THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

The Dark side of Obesity:
Multi-omics analysis of the dysmetabolic morbidities spectrum

DIMITRA LAPPA

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Life Sciences

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2023

The Dark side if Obesity: Multi-omis analysis of the dysmetabolic morbidities spectrum
Dimitra Lappa

ISBN 978-91-7905-815-9

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5281
ISSN 0346-718X

Division of Systems and Synthetic Biology
Department of Life Sciences
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Cover:
The spectrum of obesity always comes with a dark side. Here, multi-omics assist in the better comprehension of the dysmetabolic morbidities spectrum.

Printed by Chalmers Reproservice
Gothenburg, Sweden 2023

The Dark side of Obesity: Multi-omics analysis of the dysmetabolic morbidities spectrum

Dimitra Lappa
Department of Life Sciences (former Biology and Biological Engineering)
Chalmers University of Technology

Abstract

Obesity is one of the most prevalent clinical conditions worldwide and is associated with a wide spectrum of dysmetabolic comorbidities. Complex cardio-metabolic disease cohorts, such as obesity cohorts are characterised by population heterogeneity, multiple underlying diseases status and different comorbidities' treatment regiments. The systematic collection of multiple types of clinical and biological data from such cohorts and the data-analysis in an integrative manner is a challenging task due to the variables' dimensionality and the lack of standardised know-how of post-processing.

The main resource of this thesis has been the BARIA cohort, a detailed collection over time of multiple omics and demographic data from participants in bariatric surgery. BARIA datasets included plasma metabolites, RNA from hepatic, jejunal, mesenteric and subcutaneous adipose tissues and gut microbial metagenome, besides biometric data. The work presented in this thesis included the development of a systems biology integrative framework based on BARIA that (i) utilised unsupervised machine learning algorithms, self-organizing maps in particular, and multi-omics integrative frameworks, the DIABLO library, in order to stratify the BARIA heterogeneous obesity cohort and predict the bariatric surgery's outcome. The thesis covered how BARIA can be the onset for (ii) studying molecular mechanisms related to type 2 diabetes (T2D) and G-protein coupled receptors (GPCRs) and for identifying a minimal set of biomarkers for obesity's comorbidities such as (iii) non-alcoholic fatty liver disease (NAFL) and (iv) gallstones formation after bariatric surgery.

The results indicated that the metabotypes comprising a bariatric surgery cohort exhibited a concrete metabolic status and different responses over time after the bariatric surgery. It has been demonstrated how obesity and T2D associated metabolites, such as 3-hydroxydecanoate, can increase inflammatory responses via GPCRs molecular activation and signalling. Last but not least, minimal sets of both evasive and non-evasive multi-omic discriminatory biomarkers for obesity's dysmetabolic morbidities (NAFLD and gallstones after bariatric surgery) were obtained. Taking into consideration all the findings, this thesis presented how data-driven approaches can be used for studying in-depth heterogeneous cohorts, hereby facilitating early diagnosis and enabling potential preventive actions.

Keywords: Obesity, systems biology, multi-omics integration, metabotyping, self-organizing maps, biomarkers, bariatric surgery, non-alcoholic fatty liver disease, gallstones, GPCR receptors

List of Publications:

This thesis is based on the work contained in the following papers:

Paper I: Van Olden CC, Van de Laar AW, Meijnikman AS, Aydin O, Van Olst N, Hoozemans JB, De Brauw LM, Bruin SC, Acherman YIZ, Verheij J, Pyykkö JE, Hagedoorn M, Sanderman R, Bosma NC, Tremaroli V, Lundqvist A, Olofsson LE, Herrema H, **Lappa D**, Hjorth S, Nielsen J, Schwartz T, Groen AK, Nieuwdorp M, Bäckhed F, Gerdes VEA. *A systems biology approach to understand gut microbiota and host metabolism in morbid obesity: design of the BARIA Longitudinal Cohort Study*. J Intern Med. 2021;289(3):340–54.

Paper II: **Lappa D**, Meijnikman AS, Krautkramer KA, Olsson LM, Aydin Ö, Van Rijswijk AS, Acherman YIZ, De Brauw ML, Tremaroli V, Olofsson LE, Lundqvist A, Hjorth SA, Ji B, Gerdes VEA, Groen AK, Schwartz TW, Nieuwdorp M, Bäckhed F, Nielsen J. *Self-organized metabotyping of obese individuals identifies clusters responding differently to bariatric surgery*. PLoS One. 2023;18(3):e0279335.

Paper III: Mikkelsen RB, Arora T, Trošt K, Dmytriyeva O, Jensen SK, Meijnikman AS, Olofsson LE, **Lappa D**, Aydin Ö, Nielsen J, Gerdes V, Moritz T, van de Laar A, de Brauw M, Nieuwdorp M, Hjorth SA, Schwartz TW, Bäckhed F. *Type 2 diabetes is associated with increased circulating levels of 3-hydroxydecanoate activating GPR84 and neutrophil migration.* iScience. 2022;25(12).

Paper IV: Meijnikman AS*, **Lappa D***, Herrema H, Aydin O, Krautkramer KA, Tremaroli V, Olofsson LE, Lundqvist A, Bruin S, Acherman Y, Verheij J, Hjorth S, Gerdes VEA, Schwartz TW, Groen AK, Bäckhed F, Nielsen J, Nieuwdorp M. *A systems biology approach to study non-alcoholic fatty liver (NAFL) in women with obesity.* iScience. 2022;25(8).

Paper V: Guman MSS, Hoozemans JB, Haal S, de Jonge PA, Aydin Ö, **Lappa D**, Meijnikman AS, Westerink F, Acherman Y, Bäckhed F, de Brauw M, Nielsen J, Nieuwdorp M, Groen AK, Gerdes VEA. *Adipose Tissue, Bile Acids, and Gut Microbiome Species Associated With Gallstones After Bariatric Surgery.* J Lipid Res. 2022;63(11):100280.

**Contributed equally*

Additional papers not included in this thesis:

Paper VI: Meijnikman AS, van Olden CC, Aydin Ö, Herrema H, Kaminska D, **Lappa D**, Männistö V, Tremaroli V, Olofsson LE, de Brauw M, van de Laar A, Verheij J, Gerdes VEA, Schwartz TW, Nielsen J, Bäckhed F, Pajukanta P, Pihlajamäki J, Tchkonia T, Kirkland JL, Kuipers F, Nieuwdorp M, Groen AK. *Hyperinsulinemia Is Highly*

Associated With Markers of Hepatocytic Senescence in Two Independent Cohorts. Diabetes. 2022 Sep 1;71(9):1929–36.

Paper VII: Li P, Sundh D, Ji B, **Lappa D**, Ye L, Nielsen J, Lorentzon M. *Metabolic Alterations in Older Women With Low Bone Mineral Density Supplemented With Lactobacillus reuteri.* JBMR Plus. 2021;5(4):1–14.

Paper VIII: Lu H, Li F, Sánchez BJ, Zhu Z, Li G, Domenzain I, Marcišauskas S, Anton PM, **Lappa D**, Lieven C, Beber ME, Sonnenschein N, Kerkhoven EJ, Nielsen J. *A consensus S. cerevisiae metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism.* Nat Commun. 2019;10(1).

Paper IX: Kumar M, Ji B, Babaei P, Das P, **Lappa D**, Ramakrishnan G, Fox TE, Haque R, Petri WA, Bäckhed F, Nielsen J. *Gut microbiota dysbiosis is associated with malnutrition and reduced plasma amino acid levels: Lessons from genome-scale metabolic modeling.* Metab Eng. 2018 Sep;49:128–42.

Contribution Summary

Paper I: computational resources, software development, data curation

Paper II, IV: conceptual design, hypothesis generation, data curation, software development, formal data analysis, visualization, main manuscript preparation

Paper III: metabolomics imputation normalization, software development, formal data analysis, data curation, manuscript writing, reviewing & editing

Paper V: metabolomics imputation-normalization, methodology-pipeline development, manuscript reviewing

Preface

This dissertation serves as partial fulfilment of the requirements to obtain the degree of Doctor of Philosophy at the Department of Life Sciences at Chalmers University of Technology. The PhD research was carried out between September 2016 and December 2022 at the division of Systems and Synthetic Biology (Sysbio) under the supervision of Jens Nielsen. The project was co-supervised by Fredrik Bäckhed and Aleksej Zelezniak and examined by Ivan Mijakovic. The project was mainly funded by the Novo Nordisk Foundation (NNF15OC0016798) and partially by the Knut and Alice Wallenberg Foundation.

Dimitra Lappa 2023

TABLE OF CONTENTS

Abbreviations

ADA: American Diabetes Association
ANNS: Artificial Neural Networks
ANOVA: Analysis of Variance
AUC: Area Under the Curve
BCAAs: Branched Chained Amino Acids
BDP/DS: Biliopancreatic Diversion with Duodenal Switch
BF%: Body Fat Percentage
BMI: Body Mass Index
BSH: Bile Salt Hydrolase
CDC: Center of Disease and Control
CONV-R: Conventionally Raised
DIABLO: Data Integration Analysis for Biomarker discovery using Latent cOmponents
FDR: False Discovery Rate
FG: Fasting Glucose
GERD: Gastroesophageal Reflux Disease
GF: Germ Free
GLM: Generalized Linear Models
GPCRs: G-Protein Coupled Receptors
HDL: High-Density Lipoprotein
IFG: Impaired Fasting Glucose
IL: Interleukin
KEGG: Kyoto Encyclopedia of Genes and Genomes
LC-MS: Liquid Chromatography Mass Spectrometry
LDL: Low Density Lipoprotein
LPS: Lipopolysaccharides
MCFAs: Medium Chain Fatty Acids
MMTT: Mixed-Meal Tolerance Test
NAFL: Non-alcoholic Fatty Liver
NAFLD: Non-alcoholic Fatty Liver Disease
OPLS-DA: Orthogonal Partial Least Squares -Discriminant Analysis
PCA: Principal Component Analysis
PLS: Principal Least Squares
RYGB: Roux-en-Y Gastric Bypass
SCFAs: Short Chain Fatty Acids
SOMs: Self-Organizing Maps
T2D: Type 2 Diabetes
TNF: Tumor Necrosis Factor
USA: United States of America
WHO: World Health Organization

To my family and loved ones, G & A

Background

Obesity and comorbidities spectrum

Obesity is a growing epidemic and is one if the leading preventable causes of death worldwide. Approximately 1.9 billion adults are overweight and among these 650 million are suffering from obesity[1–3]. Obesity is associated with increased risk for multiple comorbidities including cardiovascular disease, type 2 diabetes mellitus (T2D), hypertension, dyslipidemia, non-alcoholic fatty liver disease (NAFLD), gastroesophageal reflux disease (GERD) and various types of cancer[4,5].

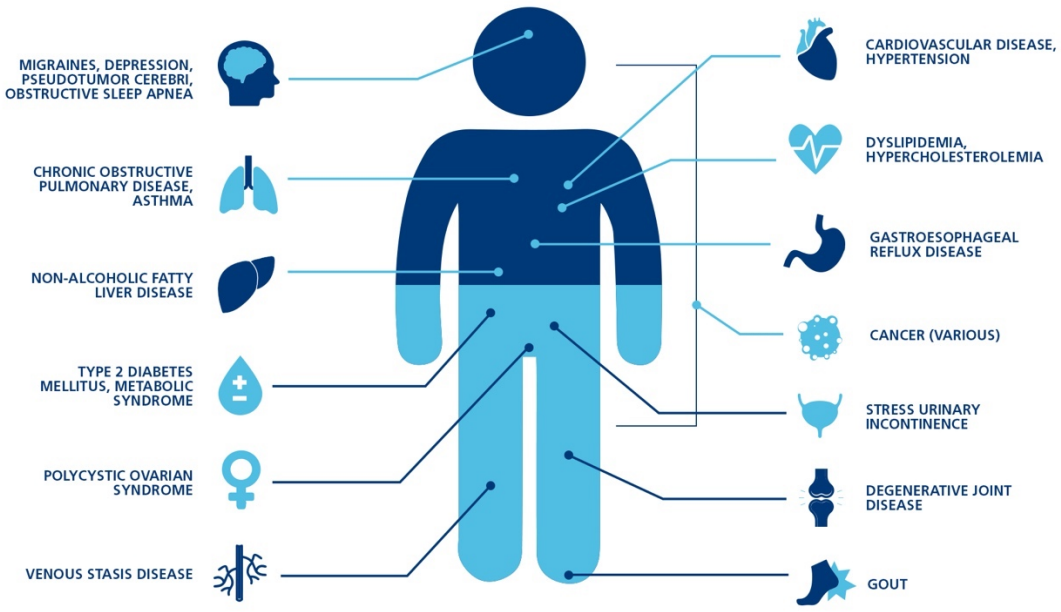

Figure 1: Obesity and associated comorbidities' spectrum.

Obesity definition

Obesity's official definition is the substantial accumulation of body fat that has a great impact in health[1]. The main classification of obese individuals is based on body mass index (BMI) – the ratio of a person's weight in kilograms to the square of their height in meters. Being overweight corresponds to a BMI of 25 or higher, whereas being obese to a BMI of 30 or higher[2]. Other subdivisions of obesity, based on BMI, include class 1 obesity with a BMI 30 to 35; class 2 obesity with BMI ranging from 35 to 40; and class 3 obesity with 40+ BMI , class 3 obesity[6]. However, BMI does not account for the interindividual variation of lean body mass. Often, within the context of medical literature, the metric for obesity besides BMI, is the body fat percentage (BF%) – the ratio of the total weight of person's fat to his or her body weight, where BMI acts as an

approximation of BF%. Indicators of obesity are considered excess BF% of 32% for women and 25% for men[7].

Causes of Obesity

The fundamental cause of obesity is the imbalance between calorie consumption and calorie expenditure. In a global scale there is increased consumption of high energy-high fat diets in combination with a more sedentary lifestyle[1]. Overprocessed food, urbanization, new means of transportation, more stationary professions are also inflicted by societal changes and associated with lack of supportive policies.

Other causes of obesity, with lower prevalence include genetic predisposition and other diseases, mostly of endocrine and psychiatric nature. The discovery of leptin provided new insights in the pathophysiological mechanisms of obesity's development. Leptin is a hormone produced peripherally in adipose tissue, but it effects the central nervous system, specifically the hypothalamus. Leptin essentially controls the appetite and leptin deficiency or leptin resistance can lead to overfeeding, which accounts for some genetic and acquired forms of obesity[8,9]. Certain physical and mental illnesses and the pharmaceutical substances used to treat them can increase the risk of obesity. These are congenital or acquired conditions such as hypothyroidism, Cushing's syndrome, hormone growth deficiency[10] and disorders such as binge eating disorder and night eating syndrome[4]. Moreover, the risk of overweight and obesity is higher in patients with psychiatric disorders, especially depression, since obesity and depression influence each other mutually[11]. Obesity can also be drug-induced, with specific medication causing weight fluctuation, such as insulin, sulfonylureas, thiazolidines, antipsychotics and antidepressants, steroids, hormonal contraceptives and anticonvulsants[4].

Main obesity comorbidities

T2D: One of the main comorbidities is T2D, a condition characterized by high blood sugar, insulin resistance, and relative inability to produce insulin, whilst some people are genetically predisposed to T2D than others[12]. The most common attributes of the disease are thirst, frequent urination, and unexplained weight loss, increased hunger, feeling tired, and sores that do not heal[13]. The constantly elevated blood sugar levels can lead to other long term complications such as heart disease, strokes, diabetic retinopathy, kidney failure, and poor blood flow in the limbs which may lead to amputations[12]. The main T2D diagnostic tools are blood tests such as fasting plasma glucose, oral glucose tolerance test, or glycated hemoglobin. T2D is a preventable disease in a similar way to obesity: by a healthy diet and regular physical activity. The main T2D treatment involves exercise and dietary changes. If high blood glucose levels persist after lifestyle changes, medication is administered, with metformin being the drug

most frequently prescribed[14,15].When the body is unable to produce insulin due to beta-cells failure[16], many T2D patients eventually require treatment with injectable insulin[17].

NAFL: NAFLD is excessive accumulation of fat in the liver without another clear cause such as alcohol use[18]. NAFLD mostly divides into two categories: non-alcoholic fatty liver (NAFL) and non-alcoholic steatohepatitis (NASH). NAFL is more benign than NASH and usually with NAFL not progressing to NASH. NASH can eventually progress to more severe conditions like cirrhosis, liver cancer, liver failure, or cardiovascular disease[19]. Risk factors for NAFLD include obesity and T2D, where over 90% of obese, 60% of diabetic, and up to 20% of normal-weight people develop NAFLD[18,20]. Unfortunately, the only way to confirm NAFLD is via evasive methods, specifically via liver biopsy[21]. NAFLD, as obesity and T2D can be treated with weight loss, dietary interventions and exercise. Severe cases of NAFL and NASH can meet improvement after bariatric surgery and administration of pioglitazone and vitamin E[22], however, since 2017 NAFLD still remains the second most common reason for liver transplantation in both US and Europe[22].

Cancer: Obesity is tightly linked to various types of cancer through a series of different molecular mechanisms[23]. From 2011 to 2015 about 37,670 new cancer cases in men (4.7%) and 74,690 new cancer cases in women (9.6%) were caused by obesity and overweight among people ages 30 and older[24]. There was a wide variation in cancer cases attributed to obesity as high as 51% for liver or gallbladder cancer and 49.2% for endometrial cancer in women and 48.8% for liver or gallbladder cancer and 30.6% for esophageal adenocarcinoma in men. In a global scale for 2012, excess body weight was responsible for approximately 3.9% of all cancers (544,300 cases), with the burden of these cancer cases higher for women (368,500 cases) than for men (175,800 cases)[25].

Obesity and treatment

The primary suggestions of the World Health Organization (WHO) for restricting the obesity epidemic are eating in a healthier manner (consuming fruit, vegetables, legumes, whole grains and nuts), increasing the frequency and the intensity of physical exercise in combination with having supportive environments, communities and policies[2,26]. One less common way to treat obesity is via pharmacological regiments. The use of medications that alter appetite or calorie absorption can potentially regulate the weight, even though it might have a multitude of side-effects. The main tools for treating obesity medically are appetite suppressant drugs with cateholamine releasing agents such as amphetamines and phentermine[27]. Tirzepatide, semaglutide, and liraglutide, all GLP-1 analogues, can affect the rate of gastric emptying and can also have neurologically-driven effects on appetite[28]. Nevertheless, the most effective way to tackle with obesity, besides diet and physical activity, is via surgical intervention.

Bariatric surgery

Bariatric surgery (weight-loss surgery) is considered the most effective obesity treatment option for durable weight loss[29] and overall reduces mortality from 40% to 23%[30]. Bariatric procedures can reduce cardiovascular risk, cause remission of T2D, reduce NAFL and lower the incidence and severity of depression syndromes, parallel to weight reduction[31]. Below are the three most common types of bariatric surgery procedures that are performed today:

Roux-en-Y Gastric Bypass Surgery (RYGB): Gastric bypass is a non-reversible surgical operation, that can also be performed laparoscopically, that helps patients reset hunger and satiety by altering how both the stomach and small intestine handle food, in order to achieve and maintain weight loss goals. RYGB is designed in a manner that restricts food intake and malabsorption properties. In addition, RYGB is regulating gut hormones and their effect in hunger and satiety, even though complete hormonal mechanisms are still to be understood[32]. RYGB reduces the size of the stomach to a small pouch, which is then appended directly to part of the small intestine. This way fewer calories and nutrients from the food get absorbed.

