasses/day in men and 0<-≤1 glasses/day in women), and heavy drinkers (>2 glasses/day in men and >1 glasses/day in women).

Assessment of covariates

Questionnaires were used to assess the participant's age, sex, and smoking status (classified as current, former, and never smokers). Furthermore, the height and weight of participants were measured with the participants standing without heavy garments or shoes. Body mass index (BMI) was computed as weight in kilograms divided by height in meters squared.

Assessment of fatty liver and hepatic steatosis using CT-scan and ultrasound

Multidetector CT scanner (Somatom Sensation 16 or 64, Siemens, Forchheim, Germany) was acquired as part of a larger project on vascular calcification. For the current project, the ECG-gated, non-contrast cardiac scan was used to assess the density of the liver, as a proxy for fatty liver disease. Detailed imaging parameters are described in detail elsewhere.[39] We assessed the density of the liver using a standardized strategy that included drawing three circular regions of interest (ROIs) in liver tissue in which the mean liver attenuation (LA) was calculated.[40] The ROIs were carefully chosen to include solely liver tissue (avoiding disruptive tissue such as focal lesions, cysts, or large blood vessels). Next, we determined the mean Hounsfield unit (HU) value from the retrieved three measurements as an indicator of the total liver fat amount. As the amount of liver fat is increased, the measured liver attenuation (LA) is decreased; therefore, a lower LA indicates a higher risk of fatty liver. All measurements were computed using Philips iSite Enterprise software (Royal Philips Electronics N.V. 2006), described in depth elsewhere.[41] In addition, we transformed liver fat (A) using exponential values (B) [$B=A^{3.5}/10000$] as it was left-skewed.[41]

Beyond the CT assessment, hepatic steatosis was determined by using the abdominal ultrasound data, generated via Hitachi HI VISION 900 by an experienced and certified technician (**Supplemental Table**

1). Steatosis was diagnosed by dichotomizing the data into the presence of hyperechogenic liver parenchyma, as reported previously.[42] More details on liver steatosis and non-alcoholic fatty liver disease (NAFLD) within the Rotterdam study can be found elsewhere.[43]

Measuring liver enzymes

Serum gamma-glutamyltransferase (GGT) and alkaline phosphatase (ALP) levels were determined within 2 weeks of collecting and stored with non-fasting and fasting blood samples at -20°C. Kit Merck Diagnostica (Merck, Whitehouse Station, NJ, USA) was used on Elan Autoanalyzer (Merck). Furthermore, considering local cut-offs, elevated GGT was defined as >34 U/L for women and >49 U/L for men, while ALP was considered elevated at >97 U/L for women and >114 U/L for men, more details can be found elsewhere.[44] To satisfy the assumption of normality of residuals, as GGT and ALP were right-skewed, we applied log transformation.

Statistical analyses

Alcohol consumption in association with alterations in miRNAs levels

Multivariable linear regression models were implemented to explore the association between alcohol consumption as the main exposure (log(glasses/day+1) and plasma miRNA levels (log2 CPM) as the outcome. For a more detailed overview of the inclusion criteria and the analysis workflow, see **Figure 1**. We tested three different models. The first model was adjusted for age, sex, and cohort, the second model was further adjusted for BMI, while in the final model we additionally adjusted for smoking status. The main results were reported from the fully adjusted model. The Bonferroni-corrected *P-value* threshold $<0.05/591=8.46\times10^{-5}$ (after adjustment based on the number of miRNAs tested) was set for our hypothesis-free approach. The assumptions of linear regression analysis including normality of residuals, normality of random effects, multicollinearity, linear relationship, and homogeneity of variance were assessed using the “performance” package in R.

Furthermore, for the alcohol-associated miRNAs, we performed a sensitivity analysis, where we treated alcohol exposure as a categorical variable. The non-drinker category was included as the reference group, where it was compared to the light- and heavy-drinkers.

Moreover, as alcohol consumption might have sex-specific differences due to differential drinking patterns[45] or alcohol metabolism,[46] we performed a sex-stratified analysis to explore potential changes in alcohol-associated miRNAs.

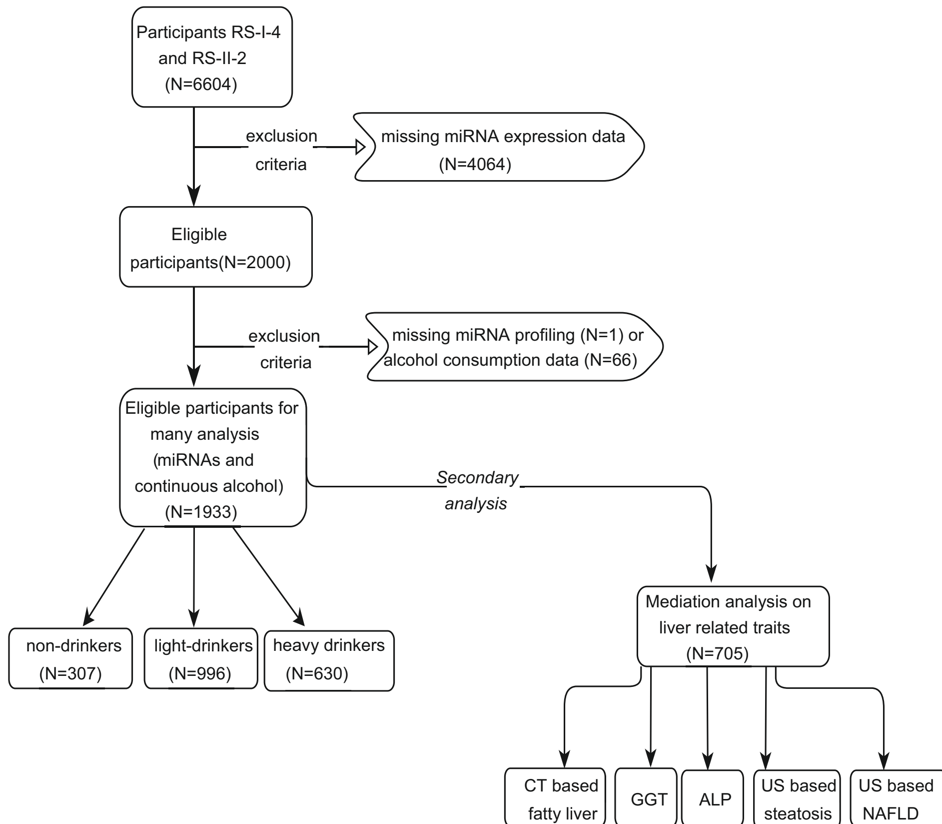


Figure 1. Overview of the study design. An overview of the flowchart summarizing sample sizes for the different analysis. The main analysis investigating the association between alcohol consumption and miRNA expression was performed on participants from RS-I-4 and RS-II-2 within Rotterdam Study, who had data available on miRNA levels and alcohol consumption (N=1933). Non-drinkers: 0 glasses/day, light drinkers: 0-2 glasses/day in men and 0-1 glasses/day in women, heavy drinkers: >2 glasses/day in men and >1 glasses/day in women. Abbreviations: miRNAs: microRNAs, ALP: Alkaline phosphatase, CT: computed tomography, GGT: gamma-glutamyl transferase, NAFLD: non-alcoholic fatty liver disease, RS: Rotterdam Study.

Mediation analyses with liver traits

In our secondary objective, we performed mediation analyses, where our exposure was always alcohol consumption, the mediators were miRNAs associated with both alcohol consumption and liver disease, and the outcomes were liver-related traits, including CT-based liver attenuation, liver enzymes (GGT and ALP), ultrasound-based hepatic steatosis, and NAFLD. For the continuous outcomes (CT-based liver attenuation, GGT, and ALP), we used linear regression, while for binary outcomes (steatosis and NAFLD), we used logistic regression. The selection criteria of potential mediators was based on a seminal paper by Baron and Kenny, stating that in order to define a variable as a mediator, there should be a significant relationship between the mediator (miRNAs) and the outcome (liver-related traits).[47] In that line, three of the alcohol-associated miRNAs (miR-193b-3p, miR-122-5p, and miR-3937) were previously associated with liver-related traits within the Rotterdam Study,[48] hence they were included as mediators in our analyses. Our mediation analyses were implemented using two-way decomposition assessing the direct and indirect effects, meaning that the overall effect of alcohol consumption on liver-related traits with miRNAs as mediators were decomposed into two main components: 1) the direct effect of alcohol consumption on liver-related traits (i.e., liver attenuation, GGT, ALP, steatosis, and NAFLD) in the absence of mediator (i.e. miR-193b-3p, miR-122-5p, or miR-3937) and 2) indirect effect. Models were adjusted for the same confounders as in the main analysis, including age, sex, cohort, BMI, and smoking status. In addition, we also assessed if there was a potential interaction effect between the exposure and the mediator. For the models that showed the presence of interaction effect ($P<0.05$), we implemented exposure and mediator interaction terms in mediation analyses. The conceptual diagram depicting the relationship between the exposure (alcohol consumption), outcomes (liver-related traits), and mediators (miRNAs) is depicted in **Figure 2**. All the confounders included in the statistical analyses were obtained at the same time point as miRNA expression data, as well as data on CT-based liver attenuation and liver enzymes (RS-I-4 and RS-II-2), while the data based on ultrasound (steatosis and NAFLD) are collected during a follow-up visit and analyzed in the longitudinal setting. We used the "mediate" function from the mediation package[49] to obtain the Average Causal Mediation Effect (ACME), Average Direct Effect (ADE), the total effect, and the proportion mediated per model. Mediation results were based on quasi-Bayesian approximation with 1000 simulations.

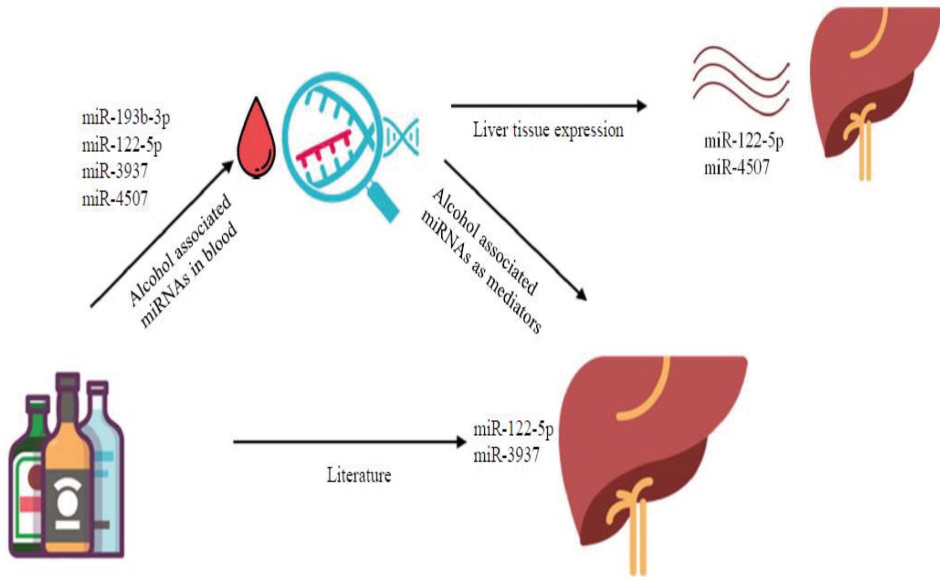


Figure 2. Conceptual diagram illustrating the relationship between alcohol consumption and liver health, and the potential mediatory role of alcohol-associated miRNAs. The conceptual diagram that depicts the relationship between exposure (alcohol consumption), outcome (liver related traits including CT-based liver attenuation, liver enzymes (GGT and ALP), and ultrasound-based hepatic steatosis and NAFLD) and the mediators (miRNA expression level). CT, computed tomography; miRNA, microRNA.

Furthermore, the mediation analyses performed assumes no unmeasured confounding. As such, we included bias analyses using the "medsens" function from the mediation package[49] to determine the ρ at which ACME is 0 per model. A value of ρ close to 0 reflects that the assumption of no additional unmeasured confounding is sensitive to violations and likely does not hold. We implemented recommended AGReMa Statement guidelines when reporting the results,[50] including reporting baseline characteristics as well as potential confounders in **Supplemental Table 1**.

Mendelian Randomization

We investigated the causal relationship between the alcohol-associated miRNAs and liver-related traits by utilizing the two-sample mendelian randomization (MR) approach. Instrumental variables (IVs) for each of the alcohol-associated miRNAs were extracted using different resources, including a genome-wide association studies (GWAS) conducted in the Rotterdam Study (n=1687), (data not shown), and publicly available GWASs on miRNAs.[51-53] We identified 10 cis-miR-eQTLs for miR-193b-3p,[53] while miR-193b-3p and miR-122-5p only had trans-miR-eQTLs.[51] The trans-eQTLs were excluded from our further analysis due to the assumption of no horizontal pleiotropy.[53, 54] Next, the cis-miR-eQTLs of miR-193b-3p were pruned at $R^2<0.01$, to remove correlated SNPs. This left us with a single SNP (rs30227) to be used as an IV. IVs on liver traits were extracted from the IEU GWAS database release (https://gwas.mrcieu.ac.uk/), where we included the following traits: liver fat percentage,[55] NAFLD (https://finngen.gitbook.io/documentation/) and liver enzymes.[56] MR was performed using the "TwoSampleMR" package in R, by implementing the Wald ratio as a single SNP was available to be used as IV.

Our analyses were performed using R software, version (V) 4.1.1. (R Core Team, 2021). Moreover we used following packages for different utilities within R, including rio (V 0.5.27)[57] for data importing/exporting, tidyverse (V 1.3.1),[58] janitor (V 2.1.0),[59] and lubridate (V 1.7.10)[60] for data manipulation and handling, stats (V 4.1.1),[61] broom (V 0.7.9),[62] performance (V 0.9.1),[63] and purrr (V 0.3.4)[64] for modelling, ggplot2 (V 3.3.5)[65] for visualization, mediation (V 4.5.0)[49] for mediation analyses, TwoSampleMR (V 0.5.6) for MR analysis,[66] and tableone (V 0.13.0)[67] for clinical characteristics.

In silico analyses of alcohol-associated miRNAs

We explored if the alcohol-associated miRNAs are expressed in the liver by using the Human miRNA tissue atlas (https://ccb-web.cs.uni-saarland.de/tissueatlas).[68, 69] More details regarding the tissue specificity index can be found elsewhere.[69] As an additional analysis, we utilized three universally used miRNA target gene prediction databases: TargetScan,[70] miRTarBase,[71] and miRDB[72] to identify putative target genes of the alcohol-associated miRNAs. Applying a cut-off based on a total context score of ≤ -0.60, we selected target genes using TargetScan, while for miRDB we applied selection on target score ≥ 60. The scores of the two databases are explained in detail elsewhere.[70, 73] In addition, we used miRTarBase[71] to select the target genes that were proven by experimental validation methods, such as reporter assay, qPCR, and western blot. We focused on genes that were available in either two out of the three above-mentioned databases. Furthermore, we investigated if any of these predicted target genes have been associated previously with alcohol consumption and/or alcohol use disorder by either a review, a genome-wide association study, an epigenome-wide association study, or a transcriptome-wide association study on alcohol consumption.[33, 38, 74, 75] Finally, the putative target genes we obtained from the analysis described above were used for gene ontology analysis to explore the biological processes

these genes might be involved in,[76] by utilizing the publically available web tool PANTHER (http://www.pantherdb.org/).[77]

Results

Characteristics of the study population (N=1933) are presented in **Table 1**. The mean (SD) age of the study population was 71.62 (±7.5 years), with a BMI of 27.65 (±4.13) kg/m^2, and the median (IQR) alcohol consumption of 0.71 glasses/day (0.07-2.00). Out of the 1933 individuals, 56.8% were women.

Table 1. Participant characteristics of the study population from RS-I-4 and RS-II-2 within the Rotterdam Study cohort

Variable	N= 1933
Age (years)	71.62 (±7.5)
Female sex (%)	1098 (56.8)
BMI (kg/m^2)	27.65 (±4.13)
Smoking	
Current (%)	260 (13.5)
Former (%)	1069 (55.3)
Never (%)	604 (31.2)
Alcohol (glasses/day)	0.71 (0.07-2.00)
Non-drinkers (%)[a]	307 (15.9)
Light-drinkers (%)[b]	996 (51.5)
Heavy drinkers (%)[c]	630 (32.6)

Variables are reported in mean (standard deviation (SD)) for continuous data and numbers (percentages) for categorical data, apart from alcohol (glasses/day) which is reported in median (IQR) due to the distribution of the variable. Alcohol categories were defined as follows: [a]non-drinkers:glasses/day=0, [b]light drinkers:0<-≤2 glasses/day in men and 0<-≤1 glasses/day in women, and [c]heavy drinkers:>2 glasses/day in men and >1 glasses/day in women.

Plasma miRNAs associated with alcohol consumption

We found four miRNAs to be significantly associated with alcohol consumption (as continuous variable) surpassing the significance threshold (*P-value* $< 8.5\times10^{-5}$). Of these, miR-193b-3p, miR-122-5p, and miR-3937 showed a positive association, while miR-4507 was inversely associated with alcohol consumption, **Table 2** and **Figure 3**. The results of our sensitivity analysis, where we explored alcohol consumption as categorical exposure, are presented in **Table 2** and **Figure 4**. The categorization of the alcohol consumption reduced the power, yet the association of miR-3937 remained statistically significant for

heavy drinkers (*P-value*= 3.02×10^{-6}) in comparison to the non-drinkers. In addition, mean expression of miR-3937 in light-drinkers increased by 0.142 compared to the mean of non-drinkers in the reference category, while it almost doubled (0.273) in heavy-drinkers. In contrast, the mean expression of miR-4507 in light-drinkers decreased by -0.029 in comparison to the mean of non-drinkers (reference), while for heavy-drinkers this drops by -0.155 (**Table 2** and **Figure 4**).

Furthermore, in the sex-stratified analysis, we observed that all the effect size estimates were in the same direction. However, most of the alcohol-associated miRNAs had stronger effect size estimates in men, except miR-4507 which showed more decrease in women in comparison to men (**Supplemental Table 2**).

Table 2. Association between MiRNAs and alcohol consumption as continuous variable (glasses/day) and categorical variable (never drinkers vs. light or heavy drinkers)

miRNA ID	Alcohol in glasses/day (N=1933)			[a]Never drinkers (N=307) VS. [b]light (N=996) or [c]heavy drinkers (N=630)			
	β	SE	P-value	Category	β	SE	P-value
miR-193b-3p	0.087	0.020	2.90×10^{-5}	light drinkers	0.026	0.031	4.07×10^{-1}
				heavy drinkers	0.086	0.033	1.02×10^{-1}
miR-122-5p	0.151	0.037	4.31×10^{-5}	light drinkers	0.015	0.056	7.77×10^{-1}
				heavy drinkers	0.125	0.060	3.75×10^{-2}
miR-3937	0.145	0.036	5.71×10^{-5}	light drinkers	0.142	0.054	8.64×10^{-3}
				heavy drinkers	0.273	0.058	3.02×10^{-6}
miR-4507	-0.110	0.027	8.36×10^{-5}	light drinkers	-0.029	0.042	4.85×10^{-1}
				heavy drinkers	-0.155	0.045	6.26×10^{-4}

On the left side of the table are the results from the linear regression with continuous data on alcohol consumption as main exposure transformed to (log(glasses/day +1)), where the analyses were adjusted for age, sex, cohort, BMI, and smoking status. Right side of the table depicts alcohol consumption stratified to categorical variable (where non-drinkers were treated as a reference) and used as main exposure for linear regression analysis, adjusted for age, sex, cohort, BMI, and smoking status. In all the analysis presented, miRNA expression levels were outcome variables, effect size reported are beta coefficients from regression analysis. Abbreviations: β:beta coefficient, SE: standard error, [a]non-drinkers: 0 glasses/day, [b]light drinkers: 0-2 glasses/day in men and 0-1 glasses/day in women, [c]heavy drinkers: >2 glasses/day in men and >1 glasses/day in women. miR; miRNA; miRNA, microRNA.

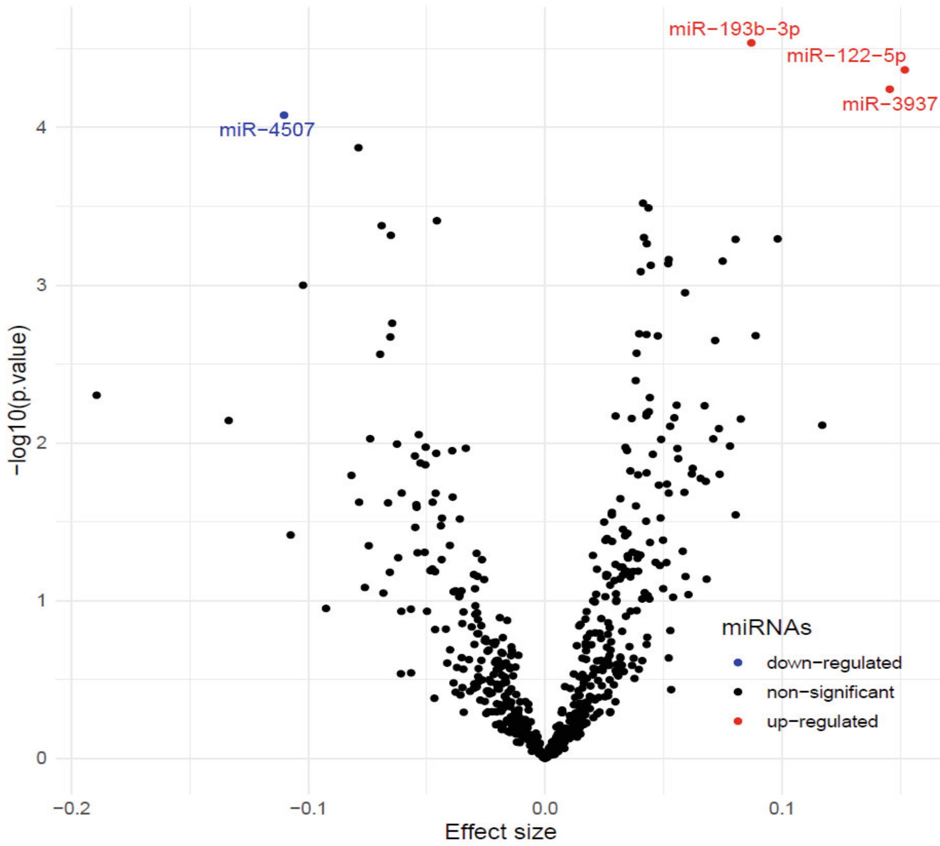


Figure 3. Plasma miRNAs associated with alcohol consumption in glasses/day(N=1933). The volcano plot depicts the measure of effect size versus magnitude of significance for the linear regression model testing the association between of miRNA expression levels and alcohol consumption, adjusted for age, sex, cohort, BMI and smoking. The dots indicate each tested miRNAs and represents the beta coefficients obtained from each linear regression analysis. Red dots indicates positively associated miRNAs, while the blue dots indicate negatively associated miRNAs, and black dots represent miRNAs that were not significantly associated. miR; miRNA; miRNA, microRNA.

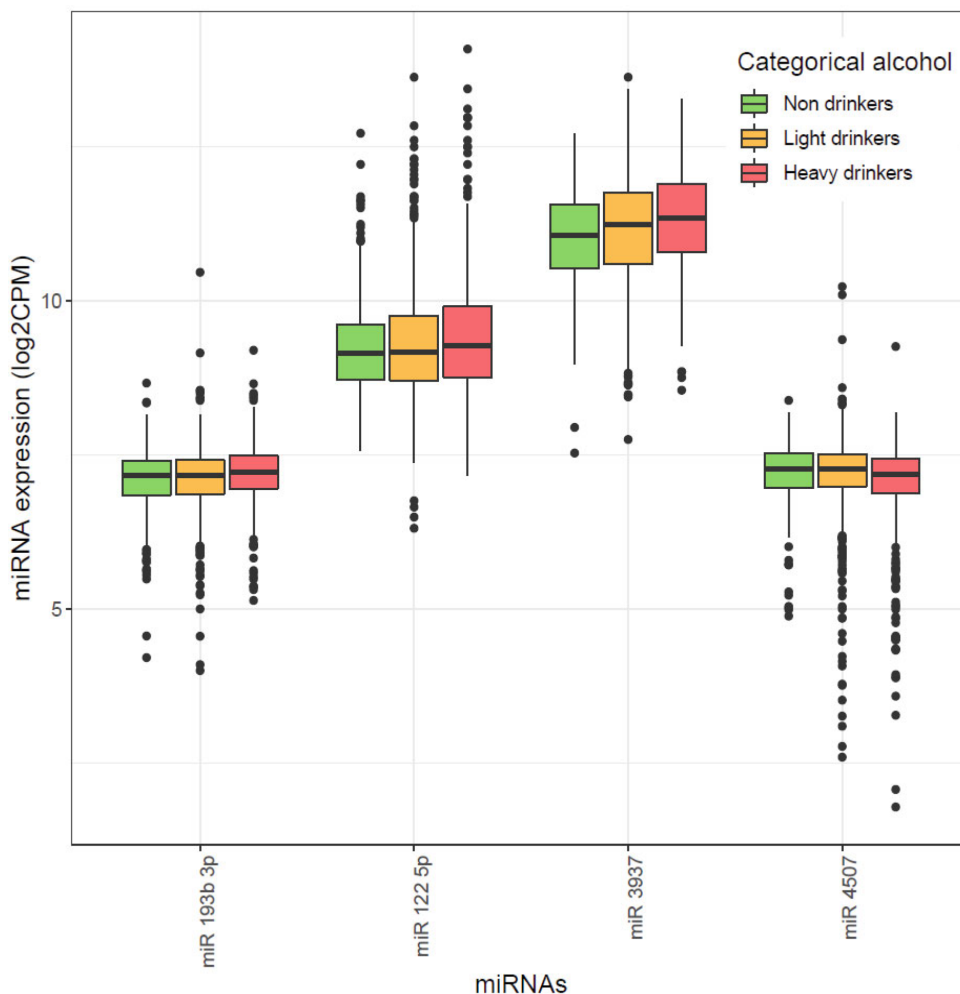


Figure 4. Distribution of the significantly associated miRNAs in the three alcohol consumption categories. In this figure, X-axis depicts significantly associated miRNAs with alcohol consumption, Y-axis depicts miRNA expression levels in log2 CPM, displaying a boxplot of median miRNA expression levels. The horizontal line within each boxplot represents the median, while the whiskers depict minimum (corresponding to Q1-1.5*IQR) and maximum value (corresponding to Q3+1.5*IQR) in the data. Different colors indicate different categories of alcohol consumption where non-drinkers are green (glasses/day=0, N=307), light-drinkers yellow (0<-≤2 glasses/day in men and 0<-≤1 glasses/day in women, N=996) and heavy-drinkers are red (>2 glasses/day in men and >1 glasses/day in women, N=630). Abbreviations: miRNA: microRNA, CPM: counts per million.

Mediation analyses for alcohol consumption, miRNA expression, and liver disease

We tested the potential mediatory role of three miRNAs previously shown to be associated with fatty liver disease (miR-193b-3p, miR-122-5p, and miR-3937)[47, 48] in the association between alcohol and liver function and disease. The descriptive characteristics of this subset of participants (N=705) are presented in **Supplemental Table 1**. We performed mediated interaction terms for all the models, of which one model suggested an interaction effect between mediator and exposure, miR-122-5p and alcohol on ALP (P=0.04) (**Supplemental Table 3)**. For this model, we included interaction terms in the main analysis, while for the other models we did not include any interaction terms (**Table 3**). Out of all the mediation analyses performed, we identified a mediatory role of miR-3937 in the association between alcohol and CT-based fatty liver as well as GGT, while miR-122-5p showed a mediatory role between alcohol and CT-based fatty liver disease, GGT, and US-based steatosis (**Table 3**). We tested the bias analysis of violating the assumption of unmeasured confounding in the mediation analyses. We conducted ρ at which ACME is 0, where we obtained ρ's in the range between -0.1 and 0.4 (**Table 3**). A value of ρ close to 0 indicates that the assumption of unmeasured confounding was sensitive to the violation.

Table 3. Mediation analysis of three alcohol-associated miRNAs with alcohol consumption and liver-related traits (CT-based liver attenuation, liver enzymes (GGT and ALP), and ultrasound-based hepatic steatosis and NAFLD) in Rotterdam Study participants (N=705)

miRNAs	Liver-related traits (N=705)	ACME (95%CI)	ADE (95%CI)	Total effect (95%CI)	*Prop. Med. (95%CI)	ρ at which ACME is 0
miR-3937	CT - based fatty liver	1.630 (0.114; 3.490)	-24.80 (-34.72; -15.66)	-23.17 (-33.25; -14.14)	-0.070 (-0.187; -0.004)	0.1
	GGT	-0.009 (-0.024; -0.0001)	0.287 (0.209; 0.375)	0.277 (0.199; 0.362)	-0.035 (-0.094; -0.0006)	-0.1
	ALP	-0.003 (-0.009; 0.00005)	-0.028 (-0.065; 0.009)	-0.031 (-0.068; 0.005)	0.117 (-0.351; 0.852)	-0.1
	US - based steatosis	-0.006 (-0.017;-0.0001)	0.063 (-0.0009;0.127)	0.057 (-0.008;0.121)	-0.108 (-0.779;0.454)	-0.1
	US - based NAFLD	-0.004 (-0.014;0.0006)	0.255 (0.189;0.317)	0.251 (0.184;0.310)	-0.019 (-0.061;0.002)	-0.1
miR-122-5p	CT - based fatty liver	-1. 394 (-3.115; -0.194)	-21.77 (-31.39; -10.92)	-23.17 (-32.77; -11.88)	0.060 (0.007; 0. 166)	-0.1
	GGT	0.036 (0.004; 0.071)	0.241 (0.166; 0.314)	0.277 (0.197; 0.360)	0.131 (0.017; 0.238)	0.4
	ALP	0.002 (-0.0008; 0.006)	-0.032 (-0.069; 0.009)	-0.028 (-0.067; 0.073)	-0.080 (-0.875; 0.644)	0
	US - based steatosis	0.008 (0.001; 0.020)	0.047 (-0.016; 0.109)	0.056 (-0.006; 0.117)	0.155 (-0.702; 1.188)	0.1
	US - based NAFLD	0.005 (-0.0001; 0.014)	0.244 (0.175; 0.305)	0.250 (0.183; 0.310)	0.022 (0.000; 0.062)	0.1
miR-193b-3p	CT - based fatty liver	-1.116 (-2.842; 0.344)	-22.05 (-31.35; -12.09)	-23.17 (-32.39; -13.49)	0.048 (-0.016; 0.144)	-0.1
	GGT	0.011 (-0.004; 0.031)	0.265 (0.188; 0.343)	0.277 (0.194; 0.363)	0.042 (-0.016; 0.114)	0.2
	ALP	-0.001 (-0.004; 0.0008)	-0.030 (-0.065; 0.004)	-0.031 (-0.067; 0.004)	0.038 (-0.079; 0.348)	0
	US - based steatosis	0.004 (-0.001; 0.012)	0.053 (-0.008; 0.117)	0.058 (-0.004; 0.121)	0.071 (-0.194; 0.467)	0.1
	US - based NAFLD	0.003 (-0.001; 0.011)	0.246 (0.177; 0.309)	0.250 (0.181; 0.314)	0.013 (-0.005; 0.044)	0.1

The table depicts results from mediation analysis where alcohol consumption was treated as the exposure, the outcomes were liver-related traits including CT-based fatty liver, liver enzymes (GGT and ALP), US-based steatosis and NAFLD, and miRNAs were the mediators. ACME reflects proportion of alcohol exposure on liver-related traits mediated through miRNA of interest, while ADE reflects the direct effect of alcohol consumption on liver-related traits. Prop. Med. reflects proportion mediated which cannot be calculated when the indirect and direct effects are in opposite directions, ρ at which ACME is 0, depicting how sensitive the tested model is to violating of unmeasured confounding. Abbreviations: β: beta coefficient, SE: standard error, ACME: Average Causal Mediation Effect, ADE: Average Direct Effect.

Mendelian randomization

We investigated the causal relationship between the alcohol-associated miR-193-5p and liver fat percentage, NAFLD, and liver enzymes (https://finngen.gitbook.io/documentation/).[51-53, 55, 56, 78] The results of the MR analysis are presented in **Supplemental Table 4**. There was no statistical evidence for a causal relationship between alcohol-associated miRNAs and liver-related traits tested.

Liver expression and target genes of alcohol-associated miRNAs

Publicly available tools were utilized to assess the expression of alcohol-associated miRNAs across a wide range of tissues (**Supplemental Table 5**). Among these, the miR-122-5p had the highest tissue specificity index (TSI) of 0.97 (where a higher score indicates miRNA is expressed in a single tissue) (**Supplemental Table 5**). In addition, miR-122-5p and miR-4507 displayed the highest expression in the liver tissue, while miR-193b-3p showed the highest expression in muscle and miR-4507 in the stomach.

Potential target genes of the alcohol-associated miRNAs are shown in **Supplemental Table 6**. Only miR-193b-3p and miR-122-5p had validated target genes by experimental methods as reported in miRTarBase (**Supplemental Table 6**).[71] By performing a literature review, we identified that several putative target genes of miR-193b-3p, miR-122-5p, and miR-3937 have been previously associated with alcohol-related traits (**Supplemental Table 7**). These include *FLI* and *SMAD3*, both putative targets of miR-193b-3p, which were previously identified in an epigenome-wide association study (EWAS) on alcohol consumption.[38] In addition, putative target genes of miR-122-5p (*XPO6* and *SLC7A11*) were identified in the same EWAS study, along with *C7orf50* a putative target gene of miR-3937, **Supplemental Table 7**.[38] Furthermore, *DCLK2*, one of the miR-3937 putative target genes, was previously associated in a trans-ethnic genome-wide association analysis of Alcohol Use Disorder Identification Test-Consumption (AUDIT) (rs4423856, *P-value*=1.48×10^{-8}).[74] Also, miR-122-5p putative target gene- *RAC1*, was previously associated with alcohol use during pregnancy.[33, 79] While *FOXP1*, another putative target gene of miR-122-5p, was previously reported in a transcriptome-wide association study on alcohol intake frequency (http://twas-hub.org/traits/).[80, 81]

Our biological processes' overrepresentation analysis with the putative target genes of alcohol-associated miRNAs is presented in **Supplemental Table 8.** The top pathways for the biological process were the following: biological regulation, biological process, and transmembrane receptor protein serine/threonine kinase signaling pathway (**Supplemental Table 8**).

Discussion

In this study, we investigated the link between plasma miRNA expression and alcohol consumption in a population-based setting. We identified plasma levels of four miRNAs to be significantly associated with alcohol consumption, including three miRNAs positively and one miRNA inversely associated. Among these, we observed a potential mediatory role of miR-122-5p and miR-3937 between alcohol consumption and liver-related traits. The identified miRNAs lay the groundwork for further investigation of miRNAs as potential mediators between modifiable lifestyle factors and disease risk.

MiRNAs could modulate gene expression in response to external influences, such as lifestyle factors (e.g. smoking, alcohol consumption, and diet).[33] It has been shown that miRNA expression was altered following exposure to maternal alcohol consumption during human embryogenesis.[82-84] Similarly, Lewohl et al.[23] have identified differential expression of 35 miRNAs in human postmortem brains between 14 alcoholics and 13 controls. However, most of the previous studies exploring the association between alcohol consumption and miRNA expression were performed on animal models.[85-87] In addition, past research has been conducted either on a subset of miRNA or had relatively modest sample sizes (with the largest sample size reported N=68).[33] Our study benefits from a greater statistical power to detect significant associations between miRNAs and alcohol consumption due to the larger sample size embedded in the population-based Rotterdam Study cohort. In addition, the RNA-sequencing method was used to measure a large number of miRNAs, enabling us to investigate a more comprehensive miRNA landscape.[88]

The most prominent association with alcohol consumption was observed for miR-193b-3p. Previous studies have identified miR-193 as a regulator of *ALDH2* gene expression across different species,[89] where the *ALDH2* gene encodes alcohol aldehyde dehydrogenase 2, a key enzyme in alcohol metabolism.[90] This miRNA has several other putative target genes, including *FLI* and *SMAD3*, previously identified in an EWAS on alcohol consumption.[38] The same study overlapped with other putative target genes of our newly identified alcohol miRNAs, including *XPO6* and *SLC7A11* of miR-122-5p and *C7orf50*- of miR-3937.[38] In addition, miR-122-5p expression has been shown to increase with moderate ethanol consumption in healthy individuals.[35] In line with this, two target genes of miR-122-5p were linked with alcohol consumption, including *RAC1* with alcohol use during pregnancy[33, 79] and *FOXP1*- in a transcriptome-wide association study on alcohol intake frequency (http://twas-hub.org/traits/).[80, 81] *DLCK2* is a target gene of miR-3937, linked with the AUDIT (rs4423856, *P-value*=1.48x10^{-8}).[74] The last alcohol-associated miRNA (miR-4507) identified in our study, was previously reported by Gardiner et al.[79] when comparing alcohol consumption to alcohol abstinence during pregnancy. Multiple target genes of alcohol-associated miRNAs were linked to alcohol consumption through other omics analyses (**Supplemental Table 7**).[33, 38, 74] In addition, most of the identified miRNAs were previously implicated in liver diseases, which is unsurprising as the liver is a primary organ for alcohol metabolism and

detoxification.[16, 91, 92] For instance, miR-193b-3p, miR-3937, and miR-122-5p were linked with fatty liver disease in the Rotterdam Study.[48] In addition, miR-122-5p is firmly recognized as a liver-specific miRNA[93] with an undeniably established role in liver function and related diseases.[93-95] These results corroborate well with the findings linking the newly identified miRNAs to alcohol consumption. When we explored the alcohol consumption as categorical exposure (non-, light-, and heavy-drinkers), despite the smaller samples size, the effect estimates were almost doubled for the alcohol-associated miRNAs when comparing heavy-drinkers and light-drinkers to the non-drinkers group (**Table 2**). In our sensitivity analysis, we identified that most of the alcohol-associated miRNAs had stronger effect estimates in men, perhaps due to the higher consumption of alcohol.

Our mediation analyses showed a potential mediatory role of miR-122-5p in the association of alcohol consumption and CT-based fatty liver disease, GGT, and US-based steatosis. Moreover, we observed a mediating effect of miR-3937 in the association between alcohol consumption and CT-based fatty liver and GGT. This may indicate a significantly estimated indirect effect of alcohol consumption on liver function or disease that is mediated partly through miR-3937 and miR-122-5p. In addition, we did not find any statistical evidence for causality between alcohol-associated miRNAs and liver-related traits. However, we believe that these results might have been hampered by the lack of strong instrumental variables (IVs), as we only found a single SNP as a valid IV. This warrants future studies to perform large-scale genome-wide association studies (GWASs) on a broad landscape of miRNAs, providing stronger IVs for estimating causal relationships.

This study has strengths as well as limitations that should be considered when interpreting the results. The strengths of our study include the large sample size, availability of clinical outcomes, and using a new RNA-sequencing-based assay with high sensitivity. Yet, it is plausible that several limitations could have influenced the results presented. First, mediation analysis requires strong assumptions whose violations might lead to spurious results, such as unmeasured confounding. In line with this, implementing mediation analysis in cross-sectional observational studies and notably in genomic studies is challenging and adds a layer of complexity.[96] We implemented bias analyses to explore if the assumption of unmeasured confounding holds. Given the cross-sectional nature of the data used for the presented study, we cannot rule out reverse causality. In line with this, data on miRNAs, alcohol consumption, fatty liver, and liver enzymes were measured at the same time point, while ultrasound data are analyzed in a longitudinal setting. Although we adjusted for potential confounders, there might still be residual confounding due to the dynamic nature of epigenetic markers - which might partially explain some of the ρ values close to 0 we obtained from bias analyses within mediation analyses. Future analyses are warranted to replicate the findings from our study and explore these findings in a longitudinal setting. In addition, future studies are needed to explore the dynamic nature of epigenetic markers such as miRNAs and explore reverse causation, especially in the context of mediation analysis. Another source of bias might have occurred from the CT scan used for liver attenuation,[97] however, we also included data on ultrasound-based

measurements. In addition, the FibroScan is currently an often used method in the clinic to determine liver fat and fibrosis while we used CT scans in the current study. Nevertheless, large cohort studies are more likely to use CT scans due to their broad implications, making it possible for direct replication of our obtained results by other studies. Also, since miRNAs are tissue-specific, we might have missed important miRNAs in relevant tissue such as liver. Though, the accessibility of plasma compared to the other tissues provides a potential benefit for identified miRNAs to serve as indicators for alcohol exposure.[98] In addition, we utilized the Tissue Atlas database (https://ccb-web.cs.uni-saarland.de/tissueatlas) [68, 69] in order to explore the expression of the alcohol -identified miRNAs across a wide range of tissues.

In addition, it is important to address the potential limitation coming from the data on alcohol consumption, as it was collected by home-administered interviews and not by food frequency questionnaires (FFQs) or other validated self-reports, such as AUDIT (https://auditscreen.org/).[99] Although the FFQs are more detailed and AUDIT is more effective in screening individuals with unhealthy alcohol use, we did not have data derived from FFQ or AUDIT on this wave of participants. In addition, participants might have underestimated their true alcohol consumption due to social desirability bias. Finally, it is important to acknowledge the potential risk of introducing the type I error in our additional analysis as we did not correct for multiple testing. Given the nature of high correlation of omics data, we believe the potential risk of introducing type I error in our additional analysis is accounted for, to a certain extent. Further studies are needed to replicate our findings in larger sample sizes and longer follow-up time as well as experimentally confirm the role of identified miRNAs in molecular pathways underlying alcohol-related diseases.

In conclusion, we showed in a population-based setting that alcohol consumption was associated with plasma levels of four miRNAs, two of which show a potential mediatory role on liver-related traits. This might provide a better understanding of the mechanism of action involved between alcohol consumption and alterations in gene expression in alcohol-related diseases.

Supplementary material

Supplementary material is a viable on

References

1. Ezzati M, Riboli E. Behavioral and dietary risk factors for noncommunicable diseases. N Engl J Med. 2013;369(10):954-64.
2. Larsson SC, Burgess S, Mason AM, Michaelsson K. Alcohol Consumption and Cardiovascular Disease: A Mendelian Randomization Study. Circ Genom Precis Med. 2020;13(3):e002814.
3. Ramstedt M. Per capita alcohol consumption and liver cirrhosis mortality in 14 European countries. Addiction. 2001;96:S19-S33.
4. Niemela O, Alatalo P. Biomarkers of alcohol consumption and related liver disease. Scand J Clin Lab Inv. 2010;70(5):305-12.
5. Chuang SC, Lee YCA, Wu GJ, Straif K, Hashibe M. O Alcohol consumption and liver cancer risk: a meta-analysis. Cancer Cause Control. 2015;26(9):1205-31.
6. Hagstrom H. Alcohol Consumption in Concomitant Liver Disease: How Much is Too Much? Curr Hepatol Rep. 2017;16(2):152-7.
7. Connor J. Alcohol consumption as a cause of cancer. Addiction. 2017;112(2):222-8.
8. Bagnardi V, Blangiardo M, La Vecchia C, Corrao G. Alcohol consumption and the risk of cancer - A meta-analysis. Alcohol Research & Health. 2001;25(4):263-70.
9. Boffetta P, Hashibe M. Alcohol and cancer. Lancet Oncol. 2006;7(2):149-56.
10. Shield KD, Soerjomataram I, Rehm J. Alcohol Use and Breast Cancer: A Critical Review. Alcoholism-Clinical and Experimental Research. 2016;40(6):1166-81.
11. Rehm J, Greenfield TK, Rogers JD. Average volume of alcohol consumption, patterns of drinking, and all-cause mortality: results from the US National Alcohol Survey. Am J Epidemiol. 2001;153(1):64-71.
12. Gutjahr E, Gmel G, Rehm J. Relation between average alcohol consumption and disease: An overview. Eur Addict Res. 2001;7(3):117-27.
13. Garofalo C, Wright AGC. Alcohol abuse, personality disorders, and aggression: The quest for a common underlying mechanism. Aggress Violent Beh. 2017;34:1-8.
14. Carnevale R, Nocella C. Alcohol and cardiovascular disease: Still unresolved underlying mechanisms. Vasc Pharmacol. 2012;57(2-4):69-71.
15. Seitz HK, Stickel F. Molecular mechanisms of alcohol-mediated carcinogenesis. Nat Rev Cancer. 2007;7(8):599-612.
16. Cederbaum AI. Alcohol metabolism. Clin Liver Dis. 2012;16(4):667-85.
17. Maher JJ. Exploring alcohol's effects on liver function. Alcohol Health Res World. 1997;21(1):5-12.
18. Singal AK, Bataller R, Ahn J, Kamath PS, Shah VH. ACG Clinical Guideline: Alcoholic Liver Disease. Am J Gastroenterol. 2018;113(2):175-94.
19. Meroni M, Longo M, Rametta R, Dongiovanni P. Genetic and Epigenetic Modifiers of Alcoholic Liver Disease. Int J Mol Sci. 2018;19(12).
20. Moosavi A, Motevalizadeh Ardekani A. Role of Epigenetics in Biology and Human Diseases. Iran Biomed J. 2016;20(5):246-58.
21. Starkman BG, Sakharkar AJ, Pandey SC. Epigenetics-beyond the genome in alcoholism. Alcohol Research. 2012;34(3):293-305.
22. Sushma, Divakar A, Kanchan S, Jha G, Mishra S, Sharma D, et al. Alcohol induced impairment/abnormalities in brain: Role of MicroRNAs. Neurotoxicology. 2021;87:11-23.
23. Lewohl JM, Nunez YO, Dodd PR, Tiwari GR, Harris RA, Mayfield RD. Up-regulation of microRNAs in brain of human alcoholics. Alcohol Clin Exp Res. 2011;35(11):1928-37.
24. Gibney ER, Nolan CM. Epigenetics and gene expression. Heredity (Edinb). 2010;105(1):4-13.
25. Ha M, Kim VN. Regulation of microRNA biogenesis. Nat Rev Mol Cell Biol. 2014;15(8):509-24.
26. O'Brien J, Hayder H, Zayed Y, Peng C. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. Front Endocrinol. 2018;9.
27. Condrat CE, Thompson DC, Barbu MG, Bugnar OL, Boboc A, Cretoiu D, et al. miRNAs as Biomarkers in Disease: Latest Findings Regarding Their Role in Diagnosis and Prognosis. Cells-Basel. 2020;9(2).
28. Jamaluddin MS, Weakley SM, Zhang LD, Kougias P, Lin PH, Yao QZ, et al. miRNAs: roles and clinical applications in vascular disease. Expert Rev Mol Diagn. 2011;11(1):79-89.
29. Visone R, Croce CM. MiRNAs and Cancer. Am J Pathol. 2009;174(4):1131-8.

30. Szabo G, Satishchandran A. MicroRNAs in Alcoholic Liver Disease. Semin Liver Dis. 2015;35(1):36-42.
31. Szabo G, Bala S. MicroRNAs in liver disease. Nat Rev Gastro Hepat. 2013;10(9):542-52.
32. Bala S, Marcos M, Szabo G. Emerging role of microRNAs in liver diseases. World J Gastroentero. 2009;15(45):5633-40.
33. Panico A, Tumolo MR, Leo CG, Donno A, Grassi T, Bagordo F, et al. The influence of lifestyle factors on miRNA expression and signal pathways: a review. Epigenomics. 2021;13(2):145-64.
34. Ten Berg PW, Shaffer J, Vliegenthart ADB, McCrae J, Sharkey N, Webb DJ, et al. Attending a social event and consuming alcohol is associated with changes in serum microRNA: a before and after study in healthy adults. Biomarkers. 2018;23(8):781-6.
35. McCrae JC, Sharkey N, Webb DJ, Vliegenthart AD, Dear JW. Ethanol consumption produces a small increase in circulating miR-122 in healthy individuals. Clinical Toxicology: The Official Journal of the American Academy of Clinical Toxicology & European Association of Poisons Centres & Clinical Toxicologists. 2016;54(1):53-5.
36. Ikram MA, Brusselle G, Ghanbari M, Goedegebure A, Ikram MK, Kavousi M, et al. Objectives, design and main findings until 2020 from the Rotterdam Study. Eur J Epidemiol. 2020;35(5):483-517.
37. Kuntsche E, Kuendig H. Beyond self-reports: drinking motives predict grams of consumed alcohol in wine-tasting sessions. Exp Clin Psychopharmacol. 2012;20(4):318-24.
38. Liu C, Marioni RE, Hedman AK, Pfeiffer L, Tsai PC, Reynolds LM, et al. A DNA methylation biomarker of alcohol consumption. Mol Psychiatry. 2018;23(2):422-33.
39. Odink AE, van der Lugt A, Hofman A, Hunink MG, Breteler MM, Krestin GP, et al. Association between calcification in the coronary arteries, aortic arch and carotid arteries: the Rotterdam study. Atherosclerosis. 2007;193(2):408-13.
40. Kodama Y, Ng CS, Wu TT, Ayers GD, Curley SA, Abdalla EK, et al. Comparison of CT methods for determining the fat content of the liver. AJR Am J Roentgenol. 2007;188(5):1307-12.
41. Wolff L, Bos D, Murad SD, Franco OH, Krestin GP, Hofman A, et al. Liver fat is related to cardiovascular risk factors and subclinical vascular disease: the Rotterdam Study. Eur Heart J-Card Img. 2016;17(12):1361-7.
42. Hamaguchi M, Kojima T, Itoh Y, Harano Y, Fujii K, Nakajima T, et al. The severity of ultrasonographic findings in nonalcoholic fatty liver disease reflects the metabolic syndrome and visceral fat accumulation. Am J Gastroenterol. 2007;102(12):2708-15.
43. Alferink LJM, Trajanoska K, Erler NS, Schoufour JD, de Knegt RJ, Ikram MA, et al. Nonalcoholic Fatty Liver Disease in The Rotterdam Study: About Muscle Mass, Sarcopenia, Fat Mass, and Fat Distribution. J Bone Miner Res. 2019;34(7):1254-63.
44. Koehler EM, Sanna D, Hansen BE, van Rooij FJ, Heeringa J, Hofman A, et al. Serum liver enzymes are associated with all-cause mortality in an elderly population. Liver International. 2014;34(2):296-304.
45. Wilsnack RW, Wilsnack SC, Kristjanson AF, Vogeltanz-Holm ND, Gmel G. Gender and alcohol consumption: patterns from the multinational GENACIS project. Addiction. 2009;104(9):1487-500.
46. Frezza M, di Padova C, Pozzato G, Terpin M, Baraona E, Lieber CS. High blood alcohol levels in women. The role of decreased gastric alcohol dehydrogenase activity and first-pass metabolism. N Engl J Med. 1990;322(2):95-9.
47. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol. 1986;51(6):1173-82.
48. Zhang X, Mens MMJ, Abozaid YJ, Bos D, Darwish Murad S, de Knegt RJ, et al. Circulatory microRNAs as potential biomarkers for fatty liver disease: the Rotterdam study. Alimentary Pharmacology & Therapeutics. 2021;53(3):432-42.
49. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. mediation: R Package for Causal Mediation Analysis. Journal of Statistical Software. 2014;59(5).
50. Lee H, Cashin AG, Lamb SE, Hopewell S, Vansteelandt S, VanderWeele TJ, et al. A Guideline for Reporting Mediation Analyses of Randomized Trials and Observational Studies: The AGReMA Statement. JAMA. 2021;326(11):1045-56.

51. Akiyama S, Higaki S, Ochiya T, Ozaki K, Niida S, Shigemizu D. JAMIR-eQTL: Japanese genome-wide identification of microRNA expression quantitative trait loci across dementia types. Database (Oxford). 2021;2021(2021).
52. Nikpay M, Beehler K, Valsesia A, Hager J, Harper ME, Dent R, et al. Genome-wide identification of circulating-miRNA expression quantitative trait loci reveals the role of several miRNAs in the regulation of cardiometabolic phenotypes. Cardiovasc Res. 2019;115(11):1629-45.
53. Huan T, Rong J, Liu C, Zhang X, Tanriverdi K, Joehanes R, et al. Genome-wide identification of microRNA expression quantitative trait loci. Nat Commun. 2015;6:6601.
54. Richardson TG, Zheng J, Davey Smith G, Timpson NJ, Gaunt TR, Relton CL, et al. Mendelian Randomization Analysis Identifies CpG Sites as Putative Mediators for Genetic Influences on Cardiovascular Disease Risk. Am J Hum Genet. 2017;101(4):590-602.
55. Liu Y, Basty N, Whitcher B, Bell JD, Sorokin EP, van Bruggen N, et al. Genetic architecture of 11 organ traits derived from abdominal MRI using deep learning. Elife. 2021;10.
56. Pazoki R, Vujkovic M, Elliott J, Evangelou E, Gill D, Ghanbari M, et al. Genetic analysis in European ancestry individuals identifies 517 loci associated with liver enzymes. Nat Commun. 2021;12(1):2579.
57. Chung-hong Chan GCC, Thomas J. Leeper, Jason Becker. rio: A Swiss-Army Knife for Data I/O, R package version 0.5.27: https://rdrr.io/cran/rio/; 2021 [Available from: https://rdrr.io/cran/rio/.
58. Hadley Wickham MA, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller,, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Wood, Hiroaki Yutani. Welcome to the {tidyverse}. Journal of Open Source Software. 2019;4:1686.
59. Firke S. janitor: Simple Tools for Examining and Cleaning Dirty Data. 2021.
60. Wickham GGaH. Dates and Times Made Easy with {lubridate}. Journal of Statistical Software. 2011;40:1--25.
61. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.: https://www.r-project.org/; 2021 [Available from: https://www.R-project.org/.
62. Couch DRaAHaS. broom: Convert Statistical Objects into Tidy Tibbles. 2021.
63. Makowski DLaMSB-SaIPaPWaD. performance: An R Package for Assessment, Comparison and Testing of Statistical Models. Journal of Open Source Software. 2021;6:3139.
64. Wickham LHaH. purrr: Functional Programming Tools. 2020.
65. Wickham H. ggplot2: Elegant Graphics for Data Analysis: Springer-Verlag New York; 2016.
66. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018;7.
67. Bartel KYaA. tableone: Create 'Table 1' to Describe Baseline Characteristics with or without Propensity Score Weights. 2021.
68. University CfCB-S. Tissue Atlas: Chair for Clinical Bioinformatics - Saarland University; May 2016 [updated July 2022. Available from: https://ccb-web.cs.uni-saarland.de/tissueatlas.
69. Ludwig N, Leidinger P, Becker K, Backes C, Fehlmann T, Pallasch C, et al. Distribution of miRNA expression across human tissues. Nucleic Acids Res. 2016;44(8):3865-77.
70. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. Elife. 2015;4.
71. Huang HY, Lin YC, Li J, Huang KY, Shrestha S, Hong HC, et al. miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. Nucleic Acids Res. 2020;48(D1):D148-D54.
72. Chen YH, Wang XW. miRDB: an online database for prediction of functional microRNA targets. Nucleic Acids Research. 2020;48(D1):D127-D31.
73. Wong N, Wang X. miRDB: an online resource for microRNA target prediction and functional annotations. Nucleic Acids Res. 2015;43(Database issue):D146-52.
74. Kranzler HR, Zhou H, Kember RL, Smith RV, Justice AC, Damrauer S, et al. Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. Nat Commun. 2019;10.

75. Mancuso N, Shi H, Goddard P, Kichaev G, Gusev A, Pasaniuc B. Integrating Gene Expression with Summary Association Statistics to Identify Genes Associated with 30 Complex Traits. Am J Hum Genet. 2017;100(3):473-87.
76. Carbon S, Douglass E, Good BM, Unni DR, Harris NL, Mungall CJ, et al. The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Research. 2021;49(D1):D325-D34.
77. Project GOPA. PANTHER: Gene Ontology Phylogenetic Annotation Project; [updated 22-02-2022.
78. FinnGen. FinnGen: FinnGen; 2020 [updated January 2022. Available from: https://finngen.gitbook.io/documentation/.
79. Gardiner AS, Gutierrez HL, Luo L, Davies S, Savage DD, Bakhireva LN, et al. Alcohol Use During Pregnancy is Associated with Specific Alterations in MicroRNA Levels in Maternal Serum. Alcohol Clin Exp Res. 2016;40(4):826-37.
80. School GLatD-FCIaHM. TWAS hub: Gusev Lab at the Dana-Farber Cancer Institute and Harvard Medical School; 06/10/2018 [updated 09/19/2018 Available from: http://twas-hub.org/traits/.
81. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet. 2016;48(3):245-52.
82. Sathyan P, Golden HB, Miranda RC. Competing interactions between micro-RNAs determine neural progenitor survival and proliferation after ethanol exposure: evidence from an ex vivo model of the fetal cerebral cortical neuroepithelium. J Neurosci. 2007;27(32):8546-57.
83. Wang LL, Zhang Z, Li Q, Yang R, Pei X, Xu Y, et al. Ethanol exposure induces differential microRNA and target gene expression and teratogenic effects which can be suppressed by folic acid supplementation. Human Reproduction. 2009;24(3):562-79.
84. Mandal C, Halder D, Jung KH, Chai YG. Maternal alcohol consumption and altered miRNAs in the developing fetus: Context and future perspectives. J Appl Toxicol. 2018;38(1):100-7.
85. Al-Shaer AE, Flentke GR, Berres ME, Garic A, Smith SM. Exon level machine learning analyses elucidate novel candidate miRNA targets in an avian model of fetal alcohol spectrum disorder. PLoS Comput Biol. 2019;15(4):e1006937.
86. Choi MR, Han JS, Jin YB, Lee SR, Choi IY, Lee H, et al. Differential expression of microRNAs in the hippocampi of male and female rodents after chronic alcohol administration. Biol Sex Differ. 2020;11(1):65.
87. Sinirlioglu ZA, Coskunpinar E, Akbas F. miRNA and mRNA expression profiling in rat brain following alcohol dependence and withdrawal. Cell Mol Biol (Noisy-le-grand). 2017;63(2):49-56.
88. Git A, Dvinge H, Salmon-Divon M, Osborne M, Kutter C, Hadfield J, et al. Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. RNA. 2010;16(5):991-1006.
89. Wakabayashi I, Sourij H, Sotoda Y, Daimon T, Groschner K, Rainer PP. Ethnic Differences in Serum Levels of microRNAs Potentially Regulating Alcohol Dehydrogenase 1B and Aldehyde Dehydrogenase 2. J Clin Med. 2021;10(16).
90. Chen YC, Peng GS, Wang MF, Tsao TP, Yin SJ. Polymorphism of ethanol-metabolism genes and alcoholism: correlation of allelic variations with the pharmacokinetic and pharmacodynamic consequences. Chem Biol Interact. 2009;178(1-3):2-7.
91. Singal AK, Mathurin P. Diagnosis and Treatment of Alcohol-Associated Liver Disease: A Review. JAMA. 2021;326(2):165-76.
92. Buchanan R, Sinclair JMA. Alcohol use disorder and the liver. Addiction. 2021;116(5):1270-8.
93. Filipowicz W, Grosshans H. The liver-specific microRNA miR-122: biology and therapeutic potential. Prog Drug Res. 2011;67:221-38.
94. Jopling CL, Norman KL, Sarnow P. Positive and negative modulation of viral and cellular mRNAs by liver-specific microRNA miR-122. Cold Spring Harb Symp Quant Biol. 2006;71:369-76.
95. Lewis AP, Jopling CL. Regulation and biological function of the liver-specific miR-122. Biochem Soc Trans. 2010;38(6):1553-7.
96. Zeng P, Shao Z, Zhou X. Statistical methods for mediation analysis in the era of high-throughput genomics: Current successes and future challenges. Comput Struct Biotechnol J. 2021;19:3209-24.
97. Zhang YN, Fowler KJ, Hamilton G, Cui JY, Sy EZ, Balanay M, et al. Liver fat imaging-a clinical overview of ultrasound, CT, and MR imaging. Br J Radiol. 2018;91(1089):20170959.

98. Szelenberger R, Kacprzak M, Saluk-Bijak J, Zielinska M, Bijak M. Plasma MicroRNA as a novel diagnostic. Clinica Chimica Acta. 2019;499:98-107.
99. a WHO-approved instrument tAiitpd. AUDIT- Alcohol Use Disorder Identification Test 2022 [Available from: https://auditscreen.org.

Chapter 3

Proteomics of fatty liver disease

Chapter 3.1

Plasma proteomic signature of fatty liver disease: the Rotterdam Study

Yasir J. Abozaid, Ibrahim Ayada*, Laurens A. van Kleef*, Costanza L Vallerga, Qiuwei Pan, Willem Pieter Brouwer, M. Arfan Ikram, Joyce Van Meurs, Robert J. de Knegt, Mohsen Ghanbari

Hepatology. 2023 Jul 1;78(1):284-294.

***Authors denotes equal contribution**

Abstract

Background & Aims: Fatty liver disease (FLD) is caused by excess fat in the liver and its global prevalence exceeds 33%. The role of protein expression on the pathogenesis of FLD and accompanied fibrosis and its potential as disease biomarker is currently not clear. Hence we aimed to identify plasma proteomics associated with FLD and fibrosis using population-based data.

Approach & Results: Blood samples were collected in 2578 participants from the population-based Rotterdam study cohort. The proximity extension assay reliably measured plasma levels of 171 cardiometabolic and inflammatory-related proteins (Olink Proteomics). FLD was assessed by ultrasound and fibrosis by transient elastography. Logistic regression models quantified the association of plasma proteomics with FLD and fibrosis. Additionally, we aimed to validate our results in liver organoids. Cross-sectional analysis identified 27 proteins significantly associated with FLD surpassing the Bonferroni-corrected $P<2.92\times10^{-4}$. The strongest association were observed for FGF-21 (β=0.45, $P=1.07\times10^{-18}$) and CES1 protein (β=0.66, $P=4.91\times10^{-40}$). Importantly, 15 of the 27 proteins significantly associated with FLD were also associated with liver fibrosis. Finally, consistent with plasma proteomic profiling, we found the expression levels of IL-18R1 and CES1 to be upregulated in a fatty liver disease model of 3D culture human liver organoids.

Conclusions: Among the general population, several inflammatory and cardiometabolic plasma proteins were associated with FLD and fibrosis. Particularly plasma levels of FGF-21, IL-18R1 and CES1 were largely dependent on the presence of FLD and fibrosis, and may therefore be important in their pathogenesis.

Introduction

Fatty liver disease (FLD) has become increasingly prevalent in the past decades with over 33% of the world's population currently affected.[1] This worrisome rise in prevalence has followed an increase in obesity and other metabolic disorders.[2] FLD is a spectrum disease ranging from simple steatosis to more severe consequences like steatohepatitis, cirrhosis, and even hepatocellular carcinoma.[3-5] Despite the high prevalence and enormous disease burden, the understanding of protein expression in FLD and fibrosis is limited as well as their potential as biomarkers to facilitate early detection.[6, 7]

Proteomics is a powerful tool for examining differences in protein expression. Promising results have been obtained previously by using sophisticated techniques to quantify protein expression. Recently, a range of proteins has been identified that were crucial for the presence and severity of alcoholic liver disease.[8] This indicates that proteomic signatures may act as potential biomarkers to predict FLD at early stage. Indeed several small studies already showed its value using the SomaScan assay to define protein expression in patient cohorts with advanced FLD.[9, 10]

Given the potential of proteomics, we performed plasma profiling of inflammatory and cardiometabolic-related proteins among a large cohort of community-dwelling individuals and thereby aimed to increase the understanding of protein expression in FLD and liver fibrosis as well as assess their potential as disease biomarkers.

Methods

Participants

This study was conducted within the prospective Rotterdam Study cohort targeting community-dwelling individuals aged ≥ 45 years living in Ommoord, a suburb of Rotterdam. The design, aims and recent achievements of the Rotterdam Study have extensively been described elsewhere.[11] For the current study, we included participants from the third Rotterdam Study extension (RS-III), collected between February 2006 and December 2008 with available data on the plasma levels of inflammatory (n=3456) and cardiometabolic-related proteins (n=3502). Exclusion criteria were missing data on FLD or covariates included in the final models **(Figure 1)**. The final sample size for the inflammatory panel was 2578 and for the cardiometabolic panel 2595.

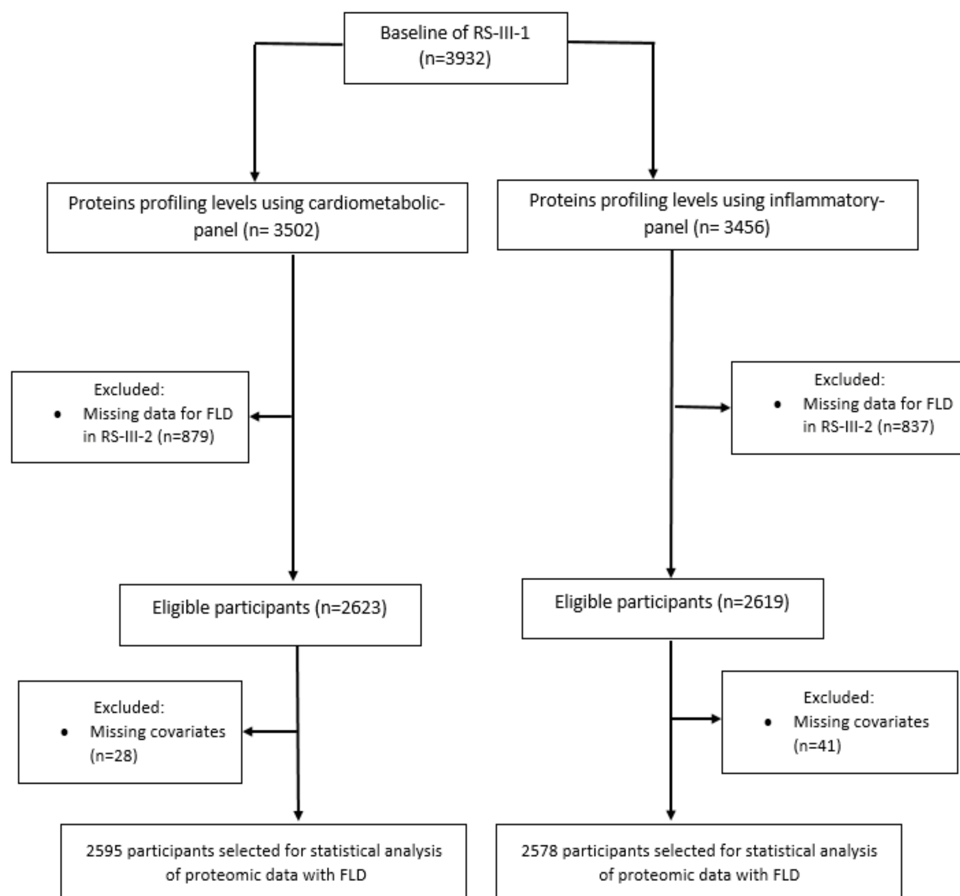


Figure 1. Flow chart of the study participants. Abbreviations: RS-III indicates the third cohort of Rotterdam Study; FLD, fatty liver disease.

Assessment of FLD and liver stiffness

FLD was defined as hyperechoic liver parenchyma compared to the kidney cortex which was assessed by abdominal ultrasonography between 2009 and 2014, 5.5 [P25-P75: 5.4-5.7] years after plasma sample collection.[12] A single experienced sonographer performed all the examinations on a Hitachi HI VISION 900. In addition to FLD, we defined NAFLD and MAFLD for sensitivity analysis. NAFLD was defined as the presence of FLD in the absence of the following secondary causes of FLD: (1) excessive alcohol consumption (>30 g/day for men and >20 g/day for women), (2) presence of viral hepatitis, and (3) use of steatogenic drugs. MAFLD was defined as FLD combined with metabolic dysfunction, comprising overweight (BMI ≥ 25 kg/m^2), type 2 diabetes mellitus or two minor metabolic dysfunction criteria such as hypertension, high waist circumference, or dyslipidemia.[13] Moreover, liver stiffness measurement (LSM) was performed using transient elastography (FibroScan; Echosens, Paris, France). At least 10 measurements were obtained with the M or XL probe. In case of an interquartile range of >30%, measurements exceeding 7.1 kPa were considered unreliable and discarded.[14] Fibrosis was defined as reliable liver stiffness measurement ≥ 8.0 kPa.[15] Plasma proteomic profiling was performed before the assessment of FLD and fibrosis.

Assessment of plasma proteomic profiling

Plasma protein levels were measured using the proximity extension immunoassay (PEA) technology (Olink Proteomics ®, Uppsala, Sweden) on the ProSeek Multiplex Inflammation (version 3021) and Cardiometabolic (version 3602) 96-plex panels. The Olink PEA technology uses a dual recognition DNA-coupled immunoassay that rapidly allows for protein identification and relative quantification with high sensitivity and specificity. The method of proteomics level measurement has been described in detail previously.[16]

Olink-generated proteomics data were background corrected, log2-transformed, and normalized to a Normalized Protein Expression (NPX) scale. NPX values represent relative quantification, meaning that protein values can be compared for the same protein across samples.

Sample-based Quality Control (QC) of the proteomics data was done as previously published[17] based on Olink reported QC warnings, i.e. samples that deviate more than +/- 0.3 NPX from the plate median. In this study, there were a total of 76 QC warnings. For assay-based QC, we excluded proteins with a high proportion of samples (>90%) with NPX values below the Limit of Detection (LOD). For the Inflammation panel, nine proteins were removed, including interleukin-1 alpha (IL-1α), interleukin-2 (IL-2), interleukin-24 (IL-24), interleukin-20 (IL-20), interleukin-33 (IL-33), leukemia inhibitory factor (LIF), thymic stromal lymphopoietin (TSLP), neurturin, NRTN, and beta-nerve growth factor (NRTN). While four proteins were removed from the cardiometabolic panel, including lysosomal Pro-X carboxypeptidase (PRCP), latent transforming growth factor-beta binding protein 2 (LTBP2), Lithostathine-1-alpha

(REG1A), and superoxide dismutase (SOD1). NPX values below LOD were treated as missing values. After QC, we retained 83 proteins from the inflammation panel and 88 proteins from the cardiometabolic panel.

Assessment of covariates

Information on age, sex, and smoking status (three categories current/former/ never), alcohol consumption (measured in grams of ethanol per day) were obtained from questionnaires.[11] Body mass index was calculated based on weight in kilograms divided the by height in meters squared. Prevalent diabetes mellitus type 2 was identified according to the World Health Organization criteria: fasting glucose levels of ≥7.0 mmol/L, nonfasting glucose levels ≥11.1 mmol/L or the use of glucose-lowering medication.[18] Fasting blood samples of participants were obtained during each visit to the research center. Full blood count measurements were performed directly after the blood sample was drawn using the COULTER® Ac·T diff2™ Hematology Analyzer (Beckman Coulter, Brea, CA, US). Laboratory measurements included absolute granulocyte, platelet, lymphocyte, and monocyte counts in 10^9 per liter. All covariates included for statistical analysis were obtained at baseline examination.

Validation in human liver organoids

A database was designed and recently published by our NAFLD/NASH group containing results on genome-wide transcriptomic profiling on human liver organoids mimicking FLD, compared to controls. Detailed information on the procedures are available elsewhere.[19] Within this database, we aimed to validate our results with the genome-wide transcriptomic profiling from this experimental model.

Statistical analysis

We used multivariable linear and logistic regression models to quantify the association between plasma levels of proteins as exposure and (1) FLD and (2) liver stiffness (continuous and categorical) as outcome. The associations were adjusted for age and sex (Model 1) and in addition for blood cell counts (RBCs, granulocyte, lymphocyte, monocyte, and platelet), BMI, alcohol, diabetes, and smoking (Model 2). In sensitivity analysis, we replaced FLD with NAFLD and MAFLD. Additional sensitivity analysis was performed using a fatty liver index (FLI) based on definition of FLD. The FLI comprised BMI, waist circumference, gamma-glutamyltransferase and triglycerides,[20] which was collected at the same time of plasma sample collection.

Further analysis focused on the potential of protein expression as a biomarker for FLD and fibrosis. The area under the curve (AUC) was calculated including the confidence intervals for the set of significantly associated proteins in the prediction of FLD and fibrosis, based on a logistic regression model including the significantly associated proteins previously identified. The diagnostic accuracy of proteins were then compared to common non-invasive markers for the prediction of FLD and fibrosis. These include FLI as

marker for FLD and the fibrosis-4 index (FIB-4), comprised of age, AST, ALT and platelets, as marker for fibrosis. Finally, proteomics were combined either with FLI or FIB-4 to assess the added value of proteins for already existing and commonly used scores.

All analyses were performed using SPSS statistical software (SPSS, version25; IBM Corp) and R software version 3.5.2 (The R Foundation for Statistical Computing). To account for multiple testing, the Bonferroni adjusted p-value for our primary aim (protein expression with FLD) was 2.92×10^{-4} based on the number of tested proteins (n = 171) and set as the threshold for statistical significance. The PANTHER (v.17.0) (protein annotation through evolutionary relationship) classification system (http://www.pantherdb.org/) was used to check the Reactome-pathways of the identified proteins in both inflammatory and cardiometabolic panels.

Results

Participant characteristics

The baseline characteristics of study participants are illustrated in **Table 1** and **S1**. A total of 2578 participants with available data on plasma protein levels and abdominal ultrasound were included. In general, the mean age was 56.3±5.9 years, 56.6% were female and metabolic comorbidity was highly common, illustrated by the mean BMI of 27.6 kg/m^2 ± 4.4 and diabetes prevalence of 8.8%. The overall prevalence of FLD was 35.2% and 4.6% had fibrosis.

Table 1. Characteristics of the Rotterdam Study participants

Characteristics	FLD-US		Liver fibrosis-TE	
	Case	**Control**	**Case**	**Control**
N	907	1671	105	2157
Age	56.72±5.68	56.14±5.94	57.85±7.03	56.05±5.63
Sex,f n(%)	465(51.3)	994(59.5)	31(29.5)	1235(57.3)
Body mass index (kg/m^2)	29.68±4.40	26.51±3.96	29.62±4.77	27.21±4.03
Alcohol consumption (g/day)	9.89±10.65	8.08±8.53	11.52±10.24	8.90±9.23
Type 2 diabetes n(%)	136(15.0)	92(5.5)	27(25.7)	147(6.8)
Smoking status n(%)				
Never	262(28.9)	589(35.2)	25(23.8)	727(33.7)
Current	183(20.2)	357(21.4)	27(25.7)	452(21.0)
Former	462(50.9)	725(43.4)	53(50.5)	978(45.3)
Blood cells counts, 10^9/l				
Red blood cells	5.09±0.35	4.80±0.37	5.07±0.38	4.87±0.37
Granulocyte	4.17±1.40	4.01±1.46	4.19±1.33	4.05±1.46
Lymphocyte	2.45±0.66	2.46±0.72	2.50±0.70	2.44±0.69
Monocyte	0.43±0.22	0.43±0.23	0.43±0.16	0.43±0.23
Platelets	0.46±0.03	0.44±0.03	0.46±0.03	0.44±0.03

Note: The characteristics of study participants in both inflammatory and cardiometabolic proteomics panels; Variables are represented as mean (± standard deviation), or number (%). Abbreviations: US, ultrasound; TE, transient elastography; N, Sample size.

Plasma proteins associated with FLD

We found 8 out of 83 inflammatory-related proteins to be significantly associated with FLD in fully adjusted models, taking into account multiple testing correction ($P<2.92\times10^{-4}$). Among them, 5 proteins (FGF-21, HGF, CDCP1, TNFSF14, and IL-18R1) were positively associated, whereas 3 proteins (CX3CL1, SCF, and TWEAK) were negatively associated with FLD (**Table 2A** and **Figure 2A**). Regarding the cardiometabolic panel, we identified 19 out of 88 proteins to be significantly associated with FLD. Of these, 12 proteins (CES1, F7, SERPINA5, THBS4, PROC, FCN2, ICAM1, CNDP1, MEGF9, LILRB1, TGFBI, and SERPINA7) were positively associated and 7 proteins (MET, KIT, PLTP, PTPRS, CHL1, TGFBR3, and REG1A) were negatively associated with FLD (**Table 2B** and **Figure 2B)**. Importantly, 25/27 proteins significantly associated with ultrasound defined FLD could be validated using FLD based on FLI which was calculated at the same time point as the proteomic data (**Table S2**).

Plasma proteins associated with fibrosis

Our analysis showed 7 out of the 8 inflammatory proteins significantly associated with FLD to be also associated with fibrosis (on a categorical scale ≥ 8.0 kPa) and liver stiffness (on a continuous scale). For the cardiometabolic panel, 8 out of the 19 proteins significantly associated with FLD were associated with fibrosis, and 10 out of the 19 proteins were associated with liver stiffness on a continuous scale in fully adjusted models **(Table 2** and **Table S3)**.

Table 2 Plasma proteins significantly associated with FLD and liver fibrosis

	Fatty liver disease			liver fibrosis		
Proteins ID	**Beta**	**SE**	**P-value**	**Beta**	**SE**	**P-value**
A- Inflammatory-related proteins						
FGF-21	0.45	0.051	1.07×10^{-18}	0.31	0.112	6.09×10^{-03}
HGF	0.34	0.055	1.27×10^{-09}	0.50	0.113	8.58×10^{-06}
CX3CL1	-0.23	0.048	1.63×10^{-06}	-0.06	0.111	6.02×10^{-01}
SCF	-0.23	0.049	3.01×10^{-06}	-0.31	0.098	1.78×10^{-03}
CDCP1	0.23	0.05	3.50×10^{-06}	0.42	0.101	4.06×10^{-05}
TNFSF14	0.22	0.049	$4.80 \times 10\text{-}^{06}$	0.26	0.107	1.42×10^{-02}
TWEAK	-0.23	0.051	5.64×10^{-06}	-0.31	0.115	7.46×10^{-03}
IL-18R1	0.23	0.051	7.25×10^{-06}	0.26	0.116	2.58×10^{-02}
B- Cardiometabolic-related proteins						
CES1	0.66	0.05	4.91×10^{-40}	0.37	0.090	3.33×10^{-05}
F7	0.37	0.046	1.19×10^{-15}	0.08	0.106	4.36×10^{-01}
MET	-0.32	0.044	4.62×10^{-13}	-0.02	0.108	8.34×10^{-01}
SERPINA5	0.30	0.046	5.62×10^{-11}	-0.08	0.112	4.86×10^{-01}
KIT	-0.29	0.045	1.72×10^{-10}	-0.15	0.108	1.64×10^{-01}
THBS4	0.25	0.044	1.99×10^{-08}	0.33	0.103	1.39×10^{-03}
PLTP	-0.23	0.044	1.64×10^{-07}	-0.01	0.107	9.16×10^{-01}
PROC	0.22	0.044	6.19×10^{-07}	0.05	0.105	6.62×10^{-01}
FCN2	0.22	0.045	6.22×10^{-07}	0.41	0.106	1.23×10^{-04}
ICAM1	0.22	0.045	1.65×10^{-06}	0.24	0.112	3.38×10^{-02}
CNDP1	0.22	0.047	3.33×10^{-06}	-0.16	0.107	1.34×10^{-01}
MEGF9	0.20	0.043	4.52×10^{-06}	-0.22	0.099	2.64×10^{-02}
LILRB1	0.20	0.044	8.02×10^{-06}	0.27	0.101	7.78×10^{-03}
TGFBI	0.19	0.045	2.03×10^{-05}	0.42	0.110	1.31×10^{-04}
PTPRS	-0.18	0.043	2.28×10^{-05}	0.11	0.101	2.80×10^{-01}
CHL1	-0.18	0.043	3.03×10^{-05}	0.31	0.121	1.05×10^{-02}
SERPINA7	0.18	0.047	1.16×10^{-04}	-0.07	0.105	5.16×10^{-01}

TGFBR3	-0.16	0.042	1.30×10^{-04}	-0.06	0.106	5.89×10^{-01}
REG1A	-0.17	0.046	2.21×10^{-04}	-0.12	0.113	3.03×10^{-01}

Note: **A** and **B** are sorted based on Bonferroni-corrected P-value association of proteins with FLD and nominal P-value with liver fibrosis. Model 1 is adjusted only for age and sex and Model 2 is further adjusted for RBCs, granulocyte, lymphocyte, monocyte, platelets, body mass index, smoking status, alcohol consumption and type 2 diabetes. The Bonferroni-corrected significance threshold is $P < 2.92\times10^{-4}$ (0.05/171 proteins) for hepatic steatosis, and the nominal association with $P < 0.05$ for liver fibrosis. Abbreviations: FLD, fatty liver disease; SE, standard error.

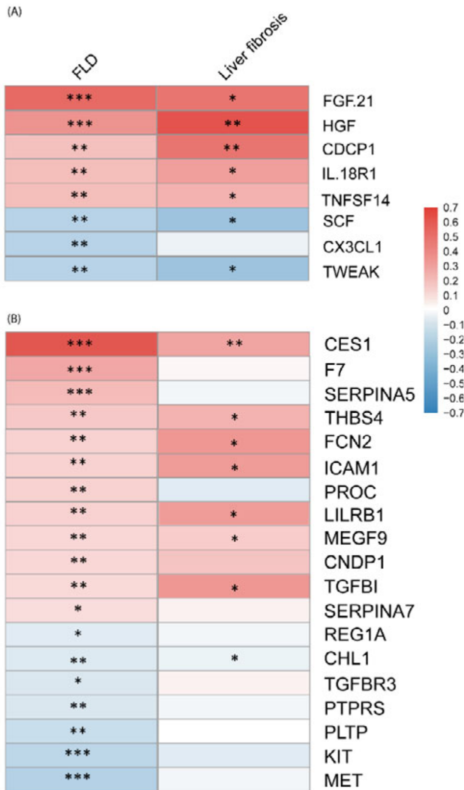


Figure 2. Heatmap visualization of protein profiles of FLD and liver fibrosis in (A) Inflammatory-related protein, (B) Cardiometabolic-related protein based on the multi-logistic regression model. Model 2 is adjusted for age, sex, blood cells count (RBCs, granulocyte, lymphocyte, monocyte, platelets), body mass index, smoking status, alcohol consumption and type2 diabetes. The color in columns represent the standardized effect estimates (betas) red color– positive relationship, blue color– negative relationship. *Significant relationship of proteins with FLD after adjusting for Bonferroni-correction ($P < 2.92\times10^{-4}$) and with liver fibrosis at the P-value <0.05. Abbreviations: NAFLD, nonalcoholic fatty liver disease; MAFLD, Metabolic associated fatty liver disease. P-value< 10^{-5}(*), P-value between 10^{-5}-10^{-9} (**), and P-value> 10^{-9} (***).