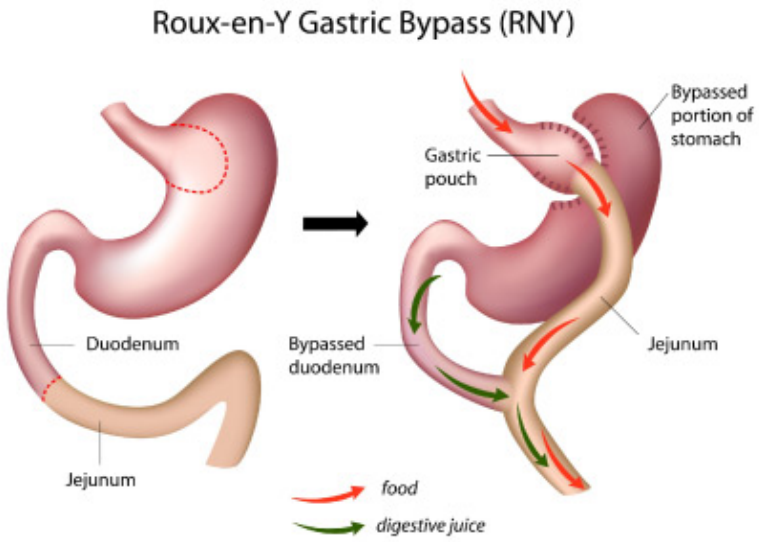

Figure 2: RYGB procedure. *Creating a gastric pouch, by-passing portion of the stomach and appending the pouch straight to jejunum.*

This procedure leads to an increase in baseline satiety hormones, so that the patient feels full by consuming a smaller amount of food[31] RYGB is one of the most popular bariatric surgery procedures, with approximately 140,000 performed in 2005[31,33]. RYGB post-surgical guidelines call for strict adherence to a healthy pattern of eating (low-sugar and low-starch diet) and is considered the method with the most rapid weight loss with comorbidities improving even prior to massive weight-loss, such as T2D remission.

Sleeve gastrectomy: Sleeve gastrectomy (gastric sleeve) is a bariatric surgery procedure where the stomach size is reduced by 85%, by removing completely a large part of it. The open edges are reattached thus creating the stomachs tubular or “sleeve”, shape. Recent research has that gastric sleeve affects the gut signaling hormones[34], besides reducing the size of the stomach. The procedure is performed laparoscopically, is irreversible and has less risk of side effects like ulcers or intestinal strictures[31]. Gastric sleeve post-operative weight loss is comparable to RYGB, although it is not as effective as RYGB at treating GERD or T2D.

Biliopancreatic Diversion with Duodenal Switch (BPD/DS): BPD/DS is a less common bariatric procedure, anatomically and functionally irreversible[33]. The stomach gets re-sected and the remaining stomach part is "tubulized", disconnected from the duodenum and appended to the small intestine. Compared to the Sleeve Gastrectomy and RYGB, BPD/DS produces the best results in terms of durable weight loss and resolution of T2D. Similar to RYGB, BPD/DS causes significant alteration in gut hormones that control hunger and satiety, complementary to its restriction and malabsorption properties[35].

Bariatric surgery complications

Bariatric surgery, as all surgical interventions can have post-operative complications. The most common side effects are nutrient malabsorption, nutritional deficiencies and is dumping syndrome [36], where food moves too quickly from the stomach to the small intestine. Other complications are osteopenia and hyperparathyroidism, due to low calcium absorption in the blood stream since the food is by-passing the stomach proceeding to the small intestine[34]. Bariatric surgery can affect kidney function causing nephrolithiasis (kidney stones) and in severe cases even renal failure[31,34]. Gallstones can also be formed in the post-surgical rapid weight-loss phase, which also require evasive treatment via cholesystectomy (gallbladder removal)[37].

Cost of obesity and challenges in healthcare

Given the prevalence of obesity and all the diseases that are directly associated and caused by excessive weight, once can infer that there is an increasing burden in health care. Approximately 41.9% of United States (US) adults are obese (as measured from 2017-2020), a rate that has increased significantly from a level of 30.5% in 1999-2000, according to Center of Disease Control (CDC)[38,39]. Obesity-related illnesses lead to increased risk for hospitalization and treatment. Obesity’s comorbidities account for 14.3% of US healthcare spending, and also result in significant losses to the economy through decreased productivity[31]. The lack of early, non-evasive, diagnostic tools for obesity’s comorbidities can further add to the cost of healthcare, seeing that comorbidities such as NAFL require evasive biopsies for confirmation. Given that bariatric surgery is the most effective way to deal with obesity, once needs to account

for the costs related, ranging from $11,500 to $26,000 in the US[39]. Additionally, the cost for postoperative complications can add up to the expenses, even though the long-term health benefits from weight loss are estimated to outweigh the direct cost of obesity and required treatment for directly linked diseases. Considering the risks, costs and effects of obesity there is an equivocal need for the [40]obesity epidemic to be contained and the healthcare costs implied to be restricted. The means of materializing this is getting a deeper understanding of the underlying mechanisms of obesity and identifying other ways for early diagnosis and treatments. This is achievable by conducting more studies on larger population obese cohorts and by emphasizing the generation of extensive multilayered datasets.

Challenges posed by obesity longitudinal cohorts

There is significant variation among obese individuals in the development and severity of these comorbidities. This complexity, poses a substantial challenge to systematization and analysis of obesity, particularly since appropriate stratification of study participants is often complicated by multiple confounding comorbidities and medications. Confounding factors in combination with the natural interindividual variation in obesity can cause multiple metabolic perturbations. These perturbations of metabolism can conceal molecular mechanisms and signaling across different tissues, hence making it more difficult to identify and comprehend the pathophysiology behind obesity and comorbidities. From a clinical perspective, it is common to group individuals with obesity based on metabolic syndrome categories or treatment(s) for already existing comorbidities. Still, the emerging complexity of the contributing factors does not provide for a comprehensive framework to study obese populations. Using clinical parameters, numerous approaches have been proposed to model obesity and predict cardiovascular risk[41,42]. Once more, these approaches focused on identifying comorbidities associated with obesity, such as T2D and traditional clinical diagnostic markers. However, most of these studies did not account for the strong confounding effects from lifestyle or medication.

Systems Biology for studying in-depth complex cohorts

A systems biology approach offers more possibilities for obtaining an in-depth phenotypic profile of obese individuals by using omics analysis. Systems biology encompasses the computational and mathematical analysis and modeling of complex biological systems. It is a relatively new, emerging, interdisciplinary field that starting from biology, focuses on multiplex interactions within biological systems, using a holistic approach [43]. Systems biology primary goal is employing various techniques of modeling complex biological systems for discovering the emergent properties

of cells, tissues and organisms functioning as a system, such as metabolic networks or cell signaling networks[43,44].

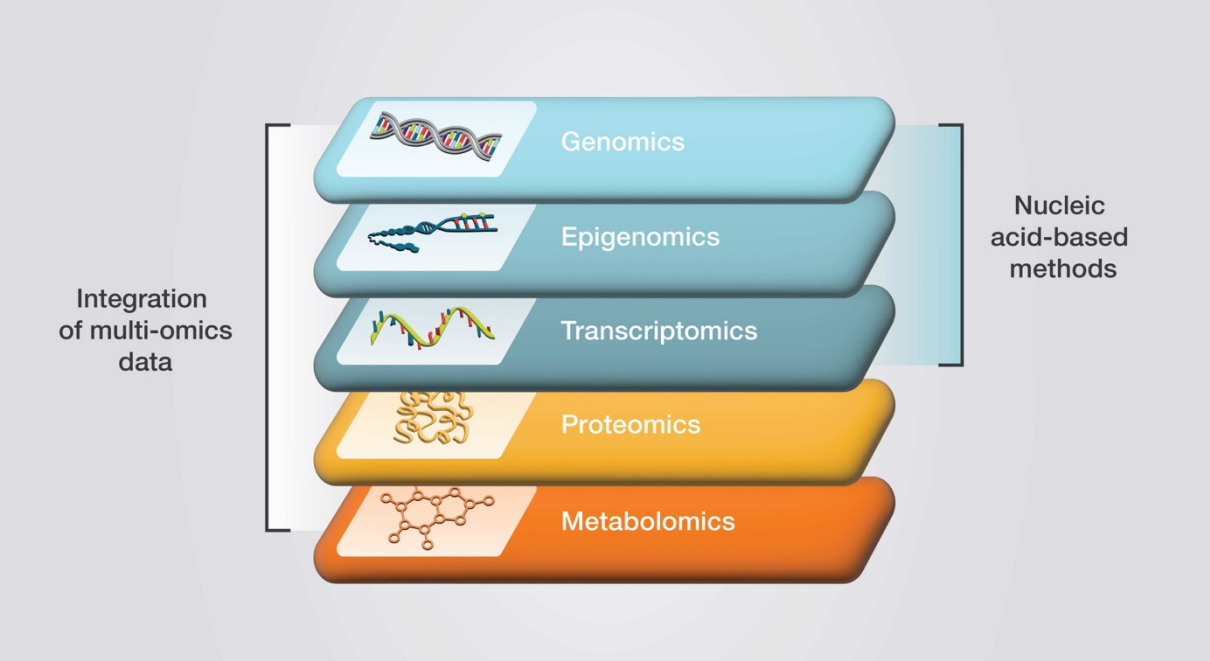

Figure 3. Types of omics that contribute to integrated multi-omics. *The top three omics are nucleic acid-based omics methods.*

The most common methods for collecting quantitative data for the construction and validation of systems biology models are genomics, epigenomics, transcriptomics, metabolomics, proteomics and high-throughput techniques[45]. Metabolomics is "systematic study of the unique chemical fingerprints that specific cellular processes leave behind"[46]. The metabolome represents the complete set of small compound molecules in a biological cell, tissue, organ, or organism, which are the end products of cellular processes[47]. Metabolomics, in the context of obesity, is consistently being used as a means of evaluating metabolic health, measuring the effect dietary intervention strategies and to identify predictive biomarkers characterizing a specific condition[48]. Transcriptomics is the study of the 'transcriptome,' an entire set of transcripts, this includes their transcription and expression levels, functions, locations, trafficking, degradation and structure of the parent genes[49]. The expression of multiple transcripts in different physiological or pathological conditions via high-throughput methods is used to investigate the relationships between transcriptome and phenotype in living entities[50]. The gut microbiome and its interactions with the host is another contributing factor to the obesity complexity spectrum[51–54]. The causal role of the gut microbiome in obesity and insulin resistance has been investigated via experimental models in animals[55,56]. Animal-studies have, therefore, been extensively used to investigate the complex interaction between genes, diet, lifestyle, and the gut microbiota as well as the molecular mechanisms underlying obesity[57–59]. However, one of the main limitations of animal studies is the extent to which these results translate to humans[60]. These efforts have included not only biochemical profiling, but a

combination of metabolites and specific clinical parameters, with the aim to identify optimal candidate groups for further interventions[61,62]. Studies in humans have also identified associations between the microbiota and cardiometabolic disorders, but a causal role for the microbiome has not yet been established[63]. Recent studies have pinpointed the production and regulation of specific metabolites of bacterial origin in humans and shown that these play an important role in metabolic diseases such as T2D, cardiovascular disease and other pathogenic conditions[53,54,64,65].

Given these complex interactions between the gut microbiota, host, and multiple obesity-associated comorbidities, there is a clear need to apply a more holistic systems biology approach to population-based studies of obesity to not only improve the identification of distinct subpopulations with obesity but also drive the development of personalized interventions. Basing such analysis on multi-omics datasets instead of clinical metadata is preferable since clinical metadata can act as confounders[66,67].

Challenges in multiple data-types integration

Although, systems biology can enable identification of novel biomarkers it has been challenging to integrate multi-omics data to gain novel mechanistic insights into the pathogenesis of disease, particularly since analysis of such data has traditionally been done by comparing predefined groups based on specific clinical parameters[68,69]. Due to natural population variation, comorbidities, and the heterogeneity of complex multigenic diseases like obesity and T2D it has, however, been difficult to extract molecular information from this approach[68].

Till recently there were limited options of computational tools that could perform “real” multiple omics type integration in a systematic manner. By way of high-throughput techniques and a rising trend in the consolidation of larger cohorts with multiple data types the development of integrative omics tools has escalated[70]. The use of machine learning methods is therefore gaining more attention for the integration and analysis of various omics datasets thus providing the means for the discovery of new biomarkers, more accurate disease prediction and delivery of precision medicine[71]. Machine learning methods in multi-omics integration can deconvolute multifactorial disease[72,73], in particular as it enables stratification of individuals in a given cohort, without a priori knowledge of clinical labels.

Aim and Scope

We have established that stratifying obesity cohorts characterised by multiple comorbidities, various pharmaceutical treatments along with analysing their complex clinical and biological datasets is a very challenging task. This thesis aims to address the systematic modelling and studying in-depth such cohorts, based solely on biological data and not only clinical diagnostic criteria. Additional goals are discovering and comprehending molecular mechanisms behind metabolic diseases linked to obesity and obtaining minimal sets of biomarkers for predicting the occurrence of these diseases that need evasive biopsies and treatments.

Therefore, through the detailed collection over time of multiple types of omics datasets, demographic data, biometric data from the BARIA cohort (**Paper I**), we developed a systems biology integrative framework that utilises Artificial neural Networks (ANNs) and unsupervised machine learning techniques (SOMs), for stratifying a heterogeneous obesity cohort and predicting the bariatric surgery's outcome (**Paper II**). In order to identify, study and comprehend the molecular mechanisms that link T2D and obesity we employed animal models and *in vitro experiments* (**Paper III**). Lastly, we constructed a pipeline that can identify a minimal set of prognostic biomarkers for obesity's comorbidities such as NAFLD and gallstones formation after bariatric surgery (**Papers IV** and **V**).

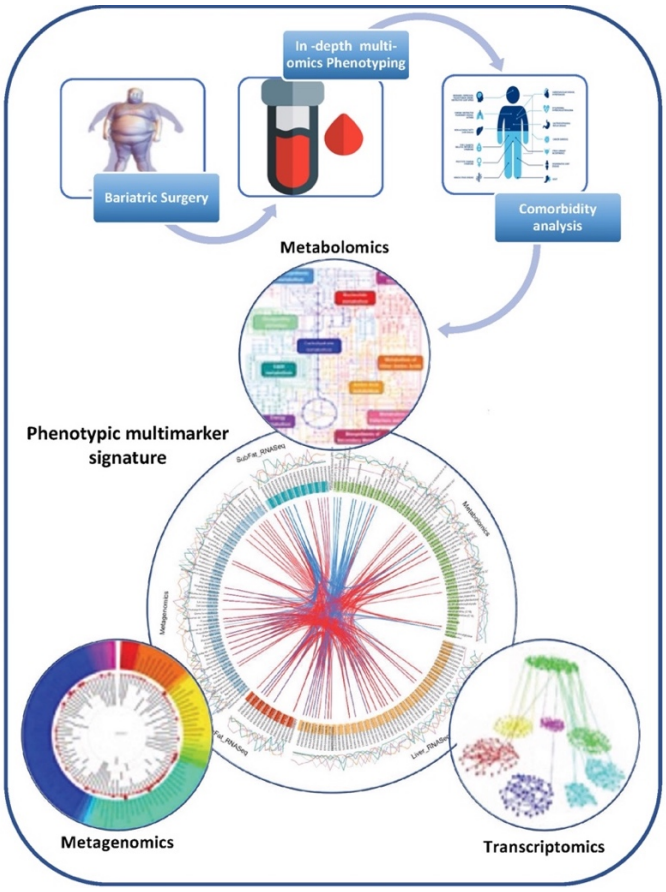

Figure 4: Thesis' aim, scope and structure

PART I: Cohort, methods and computational tools

The BARIA Cohort

The material for analysing a complex cohort, that includes obese individuals in a longitudinal follow up comes from the BARIA study[74]. The BARIA participants are patients with morbid obesity scheduled for bariatric surgery. The scope of the study is assessing how gut microbial species and microbially produced metabolites affect the transcription in key tissues and and how baseline anthropometric and metabolic characteristics determine weight loss and glucose homeostasis after bariatric surgery[74]. The study was performed in accordance with the Declaration of Helsinki and was approved by the Ethical Review Board of the Academic Medical Center, Amsterdam (approval code: NL55755.018.15). All participants provided written informed consent.

Paper I: BARIA

The BARIA study aims to create and implement a systems biology approach for studying obesity and directly associated morbidities (T2D and NAFLD) by identifying gut microbial, immunological and metabolic markers and novel pathways in the pathogenesis of obesity in a large and well phenotyped bariatric surgery cohort (BARIA study). All the BARIA participants will be followed-up prospectively so as to identify mechanisms affecting the surgical outcome.

During the bariatric surgery, different types of tissues are excised and are assessed for histology (paraffin embedded), gene regulation (RiboNucleic Acid (RNA)-sequencing) and protein expression (immunoblotting):

- Hepatic tissue from the diaphragmatic surface of segment three or five of the liver
- Jejunal tissue at the site of the jejunojejunostomy, approximately 50 cm from the Treitz ligament
- Subcutaneous tissue (from one of the laparoscopic incisions in the upper abdomen)
- Greater omentum and visceral fat tissue (omental appendices of the transverse colon)

Furthermore, EDTA plasma metabolomic datasets are collected at baseline, scheduled follow-up visits and during the bariatric surgery. These include:

- Peripheral Blood from individuals in fasting conditions
- Peripheral Blood from individuals, two hours after the ingestion of a Mixed-Meal Tolerance Test (MMTT), so as to capture the post-prandial metabolic response
- Portal vein blood samples, taken at the beginning of the surgery, only if considered safe by the surgeon.