Association of plasma proteins with MAFLD and NAFLD

In our sensitivity analysis, consistent results were obtained when MAFLD replaced FLD. Additionally, despite a drop in the sample size, after replacing FLD with NAFLD the result remained significant ($P<2.92\times10^{-4}$) in all 8 inflammatory proteins while the result for 4 of the 19 cardiometabolic-related proteins lost the statistical significance (**Table S4** and **S5)**. It should be noted that despite the loss of statistical significance, similar direction of effects were observed in these 4 proteins as well.

Pathway analysis for proteins associated with FLD

We then performed the Reactome-pathways analysis, which aims to extract information on proteins through data mining and literature enrichment analysis. We arranged the pathways by fold-enrichment values and selected for further analysis with fold changes >9.5. We found TWEAK and TNFS14 proteins to be involved in the non-canonical NF-kB pathway **(Table S6A).** Additionally, F7, SERPINA5 and PROC proteins were involved in the Formation of Fibrin Clot (Clotting Cascade) **(Table S6B).**

Validation of associations in liver organoids

Additionally, we aimed to validate the results regarding protein expression and FLD using our in-vitro data on gene expression frequency in 3D human liver-derived organoids. Of the investigated proteins, 25 out of 27 were tested in the liver organoids and 16 could be detected in our model. Interestingly, the expression levels of IL-18R1 and CES1 were upregulated in the organoids mimicking FLD. The other inflammatory and cardiometabolic-related proteins could not be replicated in this FLD model **(Table S7**).

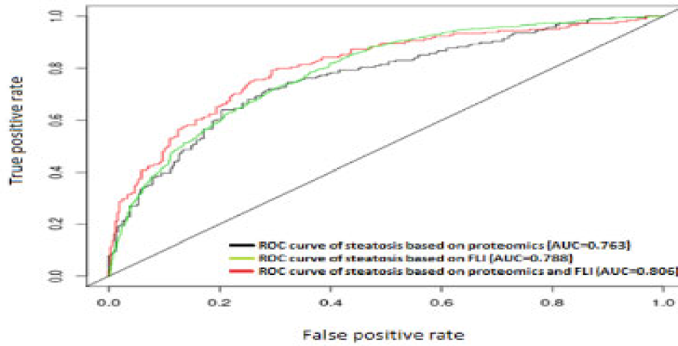
Plasma proteins and their potential as FLD biomarkers

Finally, we performed AUC-analysis to test if the identified proteins have the biomarker potential and could improve the prediction model of FLD. All prediction models demonstrated fair to good diagnostic performance for FLD: AUC 0.763 (95%CI 0.745 – 0.781) for proteomics alone, 0.788 (95%CI 0.771 – 0.805) for FLI alone and 0.806 (95%CI 0.790 – 0.822) for FLI combined with proteomics. Although proteomics alone had inferior diagnostic accuracy compared to the FLI, adding proteomics to the FLI algorithm slightly improved the AUC from 0.788 to 0.806 **Figure 3A**, but this was not statistical significant. Regarding fibrosis, proteomics alone (AUC 0.758, 95%CI 0.720 – 0.796) significantly outperformed the FIB-4 (AUC 0.634, 95%CI 0.584 – 0.684) and combining proteomics with FIB-4 yielded the best results (AUC 0.779, 95%CI 0.743 – 0.815 [**Figure 3B**]), but not significantly better than the proteomics alone.

In order to link our findings to protein cascades, we also performed pathway analysis. These results indicated that TWEAK and TNFSF14 belong to the non-canonical NF-κB signaling pathway, which regulates inflammation and host immune response.[21] In an animal model, both fibroblast growth factor-

inducible 14 (Fn14) and TWEAK in knockout mice and wild-type mice showed proliferation in hepatocytes and cholangiocytes after partial hepatectomy, indicating that the proteins signal is essential for mouse liver regeneration after hepatectomy.[21, 22] It has also been shown that TNFSF14, which increases adipose tissue inflammation, responds through its interaction with HVEM (herpes simplex virus glycoprotein D for herpes virus entry mediator).[23]

(A)



(B)

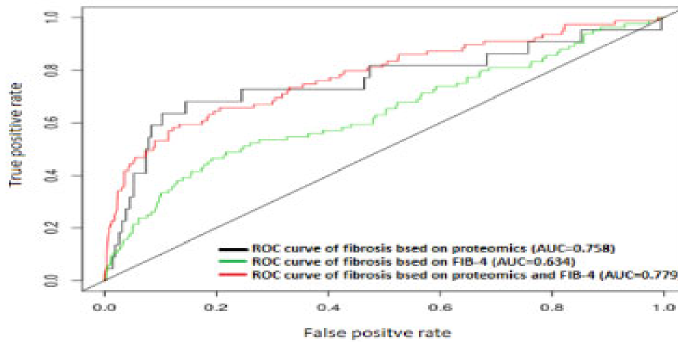


Figure 3. Area under the receiver operator characteristics (AUC-ROC curve) of each prediction model for predicting fatty liver disease **(A)** and fibrosis **(B)**.

Discussion

We investigated the association between 171 plasma proteins and FLD among a large population-based cohort study. We found 8 inflammatory and 19 cardiometabolic-related proteins to be significantly associated with FLD, of which FGF-21, CES1, and IL-18R1 will be discussed to a greater extent here, while the remaining proteins are briefly discussed in **Table S8.** Several of the proteins associated with FLD were also associated with fibrosis. These findings are important for further unraveling of the pathophysiology of FLD and the progression towards fibrosis. Moreover, these findings imply that plasma protein levels may have the potential to be used as biomarkers.

Higher plasma level of FGF-21 was associated with increased risk of FLD and liver fibrosis, and therefore may play a crucial role in the pathophysiology of disease and its progression towards fibrosis. FGF-21 is known for its role in many metabolic processes, including insulin sensitivity, glucose and lipid metabolism, and energy homeostasis.[24] Key aspects are increased with GLUT1 expression through stimulation of FGF-21, which induces glucose absorption in adipocytes.[25] Despite we could not validate this finding in liver organoids (due to not being able to reach the detection limit in FLD mimicking organoids and controls), it is supported by various previous studies that explored the association of plasma inflammatory and cardiometabolic proteins with FLD. For example, the expression of FGF-21 was an independent predictor of NAFLD and plasma levels of FGF-21 reflect fat accumulation and dysregulation of metabolic pathways in the liver.[26, 27] Interestingly, FGF-21 has been shown previously to be associated with fatty liver disease progression. For example, FGF-21 was related to lobular inflammation in patients with NAFLD, independent of steatosis grade and BMI.[28] Moreover, serum FGF-21 expression increased with the severity of liver fibrosis.[29] Overall, there is emerging evidence showing that FGF-21 is an important link to FLD and fibrosis. Our study highlights the relevance of these concepts in the general population.

Another important protein is CES1, for which upregulation was significantly associated with a higher risk of FLD and fibrosis. In liver organoids, we also demonstrated upregulation of CES1 in the models mimicking FLD, aligning our main outcomes. CES1 is involved in catalyzing the hydrolysis of esters, amides, thioesters, and carbamates, through which CES1 is important in drug metabolism.[30] Despite few studies being available that investigated the correlation between CES1 protein and FLD, the expression of *CES1* gene has been linked to FLD in humans.[31] This finding aligns with the results obtained in mouse models that demonstrated CES1 might be a potential treatment target for hyperlipidemia and NAFLD management.[32] Further research is warranted to unravel the role of CES1 in the pathogenesis of NAFLD and fibrosis and whether it could improve current non-invasive assessment of fibrosis.

Moreover, we found that upregulation of IL-18R1 is associated with FLD and fibrosis, which was further validated in our experimental study with liver organoids. IL-18R1 specifically binds interleukin 18 (IL-18) and is essential for IL-18-mediated signal transduction.[33] IL-18 is an important inflammatory cytokine which initiates a signaling complex by binding to the IL-18 alpha chain (IL-18Rα) and co-receptor, termed IL-18 receptor beta chain (IL-18Rβ), forming a high-affinity complex.[34] Many studies have primarily focused on the role of ligand (IL-18) in FLD rather than the receptor, however, a previous study in humans found that IL-18R1 was significantly correlated with a specific pattern of fat depots in the liver.[35] Moreover, in an animal model, IL-18R-dependent signaling was identified as a modulator of early liver damage among subjects with fatty liver disease. Interestingly, this could already be detected preceding the development of histologic NASH through activation of the NLR family pyrin domain containing- 3 (NLRP3).[36] Together with IL-12, IL-18 may cause impairment of the hepatic microcirculation, creating an ischemic state in the liver, and inhibiting oxidative phosphorylation.[37] It has also been shown that IL-18 could predict advanced FLD in children.[38] Validation of these results is warranted and further investigation is required on the interactions between IL-18 and its receptor (IL-18R1) in FLD and their potential as disease biomarkers.

In order to link our findings to protein cascades, we also performed pathway analysis. These results indicated that TWEAK and TNFSF14 belong to the non-canonical NF-κB signaling pathway, which regulates inflammation and host immune response.[21] In an animal model, both fibroblast growth factor-inducible 14 (Fn14) and TWEAK in knockout mice and wild-type mice showed proliferation in hepatocytes and cholangiocytes after partial hepatectomy, indicating that the proteins signal is essential for mouse liver regeneration after hepatectomy.[21, 22] It has also been shown that TNFSF14, which increases adipose tissue inflammation, responds through its interaction with HVEM (herpes simplex virus glycoprotein D for herpes virus entry mediator).[23]

Moreover, we found F7, SERPINA5, and PROC to be involved in Fibrin Clot Formation pathways. Among them, SERPINA5 protein is involved in the common Fibrin Clot Formation pathways through inactive complexes with protein C inhibitor that inhibits thrombin in complex with thrombomodulin.[39] F7 is a vitamin K-dependent glycoprotein formed by the liver, which can initiate the extrinsic pathway of blood coagulation.[40] The vitamin K-dependent protein C (PROC) pathway exerts its anticoagulant circulation properties as a proenzyme to an anticoagulant serine protease and is activated by thrombin.[41] Although clearly related to the liver, these identified pathways warrant further explanation as it remains unclear how exactly these protein cascades are linked to FLD and disease progression.

Finally, our AUC-analyses illustrated that proteomics has the potential to improve currently available non-invasive markers (FLI and FIB-4) in the prediction of FLD or fibrosis. Although we identified a slight improvement in the prediction models, the illustrated potential of proteomics in non-invasive tests aligns with previous studies in small cohorts.[9, 10]

This study has several important strengths, including the large sample size, use of proximity extension assay technology and availability of liver stiffness data in addition to ultrasound, yet the following limitations need to be considered for the interpretation of our findings. First, collecting data on FLD was 5.54 [p25-p75: 5.4-5.7] years after the plasma sample collection for the proteomic assessment. However, the impact on our results might be limited, as the time gap would result in 'dilution towards the null' and would make it increasingly difficult to demonstrate associations. Nonetheless, this may have resulted in missing out on some proteins associated with FLD and fibrosis. Reassuringly, an additional analysis using an FLI-based definition of FLD could confirm 25 out of 27 proteins that were identified in the main analysis. Second, we used surrogate markers for fatty liver disease and fibrosis, since it is unethical to perform liver biopsy in apparently healthy volunteers exposing them to risk of serious complications. Third, our results warrant validation in future studies as only few studies with different study designs are available to relate to our current findings. Aiming at validation we used liver organoids, however, not all identified proteins could be validated. This may partially be attributed to differences in protein assessment, but also due to the fact that liver organoids cannot entirely mimic the total protein expression in humans as was measured in our cohort study. Despite the differences, it was reassuring that CES1 and IL-18R1 were validated providing strong evidence that these proteins play an important role in FLD and its progression to fibrosis. Fourth, due to the nature of this study, we could not address causality. Lastly, this study did not contain longitudinal analysis because multiple measurements of the proteomics and outcome were not available in the follow-up visits.

In conclusion, we demonstrate that plasma levels of several inflammatory and cardiometabolic proteins are significantly associated with fatty liver disease and fibrosis at a population level. Our findings may help better understanding of the pathophysiology of FLD and the disease progression into fibrosis. In particular FGF-21, CES1 and IL-18R1 seem to contribute to the pathogenesis of FLD and also fibrosis. Moreover, given an unmet need for accurate and widely available non-invasive tools for risk stratification and referral strategies, our findings may improve the algorithms currently under development.

References

1. **Le MH**, **Yeo YH**, Li X, Li J, Zou B, Wu Y, et al. 2019 Global NAFLD Prevalence: A Systematic Review and Meta-analysis. Clinical Gastroenterology and Hepatology. 2021;Epub ahead of print.
2. Wong RJ, Ahmed A. Obesity and non-alcoholic fatty liver disease: Disparate associations among Asian populations. World journal of hepatology. 2014;6(5):263.
3. Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, et al. The diagnosis and management of nonalcoholic fatty liver disease: practice guidance from the American Association for the Study of Liver Diseases. Hepatology. 2018;67(1):328-57.
4. Monetti M, Levin MC, Watt MJ, Sajan MP, Marmor S, Hubbard BK, et al. Dissociation of hepatic steatosis and insulin resistance in mice overexpressing DGAT in the liver. Cell metabolism. 2007;6(1):69-78.
5. Angulo P. Nonalcoholic fatty liver disease. New England Journal of Medicine. 2002;346(16):1221-31.
6. Younossi ZM, Blissett D, Blissett R, Henry L, Stepanova M, Younossi Y, et al. The economic and clinical burden of nonalcoholic fatty liver disease in the United States and Europe. Hepatology. 2016;64(5):1577-86.
7. Lazarus JV, Mark HE, Anstee QM, Arab JP, Batterham RL, Castera L, et al. Advancing the global public health agenda for NAFLD: a consensus statement. Nature Reviews Gastroenterology & Hepatology. 2022;19(1):60-78.
8. Niu L, Thiele M, Geyer PE, Rasmussen DN, Webel HE, Santos A, et al. Noninvasive proteomic biomarkers for alcohol-related liver disease. Nature Medicine. 2022:1-11.
9. Corey KE, Pitts R, Lai M, Loureiro J, Masia R, Osganian SA, et al. ADAMTSL2 protein and a soluble biomarker signature identify at-risk non-alcoholic steatohepatitis and fibrosis in adults with NAFLD. Journal of Hepatology. 2022;76(1):25-33.
10. Luo Y, Wadhawan S, Greenfield A, Decato BE, Oseini AM, Collen R, et al. SOMAscan proteomics identifies serum biomarkers associated with liver fibrosis in patients with NASH. Hepatology Communications. 2021;5(5):760-73.
11. Ikram MA, Brusselle G, Ghanbari M, Goedegebure A, Ikram MK, Kavousi M, et al. Objectives, design and main findings until 2020 from the Rotterdam Study. European journal of epidemiology. 2020;35(5):483-517.
12. Hamaguchi M, Kojima T, Itoh Y, Harano Y, Fujii K, Nakajima T, et al. The severity of ultrasonographic findings in nonalcoholic fatty liver disease reflects the metabolic syndrome and visceral fat accumulation. Official journal of the American College of Gastroenterology| ACG. 2007;102(12):2708-15.
13. Eslam M, Newsome PN, Sarin SK, Anstee QM, Targher G, Romero-Gomez M, et al. A new definition for metabolic dysfunction-associated fatty liver disease: An international expert consensus statement. Journal of hepatology. 2020;73(1):202-9.
14. Boursier J, Vergniol J, Guillet A, Hiriart J-B, Lannes A, Le Bail B, et al. Diagnostic accuracy and prognostic significance of blood fibrosis tests and liver stiffness measurement by FibroScan in non-alcoholic fatty liver disease. Journal of hepatology. 2016;65(3):570-8.
15. Roulot D, Costes JL, Buyck JF, Warzocha U, Gambier N, Czernichow S, et al. Transient elastography as a screening tool for liver fibrosis and cirrhosis in a community-based population aged over 45 years. Gut. 2011;60(7):977-84.
16. PEA – a high-multiplex immunoassay technology with qPCR or NGS readout, 2020, [Available from: https://www.olink.com/content/uploads/2021/09/olink-white-paper-pea-a-high-multiplex-immunoassay-technology-with-qpcr-or-ngs-readout-v1.0.pdf.
17. Szilagyi IA, Vallerga CL, Boer CG, Schiphof D, Ikram MA, Bierma-Zeinstra SMA, et al. Plasma proteomics identifies CRTAC1 as a biomarker for osteoarthritis severity and progression. Rheumatology (Oxford). 2022.
18. World Health O. Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: report of a WHO/IDF consultation. 2006.
19. Wang L, Li M, Yu B, Shi S, Liu J, Zhang R, et al. Recapitulating lipid accumulation and related metabolic dysregulation in human liver-derived organoids. Journal of Molecular Medicine. 2022;100(3):471-84.

20. Bedogni G, Bellentani S, Miglioli L, Masutti F, Passalacqua M, Castiglione A, et al. The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population. BMC gastroenterology. 2006;6(1):1-7.
21. Chen Q, Lu X, Zhang X. Noncanonical NF-κB signaling pathway in liver diseases. Journal of clinical and translational hepatology. 2021;9(1):81.
22. Karaca G, Swiderska-Syn M, Xie G, Syn W-K, Krüger L, Machado MV, et al. TWEAK/Fn14 signaling is required for liver regeneration after partial hepatectomy in mice. PloS one. 2014;9(1):e83987.
23. Kim H-M, Jeong C-S, Choi H-S, Kawada T, Yu R. LIGHT/TNFSF14 enhances adipose tissue inflammatory responses through its interaction with HVEM. FEBS letters. 2011;585(3):579-84.
24. Tucker B, Li H, Long X, Rye K-A, Ong KL. Fibroblast growth factor 21 in non-alcoholic fatty liver disease. Metabolism. 2019;101:153994.
25. Kharitonenkov A, Shiyanova TL, Koester A, Ford AM, Micanovic R, Galbreath EJ, et al. FGF-21 as a novel metabolic regulator. The Journal of clinical investigation. 2005;115(6):1627-35.
26. Rusli F, Deelen J, Andriyani E, Boekschoten MV, Lute C, van den Akker EB, et al. Fibroblast growth factor 21 reflects liver fat accumulation and dysregulation of signalling pathways in the liver of C57BL/6J mice. Scientific reports. 2016;6(1):1-16.
27. Li H, Dong K, Fang Q, Hou X, Zhou M, Bao Y, et al. High serum level of fibroblast growth factor 21 is an independent predictor of non-alcoholic fatty liver disease: a 3-year prospective study in China. Journal of hepatology. 2013;58(3):557-63.
28. Shen J, Chan HL-Y, Wong GL-H, Choi PC-L, Chan AW-H, Chan H-Y, et al. Non-invasive diagnosis of non-alcoholic steatohepatitis by combined serum biomarkers. Journal of hepatology. 2012;56(6):1363-70.
29. Li D, Lu T, Shen C, Liu Y, Zhang J, Shan Y, et al. Expression of fibroblast growth factor 21 in patients with biliary atresia. Cytokine. 2016;83:13-8.
30. Casey Laizure S, Herring V, Hu Z, Witbrodt K, Parker RB. The role of human carboxylesterases in drug metabolism: have we overlooked their importance? Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy. 2013;33(2):210-22.
31. Ashla AA, Hoshikawa Y, Tsuchiya H, Hashiguchi K, Enjoji M, Nakamuta M, et al. Genetic analysis of expression profile involved in retinoid metabolism in non-alcoholic fatty liver disease. Hepatology Research. 2010;40(6):594-604.
32. Lian J, Bahitham W, Panigrahi R, Nelson R, Li L, Watts R, et al. Genetic variation in human carboxylesterase CES1 confers resistance to hepatic steatosis. Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids. 2018;1863(7):688-99.
33. Guo F, Wu S. Antisense IRAK-1 oligonucleotide blocks activation of NF-κB and AP-1 induced by IL-18. Immunopharmacology. 2000;49(3):241-6.
34. Dinarello CA, Novick D, Kim S, Kaplanski G. Interleukin-18 and IL-18 binding protein. Frontiers in immunology. 2013;4:289.
35. Lovric A, Granér M, Bjornson E, Arif M, Benfeitas R, Nyman K, et al. Characterization of different fat depots in NAFLD using inflammation-associated proteome, lipidome and metabolome. Scientific reports. 2018;8(1):1-14.
36. Hohenester S, Kanitz V, Schiergens T, Einer C, Nagel J, Wimmer R, et al. IL-18 but not IL-1 signaling is pivotal for the initiation of liver injury in murine non-alcoholic fatty liver disease. International journal of molecular sciences. 2020;21(22):8602.
37. Kaneda M, Kashiwamura S-I, Ueda H, Sawada K, Sugihara A, Terada N, et al. Inflammatory liver steatosis caused by IL-12 and IL-18. Journal of interferon & cytokine research. 2003;23(3):155-62.
38. Flisiak-Jackiewicz M, Bobrus-Chociej A, Tarasów E, Wojtkowska M, Białokoz-Kalinowska I, Lebensztejn DM. Predictive role of interleukin-18 in liver steatosis in obese children. Canadian Journal of Gastroenterology and Hepatology. 2018;2018.
39. Davie EW, Fujikawa K, Kisiel W. The coagulation cascade: initiation, maintenance, and regulation. Biochemistry. 1991;30(43):10363-70.
40. Manzano ML, Arocena C, Tomas JF, Hedner U, Quiroga JA, Carreno V. Measurement of the procoagulant activity of factor VII in patients with liver cirrhosis and normal prothrombin activity: evaluation of the bleeding risk. Blood coagulation & fibrinolysis. 2000;11:S95-S9.

41. Dahlbäck B, Villoutreix BO. The anticoagulant protein C pathway. FEBS letters. 2005;579(15):3310-6.

Chapter 4

Metabolomics of fatty liver disease

Chapter 4.1

Circulating metabolites associated with liver enzymes and fatty liver disease: a comprehensive multi-platform populations-based study

Yasir J. Abozaid, Ibrahim Ayada, Laurens A. van Kleef, Neil J Goulding, Jessica S. Williams-Nguyen, Robert C Kaplan, Robert J. de Knegt, Lynne E. Wagenknecht, Nicholette D. Palmer, Deborah A. Lawlor, Nicholas J Timpson, Jill M. Norris, Yii-Der Ida Chen, M. Arfan Ikram, Willem Pieter Brouwer, Mohsen Ghanbari

(In submission)

Abstract

Background: Non-alcoholic fatty liver disease (NAFLD) is the most common chronic liver disease strongly associated with metabolic dysfunction. However, its pathogenesis remains incompletely understood. Exploring circulating metabolites may help in elucidating underlying mechanisms and identifying new biomarkers for NAFLD.

Methods: We examined the cross-sectional association between plasma metabolites and NAFLD as well as liver enzymes using data from four population-based cohort studies (Rotterdam study, ALSPAC, IRASFS, and SOL). Metabolites on the Nightingale platform (225 metabolites) were assessed by NMR spectroscopy and on the Metabolon platform (1111 metabolites) by UHPLC-mass spectrometry. Serum levels of liver enzymes (ALT, AST, GGT) were measured and NAFLD was diagnosed by ultrasound or CT-scan. Logistic and linear regression models were performed per cohort and meta-analyzed to identify metabolites associated with liver enzymes and NAFLD. A false discovery rate (FDR) < 0.05 was considered statistically as significant threshold.

Results: Several metabolites were significantly associated with NAFLD and liver enzymes. These include 21 metabolites that were overlapping associated with both conditions. The most important were phenylalanine, triglycerides in (HDL, IDL, and small LDL), fatty acid (FA) ratios of (18:2 linoleic acid-to-total FA, omega 6 FA-to-total FA, and polyunsaturated FA-to-total FA) from the Nightingale platform. Moreover, glutamate and sphingomyelin (d18:0/18:0, d19:0/17:0) showed the strongest association with both NAFLD and liver enzymes on the Metabolon platform. Other associated metabolites were mainly involved in lipid, amino acid, carbohydrates and peptide metabolism.

Conclusions: Our study identified a landscape of circulating metabolites associated with NAFLD. These metabolites will contribute to a better understanding of the metabolic pathways underlying NAFLD and may hold promising potential for biomarkers aiding early diagnosis and monitoring of the disease in clinical practice.

Introduction

Nonalcoholic fatty liver disease (NAFLD) is the most common chronic liver disease, with a prevalence of over 30% of the global population, and generally represents the hepatic manifestation of the metabolic syndrome.[1, 2] This could progress to non-alcoholic steatohepatitis (NASH) in approximately 20-25% of patients.[3] This could further lead to liver cirrhosis, hepatocellular carcinoma development, and eventually liver transplantation. Among patients listed for liver transplantation, NAFLD, and specifically NASH, nowadays is the most rapidly increasing indications for transplantation.[4]

NAFLD is characterized by the accumulation of lipids, mainly triacylglycerol, in hepatocytes.[5] Hepatic lipid metabolism is regulated by a combination of the uptake and export of fatty acids, de novo lipogenesis, and fat utilization by β-oxidation.[6] When there is a disruption in the balance between these pathways, it can result in the accumulation of hepatic lipids, leading to NASH reflected by elevated liver enzymes.[7] This disruption also triggers prolonged activation of inflammatory and fibrotic pathways, potentially advancing to more severe stages of liver disease. Currently, evidence shows that NAFLD is strongly associated with metabolic risk factors such as obesity, insulin resistance and dyslipidemia.[8, 9] The molecular mechanisms underlying NAFLD pathophysiology are complex, multifactorial and still incompletely understood. As a consequence, therapeutic options are predominantly limited to lifestyle interventions, with no currently approved medications available.[10, 11] Nevertheless, ample novel compounds for the treatment of NASH are currently under investigation; yet, many failed to achieve the desired endpoints.

Given the lack of effective therapeutic options and the progressive detrimental natural course of NAFLD, it is of great importance to diagnose the disease at an early stage. Although several non-invasive tests are currently available to aid in the diagnosis of NAFLD, the gold standard for the diagnosis remains liver biopsy, which is hampered by its invasive nature.[12, 13] Moreover, important knowledge gaps remain with regard to the disease pathophysiology. Therefore, there is a great need to further unravel important underlying pathways which could help to the identification of non-invasive biomarkers for early detection and classification of NAFLD and may reveal new avenues for therapeutic development.

Recent metabolomics studies have shown promise in identifying metabolic signatures that may reveal the pathological underpinnings of complex diseases (i.e., the major impact of genetic variations and metabolic alterations), including NAFLD.[14, 15]

These studies have demonstrated several potential pathways which might contribute to the development of NAFLD, yet the majority of them were limited by either the small number of cases (typically<50)[16] or the number of metabolites investigated (typically100-200)[17]. To date, only a few of these metabolites have been validated at the population level and even very few have been suggested as biomarkers to be used in clinical practice.[18, 19]

Hence, this study aimed to investigate the cross-sectional association between an extensive array of circulating metabolites and NAFLD, as well as liver enzymes, using data from population-based cohort studies, which may help to elucidate the pathophysiological mechanisms of NAFLD and also for identification of potential biomarkers for early diagnosis.

Methods

Study population

Rotterdam study (RS)

The RS consists of 14,926 individuals aged 45 years or older from the Ommoord region of Rotterdam. It is a population-based study that comprises three distinct cohorts: RS-I (established in 1989), RS-II (established in 2000), and RS-III (established in 2006). The objectives and design of the Rotterdam Study have been published elsewhere.[20] The study includes multiple follow-up visits for each cohort. At each visit, participants completed questionnaires, underwent physical examinations, and provided fasting blood samples. To conduct this analysis, we included three distinct samples from different cohorts within the RS study, where plasma samples were assessed using the Nightingale platform. Sample 1 comprised participants from visit 4 of RS-I, designated as RS-I-4. Sample 2, referred to as RS-Bios in this manuscript, consisted of a combined sample of participants from visit 5 of RS-I (RS-I-5), visit 3 of RS-II (RS-II-3), and visit 2 of RS-III (RS-III-2). Sample 3 represented another independent set of participants specifically from RS-III-2.

Within both sample 2 and 3, 1594 participants had data on NAFLD based on abdominal ultrasound. Furthermore, plasma metabolites measured by the Metabolon platform were available in subsamples from the fourth visit of the first cohort, RS-I-4 (N=395), where NAFLD was measured only based on CT-scan, and the second visit of the third cohort, RS-III-2 (N=655), where NAFLD was measured based on abdominal ultrasound. A comprehensive illustration of the study approach and design is depicted in overview **Figure 1**.

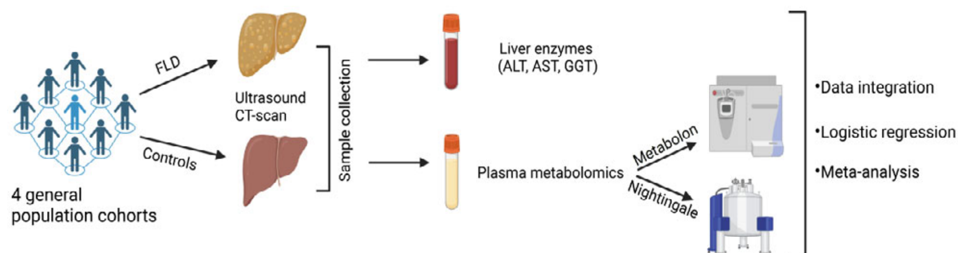


Figure 1 Graphical abstract and flowchart of the study participants per cohorts. FLD, fatty liver disease; US, ultrasound; CT, computed tomography; RS, Rotterdam Study; ALSPAC, Avon Longitudinal Study of Parents and Children; IRAFS (HA+AA), The Insulin Resistance Atherosclerosis Family Study (Hispanic American + African American); SOL, Study of Latinos ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma-glutamyl transferase; NMR, nuclear magnetic resonance; MS, mass spectrometry; N, sample sizes; metab., metabolites.
(*) indicate independent cohort of Rotterdam study used only in metabolites overlapping.

Avon Longitudinal Study of Parents and Children (ALSPAC)

The study invited pregnant women residing in Avon, UK, with expected delivery dates between April 1, 1991, and December 31, 1992, to participate. A total of 14,541 pregnancies were initially enrolled,

resulting in 14,676 fetuses. From these pregnancies, there were 14,062 live births, and at 1 year of age, 13,988 children were still alive. the design of the ALSPAC cohort have been described in details in supplementary data.

The Insulin Resistance Atherosclerosis Family Study (IRASFS)

A comprehensive description of the study design, recruitment process, and phenotyping in the Insulin Resistance Atherosclerosis Study Family Study (IRASFS) have been described in detail in elsewhere.[21] In summary, this multicenter study aimed to identify the genetic factors contributing to insulin resistance and adiposity. Further details regarding the design of the IRASFS are available in the supplementary data.

Study of Latinos (SOL)

The Hispanic Community Health Study/Study of Latinos (SOL) is a prospective cohort study conducted in four urban field centers across the United States. These centers were deliberately chosen to ensure diversity in terms of national background and behaviors, including diet. A total of 16,415 individuals who self-identified as having Hispanic and/or Latino backgrounds, including South Americans, Central Americans, Mexicans, Puerto Ricans, Cubans, and Dominicans, were recruited and participated in data collection between June 2008 and July 2011. the details of these cohort are described in the supplementary data.

Assessment of plasma levels of metabolites

Metabolomics in the Rotterdam study (Nightingale and Metabolon)

The metabolites levels in 1594 participants of the Rotterdam Study (RS-Bios and RS-III-2) were quantified using the high-throughput 1H-NMR Nightingale platform (Nightingale Ltd., Helsinki, Finland), after the collection of fasting EDTA plasma samples. The RS-Bios and RS-III-2 cohorts are independent of each other meaning that the participants are not overlapping and the assessment of metabolites was done in different years. The metabolite profiling was done as part of the 4th Rainbow Project of BioBanking for Medical Research Infrastructure of the Netherlands (BBMRI-NL) (https://www.bbmri.nl/omics-metabolomics/). All samples were stored at -80 °C, which ensured biological stability. Details on this methodology have been described previously.[22-24] In total 225 plasma metabolites were quantified, which belong to lipid fractions, including various sizes of lipoproteins, cholesterol levels, glycerides, lipids particles, as well as fatty acids, branched-chain amino-acids, glycolysis-related metabolites, ketone bodies, and metabolites related to inflammation and glycolysis. To ensure normal distribution of the metabolites per cohort, all metabolites belonging to the Nightingale platform with skewed distribution were transformed by natural logarithm, or rank transformation was applied before the analysis. The metabolite measurements were then scaled to standard deviation units

(mean 0, SD 1) to enable the comparison of results for measures with different units. The missing data of all variables were excluded from the statistical analysis.

Moreover, we profiled blood metabolites of 1082 participants of RS-I-4 using the untargeted Metabolon HD4 platform. The Metabolon platform includes 1387 metabolites of different biochemical pathways (including Lipids, amino acids, xenobiotics, nucleotides, cofactors and vitamins, peptides, carbohydrates, energy-related metabolites, and uncharacterized metabolites). We performed preprocessing of the metabolomics data before the analyses. To this end, we first excluded 14 participants which had missingness greater than 5 times the standard deviation (SD) of the mean missingness in overall participants. Then, we excluded metabolites with missingness 5 times SD of the mean missingness in metabolites, and coefficient of variance (CV) greater than 30% in internal control samples (NIST Standard Reference Material). Subsequently, we log2 transformed the values of remaining metabolites (N = 1111). We imputed the missing data with the lowest limit of detection. Before performing the imputation, we further removed 120 metabolites with missingness greater than 30% in metabolites.

Metabolomics in ALSPAC (Nightingale platform)

Over 220 quantified metabolomics measures were obtained per fasted sample of EDTA-plasma/serum, using a 1D proton (1H) NMR spectroscopy-based platform described previously.[22, 25] All samples were stored at -80 °C. To be comparable with results from the Rotterdam study, metabolite values were transformed by natural logarithm and scaled to standard deviation units (mean 0, SD 1).

Metabolomics in IRASFS (Metabolon platform)

Metabolite profiling was performed on stored (at -80 °C) fasting plasma samples collected at the 1999–2002 baseline survey. Metabolite detection and quantification was conducted by Metabolon, Inc. (Durham, North Carolina) using untargeted liquid chromatography-mass spectroscopy (MS) (DiscoveryHD4 panel). The method of metabolites level measurement has been described in details in supplementary data.

Metabolomics in SOL (Metabolon platform)

Fasting serum samples were collected, processed, and stored at −70 °C from the time of collection until metabolic profiling. Metabolites were quantified using an untargeted liquid chromatography-mass spectrometry (MS)-based metabolomics quantification protocol at Metabolon (Durham, NC, USA) using the Discovery HD4 platform. A comprehensive description of the methodology used to measure the levels of metabolites can be found in the supplementary data.

Assessment of NAFLD by abdominal ultrasound and CT scan

Abdominal ultrasonography performed by trained technicians on Hitachi HI VISION 900 in all study participants of the Rotterdam Study. The diagnosis of fatty liver was determined by the ultrasound technician according to the protocol by Hamaguchi et al.[26] In short, the diagnosis of liver steatosis is based on ultrasonographic liver brightness, and hepatorenal echo contrast. NAFLD is diagnosed by the presence of hepatic steatosis and the absence of the following secondary causes of fatty liver: (1) excessive alcohol consumption (>30 g/day for men and >20 g/day for women), (2) presence of viral hepatitis, (3) use of steatogenic agents, and (4) recent bariatric surgery.

Moreover, from February 2002 onwards trained technicians performed CT-scan in the Rotterdam Study participants. Both 16-slice and 64-slice (N=695) multi-detector CT scanner (Somatom Sensation 16 or 64, Siemens, Forchheim, Germany) was used to perform non-enhanced CT scanning. Detailed information regarding the imaging parameters of the scans is provided elsewhere.[27] Using a cardiac CT scan, we evaluated the liver fat content using liver attenuation (LA) using a standardized procedure. First, we placed three circular regions of interest (in cm^2) in the liver and calculated the mean Hounsfield unit (HU) within these regions [28]. These regions of interest were drawn throughout the imaged liver tissue (including both the left and right liver lobes) and were carefully chosen to include only liver tissue, and no disruptive tissue such as large blood vessels, cysts, or focal lesions. Next, we calculated the mean HU value from these three measurements as a marker of the total liver fat content.[28] Lower LA means a higher risk of fatty liver, so we categorized these variables as well using a threshold of 40HU. Each value below 40HU was considered to be positive for NAFLD while equal to or above the threshold of 40HU was included in the control group. All measurements were done using Philips iSite Enterprise software (Royal Philips Electronics NV 2006) and described in detail elsewhere.[29] The assessment of NAFLD in IRASF was performed using the same CT scan approach, more information are provided in the supplementary data.

Assessment of liver stiffness and NASH (Rotterdam Study)

Liver stiffness measurement (LSM) was conducted utilizing transient elastography with the FibroScan device (Echosens, Paris, France). The M or XL probe was used to obtain a minimum of 10 measurements. If the interquartile range exceeded 30%, measurements surpassing 7.1 kPa were deemed unreliable and excluded.[30] Fibrosis was defined as a reliable liver stiffness measurement of ≥8.0 kPa.[31] Furthermore, an additional sensitivity analysis was performed, using a positive NAFLD diagnosis based on ultrasound findings and elevated serum alanine aminotransferase levels ≥ 35 IU/L as an approximation NAFLD related steatohepatitis (NASH).[32]

Assessment of liver enzymes

In the Rotterdam Study, blood samples were collected while participants were fasting. Serum aspartate aminotransferase (AST), alanine aminotransferase (ALT), and gamma-glutamyl transpeptidase (GGT)

levels were determined within two weeks after collection using a Merck Diagnostica kit (Merck, Whitehouse Station, NJ, USA) on an Elan Auto-analyzer (Merck). According to local cut-offs, the elevation of GGT was defined as >34 U/L for women and >49 U/L for men. Elevation of ALT was defined as >30 U/L for women and >40 U/L for men, and elevation of AST was defined as >30 U/L for women and >36 U/L for men.[33] Serum levels of AST, ALT and GGT in other cohorts (ALSPAC, IRASFS, and SOL) were Measured similarly, more information are provided in the supplementary data.

Definitions of covariates

Potential confounders were considered to be any factors that were known or plausible causal factors for metabolites (or hypothesized exposures) and NAD/liver enzymes (hypothesized outcomes): age, sex, smoking, alcohol, BMI and lipid lowering medication.

In all four cohorts, information on smoking status (three categories: never, ex, current), and alcohol consumption (measured in grams of ethanol per day and then categorized to alcohol use yes/no) were obtained through questionnaire.[20] The height and weight of participants were measured while standing up without shoes and heavy outer garments. BMI (kg/m^2) was calculated as weight divided by squared height in meters. Information regarding the use of lipid-lowering medication (LLM) was obtained from pharmacy records and home interviews. All the covariates included in the statistical analyses were obtained at baseline, at the same time point as the blood collection for metabolites assessment. The covariates definition of other cohorts (ALSPAC, and SOL) are provided in the supplementary data.

Statistical analyses

Per cohort and within the Nightingale platform, the relationship between 225 metabolites and NAFLD was assessed by multivariable logistic regression models. Natural log transformed metabolites were used as exposure and the dichotomous NAFLD diagnosed by ultrasound was used as an outcome in the logistic regression models. In model 1, we adjusted for age and sex. In model 2, we additionally adjusted for other confounders, i.e. for the lipid-lowering medication (yes, no), alcohol consumption (yes, no), body mass index (kg/m^2), smoking status (current, ex-, or never smokers). To allow comparison between the effect sizes in model 1 and 2, missing values from any covariate were excluded. Regarding the Rotterdam Study samples, the results of two sub-cohorts (RS-III-2 and RS-Bios) were meta-analyzed using fixed-effect models in “METAL” software.[34]

Additionally, we performed multivariable logistic regression to investigate the association between metabolites in the Metabolon platform and NAFLD diagnosed by CT-scan (dichotomous data) and Hounsfield unit (HU) measurement (continues data). The number of common metabolites included in the final meta-analysis among the two cohorts (RS and IRASFS) was 671.

In the Rotterdam Study, a sensitivity analysis was conducted where the outcome variable was replaced with liver stiffness (categorical) and NASH to examine the relationship between the metabolites and NAFLD as a spectrum disease.

Multivariable linear regression models were used to examine the association of metabolites in Nightingale and Metabolon platforms with liver enzymes (GGT, ALT, and AST), using the same model 1 and model 2 adjustments and comparisons as described above (**Figure 1**).

Lastly, the results of metabolomics analysis with liver enzymes from the two RS subcohorts were meta-analyzed with the results from ALSPAC, IRASFS and SOL cohorts using inverse-variance weighted fixed-effect models. In all pooled results, a false discovery rate corrected P<0.05 (5%) was used as a significant threshold. The effect of heterogeneity was estimated by I^2, which describes the percentage of the total variation in the meta-analysis attributable to study heterogeneity.[35] All remaining analyses were performed in SPSS statistical software (SPSS, version 25; IBM Crop) and R software version 3.5.2 (The R Foundation for Statistical Computing). The metabolic pathway analysis was conducted using the MetaboAnalyst (V5.0) software (https://www.metaboanalyst.ca/home.xhtml) to check the pathways related to the identified metabolites (Metabolon platform) in both ultrasound and CT scan based NAFLD.