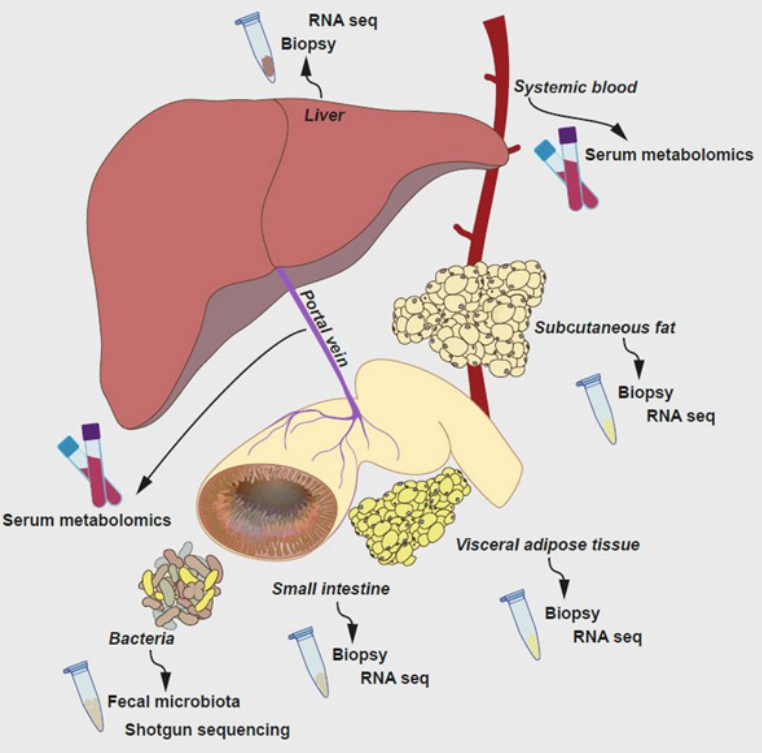

Figure 4: A systems biology approach, identifying gut microbial, immunological and metabolic markers in a large and well phenotyped bariatric surgery cohort

Additional types of data collected from the patients at both baseline and during post-operative follow up controls are:

- Faecal samples obtained at several time points, to analysed by shotgun sequencing
- Demographic and lifestyle metadata
- Biometric metadata
- MMTT: Within three months before surgery, a 2-hour MMTT was performed to assess insulin resistance and investigate dynamic alterations in circulating metabolites. The MMTT consisted of a compact 125ml drink (Nutricia®) containing in total 23.3 grams fat, 74.3 grams carbohydrates (of which 38.5 grams sugar) and 24.0 grams protein. The participants received this meal after fasting for a minimum of nine hours. Time-point zero refers to the moment at which the participant had fully consumed the meal. Blood samples were drawn *via* an intravenous line at baseline, 10, 20, 30, 60, 90 and 120 minutes. All samples were stored at -80^{o}C until further processing.
- Psychological questionnaires, prior during and after a MMTT

BARIA study preliminary results

The results of the preoperative MMTT of the patients included and operated in the first two years of the study were used to validate the reproducibility of the MMTT stimulated postprandial glucose, triglycerides and insulin curves. The BARIA individuals were

stratified into groups based on their glycaemic control as formulated in the American Diabetes Association (ADA) criteria:

- normoglycemia (fasting glucose (FG) <100 mg dL^{-1}; <5.6 mmol L^{-1})
- impaired Fasting Glucose (FG) (100–125 mg dL^{-1}; 5.6–6.9 mmol L^{-1})
- and / or increased haemoglobin A1c (5.7–6.4%; 39–47 mmol mol^{-1})
- diabetes mellitus (FG ≥126 mg dL^{-1}; ≥7.0 mmol L^{-1})

MMTTs prior to the bariatric surgery from 170 patients were analysed and clear differences were observed in glucose homeostasis between individuals. Individuals belonging to different groups of glycaemic control exhibited different profiles for MMTT stimulated plasma insulin, glucose and triglycerides. Triglycerides were higher at baseline and all following time points in individuals with Impaired Fasting Glucose (IFG), with or without increased Hba1c, as can be seen in **Table 1:**

Table 1. Baseline characteristics and results of mixed meal test in 170 participants in the first two years of inclusion in the BARIA longitudinal cohort study.

	healthy	**IFG**	**IHbA1c**	**Comb**	**T2D**
n	57	21	19	26	47
age (years)	41.4 (11.1)	46.8 (11.7)	44.6 (9.5)	49.2 (9.2)	49.5 (10.2)
sex (female)	45 (78.9)	20 (95.2)	17 (89.5)	16 (61.5)	31 (66.0)
BMI	39.5 (3.9)	39.4 (3.1)	40.6 (7.1)	40.6 (3.6)	39.2 (4.5)
hypertension	8 (14.0)	5 (23.8)	3 (15.8)	8 (30.8)	25 (53.2)
Systolic BP (mmHg)	129.5 (16.6)	130.6 (13.6)	134.2 (15.8)	133.2 (12.0)	132.1 (13.7)
Diastolic BO (mmHg)	80.1 (11.3)	80.5 (8.2)	78.1 (13.2)	84.0 (7.9)	82.6 (9.4)
insulin use					10 (21.3)
glucose (mmol/l)	5.1 (0.4)	5.9 (0.2)	5.2 (0.2)	6.1 (0.4)	7.4 (1.5)
insulin (pmol/l)	84.8 (48.0)	89.4 (46.5)	79.2 (37.2)	111.2 (46.9)	180.2 (222.5)
HbA1c (%)	5.31 (0.23)	5.41 (0.19)	5.79 (0.09)	5.88 (0.17)	7.10 (1.14)
HOMA2 IR	1.60 (0.90)	1.71 (0.83)	1.48 (0.67)	2.14 (0.85)	2.44 (1.24)
HOMA2 Beta (%)	125.4 (50.9)	98.1 (37.2)	112.6 (33.9)	105.8 (38.7)	87.3 (37.2)
AUC glucose (mmol/l)	137.1 (109.5)	122.5 (85.9)	194.6 (112.9)	211.7 (105.0)	386.3 (193.7)
AUC insulin (mmol/l)	42.3 (30.4)	46.0 (29.4)	48.7 (21.4)	50.8 (20.8)	37.6 (31.5)
eGFR (MDRD ml/min/1,73m^2)	94.5 (18.0)	92.7 (19.8)	95.6 (21.7)	94.7 (19.7)	95.7 (17.6)
ASAT (U/l)	23.6 (4.9)	23.5 (6.5)	25.1 (5.5)	25.3 (4.9)	29.9 (14.0)
ALAT (U/l)	28.6 (13.4)	28.3 (14.7)	33.7 (18.5)	30.4 (10.1)	42.1 (25.8)
Cholesterol (mmol/l)	4.6 (1.0)	5.1 (1.2)	5.2 (1.0)	4.8 (1.1)	4.1 (0.9)
HDLc (mmol/l)	1.12 (0.29)	1.13 (0.23)	1.16 (0.16)	1.08 (0.29)	1.05 (0.23)
Triglycerides (mmol/L)	1.08 (0.44)	1.58 (0.91)	1.10 (0.42)	1.79 (1.17)	1.40 (0.62)

Inclusions stratified by glycaemic classification, as formulated in the American Diabetes Association criteria: normoglycemic (Healthy), impaired fasting glucose (IFG), increased haemoglobin A1c (IHbA1c), combination of IFG and IHbA1c (Comb) and Type-2 Diabetes Mellitus (T2DM). Categorical variables are displayed as absolute numbers (percentage), continuous variables as means (SD).

MMTTs were repeated for in 10 BARIA individuals and showed satisfactory intra-individual reproducibility, with differences in plasma glucose, insulin and triglycerides within 20% of the mean difference. This constitutes the MMTT as a better estimate for glycaemic regulation than the oral glucose tolerance test[75]

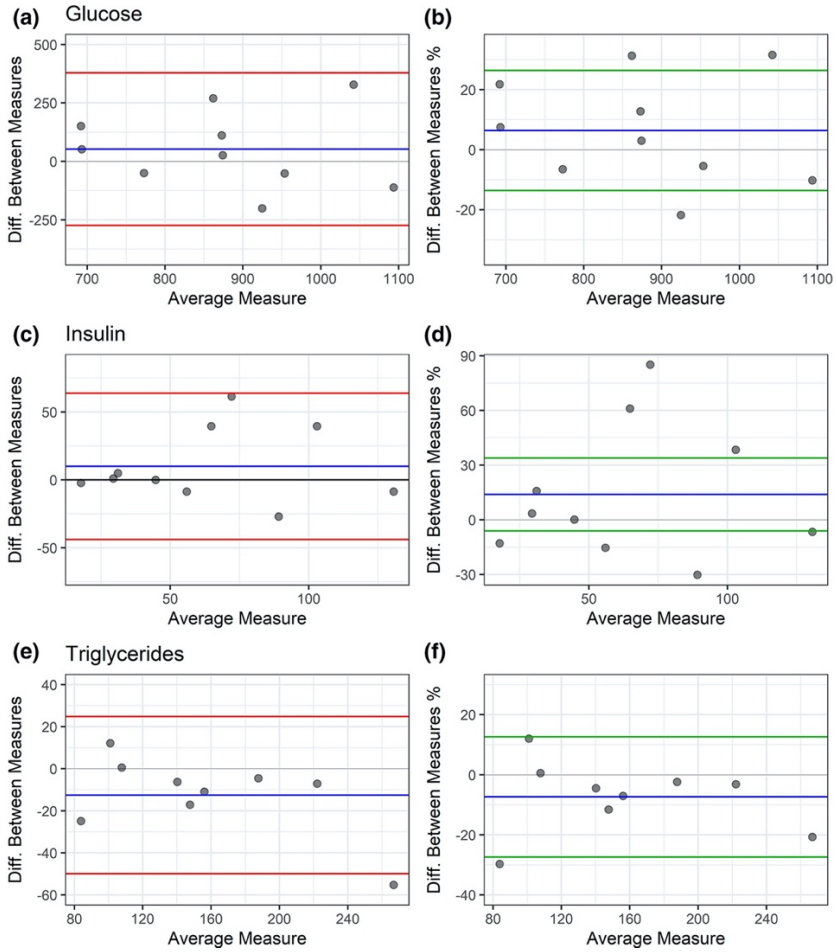

Figure 5: Reproducibility of mixed meal test (MMT). *Bland Altman plots of MMT (repeated within 1 week) for glucose, insulin and triglycerides. Blue line is mean of difference between measurements, red line is +/- 1.96*SD of mean difference, green line is +/- 20% of mean difference. A) Glucose area under the curve (AUC) in mmol/L*time. B) Glucose AUC percent change. C) Insulin AUC in mmol/L*time. D) Insulin AUC percent change. E) Triglycerides AUC in mmol/L*time. F) Triglycerides AUC percent change.*

BARIA scope

The BARIA cohort study is the means of generating a large phenomic systems biology database for subjects with morbid obesity, prior and after bariatric surgery. This knowledge database in combination with ANNs and machine learning will be the main platform for selecting microbiome-produced metabolites and for identifying their receptors in target tissues to deepen our understanding for obesity and its associated dysmetabolic morbidities.

Multi-omics extraction protocols and statistical analysis tools

Metabolome

EDTA plasma samples under both fasting and post-prandial conditions were collected BARIA participants. Samples were shipped to METABOLON (Morisville, NC, USA) for performing analysis using ultra high-performance liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) untargeted metabolomics, as previously described[76]. The metabolomic counts obtained, underwent significant curation via metabolites' pre-filtering, imputation for subsets of metabolites' missing values and data normalization, in order to minimize the effect of artifacts in the downstream analysis. Metabolomics prefiltering and imputation were performed by utilizing a variation of the Perseus platform[77]. Essentially, data has been pre-filtered so as to have a maximum of 25% missing values for a metabolite across all samples. This was followed by a log transformation of all the measured metabolites' raw intensities across the entire dataset. Then, we calculated the total data mean and standard deviation (by omitting missing values). Taking into account that the metabolite intensities distribution is approximately following normality, we chose a small distribution 2.5 standard deviations away from the original data mean towards the left tail of the original data distribution, with 0.5 standard deviations width. This new shrunken range corresponds to the actual lowest level of detection by the spectrometer. Here by drawing random values from this mini distribution, we filled the missing prefiltered data of choice. Normalization was conducted to the total signal for each sample, since each sample is a separate injection on the mass spectrometer. Effective control for changes in sample matrix affects ionization efficiency, hence there can be inevitable differences in how much each sample is loaded onto the column with each injection, etc. Therefore, we summed up the total ion intensity (i.e. total signal) for each of the samples and identified the sample with the lowest total signal. After this we could proceed to calculating the correction factor for each sample by dividing the total signal with the lowest total signal, $\boldsymbol{CorrectionFactor_i = \frac{Total\ signal\ for\ each\ individual\ sample_i}{Lowest\ total\ signal\ intensity}}$. The next step is to divide each individual metabolite within a sample with the respective $CorrectionFactor_i$. All the calculations for imputing and normalizing the metabolomics dataset have been conducted with MATLAB_R2018b and the standard built-in packages. Differential significance analysis was conducted in R (version 3.6.3) and RStudio (version 1.2.5033). Statistical analysis has been performed with two methods: ANOVA(Analysis of Variance) and Kruskal Wallis test, with the use of HybridMTest package[78]. HybridMTest performs hybrid multiple hypothesis testing using empirical Bayes probability. The significance level and cut-off used for the dataset of fasting peripheral plasma was $P<0.05$ and was applied to metabolites that were significantly differential with both ANOVA and Kruskal Wallis methods.

Transcriptome

Biopsies from liver, jejunum, mesenteric adipose fat and subcutaneous adipose fat were collected at the time of the bariatric surgery[74]. RNA was extracted from biopsies using TriPure Isolation Reagent (Roche, Basel, Switzerland) and Lysing Matrix D, 2 mL tubes (MP Biomedical, Irvine, CA, USAs) in a FastPrep®-24 Instrument (MP Biomedical, Irvine, CA, USAs) with homogenization for 20 seconds at 4.0 m/sec, with repeated bursts until no tissue was visible; homogenates were kept on ice for 5 minutes between homogenization bursts if multiple cycles were needed. RNA was purified with chloroform (Merck, Darmstadt, Germany) in phase lock gel tubes (5PRIME) with centrifugations at 4°C, and further purified and concentrated using the RNeasy MinElute kit (Qiagen, Venlo, The Netherlands). The quality of RNA was analysed on a BioAnalyzer instrument (Agilent), with quantification on Nanodrop (Thermo Fisher Scientific, Waltham, MA, USA). Due to degradation of the RNA, libraries for RNAseq sequencing were prepared by rRNA depletion; library preparation and sequencing were performed at Novogene (Nanjing, China) on an HiSeq instrument (Illumina Inc., San Diego, CA, USA) with 150 bp paired-end reads and 10G data/sample. The average read count per sample from liver and jejunum tissues were 42 ± 15 million. For mesenteric and subcutaneous fat, the average read count per sample were 43.2 ± 20 million. The extracted fastq files were analyzed with nf-core/rnaseq[79], a bioinformatics analysis pipeline used for RNA sequencing data. The workflow processed raw data from FastQ inputs (FastQC, TrimGalore!), aligned the reads (STAR) with *Homo sapiens* GRCh38 as reference genome, generates gene counts (featureCounts, StringTie) and performed extensive quality-control on the results (RSeqQC, dupRadar, Preseq, edgeR, multiQC). The pipeline was built using Nextflow. Differential gene expression analysis has been performed for liver, jejunum, subcutaneous adipose and mesenteric adipose tissues, respectively, in R (version 3.6.3) and RStudio (version 1.2.5033) with DESeq2[80] package. The statistical analysis method for calculating differential expression rates was the LRT test (log-ratio test). After False Discovery Rate (FDR) correction (FDR 5%) with multiple hypothesis testing with IHW[81] package, we analyzed genes with $P<0.05$ by DEGreport's[82] degPatterns function, so as to identify subgroups of co-expressed genes. For these differentially significant co-expressed genes we performed gene enrichment with Enrichr platform[83,84] using KEGG (Kyoto encyclopedia of genes and genomes) metabolic pathways[85].

Gut Microbial Metagenome

Fecal samples were collected on the day of surgery and immediately frozen at -80C. Total fecal genomic DNA was extracted from 100 mg of feces using a modification of the IHMS DNA extraction protocol Q[86]. Briefly, fecal samples were extracted in Lysing Matrix E tubes (MP Biomedical, Irvine, CA, USA) containing ASL buffer (Qiagen), and lysis of cells was obtained, after homogenization by vortexing for 2 minutes, by two cycles of heating at 90°C for 10 minutes followed by three bursts of bead beating at 5.5 m/sec for 60 seconds in a FastPrep®-24 Instrument (MP Biomedical,

Irvine, CA, USAs). After each bead-beating burst, samples were placed on ice for 5 minutes. The supernatants containing fecal DNA were collected after the two cycles by centrifugation at 4°C. Supernatants from the two centrifugations steps were pooled and a 600 µL aliquot from each sample was purified using the QIAamp DNA Mini kit (Qiagen, Venlo, The Netherlands) in the QIAcube (Qiagen, Venlo, The Netherlands) instrument using the procedure for human DNA analysis. Samples were eluted in 200 µL of AE buffer (10 mmol/L Tris·Cl; 0.5 mmol/L EDTA; pH 9.0). Libraries for shotgun metagenomic sequencing were prepared using a PCR-free method; library preparation and sequencing were performed at Novogene (Nanjing, China) on an HiSeq instrument (Illumina Inc., San Diego, CA, USA) with 150 bp paired-end reads and 6G data/sample. MEDUSA[51] pipeline was used for pre-processing of raw shotgun metagenomics sequence data. MEDUSA is an integrated pipeline for analysis of short metagenomic reads, which maps reads to reference databases, combines output from several sequencing runs and manipulates tables of read counts. The taxon ids were input to taxize[87] package, so as to get full taxonomic information and ranking for the species. This dataset was input to DESeq2 and phyloseq[88] packages for conducting downstream differential statistical analysis. Similar to the BARIA transcriptomics counts, log normalization has been conducted based on gene counts geometric distribution.

PART II: Obesity, multi-omics and metabotypes

It has been established that one of the ways to tackle morbid obesity is bariatric surgery. This weight-loss surgical procedure via direct interventions in the gastric tract can effectively assist in reducing both weight and the risk or occurrence of other life-threatening conditions directly associated with obesity[33]. However, there is a great variability in the response to bariatric surgery. Many studies have shown that some patients regain weight 2-10 years post operatively[89–91].

Paper II: SOM

In this paper we aimed to address the issue of post-operative weight loss responses of the BARIA participants. Thus, we developed an integrative framework that utilizes SOMs and multiple omics datasets. This approach resulted in the identification of distinct metabotypes within our BARIA inclusions, with a characteristic minimal omic biomarkers signature, that exhibited different response to surgery over time.

Here, 106 BARIA participants were recruited and underwent a complete metabolic and omic work-up at the start of the bariatric surgery process. In the analysis we included the biometric and clinical metadata in baseline, presented in **Table 2,** along with the metabolomics, transcriptomics and gut microbial metagenomics datasets extracted and analysed as described in **Part I.**

Table 2. BARIA cohort: 106 participants clinical metadata summary for SOM-defined clusters.

Clinical Metadata	***BARIA population***
Demographic	
Participants (%)	106(100%)
Female (% Total Participants, % of SOM Cluster)	84(79.2%)
Male (% Total Participants, % of SOM Cluster)	22(20.8%)
Anthropometric & clinical lab values	
Age (years)	46(20-14)
BMI (kg/m^2)	39.42(32.9-70)
Waist circumference(cm)	84.3 ± 57.7
Upper thigh circumference (cm)	122.5(103-165)
Total Body Fat (%)	51(31.7-104.8)

Fat Free Mass (kg)	58.9(47.5-93.8)
Systolic blood Pressure (mmHg)	132.5(102-193)
Diastolic blood Pressure (mmHg)	81(45-121)
Fasting glucose (mmol/l)	5.8(4.5-14.8)
HbA1c (mmol/mol)	5.7(4.6-9.8)
HOMA-IR	1.6(0.6-6.9)
HOMA2-β	93.5(29.1-357.8)
Total Cholesterol (mmol/l)	4.9 ± 1.1
Triglycerides (mmol/l)	1.4(0.6-6)
HDL Cholesterol (mmol/l)	1.6(0.2-2.7)
LDL Cholesterol (mmol/l)	3 ±1.1
Creatinine (μmol/l)	66(46-172)
Glomerular Filtration Rate (kl/1.73m2)	88.5(26-91)

Baseline characteristics of the 106 BARIA participants included in the study. Data is expressed as mean ± standard deviation. Categorical variables are presented as numbers and percentages. Non-normally distributed variables are presented as median with interquartile range. BMI: Body Mass Index, HbA1c: Hemoglobin A1c, HOMA-IR: Homeostatic Model Assessment of Insulin Resistance, HOMA-β: Homeostatic Model Assessment of beta-cell function, LDL: Low-Density Lipoprotein, HDL: High-Density Lipoprotein.