Results

Descriptive characteristics of cohorts

The baseline characteristics of the study participants in different cohorts are presented in **Table 1A** and **1B**. Compared to the Rotterdam Study, ALSPAC participants were younger and had a higher current smokers 28.3%. The prevalence of NAFLD among the Rotterdam Study subcohorts was 29.3% in RS-Bios and 26.0% in RS-III-2. Notably, the Rotterdam Study subcohorts had higher BMI compared to the ALSPAC participants, likely as a result of the age differences between the participating cohorts, (i.e. 62.85, 24.5, 43.37, and 45.82 years, respectively in RS, ALSPAC, IRASFS, and SOL). Comparing the cohorts that have used a CT scan-based NAFLD diagnosis, the NAFLD prevalence was 6.1% in the RS cohort, 4.5% within the IRASFS-AA and 15.5% in the IRASFS-HA (as shown in **Table 1B)**.

Metabolites associated with NAFLD

Through a meta-analysis of the data from the Rotterdam Study sub-cohorts (RS-Bios and RS-III-2), we found 100 out of 225 metabolites as well as their ratios to be significantly associated with NAFLD, after adjusting for multiple testing (false discovery rate; FDR< 0.05; **Table S1**). While, in the Rotterdam Study (RS-I-4), where NAFLD was diagnosed using CT-scan, we identified a significant association with 51 metabolites (**Table S2-S3**).

On the Metabolon platform, we found 81 out of 991 metabolites to be significantly associated with NAFLD diagnosed by ultrasound in the Rotterdam Study (**Table S4**). In the case of NAFLD diagnosed based on CT-scan, a meta-analysis combining the results from the Rotterdam Study with IRASFS revealed 110 metabolites to be significantly associated with NAFLD in dichotomous data (FDR<0.05) (**Table S5-S6**).

Table 1 Characteristics of study participants

(A) NAFLD-based on ultrasound and liver enzymes per cohort in the metabolomics analysis with the Nightingale platform.

	RS-Bios-US			RS-III-2-US			RS-Bios (Liver enzymes)	RS-III-2 (Liver enzymes)	ALSPAC (Liver enzymes)
Characteristic	**All**	**NAFLD cases**	**Controls**	**All**	**NAFLD cases**	**Controls**	**All**	**All**	**All**
Number	663	29.3%(194)	70.7%(469)	931	26.0%(242)	74.0%(689)	545	1476	2687
Age, years	68.31±5.7	68.95±5.6	68.05±5.7	62.06±5.2	62.33±5.2	61.96±5.3	68.39±3.7	62.75±5.8	24.5±0.8
Females, %(N)	58.07(385)	59.28(115)	57.87(270)	58.22(542)	59.09(143)	57.91(399)	58.8(267)	57.7(852)	59.1(1588)
Alcohol consumption, %(N)	86.43(573)	87.63(170)	85.93(403)	87.54(815)	85.54(207)	88.24(608)	85.0(386)	88.4(1305)	96.1(2581)
Lipid lowering medication, %(N)	31.83(211)	34.54(67)	30.70(144)	20.30(189)	26.45(64)	18.14(125)	32.8(149)	22.1(326)	NA
Smoking status, % (N)									
Never smokers	33.33(221)	30.41(59)	34.54(162)	35.34(329)	33.88(82)	35.85(247)	31.5(143)	36.1(533)	36.5(981)
Ex-smokers	57.16(379)	62.37(121)	55.01(258)	51.34(478)	54.55(132)	50.22(346)	59.0(268)	50.2(741)	35.2(947)
Current smokers	9.50(63)	7.22(14)	10.45(49)	13.32(124)	11.57(28)	13.93(96)	9.5(43)	13.7(202)	28.3(759)
Body mass index (kg/m2)	27.75±4.2	30.07±4.0	26.79±3.9	27.37±4.4	30.05±4.6	26.43±4.0	27.76±4.2	27.4±4.5	24.7±4.8
Laboratory data									
AST (U/L)	26.15±10.4	27.43±13.5	25.6±8.7	25.41±14.8	26.09±9.2	25.17±16.3	26.85±12.4	25.7±21.1	27.2±23.4
ALT (U/L)	20.9±11.1	24.1±11.9	19.54±10.5	22.4±14.0	24.7±14.0	21.57±13.9	20.96±12.3	22.6±17.7	25.8±22.7
GGT (U/L)	80.28±64.9	75.57±59.4	82.30±67.3	31.9±34.8	38.2±44.4	29.67±30.1	33.59±30.6	32.6±49.4	20.1±19.6

The table shows characteristics of the study participants within the Rotterdam Study with ultrasonography and the study participants in the RS and ALSPAC with liver enzymes data. Values are represented as mean ± SD), or sample sizes and percentages. Abbreviations: RS-Rotterdam Study; RS-Bios consists of RS-I-5, RS-II-3 and RS-III-2 sub-cohorts of the Rotterdam Study from the same visit with metabolomics data; RS-III-2, an independent Rotterdam sub-cohort with metabolomics data; ALSPAC- Avon Longitudinal Study of Parents and Children; NAFLD-Non-alcoholic fatty liver disease; BMI-body mass index; AST-Aspartate transaminase, ALT-Alanine transaminase; GGT- Gamma-glutamy.ltransferase.

(B) NAFLD-based on CT scan and liver enzymes per cohort in the metabolomics analysis with the Metabolon platform.

	RS-I-4 (CT-scan)			IRASFS-AA(CT-scan)			IRASFS-HA (CT-scan)			SOL (Liver enzymes)
Characteristic	**All**	**NAFLD cases**	**Controls**	**All**	**NAFLD cases**	**Controls**	**All**	**NAFLD cases**	**Controls**	**All**
Number	395	24	371	394	18	340	823	128	638	3833
Age, years	62.85±5.2	60.3±4.21	63.01±5.23	43.37±13.54	44.25±12.20	42.99±13.62	41.11±13.38	41.61±11.24	41.05±13.83	45.82 ±13.80
Females, N (%)	273 (69.1)	16(66.7)	265 (71.4)	242 (61.42)	13 (72.22)	210 (61.76)	513 (62.33)	67 (52.34)	414 (64.89)	2206 (57.6)
Alcohol consumption, N (%)	344 (87.1)	23 (95.8)	321 (86.5)	216 (54.82)	11 (61.11)	184 (54.12)	442 (53.71)	66 (51.56)	349 (54.70)	1884 (49.2)
Lipid lowering medication, N (%)	53 (13.3)	3 (12.5)	50 (13.5)	29 (7.36)	2 (11.11)	26 (7.65)	33 (4.01)	5 (3.91)	24 (3.76)	450 (11.7)
Smoking status, N (%)										
Never smokers	99 (25.1)	5 (20.8)	94 (25.3)	212 (53.81)	12 (66.67)	186 (54.71)	488 (59.30)	79 (61.72)	382 (59.87)	2252 (58.8)
Ex-smokers	259 (65.6)	16 (66.7)	243 (65.5)	92 (23.35)	3 (16.67)	80 (23.53)	151 (18.35)	25 (19.53)	109 (17.08)	766 (20.0)
Current smokers	37 (9.4)	3 (12.5)	34 (9.2)	90 (22.84)	3 (16.67)	74 (21.76)	184 (22.36)	24 (18.75)	147 (23.04)	815 (21.3)
Body mass index (kg/m2)	27.46±3.65	30.5±4.13	27.26±3.53	29.86±6.65	33.51±6.37	29.72±6.64	29.12±6.23	32.27±6.02	28.34±5.89	29.78 ±6.04
Mean liver attenuation (HU)	59.44±9.67	34.1±6.72	61.1±7.23	55.77±8.47	29.01±9.89	57.19±5.54	51.45±12.60	28.71±8.45	56.01±7.20	NA
Laboratory data										
AST (U/L)	NA	NA	NA	18.12±6.98	18.61±4.68	18.18±7.23	19.00±9.39	21.54±10.29	18.53±9.39	24.88±19.71
ALT (U/L)	NA	NA	NA	7.75±5.39	9.22±5.63	7.73±5.43	11.12±10.01	12.88±10.22	10.88±10.19	27.53±24.68
GGT (U/L)	30.92±27.4	32.67±18.03	27.6±16.5	37.40±38.50	35.67±14.50	37.53±40.57	38.28±39.57	46.41±47.40	36.23±38.34	34.92 ±55.79

The table shows characteristics of the study participants within the Rotterdam Study and IRASFS with CT-scan and the study participants in the RS, IRASFS and LOS with liver enzymes data. Values are represented as mean ± (SD), or sample sizes and percentages. Abbreviations: RS-Rotterdam Study; IRASFS-AA, Insulin Resistance Atherosclerosis Family Study-African American; IRASFS-HA, Insulin Resistance Atherosclerosis Family Study-Hispanic American; SOL, Study of Latinos; NAFLD-Non-alcoholic fatty liver disease; BMI-body mass index; NA-not applicable; AST-Aspartate transaminase, ALT-Alanine transaminase; GGT- Gamma-glutamyltransferase; N, sample size.

Metabolites associated with liver enzymes

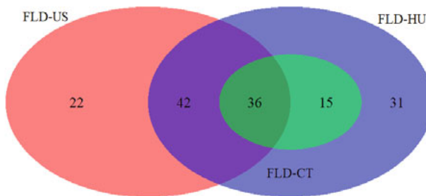
The analysis of the Nightingale platform results in the Rotterdam Study sub-cohorts, along with the ALSPAC replicates and the subsequent meta-analysis of all these cohorts, revealed that the number of metabolites associated with liver enzymes remained relatively consistent. Specifically, we observed a similar number of associated metabolites for ALT (N=107), AST (N=37), and GGT (N=118) (**Table S7-S9**).

In contrast, when analyzing the Metabolon results from the Rotterdam Study, IRASFS, and SOL in relation to liver enzymes, a broader range of associations was observed. The meta-analysis across these cohorts revealed that 148 metabolites were significantly associated with ALT, 197 metabolites were associated with AST, and 243 metabolites were associated with GGT (detailed in **Table S10-S12**).

Overlapping metabolites associated with NAFLD and liver enzymes

The metabolites identified in the Nightingale platform were examined for their association with the prevalence of NAFLD using ultrasound or CT-scan, as well as their relationship with liver enzymes (ALT, AST, and GGT). Through this analysis, we aimed to identify any overlapping metabolites among these different measures of liver function and disease. The comparison revealed 36 overlapping metabolites that were associated with NAFLD diagnosed by both US and CT-scan, as displayed **in Figure 2A**. Additionally, there were 16 overlapping metabolites observed among the different liver enzymes, as presented in **Figure 2B**. Considering all the different outcomes, including NAFLD based on US and CT-scan, ALT, AST, and GGT, a total of 7 metabolites were found to be overlapping in all traits. These 7 metabolites, namely Phenylalanine, Triglycerides in HDL, Triglycerides in small LDL, Triglycerides in IDL (positive association with NAFLD), Ratio of 18:2 linoleic acids to total fatty acid, Ratio of omega 6 fatty acid to total fatty acid, and Ratio of polyunsaturated fatty acid to total fatty acid (negative association with NAFLD) are shown in **Figure 3A** and **Table 2**.

A-

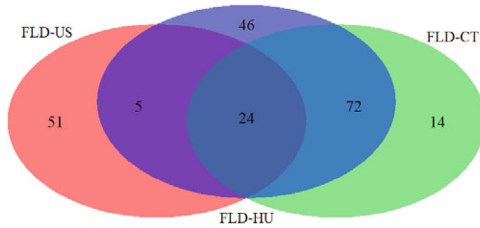


B-

Figure 2. Venn diagram showing the overlapping metabolites assessed by the Nightingale platform in (A) NAFLD diagnosing utilities, (B) liver enzymes. A Venn diagram in (A) showing the number of metabolites (FDR<0.05) that were significantly associated with NAFLD-US (n=100), NAFLD-CT scan (n=51), and NAFLD-HU (n=124) revealed 36 overlapping metabolites whose expression was up/down regulated. While in (B) showing the number of metabolites associated with ALT (n=107), AST (n=37), and GGT (n=118) overlapping revealed 16 metabolites. Abbreviation: NAFLD, non-alcoholic fatty liver disease; US, Ultrasound; CT scan, computed tomography scan, HU, Hounsfield unit; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma-glutamyl transferase; n, number of metabolite.

Similarly, we followed the same procedure for the Metabolon platform. In this case, we found 24 common metabolites that were associated with both US and CT-scan, as depicted in **Figure 2C**. Additionally, there were 87 metabolites that overlapped among the three liver enzymes, as shown in **Figure 2D**. Overall, there were 14 overlapping metabolites across all the different outcomes (NAFLD based on US and CT-scan, ALT, AST, and GGT), as illustrated in **Figure 3B** and **Table 3**. These metabolites mainly belong to the metabolism of lipids, peptides, amino acids, energy, and carbohydrates.

A-



B-

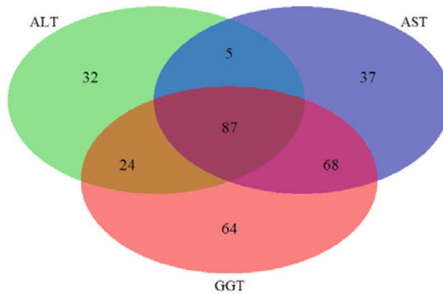


Figure 3. Venn diagram showing the overlapping of metabolites assessed by metabolon platform in (A) NAFLD diagnosing utilities, (B) liver enzymes. The diagram in (A) showing the number of metabolites (FDR<0.05) that that were significantly associated with NAFLD-US (n=81), NAFLD-CT scan (n=110), and NAFLD-HU (n=147) revealed 24 overlapping metabolites whose expression was up/down-regulated. While in (B) showing the number of metabolites associated with ALT (n=148), AST (n=197), and GGT (n=243) overlapping revealed 87 metabolites. Abbreviation: NAFLD, non-alcoholic fatty liver disease; US, Ultrasound; CT scan, computed tomography scan, HU, Hounsfield unit; ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma-glutamyl transferase; n, number of metabolite.

Table 2. Seven metabolites from the Nightingale platform significantly associated with NAFLD and liver enzymes

	ALT		AST		GGT		FLD-HU		FLD-CT		FLD-US	
Metabolites	**Beta**	**P-value**	**Beta**	**P-value**	**Beta**	**P-value**	**Beta**	**P-value**	**Beta**	**P-value**	**Beta**	**P-value**
Phenylalanine	0.954	1.45E-03	0.868	5.16E-03	1.303	1.91E-04	-2.192	6.78E-09	0.460443	1.13E-02	0.156	2.27E-02
Triglycerides in HDL	1.450	1.73E-06	0.791	9.50E-03	3.188	1.30E-20	-1.104	3.54E-03	0.432096	1.05E-02	0.250	1.09E-04
Triglycerides in small LDL	1.386	5.62E-06	0.816	7.90E-03	3.347	1.73E-22	-1.519	1.07E-04	0.471733	5.19E-03	0.290	5.59E-06
Triglycerides in IDL	1.824	7.08E-09	1.322	1.65E-05	3.325	6.16E-22	-0.985	9.33E-03	0.392824	2.47E-02	0.196	2.18E-03
Ratio of 18:2 linoleic acid to total fatty acid	-1.582	1.35E-06	-0.970	2.90E-03	-2.439	2.50E-12	2.780	1.81E-12	-1.22635	3.76E-06	-0.286	5.36E-05
Ratio of omega 6 fatty acid to total fatty acid	-1.822	1.69E-08	-0.809	1.20E-02	-3.181	7.22E-20	3.103	4.84E-18	-1.1617	1.57E-06	-0.431	5.78E-10
Ratio of polyunsaturated fatty acid to total fatty acid	-1.553	1.42E-06	-0.659	4.06E-02	-2.996	1.02E-17	3.096	3.79E-16	-1.14286	9.28E-07	-0.409	2.69E-09

The table shows 7 metabolites assessed by the Nightingale platform that were significantly associated and overlapped with all studied traits (incl. NAFLD diagnosed by Ultrasound and CT-scan and three liver enzymes). FLD-US shows the combined results of the two Rotterdam study sub-cohorts (RS-Bios and RS-III-2). The results for liver enzymes are after meta-analysis of RS-III-2, RS-Bios,RS-1-4 and ALSPAC cohorts. The table shows the results of Model 2: adjusted for age, sex, lipid lowering medication, alcohol consumption, body mass index, and smoking status. The significant P-value set at < 0.05 (FDR). Abbreviations: ALT: Alanine aminotransferase; GGT: Gamma-glutamyl Transferase; AST: Aspartate aminotransferase; CT, Computed tomography; HU, Hounsfield unit; US, Ultra-sound, HDL: High-density lipoprotein; LDL: low-density lipoprotein; IDL: intermediate-density lipoprotein

Table 3. Fourteen metabolites from the Metabolon platform significantly associated with NAFLD and liver enzymes

	ALT		AST		GGT		FLD-CT		FLD-HU		FLD-US	
Metabolites	**Beta**	**P-value**	**Beta**	**P-value**	**Beta**	**P-value**	**Beta**	**P-value**	**Beta**	**P-value**	**Beta**	**P-value**
1-myristoyl-2-palmitoyl-GPC (14:0/16:0)	0.034	3.01E-06	0.019	6.32E-06	0.056	6.11E-16	0.149	9.14E-03	-1.441	2.17E-05	0.596	4.76E-02
1-palmitoyl-2-arachidonoyl-GPI (16:0/20:4)	0.042	2.04E-08	0.019	8.12E-06	0.053	2.62E-14	0.173	1.58E-03	-1.273	1.88E-04	0.879	4.48E-03
1-palmitoyl-2-linoleoyl-GPE (16:0/18:2)	0.031	6.30E-05	0.011	1.37E-02	0.044	1.26E-09	0.168	4.21E-03	-1.549	5.40E-06	0.633	2.46E-02
1-stearoyl-2-linoleoyl-GPE (18:0/18:2)	0.042	1.47E-08	0.012	5.81E-03	0.055	2.62E-14	0.159	5.03E-03	-1.571	4.92E-06	0.868	2.55E-03
1-stearoyl-2-oleoyl-GPE (18:0/18:1)	0.051	9.05E-12	0.016	5.07E-04	0.066	6.00E-20	0.196	8.44E-04	-1.925	1.94E-08	0.896	8.28E-04
3-methyl-2-oxovalerate	0.041	6.25E-06	0.012	3.35E-02	0.047	5.96E-08	0.269	9.64E-05	-2.142	8.84E-08	1.399	4.35E-03
4-methyl-2-oxopentanoate	0.050	2.28E-08	0.014	9.34E-03	0.050	7.99E-09	0.231	8.29E-04	-1.813	4.25E-06	1.111	2.85E-02
aconitate [cis or trans]	0.042	7.83E-08	0.041	3.52E-21	0.050	3.16E-11	0.143	1.76E-02	-1.007	1.47E-02	1.305	1.08E-02
gamma-glutamylisoleucine	0.028	2.38E-03	0.013	1.09E-02	0.059	1.19E-12	0.166	1.19E-02	-1.259	2.67E-03	1.294	2.60E-03
gamma-glutamylleucine	0.033	4.07E-04	0.016	3.33E-03	0.058	3.05E-11	0.205	2.21E-03	-1.561	2.43E-04	1.330	1.37E-02
gamma-glutamylvaline	0.037	1.30E-05	0.020	4.80E-05	0.071	8.09E-20	0.155	1.19E-02	-1.431	1.44E-04	0.857	3.10E-02
glutamate	0.059	4.28E-13	0.032	1.17E-12	0.101	1.37E-44	0.160	1.15E-02	-1.893	2.47E-07	1.438	3.27E-05
lactate	0.022	8.36E-03	0.014	2.35E-03	0.045	4.06E-10	0.188	9.97E-04	-1.660	1.84E-06	0.970	3.52E-03
sphingomyelin (d18:0/18:0, d19:0/17:0)	0.070	4.38E-19	0.033	4.55E-13	0.078	9.03E-24	0.198	9.87E-04	-2.323	2.34E-10	0.846	2.52E-03

The table shows 14 metabolites assessed by the Metabolon platform that were significantly associated and overlapped among all the studied traits (incl. NAFLD diagnosed by Ultrasound and CT-scan and three liver enzymes). The results of NAFLD diagnosed by CT-Scan and liver enzymes are after meta-analysis of metabolomics data from the RS, IRASFS-AA, IRASFS-HA and SOL cohorts. The results are based on Model 2: adjusted for age, sex, lipid lowering medication, alcohol consumption, body mass index, and smoking status. The significant P-value set at < 0.05 (FDR). Abbreviations: ALT: Alanine aminotransferase; GGT: Gamma-glutamyl Transferase; AST: Aspartate aminotransferase; CT, Computed tomography; HU, Hounsfield unit.

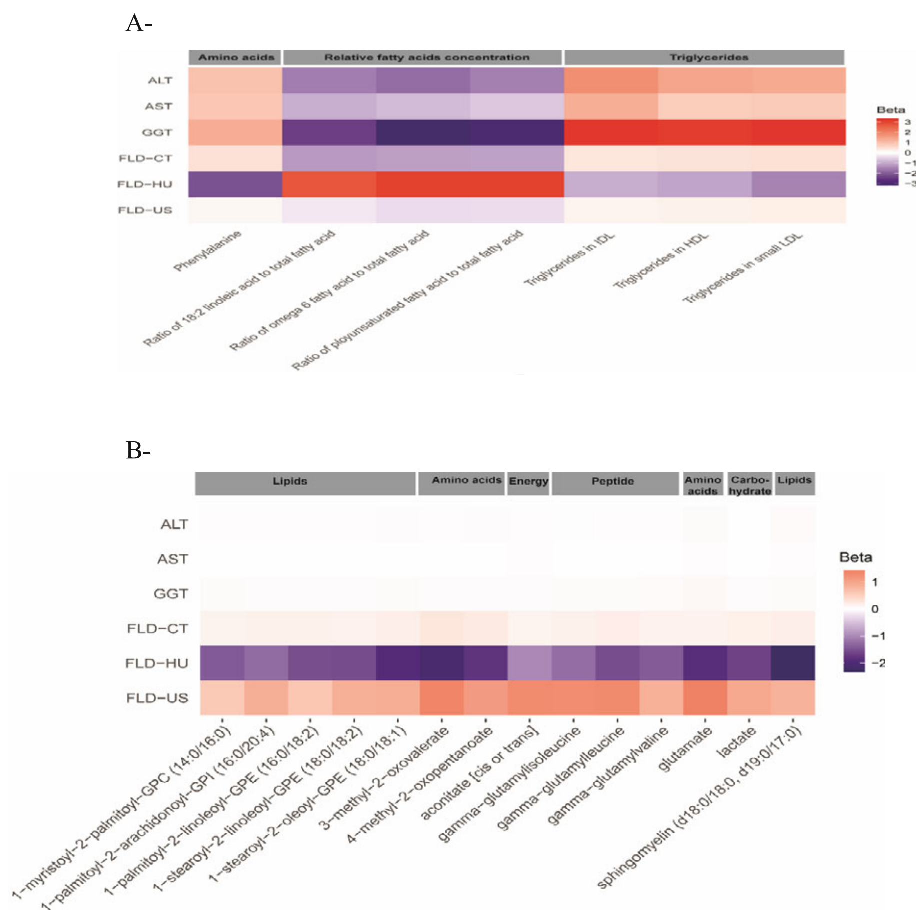


Figure 4. Heatmap visualization of metabolites profiles that were significantly associated and overlapping with NAFLD based on either US or CT scan and liver enzymes (A) in the Nightingale platform, and (B) Metabolon platform. The color in columns represent the standardized effect estimates (betas) red color– positive relationship, blue color– negative relationship. The significant threshold is corrected for a false discovery rate (FDR) adjusted P<0.05. Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma-glutamyl transferase; FLD, fatty liver disease; HU, Hounsfield unit; US, Ultra-sound; HDL, high-density lipoprotein; LDL, low-density lipoprotein; IDL, intermediate-density lipoprotein.

Metabolites associated with fibrosis and NASH

In our sensitivity analysis within the Rotterdam Study, we observed that among the 255 metabolites assessed using the Nightingale platform, 53 metabolites were nominally associated with liver fibrosis (defined as ≥ 8.0 kPa on a categorical scale), while 103 metabolites were associated with NASH. We further investigated the overlap between the metabolites associated with NAFLD diagnosed by ultrasound and those associated with both liver fibrosis and NASH. Our findings revealed that 40 metabolites were common among all three traits (**Table S13**).

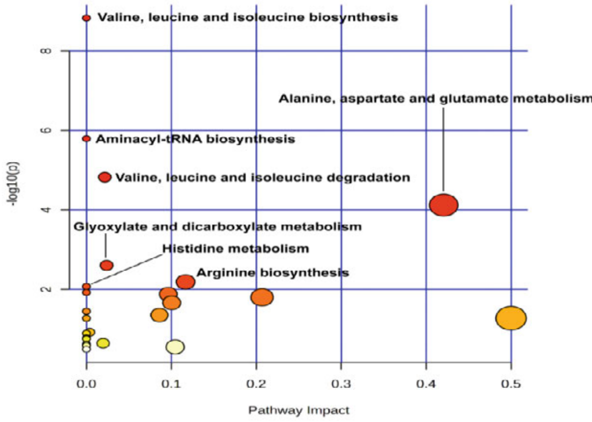
Similarly, we explored the association between metabolites measured using the Metabolon platform and both liver fibrosis and NASH. Our analysis revealed that out of the 991 metabolites examined, 118 were associated with NASH, while 40 metabolites showed an association with liver fibrosis. Among these, three metabolites were found to overlap between NAFLD based on ultrasound and NASH, while two metabolites overlapped with liver fibrosis. Additionally, we observed that three metabolites were common to both NASH and liver fibrosis, while no metabolites overlapped among all three traits (**Table S14**).

Metabolic pathway analysis

We conducted metabolic pathway analysis (MetPA) based on the identified overlapping metabolites (in both NAFLD diagnosed by ultrasound and CT scan (N=24), and NAFLD diagnosed by ultrasound and CT scan, and liver enzymes (N=14)) and the results are shown in **Fig 5A** and **B**.

The identified metabolites were significantly enriched in specific metabolic pathways, including biosynthesis and degradation of the amino acids valine, leucine, and isoleucine. Another important pathway was related to metabolic reactions involved in the synthesis, utilization and/or degradation of glyoxylate and dicarboxylate. Finally, we found D-Glutamine and D-glutamate metabolism, glycosylphosphatidylinositol (GPI)-anchor biosynthesis, and arginine biosynthesis (**Fig 5B**). While for common metabolites that were associated and overlapping with NAFLD based on ultrasound and CT scan, we found the identified metabolites to be enriched in metabolic pathways such as aminacyl-tRNA biosynthesis, alanine, aspartate, glutamate, and histidine metabolism, in addition to the biosynthesis and degradation of the amino acids valine, leucine and isoleucine, arginine biosynthesis, and histidine metabolism **(Fig 5A)**.

A-



B-

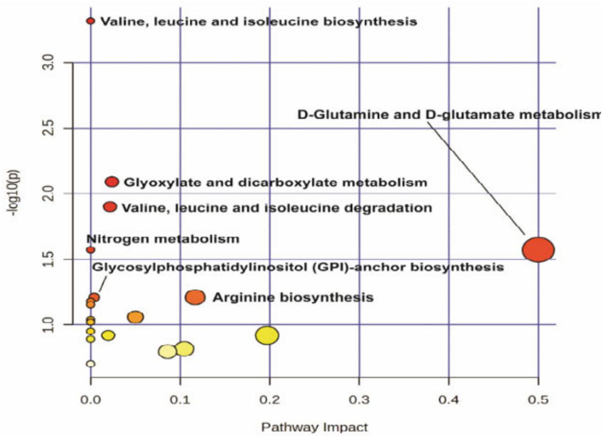


Figure 5. Pathway analysis of the identified metabolites in the Metabolon platform using the over-representation method. (A) Metabolites associated with NAFLD-based CT scan and ultrasound (N=24) and (B) Metabolites associated with all traits including NAFLD-based CT scan and ultrasound as well as liver enzymes (N=14). In this metabolome figure, each circle represents a pathway; circle size and color shade are based on the pathway impact and *p*-value (red circle indicates significant and yellow circle indicates not significant), respectively. The Human KEGG pathway library was used. Abbreviation: NAFLD, Nonalcoholic fatty liver disease.

Discussion

In this study, we explored the association between a wide array of circulating metabolites evaluated through two established platforms (Nightingale and Metabolon), with NAFLD and liver enzymes across four distinct population-based settings. We found a range of circulating metabolites deriving from lipids, fatty acids, and amino acids, that could reflect the presence of fatty liver disease independently of other metabolic risk factors. Subsequently, we compared these identified metabolites with those associated with three liver enzymes, aiming to identify the most significant metabolites implicated in liver steatosis pathogenesis. Within discussion of our findings, we highlight 8 out of the 21 overlapping metabolites associated with all the studied traits, while the remaining 13 metabolites- mainly involved in lipid, amino acid, and carbohydrate metabolism- are briefly discussed in **Table S15**.

Metabolomics studies, especially within the field of hepatology, represent an important and rapidly evolving area of research which has the potential to disclose new insights into the pathophysiology of NAFLD. By analyzing metabolic profiles, it becomes possible to identify metabolic dysfunctions that precede the diagnosis of NAFLD. This notion finds support in a significant population-based study conducted in Finland[36] that the researchers through a longitudinal study noted similar alterations in plasma metabolic profiles up to a decade before the identification of fatty liver. Moreover, within the same study, cross-sectional analysis revealed correlations among a multitude of metabolites and NAFLD, including notable associations with triglyceride particles within small LDL. Importantly, our research outcomes align closely with the findings of this study. Previous studies have also shown the association of several metabolites, including triglycerides, with NAFLD and suggested their potential to be used as potential biomarkers for the disease, but few of them have been validated.[17, 37] Dysregulated lipid metabolism in NAFLD leads to the formation of small, dense LDL particles characterized by increased triglycerides.[38] These altered LDL particles are more prone to oxidative modification and glycation, eventually promoting inflammation and endothelial dysfunction, which are key factors in cardiovascular disease development. Moreover, a study found that TG is the strongest predictor for NAFLD, as compared to other markers of metabolic dysfunction such as HDL, LDL, and serum glucose.[39] Likewise, previous study has shown that LDL particles are associated positively with both fatty liver and liver enzymes.[40]

On the other hand, and in line with our results, emerging evidence suggests that specific fatty acids, particularly omega 6 fatty acids (= n–6 polyunsaturated fatty acids) and linoleic acid, may play a protective role in the development of NAFLD.[41] This observation is reinforced by another study which showed that individuals with steatosis exhibited lower blood levels of omega 6 and omega 3 fatty acids than those with normal liver function. Finally, it's noteworthy that omega 6 fatty acids demonstrated stronger inverse associations with NAFLD compared to omega 3 fatty acids.[36, 42] This evidence provides an insight on the distribution of fatty acids in the onset of NAFLD.

Compared to the commonly used Nightingale metabolomics platform, diverse metabolites derived from the Metabolon platform have been studied to a lesser extent in the context of NAFLD. Here, we found a multitude of metabolites from various metabolic pathways within the Metabolon platform that exhibit significant associations with both NAFLD and liver enzymes. These metabolites are primarily involved in lipids, amino acids, peptides, and carbohydrate metabolism. Similar to our findings, Lind et al. identified several metabolites, including 3-methyl-2-oxovalerate, that showed a positive correlation with liver fat based on proton density fat fraction.[43] Additionally, other studies have demonstrated that alterations in amino acid concentrations, such as gamma-glutamyl leucine and glutamate, are positively linked with NAFLD.[44-46] Our analysis also revealed the positive associations of glycolysis-related metabolites such as lactate which previously reflected the presence of NAFLD.[40] This finding was supported by a previous study, which established higher lactate and triglyceride levels as associated with hepatocellular carcinoma (HCC) in cirrhotic livers.[47] Explicit evidence shows that sphingolipid metabolism is altered in the course of NAFLD and these changes might contribute to NAFLD progression.[48] A previous study reviewed that some sphingolipid species, such as ceramides, may have the potential as biomarkers for NAFLD.[49] Altogether, these findings illustrate how multiple pathways in systemic metabolism are troubled before the development of NAFLD. This insight may aid the prevention or progression of NAFLD to its severe stages.

Finally, our pathway analysis showed connections between various identified metabolites and multiple pathways. Predominantly, these pathways are centered around amino acid metabolism, particularly valine, leucine, and isoleucine biosynthesis and degradation, as well as alanine, aspartate, and glutamate metabolism. Another major pathway identified in our study is that of the lipid metabolism. Specifically metabolism of glyoxylate, dicarboxylate, and glycosylphosphatidylinositol (GPI)-anchor biosynthesis, are also associated with NAFLD irrespective of the extent of obesity and insulin resistance.

Elevated plasma levels of branched-chain amino acids, including valine, leucine, and isoleucine as confirmed in our study, have been previously linked to intra-hepatic fat accumulation. These findings suggest a potential connection between hepatic insulin resistance and NAFLD.[50] Previous studies have noted that increased plasma amino acid concentrations are most commonly observed in individuals with obesity and NAFLD, possibly due to heightened insulin resistance and protein catabolism.[44] Moreover, the genome-scale metabolic modeling of the human gut microbiome has highlighted alterations in glyoxylate and dicarboxylate metabolism within metabolic disorders.[51] Similarly, pathways associated with lipids, including glycosylphosphatidylinositol, have been found to exhibit associations with changes in metabolites related to NAFLD.[52]

Our study has several strengths. First and foremost, it draws upon relatively large sample size derived from different population-based cohort studies (RS, ALSPAC, IRASFS, and SOL). Second, the study was conducted through the quantification of a vast number of metabolites utilizing two high-throughput widely

used metabolomics platforms: Nightingale and Metabolon, known for their non-destructive, fast, and generating highly reproducible outcomes.[53] Third, we incorporated liver enzyme parameters alongside NAFLD data in our investigation, aiming to provide supplementary evidence for the observed associations. Nonetheless, this study has some limitations that must be considered. First, NAFLD based-ultrasound was highly prevalent (26%) within the Rotterdam Study. Due to the unavailability of additional cohorts for replication, our focus was solely directed towards conducting a meta-analysis using the two subcohorts of the Rotterdam Study. Second, the included cohorts are not entirely comparable. The difference in age between cohorts particularly ALSPAC cohort might cause distortion within the metabolite profiles. Third, given that ultrasonography is not highly sensitive to detect mild steatosis,[54] we additionally used CT-scan based NAFLD diagnosis data in the Rotterdam Study as well as meta analyzing this with the IRASFS cohort. Finally, in a cross-sectional observational study, the ability to assess causality is limited. Therefore, future studies warrant to assess the relationship between plasma levels of identified metabolites and the risk of NAFLD development longitudinally by incorporating additional independent cohorts, preferably from different ethnic groups.

In conclusion, this population-based study demonstrates that plasma levels of several metabolites are significantly associated with liver function and disease. Our findings may help to elucidate metabolic pathways involved in NAFLD pathogenesis, and the identified metabolites might be considered as potential biomarkers for early diagnosis of the disease. Future studies are yet to be conducted to further validate our findings and more importantly, with a longitudinal design to elucidate whether metabolites may be used as diagnostic or prognostic biomarkers for NAFLD.

References

1. Fotbolcu H, Zorlu E. Nonalcoholic fatty liver disease as a multi-systemic disease. World journal of gastroenterology. 2016;22(16):4079.
2. Le MH, Yeo YH, Li X, Li J, Zou B, Wu Y, et al. 2019 Global NAFLD Prevalence: A Systematic Review and Meta-analysis. Clin Gastroenterol Hepatol. 2022;20(12):2809-17 e28.
3. Peng C, Stewart AG, Woodman OL, Ritchie RH, Qin CX. Non-alcoholic steatohepatitis: a review of its mechanism, models and medical treatments. Frontiers in Pharmacology. 2020;11:603926.
4. Burra P, Becchetti C, Germani G. NAFLD and liver transplantation: disease burden, current management and future challenges. JHEP reports. 2020;2(6):100192.
5. Kawano Y, Cohen DE. Mechanisms of hepatic triglyceride accumulation in non-alcoholic fatty liver disease. Journal of gastroenterology. 2013;48(4):434-41.
6. Ipsen DH, Lykkesfeldt J, Tveden-Nyborg P. Molecular mechanisms of hepatic lipid accumulation in non-alcoholic fatty liver disease. Cellular and molecular life sciences. 2018;75:3313-27.
7. Badmus OO, Hillhouse SA, Anderson CD, Hinds Jr TD, Stec DE. Molecular mechanisms of metabolic associated fatty liver disease (MAFLD): Functional analysis of lipid metabolism pathways. Clinical Science. 2022;136(18):1347-66.
8. Speliotes EK, Massaro JM, Hoffmann U, Vasan RS, Meigs JB, Sahani DV, et al. Fatty liver is associated with dyslipidemia and dysglycemia independent of visceral fat: the Framingham Heart Study. Hepatology. 2010;51(6):1979-87.
9. Hamaguchi M, Kojima T, Takeda N, Nakagawa T, Taniguchi H, Fujii K, et al. The metabolic syndrome as a predictor of nonalcoholic fatty liver disease. Annals of internal medicine. 2005;143(10):722-8.
10. Cohen JC, Horton JD, Hobbs HH. Human fatty liver disease: old questions and new insights. Science. 2011;332(6037):1519-23.
11. Suomela E, Oikonen M, Virtanen J, Parkkola R, Jokinen E, Laitinen T, et al. Prevalence and determinants of fatty liver in normal-weight and overweight young adults. The Cardiovascular Risk in Young Finns Study. Annals of Medicine. 2015;47(1):40-6.
12. Shackel NA, McCaughan GW. Liver biopsy: is it still relevant? : Wiley Online Library; 2006. p. 689-91.
13. Barrera-Saldaña HA, Fernández-Garza LE, Barrera-Barrera SA. Liquid biopsy in chronic liver disease. Annals of Hepatology. 2021;20:100197.
14. Zhou Y, Orešič M, Leivonen M, Gopalacharyulu P, Hyysalo J, Arola J, et al. Noninvasive detection of nonalcoholic steatohepatitis using clinical markers and circulating levels of lipids and metabolites. Clinical Gastroenterology and Hepatology. 2016;14(10):1463-72. e6.
15. Luukkonen PK, Zhou Y, Sädevirta S, Leivonen M, Arola J, Orešič M, et al. Hepatic ceramides dissociate steatosis and insulin resistance in patients with non-alcoholic fatty liver disease. Journal of hepatology. 2016;64(5):1167-75.
16. Craven L, Rahman A, Parvathy SN, Beaton M, Silverman J, Qumosani K, et al. Allogenic fecal microbiota transplantation in patients with nonalcoholic fatty liver disease improves abnormal small intestinal permeability: a randomized control trial. Official journal of the American College of Gastroenterology| ACG. 2020;115(7):1055-65.
17. Alferink LJM, Radjabzadeh D, Erler NS, Vojinovic D, Medina-Gomez C, Uitterlinden AG, et al. Microbiomics, Metabolomics, Predicted Metagenomics, and Hepatic Steatosis in a Population-Based Study of 1,355 Adults. Hepatology. 2021;73(3):968-82.
18. Pang Y, Kartsonaki C, Lv J, Millwood IY, Fairhurst-Hunter Z, Turnbull I, et al. Adiposity, metabolomic biomarkers, and risk of nonalcoholic fatty liver disease: a case-cohort study. The American Journal of Clinical Nutrition. 2022;115(3):799-810.
19. Gorden DL, Myers DS, Ivanova PT, Fahy E, Maurya MR, Gupta S, et al. Biomarkers of NAFLD progression: A lipidomics approach to an epidemic1 [S]. Journal of lipid research. 2015;56(3):722-36.
20. Ikram MA, Brusselle G, Ghanbari M, Goedegebure A, Ikram MK, Kavousi M, et al. Objectives, design and main findings until 2020 from the Rotterdam Study. European journal of epidemiology. 2020;35(5):483-517.
21. Henkin L, Bergman RN, Bowden DW, Ellsworth DL, Haffner SM, Langefeld CD, et al. Genetic epidemiology of insulin resistance and visceral adiposity: the IRAS Family Study design and methods. Annals of epidemiology. 2003;13(4):211-7.

22. Soininen P, Kangas AJ, Würtz P, Suna T, Ala-Korpela M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. Circulation: cardiovascular genetics. 2015;8(1):192-206.
23. Kettunen J, Demirkan A, Würtz P, Draisma HHM, Haller T, Rawal R, et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. Nature communications. 2016;7(1):1-9.
24. Kujala UM, Mäkinen V-P, Heinonen I, Soininen P, Kangas AJ, Leskinen TH, et al. Long-term leisure-time physical activity and serum metabolome. Circulation. 2013;127(3):340-8.
25. Soininen P, Kangas AJ, Würtz P, Tukiainen T, Tynkkynen T, Laatikainen R, et al. High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. Analyst. 2009;134(9):1781-5.
26. Hamaguchi M, Kojima T, Itoh Y, Harano Y, Fujii K, Nakajima T, et al. The severity of ultrasonographic findings in nonalcoholic fatty liver disease reflects the metabolic syndrome and visceral fat accumulation. Official journal of the American College of Gastroenterology| ACG. 2007;102(12):2708-15.
27. Odink AE, van der Lugt A, Hofman A, Hunink MGM, Breteler MMB, Krestin GP, et al. Association between calcification in the coronary arteries, aortic arch and carotid arteries: the Rotterdam study. Atherosclerosis. 2007;193(2):408-13.
28. Kodama Y, Ng C, Wu T-T, Ayers G, Curley S, Abdalla E, et al. Comparison of CT methods for determining the fat content of the liver. 2007.
29. Wolff L, Bos D, Murad SD, Franco OH, Krestin GP, Hofman A, et al. Liver fat is related to cardiovascular risk factors and subclinical vascular disease: the Rotterdam Study. European Journal of Echocardiography. 2016;17(12):1361-7.
30. Boursier J, Vergniol J, Guillet A, Hiriart J-B, Lannes A, Le Bail B, et al. Diagnostic accuracy and prognostic significance of blood fibrosis tests and liver stiffness measurement by FibroScan in non-alcoholic fatty liver disease. Journal of hepatology. 2016;65(3):570-8.
31. Roulot D, Costes J-L, Buyck J-F, Warzocha U, Gambier N, Czernichow S, et al. Transient elastography as a screening tool for liver fibrosis and cirrhosis in a community-based population aged over 45 years. Gut. 2011;60(7):977-84.
32. Verma S, Jensen D, Hart J, Mohanty SR. Predictive value of ALT levels for non-alcoholic steatohepatitis (NASH) and advanced fibrosis in non-alcoholic fatty liver disease (NAFLD). Liver International. 2013;33(9):1398-405.
33. Koehler EM, Sanna D, Hansen BE, van Rooij FJ, Heeringa J, Hofman A, et al. Serum liver enzymes are associated with all-cause mortality in an elderly population. Liver International. 2014;34(2):296-304.
34. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190-1.
35. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Bmj. 2003;327(7414):557-60.
36. Kaikkonen JE, Würtz P, Suomela E, Lehtovirta M, Kangas AJ, Jula A, et al. 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):491-500.
37. Masoodi M, Gastaldelli A, Hyötyläinen T, Arretxe E, Alonso C, Gaggini M, et al. Metabolomics and lipidomics in NAFLD: biomarkers and non-invasive diagnostic tests. Nature Reviews Gastroenterology & Hepatology. 2021;18(12):835-56.
38. Deprince A, Haas JT, Staels B. Dysregulated lipid metabolism links NAFLD to cardiovascular disease. Molecular metabolism. 2020;42:101092.
39. Tomizawa M, Kawanabe Y, Shinozaki F, Sato S, Motoyoshi Y, Sugiyama T, et al. Triglyceride is strongly associated with nonalcoholic fatty liver disease among markers of hyperlipidemia and diabetes. Biomedical reports. 2014;2(5):633-6.
40. Hartley A, Ferreira DLS, Anderson EL, Lawlor DA. Metabolic profiling of adolescent non-alcoholic fatty liver disease. Wellcome Open Research. 2018;3.

41. Mäkelä TNK, Tuomainen T-P, Hantunen S, Virtanen JK. Associations of serum n–3 and n–6 polyunsaturated fatty acids with prevalence and incidence of nonalcoholic fatty liver disease. The American journal of clinical nutrition. 2022;116(3):759-70.
42. Walle P, Takkunen M, Männistö V, Vaittinen M, Lankinen M, Kärjä V, et al. Fatty acid metabolism is altered in non-alcoholic steatohepatitis independent of obesity. Metabolism. 2016;65(5):655-66.
43. Lind L, Salihovic S, Risérus U, Kullberg J, Johansson L, Ahlström H, et al. The plasma metabolomic profile is differently associated with liver fat, visceral adipose tissue, and pancreatic fat. The Journal of Clinical Endocrinology & Metabolism. 2021;106(1):e118-e29.
44. Gaggini M, Carli F, Rosso C, Buzzigoli E, Marietti M, Della Latta V, et al. Altered amino acid concentrations in NAFLD: Impact of obesity and insulin resistance. Hepatology. 2018;67(1):145-58.
45. Lovric A, Granér M, Bjornson E, Arif M, Benfeitas R, Nyman K, et al. Characterization of different fat depots in NAFLD using inflammation-associated proteome, lipidome and metabolome. Scientific reports. 2018;8(1):1-14.
46. Lo EKK, Xu J-H, Zhan Q, Zeng Z, El-Nezami H. The emerging role of branched-chain amino acids in liver diseases. Biomedicines. 2022;10(6):1444.
47. Moon C-M, Shin SS, Heo SH, Jeong YY. Metabolic alterations associated with early-stage hepatocellular carcinoma and their correlation with aging and enzymatic activity in patients with viral hepatitis-induced liver cirrhosis: A preliminary study. Journal of Clinical Medicine. 2020;9(3):765.
48. Nikolova-Karakashian M. Sphingolipids at the crossroads of NAFLD and senescence. Advances in Cancer Research. 2018;140:155-90.
49. Régnier M, Polizzi A, Guillou H, Loiseau N. Sphingolipid metabolism in non-alcoholic fatty liver diseases. Biochimie. 2019;159:9-22.
50. Goffredo M, Santoro N, Tricò D, Giannini C, D'Adamo E, Zhao H, et al. A branched-chain amino acid-related metabolic signature characterizes obese adolescents with non-alcoholic fatty liver disease. Nutrients. 2017;9(7):642.
51. Proffitt C, Bidkhori G, Lee S, Tebani A, Mardinoglu A, Uhlen M, et al. Genome-scale metabolic modelling of the human gut microbiome reveals changes in the glyoxylate and dicarboxylate metabolism in metabolic disorders. Iscience. 2022;25(7):104513.
52. Chalasani N, Vuppalanchi R, Raikwar NS, Deeg MA. Glycosylphosphatidylinositol-specific phospholipase d in nonalcoholic Fatty liver disease: a preliminary study. The Journal of Clinical Endocrinology & Metabolism. 2006;91(6):2279-85.
53. Pan Z, Raftery D. Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics. Analytical and bioanalytical chemistry. 2007;387(2):525-7.
54. Lee C-m, Yoon EL, Nakajima A, Yoneda M, Toyoda H, Yasuda S, et al. A reappraisal of the diagnostic performance of B-mode ultrasonography for mild liver steatosis. Official journal of the American College of Gastroenterology| ACG. 2022:10.14309.

Supplementary material

Study population

Avon Longitudinal Study of Parents and Children (ALSPAC)

Pregnant women resident in Avon, UK with expected dates of delivery between 1st April 1991 and 31st December 1992 were invited to take part in the study.[1-3] The initial number of pregnancies enrolled was 14,541. Of the initial pregnancies, there was a total of 14,676 foetuses, resulting in 14,062 live births and 13,988 children who were alive at 1 year of age. When the oldest children were approximately 7 years of age, an attempt was made to bolster the initial sample with eligible cases who had failed to join the study originally. The total sample size for analyses using any data collected after the age of seven is therefore 15,447 pregnancies, resulting in 15,658 foetuses. Of these 14,901 children were alive at 1 year of age. A total of 14,833 unique women (G0 mothers) enrolled in ALSPAC as of September 2021. Study data were collected and managed using REDCap electronic data capture tools hosted at the University of Bristol.[4] REDCap (Research Electronic Data Capture) is a secure, web-based software platform designed to support data capture for research studies. The study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool: http://www.bristol.ac.uk/alspac/researchers/-our-data/. For this project, data was used from N=2896 offspring at age 24 who had both metabolite data (from the Nightingale metabolomics platform) and liver enzyme data available.

The Insulin Resistance Atherosclerosis Family Study (IRASFS)

Study design, recruitment and phenotyping in the Insulin Resistance Atherosclerosis Study Family Study (IRASFS) have been described in detail.[5] Briefly, this multicenter study was designed to identify genetic determinants of insulin resistance and adiposity. The design of the IRASFS have been described in the supplementary data. The family-based sample includes individuals of self-reported Hispanic American ethnicity from San Antonio, TX, and San Luis Valley, CA, and individuals of self-reported African American ethnicity from Los Angeles, CA. Two clinical examinations were conducted at a 5-year interval. The examinations included an interview for health behaviors and medical history, anthropometric measurements, resting blood pressure and blood drawing for plasma and DNA. Usual consumption of beer, wine and liquor in the past year was assessed by self-report. Exclusions were made for usual alcohol consumption, which exceeded two drinks/day in men and one drink/day in women. History of liver disease was not collected.

Study of Latinos (SOL)

The Hispanic Community Health Study/Study of Latinos (SOL) is a population-based prospective cohort study conducted at 4 urban field centers in the United States, which were selected to provide diversity in national background and behaviors such as diet. SOL subject recruitment and the study design have been described previously.[6, 7] In total, 16,415 individuals who self-identified as Hispanic and/or Latino backgrounds (South Americans, Central Americans, Mexicans, Puerto Ricans, Cubans, and Dominicans) were recruited and completed data collection between June 2008 and July 2011. Detailed data were collected including demographics, socioeconomic status, acculturation, health-related behaviors, medical history, blood laboratory testing in the fasting state, and clinical measurements. Metabolomic profiling was conducted using stored serum samples for a random sample presenting approximately one-third of the cohort (N= 3972). Of these, 3833 participants had information on liver enzymes and covariates.

Assessment of blood levels of metabolites

Metabolomics in IRASFS (Metabolon platform)

Metabolite profiling was performed on stored (at -80 °C) fasting plasma samples collected at the 1999–2002 baseline survey. Metabolite detection and quantification was conducted by Metabolon, Inc. (Durham, North Carolina) using untargeted liquid chromatography-mass spectroscopy (MS) (DiscoveryHD4 panel). Samples were prepared using the automated MicroLab STAR system (Hamilton Company, Salt Lake City, UT). A methanol extraction was used to remove protein, dissociate small molecules bound to protein or trapped in the precipitated protein matrix, and to recover chemically diverse metabolites. The resulting extract was divided into five fractions: two for analysis by two separate reverse phase/ultra-performance liquid chromatography-MS/MS methods with positive ion mode electrospray ionization (ESI), one for analysis by reverse phase/ultra-performance liquid chromatography-MS/MS with negative ion mode ESI, one for analysis by hydrophilic interaction liquid chromatography/ultra-performance liquid chromatography-MS/MS with negative ion mode ESI and one sample was reserved for backup. All methods utilized 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 source and Orbitrap mass analyzer operated at 35,000 mass resolution. Raw data were extracted, peak-identified and quality control processed using Metabolon's hardware and software. Compounds were identified by comparison to library entries of purified standards or recurrent unknown entities. Peaks were quantified using area under the curve. Several types of controls were analyzed in addition to experimental samples: a technical replicate, pooled matrix sample generated from a small volume of each experimental sample; process blanks, extracted water samples; and QC standards, a cocktail of QC standards chosen not to interfere with the measurement of endogenous compounds were spiked into every analyzed sample, allowed instrument performance monitoring and aided

chromatographic alignment. This panel identified and provided relative quantification of known chemical compounds among amino acid, carbohydrate, energy, lipid, nucleotide, and peptide super pathways. In addition to individual named biochemicals; super- and sub-pathways were annotated based on a combination of pathway and chemical structure similarities to serve as a guide for interpretation. Prior to return, data were block corrected for a run day, normalized by batch, and volume extracted. Missing data for metabolites were imputed to the minimum value for the respective metabolite. Each metabolite in original scale was rescaled to set the median equal to one.

Metabolomics in SOL (Metabolon platform)

Fasting serum samples were collected, processed, and stored at −70 °C from the time of collection until metabolic profiling. Metabolites were quantified using an untargeted liquid chromatography-mass spectrometry (MS)-based metabolomic quantification protocol at Metabolon (Durham, NC, USA) using the Discovery HD4 platform. Detailed procedures are described in elsewhere.[8] The platform captures information for a total of 1136 metabolites, including 782 metabolites with known structural identities and 354 unknown metabolites. Metabolites with missing values for 20% or more of the participants were excluded. For the remaining values, missing values were imputed to half of the limit of detection. Metabolite values were tested for skewness and, if significantly skewed, log10 transformed. Outlier metabolite values (>10 standard deviations from the mean) were set to missing.

Assessment of NAFLD by CT scan (IRASFS)

The assessment of NAFLD in IRASFS was performed using CT imaging under a standardized protocol and scans were read centrally at the University of Colorado School of Medicine, Department of Radiology, Bio-Imaging Research Laboratory. Participants received a scout view of the abdomen and pelvis followed by three axial images all during suspended respiration. The three 10-mm-thick images were obtained through the L2–L3, L4–L5 and T11–T12 disc spaces. If the T11–T12 image did not include liver and spleen, a fourth image was obtained by using the scout to determine an appropriate intervertebral disc location. Liver and spleen density were then quantified in Hounsfield Units in the entire liver and spleen as visualized in the slice, excluding any visible vasculature.[9] The image obtained at the L4–L5 disc space was used for the determination of visceral adipose tissue (VAT) area; bowel fat is excluded from measurement. All CT images were coded for pathology and image quality; poor-quality studies were excluded from analysis.

Assessment of liver enzymes (IRASFS)

In IRASFS, alanine transaminase (ALT) and aspartate transaminase (AST) were determined by enzymatic colourimetric assays using a Chemistry Analyzer Model ATAC 8000 (Elan Diagnostic Co., Lakewood, NJ, USA) during the follow-up visit from 2005-2006.

Assessment of liver enzymes (ALSPAC)

Blood samples were collected whilst the ALSPAC 24-year-old offspring were fasting. GGT, ALT and AST were measured using commercially available enzymatic colorimetric assay kits, manufactured and supplied by Roche diagnostics GmbH, Sandhofer Strasse 116, D-68305 Mannheim.

Assessment of liver enzymes (SOL)

Serum ALT, AST and GGT were measured in the fasting state using a Roche Modular P Chemistry Analyzer (Roche Diagnostics) with an α-ketoglutaratic enzymatic method as described previously.[10, 11] The coefficients of variation were 6.0–7.0% for liver enzymes.

Definitions of covariates (ALSPAC)

In ALSPAC, we used data collected from the clinic at age 24. Participants were asked for the frequency they had a drink containing alcohol in the past year. We recoded 'never' or 'never had a whole drink' as no and 'monthly or less'/'2-4 times a month'/'2-3 times a week'/'>4 times a week' as yes. The three smoking categories of never/ex/current were derived from questions asking if they had ever smoked a cigarettes and whether they had smoked a cigarette in the last 30 days.

Definitions of covariates (SOL)

In SOL, alcohol use was derived from participant self-report of current alcohol use (question: "Do you presently drink alcoholic beverages?"). In SOL, participants were instructed to bring containers of those medications taken in the prior month. Medications were inventoried and therapeutically classified as described previously.[12] Participants were classified as taking a lipid-lowering medication if any medication in their individual inventory was coded as "39xxxx" (antihyperlipidemics) or "409925" (calcium channel blocker & HMG CoA reductase inhibitor combined).

References

1. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, et al. Cohort profile: the 'children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. International journal of epidemiology. 2013;42(1):111-27.
2. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, et al. Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. International journal of epidemiology. 2013;42(1):97-110.
3. Northstone K, Lewcock M, Groom A, Boyd A, Macleod J, Timpson N, et al. The Avon Longitudinal Study of Parents and Children (ALSPAC): an update on the enrolled sample of index children in 2019. Wellcome open research. 2019;4.
4. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. Journal of biomedical informatics. 2009;42(2):377-81.
5. Henkin L, Bergman RN, Bowden DW, Ellsworth DL, Haffner SM, Langefeld CD, et al. Genetic epidemiology of insulin resistance and visceral adiposity: the IRAS Family Study design and methods. Annals of epidemiology. 2003;13(4):211-7.
6. Sorlie PD, Avilés-Santa LM, Wassertheil-Smoller S, Kaplan RC, Daviglus ML, Giachello AL, et al. Design and implementation of the Hispanic Community Health Study/Study of Latinos. Ann Epidemiol. 2010;20(8):629-41.
7. LaVange LM, Kalsbeek WD, Sorlie PD, Avilés-Santa LM, Kaplan RC, Barnhart J, et al. Sample design and cohort selection in the Hispanic Community Health Study/Study of Latinos. Annals of epidemiology. 2010;20(8):642-9.
8. Kaplan RC, Williams-Nguyen JS, Huang Y, Mossavar-Rahmani Y, Yu B, Boerwinkle E, et al. Identification of Dietary Supplements Associated with Blood Metabolites in the Hispanic Community Health Study/Study of Latinos Cohort Study. The Journal of Nutrition. 2023;153(5):1483-92.
9. Davidson LE, Kuk JL, Church TS, Ross R. Protocol for measurement of liver fat by computed tomography. Journal of applied physiology. 2006;100(3):864-8.
10. Kallwitz ER, Daviglus ML, Allison MA, Emory KT, Zhao L, Kuniholm MH, et al. Prevalence of suspected nonalcoholic fatty liver disease in Hispanic/Latino individuals differs by heritage. Clinical Gastroenterology and Hepatology. 2015;13(3):569-76.
11. Thyagarajan B, Howard AG, Durazo-Arvizu R, Eckfeldt JH, Gellman MD, Kim RS, et al. Analytical and biological variability in biomarker measurement in the Hispanic Community Health Study/Study of Latinos. Clinica Chimica Acta. 2016;463:129-37.
12. Faurot KR, Siega-Riz AM, Gardiner P, Rivera JO, Young LA, Poole C, et al. Comparison of a medication inventory and a dietary supplement interview in assessing dietary supplement use in the Hispanic community health study/study of Latinos. Integrative Medicine Insights. 2016;11:IMI. S25587.

Table S1. List of top 30 metabolites significantly associated with NAFLD-US meta-analysis between RS subcohorts in model 2 (N=1594)

	Model2				
Metabolites	**Beta**	**SE**	**P-value**	**Direction**	**FDR**
Mean diameter for HDL particles	-0,802	0,084	7,62E-22	--	1,19E-19
Total cholesterol in large HDL	-0,775	0,082	4,59E-21	--	3,58E-19
Cholesterol esters in large HDL	-0,775	0,083	6,18E-21	--	3,21E-19
Phospholipids in very large HDL	-0,820	0,088	7,40E-21	--	2,89E-19
Free cholesterol in large HDL	-0,806	0,087	1,83E-20	--	5,71E-19
Total lipids in large HDL	-0,722	0,078	1,92E-20	--	4,99E-19
Concentration of large HDL particles	-0,772	0,085	6,49E-20	--	1,45E-18
Total cholesterol in HDL2	-0,718	0,080	2,48E-19	--	4,84E-18
Phospholipids in large HDL	-0,697	0,078	2,64E-19	--	4,58E-18
Total cholesterol in HDL	-0,718	0,080	3,81E-19	--	5,94E-18
Isoleucine	0,676	0,077	1,23E-18	++	1,74E-17
Ratio of triglycerides to phosphoglycerides	0,621	0,072	5,26E-18	++	6,84E-17
Total lipids in very large HDL	-0,710	0,083	9,06E-18	--	1,09E-16
Concentration of very large HDL particles	-0,750	0,089	2,50E-17	--	2,79E-16
Free cholesterol in very large HDL	-0,724	0,087	8,00E-17	--	8,32E-16
Mean diameter for VLDL particles	0,579	0,071	3,41E-16	++	3,32E-15
Triglycerides in VLDL	0,553	0,069	1,56E-15	++	1,43E-14
Triglycerides in small HDL	0,538	0,068	2,99E-15	++	2,59E-14
Ratio of apolipoprotein B to apolipoprotein A_I	0,536	0,070	1,43E-14	++	1,17E-13
Total lipids in medium VLDL	0,499	0,066	3,02E-14	++	2,36E-13
Triglycerides in medium VLDL	0,504	0,066	3,10E-14	++	2,30E-13
Serum total triglycerides	0,507	0,067	3,77E-14	++	2,67E-13
Total lipids in large VLDL	0,499	0,066	4,95E-14	++	3,36E-13
Triglycerides in small VLDL	0,511	0,068	8,74E-14	++	5,68E-13
Triglycerides in large VLDL	0,493	0,067	1,70E-13	++	1,06E-12
Glucose	0,503	0,068	1,84E-13	++	1,10E-12
Glycoprotein acetyls_ mainly a1_acid glycoprotein	0,482	0,066	1,99E-13	++	1,15E-12
Total cholesterol in very large HDL	-0,586	0,080	2,12E-13	--	1,18E-12
Leucine	0,550	0,075	3,07E-13	++	1,65E-12
Total lipids in small VLDL	0,480	0,067	8,04E-13	++	4,18E-12

The table shows metabolites significantly assocaited with NAFLD-US ordered by the P-value of Model 2 (FDR <0.05). Model 2: NAFLD~Metabolite + age + sex+lipid lowering medication + alcohol consumption+ body mass index + smoking status. Abbreviation: RS, Rotterdam study; US, Ultra-sound; HDL, High-density lipoprotein; LDL, Low-density lipoprotein; IDL, Intermediate-density lipoprotein; VLDL, Very low- density lipoprotein; N, sample size

Table S2. List of top 20 metabolites significantly associated with NAFLD based-CT scan (RS-I-4)

	Model1			Model2			
Metabolite	**Beta**	**SE**	**P-value**	**Beta**	**SE**	**P-value**	**N**
Ratio of polyunsaturated fatty acids to total fatty acids	-1,098	0,198	2,82E-08	-1,143	0,233	9,28E-07	680
Ratio of omega_6 fatty acids to total fatty acids	-1,114	0,203	4,36E-08	-1,162	0,242	1,57E-06	680
Ratio of 18:2 linoleic acid to total fatty acids	-1,162	0,223	1,97E-07	-1,226	0,265	3,76E-06	680
Estimated degree of unsaturation	-0,990	0,208	1,91E-06	-0,935	0,225	3,19E-05	680
Ratio of monounsaturated fatty acids to total fatty acids	0,846	0,189	7,51E-06	0,800	0,209	1,25E-04	680
Leucine	0,645	0,168	1,22E-04	0,660	0,193	6,17E-04	681
Ratio of saturated fatty acids to total fatty acids	0,675	0,190	3,89E-04	0,626	0,209	2,73E-03	680
Phospholipids in small HDL	0,602	0,175	5,84E-04	0,579	0,196	3,17E-03	681
Triglycerides in small HDL	0,596	0,164	2,70E-04	0,503	0,174	3,83E-03	681
Triglycerides in small LDL	0,544	0,159	6,25E-04	0,472	0,169	5,19E-03	681
Estimated description of fatty acid chain length_ not actual carbon number	0,568	0,171	8,72E-04	0,538	0,197	6,41E-03	680
Monounsaturated fatty acids	0,519	0,155	7,99E-04	0,428	0,164	8,92E-03	680
Cholesterol esters in small HDL	-0,532	0,194	5,98E-03	-0,553	0,212	9,08E-03	681
Phospholipids in large VLDL	0,467	0,134	4,93E-04	0,372	0,144	9,64E-03	680
Triglycerides in HDL	0,489	0,164	2,85E-03	0,432	0,169	1,05E-02	681
Phenylalanine	0,472	0,165	4,23E-03	0,460	0,182	1,13E-02	681
Triglycerides in medium HDL	0,509	0,160	1,44E-03	0,415	0,165	1,21E-02	681
Isoleucine	0,545	0,163	8,45E-04	0,462	0,186	1,29E-02	681
Glutamine	-0,615	0,187	1,02E-03	-0,490	0,198	1,35E-02	681
Phospholipids in very large VLDL	0,431	0,127	6,71E-04	0,332	0,135	1,36E-02	681

The table shows metabolites associated with NAFLD based CT scan in Model 2. Model 1: NAFLD~Metabolite + age + sex. Model 2: Model 1 + lipid-lowering medication + alcohol consumption+ body mass index + smoking status. Abbreviation: RS, Rotterdam study; SE, standard error; N, sample size; HDL, High-density lipoprotein; LDL, Low-density lipoprotein; IDL, Intermediate-density lipoprotein; VLDL, Very low-density lipoprotein.

Table S3. List of top 30 metabolites significantly associated with NAFLD based-CT scan (HU) (RS-I-4)

	Model1			Model2				
Metabolite	**Beta**	**SE**	**P-value**	**Beta**	**SE**	**P-value**	**FDR**	**N**
Ratio of polyunsaturated fatty acids to total fatty acids	3,503	0,334	6,00E-24	3,096	0,344	2,35E-18	3,79E-16	680
Ratio of omega_6 fatty acids to total fatty acids	3,520	0,334	3,69E-24	3,103	0,348	4,84E-18	3,89E-16	680
Estimated degree of unsaturation	2,869	0,341	2,55E-16	2,581	0,332	2,75E-14	1,47E-12	680
Ratio of 18:2 linoleic acid to total fatty acids	3,166	0,342	2,75E-19	2,780	0,361	4,51E-14	1,81E-12	680
Leucine	-3,194	0,381	2,77E-16	-2,675	0,391	1,71E-11	5,50E-10	681
Isoleucine	-3,133	0,361	3,07E-17	-2,579	0,378	1,89E-11	5,08E-10	681
Phospholipids in small HDL	-2,638	0,354	3,05E-13	-2,415	0,360	4,36E-11	1,00E-09	681
Ratio of saturated fatty acids to total fatty acids	-2,386	0,347	1,33E-11	-2,171	0,334	1,58E-10	3,18E-09	680
Phenylalanine	-2,508	0,351	2,41E-12	-2,192	0,345	3,79E-10	6,78E-09	681

Tyrosine	-2,703	0,345	1,80E-14	-2,174	0,345	5,57E-10	8,96E-09	681
Ratio of monounsaturated fatty acids to total fatty acids	-2,512	0,348	1,36E-12	-2,015	0,352	1,59E-08	2,33E-07	680
Concentration of small HDL particles	-2,476	0,370	4,47E-11	-2,096	0,369	1,93E-08	2,59E-07	681
Triglycerides in large VLDL	-2,459	0,350	5,46E-12	-1,927	0,354	7,35E-08	9,11E-07	680
Phospholipids in large VLDL	-2,467	0,351	4,83E-12	-1,925	0,354	7,44E-08	8,56E-07	680
Triglycerides in small HDL	-2,425	0,349	9,01E-12	-1,909	0,353	8,71E-08	9,35E-07	681
Triglycerides in very large VLDL	-2,481	0,350	3,51E-12	-1,904	0,354	1,01E-07	1,01E-06	681
Concentration of large VLDL particles	-2,407	0,351	1,59E-11	-1,867	0,354	1,77E-07	1,68E-06	680
Total lipids in large VLDL	-2,388	0,351	2,33E-11	-1,846	0,354	2,40E-07	2,14E-06	680
Total lipids in small HDL	-2,276	0,373	1,71E-09	-1,924	0,369	2,44E-07	2,07E-06	681
Concentration of very large VLDL particles	-2,412	0,351	1,44E-11	-1,835	0,353	2,71E-07	2,18E-06	681
Phospholipids in very large VLDL	-2,382	0,352	2,84E-11	-1,817	0,353	3,45E-07	2,64E-06	681
Total lipids in very large VLDL	-2,383	0,351	2,57E-11	-1,807	0,353	4,06E-07	2,97E-06	681
Serum total triglycerides	-2,334	0,352	6,44E-11	-1,805	0,354	4,31E-07	3,02E-06	681
Triglycerides in VLDL	-2,337	0,351	5,61E-11	-1,801	0,354	4,79E-07	3,21E-06	681
Free cholesterol in large VLDL	-2,332	0,353	7,76E-11	-1,786	0,354	5,72E-07	3,69E-06	680
Mean diameter for VLDL particles	-2,336	0,351	6,00E-11	-1,798	0,359	6,92E-07	4,29E-06	681
Triglycerides in medium VLDL	-2,278	0,351	1,75E-10	-1,756	0,354	9,00E-07	5,37E-06	681
Free cholesterol in chylomicrons and extremely large VLDL	-2,298	0,353	1,49E-10	-1,741	0,352	9,65E-07	5,55E-06	680

The table shows metabolites assocaited with NAFLD based CT-scan in Model 2. Model 1: NAFLD~Metabolite + age + sex. Model 2: Model 1 + lipid-lowering medication + alcohol consumption + body mass index + smoking status. Abbreviation: RS, Rotterdam study; HU, Hounsfield unit; SE, standard erorr; N, sample size; HDL, High-density lipoprotein; LDL, Low-density lipoprotein; IDL, Intermediate-density lipoprotein; VLDL, Very low-density lipoprotein.

Table S4. List of top 25 metabolites significantly associated with NAFLD based-US (RS-III-2)

	Model1			Model2				
Metabolite	**Beta**	**SE**	**P-value**	**Beta**	**SE**	**P-value**	**FDR**	**N**
4-hydroxyglutamate	1,302	0,160	3,48E-16	1,059	0,170	4,84E-10	4,80E-07	654
2-oxoarginine*	1,319	0,200	3,88E-11	1,149	0,206	2,63E-08	1,30E-05	654
behenoyl dihydrosphingomyelin (d18:0/22:0)*	1,645	0,253	7,81E-11	1,433	0,260	3,37E-08	1,11E-05	654
1-carboxyethylleucine	1,466	0,215	8,54E-12	1,200	0,221	5,49E-08	1,36E-05	654
glutamate	1,974	0,261	4,36E-14	1,438	0,275	1,65E-07	3,27E-05	654
1-carboxyethylphenylalanine	1,324	0,204	8,27E-11	1,093	0,214	3,37E-07	5,57E-05	654
1-carboxyethylisoleucine	1,000	0,163	7,58E-10	0,831	0,170	1,07E-06	1,51E-04	654
isoleucine	3,678	0,571	1,15E-10	2,837	0,589	1,45E-06	1,80E-04	654
sphingomyelin (d18:0/20:0, d16:0/22:0)*	1,430	0,249	9,23E-09	1,203	0,252	1,75E-06	1,93E-04	654
1-carboxyethylvaline	1,243	0,204	1,00E-09	0,990	0,208	2,01E-06	1,99E-04	654
palmitoyl-arachidonoyl-glycerol (16:0/20:4) [2]*	0,648	0,111	5,04E-09	0,503	0,110	4,80E-06	4,32E-04	654
palmitoyl-oleoyl-glycerol (16:0/18:1) [1]*	0,837	0,133	3,26E-10	0,630	0,139	5,35E-06	4,42E-04	654
myristoyl-linoleoyl-glycerol (14:0/18:2) [2]*	0,575	0,095	1,62E-09	0,452	0,100	6,23E-06	4,75E-04	654

metabolonic lactone sulfate	1,078	0,159	1,26E-11	0,751	0,167	6,50E-06	4,60E-04	654
aspartate	2,229	0,364	9,45E-10	1,617	0,364	8,78E-06	5,80E-04	654
X - 12456	0,682	0,111	8,27E-10	0,501	0,115	1,34E-05	8,30E-04	654
1-stearoyl-2-oleoyl-GPE (18:0/18:1)	1,159	0,201	8,53E-09	0,896	0,206	1,42E-05	8,28E-04	654
1-carboxyethyltyrosine	1,034	0,176	4,70E-09	0,799	0,186	1,72E-05	9,47E-04	654
X - 22834	0,443	0,089	7,09E-07	0,412	0,097	2,01E-05	1,05E-03	654
palmitoyl-oleoyl-glycerol (16:0/18:1) [2]*	0,955	0,158	1,52E-09	0,690	0,164	2,65E-05	1,31E-03	654
palmitoyl-linoleoyl-glycerol (16:0/18:2) [2]*	0,997	0,183	5,32E-08	0,755	0,184	4,07E-05	1,92E-03	654
fructosyllysine	1,724	0,312	3,30E-08	1,265	0,314	5,55E-05	2,50E-03	654
tetrahydrocortisone glucuronide (5)	1,147	0,211	5,63E-08	0,868	0,216	5,80E-05	2,50E-03	654
cystathionine	0,719	0,144	6,24E-07	0,604	0,151	6,38E-05	2,63E-03	654
1-stearoyl-2-linoleoyl-GPE (18:0/18:2)*	0,982	0,215	5,07E-06	0,868	0,217	6,44E-05	2,55E-03	654

The table shows metabolites significantly associated with NAFLD-US ordered by the P-value of Model 2 (FDR <0.05). Model 1: NAFLD~Metabolite + age + sex. Model 2: Model 1 + lipid-lowering medication + alcohol consumption + body mass index + smoking status. Abbreviation: RS, Rotterdam study; US, Ultrasound; SE, standard error; N, sample size.

Table S5. List of top 30 metabolites significantly associated with NAFLD based CT-scan a meta-analysis of the RS, IRASFS-AA and IRASFS-HA (N=1451)

	Model2						
Metabolites	**Beta**	**SE**	**P-value**	**Direction**	I^2	**Het.P-value**	**FDR**
1-palmitoyl-2-linoleoyl-GPI (16:0/18:2)	0,224	0,054	3,59E-05	+++	79,4	7,79E-03	1,85E-02
3-methyl-2-oxovalerate	0,269	0,069	9,64E-05	-++	0	4,85E-01	2,48E-02
cinnamoylglycine	-0,239	0,065	2,45E-04	---	10,5	3,27E-01	4,20E-02
isoleucine	0,235	0,065	2,66E-04	+++	0	7,15E-01	3,43E-02
leucine	0,236	0,068	5,07E-04	+++	34,3	2,18E-01	5,22E-02
4-methyl-2-oxopentanoate	0,231	0,069	8,29E-04	+++	0	5,54E-01	7,11E-02
1-stearoyl-2-oleoyl-GPE (18:0/18:1)	0,196	0,059	8,44E-04	+++	68,6	4,12E-02	6,21E-02
sphingomyelin (d18:0/18:0, d19:0/17:0)*	0,198	0,060	9,87E-04	+++	0	4,62E-01	6,35E-02
betaine	-0,189	0,057	9,89E-04	---	0	6,65E-01	5,66E-02
urate	0,237	0,072	9,94E-04	+++	0	4,84E-01	5,12E-02
lactate	0,188	0,057	9,97E-04	-++	0	3,86E-01	4,67E-02
trigonelline (N'-methylnicotinate)	-0,210	0,064	1,05E-03	---	39,2	1,93E-01	4,52E-02
1-palmitoleoylglycerol (16:1)*	0,211	0,064	1,07E-03	++?	0	8,60E-01	4,22E-02
dihomo-linolenate (20:3n3 or n6)	0,180	0,055	1,07E-03	-++	22,8	2,74E-01	3,92E-02
1-stearoyl-2-docosahexaenoyl-GPE (18:0/22:6)*	0,195	0,060	1,12E-03	+++	62	7,19E-02	3,84E-02
S-methylcysteine	-0,174	0,054	1,21E-03	---	0	8,67E-01	3,88E-02
3-methyl-2-oxobutyrate	0,198	0,061	1,24E-03	-++	7,9	3,38E-01	3,74E-02
1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1)*	0,179	0,056	1,35E-03	+++	77,7	1,12E-02	3,87E-02
lysine	0,188	0,059	1,50E-03	+++	0	5,36E-01	4,06E-02
X - 24295	0,187	0,059	1,52E-03	-++	0	6,80E-01	3,90E-02
N-acetylaspartate (NAA)	-0,170	0,054	1,55E-03	---	0	6,65E-01	3,79E-02
aspartate	0,161	0,051	1,57E-03	+++	0	5,47E-01	3,68E-02

1-palmitoyl-2-arachidonoyl-GPI (16:0/20:4)*	0,173	0,055	1,58E-03	+++	30,6	2,37E-01	3,54E-02
2-hydroxy-3-methylvalerate	0,209	0,067	1,69E-03	+++	0	5,97E-01	3,63E-02
X - 11308	-0,179	0,058	2,20E-03	---	7,7	3,39E-01	4,53E-02
gamma-glutamylleucine	0,205	0,067	2,21E-03	+++	18,7	2,93E-01	4,38E-02
X - 12104	-0,179	0,059	2,28E-03	---	0	1,00E+00	4,34E-02
tryptophan	0,214	0,070	2,30E-03	+++	0	5,52E-01	4,23E-02
valine	0,192	0,064	2,61E-03	+++	25	2,63E-01	4,63E-02
3-methoxytyrosine	-0,166	0,056	3,13E-03	+--	0	5,44E-01	5,38E-02

Ordered by the P-value of Model2 (FDR <0.05). Model 2: NAFLD~Metabolite + age + sex+lipid lowering medication + alcohol consumption+ body mass index + smoking status.; Direction: Direction of the effect estimate in each study of the meta-analysis (order – RS-I-4, IRASFS-AA, IRASFS-HA); I^2: heterogeneity parameter, Het. P-value: significance of the heterogeneity; FDR, False discovery rate.

Table S6. List of top 50 metabolites significantly associated with NAFLD based CT-scan (HU) a meta-analysis of the RS, IRASFS-AA and IRASFS-HA (N=1451)

	Model2						
Metabolites	**Beta**	**SE**	**P-value**	**Direction**	I^2	**Het.P-value**	**FDR**
sphingomyelin (d18:0/18:0, d19:0/17:0)*	-2,323	0,321	4,46E-13	---	78,9	8,69E-03	2,34E-10
tyrosine	-2,107	0,320	4,42E-11	---	62,8	6,79E-02	1,16E-08
sphingomyelin (d18:0/20:0, d16:0/22:0)*	-2,071	0,320	9,01E-11	---	77,2	1,25E-02	1,57E-08
1-stearoyl-2-oleoyl-GPE (18:0/18:1)	-1,925	0,300	1,48E-10	---	71,2	3,10E-02	1,94E-08
1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1)*	-1,923	0,303	2,25E-10	---	88,3	1,89E-04	2,36E-08
leucine	-2,261	0,370	9,50E-10	---	86,5	6,20E-04	8,30E-08
isoleucine	-2,052	0,336	9,89E-10	---	89,9	5,06E-05	7,40E-08
3-methyl-2-oxovalerate	-2,142	0,353	1,35E-09	---	72,4	2,68E-02	8,84E-08
glutamate	-1,893	0,322	4,25E-09	---	89,3	8,41E-05	2,47E-07
valine	-1,949	0,338	8,10E-09	---	71,9	2,86E-02	4,24E-07
1-stearoyl-2-docosahexaenoyl-GPE (18:0/22:6)*	-1,862	0,324	9,34E-09	---	81,4	4,67E-03	4,45E-07
X - 24295	-1,642	0,297	3,33E-08	---	61,3	7,53E-02	1,45E-06
2-oxoarginine*	-1,754	0,319	3,68E-08	---	0	5,20E-01	1,48E-06
lactate	-1,660	0,304	4,92E-08	---	61,1	7,64E-02	1,84E-06
1-palmitoyl-2-linoleoyl-GPI (16:0/18:2)	-1,542	0,286	7,28E-08	---	85,7	9,07E-04	2,54E-06
X - 19438	-1,477	0,278	1,08E-07	---	0	3,80E-01	3,54E-06
cinnamoylglycine	1,266	0,239	1,10E-07	+++	83,5	2,30E-03	3,39E-06
homoarginine	-1,659	0,315	1,41E-07	---	0	7,57E-01	4,10E-06
4-methyl-2-oxopentanoate	-1,813	0,346	1,54E-07	---	76,3	1,48E-02	4,25E-06
gamma-glutamylglycine	1,609	0,309	1,89E-07	+++	69,7	3,69E-02	4,95E-06
1-stearoyl-2-linoleoyl-GPE (18:0/18:2)*	-1,571	0,302	1,97E-07	---	49,6	1,37E-01	4,92E-06
glycine	1,596	0,307	2,01E-07	+++	83,9	2,01E-03	4,79E-06
1-palmitoyl-2-linoleoyl-GPE (16:0/18:2)	-1,549	0,300	2,37E-07	---	59,1	8,66E-02	5,40E-06
tryptophan	-1,682	0,327	2,73E-07	---	60,1	8,18E-02	5,96E-06
N-acetylglycine	1,960	0,386	3,69E-07	+++	22,2	2,76E-01	7,73E-06
trigonelline (N'-methylnicotinate)	1,443	0,286	4,41E-07	+++	0	9,19E-01	8,89E-06

gamma-glutamyltyrosine	-1,782	0,358	6,20E-07	---	87,2	4,08E-04	1,20E-05
1-myristoyl-2-palmitoyl-GPC (14:0/16:0)	-1,441	0,296	1,16E-06	---	70,4	3,41E-02	2,17E-05
aspartate	-1,487	0,311	1,73E-06	---	91,1	1,28E-05	3,13E-05
1-palmitoyl-2-arachidonoyl-GPE (16:0/20:4)*	-1,433	0,304	2,50E-06	---	33,9	2,20E-01	4,37E-05
alpha-hydroxyisovalerate	-1,488	0,318	2,88E-06	---	58,6	8,95E-02	4,87E-05
2-hydroxy-3-methylvalerate	-1,552	0,332	3,00E-06	---	66,5	5,07E-02	4,91E-05
3-(4-hydroxyphenyl)lactate	-1,608	0,349	3,98E-06	---	83,7	2,20E-03	6,32E-05
alpha-ketobutyrate	-1,344	0,292	4,18E-06	---	76,9	1,32E-02	6,44E-05
pyruvate	-1,376	0,300	4,52E-06	---	73,1	2,43E-02	6,77E-05
1-palmitoylglycerol (16:0)	-1,319	0,288	4,61E-06	---	0	8,36E-01	6,71E-05
1-(1-enyl-palmitoyl)-2-oleoyl-GPC (P-16:0/18:1)*	1,456	0,322	6,01E-06	+++	78,1	1,03E-02	8,51E-05
X - 16087	-1,374	0,307	7,41E-06	---	59,7	8,39E-02	1,02E-04
X - 12216	1,164	0,263	9,73E-06	+++	28	2,49E-01	1,31E-04
gamma-glutamylvaline	-1,431	0,325	1,10E-05	---	37,1	2,04E-01	1,44E-04
N-acetylmethionine	1,314	0,303	1,46E-05	+++	79,2	8,24E-03	1,87E-04
1-palmitoyl-2-arachidonoyl-GPI (16:0/20:4)*	-1,273	0,294	1,51E-05	---	46,6	1,54E-01	1,88E-04
propionylglycine	1,547	0,359	1,65E-05	++?	0	8,14E-01	2,01E-04
linoleoylcarnitine (C18:2)*	1,357	0,315	1,65E-05	+++	84,9	1,34E-03	1,97E-04
1-oleoyl-2-docosahexaenoyl-GPC (18:1/22:6)*	1,382	0,321	1,68E-05	+++	48,3	1,44E-01	1,96E-04
S-methylcysteine	1,218	0,284	1,77E-05	+++	0	6,25E-01	2,02E-04
gamma-glutamylleucine	-1,561	0,368	2,18E-05	---	79,8	7,14E-03	2,43E-04
3-methyl-2-oxobutyrate	-1,534	0,364	2,54E-05	---	50,3	1,34E-01	2,77E-04
methionine	-1,350	0,321	2,67E-05	---	0	6,22E-01	2,86E-04
xanthurenate	-1,206	0,291	3,45E-05	---	42,3	1,77E-01	3,62E-04

Ordered by the P-value of Model2 (FDR <0.05).Model 2: NAFLD~Metabolite + age + sex+lipid lowering medication + alcohol consumption+ body mass index + smoking status.; Direction: Direction of the effect estimate in each study of the meta-analysis (order – RS-I-4, IRASFS-AA, IRASFS-HA); I^2: heterogeneity parameter, Het. P-value: significance of the heterogeneity; HU, Hounsfield unit; FDR, False discovery rate.