Metabotyping with SOMs

In order to be able to study the 106 BARIA inclusions in depth, without being biased from the bariatric surgery population's underlying diagnosis and medical treatments, we had to choose criteria for stratifying the patient inclusions. The stratification concept that we implemented is metabotyping, phenotyping based on the metabolome[62]. Since the metabolome is the most direct phenotypic reader, that captures crosstalk between metabolic regulation and gut microbiota, metabotyping has recently been employed to better understand heterogeneous cohorts and link different metabotypes to obesity and its comorbidities[61,92–94].

The fasting peripheral metabolome was used as a scaffold for metabotyping. The dataset was given as input to the SOM toolbox[95,96]. SOM conducted unsupervised competitive learning and produced low-dimensionality visualizations by employing vector quantization[96,97]. Every data item, mapped into one point (node or neuron) in the map[98]. The SOM was trained with the batch training algorithm, where the assignment of the input vectors to the ANN's nodes, was done by the calculation of the Euclidean distances among prototype vectors and each neuron. In each step of the

training, the node with the smallest Euclidean distance from the input vector was the "winner". Then, the input vector got assigned to the winner node and the weights of the prototype vector (and its spatial neighbors) got updated[99,100]. In this manner all the local re-arrangements got propagated in the grid, during the training epochs. Consequently, the more similar data items are placed closer in the map. This resulted in 48 nodes, which is a concrete reduction in the original dimensions of the dataset, nevertheless not as interpretable.

We then subjected the map into further partitioning with $k-means$ clustering, aiming to get an even more comprehensible stratification. $K-means$ is sensitive to initialization, so to avoid this pitfall we ran a cross validation simulation for 100 times for each k (starting from 48 nodes of the SOM going to 1 node with step of -1) for each with different random initializations. The best partitioning for each k was selected based on Davies-Bouldin cluster validity index[101], for minimizing the partitioning error, resulted in 5 clusters.

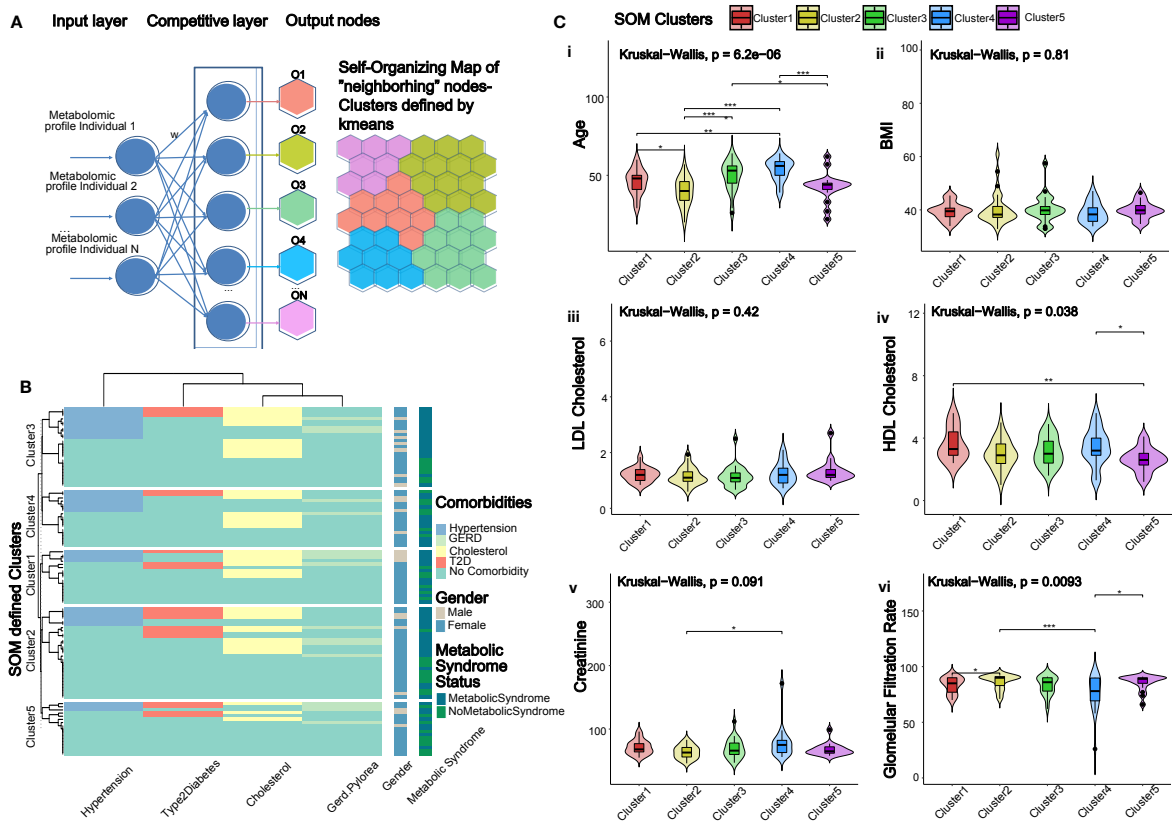

Figure 6: Self-organizing maps reveal five distinct metabotypes within BARIA cohort. (A) *Architecture of a competitive artificial neural network. Each individual's complete metabolomic profile is assigned a weight. The weights are in turn assigned to neurons in the competitive layer of the neural network. In the competitive layer, SOM algorithm calculates the similarity metric (here Euclidean distance) between each metabolomic profile and the neurons and then updates the weights. After training, the network assigns the individual's metabolomic profile to the "winner" output node, the node that is essentially more similar to the input metabolomic profile. Once this step is complete, all the nodes are comprising the SOM. Finally, all the nodes of the SOM are subjected to k-means clustering resulting in the partitioned topology, the metabotypes (SOM & k-means defined clusters).* ***(B)*** *Clustergram of hierarchical cluster analysis depicting the distribution of medically treated cardiometabolic comorbidities of the individuals in each of the metabotypes*

(SOM & k-means defined clusters). The treated comorbidities are: hypertension, T2D, GERD and cholesterol. In parallel columns are the gender and metabolic syndrome status of each individual, respectively. ***(C)*** *Clinical variables associated with obesity and their statistical significance across the metabotypes (SOM & k-means defined clusters): age* ***(C. i),*** *BMI* ***(C. ii)****, HDL cholesterol* ***(C. iii),*** *LDL cholesterol* ***(C. iv)****, creatinine and* ***(C. v)****, glomerular filtration rate* ***(c. vi)****; statistical significance among metabotypes is calculated with Kruskal-Wallis test; the symbols indicating significance among metabotypes are '*': P<=0.05, '**': P<=0.01, '***': P<=0.001.*

Significant differences among metabotypes in the multi-omics datasets

The metabolomics analysis revealed pronounced changes in lysophospholipids, phosphatidylcholines, dicarboxylate fatty acids, sphingomyelins, and branched-chain amino acid metabolites among the five different metabotypes, especially metabotypes 2 and 3 were most abundant in lipids (especially lysophospholipids and sphingomyelins) and amino acids (urea, arginine and proline metabolism). When the transcriptomics analysis results were subjected to gene enrichment analysis, they showed that KEGG metabolic pathways related to immune functions, fatty acid biosynthesis and elongation, protein-signaling and pathogenic pathways were regulated in different ways for each metabotype. Specifically, consistent upregulation of amino acid metabolic pathways was noted for metabotypes 4 and 5. In the gut microbial metagenome, the abundance of *Prevotella* and *Lactobacillus* species varied the most between the metabotypes, and metabotypes 4 and 5 had a lower abundance compared to metabotypes 2 and 3.

mixOmics: DIABLO

Data Integration Analysis for Biomarker discovery using Latent cOmponents (DIABLO)[103] performed supervised multi-omics data integration, by maximizing the correlation between co-expressed elements in the multi-omics datasets. DIABLO took as input all the differentially significant components from the omics datasets (289 metabolites, 119 microbial species and 776 genes) and produced as output a minimal signature of total 113 markers that distinguish the 5 metabotypes.

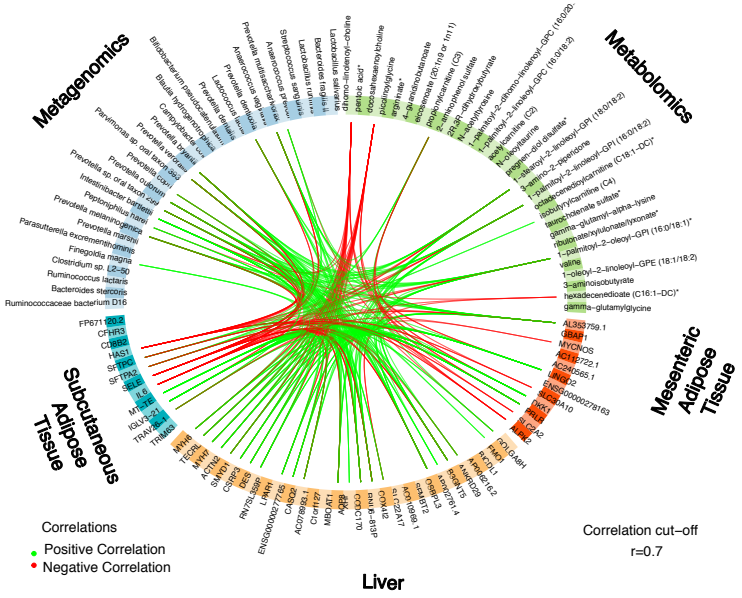

Figure 7: DIABLO analysis and correlations among multiple omics datasets for the five defined SOM clusters. *Circular correlation plot by DIABLO, for selecting top contributing components from each omics dataset (metabolites, genes, bacterial species). Correlation cut-off was r=0.7. The chosen elements constituted a highly correlated discriminatory signature for the five metabotypes. This signature involves a series of: i) Prevotella species (P. veroralis, P. copri, P. multisaccharivorax, P. oulorum, P. denticola, P. sp. oral taxon 299, P .bryantii, P. melaninogenica), Intestinibacter bartlettii, Anaerococcus prevotii; ii) lipid metabolites (especially phospatidylcholines); iii) liver genes enriched in oxidative phosphorylation, lipid metabolism and cardiomyopathy pathways; iv) subcutaneous adipose fat IL6 and SELE genes involved in inflammatory and immune system pathways; v) mesenteric adipose fat genes enriched in prolactine signaling, T2DM and PI3K-Akt signaling pathways.*

Bariatric surgery outcomes and metabotypes

For assessing the stratification into metabotypes, we evaluated their responses over time to bariatric surgery by post-operative biometric controls (weight, waist-upper leg circumference) at three time points: three months, six months and 12 months after surgery.

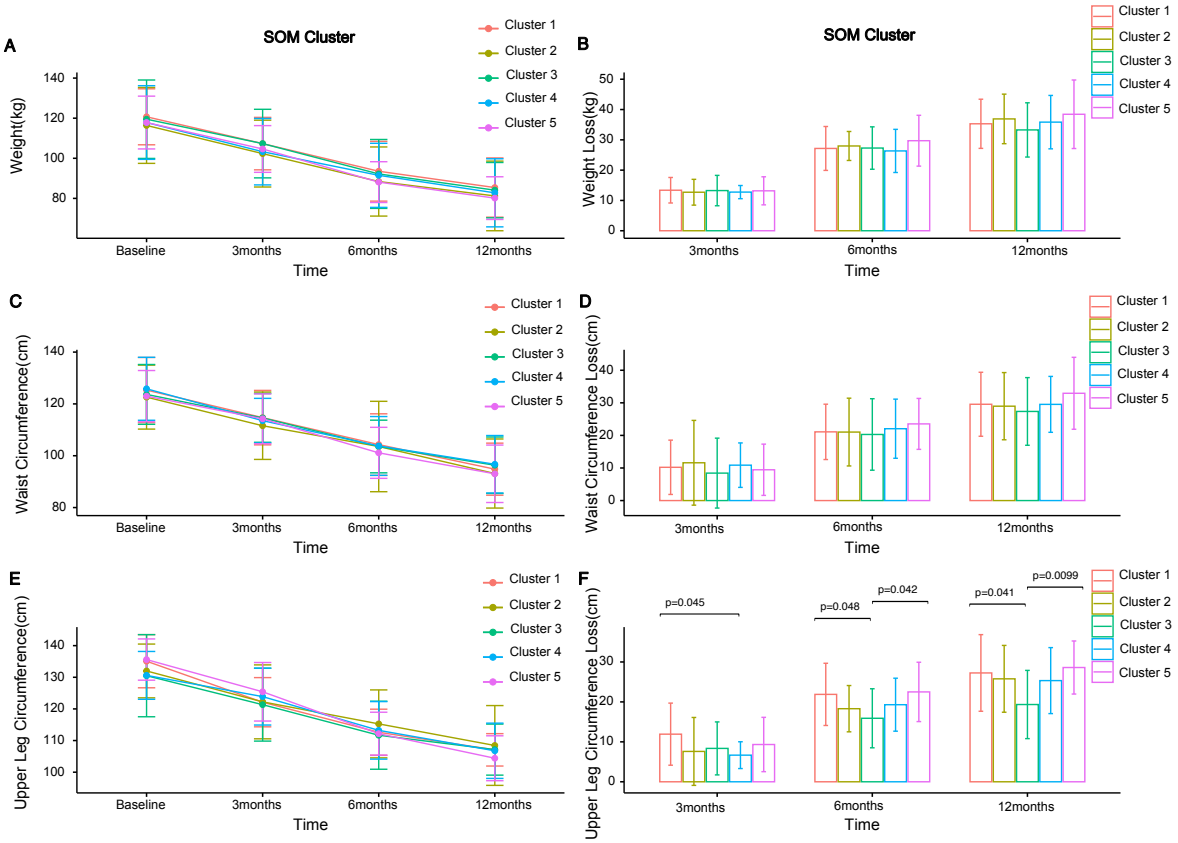

Figure 8: Weight and fat loss progression at distinct time points after bariatric surgery for the five defined SOM clusters (metabotypes). *(A) Weight (kg) of BARIA individuals at baseline, three months, six months and one year after bariatric surgery for each metabotype. (B) Weight loss(kg) of BARIA individuals at baseline, three months, six months and one year after bariatric surgery for each metabotype. (C) Waist circumference (cm) of BARIA individuals at baseline, three months, six months and one year after bariatric surgery for each metabotype. (D) Reduction of waist circumference(cm) of BARIA individuals at baseline, three months, six months and one year after bariatric surgery for each metabotype. (E) Upper leg circumference (cm) of BARIA individuals at baseline, three months, six months and one year after bariatric surgery for each metabotype. (F) Reduction of upper leg circumference(cm)of BARIA individuals at baseline, three months, six months and one year after bariatric surgery for each metabotype. Statistical significance among metabotypes is calculated with t-test and adjusted with FDR; the symbols indicating significance among metabotypes are '*': P<=0.05, '**': P<=0.01, '***': P<=0.001*

Metabotypes 2 and 5 have the highest weight loss one year post-operatively (35kg and 38 kg in average, respectively), metabotype 2 exhibits the largest waist circumference loss at three months after surgery (12cm) while metabotypes 1 and 5 are the best responders when it comes to upper leg circumference reduction, with the loss being consistent at all three time points. This trend is the same for weight loss, even if it was not confirmed by statistical significance testing.

Metabotyping challenges and significance

Metabotyping the BARIA individuals was done based on the results of unsupervised machine learning stratification. As a consequence, it was very difficult to directly compare the results with more traditional clinical classifiers. BARIA participants were

also stratified based on the presence or absence of metabolic syndrome and the same pipeline for analyzing the multi-omics datasets was applied. There were no notable statistically significant differences in the metabolome, transcriptome, weight and adiposity loss in none of the three time points for the metabolic syndrome classification. Our analysis showed that metabolic syndrome diagnosis can indeed capture a fraction of the microbial variability within obesity. Even so, our suggested metabotyping approach can identify more gut microbial species across the spectrum of obesity and its related comorbidities. When it comes to heterogenous cohorts, there is always the issue of the confounders effect, such as age, gender, multiple medication etc. In this analysis the metabotyping effect remained in both the metabolome and the gut microbial metagenome even after regressing out confounders like age and gender. Rather than traditional clinical disease classifiers, this grouping method may reduce the confounding effects of such clinical metadata[66,67].

ANNS in combination with a multi-omics integrative framework were able to effectively stratify this bariatric surgery complex cohort and reduce the original dimensionality of the BARIA dataset. The metabotypes were also associated with different responses in terms of weight loss and reduction of waist and upper leg circumference to bariatric surgery. The main advantage of this methodology is the SOMs and k-means topological projection is reusable for projecting new metabolomes, without further training of the map. Given the lack of an external independent validation cohort, we can utilize this map not only for comparing the metabolic distance to new subjects but also to post-operative metabolomes. This way we can identify the drivers of metabolism differences within this heterogeneous cohort. Moreover, this framework outperformed traditional clinical classifiers, such as metabolic syndrome, in capturing more metabolites, genes and gut microbial species that differed in the study population.

PART III: Dysmetabolic morbidities of obesity

Paper III: Obesity, T2D & GPCRs

One of the main comorbidities directly associated with obesity is T2D, where both diseases are linked with inflammation[104] and altered plasma levels of several metabolites[105]. Nevertheless, the molecular mechanisms behind these effects are not yet elucidated.

G-Protein Coupled Receptors (GPCRs) are receptors expressed in e.g. immune cells, adipocytes, and endocrine cells and are involved in functions such as lipolysis, gut hormone secretion, insulin secretion, and chemotaxis[106,107]. GPCRs have been associated with different levels of SCFAs and other lipid metabolites[108–110] and have exhibited pro- or anti-inflammatory functions[106,111,112]. GPR84 is a medium-chain fatty acid (MCFA) predominantly pro-inflammatory receptor involved in inflammatory gene expression, cytokine release, and neutrophil migration[46–48]. Within the context of this Thesis, 3-hydroxydecanoate is a metabolite that was demonstrated to increase inflammatory responses via GPR84 activation and neutrophil migration and may potentially contribute to the chronic inflammation observed in T2D.

Different levels of 3-hydroxydecanoate in T2D

By analyzing the fasting and two-hour post-prandial metabolome of a MMTT from 106 BARIA individuals, we identified the 3-hydroxydecanoate metabolite to be enriched in the circulation of obese individuals with T2D compared with non T2D controls. This difference is a strong indication of differential metabolism or absorption of 3-hydroxydecanoate in T2D.