Table S7. List of top 25 metabolites significantly associated with liver enzymes (ALT) meta-analysis between RS cohorts and ALSPAC

	Model2						
Metabolite	**Beta**	**SE**	**P-value**	**Direction**	I^2	**Het.P-value**	**FDR**
Triglycerides in large LDL	1,916	0,284	1,47E-11	++?+	58,9	8,80E-02	2,28E-09
Triglycerides in IDL	1,822	0,277	4,72E-11	++?+	74,5	1,99E-02	3,66E-09
Triglycerides in LDL	1,824	0,284	1,37E-10	++?+	61,7	7,35E-02	7,08E-09
Ratio of omega_6 fatty acids to total fatty acids	-1,822	0,292	4,36E-10	--?-	0	9,64E-01	1,69E-08
Triglycerides in medium LDL	1,660	0,287	7,18E-09	++?+	63	6,69E-02	2,23E-07
Tyrosine	1,622	0,283	9,73E-09	++?+	51,3	1,28E-01	2,51E-07
Saturated fatty acids	1,533	0,282	5,32E-08	++?+	62,2	7,09E-02	1,18E-06
Triglycerides in very large HDL	1,506	0,277	5,54E-08	+-?+	87,3	3,74E-04	1,07E-06
Concentration of very small VLDL particles	1,529	0,283	6,82E-08	++?+	78,7	9,20E-03	1,17E-06

Ratio of 18:2 linoleic acid to total fatty acids	-1,582	0,296	8,72E-08	--?-	0	8,44E-01	1,35E-06
Ratio of polyunsaturated fatty acids to total fatty acids	-1,553	0,292	1,01E-07	--?-	0	9,69E-01	1,42E-06
Triglycerides in HDL	1,450	0,275	1,34E-07	++?+	69,1	3,93E-02	1,73E-06
Triglycerides in very small VLDL	1,475	0,281	1,54E-07	++?+	74,4	2,00E-02	1,84E-06
Total lipids in very small VLDL	1,485	0,284	1,68E-07	++?+	78,6	9,39E-03	1,86E-06
Cholesterol esters in very small VLDL	1,508	0,291	2,19E-07	++?+	81,5	4,50E-03	2,26E-06
Triglycerides in small LDL	1,386	0,277	5,80E-07	++?+	68,8	4,04E-02	5,62E-06
Ratio of saturated fatty acids to total fatty acids	1,330	0,268	7,13E-07	++?+	67,1	4,80E-02	6,50E-06
Total cholesterol in very small VLDL	1,415	0,288	9,36E-07	++?+	81,3	4,78E-03	8,06E-06
Cholesterol esters in chylomicrons and extremely large VLDL	1,304	0,276	2,24E-06	++?+	70,9	3,21E-02	1,83E-05
Leucine	1,489	0,318	2,87E-06	++?+	12,9	3,17E-01	2,22E-05
Total cholesterol in VLDL	1,317	0,284	3,44E-06	++?+	77,2	1,25E-02	2,54E-05
Omega_3 fatty acids	1,264	0,275	4,38E-06	++?+	15,9	3,05E-01	3,09E-05
RemNAt cholesterol (non_HDL_ non_LDL _cholesterol)	1,291	0,286	6,50E-06	++?+	80,2	6,34E-03	4,38E-05
Total cholesterol in chylomicrons and extremely large VLDL	1,222	0,275	8,70E-06	++?+	64,9	5,79E-02	5,62E-05
Monounsaturated fatty acids	1,235	0,280	1,02E-05	+-?+	68	4,41E-02	6,32E-05

The table shows metabolites significantly associated with liver enzyme (ALT) ordered by the P-value of Model 2 (FDR <0.05). Model 2: NAFLD~Metabolite + age + sex+lipid lowering medication + alcohol consumption+ body mass index + smoking status. Abbreviation: RS, Rotterdam study; ALSPAC, Avon Longitudinal Study of Parents and Childreny; ALT, Alanine aminotransferase; HDL, High-density lipoprotein; LDL, Low-density lipoprotein; IDL, Intermediate-density lipoprotein; VLDL, Very low-density lipoprotein; N, sample size; ?, not applicable cohort from RS-I-4.

Table S8. List of top 25 metabolites significantly associated with liver enzymes (AST) meta-analysis between RS cohorts and ALSPAC

	Model2						
Metabolite	**Beta**	**SE**	**P-value**	**Direction**	I^2	**Het.P-Value**	**FDR**
Triglycerides in large HDL	1,624	0,338	1,52E-06	++?+	93	6,89E-07	2,36E-04
Triglycerides in large LDL	1,483	0,315	2,44E-06	++?+	65	5,72E-02	1,89E-04
Triglycerides in medium LDL	1,489	0,318	2,81E-06	+-?+	79,7	7,21E-03	1,45E-04
Triglycerides in LDL	1,385	0,315	1,11E-05	++?+	70,4	3,43E-02	4,30E-04
Mean diameter for LDL particles	1,316	0,304	1,46E-05	++?-	92,1	3,33E-06	4,53E-04
Triglycerides in IDL	1,322	0,307	1,65E-05	++?+	64,3	6,07E-02	4,26E-04
Triglycerides in very large HDL	1,037	0,307	7,21E-04	+-?+	89	1,17E-04	1,60E-02
Citrate	1,007	0,302	8,47E-04	++?+	36,9	2,05E-01	1,64E-02
Tyrosine	0,976	0,314	1,88E-03	++?+	62,2	7,11E-02	3,24E-02
Concentration of very large HDL particles	1,059	0,350	2,51E-03	+-?+	84	1,91E-03	3,89E-02
Ratio of 18:2 linoleic acid to total fatty acids	-0,970	0,326	2,90E-03	--?-	0	9,55E-01	4,09E-02
Mean diameter for HDL particles	1,027	0,353	3,57E-03	++?+	83,6	2,21E-03	4,61E-02
Concentration of large HDL particles	1,008	0,353	4,29E-03	++?+	76,7	1,36E-02	5,11E-02
Cholesterol esters in very large HDL	0,971	0,342	4,51E-03	+-?+	74,6	1,95E-02	4,99E-02
Phenylalanine	0,868	0,310	5,16E-03	++?+	29,9	2,40E-01	5,33E-02
Free cholesterol in large HDL	0,938	0,352	7,68E-03	++?+	76	1,56E-02	7,44E-02

Triglycerides in small LDL	0,816	0,307	7,90E-03	+-?+	73	2,47E-02	7,20E-02
Total cholesterol in very large HDL	0,917	0,346	8,01E-03	+-?+	75,1	1,82E-02	6,90E-02
Cholesterol esters in large HDL	0,930	0,351	8,09E-03	++?+	76,6	1,39E-02	6,60E-02
Ratio of saturated fatty acids to total fatty acids	0,778	0,296	8,60E-03	++?-	84,4	1,66E-03	6,66E-02
Phospholipids in very large HDL	0,918	0,353	9,37E-03	+-?+	82,2	3,62E-03	6,92E-02
Total cholesterol in large HDL	0,912	0,352	9,48E-03	++?+	75,9	1,59E-02	6,68E-02
Triglycerides in HDL	0,791	0,305	9,50E-03	++?+	80,8	5,47E-03	6,40E-02
Total lipids in very large HDL	0,902	0,352	1,03E-02	+-?+	79,9	6,86E-03	6,67E-02
Ratio of omega_6 fatty acids to total fatty acids	-0,809	0,322	1,20E-02	--?-	0	8,96E-01	7,46E-02

The table shows metabolites significantly associated with liver enzyme (AST) ordered by the P-value of Model 2 (FDR <0.05). Model 2: NAFLD~Metabolite + age + sex+lipid lowering medication + alcohol consumption+ body mass index + smoking status. Abbreviation: RS, Rotterdam study; ALSPAC, Avon Longitudinal Study of Parents and Children; AST, Aspartate aminotransferase; HDL, High-density lipoprotein; LDL, Low-density lipoprotein; IDL, Intermediate-density lipoprotein; VLDL, Very low- density lipoprotein; N, sample size; ?, not applicable cohort from RS-I-4.

Table S9. List of top 25 metabolites significantly associated with liver enzymes (GGT) meta-analysis between RS cohorts and ALSPAC

	Model2						
Metabolite	**Beta**	**SE**	**P-value**	**Direction**	I^2	**Het.P-Value**	**FDR**
Total cholesterol in chylomicrons and extremely large VLDL	3,657	0,340	4,97E-27	++++	50	1,12E-01	8,00E-25
Cholesterol esters in chylomicrons and extremely large VLDL	3,629	0,339	9,08E-27	++++	55,5	8,07E-02	7,31E-25
Free cholesterol in very large VLDL	3,568	0,341	1,40E-25	++++	43,7	1,49E-01	7,51E-24
Free cholesterol in chylomicrons and extremely large VLDL	3,539	0,341	2,66E-25	++++	44	1,48E-01	1,07E-23
Phospholipids in very large VLDL	3,544	0,343	4,92E-25	++++	36,5	1,93E-01	1,58E-23
Phospholipids in chylomicrons and extremely large VLDL	3,528	0,342	5,28E-25	++++	35,9	1,97E-01	1,42E-23
Total cholesterol in very large VLDL	3,541	0,343	5,65E-25	++++	43,6	1,50E-01	1,30E-23
Total lipids in chylomicrons and extremely large VLDL	3,524	0,345	1,48E-24	++++	33,1	2,13E-01	2,98E-23
Cholesterol esters in very large VLDL	3,504	0,344	2,47E-24	++++	45,5	1,39E-01	4,42E-23
Glycoprotein acetyls_ mainly a1_acid glycoprotein	3,589	0,355	4,95E-24	++++	0	5,48E-01	7,97E-23
Concentration of very large VLDL particles	3,493	0,346	5,32E-24	++++	50,9	1,06E-01	7,79E-23
Concentration of chylomicrons and extremely large VLDL particles	3,471	0,345	7,57E-24	++++	21	2,84E-01	1,02E-22
Triglycerides in chylomicrons and extremely large VLDL	3,474	0,346	8,93E-24	++++	32,1	2,20E-01	1,11E-22
Triglycerides in small LDL	3,347	0,335	1,50E-23	++++	78,1	3,31E-03	1,73E-22
Total lipids in very large VLDL	3,444	0,347	3,00E-23	++++	19,7	2,91E-01	3,22E-22
Triglycerides in IDL	3,325	0,337	6,12E-23	++++	82,1	7,93E-04	6,16E-22
Free cholesterol in large VLDL	3,415	0,347	6,45E-23	++++	37	1,90E-01	6,11E-22
Monounsaturated fatty acids	3,305	0,336	7,36E-23	++++	63,7	4,09E-02	6,58E-22
Triglycerides in LDL	3,360	0,343	1,24E-22	++++	83,9	3,29E-04	1,05E-21
Triglycerides in very large VLDL	3,380	0,347	2,28E-22	++++	15	3,17E-01	1,84E-21
Phospholipids in large VLDL	3,371	0,351	6,60E-22	++++	31,6	2,23E-01	5,06E-21

Total cholesterol in large VLDL	3,361	0,350	7,43E-22	++++	44,4	1,45E-01	5,44E-21
Triglycerides in large LDL	3,271	0,344	1,94E-21	++++	83,3	4,53E-04	1,36E-20
Triglycerides in HDL	3,188	0,335	1,94E-21	++++	90,4	7,57E-07	1,30E-20
Serum total triglycerides	3,290	0,349	3,83E-21	++++	36,6	1,93E-01	2,47E-20

The table shows metabolites significantly associated with liver enzyme (GGT) ordered by the P-value of Model 2 (FDR <0.05). Model 2: NAFLD~Metabolite + age + sex+lipid lowering medication + alcohol consumption+ body mass index + smoking status. Abbreviation: RS, Rotterdam study; GGT, Gamma-glutamyl Transferase; HDL, High-density lipoprotein; LDL, Low-density lipoprotein; IDL, Intermediate-density lipoprotein; VLDL, Very low- density lipoprotein; N, sample size

Table S10. List of top 50 metabolites significantly associated with liver enzymes (ALT) a meta-analysis of the RS, IRASFS-AA, IRASFS-HA and SOL (N=6059)

	Model2						
Metabolites	**Beta**	**SE**	**P-value**	**Direction**	I^2	**Het.P-value**	**FDR**
N-acetylmethionine	-0,075	0,007	2,12E-30	-+--	97,5	1,26E-25	1,00E-27
sphingomyelin (d18:0/18:0, d19:0/17:0)*	0,070	0,007	1,85E-21	+-++	80,7	1,41E-03	4,38E-19
sphingomyelin (d18:0/20:0, d16:0/22:0)*	0,069	0,007	2,80E-21	++++	94,3	1,76E-11	4,41E-19
glutamate	0,059	0,007	3,62E-15	+-++	66,9	2,83E-02	4,28E-13
1-stearoyl-2-oleoyl-GPE (18:0/18:1)	0,051	0,007	9,57E-14	+-++	57,1	7,23E-02	9,05E-12
1-stearoyl-2-arachidonoyl-GPE (18:0/20:4)	0,045	0,006	1,13E-12	+-++	30,3	2,30E-01	8,91E-11
cholesterol	0,049	0,007	1,39E-12	-+++	95,3	1,04E-13	9,39E-11
N-palmitoyl-sphingosine (d18:1/16:0)	0,047	0,007	5,56E-12	++++	29,7	2,34E-01	3,29E-10
1-stearoyl-2-arachidonoyl-GPC (18:0/20:4)	0,049	0,007	1,41E-11	+-++	70,3	1,77E-02	7,41E-10
X - 21471	0,046	0,007	3,08E-11	+-++	89,8	1,71E-06	1,46E-09
1-stearoyl-2-oleoyl-GPC (18:0/18:1)	0,044	0,007	3,68E-11	+-++	89,9	1,65E-06	1,58E-09
diacylglycerol (16:1/18:2 [2], 16:0/18:3 [1])*	0,045	0,007	4,07E-11	++++	0	4,88E-01	1,60E-09
1-stearoyl-2-linoleoyl-GPE (18:0/18:2)*	0,042	0,007	4,03E-10	+-++	81	1,24E-03	1,47E-08
1-palmitoyl-2-arachidonoyl-GPI (16:0/20:4)*	0,042	0,007	6,03E-10	++++	0	4,63E-01	2,04E-08
1-stearoyl-2-linoleoyl-GPC (18:0/18:2)*	0,040	0,007	6,45E-10	+-++	94,9	1,14E-12	2,03E-08
4-methyl-2-oxopentanoate	0,050	0,008	7,72E-10	++++	93,8	1,95E-10	2,28E-08
myristoyl dihydrosphingomyelin (d18:0/14:0)*	0,042	0,007	1,77E-09	+-++	6,4	3,61E-01	4,92E-08
aconitate [cis or trans]	0,042	0,007	2,98E-09	+-++	72,4	1,24E-02	7,83E-08
alpha-ketoglutarate	0,041	0,007	3,05E-09	++++	77,6	3,83E-03	7,59E-08
1-stearoyl-2-arachidonoyl-GPI (18:0/20:4)	0,039	0,007	5,06E-09	+-++	0	8,67E-01	1,20E-07
2-oxoarginine*	0,042	0,007	5,50E-09	++++	83,5	3,97E-04	1,24E-07
X - 24588	0,043	0,007	7,37E-09	+-++	51	1,06E-01	1,58E-07
sphingomyelin (d18:1/20:0, d16:1/22:0)*	0,040	0,007	8,25E-09	-+++	87,2	3,41E-05	1,70E-07
1-palmitoyl-2-linoleoyl-GPI (16:0/18:2)	0,038	0,007	1,04E-08	++++	0	7,26E-01	2,05E-07
oleoyl-linoleoyl-glycerol (18:1/18:2) [2]	0,039	0,007	1,20E-08	+-++	82,5	6,69E-04	2,27E-07
1-palmitoyl-2-stearoyl-GPC (16:0/18:0)	0,039	0,007	2,01E-08	-+++	70,4	1,74E-02	3,66E-07

1-palmitoyl-2-arachidonoyl-GPC (16:0/20:4n6)	0,038	0,007	3,62E-08	+-++	61,6	5,01E-02	6,34E-07
leucine	0,045	0,008	3,84E-08	++++	94,3	2,15E-11	6,49E-07
1-(1-enyl-palmitoyl)-2-arachidonoyl-GPE (P-16:0/20:4)*	0,038	0,007	4,17E-08	+-++	50,4	1,09E-01	6,80E-07
oleoyl-linoleoyl-glycerol (18:1/18:2) [1]	0,036	0,007	1,41E-07	+-++	0	9,81E-01	2,22E-06
sphingomyelin (d18:1/14:0, d16:1/16:0)*	0,035	0,007	1,70E-07	+-++	60,3	5,59E-02	2,59E-06
xanthine	0,036	0,007	1,75E-07	++++	66,4	3,04E-02	2,59E-06
cerotoylcarnitine (C26)*	0,039	0,008	1,82E-07	+-++	7,2	3,57E-01	2,61E-06
valine	0,040	0,008	2,01E-07	+-++	79,3	2,30E-03	2,80E-06
1-palmitoyl-2-oleoyl-GPC (16:0/18:1)	0,035	0,007	3,19E-07	+-++	0	6,74E-01	4,31E-06
tyrosine	0,036	0,007	3,66E-07	+-++	62,4	4,62E-02	4,81E-06
1-palmitoyl-2-linoleoyl-GPC (16:0/18:2)	0,033	0,007	4,15E-07	++++	74,8	7,75E-03	5,31E-06
3-methyl-2-oxovalerate	0,041	0,008	5,02E-07	++++	82	8,29E-04	6,25E-06
5alpha-pregnan-3beta,20alpha-diol monosulfate (2)	-0,034	0,007	1,02E-06	----	90,5	6,93E-07	1,24E-05
gamma-glutamylvaline	0,037	0,008	1,10E-06	++++	87,5	2,46E-05	1,30E-05
1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1)*	0,033	0,007	1,39E-06	++++	73,4	1,04E-02	1,60E-05
glycosyl-N-palmitoyl-sphingosine (d18:1/16:0)	0,032	0,007	1,72E-06	-+++	98,3	1,44E-38	1,94E-05
1-palmitoyl-2-arachidonoyl-GPE (16:0/20:4)*	0,033	0,007	1,73E-06	+-++	12,5	3,30E-01	1,90E-05
X - 16087	0,035	0,007	1,77E-06	++++	46,4	1,33E-01	1,90E-05
X - 21467	0,033	0,007	1,83E-06	++++	64,5	3,75E-02	1,92E-05
1-stearoyl-2-linoleoyl-GPI (18:0/18:2)	0,031	0,007	2,71E-06	+-++	0	5,17E-01	2,79E-05
gamma-glutamylglutamate	0,035	0,008	3,21E-06	+-++	54,3	8,70E-02	3,23E-05
palmitoyl dihydrosphingomyelin (d18:0/16:0)*	0,032	0,007	3,49E-06	--++	35,7	1,98E-01	3,44E-05
2-hydroxybutyrate/2-hydroxyisobutyrate	0,032	0,007	3,73E-06	+-++	80,4	1,57E-03	3,60E-05
glycochenodeoxycholate glucuronide (1)	0,031	0,007	3,78E-06	++++	81,6	9,68E-04	3,58E-05
trans-urocanate	-0,031	0,007	4,18E-06	----	87	3,82E-05	3,88E-05

Ordered by the P-value of Model 2 (FDR <0.05). Model 2: Liver enzymes (ALT)~Metabolite + age + sex+lipid lowering medication + alcohol consumption+ body mass index + smoking status; Direction: Direction of the effect estimate in each study of the meta-analysis (order – RS-III-2, SOL, IRASFS-HA, and IRASFS-AA); ALT: Alanine aminotransferase; I^2: heterogeneity parameter, Het. P-value: significance of the heterogeneity; FDR, False discovery rate.

Table S11. List of top 50 metabolites significantly associated with liver enzymes (AST) a meta-analysis of the RS, IRASFS-AA, IRASFS-HA and SOL (N=6092)

	Model2						
Metabolites	**Beta**	**SE**	**P-value**	**Direction**	I^2	**Het.P-value**	**FDR**
N-acetylmethionine	-0,038	0,004	1,68E-24	----	99,8	0,00E+00	8,00E-22
aconitate [cis or trans]	0,041	0,004	1,48E-23	++++	89,8	1,75E-06	3,52E-21
alpha-ketoglutarate	0,034	0,004	1,34E-18	++++	98,8	2,29E-52	2,13E-16
adenosine 5'-monophosphate (AMP)	0,032	0,004	9,18E-17	+-++	33,7	2,10E-01	1,09E-14
xanthine	0,031	0,004	1,29E-15	++++	98,7	6,73E-50	1,23E-13
sphingomyelin (d18:0/18:0, d19:0/17:0)*	0,033	0,004	5,73E-15	++++	97,5	3,60E-26	4,55E-13

sphingomyelin (d18:0/20:0, d16:0/22:0)*	0,031	0,004	9,27E-15	++++	96,1	1,57E-16	6,30E-13
glutamate	0,032	0,004	1,96E-14	++++	97,7	2,02E-28	1,17E-12
myristoyl dihydrosphingomyelin (d18:0/14:0)*	0,029	0,004	1,41E-13	++++	96,1	1,25E-16	6,71E-12
palmitoyl dihydrosphingomyelin (d18:0/16:0)*	0,029	0,004	1,29E-13	++++	97,1	5,77E-22	6,82E-12
glycocholate	0,027	0,004	3,67E-13	++++	98,7	1,18E-51	1,59E-11
3-ureidopropionate	0,027	0,004	6,54E-12	++++	99,5	2,62E-128	2,59E-10
myristoylcarnitine (C14)	0,026	0,004	1,01E-11	++++	96,6	4,10E-19	3,70E-10
X - 24588	0,029	0,004	1,67E-11	++++	89	4,95E-06	5,68E-10
beta-citrylglutamate	0,026	0,004	8,73E-11	-+++	69	2,16E-02	2,77E-09
1,2-dipalmitoyl-GPC (16:0/16:0)	0,025	0,004	1,04E-10	++++	97,8	4,12E-29	2,91E-09
palmitoylcarnitine (C16)	0,026	0,004	9,87E-11	++++	94,7	2,52E-12	2,94E-09
gamma-glutamylglutamate	0,027	0,004	1,69E-10	-+++	71	1,59E-02	4,47E-09
1-palmitoyl-2-linoleoyl-GPI (16:0/18:2)	0,024	0,004	2,49E-10	++++	96,7	7,19E-20	6,24E-09
1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1)*	0,024	0,004	2,80E-10	++++	98	7,28E-33	6,66E-09
X - 21470	0,025	0,004	3,26E-10	++++	97,2	2,02E-23	7,39E-09
1-stearoyl-2-oleoyl-GPC (18:0/18:1)	0,024	0,004	3,70E-10	++++	96,4	3,57E-18	8,01E-09
X - 21471	0,024	0,004	7,22E-10	++++	98,3	2,00E-37	1,49E-08
hexanoylglutamine	0,024	0,004	9,29E-10	++++	97,4	4,24E-25	1,84E-08
glycosyl-N-palmitoyl-sphingosine (d18:1/16:0)	0,023	0,004	1,13E-09	++++	95,4	4,01E-14	2,15E-08
choline phosphate	0,024	0,004	1,70E-09	+-++	90,3	8,80E-07	3,11E-08
X - 21796	0,023	0,004	1,78E-09	++++	94,1	4,46E-11	3,14E-08
glycochenodeoxycholate	0,022	0,004	2,46E-09	++++	98,3	5,18E-38	4,18E-08
sphingosine 1-phosphate	0,026	0,004	2,62E-09	+-++	28,6	2,41E-01	4,30E-08
acetylcarnitine (C2)	0,022	0,004	6,15E-09	++++	95,4	4,15E-14	9,76E-08
taurochenodeoxycholate	0,023	0,004	7,39E-09	++++	99,1	4,11E-74	1,13E-07
tyrosine	0,024	0,004	1,06E-08	++++	97,7	8,66E-28	1,58E-07
aspartate	0,023	0,004	1,21E-08	++++	97	2,38E-21	1,75E-07
gamma-glutamyltyrosine	0,024	0,004	1,49E-08	++++	97,3	2,98E-24	2,09E-07
1-palmitoyl-2-oleoyl-GPC (16:0/18:1)	0,022	0,004	1,54E-08	++++	95,3	6,71E-14	2,09E-07
nicotinamide	0,022	0,004	1,60E-08	++++	80	1,82E-03	2,12E-07
docosadienoate (22:2n6)	0,022	0,004	3,76E-08	++++	97,1	1,56E-22	4,84E-07
glucuronate	0,022	0,004	3,87E-08	++++	92	3,67E-08	4,85E-07
palmitoleoylcarnitine (C16:1)*	0,021	0,004	5,37E-08	++++	97,3	1,08E-23	6,55E-07
caprylate (8:0)	0,020	0,004	1,26E-07	++++	95	6,06E-13	1,50E-06
sphingomyelin (d18:2/18:1)*	-0,022	0,004	1,81E-07	----	96,9	1,21E-20	2,10E-06
X - 12026	-0,023	0,004	2,49E-07	----	94,2	2,77E-11	2,82E-06
glycochenodeoxycholate glucuronide (1)	0,020	0,004	2,66E-07	++++	97,3	1,18E-23	2,94E-06
cholesterol	0,020	0,004	3,44E-07	++++	85,5	1,19E-04	3,72E-06
3-(4-hydroxyphenyl)lactate	0,022	0,004	3,61E-07	++++	97,9	2,05E-31	3,82E-06
orotidine	0,020	0,004	4,54E-07	-+++	49,7	1,14E-01	4,70E-06
N-palmitoyl-sphingosine (d18:1/16:0)	0,020	0,004	5,37E-07	++++	95,6	1,19E-14	5,44E-06
1-palmitoyl-2-stearoyl-GPC (16:0/18:0)	0,020	0,004	7,41E-07	++++	70,5	1,71E-02	7,35E-06

X - 12462	0,019	0,004	7,89E-07	--++	0	6,25E-01	7,66E-06
1-palmitoyl-2-arachidonoyl-GPI (16:0/20:4)*	0,019	0,004	8,53E-07	++++	89,3	3,41E-06	8,12E-06

Ordered by the P-value of Model 2 (FDR <0.05). Model 2: Liver enzymes (AST)~Metabolite + age + sex+lipid lowering medication + alcohol consumption+ body mass index + smoking status.; Direction: Direction of the effect estimate in each study of the meta-analysis (order – RS-III-2, SOL, IRASFS-HA, and IRASFS-AA); AST: Aspartaat aminotransferase; I^2: heterogeneity parameter, Het. P-value: significance of the heterogeneity; FDR, False discovery rate.

Table S12. List of top 50 metabolites significantly associated with liver enzymes (GGT) a meta-analysis of the RS, IRASFS-AA, IRASFS-HA and SOL (N=6358)

	Model2						
Metabolites	**Beta**	**SE**	**P-value**	**Direction**	I^2	**Het.P-value**	**FDR**
glutamate	0,101	0,007	2,96E-47	+++++	99,1	2,05E-97	1,37E-44
sphingomyelin (d18:0/18:0, d19:0/17:0)*	0,078	0,007	3,91E-26	+++++	96,2	3,95E-22	9,03E-24
1-stearoyl-2-oleoyl-GPC (18:0/18:1)	0,067	0,007	2,12E-24	+++++	97,9	3,70E-40	3,26E-22
1-palmitoyl-2-oleoyl-GPC (16:0/18:1)	0,064	0,007	2,58E-22	+++++	97,6	9,06E-35	2,98E-20
1-stearoyl-2-oleoyl-GPE (18:0/18:1)	0,066	0,007	6,49E-22	+++++	97,8	5,94E-39	6,00E-20
N-palmitoyl-sphingosine (d18:1/16:0)	0,065	0,007	6,92E-22	+-+++	96,4	5,14E-23	5,33E-20
sphingomyelin (d18:0/20:0, d16:0/22:0)*	0,068	0,007	8,71E-22	+++++	93,6	7,24E-13	5,75E-20
gamma-glutamylvaline	0,071	0,007	1,40E-21	+++++	86,1	9,05E-06	8,09E-20
1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1)*	0,063	0,007	1,89E-21	+++++	98,6	7,09E-63	9,70E-20
alpha-ketoglutarate	0,062	0,007	1,65E-20	+++++	98,4	6,17E-52	7,62E-19
X - 24588	0,067	0,007	5,04E-20	+++++	92,2	2,17E-10	2,12E-18
1-stearoyl-2-arachidonoyl-GPE (18:0/20:4)	0,060	0,007	1,89E-18	+++++	96,3	9,09E-23	7,28E-17
taurochenodeoxycholate	0,058	0,007	1,46E-17	+++++	98,2	1,98E-47	5,19E-16
cholesterol	0,058	0,007	5,06E-17	+++++	94,4	1,52E-14	1,67E-15
glycocholate	0,053	0,007	2,30E-16	+-+++	97,8	8,71E-39	7,08E-15
1,2-dipalmitoyl-GPC (16:0/16:0)	0,055	0,007	3,73E-16	+-+++	96,9	2,87E-27	1,08E-14
1-palmitoyl-2-arachidonoyl-GPI (16:0/20:4)*	0,053	0,007	9,63E-16	+++++	94,3	1,77E-14	2,62E-14
1-stearoyl-2-linoleoyl-GPE (18:0/18:2)*	0,055	0,007	1,02E-15	+++++	96,8	4,49E-26	2,62E-14
X - 21470	0,056	0,007	1,25E-15	+-+++	98,2	8,06E-48	3,04E-14
N-acetylmethionine	-0,053	0,007	3,90E-15	-----	98,9	1,05E-79	9,01E-14
1-palmitoyl-2-linoleoyl-GPI (16:0/18:2)	0,051	0,007	4,14E-15	+++++	95,4	6,90E-18	9,11E-14
1-palmitoyl-2-stearoyl-GPC (16:0/18:0)	0,054	0,007	4,87E-15	+++++	91	5,72E-09	1,02E-13
1-stearoyl-2-arachidonoyl-GPC (18:0/20:4)	0,055	0,007	4,18E-14	+++++	80,5	3,98E-04	8,40E-13
gamma-glutamylisoleucine*	0,059	0,008	6,18E-14	+++++	90	4,24E-08	1,19E-12
myristoyl dihydrosphingomyelin (d18:0/14:0)*	0,051	0,007	9,39E-14	+++++	94,4	9,67E-15	1,74E-12
X - 21441	0,051	0,007	2,58E-13	+-+++	94,6	4,28E-15	4,58E-12
gamma-glutamylglutamine	-0,047	0,007	1,11E-12	-+---	98,4	1,23E-54	1,90E-11
1-palmitoyl-2-arachidonoyl-GPE (16:0/20:4)*	0,049	0,007	1,33E-12	+++++	97,1	9,89E-29	2,19E-11
palmitoylcarnitine (C16)	0,049	0,007	1,50E-12	+++++	95,3	1,75E-17	2,39E-11

gamma-glutamylleucine	0,058	0,008	1,98E-12	-++++	41,7	1,44E-01	3,05E-11
aconitate [cis or trans]	0,050	0,007	2,12E-12	+++++	76,3	2,05E-03	3,16E-11
1-palmitoyl-2-arachidonoyl-GPC (16:0/20:4n6)	0,050	0,007	2,57E-12	+++++	86,6	5,52E-06	3,71E-11
arabitol/xylitol	0,049	0,007	8,26E-12	+++++	97,1	2,18E-28	1,16E-10
1-stearoyl-GPE (18:0)	0,047	0,007	9,57E-12	+++++	95,7	2,82E-19	1,30E-10
docosadienoate (22:2n6)	0,047	0,007	2,95E-11	+-+++	98	4,11E-41	3,89E-10
N-acetylphenylalanine	0,047	0,007	3,09E-11	+++++	96,8	6,67E-26	3,97E-10
gamma-glutamyltyrosine	0,050	0,008	3,30E-11	+++++	85,3	1,83E-05	4,12E-10
lactate	0,045	0,007	3,34E-11	+-+++	79,4	6,61E-04	4,06E-10
sphingomyelin (d18:1/14:0, d16:1/16:0)*	0,045	0,007	3,73E-11	+++++	71,1	7,82E-03	4,42E-10
glycochenodeoxycholate	0,043	0,007	3,91E-11	+++++	96,8	6,80E-26	4,52E-10
1-palmitoyl-2-linoleoyl-GPE (16:0/18:2)	0,044	0,007	1,12E-10	+++++	96,3	2,96E-22	1,26E-09
1-palmitoyl-GPI (16:0)	0,046	0,007	1,19E-10	+++++	93,2	4,27E-12	1,31E-09
3-ureidopropionate	0,045	0,007	1,25E-10	+++++	98,3	1,43E-48	1,34E-09
glucuronate	0,046	0,007	1,26E-10	+++++	89,2	1,89E-07	1,32E-09
2-hydroxy-3-methylvalerate	0,049	0,008	1,88E-10	+++++	98,5	5,49E-55	1,93E-09
alpha-hydroxyisovalerate	0,047	0,007	2,31E-10	+++++	98,6	1,48E-62	2,32E-09
docosapentaenoate (n6 DPA; 22:5n6)	0,044	0,007	2,58E-10	+++++	95,9	4,67E-20	2,54E-09
X - 21471	0,044	0,007	3,06E-10	+++++	96,5	1,70E-23	2,95E-09
1-stearoyl-2-linoleoyl-GPC (18:0/18:2)*	0,042	0,007	3,40E-10	+++++	91,1	4,31E-09	3,21E-09
3-methyl-2-oxobutyrate	0,045	0,007	3,57E-10	+-+++	89,8	6,56E-08	3,30E-09

Ordered by the P-value of Model 2 (FDR <0.05). Model 2: Liver enzymes (GGT)~Metabolite + age + sex+lipid lowering medication + alcohol consumption+ body mass index + smoking status.; Direction: Direction of the effect estimate in each study of the meta-analysis (order – RS-III-2, RS-I-4, SOL, IRASFS-HA and IRASFS-AA); GGT: Gamma-glutamyl Transferase; I^2: heterogeneity parameter, Het. P-value: significance of the heterogeneity; FDR, False discovery rate.

Table S13. Metabolites significantly associated and overlapping among NAFLD-US, NASH and liver fibrosis in Rotterdam Study (Nightingale platform)

	NAFLD-US			NASH			Liver fibrosis		
Metabolites	**Beta**	**SE**	**P-value**	**Beta**	**SE**	**P-value**	**Beta**	**SE**	**FDR**
Apolipoprotein B	0,320	0,065	2,04E-06	0,340	0,140	1,54E-02	0,318	0,138	2,14E-02
Glucose	0,503	0,068	1,10E-12	0,389	0,119	1,12E-03	0,372	0,106	4,68E-04
Lactate	0,308	0,061	7,49E-07	0,316	0,128	1,36E-02	0,282	0,127	2,70E-02
Tyrosine	0,270	0,063	3,47E-05	0,375	0,138	6,50E-03	0,304	0,134	2,30E-02
Cholesterol esters in chylomicrons and extremely large VLDL	0,348	0,067	4,10E-07	0,252	0,100	1,13E-02	0,380	0,172	2,74E-02
Cholesterol esters in large HDL	-0,775	0,083	3,21E-19	-0,406	0,175	2,00E-02	-0,385	0,152	1,12E-02
Cholesterol esters in large VLDL	0,418	0,066	6,57E-10	0,248	0,105	1,88E-02	0,400	0,164	1,48E-02
Cholesterol esters in medium VLDL	0,364	0,062	1,31E-08	0,274	0,124	2,68E-02	0,342	0,152	2,42E-02
Cholesterol esters in very large VLDL	0,398	0,068	1,31E-08	0,237	0,096	1,38E-02	0,374	0,168	2,58E-02

Concentration of large HDL particles	-0,772	0,085	1,45E-18	-0,387	0,176	2,81E-02	-0,399	0,152	8,75E-03
Concentration of large VLDL particles	0,445	0,069	3,58E-10	0,249	0,096	9,68E-03	0,343	0,159	3,11E-02
Concentration of medium VLDL particles	0,449	0,067	7,43E-11	0,315	0,124	1,09E-02	0,374	0,149	1,19E-02
Concentration of small VLDL particles	0,472	0,067	8,70E-12	0,383	0,145	8,42E-03	0,310	0,141	2,77E-02
Free cholesterol in large HDL	-0,806	0,087	5,71E-19	-0,439	0,189	1,99E-02	-0,357	0,153	1,95E-02
Free cholesterol in large VLDL	0,443	0,068	2,12E-10	0,268	0,098	5,96E-03	0,318	0,154	3,96E-02
Free cholesterol in medium VLDL	0,454	0,066	1,96E-11	0,341	0,126	6,86E-03	0,333	0,144	2,03E-02
Glycoprotein acetyls_ mainly a1_acid glycoprotein	0,482	0,066	1,15E-12	0,348	0,131	7,90E-03	0,358	0,145	1,38E-02
Mean diameter for HDL particles	-0,802	0,084	1,19E-19	-0,411	0,173	1,76E-02	-0,300	0,150	4,61E-02
Mean diameter for VLDL particles	0,579	0,071	3,32E-15	0,306	0,138	2,68E-02	0,291	0,143	4,24E-02
Phenylalanine	0,156	0,063	2,27E-02	0,265	0,130	4,12E-02	0,297	0,130	2,28E-02
Phospholipids in large HDL	-0,697	0,078	4,58E-18	-0,330	0,156	3,44E-02	-0,406	0,155	8,66E-03
Phospholipids in large VLDL	0,461	0,067	3,34E-11	0,278	0,101	5,99E-03	0,329	0,154	3,22E-02
Phospholipids in medium VLDL	0,465	0,066	5,99E-12	0,338	0,125	7,07E-03	0,358	0,146	1,39E-02
Ratio of apolipoprotein B to apolipoprotein A_I	0,536	0,070	1,17E-13	0,380	0,151	1,15E-02	0,354	0,142	1,25E-02
Ratio of triglycerides to phosphoglycerides	0,621	0,072	6,84E-17	0,360	0,147	1,41E-02	0,294	0,145	4,31E-02
Total cholesterol in chylomicrons and extremely large VLDL	0,371	0,068	1,17E-07	0,253	0,094	7,06E-03	0,327	0,164	4,59E-02
Total cholesterol in large HDL	-0,775	0,082	3,58E-19	-0,408	0,174	1,91E-02	-0,380	0,152	1,24E-02
Total cholesterol in large VLDL	0,439	0,066	1,25E-10	0,269	0,103	9,20E-03	0,354	0,157	2,38E-02
Total cholesterol in medium VLDL	0,419	0,063	1,48E-10	0,309	0,124	1,29E-02	0,341	0,146	1,93E-02
Total cholesterol in very large VLDL	0,398	0,068	1,28E-08	0,249	0,095	8,49E-03	0,344	0,163	3,53E-02
Total cholesterol in VLDL	0,418	0,066	8,87E-10	0,363	0,143	1,12E-02	0,271	0,133	4,20E-02
Total lipids in large HDL	-0,722	0,078	4,99E-19	-0,360	0,158	2,28E-02	-0,402	0,154	8,92E-03
Total lipids in large VLDL	0,499	0,066	3,36E-13	0,320	0,114	4,87E-03	0,323	0,146	2,74E-02
Total lipids in medium VLDL	0,499	0,066	2,36E-13	0,349	0,131	7,91E-03	-0,358	0,142	1,18E-02
Total lipids in small VLDL	0,480	0,067	4,18E-12	0,381	0,146	8,81E-03	0,305	0,140	3,00E-02
Total lipids in very large VLDL	0,442	0,067	1,72E-10	0,272	0,099	5,98E-03	0,299	0,150	4,69E-02
Triglycerides in large VLDL	0,493	0,067	1,06E-12	0,304	0,109	5,42E-03	0,326	0,149	2,86E-02
Triglycerides in medium VLDL	0,504	0,066	2,30E-13	0,347	0,132	8,45E-03	0,373	0,144	9,32E-03
Triglycerides in small VLDL	0,511	0,068	5,68E-13	0,392	0,147	7,68E-03	0,319	0,143	2,59E-02

Triglycerides in VLDL	0,553	0,069	1,43E-14	0,422	0,146	3,92E-03	0,314	0,141	2,58E-02

The table shows metabolites assessed by Nightingale significantly associated and overlapping among NAFLD-US, NASH, and liver fibrosis. The significant threshold in NAFLD-US is (FDR <0.05) while for NASH and liver fibrosis is nominal (P<0.05). Model 2: adjusted for age + sex + lipid-lowering medication + alcohol consumption + body mass index + smoking status. Abbreviation: SE, standard error; NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis; US, Ultra-sound; HDL, High-density lipoprotein; LDL, Low-density lipoprotein; IDL, Intermediate-density lipoprotein; VLDL, Very low- density lipoprotein; FDR, false discovery rate.