For investigating whether 3-hydroxydecanoate levels in fasting and post-prandial states is regulated, an experimental setting involving mice was established. The mice were divided into two groups: conventionally raised (CONV-R) and germ free (GF). Mice were fasted overnight and then fed chow diet (not including 3-hydroxydecanoate). The levels of 3-hydroxydecanoate detected in the plasma serum of the mice were significantly reduced following the refeeding (Figure 9C), which could suggest that 3-hydroxydecanoate is released during fasting-induced β-oxidation[113]. GF mice had higher levels of 3-hydroxydecanoate when compared with CONV-R mice (Figure 9D), suggesting that 3-hydroxydecanoate is not produced by the microbiota, but potentially regulated by it.

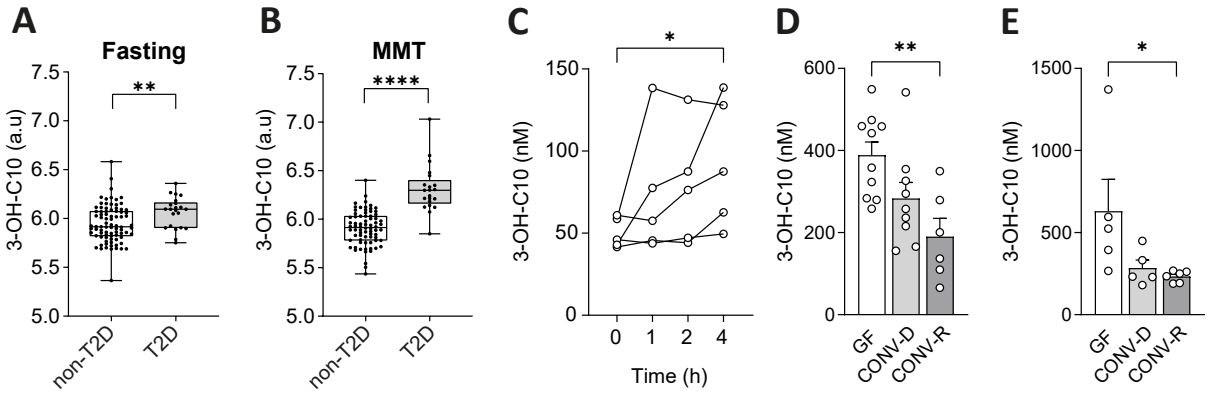

Figure 9: 3-hydroxydecanoate is enriched in obese patients with T2D and fasting induced in mice*. (A and B) Peripheral plasma levels of 3-hydroxydecanoate (3-OH-C10) in obese individuals with or without T2D at A) fasting and B) after a 2 h mixed meal test (MMT). The data are presented as boxplots where the box shows the 25th, median and 75th percentiles. The whiskers show minimum and maximum. (C) Plasma levels of 3-hydroxydecanoate in eye blood from CONV-R (N = 14) mice. The mice were fasted for 16 h followed by refeeding and sampling for 4 h. (D) Plasma levels of 3-hydroxydecanoate in eye blood from GF and CONV-R mice after fasting for 4 h (N = 6–10 mice/group). Data are shown as mean ± SEM (C and D). **p < 0.01, ***p < 0.001, ****p < 0.0001. p values were determined by two-tailed Mann-Whitney test (A, B, and D) or Friedman's test with Dunn's post-hoc analysis (C).*

3-hydroxydecanoate increases fasting insulin resistance and adipose tissue inflammation in mice

The next step was to assess if 3-hydroxydecanoate could affect glucose metabolism in mice. Mice were treated with daily intraperitoneal injections of 3-hydroxydecanoate for seven days. Fasting glucose did not differ between the 3-hydroxydecanoate treated and untreated mice groups. On the other hand, fasting insulin was significantly increased in the 3-hydroxydecanoate-treated group. Logically, 3-hydroxydecanate was next assessed for its effected in glucose tolerance via intraperitoneal glucose tolerance test, but no differences in glucose tolerance were noticeable. Based on this observation, 3-hydroxydecanoate may affect early stages of impaired glucose metabolism that has not yet developed into glucose intolerance.

It was also investigated if 3-hydroxydecanoate can modulate inflammation and affect inflammatory gene expression in adipose tissue and the liver. The assessment was done by looking into the expression of *Tnf*, *Ccl2*, *Il6* and *Il1b*. The expression levels for all genes were increased in adipose tissue, implying that 3-hydroxydecanoate increases inflammation there. The subsequent step was looking into whether 3-hydroxydecanoate could increase the accumulation of immune cells in the tissues via marker staining (macrophage marker CD68, neutrophil marker Ly6G). Apparently, 3-hydroyxdecanoate appears to increase infiltration of neutrophils and macrophages in adipose tissue. Transwell migration experiments (including primary murine immune

cells) showed that neutrophils migrated toward 3-hydroxydecanoate, but monocyte migration was not affected.

3-hydroxydecanoate signals through GPR84- $G\alpha_i$

Due to the immuno-metabolic response of 3-hydroxydecanoate in mice, it was crucial to identify potential receptors for further studying the mechanisms behind its effects. xCELLigence[114,115] label-free signaling assay was performed and the result showed that 3-hydroxydecanoate can indeed activate cellular responses. A series of experiments with pre-incubated cells with inhibitors for different GPCRs was conducted and the outcome suggested that 3-hydroxydecanoate signals through both $G\alpha_i$ and $G\alpha_q$ pathways. Next the PRESTO-Tango assay was used to verify potential GPCR targets for 3-hydroxydecanoate, where the metabolite was identified to be a GPR84 agonist. With the help of the inositol trisphosphate (IP3) accumulation assays it was possible to confirm that 3-hydroxydecanoate signals through GPR84- $G\alpha_i$.

Since the mechanisms relating 3-hydroxydecanoate to neutrophil migration and signaling through GPR84 and $G\alpha_i$ were established in animal models, we next investigated if 3-hydroxydecanoate could mediate migration of human peripheral neutrophils. A series of experiments with Transwell migration assays, pre-incubated neutrophils with or without PTX and human monocyte cell lines THP-1 could confirm that 3-hydroxydecanoate mediates neutrophil migration through GPR84- $G\alpha_i$. but no monocyte migration.

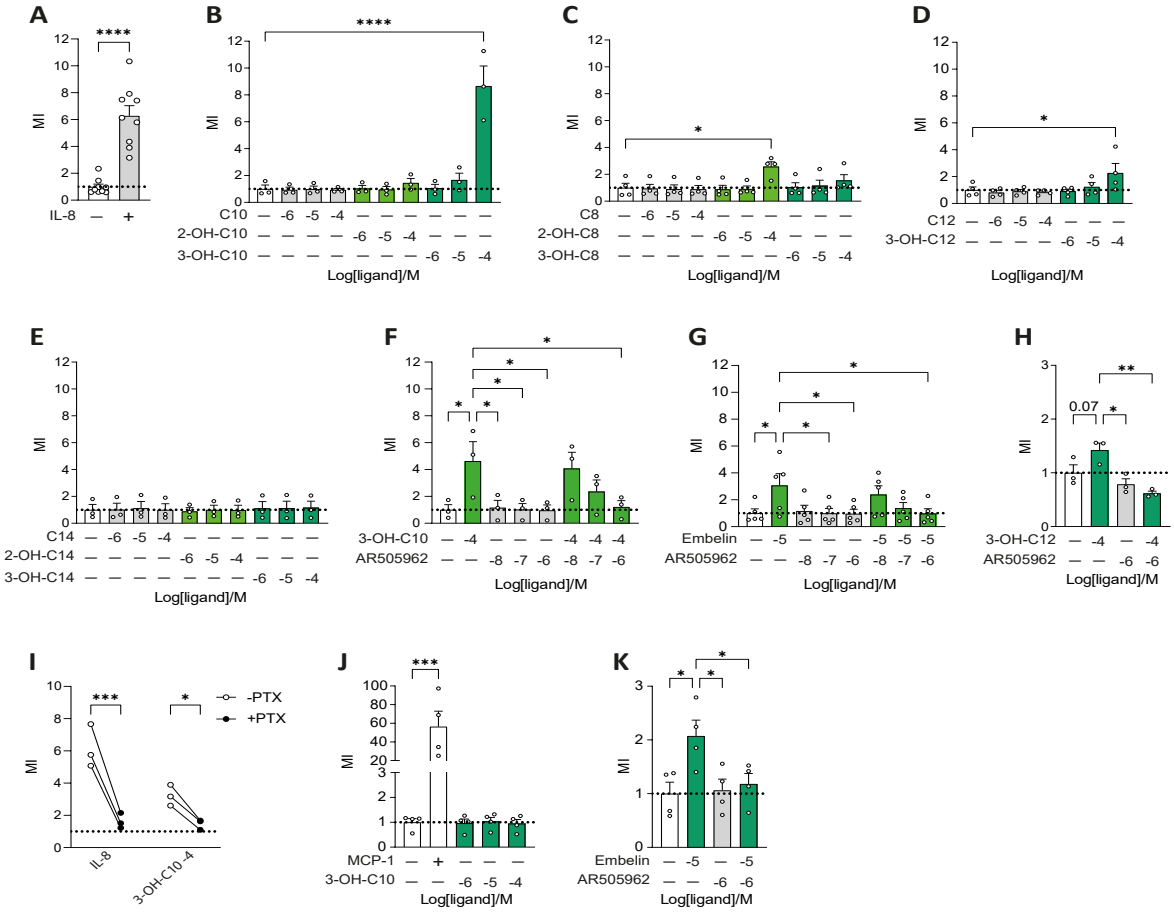

Figure 10: 3-Hydroxydecanoate mediates neutrophil migration through GPR84 and Gα$_i$. *(A) Migration of human primary neutrophils toward 10 ng/mL IL-8 or DMSO vehicle (N = 9). (B) Migration of human primary neutrophils toward 3-hydroxydecanoate (3-OH-C10), 2-hydroxydecanoate (2-OH-C10), and decanoate (C10) (N = 3). (C and D) Migration of human primary neutrophils toward (C) 3-hydroxydecanoate or (D) embelin in the absence and presence of increasing concentrations of the GPR84 antagonist AR505962 (N = 3–5). (E–G) Migration of human primary neutrophils toward (E) octanoate (C8), (F) laurate (C12), or (G) myristate (C14) as well as their 2- and 3-hydroxy derivatives (N = 3–4). (H) Migration toward 3-hydroxylaurate (3-OH-C12) in the absence or presence of 1 μM (–6) AR505962 (N = 3). (I) Paired comparison of MI toward IL-8 or 3-hydroxydecanoate after pre-incubation of human primary neutrophils with or without 200 ng/mL PTX for 16 h (N = 3). (J) Migration of the human monocyte cell line THP-1 toward 20 ng/mL MCP-1 or 3-hydroxydecanoate (N = 4). (K) Migration of THP-1 cells toward embelin in the absence or presence of the GPR84 antagonist AR505962 (N = 4). Data are shown as mean migration indexes (MI) ± SEM MI was calculated as the relative migration of samples compared to the vehicle control. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001. p values were determined by two-tailed Mann-Whitney test (A), 1-way ANOVA followed by Dunnett's multiple comparisons test (B–H and J–K), and two-way repeated measures ANOVA followed by Sidak's multiple comparison test (I).*

To summarize, we observed increased circulating levels of 3-hydroxydecanoate in individuals with T2D compared to obese controls without T2D in fasting and post-prandial metabolomics datasets. Yet, the source of 3-hydroxydecanoate, as well as its regulatory mechanisms in vivo is not yet established. The compound molecule of 3-hydroxydecanoate has been detected in dairy dietary sources[116]. Nonetheless experiments conducted in CONV-R and GF mice confirmed that 3-hydroxydecanoate

is present in plasma and liver but not in adipose tissue or their diet. This finding suggests that the metabolite might be produced during mitochondrial β-oxidation[113] in the liver. Furthermore, Gram-negative bacteria in the gut utilize 3-hydroxydecanoate for Lipopolysachcharide (LPS) biosynthesis[117]. Altogether, our data support that 3-hydroxydecanoate may be produced as an intermediate molecule in mitochondrial β-oxidation in the liver and that its bioavailability may be influenced by the gut microbiota. Our *in vivo* and *in vitro* experimental settings could confirm that administration of 3-hydroxydecanoate to mice could increase fasting serum insulin levels, not impair their glucose tolerance and most importantly could promote tissue inflammation and immune cell migration adipose tissue. Last but not least, with a series of assays the mechanism for molecular signal transduction of 3-hydroxydecanoate mediated via GPR84- $G\alpha_i$ was spotted, which is of great relevance for potential drug design since targeting specific signaling pathway(s) allows for more precise cellular responses and reduces the risk of unwanted side effects.

Paper IV: NAFLD

NAFLD prevalence and connection to obesity

NAFLD is now recognized as the most prevalent chronic liver disease worldwide and this prevalence increases to over 80% in individuals with obesity[18]. NAFLD comprises a range of clinical and histopathological abnormalities, caused by substantial fat accumulation in the liver. This fat build-up has been characterised as relatively benign, however, an estimated 30% of people with NAFL will develop NASH[118], a progressive form of liver disease that leads to fibrosis, cirrhosis, and hepatocellular carcinoma, that might even require liver transplantation. Women tend to have a lower risk of developing NAFLD, but once the disease is established, women have a higher risk of disease progression[119].NAFLD prevalence is growing, and the lack of effective treatment options could increase the obesity-related burden on public health and economies. Therefore, it is crucial to develop appropriate, sex-specific, non-invasive diagnostic methods and treatment options for this disease. Here, we used a systems biology approach to identify factors that may contribute to NAFL development by deep phenotyping of 55 women, all participants in the BARIA study. We analysed 6 omic datasets including faecal metagenomics, plasma metabolomics, and liver, subcutaneous, and mesenteric adipose tissue transcriptomics, where the only difference between the subjects was the presence/absence of hepatic steatosis. Omics datasets included fasting and 2-hour MMTT plasma metabolome, liver and adipose tissue (subcutaneous and mesenteric) transcriptome, along with gut microbial metagenome. All the omics collected, extracted and analysed as described in **Part I** of the Thesis. For investigating differences in glucose metabolism between women with and without NAFL, glucose and insulin responses during a MMTT prior to bariatric surgery and one-year post-intervention were also analysed.

Significant differences in omic datasets between women diagnosed with and without NAFL

Microbial alpha diversity between women with and without NAFL was similar in the two groups, albeit large interindividual variation in the gut microniome composition was observed. The microbiome was dominated by Firmicutes in NAFL diagnosed BARIA individuals, while Bacteroidetes was the most dominant phylum in individuals without NAFL (Figure 11A). 57 bacterial species were differentially significant among individuals with and without NAFL (Figure 11A). Three bacterial species were at least twice as abundant in individuals with NAFL (*Collinsela stercoris, Lactobacillus buchneri, Lactobacillus iners*). In individuals without NAFL, 11 bacterial species were at least twice as abundant compared to individuals with NAFL (*Prevotella oulorum, Prevotella sp. oral taxon 317, Prevotella sp. Oral taxon 472, Prevotella multisaccharivorax, Prevotella dentalis, and Prevotella bryanti, Lactobacillus delbrueckii, Enterococcus casseliflavus, Citrobacter rodentium, Yersinia enterocolitica*, and *Haemophilus pittmaniae*).

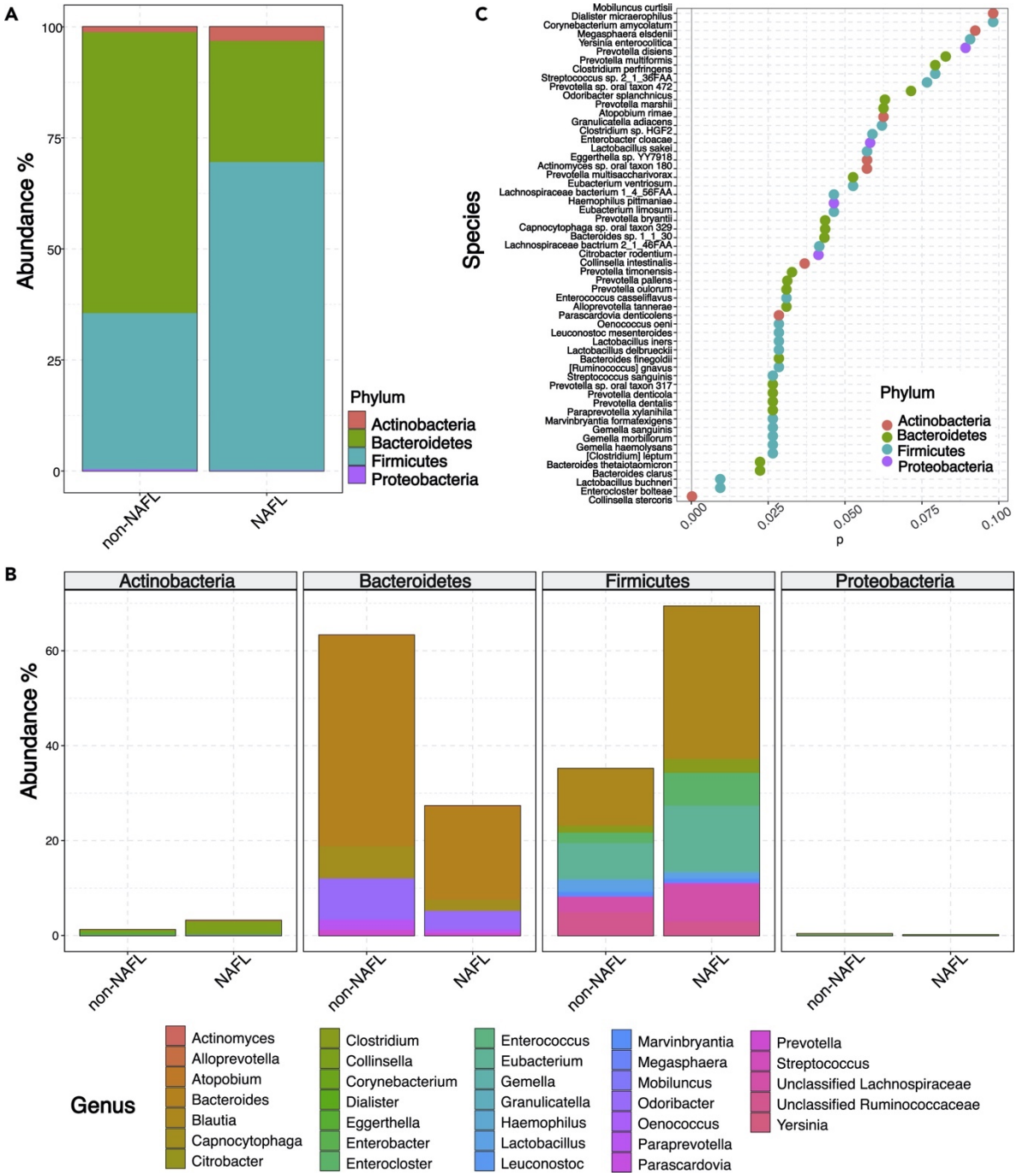

Figure 11: Microbial species and phyla between individuals with and without NAFL*. (A) Difference in total abundance of bacterial species indicated at the phylum level between individuals with and without NAFL. (B) Relative abundance and distribution within of differentially significant microbial species between individuals with and without NAFL. (C) 57 differentially significant microbial species between individuals with and without NAFL, after differential microbial species analysis with DESeq2 (adjusted $p < 0.1$) Likelihood Ratio Test for significance.*

Phosphathidylcholine 1-palmitoyl-2-arachidonoyl-GPC (16:0/20:4n6) was lower in women with NAFL and was the only differentially significant metabolite in the fasting metabolome dataset. In the post-prandial metabolome two sphingomyelin metabolites were decreased in individuals with NAFL whereas diacylglycerol, a signaling lipid

previously linked to hepatic insulin resistance and NAFLD[120] was more abundant the NAFL diagnosed women. 1-carboxyethylisoleucine and 1-carboxyethylvaline, both Branched-Chained Amino Acids (BCAAs) derivative metabolites, were increased in individual with NAFL. Since BCAAS are associated with insulin resistance[121] we looked into the MMTT results. Insulin and glucose levels did not differ during the MMTT, suggesting that these alterations may be independent of altered glucose metabolism.