Table S14. Metabolites significantly associated and overlapping among NAFLD-US, NASH and liver fibrosis in Rotterdam Study (Metabolon platform)

	NAFLD-US			NASH			Liver fibrosis		
Metabolites	**Beta**	**SE**	**P-value**	**Beta**	**SE**	**P-value**	**Beta**	**SE**	**FDR**
1-carboxyethylisoleucine	0,831	0,170	1,51E-04	0,534	0,264	4,28E-02	0,108	0,226	6,33E-01
1-carboxyethylleucine	1,200	0,221	1,36E-05	0,614	0,282	2,98E-02	0,179	0,261	4,93E-01
4-hydroxyglutamate	1,059	0,170	4,80E-07	0,670	0,214	1,72E-03	-0,049	0,189	7,94E-01
argininate*	0,683	0,188	6,65E-03	-0,114	0,231	6,22E-01	0,727	0,257	4,63E-03
gamma-glutamylisoleucine*	1,294	0,325	2,60E-03	-0,339	0,279	2,25E-01	0,868	0,393	2,71E-02
1-linoleoyl-2-arachidonoyl-GPC (18:2/20:4n6)*	-0,115	0,278	8,61E-01	1,122	0,549	4,11E-02	1,164	0,546	3,29E-02
orotidine	0,249	0,148	3,24E-01	0,676	0,301	2,49E-02	0,431	0,203	3,43E-02
p-cresol glucuronide*	0,093	0,079	5,26E-01	-0,259	0,109	1,78E-02	-0,235	0,111	3,46E-02

The table shows metabolites assessed by metabolon platform significantly associated and overlapping among NAFLD-US, NASH, and liver fibrosis. The significant threshold in NAFLD-US is (FDR <0.05) while for NASH and liver fibrosis is nominal (P<0.05). Model 2: adjusted for age + sex + lipid-lowering medication + alcohol consumption + body mass index + smoking status. Abbreviation: SE, standard error; NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis; US, Ultra-sound; FDR, false discovery rate.

Table S15. Supporting evidence for the association between the 21 overlapping metabolites and NAFLD reported in previous studies

Metabolites	**Expression at NAFLD**	**Role**	**Reference**
A- Nightingale platform			
Triglycerides in HDL	Up-regulated	Insulin resistance might reduce HDL levels by affecting its assembly. HDL forms in plasma at the hepatocyte surface, involving ApoA-1 and ATP-binding cassette transporter A1 (ABCA1) interaction. Insulin resistance interferes with HDL formation by increasing ABCA1 degradation and reducing its activity. It can also lead to triglyceride-rich HDL formation via cholesterol ester transfer protein (CETP) action.	Kaikkonen JE et al. Cali AMG et al. Arvind A et al. Zannis VI et al. Nonomura K et al. Lucero D et al.

Triglycerides in IDL	Up-regulated	IDLs are intermediate between very-low-density lipoproteins (VLDL) and low-density lipoproteins (LDL). These particles are involved in transporting triglycerides and cholesterol throughout the body. In the context of NAFLD, elevated IDL levels can contribute to the deposition of triglycerides and cholesterol in the liver, further exacerbating fat accumulation.	Kaikkonen JE et al. Miksztowicz, V et al. Tulenko, T.N. et al.
Phenylalanine	Up-regulated	Elevated levels of phenylalanine have been linked to insulin resistance, oxidative stress, inflammation, and adipose tissue dysfunction. These factors are key players in the development and progression of NAFLD. Additionally, phenylalanine is a precursor for several metabolites involved in pathways related to lipid metabolism and insulin signaling, further influencing the pathogenesis of NAFLD.	Kaikkonen JE et al. Kalhan SC et al. Lake AD et al. Cheng Set al. Gaggini M et al.
Ratio of 18:2 linolieic acid to total fatty acid	Down-regualted	Dietary linoleic acid, an essential omega-6 polyunsaturated fatty acid, plays a role in modulating inflammation and oxidative stress within the liver. It promotes the production of anti-inflammatory molecules and can inhibit pro-inflammatory pathways, contributing to a reduced inflammatory environment in the liver. Additionally, linoleic acid has been shown to influence hepatic lipid metabolism, facilitating the breakdown of fatty acids and potentially reducing lipid accumulation in the liver, a key factor in the progression of fibrosis.	Gaggini M et al.
Ratio of omega 6 fatty acid to total fatty acid	Down-regualted	Omega-6 fatty acids are essential nutrients and play important roles in various physiological processes, an imbalance in their intake, particularly when combined with other dietary and lifestyle factors, can contribute to inflammation, oxidative stress, and insulin resistance, all of which are associated with NAFLD development and progression.	Bonafini S et al.
B- Metabolon platform			
1-myristoyl-2-palmitoyl-GPC (14:0/16:0) 1-palmitoyl-2-arachidonoyl-GPI (16:0/20:4) 1-palmitoyl-2-linoleoyl-GPE (16:0/18:2) 4-methyl-2-oxopentanoate	Up-regulated	They concluded that these lipids and amino acids are associated with dynamic and basal measures of glucose homeostasis, which provides expanded insight into the metabolic basis of insulin resistance.	Palmer ND et al.
1-stearoyl-2-linoleoyl-GPE (18:0/18:2) 1-stearoyl-2-oleoyl-GPE (18:0/18:1)	Up-regulated	Its potential mechanism in the context of non-alcoholic fatty liver disease (NAFLD) may involve its role in lipid metabolism and inflammation and oxidative stress.	Lind L et al. Marchisello, S et al.
aconitate (cis or trans)	Up-regulated	Its to be positively correlated with the accumulation of ectopic fat depots in non-diabetic male subjects with NAFLD	Lovric A et al.
gamma-glutamyl isoleucine gamma-glutamyl valine	Up-regulated	the mechanism in NAFLD involves its association with oxidative stress, and potential disruption of hepatic lipid homeostasis, contributing to the progression of liver steatosis and related metabolic dysregulation.	Gaggini M et al. Lovric A et al. Lo EKK et al.

The table shows the metabolites with its role that are significantly increased or decreased in both (A) Nightingale platform, and (B) Metabolon platform with NAFLD in model 2. Model 2 is adjusted for age, sex, , lipid-lowering medication + alcohol consumption+ body mass index + smoking status. Significant for the metabolites is sorted based on the FDR ($P < 0.05$).

Chapter 5

General discussion

Obesity has become a major epidemic in the 21st century, leading to an increased risk of various health conditions such as dyslipidemia, hypertension, type 2 diabetes, and fatty liver disease (FLD).[1] FLD is believed to be involved in the pathogenesis of common disorders such as of type 2 diabetes and cardiovascular disease (CVD), and the global burden of FLD parallels the increase in obesity rates across the world.[2] Non-alcoholic fatty liver disease (NAFLD) is currently the leading cause of chronic liver disease in Western countries and represents an alarming global health crisis, affecting >33% of the population,[3-5] To complicate matters further, it is also predicted to become also the most frequent indication for liver transplantation by 2030.[6]

In recent years, the advent of the "omics revolution" has enabled population-based studies to adopt multi-omics approaches, integrating data from different omics levels to gain a better understanding of mechanisms underlying complex diseases. In this thesis I used multi-omics data mainly from the population-based Rotterdam Study cohort to further investigate the etiology of obesity and FLD. In particular, transcriptomics, proteomics, and metabolomics were used to elucidate the molecular mechanisms underlying these metabolic diseases and to identify new non-invasive biomarkers for their early diagnosis. Leveraging large-scale omics data and applying advanced molecular epidemiological approaches could help to a deeper understanding of the molecular aspects of obesity and FLD, paving the way for improved diagnostic methods and potential therapeutic targets.

Main finding

Epigenetic of obesity and fatty liver disease

Recent advances in the study of epigenetic modifications have considerably increased our understanding of the function of genes and epigenetic mechanisms in regulating energy metabolism and expenditure in obesity and other metabolic diseases.[7, 8] These epigenetic modifications, unlike genetic variation, involve dynamic changes and are potentially reversible, this reversibility makes them modifiable to modification through lifestyle changes and other therapeutic interventions.

In **Chapter 2.1** of this thesis, I conducted a genome-wide profile of circulating miRNAs in the plasma of 1208 participants from the population-based Rotterdam Study cohort (RS-I-4 and RS-II-2) to investigate the association of circulating miRNAs with obesity, body fat distribution, and fat mass. I found plasma levels of 65 miRNAs to be associated with body mass index (BMI), 40 miRNAs with waist to hip ratio (WHR), 65 miRNAs with fat mass index (FMI), and 15 miRNAs with android fat to gynoid fat ratio (AGR), after correcting for multiple testing. Notably, a total of 12 miRNAs showed significant associations with all the traits examined. In addition, 4 miRNAs were specifically associated with WHR, 3 miRNAs with FMI, and miR-378i showed an association with AGR. Among these overlapping miRNAs, the most significant association was observed with miR-193a-5p. This particular miRNA has previously been linked to the risk of developing type 2 diabetes and FLD. These associations were

identified through independent studies conducted using miRNA data from the Rotterdam Study.[9, 10] Additionally, five of the obesity-associated miRNAs and two of the body fat distribution miRNAs have been correlated previously to cardiovascular diseases.[11-15] In females, the vast majority of the 12 common miRNAs were strongly associated and overlapped with all traits, whereas in males, these 12 miRNAs were less significant. Interestingly, we identified five new miRNAs (miR-185-5p, miR-19a-3p, miR-19b-3p, miR-7150, and miR-93-5p) that were specifically associated with the four traits only in females, surpassing the Bonferroni-corrected p-value threshold ($P < 8.46 \times 10^{-5}$). Moreover, our study identified seven previously unreported miRNAs, namely miR-3937, miR-4433b-5p, miR-4478, miR-6088, miR-6799-5p, miR-6803-5p, miR-6821-5p, and miR-1304-3p, that showed associations with all obesity-related traits, with miR-1304-3p specifically linked to fat mass. To our knowledge, our study is the first to report these associations with obesity and fat mass. It is noteworthy that several of these identified miRNAs have also been associated with cancer, as previous studies have indicated that approximately 20% of common cancers can be attributed to excessive body fat accumulation.[16, 17] Furthermore, in the longitudinal analysis, we found 24 miRNAs to be significantly associated with the prevalence of metabolic syndrome, of these, 10 miRNAs overlapped with the 12 common miRNAs associated with all four traits.

In **Chapter 2.2**, plasma levels of circulating miRNAs in the Rotterdam Study participants was used to investigate the association between miRNAs and FLD diagnosed by CT scan and liver enzymes. To understand if the miRNAs are involved in the pathways underlying liver diseases, we performed subsequent analysis and linked the identified miRNAs with the risk of FLD based-ultrasound and FibroScan-based liver fibrosis. In the cross-sectional analysis, I found significant associations between 61 miRNAs and serum levels of gamma-glutamyl transferase and/or alkaline phosphatase ($P < 8.46 \times 10^{-5}$ after Bonferroni correction). Additionally, among these miRNAs, 17 were significantly associated with CT-based fatty liver disease ($P < 8.46 \times 10^{-5}$), with 14 of them overlapping with miRNAs associated with liver enzymes. In the longitudinal analysis, I found that 4 out of the 14 identified miRNAs (miR-193a-5p, miR-122-5p, miR-378d, and miR-187-3p) were significantly associated with hepatic steatosis ($P < 3.57 \times 10^{-3}$), while three miRNAs (miR-193a-5p, miR-122-5p, and miR-193b-3p) showed nominal associations with liver fibrosis ($P < 0.05$). Furthermore, nine out of the 14 identified miRNAs were involved in pathways related to liver diseases. Additionally, I conducted a search in the Human miRNA tissue atlas and the miRmine database to examine the expression of the 14 plasma-associated miRNAs in the liver. Among these miRNAs, miR-122-5p was identified as a highly expressed miRNA in the liver and exhibited a high tissue specificity index (TSI) of 0.97, indicating its specific expression in this tissue.

Notably, miR-193a-5p that was the most significant miRNA associated with obesity-related traits in **Chapter 2.1**, was also associated with FLD in **Chapter 2.2**. This may indicate miR-193a-5p as a common miRNA to be involved in regulating metabolic pathways. This finding might be promising to be explored further in other population studies and in experimental settings to elucidate the pathophysiology of obesity-related FLD.

In **Chapter 2.3**, a genome-wide screening was conducted to investigate the association of plasma miRNAs with alcohol consumption, and to explore whether there is a mediating effect for the alcohol-associated miRNAs with liver function and disease. Recent studies have reported a vital role for epigenetic factors, which modulate gene expression in the absence of changes in DNA sequence, in the onset and progression of liver disorders toward hepatic fibrogenesis and cirrhosis.[18] Mounting findings have also delineated that alcohol consumption extensively modulates liver epigenetics, thus, prompting the etiology of alcoholic liver disease (ALD). In addition, increasing evidence suggests that miRNAs are involved in inflammation, lipid metabolism, and oncogenetics are affected by excessive alcohol administration in mouse models of ALD.[19] Our study explained in **Chapter 2.3** found plasma concentrations of miR-193b-3p, miR-122-5p, miR-3937, and miR-4507 to be significantly associated with alcohol consumption surpassing the Bonferroni-corrected $P < 8.46 \times 10^{-5}$. Notably, I also found that effect size estimates of alcohol-associated miRNAs were nearly doubled when comparing the mean consumption of heavy drinkers to non-drinkers. Most of the alcohol-associated miRNAs exhibited stronger effect sizes in men, except for miR-4507, which showed a more significant decrease in women with alcohol consumption. With the liver being primarily responsible for alcohol detoxification and metabolism, the study delved deeper into investigating the potential role of alcohol-associated miRNA levels as mediators in liver-related diseases. Among the four alcohol-associated miRNAs, miR-3937 and miR-122-5p demonstrated a potential mediating role in the association between alcohol consumption and CT-based fatty liver disease, GGT, and US-based steatosis. However, the Mendelian randomization (MR) analysis did not provide sufficient statistical evidence to establish a causal relationship between alcohol-associated miRNAs and liver-related traits. This could potentially be attributed to the lack of robust instrumental variables (IVs). After pruning the trans-miR-QTLs and correlated single-nucleotide polymorphisms (SNPs), only one SNP (rs30227) of miR-193bp remained as an instrumental variable. Consequently, the limited evidence of causality may have been hindered by the scarcity of strong IVs, particularly considering that only a single miRNA had a cis-expression quantitative trait locus (eQTL) available. This emphasizes the need for future studies to conduct genome-wide association studies (GWAS) on larger sample sizes and explore a broader landscape of miRNAs to generate more robust instrumental variables for estimating causal relationships.

Our research in **Chapter 2.2** revealing a significant association between miR122-5p and miR193b-3p with FLD are in line with our findings regarding miRNAs linked to alcohol consumption in **Chapter 2.3**. Furthermore, in **Chapter 2.3**, we discovered that miR-3937, which was previously identified as an overlapping miRNA in relation to obesity-related traits in **Chapter 2.1**, may play a mediating role between alcohol consumption and liver function. These findings further highlight the interconnected nature of miRNAs, alcohol consumption, and liver-related conditions, emphasizing the need for deeper investigation into their mechanisms and potential implications in future studies.

Proteomics and fatty liver disease

The advancement of proteomics analysis has provided us with powerful tools for studying FLD mechanisms and the opportunity for discovering novel biomarkers. Previous proteomics studies detected changes in the plasma proteome of patients with cirrhosis and NAFLD that are clearly linked to the underlying disease manifestations and clinical observations.[20] Furthermore, other studies reported several plasma and tissue specific proteins in FLD which have been identified as important diagnostic biomarkers for patients with cirrhosis and hepatocellular carcinoma.[21, 22] However, it is important to note that these studies often had relatively small sample sizes, which can increase the risk of missing relevant biomarkers in addition to the lack of disease stratification particularly, in the early stage of FLD (e.g. steatosis). In the Rotterdam Study, however, there was a large-scale plasma profiling of 171 proteins (two Olink panels, cardiometabolic and inflammatory-related proteins) with long-time follow-up for FLD available. Therefore in **Chapter 3.1**, I used these plasma proteomics data from 2,578 participants to identify proteins associated with FLD-based ultrasound and liver fibrosis. In our analysis, I found 27 proteins significantly associated with FLD surpassing the Bonferroni-corrected $P < 2.92\times10^{-4}$, the strongest association was observed for FGF-21 and CES1 proteins. Importantly, 15 of the 27 proteins significantly associated with FLD were also associated with liver fibrosis. In the Reactome pathway analysis, I found TWEAK and TNFS14 proteins to be involved in the non-canonical NF-kB pathway. While, F7, SERPINA5, and PROC proteins were involved in the formation of fibrin clots (clotting cascade). Interestingly, our in vitro experiments confirmed that the gene expression levels of *IL-18R1* and *CES1* were upregulated in 3D cultured primary liver organoids mimicking FLD. Finally, the area under the curve (AUC) analysis illustrated that proteomics has the potential to improve currently available non-invasive markers (FLI and FIB-4) in the prediction of FLD or fibrosis. The findings of this study could help towards a better understanding of the pathophysiology of FLD and the disease progression into fibrosis. In particular, FGF-21, CES1 and IL-18R1 seem to contribute to the pathogenesis of FLD and also fibrosis. Moreover, given an unmet need for accurate and widely available non-invasive tools for risk stratification and referral strategies, our findings may improve the algorithms currently under development.

Metabolites and risk of fatty liver disease

The burden of obesity and NAFLD is rising globally, but little is known about the metabolic pathways underlying these diseases at the population level.[23] Recent developments in high-throughput analysis, curation, and robust statistical analysis have allowed investigators to understand the changes in cellular and tissue metabolism based on metabolomics data.[24] In **Chapter 4.1**, I used both nuclear magnetic resonance (NMR) spectroscopy from the Nightingale platform (almost 250 metabolites) and mass spectrometry (MS) from the Metabolon platform with more metabolites (>1000), to investigate the association of plasma metabolites with NAFLD and liver enzymes. I used data from four population-based cohorts including the Rotterdam study, Avon Longitudinal Study of Parents and Children (ALSPAC) from the UK, and two USA based cohorts called The Insulin Resistance Atherosclerosis Family Study (IRASFS) and Study of Latino (SOL). Numerous metabolites from both platforms were significantly associated and overlapping with NAFLD and liver enzymes after adjusting for potential confounders and multiple testing corrections (FDR < 0.05) . Of these, 21 metabolites were jointly associated with NAFLD (diagnosed by ultrasound and CT-scan) and liver enzymes (ALT, AST, and GGT), among them 7 metabolites within Nightingales platform such as phenylalanine, triglycerides in (HDL, IDL, and small LDL), fatty acids (FAs) ratios of (18:2 linoleic acid to total FA, omega 6 FA to total FA, and polyunsaturated FA to total FA. In the Metabolon platform, I also found 14 common metabolites including glutamate and sphingomyelin to have the strongest association with both NAFLD and liver enzymes. Other associated metabolites were mainly involved in lipid, amino acid, carbohydrates, and peptide metabolism. A previous Finnish study (n=2,002) conducted based on young and middle age participants, using Nightingale platforms (N=68), has shown numerous metabolites to be significantly associated with fatty liver disease based on ultrasound. These metabolites included lipoprotein particles, triglycerides, fatty acids, amino acids, and glycolysis-related metabolites.[25] Our findings align closely with these associations, demonstrating consistent results in terms of metabolite associations with fatty liver disease. Furthermore, our results based on the Metabolon platform are also consistent with previous studies that have identified several metabolites associated with fatty liver disease as diagnosed by ultrasound and CT-scan.[26-29] Importantly, the majority of these metabolites are implicated in various known metabolic pathways, including lipid, amino acids, carbohydrate, and peptide metabolism. This convergence of results across studies further supports the relevance of these metabolic pathways in the development and progression of fatty liver disease. As a result, my study revealed a wide range of plasma metabolites associated with NAFLD, providing valuable insights into the underlying mechanisms of the disease.

Methodological considerations

Study design

The studies in this thesis were conducted mainly in the Rotterdam Study, a prospective population-based cohort study consisting of participants aged 45 years or older.[30] Population-based cohort studies play a crucial role in epidemiological research, allowing for the study of disease incidence, etiology, and long-term health outcomes in representative populations. A notable advantage of cohort studies, in general, is the ability to study multiple exposures and multiple outcomes within a single cohort.[31] Despite the advantages, it is important to note that population-based cohort studies might subject to certain biases that should be considered (including selection bias, information bias, and confounding bias).[32] Selection bias occurs when selection of exposed and unexposed subjects is not independent of outcome being studied. Information bias occurs when there is a distortion in the measure of association due to inaccurate measurements of key study variables. While confounding bias occurs when the relationship between an exposure and an outcome is influenced by a third variable. It is therefore important for researchers to be aware of these potential biases and take steps to minimize or account for them in their study design, data collection, and analysis. Additionally, critical evaluation of study findings should always consider the potential sources of bias to assess the reliability and generalizability of the results. Nevertheless, the Rotterdam Study is a widely recognized and established cohort study which has gained a significant reputation in the field of epidemiology and medical research. With its rigorous methodology, relatively large samples size, and long-term follow-up, the Rotterdam Study has become a cornerstone in population-based research. Its robust design and comprehensive data collection make it a reliable and trusted source for investigating various health outcomes and informing evidence-based healthcare practices.

In this thesis, the majority of the studies are of cross-sectional study design, which is generally considered less informative for testing causal inference when the exposure is not an inherent trait but one that developed over time. Nonetheless, despite its limitations, cross-sectional studies can provide strong evidence for and show the strength of an association between a disease and putative causative factors. This especially refers to **Chapter 3.1** and **Chapter 4.1**, in which I detected the relationships of plasma metabolites and proteins with FLD. However, I believe that epidemiologic research work complementary in revealing possibly new relationships that can further be studied in more details by experimental or fundamental studies.

Precision of outcome

Liver biopsy is currently considered the gold standard for the diagnosis and histological assessment of FLD. In this thesis, we used CT scan, ultrasound, and liver enzymes as measurements for liver health and disease. Although liver biopsy remains the gold standard for diagnosing FLD, it is invasive and may cause adverse risks.[33] Moreover, due to ethical concerns and the unneglectable risk of complications, it will be unacceptable to expose several thousand of apparently healthy individuals to liver biopsy. Abdominal ultrasound is a rather accurate marker for the presence of fatty liver disease, especially when steatosis affects >20% of the hepatocytes. A meta-analysis demonstrated an 84.8% sensitivity to detect moderate to severe steatosis (compared to liver biopsy).[34] In the Rotterdam Study, liver fibrosis was assessed by using transient elastography. In **Chapter 2.2**, we used liver stiffness measurement (LSM) ≥9.0 kPa as a cut-off suggesting clinically relevant liver fibrosis, while previous studies preferred the threshold of 8.0 kPa in the general population to select participants at high risk of fibrosis.[35, 36] For this reason, In **Chapter 3.1** and **Chapter 4.1,** I used 8.0 kPa as a cut-off value of significant fibrosis. To come back to **Chapter 2.2** and **Chapter 4.1**, FLD was based on CT with a cut-off value of less than 40 Hounsfield Unit (HU) attenuation, which is somewhat stricter than other studies utilizing mean liver attenuation ≤ 51 HU.[37] This would decrease the prevalence of FLD, but improve the accuracy of identifying participants with FLD. Although earlier studies have shown various cut-offs, the most clinical indication for moderate-to-severe steatosis is a cut-off value of 40 HU on non-enhanced CT.[38, 39]

On the other hand, we used BMI, waist-to-hip ratio (WHR), android-fat to gynoid-fat ratio (AGR), and fat mass index (FMI) measured by anthropometrics and Dual X-ray Absorptiometry as parameters of obesity and fat distribution. BMI is the simplest and the most widely used parameter for measuring obesity.[40] It is calculated by dividing body weight in kilograms by height in meters squared (kg/m^2). BMI is widely used as a risk factor for the development of or the prevalence of several health conditions.[41] According to the World Health Organization (WHO), a BMI of greater than or equal to 25 is classified as overweight and a BMI of greater than or equal to 30 is classified as obese.[42] This is the most useful population-level measure of overweight and obesity. However, it is increasingly clear that BMI is a rather poor indicator of the percent of body fat.[43] Importantly, the BMI also does not capture information on the mass of fat in different body sites. While WHR, a measure of central obesity and visceral fat, may be a better indicator of obesity than other anthropometric measures,[44] it looks at the proportion of fat stored on your body around your waist and hip. It is found to be associated with CVD, particularly in women.[45] Furthermore, dual-energy X-ray absorptiometry (DEXA) is considered the gold standard method for measuring body composition,[46] it provides an in-depth analysis of body fat composition. It has the advantages of low cost, low radiation, and quick scan times, making it a quick and practical method for assessing body composition in large longitudinal studies.[47] FMI is defined as the total DEXA fat mass normalized by height squared (Fat mass/ $Hight^2$), it has a distinct advantage over BMI for defining obesity status since it is independent of lean mass status,[48] while the AGR is the ratio between abdominal, or

android fat pattern, and lower limb fat percentage, or gynoid fat pattern.[49] Moreover, previous studies have shown important relationships between AGR with metabolic and CVD risks in healthy adults.[50, 51] DEXA measurements of fat mass can play a vital role in predicting cardiometabolic diseases including CVD and T2D. In this line, in **Chapter 2.1**, we found some of the identified miRNAs with DEXA parameters (FMI and AGR) were also associated with cardiovascular and metabolic diseases reported in previous studies.[9, 14, 15]

Precision of exposure

MicroRNA profiling

In recent years, the link between regulatory miRNAs and diseases has been the object of intensive research. Intriguingly, it has recently been discovered that a number of miRNAs regulate the development and metabolism of adipose tissue, as well as insulin secretion and action. As a result, an imbalance in these miRNAs may contribute to the development of obesity and its associated metabolic complications.[52-54] Previous genetic association studies have demonstrated the association of SNPs in miRNAs with obesity and fat distribution.[55, 56] Furthermore, miRNA profiling studies based on qPCR assay has investigated the link between miRNA levels and metabolic diseases but with a limited number of miRNAs and a small sample size. My research in **Chapter 2.** was conducted in the Rotterdam study with much larger sample sizes, based on a new genome-wide miRNA-seq based assay (HTG EdgeSeq miRNA whole Transcriptome Assay), which measured the expression of 2083 mature human miRNAs and demonstrated higher accuracy, sensitivity, and specificity compared to qPCR method. Although the use of the small RNA sequencing method in epidemiological studies is still limited, it has been shown to measure the expression levels of miRNAs over a wide dynamic range with the ability to identify novel miRNAs with the least bias detection.[57, 58] Despite the absence of defined guidelines for the use of miRNAs in current clinical practice, there is promising evidence that they represent a reliable tool for use in the future as a biomarker.[59] The intriguing aspect of miRNAs in biomarker discovery is their stability in human fluids, even after multiple freeze-thaw cycles.[60]

Proteome

The proteome analysis offers a compelling goal for the identification of therapeutic and biomarker targets for complex diseases. The easiest tissue to collect and the one with the most comprehensive human proteome is blood. Interestingly, several studies have identified protein expression in blood by using sophisticated techniques (e.g. Olink panels or SomaScan assay) to discover potential biomarkers in a number of metabolic and inflammatory diseases including FLD,[61-63] using disease stratification with small sample sizes. In **Chapter 3.1**, the research was embedded within the Rotterdam Study with a large sample size, based on the high-throughput Proximity Extension Assay (PEA) technique (Olink Proteomics AB, Uppsala, Sweden). The Olink technology combines the best of antibody- and DNA-based methodologies

to provide unique, enabling tools for protein biomarker discovery and development.[64] The Olink platforms offer several key advantages that set them apart from other techniques like SomaScan. Firstly, they demonstrate high sensitivity and specificity, enabling accurate measurement of low-abundance proteins. This is particularly valuable for detecting subtle changes in protein expression levels. Additionally, Olink platforms require small sample volumes, making them well-suited for working with precious or limited samples, in contrast to other techniques. Furthermore, it provides a diverse range of panels designed to analyze protein biomarkers specific to various research areas and disease conditions, including cardiovascular disease, oncology, neurology, and immunology. This targeted approach enhances the relevance and applicability of the platform in different fields. Overall, Olink platforms serve as valuable tools for biomarker discovery, validation, and research, offering a comprehensive and multiplexed approach to protein analysis. In our study, the utilized assay reliably measured plasma levels of 171 cardiometabolic and inflammatory-related proteins after excluding proteins that were below the limit of Detection during the quality control procedures.

Metabolomics

Metabolomics has been widely used for investigating the biological functions of disease expression and has the potential to discover biomarkers in circulating biofluids or tissue extracts that reflect in phenotypic changes.[65] Mass spectrometry (MS) and nuclear magnetic resonance (NMR) are the two main analytical technologies for the analysis of metabolite pools, the former in combination with separation techniques such as gas and liquid chromatography, and has contributed to the discovery of recent disease biomarkers. These methods enable the generation of large amounts of data, and sophisticated chemometric analysis provides details on the wide range of metabolites.[66] Both targeted and untargeted mass spectrometry approaches are among the most commonly used in metabolome analysis and it has become the indispensable tool in metabolome analysis.[67] Moreover, MS is intrinsically a highly sensitive method for the detection, quantitation, and structure elucidation of upwards of several hundred metabolites in a single measurement.[68] Likewise, NMR plays a valuable tool that can be used for quantitative fingerprinting to identify different metabolites within tissue engineering scaffolds.[69] Moreover, NMR can also be exploited for targeted and untargeted human metabolic phenotype diversity. Proton (1H)-NMR-based metabolomics profiling first became a popular technology because it could be used to identify organic compounds within biological fluids.[70] In addition, MS and NMR methods are both supplementary and complementary to one another. In **Chapter 4.1**, I used both the Nightingale and Metabolon platforms for linking metabolites to FLD and liver enzymes. Nightingale platform is valued for being non-invasive, non-destructive, fast, and for providing highly reproducible results.[71] As well, the Metabolon platform provides a snapshot of the metabolic state of the entire organism as well as individual tissues, by using technologies able to detect a wide number of metabolites.[72] So, the results of our study using these two platforms able to generate a more comprehensive understanding of the metabolic profiles linked to FLD

and liver enzyme activity. This combined approach shed light on the underlying metabolic pathways involved in the development and progression of the disease, leading to valuable insights in the field.

Potential implications and future directions

In this section, I describe the future directions of using population-level omics data in obesity and FLD, which may open new avenues for understanding the risk and pathophysiology of these diseases and also potential biomarkers. As explained in this thesis, it has been established that genetic predispositions play an essential role in the unbalanced energy metabolism associated with obesity. Emerging evidence suggests that epigenetic modifications including miRNAs have emerged as key mediators of metabolic processes, playing crucial roles in maintaining/altering physiological processes, including energy balance and metabolic homeostasis.[54] Dysregulation of miRNAs may affect the status and functions of different tissues and organs, possibly contributing to metabolic abnormalities associated with obesity and obesity-related diseases. More recently, the discovery of circulating miRNAs easily detectable in plasma and other body fluids has emphasized their potential as both endocrine signaling molecules and disease indicators. In **Chapter 2.1**, despite having a comprehensive global profile of 591 well-expressed miRNAs obtained through RNA sequencing-based assay, studying the relationship between miRNAs and obesity-related traits presents several challenges. Firstly, the complex nature of obesity, influenced by a combination of genetic, environmental, and lifestyle factors, introduces difficulties in discerning the specific role of miRNAs in obesity and its associated traits. Moreover, miRNAs exhibit tissue-specific expression patterns, making it intricate to determine their significance in specific tissues implicated in obesity, such as adipose tissue or liver. To capture the dynamic changes in miRNA expression associated with obesity and its progression, longitudinal studies with repeated measurements are required. It is crucial to have well-characterized and diverse study populations encompassing individuals from different ethnicities and age groups to understand the generalizability of miRNA-obesity associations. Additionally, conducting long-term follow-up studies is necessary to explore the causality and temporal relationships between miRNAs and obesity-related traits, as well as to assess their potential as therapeutic targets or biomarkers. Overall, addressing these challenges is crucial for advancing our understanding of the involvement of miRNAs in obesity-related traits and their potential as therapeutic targets or biomarkers.

Among the identified miRNAs, miR-193a-5p was the top miRNAs associated with obesity-related traits in **Chapter 2.1** and also associated with FLD in **Chapter 2.2**. Currently, one study based on biopsies that confirmed NAFLD and assessed 2083 serum miRNAs in a discovery cohort indicated that serum miR-193a-5p levels substantially related with NAFLD activity grade and fibrosis stage. Additionally, miR-193a-5p may contribute to the liver's response to oxidative stress and may be a therapeutically useful circulating biomarker for progressive NAFLD.[73] Therefore, further research is warranted on the role of miR-193a-5p throughout the body, particularly to establish its direct effects on certain target genes and to

identify the cells and potential extrahepatic tissues where these interactions and subsequent consequences might be seen.

Considering the strong association between FLD and obesity, adiposity, as determined by obesity parameters such as waist circumference and waist-to-hip ratio, has been identified as the most influential predictor of NAFLD risk on a global scale.[74-76] Consequently, further investigation is warranted to better understand the underlying mechanisms linking adiposity and NAFLD, as well as to explore potential interventions targeting this relationship. Extensive research has been conducted on the mechanisms underlying the close relationship between obesity and NAFLD, including chronic inflammation, oxidative stress, and insulin resistance.[74, 75, 77] However, the specific mechanisms by which metabolic imbalances resulting from obesity contribute to FLD development remain unclear. Therefore, future studies should focus on investigating the role of epigenetic factors, such as miRNAs, in metabolic dysfunction and their connection to NAFLD/MAFLD. It is still unclear why only a certain population with metabolic dysfunction will develop FLD while others with similar metabolic dysfunctions do not. This aspect deserves particular attention, with a focus on identifying factors that might induce the chronic inflammatory state, which is a key determinant of advanced fibrosis.

In **Chapter 2.3**, we explored the potential impact of lifestyle factors, specifically alcohol consumption, on the alteration of plasma miRNA concentrations and their association with liver disease. However, we encountered some challenges throughout this study. Firstly, as mentioned previously, it is important to note that the majority of our research followed a cross-sectional design, which inherently introduces the limitation of potential reverse causality. Additionally, the application of Mendelian randomization (MR) presents a promising approach to establish a causal connection between lifestyle factors, like alcohol consumption, and the risk of developing illnesses through epigenetic markers. Nevertheless, unlike chronic diseases, lifestyle factors have a limited genetic component, which resulted in the identification of inadequate instrumental variables (IVs) in many genome-wide association studies (GWAS) investigating lifestyle factors.[78] Consequently, the utilization of the MR methodology to elucidate the causal pathway between lifestyles and epigenetics can present challenges. Furthermore, we collected data on alcohol consumption through home-administered interviews rather than employing validated self-reporting methods like food frequency questionnaires (FFQs) or other reliable measures. This introduces the possibility of participants underestimating their true alcohol consumption due to social desirability bias. To address this limitation, Further studies are needed by using larger sample sizes and longer follow-up times with alcohol consumption collected by FFQs as well as in to experimentally confirm the role of identified miRNAs in molecular pathways underlying alcohol-related diseases. Finally, tissue specificity poses another challenge when translating our findings into clinical practice. The majority of available data are derived from easily accessible tissues such as blood or plasma, but conducting miRNA association studies on different tissues may yield divergent results. It is therefore crucial to strive for a more comprehensive collection of epigenetic data across various tissues.

Proteomics is a powerful tool for studying changes in the protein expression levels in diverse patient populations, providing a robust and impactful tool for understanding disease mechanisms. In particular, the liver, as a key secretory organ in the human body, is responsible for synthesizing and producing the majority of plasma proteins that may play a direct role in circulation.[20] Among these proteins, Albumin is the most abundant protein in the blood and accounts to approximately 60% of all plasma proteins.[79] In **Chapter 3.1**, I performed a cross-sectional study to assess the protein expression in hepatic steatosis and fibrosis and the possibility to use it as a biomarker for disease diagnosis. The gap between protein measurement and FLD diagnosed by ultrasounds, which makes conclusions regarding dynamic changes of fatty liver difficult would result in ‘dilution towards the null’. Recently, several small studies define protein levels in patient cohorts with advanced FLD by using sophisticated techniques (e.g. SomaScan assay or Olink).[61, 63] Variability in research design, sample selection, storage and processing, instrumental analysis, data analysis, and most critically, might result in inconsistent and conflicting results. Future large-scale proteomics studies of FLD would benefit from a cross-sectional design measuring proteins in blood and FLD diagnosed by ultrasound collected at the same time point. In addition, a follow-up study design of the same people at multiple time points.

Finally, metabolomics offers new opportunities for biomarker discovery in complex diseases and may provide a pathological understanding of diseases beyond traditional technologies.[80] Several studies have assessed the associations of metabolomics with NAFLD or liver fat content.[81] These studies have identified several pathways underlying the development of NAFLD. However, the majority of these studies have variability in study design, involved a small number of cases, and measured a limited set of metabolomic biomarkers.[24, 82] In **Chapter 4.1**. I used comprehensive metabolomics profiling with two well-established platforms (Nightingale and Metabolon) measured in several population-based studies. In addition, different modalities were used to evaluate FLD (CT scan and ultrasound). Currently, few studies have systematically detected metabolites associated with FLD at the population level. Therefore, my findings in this chapter may help for elucidating metabolic pathways involved in NAFLD etiology and the identified metabolites might be considered as potential biomarkers for early diagnosis of the disease. Though future research is warranted to include more population-based cohorts from different ethnicities and age groups to assess better metabolites profiling in FLD.

Conclusions

In this thesis, I conducted comprehensive analyses of different omics data (incl. transcriptomics, proteomics, and metabolomics) from population-based studies to elucidate the underlying molecular mechanisms of obesity and fatty liver disease and to identify potential biomarkers for the early diagnosis of these diseases. Particularly, the results of my studies may contribute to better understanding of the pathogenesis of FLD, yet much more clinical and experimental research need to be conducted in order to effectively prevent and treat FLD. FLD is a growing concern with increasing awareness, and I believe that significant progress will be made in unraveling its pathophysiology in the upcoming decade, through advancements in research techniques, collaborative efforts, access to large-scale datasets, and the pursuit of innovative therapeutic strategies.

References

1. Drozdz D, Alvarez-Pitti J, Wójcik M, Borghi C, Gabbianelli R, Mazur A, et al. Obesity and cardiometabolic risk factors: From childhood to adulthood. Nutrients. 2021;13(11):4176.
2. Buechler C, Wanninger J, Neumeier M. Adiponectin, a key adipokine in obesity related liver diseases. World journal of gastroenterology: WJG. 2011;17(23):2801.
3. Quek J, Chan KE, Wong ZY, Tan C, Tan B, Lim WH, et al. Global prevalence of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in the overweight and obese population: a systematic review and meta-analysis. The Lancet Gastroenterology & Hepatology. 2022.
4. Le MH, Yeo YH, Li X, Li J, Zou B, Wu Y, et al. 2019 global NAFLD prevalence: a systematic review and meta-analysis. Clinical Gastroenterology and Hepatology. 2022;20(12):2809-17. e28.
5. Lim GEH, Tang A, Ng CH, Chin YH, Lim WH, Tan DJH, et al. An observational data meta-analysis on the differences in prevalence and risk factors between MAFLD vs NAFLD. Clinical Gastroenterology and Hepatology. 2023;21(3):619-29. e7.
6. Byrne CD, Targher G. NAFLD: a multisystem disease. Journal of hepatology. 2015;62(1):S47-S64.
7. Sadashiv, Modi A, Khokhar M, Sharma P, Joshi R, Mishra SS, et al. Leptin DNA Methylation and Its Association with Metabolic Risk Factors in a Northwest Indian Obese Population. J Obes Metab Syndr. 2021;30(3):304-11.
8. Hyun J, Jung Y. DNA methylation in nonalcoholic fatty liver disease. International journal of molecular sciences. 2020;21(21):8138.
9. Mens MMJ, Mustafa R, Ahmadizar F, Ikram MA, Evangelou M, Kavousi M, et al. MiR-139-5p is a causal biomarker for type 2 diabetes; Results from genome-wide microRNA profiling and Mendelian randomization analysis in a population-based study. medRxiv. 2021.
10. Zhang X, Mens MMJ, Abozaid YJ, Bos D, Darwish Murad S, de Knegt RJ, et al. Circulatory microRNAs as potential biomarkers for fatty liver disease: the Rotterdam study. Alimentary Pharmacology & Therapeutics. 2021;53(3):432-42.
11. Tao Z, Zheng S, He X, Sun J, He C, Zhang L. Hsa_circ_0037897 may be a risk factor for essential hypertension via hsa-miR-145-5p. Clinical and Experimental Hypertension. 2021;43(3):281-6.
12. Zhou Q, Li D, Zheng H, He Z, Qian F, Wu X, et al. A novel lncRNA-miRNA-mRNA competing endogenous RNA regulatory network in lung adenocarcinoma and kidney renal papillary cell carcinoma. Thoracic cancer. 2021;12(19):2526-36.
13. Sonoda T, Matsuzaki J, Yamamoto Y, Sakurai T, Aoki Y, Takizawa S, et al. Serum microRNA-based risk prediction for stroke. Stroke. 2019;50(6):1510-8.
14. Zhang H, Hao J, Sun X, Zhang Y, Wei Q. Circulating pro-angiogenic micro-ribonucleic acid in patients with coronary heart disease. Interactive Cardiovascular and Thoracic Surgery. 2018;27(3):336-42.
15. Chen L, Bai J, Liu J, Lu H, Zheng K. A four-microRNA panel in peripheral blood identified as an early biomarker to diagnose acute myocardial infarction. Frontiers in Physiology. 2021;12.
16. Lancet T. The link between cancer and obesity. 2017. p. 1716.
17. De Pergola G, Silvestris F. Obesity as a major risk factor for cancer. Journal of obesity. 2013;2013.
18. Ajoolabady A, Aslkhodapasandhokmabad H, Zhou Y, Ren J. Epigenetic modification in alcohol-related liver diseases. Medicinal Research Reviews. 2022.
19. Xu T, Li L, Hu HQ, Meng XM, Huang C, Zhang L, et al. MicroRNAs in alcoholic liver disease: Recent advances and future applications. Journal of cellular physiology. 2019;234(1):382-94.
20. Niu L, Geyer PE, Wewer Albrechtsen NJ, Gluud LL, Santos A, Doll S, et al. Plasma proteome profiling discovers novel proteins associated with non-alcoholic fatty liver disease. Molecular systems biology. 2019;15(3):e8793.
21. Mas VR, Maluf DG, Archer KJ, Yanek K, Bornstein K, Fisher RA. Proteomic analysis of HCV cirrhosis and HCV-induced HCC: identifying biomarkers for monitoring HCV-cirrhotic patients awaiting liver transplantation. Transplantation. 2009;87(1):143.
22. Chaerkady R, Harsha HC, Nalli A, Gucek M, Vivekanandan P, Akhtar J, et al. A quantitative proteomic approach for identification of potential biomarkers in hepatocellular carcinoma. Journal of proteome research. 2008;7(10):4289-98.

23. Pang Y, Kartsonaki C, Lv J, Millwood IY, Fairhurst-Hunter Z, Turnbull I, et al. Adiposity, metabolomic biomarkers, and risk of nonalcoholic fatty liver disease: a case-cohort study. The American Journal of Clinical Nutrition. 2022;115(3):799-810.
24. Kalhan SC, Guo L, Edmison J, Dasarathy S, McCullough AJ, Hanson RW, et al. Plasma metabolomic profile in nonalcoholic fatty liver disease. Metabolism. 2011;60(3):404-13.
25. Kaikkonen JE, Würtz P, Suomela E, Lehtovirta M, Kangas AJ, Jula A, et al. 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):491-500.
26. Lind L, Salihovic S, Risérus U, Kullberg J, Johansson L, Ahlström H, et al. The plasma metabolomic profile is differently associated with liver fat, visceral adipose tissue, and pancreatic fat. The Journal of Clinical Endocrinology & Metabolism. 2021;106(1):e118-e29.
27. Lovric A, Granér M, Bjornson E, Arif M, Benfeitas R, Nyman K, et al. Characterization of different fat depots in NAFLD using inflammation-associated proteome, lipidome and metabolome. Scientific reports. 2018;8(1):1-14.
28. Lo EKK, Xu J-H, Zhan Q, Zeng Z, El-Nezami H. The emerging role of branched-chain amino acids in liver diseases. Biomedicines. 2022;10(6):1444.
29. Palmer ND, Okut H, Hsu F-C, Ng MCY, Chen Y-DI, Goodarzi MO, et al. Metabolomics identifies distinctive metabolite signatures for measures of glucose homeostasis: the Insulin Resistance Atherosclerosis Family Study (IRAS-FS). The Journal of Clinical Endocrinology & Metabolism. 2018;103(5):1877-88.
30. Ikram MA, Brusselle G, Ghanbari M, Goedegebure A, Ikram MK, Kavousi M, et al. Objectives, design and main findings until 2020 from the Rotterdam Study. European journal of epidemiology. 2020;35(5):483-517.
31. Euser AM, Zoccali C, Jager KJ, Dekker FW. Cohort studies: prospective versus retrospective. Nephron Clinical Practice. 2009;113(3):c214-c7.
32. Kesmodel US. Information bias in epidemiological studies with a special focus on obstetrics and gynecology. Acta Obstetricia et Gynecologica Scandinavica. 2018;97(4):417-23.
33. Musaddaq G, Shahzad N, Ashraf MA, Arshad MI. Circulating liver-specific microRNAs as noninvasive diagnostic biomarkers of hepatic diseases in human. Biomarkers. 2019;24(2):103-9.
34. Hernaez R, Lazo M, Bonekamp S, Kamel I, Brancati FL, Guallar E, et al. Diagnostic accuracy and reliability of ultrasonography for the detection of fatty liver: a meta-analysis. Hepatology. 2011;54(3):1082-90.
35. Roulot D, Costes JL, Buyck JF, Warzocha U, Gambier N, Czernichow S, et al. Transient elastography as a screening tool for liver fibrosis and cirrhosis in a community-based population aged over 45 years. Gut. 2011;60(7):977-84.
36. van Kleef LA, Ayada I, Alferink LJM, Pan Q, de Knegt RJ. Metabolic dysfunction–associated fatty liver disease improves detection of high liver stiffness: The Rotterdam Study. Hepatology. 2022;75(2):419-29.
37. Zeb I, Li D, Nasir K, Katz R, Larijani VN, Budoff MJ. Computed tomography scans in the evaluation of fatty liver disease in a population based study: the multi-ethnic study of atherosclerosis. Academic radiology. 2012;19(7):811-8.
38. Kodama Y, Ng C, Wu T-T, Ayers G, Curley S, Abdalla E, et al. Comparison of CT methods for determining the fat content of the liver. 2007.
39. Park SH, Kim PN, Kim KW, Lee SW, Yoon SE, Park SW, et al. Macrovesicular hepatic steatosis in living liver donors: use of CT for quantitative and qualitative assessment. Radiology. 2006;239(1):105-12.
40. Mahadevan S, Ali I. Is body mass index a good indicator of obesity? : Springer; 2016. p. 140-2.
41. Kivimäki M, Strandberg T, Pentti J, Nyberg ST, Frank P, Jokela M, et al. Body-mass index and risk of obesity-related complex multimorbidity: an observational multicohort study. The lancet Diabetes & endocrinology. 2022;10(4):253-63.
42. World Health O. Obesity: preventing and managing the global epidemic. 2000.
43. Nuttall FQ. Body mass index: obesity, BMI, and health: a critical review. Nutrition today. 2015;50(3):117.
44. Anand SS, Islam S, Rosengren A, Franzosi MG, Steyn K, Yusufali AH, et al. Risk factors for myocardial infarction in women and men: insights from the INTERHEART study. European heart journal. 2008;29(7):932-40.

45. Cao Q, Yu S, Xiong W, Li Y, Li H, Li J, et al. Waist-hip ratio as a predictor of myocardial infarction risk: A systematic review and meta-analysis. Medicine. 2018;97(30).
46. Hussain Z, Jafar T, uz Zaman M, Parveen R, Saeed F. Correlations of skin fold thickness and validation of prediction equations using DEXA as the gold standard for estimation of body fat composition in Pakistani children. BMJ open. 2014;4(4):e004194.
47. Chen W, Wilson JL, Khaksari M, Cowley MA, Enriori PJ. Abdominal fat analyzed by DEXA scan reflects visceral body fat and improves the phenotype description and the assessment of metabolic risk in mice. American Journal of Physiology-Endocrinology and Metabolism. 2012.
48. Kelly TL, Wilson KE, Heymsfield SB. Dual energy X-Ray absorptiometry body composition reference values from NHANES. PloS one. 2009;4(9):e7038.
49. Okosun IS, Seale JP, Lyn R. Commingling effect of gynoid and android fat patterns on cardiometabolic dysregulation in normal weight American adults. Nutrition & Diabetes. 2015;5(5):e155-e.
50. Walton C, Lees B, Crook D, Worthington M, Godsland IF, Stevenson JC. Body fat distribution, rather than overall adiposity, influences serum lipids and lipoproteins in healthy men independently of age. The American journal of medicine. 1995;99(5):459-64.
51. Kang SM, Yoon JW, Ahn HY, Kim SY, Lee KH, Shin H, et al. Android fat depot is more closely associated with metabolic syndrome than abdominal visceral fat in elderly people. PloS one. 2011;6(11):e27694.
52. Hossain MM, Sohel MMH, Schellander K, Tesfaye D. Characterization and importance of microRNAs in mammalian gonadal functions. Cell and tissue research. 2012;349:679-90.
53. Pallante P, Battista S, Pierantoni GM, Fusco A. Deregulation of microRNA expression in thyroid neoplasias. Nature Reviews Endocrinology. 2014;10(2):88-101.
54. Iacomino G, Siani A. Role of microRNAs in obesity and obesity-related diseases. Genes & nutrition. 2017;12:1-16.
55. Hilton C, Neville MJ, Wittemans LBL, Todorcevic M, Pinnick KE, Pulit SL, et al. MicroRNA-196a links human body fat distribution to adipose tissue extracellular matrix composition. EBioMedicine. 2019;44:467-75.
56. Ghanbari M, Sedaghat S, De Looper HWJ, Hofman A, Erkeland SJ, Franco OH, et al. The association of common polymorphisms in mi R-196a2 with waist to hip ratio and mi R-1908 with serum lipid and glucose. Obesity. 2015;23(2):495-503.
57. Eminaga S, Christodoulou DC, Vigneault F, Church GM, Seidman JG. Quantification of microRNA expression with next-generation sequencing. Current protocols in molecular biology. 2013;103(1):4.17. 1-4.. 4.
58. Godoy PM, Barczak AJ, DeHoff P, Srinivasan S, Etheridge A, Galas D, et al. Comparison of Reproducibility, Accuracy, Sensitivity, and Specificity of miRNA Quantification Platforms. Cell Rep. 2019;29(12):4212-22 e5.
59. Grandjean P. Biomarkers in epidemiology. Clin Chem. 1995;41(12 Pt 2):1800-3.
60. Glinge C, Clauss S, Boddum K, Jabbari R, Jabbari J, Risgaard B, et al. Stability of circulating blood-based microRNAs–pre-analytic methodological considerations. PloS one. 2017;12(2):e0167969.
61. Luo Y, Wadhawan S, Greenfield A, Decato B, Oseini A, Collen R, et al. SOMAscan Proteomics Identifies Serum Biomarkers Associated With Liver Fibrosis in Patients With NASH. Hepatology Communications. 2021;5.
62. Hanff TC, Cohen JB, Zhao L, Javaheri A, Zamani P, Prenner SB, et al. Quantitative proteomic analysis of diabetes mellitus in heart failure with preserved ejection fraction. Basic to Translational Science. 2021;6(2):89-99.
63. Corey KE, Pitts R, Lai M, Loureiro J, Masia R, Osganian SA, et al. ADAMTSL2 protein and a soluble biomarker signature identify at-risk non-alcoholic steatohepatitis and fibrosis in adults with NAFLD. Journal of Hepatology. 2022;76(1):25-33.
64. Assarsson E, Lundberg M, Holmquist G, Björkesten J, Bucht Thorsen S, Ekman D, et al. Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. PloS one. 2014;9(4):e95192.
65. Saigusa D, Matsukawa N, Hishinuma E, Koshiba S. Identification of biomarkers to diagnose diseases and find adverse drug reactions by metabolomics. Drug metabolism and pharmacokinetics. 2021;37:100373.

66. Tyagi R, Kumar P, Sharma U. Metabolomics techniques: A brief update. Epigenetics and Metabolomics: Elsevier; 2021. p. 1-29.
67. Defossez E, Bourquin J, von Reuss S, Rasmann S, Glauser G. Eight key rules for successful data-dependent acquisition in mass spectrometry-based metabolomics. Mass Spectrometry Reviews. 2023;42(1):131-43.
68. Gowda GAN, Djukovic D. Overview of mass spectrometry-based metabolomics: opportunities and challenges. Mass Spectrometry in Metabolomics: Methods and Protocols. 2014:3-12.
69. Raja G, Jung Y, Jung SH, Kim T-J. 1H-NMR-based metabolomics for cancer targeting and metabolic engineering–A review. Process Biochemistry. 2020;99:112-22.
70. Emwas A-H, Roy R, McKay RT, Tenori L, Saccenti E, Gowda GAN, et al. NMR spectroscopy for metabolomics research. Metabolites. 2019;9(7):123.
71. Pan Z, Raftery D. Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics. Analytical and bioanalytical chemistry. 2007;387:525-7.
72. Smirnov KS, Maier TV, Walker A, Heinzmann SS, Forcisi S, Martinez I, et al. Challenges of metabolomics in human gut microbiota research. International Journal of Medical Microbiology. 2016;306(5):266-79.
73. Johnson K, Leary PJ, Govaere O, Barter MJ, Charlton SH, Cockell SJ, et al. Increased serum miR-193a-5p during non-alcoholic fatty liver disease progression: Diagnostic and mechanistic relevance. JHEP Reports. 2022;4(2):100409.
74. Zhou F, Zhou J, Wang W, Zhang XJ, Ji YX, Zhang P, et al. Unexpected rapid increase in the burden of NAFLD in China from 2008 to 2018: a systematic review and meta-analysis. Hepatology. 2019;70(4):1119-33.
75. Li L, Liu DW, Yan HY, Wang ZY, Zhao SH, Wang B. Obesity is an independent risk factor for non-alcoholic fatty liver disease: evidence from a meta-analysis of 21 cohort studies. Obesity reviews. 2016;17(6):510-9.
76. Yki-Järvinen H. Non-alcoholic fatty liver disease as a cause and a consequence of metabolic syndrome. The lancet Diabetes & endocrinology. 2014;2(11):901-10.
77. Milić S, Lulić D, Štimac D. Non-alcoholic fatty liver disease and obesity: biochemical, metabolic and clinical presentations. World journal of gastroenterology: WJG. 2014;20(28):9330.
78. Cornelis MC, Munafo MR. Mendelian randomization studies of coffee and caffeine consumption. Nutrients. 2018;10(10):1343.
79. 5 Human Albumin. Transfus Med Hemother. 2009;36(6):399-407.
80. Zhang A, Sun H, Yan G, Wang P, Wang X. Metabolomics for biomarker discovery: moving to the clinic. Biomed research international. 2015;2015.
81. Vilar-Gomez E, Chalasani N. Non-invasive assessment of non-alcoholic fatty liver disease: Clinical prediction rules and blood-based biomarkers. Journal of hepatology. 2018;68(2):305-15.
82. Alferink LJM, Radjabzadeh D, Erler NS, Vojinovic D, Medina-Gomez C, Uitterlinden AG, et al. Microbiomics, Metabolomics, Predicted Metagenomics, and Hepatic Steatosis in a Population-Based Study of 1,355 Adults. Hepatology. 2021;73(3):968-82.

Chapter 6

Summary/ Samenvatting

English summary

Fatty liver disease (FLD) and its associated complications, such as type 2 diabetes (T2D), obesity, hypertension, and dyslipidemia, impose a significant burden on global health, leading to illness and mortality. Over the past few decades, the prevalence of FLD has been on the rise, affecting more than 33% of the world's population, largely due to the increasing rates of obesity. Among individuals with FLD, the majority develop non-alcoholic fatty liver disease (NAFLD), which stands as the most prevalent form of chronic liver disease worldwide. Although NAFLD often presents with little or no symptoms, a subset of patients can progress to advanced stages of liver disease, including end-stage liver disease and even liver cancer. Recently, an international panel of experts recommended a name change for NAFLD to metabolic dysfunction-associated fatty liver disease (MAFLD) to reflect its broader metabolic implications beyond alcohol consumption. Furthermore, in recent development, the global community chose steatotic liver disease (SLD) as an overarching term to encompass the various etiologies of steatosis. However, the underlying mechanisms of obesity-related diseases and FLD and early diagnostic strategies remain poorly understood. Gaining a comprehensive understanding of the pathophysiology of FLD is crucial for advancing prevention and treatment approaches. In this thesis, I aim to provide new insights in to the molecular mechanisms of obesity-related traits and FLD by employing various omics approaches and integrating multiple types of omics data, with a particular focus on transcriptomics, proteomics, and metabolomics.

Chapter 2, focused on exploring the epigenetic regulation of obesity-related traits and FLD. In **Chapter 2.1**, I conducted a genome-wide screening using data from the Rotterdam Study to explore the relationship between plasma levels of miRNAs and obesity, body fat distribution, and fat mass. Notably, my findings revealed the association of several miRNAs with obesity-related traits, with 12 miRNAs overlapping across all traits, including BMI, WHR, FMI, and AGR. It is worth mentioning that among these miRNAs, miR-193a-5p, the most prominent miRNA associated with obesity, was also found to be significantly associated with FLD in Chapter 2.2.

In **Chapter 2.2**, I investigated the association between miRNA plasma levels and the prevalence of fatty liver disease. I examined the relationships between miRNAs and both fatty liver disease and liver enzymes, specifically gamma-glutamyl transferase and alkaline phosphatase. The results of my analysis revealed that 61 miRNAs exhibited associations with liver enzymes. Additionally, 17 miRNAs were found to be associated with fatty liver disease, out of which 14 miRNAs overlapped with the ones associated with liver enzymes. Notably, one of the identified miRNAs was miR-122-5p, a well-established miRNA in liver tissue.

In **Chapter 2.3**, I explored the relationship between plasma miRNA expression levels and alcohol consumption. I found that miR-193p, miR-122-5p, and miR-3937 were positively associated with alcohol

consumption, while miR-4507 showed a negative association. When comparing different alcohol consumption categories (non-drinkers, light drinkers, and heavy drinkers), the mean expression levels of heavy drinkers were almost double those of light drinkers, with non-drinkers as the reference category.

Furthermore, some of the miRNAs associated with alcohol consumption were found to target genes that have been previously linked to alcohol consumption through GWAS, EWAS, TWAS, or reviews on lifestyle factors and miRNAs. Pathway analysis of these target genes revealed their involvement in biological regulation and cellular processes. Additionally, miR-3937 and miR-122-5p, among the identified miRNAs, showed potential mediatory roles in liver-related traits.

In **Chapter 3**, I explored the association between plasma proteomic and FLD and liver fibrosis. In **Chapter 3.1**, I found 27 inflammatory and cardiometabolic plasma proteins were associated with FLD, the strongest association was observed for FGF-21 and CES1 proteins. Importantly, 15 of the 27 proteins significantly associated with FLD were also associated with liver fibrosis. Through pathway analysis, I identified the involvement of TWEAK and TNFS14 proteins in the non-canonical NF-kB pathway, while F7, SERPINA5, and PROC proteins were found to play roles in the clotting cascade, specifically in the formation of fibrin clots. Interestingly, the expression levels of IL-18R1 and CES1 were upregulated in the organoids mimicking FLD. Finally, the AUC analysis illustrated that proteomics has the potential to improve currently available non-invasive markers (FLI and FIB-4) in the prediction of FLD or fibrosis.

Finally, in **Chapter 4**, I explored the association of circulating metabolites with NAFLD and liver enzymes using data from multiple population-based cohort studies. In **Chapter 4.1**, I found significant associations between several metabolites from both platforms and NAFLD, as well as liver enzymes. Notably, there were 21 metabolites that exhibited joint associations with both NAFLD and liver enzymes, including phenylalanine, triglycerides in HDL, IDL, and small LDL, and various ratios of fatty acids such as linoleic acid to total fatty acid, omega 6 fatty acid to total fatty acid, and polyunsaturated fatty acid to total fatty acid within the Nightingale platform. Within the Metabolon platform, glutamate and sphingomyelin (d18:0/18:0, d19:0/17:0) displayed the strongest associations with both NAFLD and liver enzymes. Additionally, the remaining metabolites that significantly correlated with the presence of NAFLD were involved in lipid, amino acid, and peptide metabolism.

Nederlandse samenvatting

Leververvetting (FLD, fatty liver disease) en daaraan gerateerde metabole ziekten, zoals diabetes type 2, obesitas, hypertensie en dyslipidemie, vormen een aanzienlijke belasting voor de wereldwijde gezondheid, wat leidt tot ziekte en sterfte. In de afgelopen decennia is de prevalentie van FLD toegenomen en treft het meer dan 33% van de wereldbevolking. Dit gaat parallel aan de wereldwijde obesitas epidemie. Van de mensen met FLD ontwikkelt de meerderheid niet-alcoholische leververvetting (NAFLD), de meest voorkomende vorm van chronische leverziekte is. Hoewel NAFLD vaak weinig of geen symptomen vertoont, kan een subgroep van patiënten zich ontwikkelen tot gevorderde stadia van leverziekte, wat kan leiden tot eindstadium leverfalen en zelfs leverkanker. De terminologie van leververvetting staat onder discussie. Onlangs adviseerde een internationaal panel van experts een naamswijziging door te voeren van NAFLD naar metabole disfunctie-geassocieerde leververvetting (MAFLD) om de bredere metabole implicaties weer te geven zonder het woord alcohol te gebruiken. Nadien koos de wereldwijde gemeenschap in recente ontwikkelingen de naam steatotische leverziekte (SLD) als een overkoepelende term om de verschillende etiologieën van steatose te omvatten. De onderliggende pathofysiologie die ten grondslag ligt aan obesitas en daaraan gerelateerde ziekten (zoals FLD) en vroege diagnostische strategieën blijven echter slecht begrepen. Het tot in de details begrijpen van de pathofysiologie van FLD is cruciaal voor het bevorderen van preventie- en behandelingsbenaderingen. In dit proefschrift probeer ik nieuwe inzichten te verschaffen in de moleculaire mechanismen van aan obesitas gerelateerde ziekten en FLD door verschillende omics-benaderingen toe te passen en meerdere soorten omics-gegevens te integreren. In het bijzonder focus ik me op transcriptomics, proteomics en metabolomics.

Hoofdstuk 2, gericht op het verkennen van de epigenetische regulatie van aan obesitas gerelateerde fenotypen en FLD. In **Hoofdstuk 2.1** heb ik een genoombrede screening uitgevoerd met behulp van gegevens van de Rotterdam Study om de relatie tussen miRNA's plasmaspiegels en obesitas, lichaamsvetverdeling en vetmassa te onderzoeken. Verschillende miRNA’s bleken geassocieerd te zijn met obesitas gerelateerde kenmerken waarvan 12 miRNA’s overlappend zijn met alle waaronder BMI, WHR, FMI en AGR. Het is vermeldenswaardig dat van deze miRNA's, miR-193a-5p het meest prominente miRNA geassocieerd met obesitas, ook significant geassocieerd bleek te zijn met FLD in Hoofdstuk 2.2.

In **Hoofdstuk 2.2** heb ik het verband onderzocht tussen miRNA spiegels in plasma en de prevalentie van leververvetting. Ik onderzocht de relaties tussen miRNA's zowel leververvetting als leverenzymen, met name gamma-glutamyltransferase en alkalische fosfatase. De resultaten van mijn analyse onthulden dat 61 miRNA's associaties vertoonden met leverenzymen. Bovendien bleken 17 miRNA's geassocieerd te zijn met leververvetting, waarvan 14 miRNA's overlapten met die geassocieerd waren met leverenzymen. Een van de geïdentificeerde miRNA's was met name miR-122-5p, een bekend miRNA in leverweefsel.

In **Hoofdstuk 2.3** onderzocht ik de relatie tussen miRNA-expressie in plasma en alcoholgebruik. Ik ontdekte dat miR-193p, miR-122-5p en miR-3937 positief geassocieerd waren met alcoholgebruik, terwijl miR-4507 een negatief verband vertoonde. Bij het vergelijken van verschillende alcoholconsumptiecategorieën (niet-drinkers, lichte drinkers en zware drinkers), waren de gemiddelde expressieniveaus van zware drinkers bijna het dubbele van die van lichte drinkers, met niet-drinkers als referentiecategorie. Bovendien bleken sommige van de miRNA's die verband houden met alcoholgebruik zich te richten op genen die eerder in verband werden gebracht met alcoholgebruik via GWAS, EWAS, TWAS studies. Pathway analyse van deze target genen onthulde hun betrokkenheid bij biologische regulatie en cellulaire processen. Bovendien vertoonden miR-3937 en miR-122-5p potentiële bemiddelende rollen in de lever.

In **Hoofdstuk 3** onderzocht ik de associatie tussen plasma proteoom met FLD en leverfibrose. In **Hoofdstuk 3.1** vond ik dat 27 inflammatoire en cardiometabole plasma-eiwitten geassocieerd waren met FLD, de sterkste associatie werd waargenomen voor FGF-21 en CES1 eiwitten. Belangrijk is dat 15 van de 27 eiwitten die significant geassocieerd waren met FLD ook geassocieerd waren met leverfibrose. Door middel van pathway-analyse identificeerde ik de betrokkenheid van TWEAK- en TNFS14-eiwitten in de niet-canonieke NF-kB-route. Verder speelden F7, SERPINA5 en PROC-eiwitten een rol speelden in de stollingscascade, met name bij de vorming van fibrinestolsels. Interessant is dat de expressieniveaus van IL-18R1 en CES1 werden opgereguleerd in de organoïden die FLD nabootsen. Ten slotte illustreerde de AUC-analyse dat proteomics de potentie hebben om de momenteel beschikbare niet-invasieve markers (FLI en FIB-4) te verbeteren bij de voorspelling van FLD of fibrose.

Ten slotte onderzocht ik **in Hoofdstuk 4** de associatie van circulerende metabolieten met NAFLD en leverenzymen met behulp van gegevens uit meerdere cohortstudies op populatie niveau. In **Hoofdstuk 4.1** vond ik significante associaties tussen verschillende metabolieten en NAFLD, evenals leverenzymen. Ik vond in totaal 21 metabolieten die gezamenlijke associaties vertoonden met zowel NAFLD als leverenzymen. Binnen het Nightingale platform vond ik fenylalanine, triglyceriden in HDL, IDL en klein LDL, en verschillende verhoudingen van vetzuren zoals linolzuur tot totaal vetzuur, omega 6-vetzuur tot totaal vetzuur en meervoudig onverzadigd vetzuur tot totaal vetzuur. Binnen het Metabolon-platform vertoonden glutamaat en sfingomyeline de sterkste associaties met zowel NAFLD als leverenzymen. Bovendien waren de resterende metabolieten die significant correleerden met de aanwezigheid van NAFLD betrokken bij het metabolisme van lipiden, aminozuren en peptiden.

Chapter 7

Appendices

Acknowledgement

I am deeply grateful for the help and immense support from many individuals who made this thesis possible. Alhamdulillah, here I would like to extend my heartfelt appreciation to all those who directly and indirectly contributed to the realization of this dissertation. Your contributions have been invaluable, and I am truly thankful for your support.

Foremost, I would like to express my sincere gratitude to my promotor, **Prof. dr. Arfan Ikram** for the continuous support of my PhD study and research. I am extremely grateful for the opportunity to study in the epidemiology department and conduct research in this field. Your guidance and support as my promotor have been invaluable, and I consider myself extremely fortunate to have you as my promotor. Thank you for believing in my abilities and providing me with this wonderful chance to pursue my academic and research aspirations in epidemiology.

I would like to express my deepest gratitude to my co-promotor **Dr. Mohsen Ghanbari**, for providing me with the wonderful opportunity to join his research group and conduct my doctoral research under his guidance. From the very beginning, you showed unwavering trust in me and offered firm support, especially during the initial stages when I was still getting acquainted with various epidemiological concepts. Your belief in my potential has been a driving force throughout this journey. Your patience, motivation, and enthusiasm have been truly inspiring. Your vast knowledge in the field of epidemiology has been an invaluable resource which has helped me navigate through the complexities of research. Your mentorship and guidance have played a pivotal role in shaping the content of this thesis, and I am sincerely grateful for your dedicated involvement in every step of the process. Working with you has been a privilege, and I genuinely believe that I couldn't have asked for a better co-promotor. Your support has been instrumental in bringing this research to fruition, and I am deeply appreciative of the time and effort you have invested in my academic growth. Once again, thank you Dr. Ghanbari for everything. I am honored to have had you by my side as my co-promotor during this rewarding academic pursuit.

I express my deep gratitude to the **assessment committee members** (Prof.dr. P.E. Slagboom, Prof.dr. L.J.W. van der Laan, and Prof.dr. M. Schram). I extend my sincere thanks to the members of the large doctoral committee for their valuable participation in my PhD defense. It is truly an honor to have each of you as a member of my doctoral committee.

I am deeply grateful to the **Human Capacity Development Program** (HCDP) Ministry of Higher Education-KRG-Iraq for providing me with a PhD scholarship, enabling me to study at Erasmus MC, Rotterdam, the Netherlands. I would also like to express my gratitude to my lovely paranymphs (**Ibrahim and Sam**).

I extend my heartfelt gratitude to **Dr. Ivana Prokic** for her unwavering support and guidance during my first year of my PhD journey. Her expertise and encouragement have been invaluable to my academic growth.

I am also deeply thankful to the other esteemed academics with whom I had the privilege of discussing and collaborating over the years, including **Fernando Rivadeneira, Robert J. de Knegt, Joyce Van Meurs, Willem P. Brouwer, Qiuwei Pan, Fariba Ahmadizar, Trudy Voortman, Maryam Kavousi, Daniel Bos, Frank Wolters, Costanza L. Vallerga** and other co-authors. Your valuable insights and contributions have enriched my research and profoundly influenced my journey as a researcher. Thank you all for your support and encouragement. It has been a pleasure working with all of you.

Dear **Xiaofang**, working with you has been an absolute pleasure, and our exploration of fatty liver disease has been both enjoyable and enlightening. I want to express my gratitude for all the productive meetings we had and for sharing relevant literature throughout our collaboration.

Dear **Ibrahim** and **Laurens**, I want to extend my heartfelt appreciation to both of you for the valuable chats, discussions, and productive collaborations we had during our work together. It was truly refreshing to collaborate with individuals possessing diverse skill sets, and I firmly believe that this diversity greatly contributed to the high-quality research outcomes, and most importantly, made working on these projects an enjoyable experience.

Dear **Amber**, I cherish the memory of our first meeting during the Linux course the first year of our academic study. I want to express my deep gratitude for all the support and engaging conversations we've had since then. I wish you a very bright future!

Dear **Andreas**, I want to express my sincere gratitude for your invaluable advice and help in organizing my documents during the process of applying for my residence permit multiple times throughout my PhD journey.

Dear **Dr. Ali Farzaneh**, thank you for being my office buddy and for all your help and enjoyable conversations about computer programming. Your friendly and supportive nature has been a great source of comfort. I wish you good health and a bright future ahead!

This journey would have been very different if it weren't for the continuous support I received from my colleagues within our department. I am deeply grateful to **Shahzad, Pooja, Dina, Sam, Georgia, Enping, Lisa, Soroush, Mitra, Nigin, Marlou, Banafsheh, Sven, Amy, Maxime, Mirjam, Natalie, Layal, Gennady, Hata** and everyone else whose name I might have missed (my apologies). Thank you all for being a part of this memorable experience. Dear **Maaike**, I want to express my sincere gratitude for your outstanding support and patience in handling all the complex administration matters. Your help has been vital in making this process smoother. Dear **Nano**, I appreciate your patience and help in resolving IT-related issues, and I am grateful for all the pleasant conversations we've had. Thank you for your support.

A special thanks to **Dr. Adnan M. Abdulazeez**, former president of Duhok Polytechnic University-Iraq, for his unwavering support throughout my PhD journey. His guidance and help in resolving complicated issues with my sponsorship have been invaluable. I am grateful for his continuous backing during this period.

I would like to express my deepest gratitude and love to my family for their steadfast support and encouragement throughout my PhD journey. To my **parents**, thank you for instilling in me the value of education and for always believing in my abilities. Your endless love and sacrifices have been the foundation of my success, and I am forever grateful for your guidance and encouragement. To my **brothers** and **sisters**, thank you for being my pillars of strength and for cheering me on every step of the way. Your constant support and words of motivation have meant the world to me. To my wife, **Narin** you have been my rock and source of inspiration. Your unwavering belief in me, even during the most challenging times, has given me the strength to keep pushing forward. Thank you for being my confidante, my cheerleader, and my partner in this journey. To my children, **Laween** and **Lana** thank you for being my greatest joy and motivation. Your innocent smiles and boundless love have given me the strength to overcome obstacles and pursue my dreams. You are the best gift to me from Allah. I hope to be a role model for you and show you the importance of perseverance and hard work.

To all my friends, family, and colleagues in the Netherlands and Kurdistan, thank you for your friendship, support, and encouragement throughout my PhD journey. Your presence in my life has made this experience more meaningful and enjoyable. I am grateful for the wonderful memories we have shared and for the valuable discussions and collaborations. Your belief in me has been a driving force in achieving this milestone, and I am honored to have you in my life. Thank you for being a part of this incredible journey, and hopefully for new explorations in the future.

Yasir Jameel Abozaid
Rotterdam

List of manuscripts

Abozaid, Y.J., Zhang, X., Mens, M.M., Ahmadizar, F., Limpens, M., Ikram, M.A., Rivadeneira, F., Voortman, T., Kavousi, M. and Ghanbari, M., 2022. Plasma circulating microRNAs associated with obesity, body fat distribution, and fat mass: the Rotterdam Study. *International Journal of Obesity*, 46(12), pp.2137-2144.

Zhang, X., Mens, M.M *., **Abozaid, Y.J** *., Bos, D., Darwish Murad, S., de Knegt, R.J., Ikram, M.A., Pan, Q. and Ghanbari, M., 2021. Circulatory microRNAs as potential biomarkers for fatty liver disease: the Rotterdam study. *Alimentary Pharmacology & Therapeutics, 53*(3), pp.432-442.

Karabegović, I., **Abozaid, Y**., Maas, S.C., Labrecque, J., Bos, D., De Knegt, R.J., Ikram, M.A., Voortman, T. and Ghanbari, M., 2022. Plasma microRNA signature of alcohol consumption: the Rotterdam Study. *The Journal of nutrition, 152*(12), pp.2677-2688.

Abozaid, Y.J., Ayada, I., van Kleef, L.A., Vallerga, C.L., Pan, Q., Brouwer, W.P., Ikram, M.A., Van Meurs, J., de Knegt, R.J. and Ghanbari, M., 2023. Plasma proteomic signature of fatty liver disease: the Rotterdam Study. *Hepatology, 78*(1), pp.284-294.

Yasir Abozaid, Ibrahim Ayada, Laurens A. van Kleef, Neil J Goulding, Jessica S. Williams-Nguyen, Robert C Kaplan, Robert J. de Knegt, Lynne E. Wagenknecht, Nicholette D. Palmer, Deborah A. Lawlor, Nicholas J. Timpson, Jill M. Norris, Yii-Der Ida Chen, M. Arfan Ikram, Willem Pieter Brouwer, Mohsen Ghanbari. Circulating metabolites associated with liver enzymes and fatty liver disease: a multi-platform population-based meta-analysis. (In submission)

Ayada, I., van Kleef, L.A., Zhang, H., Liu, K., Li, P., **Abozaid, Y.J**., Lavrijsen, M., Janssen, H.L., van der Laan, L.J., Ghanbari, M. and Peppelenbosch, M.P., 2023. Dissecting the multifaceted impact of statin use on fatty liver disease: A multidimensional study. *EBioMedicine,* 87.

Ibrahim Ayada1, Jiajing Li, Jiahua Zhou, Laurens A. van Kleef, Willem Pieter Brouwer, Pengfei Li, **Yasir J. Abozaid**, Maikel P. Peppelenbosch, Harry L. A. Janssen, Mohsen Ghanbari, Luc J. W. van der Laan, Robert J. de Knegt, Caroline M. den Hoed, Qiuwei Pan. Magnitude of fatty liver disease before and after liver transplantation and the impact of immunosuppressive medication: a multidimensional study. (In submission)

PhD portfolio

Name PhD student:	Yasir Jameel Abozaid
Erasmus MC Department:	Epidemiology
Research school:	Netherlands Institute for Health Sciences Molecular Medicine Post Graduated School (MolMed)
PhD period:	September 2019-September 2023
Promotor:	Prof. Dr. Arfan Ikram
Co-promotor:	Dr. Mohsen Ghanbari

PhD training	Year	ECTS*
General Courses		
Master of Science in Health Science, NIHES	2023	70
Specific Courses		
Introduction to medical writing	2019	2.0
Linux for scientists	2019	0.6
Genetic-Epidemiologic Research Methods	2019	5.1
Biostatistical methods I	2020	5.7
Biostatistical methods II	2020	4.3
Principles of Research in Medicine and Epidemiology	2020	0.7
Principles of Genetic Epidemiology	2020	0.7
Advances in Genome-Wide Association	2020	1.4
Health Economics	2020	0.7
Causal Inference	2020	1.4
The Practice of Epidemiologic Analysis	2020	0.7
Study Design	2021	4.3
Research Integrity	2021	0.3
LLS Leadership and Teamwork	2021	0.3
LLS Intervision	2021	0.4
Advances in Genomics Research	2021	0.4
Mendelian Randomization	2021	0.9
Introduction to Bayesian Methods in Clinical Research	2021	1.4
Genome Wide Association Studies	2021	1.4
An Introduction to the Analysis of the Next-Generation Sequencing Data	2022	1.4
Introduction to the Analysis of Microbiome and Metabolomics Data	2022	0.7
Introduction to the Analysis of Epigenomics and Transcriptomics Data	2022	0.7
Microbiome Data Analysis in Population-Based Studies	2022	1.4
Topics in Meta-analysis	2022	0.7
Molmed Courses		
The NGS in DNA diagnostic course	2016	1.0
The Basic and Translation Oncology	2016	1.8

The Basic Human Genetics	2016	0.6
The CLC Workbench/ Ingenuity Variant Analysis	2016	0.5
Biomedical Research Techniques XV	2016	1.5
The Galaxy for NGS	2017	1.0
The introduction in Graphpad Prism Version 6	2017	0.3
The course Bayesian Statistics and JASP	2018	0.3
The workshop on Microsoft Excel 2010: Basic	2018	0.3
The workshop on Microsoft Excel 2010: Advanced	2018	0.4
The course on Microbiomics I	2018	0.6
SNPs and Human Diseases	2020	2.0
Photoshop and Illustrator CC 2021 workshop	2021	0.3
Seminars and meetings		
Molecular system Epidemiology meetings, Erasmus MC	2019-2023	1.0
2020 meetings, department of Epidemiology, Erasmus MC	2021	1.0
NAFLD/NASH meetings, department of Gastroenterology and Hepatology, Erasmus MC	2023	1.0
Conferences		
Charge Webinar meeting	2020	0.3
Charge consortium virtual meeting	2021	0.3
ESHG conference, Vienna	2022	0.3

*1 ECTS (European Credit Transfer System) equals a workload of 28 hours.

About the author

Yasir Jameel Abozaid was born on September 12th, 1979 in Aqra, Iraq. He began his academic journey by pursuing a Bachelor's degree in Veterinary Medicine at the University of Duhok College of Veterinary Medicine in 1999, and completed it in 2005. In 2006, he joined the same college as teaching assistant. Later in 2006, he pursued a Master's program in Microbiology and Pathology, earning his MSc in 2008.

In 2015, he received a scholarship under the Human Capacity Development Program (HCDP) sponsored by the Ministry of Higher Education and Scientific Research Kurdistan region of Iraq. This opportunity led him to train at the Urology Department of Erasmus Medical Center University in Rotterdam, the Netherlands. Additionally, in 2016 he worked as a researcher at the Molecular Microbiology and Pathology Department at the same university. In September 2019, he commenced his PhD journey at the Epidemiology Department of Erasmus Medical Center, under the supervision of Prof. Dr. Arfan Ikram and Dr. Mohsen Ghanbari. Concurrently, he pursued a Master's program in Genetic and Molecular Epidemiology at the Netherlands Institute for Health Sciences (NIHES) under the supervision of Dr. Mohsen Ghanbari.

Upon completing his PhD, Yasir aspires to continue investigating the underlying mechanisms of obesity and fatty liver disease. During his academic pursuits, he was also married to Narin in 2008, and since then they have had two children: Laween in December 2010 and Lana in March 2016.