In the transcriptomics datasets, KEGG pathway enrichment analysis using EnrichR, pinpointed to specific pathways that differed among individuals with and without NAFL. More specific:

- In the hepatic transcriptome: Pathways involved in cancer proliferation, the hypoxia-inducible factor 1 (HIF-1) signaling pathway were enriched in individuals with NAFL while the only significant pathway that was enriched in individuals without NAFL was the pathway involved in arginine and proline metabolism
- In the mesenteric adipose transcriptome: carbohydrate, galactose, sucrose, pathways associated with fat digestion and absorption and protein metabolism pathways were enriched in mesenteric adipose tissue from individuals with NAFL, while pathways involved in infectious disease were not enriched
- In the subcutaneous adipose transcriptome interleukin (IL)-17, advanced glycation end products (AGE), tumor necrosis factor (TNF), signaling pathways were enriched in individuals with NAFL whereas response to oxidative stress was not enriched

NAFL multi-omics signature

The DIABLO computational framework used here for integrating various omics datasets successfully identified a highly correlated discriminatory signature for NAFL consisting of BCAA metabolites, diacylglycerol, liver genes involved in HIF-1 signalling, mesenteric adipose tissue genes involved in fat metabolism and subcutaneous adipose tissue genes that are part of mitochondrial translation/elongation.

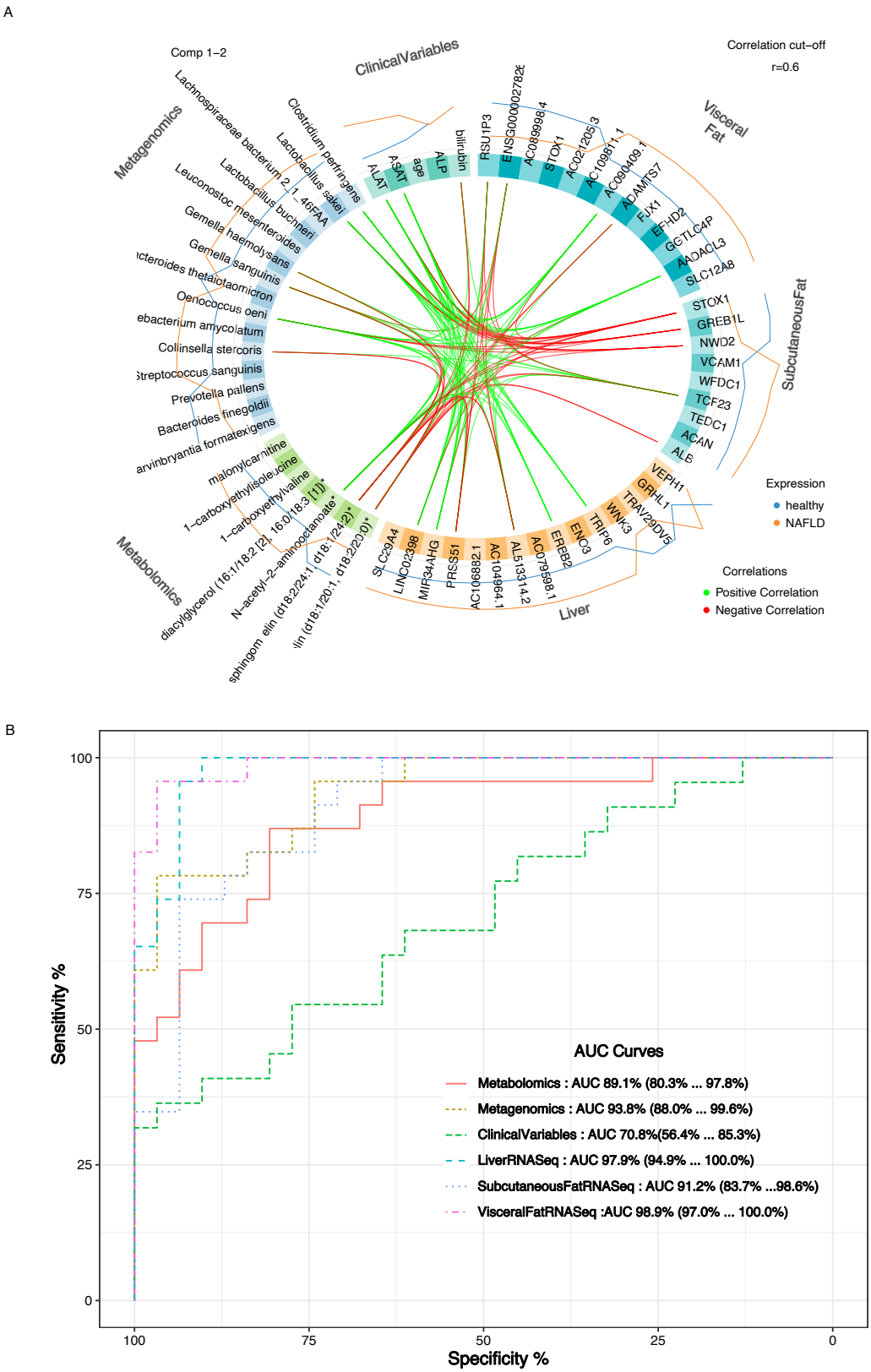

Figure 12: DIABLO analysis and AUC predictive capacity for NAFL given each omic dataset.

For evaluating the DIABLO minimal biomarker signature for NAFL, we constructed a series of generalized linear models (GLMs) aimed to investigate whether this signature could outperform the clinical variables capacity to correctly predict NAFL. All the transcriptomics datasets and the chosen genes can very accurately predict NAFL, whereas the metabolome and metagenome datasets outperform the traditional clinical variables in NAFL predictive capacity, with an Area under the Curve (AUC) AUC = 89.1% and 93.8%, respectively, versus AUC = 70.8%.

The MMTT performed one year after weight-loss surgery revealed a clear difference in glucose and insulin response between individuals with and without NAFL, contrary to baseline findings. This outcome suggests that whole body metabolism is indeed different in this early phase of NAFL and is affected by the massive weight loss after bariatric surgery.

Analysis of the gut microbiome showed that subjects with NAFL have a Firmicutes dominated microbiome, which is in line with previous literature[122,123]. However, they contradict recent reports where liver steatosis was anticorrelated with Firmicutes[124]. This might entail that there is not one unique microbiome signature for NAFLD given the multiple confounding factors such as age, sex, and disease state[125]. The subtle changes in the plasma post-prandial metabolome could emphasize that early changes in metabolism are more pronounced post meal than in fasting conditions. Since the composition of BCAA in cardiometabolic disease patients are often different, this attribute can be explained by insulin resistance in the liver or muscle[126]. Yet, insulin and glucose levels did not differ during the MMTT, so potentially these changes are independent of insulin resistance and could have been originated elsewhere, potentially the gut microbiome[127]. The analysis of the transcriptomics datasets is in line with the current concept that adipocyte dysfunction plays an important role in the pathophysiology of NAFLD[128,129]. Adipose tissue expansion (both subcutaneous and visceral) can lead to hypoxia-induced hypersecretion of adipocytokines, such as TNF and interleukin IL6, by the adipocytes as well as by the inflammatory immune cells that accumulate in adipose tissue of obese individuals [128,130]. These mediators, together with increased levels of lipid metabolites such as diacylglycerols observed during metabolic dysregulation, when they reach the liver via portal vein, can contribute to the development and progression of NAFLD[129]. The KEGG pathway enrichment of the differential significant genes of both mesenteric and subcutaneous adipose tissue revealed that pathways involved in fat and glucose metabolism and TNF signalling were upregulated in NAFL, respectively thus highlighting the potential role of the adipose tissue in the development of NAFLD.

The associative results from DIABLO omics integrative analysis stress the interrelation between metabolites, bacterial species, and genes and can be used to generate hypothesis to further study the pathophysiology of NAFL in humans.

Paper V: Gallstones formation after bariatric surgery

Gallstones: formation and relation to bariatric surgery

Decreased secretion of bile acids, hypersecretion of cholesterol, rapid phase transitions of cholesterol in bile leading to the precipitation of cholesterol crystals, and impaired gallbladder motility with hypersecretion of mucus are some of the main factors leading to gallstones formation[131]. The gut microbiome can influence bile acid metabolism conversion of primary bile acids into secondary bile acids in the gut and has therefore been suggested as a potential driver of cholesterol gallstone formation[84,132,133].

The rapid weight loss following bariatric surgery has been pinpointed as a risk factor for gallstone formation in about one-third of patients after bariatric surgery[37,134,135] and approx. 8–15% of these patients require cholecystectomy. The relationship between bariatric surgery and gallstone disease is yet to be investigated in depth. Therefore, 106 individuals from the BARIA cohort were recruited.

Gallstones and BARIA

Out of these 106 inclusions, 88 BARIA individuals formed gallstones post-surgically. 32 individuals (36.4%) had gallstones or sludge after surgery and 11 individuals underwent a cholecystectomy for symptoms of gallstone disease after a mean of 9.6 ± 4.8 months after bariatric surgery (range 5–20 months after surgery). The ultrasound showed gallstones in three of these patients. Among the remaining patients, three had sludge and 18 patients had gallstones on ultrasounds performed at one year after surgery in two patients and at two years in 19 patients. This study is the first to relate differences in metabolic activity of subcutaneous and visceral adipose tissue to the presence of gallstones in patients after bariatric surgery. Of the 88 included patients, 56 did not have gallstones at follow-up 1 ($n = 2$) or 2 ($n = 54$) years after bariatric surgery. Fecal microbiome analysis in these patients revealed species that might act protective against gallstone development. On the other hand, transcriptomic analysis of adipose tissue showed that altered lipid (cholesterol) metabolism might contribute to gallstone development after bariatric surgery. Moreover, several sulfated bile acids were higher concentrated in patients with gallstones.

Significant differences in omic datasets between BARIA individuals that did and did not develop gallstones after bariatric surgery

The multi omics datasets employed for this study included gut microbial metagenome, liver, mesenteric adipose and subcutaneous adipose tissue transcriptome as well as fasting metabolome from the BARIA individuals. All omics extractions were conducted as described in Part I of the Thesis. For the transcriptome datasets the paired-end reads first trimmed and cropped using trimmomatic[136], version 0.38 with the following settings: HEADCROP: 6, SLIDINGWINDOW: 4:15, and MINLEN: 50. The resulting read sets were then mapped using kallisto[137], version 0.46.0 against the GRCh38

assembly of the human genome with sequence bias correction, 100 bootstrap samples (options --bias, -b 100, and --rf-stranded). Downstream differential significance analysis for both the transcriptomics and microbiome data and differential abundance analyses were performed using the the DESeq2 with Wald significance testing and parametric fitting. Specifically, for the gut microbiome Significance levels were calculated using the adonis function from the vegan[138] R package, version 2.5-7 and were adjusted for sex .For the metabolomics data, differences in metabolite concentrations between the two participant groups were calculated using the Wilcoxon signed-rank test, of which P values were adjusted for multiple testing using the Benjamini-Hochberg[139] procedure.

There were no significant differences observed with alpha and beta diversity metrics in the gut microbial metagenome between the individuals that did and did not develop gallstones after bariatric surgery. However, differential abundance analyses at species level (with DeSeq2 with independent hypothesis weighting) revealed 41 bacterial species that were significantly differently abundant between groups (adjusted $P \leq 0.05$ and log2 fold change ≤ -1 or ≥ 1). *Bacteroides intestinalis*, *Finegoldia magna*, *Ruminococcus gnavus*, and *Prevotella buccalis* were more abundant in patients with gallstones, than in patients without gallstones. In patients without gallstones, higher abundance of 37 bacterial species was observed, of which the majority were members of the Lactobacillaceae (12 species) and Enterobacteriaceae (7 species), as illustrated in Figure 13B.

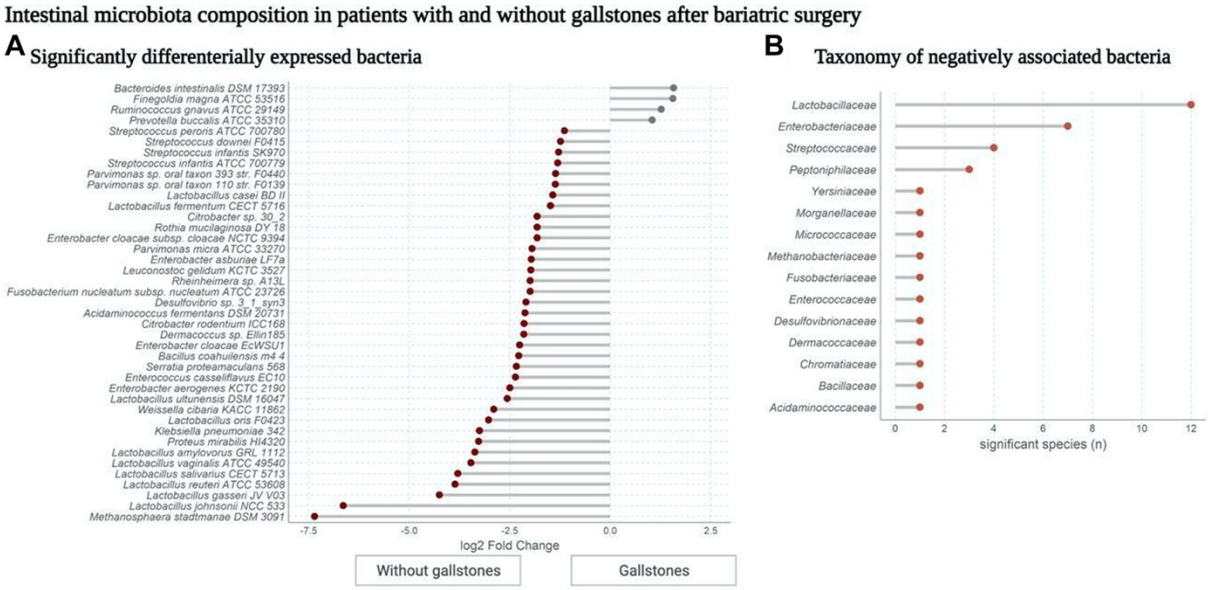

Figure 13: 1Intestinal microbiota composition in patients with and without gallstones after bariatric surgery. *A: Significantly differentially expressed bacteria. The first four species are more abundant in patients with gallstones (gray). The following 37 species were more abundant in patients without gallstones (red). B: Taxonomy of negatively associated bacteria. For example, of the 37 species that were more abundant in patients without gallstones, 12 were members of the Lactobacillaceae and 7 belonged to the Enterobacteriaceae.*

Transcriptomic analysis of liver tissue revealed a significant increased expression of four genes in patients with gallstones compared with patients without gallstones: *TEX14*, *MPPED1*, *GREB1*, and *AC005666.1.* that are involved in different pathways regulating cell division. In subcutaneous adipose tissue nine differentially significant genes were upregulated in patients with gallstones (*ALB*, *APOA1*, *TAT*, *TRPV5*, *CYP4F2*, *CTSE*, *HMGCS2*, *MOGAT2*, and *ALDOB*) and four in patients without gallstones (*DRP2*, *MT1A*, *SFRP5*, and *ANGPTL7*). Finally, in visceral adipose tissue, two genes were significantly more often expressed in patients with gallstones.

Secondary bile acids metabolites showed higher concentrations among in patients with gallstones. When comparing to patients without gallstones, the bile acids glycochenodeoxycholate 3-sulfate, glycochenodeoxycholate glucuronide, glycocholate, glycodeoxycholate 3-sulfate, glycohyocholate, glycolithocholate sulfate, taurochenodeoxycholic acid 3-sulfate, and taurolithocholate 3-sulfate were increased, as seen in Figure 14:

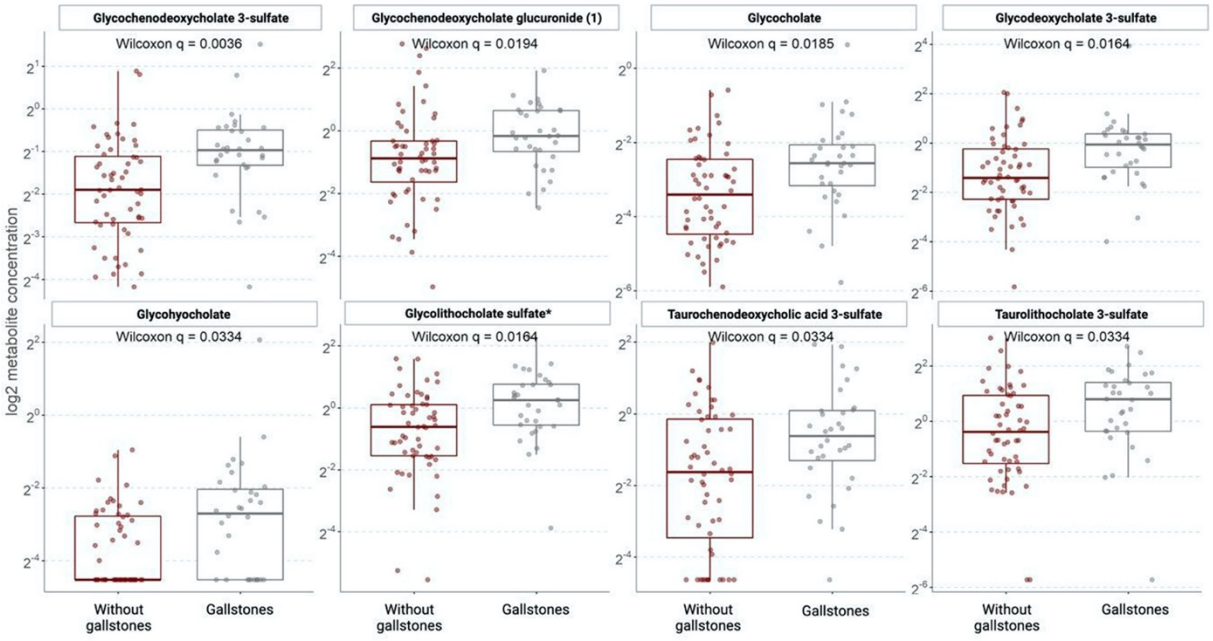

Figure 14: 3Plasma metabolites with different concentrations between patients with and without gallstones. *Bile acids are increased in patients with gallstones after* P *value adjustment per subpathway.*

The higher abundance of *Ruminococcus gnavus* in patients with gallstones after bariatric surgery, is in accordance with recent studies that identified it as a biomarker for gallstones[140]. Exogenous alcohol consumption has been associated with a decreased risk of gallstone formation[141], so the abundance of *Klebsiella pneumoniae* (Enterobacteriaceae) and Lactobacillaceae, that are able to produce

microbial ethanol[142,143] is a possible protection factor against gallstones. Bile salt hydrolase (BSH) plays a role in bile acid-mediated signaling pathways, which regulate lipid absorption, glucose metabolism, and energy homeostasis and can be produced by Lactobacillaceae species[144], which are therefore studied as possible cholesterol-lowering probiotics[145,146]. Accumulation of adiposity tissue is a known risk factor for gallstone formation in both men and women[147,148]. In visceral and subcutaneous adipose tissue of patients with gallstones, we identified metabolic pathways involved in inflammatory response and lipid metabolism, including cholesterol and fatty acid metabolism. Up to 50% of cholesterol in obese patients is stored as free cholesterol in the adipose tissue, which is the state of cholesterol when excreted via bile[149–151]. In the rapid weight loss phase following bariatric surgery, adipose tissue mass is reduced, possibly resulting in the release of a large amount of free cholesterol to the liver which can result in gallstone formation. So increased gene expression of genes involved in tissue regulation in patients without gallstones might be indicative of more adaptive tissue state, acting as a protective mechanism against gallstone formation during this post-operative weight loss stage. Also in line with previous studies, increased plasma levels of conjugated bile acids were observed in patients with gallstones, probably due to lower excretion of bile acids into the gallbladder[152,153].

Conclusively for the first time, differences in metabolic activity of subcutaneous and visceral adipose tissue were associated with the presence of gallstones in patients after bariatric surgery. This analysis revealed gut microbial species that might act protective against gallstone development. Alterations in cholesterol metabolism could affect gallstone development after bariatric surgery, whilst several sulfated bile acids were higher concentrated in patients with gallstones.

Summary and Conclusions

The contributions of this thesis can be summarized in metabotypes with different responses to bariatric surgery, the establishment of metabolites' affinity mechanisms to GPCR receptors and minimal set of evasive and non-evasive biomarkers for NAFL and gallstones formation after bariatric surgery.

The BARIA study (**Paper I**), as a tool for investigating into an obese population, can add more understanding in how specific molecule compounds and gut-microbiota affect metabolism, glucose metabolism and NAFLD. The observed differences in the BARIA participants glucose and lipidemic status indeed establishes the heterogeneity within the bariatric surgery inclusions. The intra-reproducibility of MMTT sets up its importance and as another investigative tool.

The data-driven methodology of using BARIA cohort as an anchor for combining SOMs, metabotyping and DIABLO correlation analysis (**Paper II**) enabled the identification of an underlying common yet discriminatory minimal multi-omics signature among obese individuals, that could lead to predictive markers of the bariatric surgery outcome. This approach can also constitute a valuable tool for studying other multifaceted metabolic disorders.

This association framework can be the starting point for selecting candidate compounds, such as metabolites for a more thorough examination and provide mechanistic insight into the causality of pathogenicity originating in the tissues, mediated by bacteria and materializing via metabolites and clinical metadata. Through this framework, it has been possible to directly link increased levels of a MCFA metabolite (3-hydroxydecanoate) to T2D and through detailed experimental settings to establish the GPR84 receptor activation and neutrophil migration molecular mechanisms behind this (**Paper III**).

Within the contexts of NAFLD, this computational approach suggested that there is substantial crosstalk between omics sets and that in early stages of the disease, adipocyte dysfunction is the predominant factor in disease development followed by gut microbial composition and plasma metabolites (**Paper IV**).

When it comes to post-operative complications of obesity, such as gallstones formation, the present study observed that higher abundance of specific gut microbial species in patients without gallstones prior to bariatric surgery could possibly act protectively, whereas specific lipid, inflammation, cholesterol pathways in adipose tissue, can lead in gallstone formation after bariatric surgery (**Paper V**).

Future Perspectives

It is evident that there is this ingoing trend to phenotype in detail complex populations with respect to multigenic diseases and therefore there is a need for going beyond the traditional methods of statistical analysis and binary clinical disease classifiers, to be able to effectively phenotype these patients. The volume of data has massively increased leading to a new scientific field of "Big-Data", hence the complexity and dependencies among data grow exponentially as well. Machine learning and integrative techniques are on the rise and are the only way to deconvolute the complex signals in heterogenous populations. This deconvolution is the basis for selecting out key components, such as metabolites, genes and microbial species for proceeding into detailed research and experimental setting aiming to identify pathophysiological mechanisms and then selecting targeted interventions, thus further contributing to the personalized medicine approach.

The BARIA cohort is an excellent resource for investigating obesity and its associated co-morbidities. The materials included in the papers that comprise this thesis come from the first ~100 inclusion of the cohort. Currently, more than 300 bariatric surgeries have been performed and the respective omic datasets have been collected, whereas the metabolome, transcriptome and gut microbial metagenome have been extracted and are being analysed. These new inclusions can act as a validation cohort, given the uniqueness of BARIA in multi-omics and detailed follow-up. Hence, the hypotheses, initial assumptions and preliminary outcomes from these 5 Papers of this Thesis can be extended and validated further. Identifying key factors and biomarkers for obesity and each associated comorbidity may indeed provide diagnostic and therapeutic leads to control epidemic. We might be able to predict more accurately biometric features after bariatric surgery (weight loss) and broaden our prognostic capability in the status of associated comorbidities, such as the post-surgical resolution of T2D, NAFLD and the prevention of gallstone formation. We can implement the systems biology framework in combination with into other cohorts for stratifying inclusions and then overlay the related comorbidities for deepening our knowledge of these medical conditions.

The results and methods presented within this thesis illustrate how data-driven approaches can facilitate early diagnosis and enable potential preventive actions through the study of heterogeneous cohorts.

Acknowledgements

Throughout my 8 years in Sysbio, both during my time as a Research Engineer and as a PhD candidate, I have had an amazing support network. In this small section I will try to express my gratitude to everyone that has joined me along the way. Firstly, I would like to thank my supervisor, Jens, for inviting me to be a part of the Sysbio family. I have to thank you for sharing your professional ethics and for allowing me to grow and witness the way you conduct your work. As I have said multiple times "Jens is a school in his own right". Thanks also to my co-supervisors Fredrik Bäckhed and Aleksej Zelezniak, for lengthy discussions, helpful suggestions and thorough reviews. To my former line-manager Shaq: a very big thank you. Without you I would have never arrived in Gothenburg and joined the Sysbio family.

Thanks to all members of Gut-MMM consortium for assembling the patients, coordinating the cohort and generating almost all of the data I have analyzed. Thanks to the PIs: Max Nieuwdorp and Thue Schwartz and the collaborators Siv, Randi, Louise, Valentina, Lisa, Kimberly, Annika, Viktor, Bert, Ömrüm, Yair and Maurits. Thanks to Boyang Ji for his role in the reviews and article submissions. To my direct collaborators in the BARIA cohort Stijn Meijnikman and Peishun Li: thank you for all the discussions, strategies, drafts and valuable technical and clinical input.

Thanks to all of the SysBio group (researchers and admin): Martina, Florian, Anastasia, Ed, Mikael, Joakim, Verena, Xin, Shantosh, Shadi, Lei, Hao, Oliver, Fariba, Mihai, Andrea, Iván, Cheewin, Angelo, Dany, Max, Patrick, Marta, Sandra, Yanyan, Le, Simone, Nils, Christos, Angelica, Abder, Emelie, Thomas, Erica, Gunilla, Anne-Lise plus all the new members that I haven't had the joy to meet and work with yet. Thanks to all that I have encountered during my time here: Sakda, Julia, Francesco, Petri, Leif, Michi, Benjamin, Kathi, Avlant, Promi, Parizad, Raphael, Simonas, Manish, Kate, Veronica, Anna, Doris, Honhzhong, Feiran, Yun, Carl, Johan B & Johan G., Amir, Pouyan, Shaeed, Olena, Ievgeniia, Sylvain, Jens, Marie, Ximena, Marcus, Tatiana, Leonie, Francesca, Christoph, Paulo, Piérre. Thank you for your sharing your scientific knowledge, your computational know-how, your every-day tips and tricks, your culture, your thoughts on everything in life and most of all your friendship. It truly has really been a privilege getting to know you getting to work with you, therefore I leave Sysbio grateful for all our joint experiences.

The final words are for my family. To my parents: Thank you for all your personal sacrifice and for choosing to invest in our education. Without your support through the years my journey in academia would not have been possible. This thesis is dedicated to the family we created here in Gothenburg. Giorgos, without you this thesis would not have been materialized. Thanks for your unconditional support, understanding, insights, tolerance and advice. Thank you for being present in every step of the way and for planning our future ahead. Last but not least, to Aris and our little one on the way, thank you for making our days brighter and for filling us with joy.

References

1. WHO. Obesity and overweight [Internet]. 2021 [cited 2020 Jul 29]. Available from: https://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight

2. WHO. Obesity: preventing and managing the global epidemic.WHO Technical Report Series number 894. Geneva; 2000.

3. Bentham J, Di Cesare M, Bilano V, Bixby H, Zhou B, Stevens GA, et al. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. Lancet. 2017 Dec 16;390(10113):2627–42.

4. Haslam DW, James WPT. Obesity. Lancet. 2005 Oct 1;366(9492):1197–209.

5. Engin A. The Definition and Prevalence of Obesity and Metabolic Syndrome. In: Engin AB, Engin A, editors. Obesity and Lipotoxicity. Cham: Springer International Publishing; 2017. p. 1–17.

6. CDC. Defining Adult Overweight & Obesity | Overweight & Obesity | CDC [Internet]. [cited 2023 Feb 7]. Available from: https://www.cdc.gov/obesity/basics/adult-defining.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fobesity%2Fadult%2Fdefining.html

7. Seidell JC, Flegal KM. Assessing obesity: classification and epidemiology. Britijh Med Bull. 1997;53(2):23–252.

8. Zhang Y, Proenca R, Maffei M, Barone M, Leopold L, Friedman JM. Positional cloning of the mouse obese gene and its human homologue. Nature. 1995;374(6521):479.

9. James J. Fat and water 1H MRI to investigate effects of leptin in obese mice. Obesity. 2009;17:2089–93.

10. Rosen T, Bosaeus I, Tolli J, Lindstedt G, Bengtsson BA. Increased body fat mass and decreased extracellular fluid volume in adults with growth hormone deficiency. Clin Endocrinol (Oxf). 1993 Jan 1;38(1):63–71.

11. Luppino FS, De Wit LM, Bouvy PF, Stijnen T, Cuijpers P, Penninx BWJH, et al. Overweight, Obesity, and Depression: A Systematic Review and Meta-analysis of Longitudinal Studies. Arch Gen Psychiatry. 2010 Mar 1;67(3):220–9.

12. Pearson ER. Type 2 diabetes: a multifaceted disease. Diabetologia. 2019 Jul 1;62(7):1107–12.

13. Symptoms & Causes of Diabetes | NIDDK [Internet]. [cited 2023 Feb 7]. Available from: https://www.niddk.nih.gov/health-information/diabetes/overview/symptoms-causes?dkrd=hispt0015

14. Maruthur NM, Tseng E, Hutfless S, Wilson LM, Suarez-Cuervo C, Berger Z, et al. Diabetes medications as monotherapy or metformin-based combination therapy for type 2 diabetes: A systematic review and meta-analysis. Ann Intern Med. 2016 Jun 7;164(11):740–51.

15. Saenz A, Fernandez-Esteban I, Mataix A, Segura MA, Roqué i Figuls M, Moher D. Metformin monotherapy for type 2 diabetes mellitus. Cochrane Database Syst Rev. 2005 Jul 20;2005(3).

16. Ishikawa K, Tsunekawa S, Ikeniwa M, Izumoto T, Iida A, Ogata H, et al. Long-term pancreatic beta cell exposure to high levels of glucose but not palmitate induces DNA methylation within the insulin gene promoter and represses transcriptional activity. PLoS One. 2015;10(2):1–19.

17. Krentz AJ, Bailey CJ. Oral antidiabetic agents: Current role in type 2 diabetes mellitus. Drugs. 2005 Sep 19;65(3):385–411.

18. Younossi Z, Tacke F, Arrese M, Sharma BC, Mostafa I, Bugianesi E, et al. Global Perspectives on Non-alcoholic Fatty Liver Disease and Non-alcoholic Steatohepatitis. Hepatology. 2019 Sep;69(6):2672–82.

19. Marjot T, Moolla A, Cobbold JF, Hodson L, Tomlinson JW. Nonalcoholic Fatty Liver Disease in Adults: Current Concepts in Etiology, Outcomes, and Management. Endocr Rev. 2020;41(1):bnz009.

20. Younossi ZM. Non-alcoholic fatty liver disease – A global public health perspective. J Hepatol. 2019 Mar 1;70(3):531–44.

21. RACGP - Fatty liver disease – a practical guide for GPs [Internet]. [cited 2023 Feb 7]. Available from: https://www.racgp.org.au/afp/2013/july/fatty-liver-disease/

22. Tilg H, Moschen AR, Roden M. NAFLD and diabetes mellitus. Nat Rev Gastroenterol Hepatol 2016 141. 2016 Oct 12;14(1):32–42.

23. Liu XZ, Pedersen L, Halberg N. Cellular mechanisms linking cancers to obesity. Cell Stress. 2021;5(5):55–72.

24. Islami F, Goding Sauer A, Gapstur SM, Jemal A. Proportion of Cancer Cases Attributable to Excess Body Weight by US State, 2011-2015. JAMA Oncol. 2019 Mar 1;5(3):384–92.

25. Sung H, Siegel RL, Torre LA, Pearson-Stuttard J, Islami F, Fedewa SA, et al. Global patterns in excess body weight and the associated cancer burden. CA Cancer J Clin. 2019 Dec 12;69(2).

26. WHO. Obesity and overweight [Internet]. 2023 [cited 2023 Feb 5]. Available from: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

27. Bray GA. Use and abuse of appetite-suppressant drugs in the treatment of obesity. Ann Intern Med. 1993;119(7 II):707–13.

28. Shah M, Vella A. Effects of GLP-1 on appetite and weight. Rev Endocr Metab Disord. 2014 May 9;15(3):181–7.

29. Maciejewski ML, Arterburn DE, Van Scoyoc L, Smith VA, Yancy WS, Weidenbacher HJ, et al. Bariatric Surgery and Long-term Durability of Weight Loss. JAMA Surg. 2016 Nov 1;151(11):1046–55.

30. Robinson MK. Surgical Treatment of Obesity — Weighing the Facts. https://doi.org/101056/NEJMe0904837. 2009 Jul 30;361(5):520–1.

31. English WJ, Williams DB. Metabolic and Bariatric Surgery: An Effective Treatment Option for Obesity and Cardiovascular Disease. Prog Cardiovasc Dis. 2018 Jul 1;61(2):253–69.

32. Pucci A, Batterham RL. Mechanisms underlying the weight loss effects of RYGB and SG: similar, yet different. J Endocrinol Invest. 2019 Feb 4;42(2):117–28.

33. Buchwald H, Avidor Y, Braunwald E, Jensen MD, Pories W, Fahrbach K, et al. Bariatric Surgery: A Systematic Review and Meta-analysis. JAMA. 2004 Oct 13;292(14):1724–37.

34. Cornejo-Pareja I, Clemente-Postigo M, Tinahones FJ. Metabolic and Endocrine Consequences of Bariatric Surgery. Front Endocrinol (Lausanne). 2019 Sep 19;10:626.

35. Korenkov M, Biertho L, Steffen R, Gagner M, Trelles N, Topart P, et al. Biliopancreatic Diversion With Duodenal Switch. Bariatr Surg Tech Var Complicat. 2022 Sep 19;135–67.

36. Dumping Syndrome | NIDDK [Internet]. [cited 2023 Feb 7]. Available from: https://www.niddk.nih.gov/health-information/digestive-diseases/dumping-syndrome

37. Jonas E, Marsk R, Rasmussen F, Freedman J. Incidence of postoperative gallstone disease after antiobesity surgery: population-based study from Sweden. Surg Obes Relat Dis. 2010 Jan

1;6(1):54–8.

38. Adult Obesity Facts | Overweight & Obesity | CDC [Internet]. [cited 2023 Feb 8]. Available from: https://www.cdc.gov/obesity/data/adult.html

39. Ward ZJ, Bleich SN, Long MW, Gortmaker SL. Association of body mass index with health care expenditures in the United States by age and sex. PLoS One. 2022 Mar 1;16(3 March).

40. Huang PL. A comprehensive definition for metabolic syndrome. Dis Model Mech. 2009;2(5–6):231–7.

41. Ahlqvist E, Storm P, Käräjämäki A, Martinell M, Dorkhan M, Carlsson A, et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. Lancet Diabetes Endocrinol. 2018;6(5):361–9.

42. Wang TJ. Risk Prediction in Cardiovascular Medicine Assessing the Role of Circulating , Genetic , and Imaging Biomarkers in Cardiovascular Risk Prediction. Circulation. 2011;123:551–65.

43. Klipp, Edda; Liebermeister, Wolfram; Wierling, Christoph; Kowald A. Systems Biology - A Textbook. 2nd Editio. Wiley; 2016.

44. Nielsen J. Systems Biology of Metabolism: A Driver for Developing Personalized and Precision Medicine. Cell Metab. 2017;25(3):572–9.

45. Romualdi C, Lanfranchi G. Statistical tools for gene expression analysis and systems biology and related web resources. Bioinforma Syst Biol. 2009;9781597454407:181–205.

46. Newgard CB. Review Metabolomics and Metabolic Diseases : Where Do We Stand ? Cell Metab. 2017;25(1):43–56.

47. Jordan KW, Nordenstam J, Lauwers GY, Rothenberger DA, Alavi K, Garwood M, et al. Metabolomic characterization of human rectal adenocarcinoma with intact tissue magnetic resonance spectroscopy. Dis Colon Rectum. 2009 Mar;52(3):520–5.

48. Rangel-Huerta OD, Pastor-Villaescusa B, Gil A. Are we close to defining a metabolomic signature of human obesity? A systematic review of metabolomics studies. Metabolomics. 2019 Jun 13;15(6):93.

49. Milward EA, Shahandeh A, Heidari M, Johnstone DM, Daneshi N, Hondermarck H. Transcriptomics. Encycl Cell Biol. 2016 Jan 1;4:160–5.

50. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009 Jan;10(1):57.

51. Karlsson FH, Nookaew I, Nielsen J. Metagenomic Data Utilization and Analysis (MEDUSA) and Construction of a Global Gut Microbial Gene Catalogue. PLoS Comput Biol. 2014;10(7).

52. Sommer F, Bäckhed F. The gut microbiota-masters of host development and physiology. Nat Rev Microbiol. 2013;11(4):227–38.

53. Mardinoglu A, Shoaie S, Bergentall M, Ghaffari P, Zhang C, Larsson E, et al. The gut microbiota modulates host amino acid and glutathione metabolism in mice. Mol Syst Biol. 2015;11(10):834–834.

54. Schroeder BO, Bäckhed F. Signals from the gut microbiota to distant organs in physiology and disease. Nat Med. 2016;22(10):1079–89.

55. Koh A, Molinaro A, Ståhlman M, Khan MT, Schmidt C, Mannerås-Holm L, et al. Microbially Produced Imidazole Propionate Impairs Insulin Signaling through mTORC1. Cell. 2018 Nov;175(4):947-961.e17.

56. Udayappan SD, Kovatcheva-Datchary P, Bakker GJ, Havik SR, Herrema H, Cani PD, et al.

Intestinal Ralstonia pickettii augments glucose intolerance in obesity. PLoS One. 2018;1(13):e0192339.

57. Kleinert M, Clemmensen C, Hofmann SM, Moore MC, Renner S, Woods SC, et al. Animal models of obesity and diabetes mellitus. Vol. 14, Nature Reviews Endocrinology. Nature Publishing Group; 2018. p. 140–62.

58. Renner S, Blutke A, Clauss S, Deeg CA, Kemter E, Merkus D, et al. Porcine models for studying complications and organ crosstalk in diabetes mellitus. Vol. 380, Cell and Tissue Research. Springer; 2020. p. 341–78.

59. Koya D, Kanasaki K. Biology of obesity: Lessons from animal models of obesity. J Biomed Biotechnol. 2011;(197636).

60. Yazdi FT, Clee SM, Meyre D. Obesity genetics in mouse and human: Back and forth, and back again. PeerJ. 2015;2015(3):1–69.

61. Lacruz ME, Kluttig A, Tiller D, Medenwald D, Giegling I, Rujescu D, et al. Instability of personal human metabotype is linked to all-cause mortality. Sci Rep. 2018;8:9810.

62. Riedl A, Gieger C, Hauner H, Daniel H, Linseisen J, Landstr I. Metabotyping and its application in targeted nutrition : an overview. Br J Nutr. 2017;12(117):1631–44.

63. Bouter KE, van Raalte DH, Groen AK, Nieuwdorp M. Role of the Gut Microbiome in the Pathogenesis of Obesity and Obesity-Related Metabolic Dysfunction. Gastroenterology. 2017 May 1;152(7):1671–8.

64. Savolainen O, Lind MV, Bergström G, Fagerberg B, Sandberg AS, Ross A. Biomarkers of food intake and nutrient status are associated with glucose tolerance status and development of type 2 diabetes in older Swedish women. Am J Clin Nutr. 2017;106(5):1302–10.

65. Sun L, Xie C, Wang G, Wu Y, Wu Q, Wang X, et al. Gut microbiota and intestinal FXR mediate the clinical benefits of metformin. Nat Med. 2018;24:1919–29.

66. Holmes E, Wilson ID, Nicholson JK. Forum Metabolic Phenotyping in Health and Disease. Cell. 2008;134(5):714–7.

67. Nicholson JK, Holmes E, Kinross J, Burcelin R, Gibson G, Jia W, et al. Host-gut Metabolic Interactions. Science (80-). 2012;336(June):1262–8.

68. de Maturana EL, Alonso L, Alarcón P, Martín-Antoniano IA, Pineda S, Piorno L, et al. Challenges in the Integration of Omics and Non-Omics Data. Genes (Basel). 2019 Mar 1;10(3).

69. Gomez-Cabrero D, Abugessaisa I, Maier D, Teschendorff A, Merkenschlager M, Gisel A, et al. Data integration in the era of omics: current and future challenges. BMC Syst Biol. 2014;8 Suppl 2(Suppl 2):I1.

70. Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics Data Integration, Interpretation, and Its Application. Bioinform Biol Insights. 2020;14.

71. Reel PS, Reel S, Pearson E, Trucco E, Jefferson E. Using machine learning approaches for multi-omics data analysis: A review. Biotechnol Adv. 2021 Jul 1;49.

72. Vanamala JKP, Knight R, Spector TD. Can Your Microbiome Tell You What to Eat? Cell Metab. 2015;22(6):960–1.

73. Wilmanski T, Rappaport N, Earls JC, Magis AT, Manor O, Lovejoy J, et al. Blood metabolome predicts gut microbiome α-diversity in humans. Nat Biotechnol. 2019;37(10):1217–28.

74. Van Olden CC, Van de Laar AW, Meijnikman AS, Aydin O, Van Olst N, Hoozemans JB, et al. A systems biology approach to understand gut microbiota and host metabolism in morbid obesity: design of the BARIA Longitudinal Cohort Study. J Intern Med. 2021;289(3):340–54.

75. Mooy JM, Grootenhuis PA, De Vries H, Kostense PJ, Popp-Snijders C, Bouter LM, et al. Intra-individual variation of glucose, specific insulin and proinsulin concentrations measured by two oral glucose tolerance tests in a general Caucasian population: The Hoorn Study. Diabetologia. 1996;39(3):298–305.

76. Koh A, Molinaro A, Ståhlman M, Khan MT, Schmidt C, Mannerås-Holm L, et al. Microbially Produced Imidazole Propionate Impairs Insulin Signaling through mTORC1. Cell. 2018;175(4):947-961.e17.

77. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote) omics data. Nat Methods. 2016;13(9):731–40.

78. Stan Pounds DF. HybridMTest: Hybrid Multiple Testing. 2019.

79. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, et al. The nf-core framework for community-curated bioinformatics pipelines. Nat Biotechnol 2020 383. 2020 Feb 13;38(3):276–8.

80. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014 Dec;15(15):550.

81. Ignatiadis N, Klaus B, Zaugg JB, Huber W. Data-driven hypothesis weighting increases detection power in genome- scale multiple testing. 2016;13(7).

82. Pantano L. DEGreport: Report of DEG analysis. 2019.

83. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr : interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013;14(14):128.

84. Kuleshov M V., Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016 Jul 8;44(W1):W90–7.

85. Kanehisa M, Goto S. KEGG : Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000;28(1):27–30.

86. Costea PI, Zeller G, Sunagawa S, Pelletier E, Alberti A, Levenez F, et al. A n a ly s i s Towards standards for human fecal sample processing in metagenomic studies. Nat Publ Gr. 2017;35(11):1069–76.

87. Chamberlain SA, Szöcs E. taxize: taxonomic search and retrieval in R. F1000Research. 2013;2:191.

88. Mcmurdie PJ, Holmes S. phyloseq : An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS One. 2013;8(4):e61217.

89. Magro DO, Geloneze B, Delfini R, Pareja BC, Callejas F, Pareja JC. Long-term weight regain after gastric bypass: A 5-year prospective study. Obes Surg. 2008 Jun 8;18(6):648–51.

90. Shubeck S, Dimick JB, Telem DA. Long-term outcomes following bariatric surgery. JAMA - J Am Med Assoc. 2018 Jan 16;319(3):302–3.

91. Larose JG, Leahey TM, Lanoye A, Bean MK, Fava JL, Tate DF, et al. Effect of a Lifestyle Intervention on Cardiometabolic Health among Emerging Adults: A Randomized Clinical Trial. JAMA Netw Open. 2022;E2231903.

92. Fiamoncini J, Rundle M, Gibbons H, Thomas EL, Geillinger-k K, Bunzel D, et al. Plasma metabolome analysis identifies distinct human metabotypes in the postprandial state with different susceptibility to weight loss – mediated metabolic improvements. FASEB J. 2018;32(10):5447–58.

93. Riedl A, Wawro N, Gieger C, Meisinger C, Peters A, Roden M, et al. Identification of Comprehensive Metabotypes Associated with Cardiometabolic Diseases in the Population-Based KORA Study. Mol Nutr Food Res. 2018;62(16):180017.

94. Cirulli ET, Guo L, Leon Swisher C, Shah N, Huang L, Napier LA, et al. Profound Perturbation of the Metabolome in Obesity Is Associated with Health Risk. Cell Metab. 2019;29(2):488-500.e2.

95. Vesanto J. SOM Toolbox [Internet]. Matlab. 2015. Available from: http://www.cis.hut.fi/projects/som- toolbox/

96. Kohonen TK. MATLAB Implementations and Applications of the Self-Organizing Map. 2014.

97. Kohonen T. Self-organized formation of topologically correct feature maps. Biol Cybern. 1982;43(1):59–69.

98. Kohonen T. Self-Organizing Maps. Vol. 30, Springer Series in Information Sciences. Berlin, Heidelberg: Springer International Publishing; 1995.

99. Kohonen TK. Self-Organizing Feature Maps. 1989;8:119–57.

100. Liu Y, Weisberg RH, Mooers CNK. Performance evaluation of the self-organizing map for feature extraction. J Geophys Res Ocean. 2006;111(5):1–14.

101. Davies D.L. BDW. A cluster separation measure. IEEE Trans Pattern Anal Mach Intell. PAMI-1(2):224–22.

102. Singh A, Shannon CP, Rohart F, Tebbutt SJ, Le K anh. Systems biology DIABLO : an integrative approach for identifying key molecular drivers from multi-omics assays. 2019;35(January):3055–62.

103. Singh A, Shannon CP, Rohart F, Tebbutt SJ, Le K anh. Systems biology DIABLO : an integrative approach for identifying key molecular drivers from multi-omics assays. Bioinformatics. 2019;35(17):3055–62.

104. Osborn O, Olefsky JM. The cellular and signaling networks linking the immune system and metabolism in disease. Nat Med. 2012 Mar;18(3):363–74.

105. Knebel B, Strassburger K, Szendroedi J, Kotzka J, Scheer M, Nowotny B, et al. Specific Metabolic Profiles and Their Relationship to Insulin Resistance in Recent-Onset Type 1 and Type 2 Diabetes. J Clin Endocrinol Metab. 2016 May 1;101(5):2130–40.

106. Husted AS, Trauelsen M, Rudenko O, Hjorth SA, Schwartz TW. GPCR-Mediated Signaling of Metabolites. Cell Metab. 2017 Apr 4;25(4):777–96.

107. Blad CC, Tang C, Offermanns S. G protein-coupled receptors for energy metabolites as new therapeutic targets. Nat Rev Drug Discov. 2012 Aug;11(8):603–19.

108. Irukayama-Tomobe Y, Tanaka H, Yokomizo T, Hashidate-Yoshida T, Yanagisawa M, Sakurai T. Aromatic D-amino acids act as chemoattractant factors for human leukocytes through a G protein-coupled receptor, GPR109B. Proc Natl Acad Sci U S A. 2009 Mar 10;106(10):3930–4.

109. Tang C, Ahmed K, Gille A, Lu S, Gröne HJ, Tunaru S, et al. Loss of FFA2 and FFA3 increases insulin secretion and improves glucose tolerance in type 2 diabetes. Nat Med. 2015;21(2):173–7.

110. Ahmed K, Tunaru S, Tang C, Müller M, Gille A, Sassmann A, et al. An autocrine lactate loop mediates insulin-dependent inhibition of lipolysis through GPR81. Cell Metab. 2010 Apr 7;11(4):311–9.

111. Smith PM, Howitt MR, Panikov N, Michaud M, Gallini CA, Bohlooly-Y M, et al. The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis. Science. 2013;341(6145):569–73.

112. Singh N, Gurav A, Sivaprakasam S, Brady E, Padia R, Shi H, et al. Activation of Gpr109a, receptor for niacin and the commensal metabolite butyrate, suppresses colonic inflammation and carcinogenesis. Immunity. 2014 Jan 16;40(1):128–39.

113. Jin S, Hoppel C, Chemistry KTJ of B, 1992 undefined. Incomplete fatty acid oxidation. The production and epimerization of 3-hydroxy fatty acids. ASBMB. 1992;267(1):119–25.

114. Kho D, MacDonald C, Johnson R. Application of xCELLigence RTCA biosensor technology for revealing the profile and window of drug responsiveness in real time. Biosensors(Basel). 2015;5(2):199–222.

115. Doornbos M, Linden I Van der, … LVB, 2018 U. Constitutive activity of the metabotropic glutamate receptor 2 explored with a whole-cell label-free biosensor. Biochem Pharmacol. 2018;151:201–10.

116. Jenske R, Vetter W. Concentrations of medium-chain 2- and 3-hydroxy fatty acids in foodstuffs. Food Chem. 2009 Jun 1;114(3):1122–9.

117. Masoud H, Lindner B, Weckesser J, Mayer H. The Structure of the Lipid A Component of Rhodocyclus gelatinosus Dr2 Lipopolysaccharide. Syst Appl Microbiol. 1990;13(3):227–33.

118. Farrell GC, Larter CZ. Nonalcoholic fatty liver disease: From steatosis to cirrhosis. Hepatology. 2006 Feb;43(S1):S99–112.

119. Balakrishnan M, Patel P, Dunn-Valadez S, Dao C, Khan V, Ali H, et al. Women Have a Lower Risk of Nonalcoholic Fatty Liver Disease but a Higher Risk of Progression vs Men: A Systematic Review and Meta-analysis. Clin Gastroenterol Hepatol. 2021 Jan 1;19(1):61-71.e15.

120. Samuel VT, Shulman GI. Nonalcoholic Fatty Liver Disease as a Nexus of Metabolic and Hepatic Diseases. Cell Metab. 2018 Jan;27(1):22–41.

121. 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.

122. Boursier J, Mueller O, Barret M, Machado M, Fizanne L, Araujo-Perez F, et al. The severity of nonalcoholic fatty liver disease is associated with gut dysbiosis and shift in the metabolic function of the gut microbiota. Hepatology. 2016;63(3):764–75.

123. Loomba R, Seguritan V, Li W, Long T, Klitgord N, Bhatt A, et al. Gut Microbiome-Based Metagenomic Signature for Non-invasive Detection of Advanced Fibrosis in Human Nonalcoholic Fatty Liver Disease. Cell Metab. 2017 May;25(5):1054-1062.e5.

124. Hoyles L, Fernández-Real JM, Federici M, Serino M, Abbott J, Charpentier J, et al. Molecular phenomics and metagenomics of hepatic steatosis in non-diabetic obese women. Nat Med. 2018 Jul;24(7):1070–80.

125. Meijnikman AS, Gerdes VE, Nieuwdorp M, Herrema H. Evaluating causality of gut microbiota in obesity and diabetes in humans. Endocr Rev. 2018;39(2):133–53.

126. White PJ, Newgard CB. Branched-chain amino acids in disease. Science (80-). 2019 Feb 8;363(6427):582–3.

127. Krautkramer KA, Fan J, Bäckhed F. Gut microbial metabolites as multi-kingdom intermediates. Nat Rev Microbiol. 2021 Feb 1;19(2):77–94.

128. Du Plessis J, Van Pelt J, Korf H, Mathieu C, Van Der Schueren B, Lannoo M, et al. Association of Adipose Tissue Inflammation With Histologic Severity of Nonalcoholic Fatty Liver Disease. Gastroenterology. 2015 Sep 1;149(3):635-648.e14.

129. Arab JP, Arrese M, Trauner M. Recent Insights into the Pathogenesis of Nonalcoholic Fatty Liver Disease. Annu Rev Pathol. 2018 Jan 24;13:321–50.

130. Fernandez-Real JM. Circulating Interleukin 6 Levels, Blood Pressure, and Insulin Sensitivity in Apparently Healthy Men and Women. J Clin Endocrinol Metab. 2001;86(3):1154–9.

131. Di Ciaula A, Wang DQH, Portincasa P. An update on the pathogenesis of cholesterol gallstone disease. Curr Opin Gastroenterol. 2018 Mar 1;34(2):71–80.

132. Wu T, Zhang Z, Liu B, Hou D, Liang Y, Zhang J, et al. Gut microbiota dysbiosis and bacterial community assembly associated with cholesterol gallstones in large-scale study. BMC Genomics. 2013 Oct 1;14(1).

133. Sayin SI, Wahlström A, Felin J, Jäntti S, Marschall HU, Bamberg K, et al. Gut microbiota regulates bile acid metabolism by reducing the levels of tauro-beta-muricholic acid, a naturally occurring FXR antagonist. Cell Metab. 2013 Feb 5;17(2):225–35.

134. Anveden Å, Peltonen M, Näslund I, Torgerson J, Carlsson LMS. Long-term incidence of gallstone disease after bariatric surgery: results from the nonrandomized controlled Swedish Obese Subjects study. Surg Obes Relat Dis. 2020 Oct 1;16(10):1474–82.

135. Brandão De Oliveira CI, Chaim EA, Da Silva BB. Impact of rapid weight reduction on risk of cholelithiasis after bariatric surgery. Obes Surg. 2003 Aug;13(4):625–8.

136. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114–20.

137. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34(5):525–7.

138. Oksanen Jari, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Sólymos, M. Henry H. Stevens HW. vegan: Community Ecology Package. R package; 2019.

139. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B. 1995 Jan 1;57(1):289–300.

140. Wang Q, Hao C, Yao W, Zhu D, Lu H, Li L, et al. Intestinal flora imbalance affects bile acid metabolism and is associated with gallstone formation. BMC Gastroenterol. 2020 Mar 6;20(1).

141. Wang J, Duan X, Li B, Jiang X. Alcohol consumption and risk of gallstone disease: a meta-analysis. Eur J Gastroenterol Hepatol. 2017;29(4):e19–28.

142. Elshaghabee FMF, Bockelmann W, Meske D, Vrese M de, Walte HG, Schrezenmeir J, et al. Ethanol production by selected intestinal microorganisms and lactic acid bacteria growing under different nutritional conditions. Front Microbiol. 2016;7(47).

143. Yuan J, Chen C, Cui J, Lu J, Yan C, Wei X, et al. Fatty liver disease caused by high-alcohol-producing Klebsiella pneumoniae. Elsevier.

144. Yazen A. Bile acid sulfation: a pathway of bile acid elimination and detoxification. Toxicol Sci. 2019;108(2):225–46.

145. Liang L, Yi Y, Liv Y, Qian J, Lei X, Zhang G. A Comprehensive Genome Survey Provides Novel Insights into Bile Salt Hydrolase (BSH) in Lactobacillaceae. Molecules. 2018;23(5):1157.

146. Jia B, Park D, Hahn Y, Microbes CJG, 2020 undefined. Metagenomic analysis of the human microbiome reveals the association between the abundance of gut bile salt hydrolases and host health. Taylor Fr. 2020 Sep 2;11(5):1300–13.

147. Tsai CJ, Leitzmann MF WW. Prospective study of abdominal adiposity and gallstone disease in US men. Am J Clin Nutr. 2004;80(1):38–44.

148. Tsai C, Leitzmann M, Willett W, Giovannucci E. Central adiposity, regional fat distribution, and the risk of cholecystectomy in women. Gut. 2006;55(5):708–14.

149. Krause BR HA. Adipose tissue and cholesterol metabolism. J Lipid Res. 1984;25(2):97–110.

150. Dikkers A TU. Biliary cholesterol secretion: more than a simple ABC. World J Gastroenterol. 2012;16(47):5936–45.

151. Chung s PJ. Dietary cholesterol effects on adipose tissue inflammation. Curr Opin Lipidol. 2016;27(1):19–25.

152. Cai J, Wang Z, Chen G, Li D, Liu J, Hu H, et al. Reabsorption of bile acids regulated by FXR-OATP1A2 is the main factor for the formation of cholesterol gallstone. Am J Physiol - Gastrointest Liver Physiol. 2020 Sep 1;319(3):G303–8.

153. Rudling M, Laskar A SS. Gallbladder bile supersaturated with cholesterol in gallstone patients preferentially develops from shortage of bile acids. J Lipid Res. 2019;60(3):498–505.