International PhD Programme in Biomolecular Sciences

Department of Cellular, Computational and Integrative Biology – CIBIO

XXXIV Cycle

Obesity and Health in the CHRIS Study

Advisor
Francisco S. Domingues, Dr. rer. nat.
Institute for Biomedicine, EURAC Research

Tutor
Fulvio Mattivi, Full Professor
CIBIO, University of Trento

Ph.D. Thesis by

Giulia Pontali
Institute for Biomedicine, EURAC Research
CIBIO, University of Trento

2021-2022 Academic Year

Abstract

Obesity is a major risk factor for multiple common chronic diseases. The prevalence in European countries is high and a significant public health concern. This thesis aims to explore the obesity landscape in the Cooperative Health Research in South Tyrol (CHRIS) study. The first step was to characterise the obese CHRIS population, taking into account the established body mass index (BMI) classification from the World Health Organization (WHO) and looking at metabolically healthy and unhealthy obesity. We investigated the familial aggregation of these traits. We identified several families with significant familial aggregation and observed varying degrees of overlap for these traits in different families. The focus was then on implementing and applying a Genome-Wide Polygenic Score for obese participants. These scores were computed for individuals based on the presence of different genetic variants weighted according to their measured effects in genome-wide association studies (GWAS). We then paid attention to the targeted metabolomics data of the CHRIS study, to identify different serum metabolites associated with metabolically healthy/unhealthy obesity, using logistic regression and random forest methods to explore metabolic signatures to distinguish obesity into metabolically healthy and metabolically unhealthy obesity. Several

biomarkers were shown to be related to obesity, many of which confirmed by existing evidence (such as BCAAs, tyrosine, and lysophosphatidylcholines).

Declaration

I, **Giulia Pontali**, confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Acknowledgements

It has been a challenging journey and in different ways many people have contributed to making all this possible. First and foremost, I would like to thank my supervisor, Dr. Francisco Domingues, for guiding me through my PhD, for his support, the scientific knowledge he shared with me and his patience and trust. I am also grateful to Prof. Fulvio Mattivi, all the members of the Biomedical Informatics group: Christian Weichenberger, Hagen Blankenburg, Essi Marjatta Hantikainen, Emanuela Kerschbamer, and also to Johannes Rainer, Cristian Pattaro and Luisa Foco. Their advice contributed greatly to this work.

I would also like to thank all the CHRIS study team and CHRIS participants, without whose contribution we would not have had such a valuable data resource.

My biggest thanks go to my parents, who have always trusted and supported my life decisions and have raised me to live boldly and not to be afraid of having big dreams.

The work was funded by the Tyrol-Alto Adige-Trentino Euregion, EUREGIO Environment, Food and Health (EFH) project.

Contents

Chapter 1

Background

1.1 Introduction

A worldwide increase in the obesity prevalence of obesity is currently a major concern, as overweight and obesity are known risk factors for multiple chronic diseases and obesity is associated with all-cause mortality. Many of the causes of overweight and obesity are preventable and reversible [1].

The EUREGIO Environment, Food Health (EUREGIO-EFH) initiative has established a local partnership of the neighbouring regions of Trentino and Alto Adige in Italy, and Tyrol in Austria, bringing together local strengths in the fields of biomedicine, nutrition and food chemistry to address these challenges. In this context, the goal of this thesis was to contribute to the initiative by exploring the obesity landscape in the region, based on data available from the CHRIS study. Considerable phenotypic variation is observed among obese individuals, resulting in different health outcomes. We further stratified obese individuals into metabolically healthy and unhealthy cate-

gories and investigated familial aggregation (*Chapter 2*). We implemented genomic-based methods to predict individuals at risk within the local population based on the CHRIS resource data (*Chapter 3*). With the recent advent of precision medicine, clinical metabolomics is in the spotlight due to its ability to provide molecular phenotyping of biofluids, cells or tissues. In this context, metabolomics is increasingly being applied to diagnose disease and understand disease mechanisms [2]. We investigated the metabolomic markers associated with obesity and metabolic healthy and unhealthy obese (*Chapter 4*), as we had the largest single-site targeted metabolomics data sets.

1.2 The CHRIS study

The Cooperative Health Research in South Tyrol (CHRIS) study is a population-based study taking place in the Val Venosta [3] (see Figure 1.1). The target population is relatively stable, with low residential mobility across generations and with homogeneous lifestyle and environmental conditions [3].

The study started in 2011 and the first phase was completed in 2018 [3]. In 2015, the study had 5 000 participants (CHRIS II), while in 2017 there were 10 518 participants (CHRIS III), increasing to 13 393 participants by the end of the first phase in 2018 (CHRIS baseline). The adult subjects were subjected to a core assessment protocol including an interview, self-administered questionnaires, blood and urine sampling, anthropometry, and ECG and blood pressure measurement, over an approximate duration of three hours [3]. Over the course of ten years, more than 40% of the adult population

in the target municipalities participated in the CHRIS study. The scientific objective of the CHRIS study is to investigate the molecular and genetic basis of common chronic conditions and their interaction with lifestyle and environmental factors. The study is expected to enhance the development of more effective approaches to disease prevention and early diagnosis [3].

As a prospective study, CHRIS research collected data about large families' clinical variables, medical histories, lifestyles (such as diet, physical activity, exposure to cigarette smoke, etc.) as well as genomic and metabolomic data from biobank samples, to determine if some of the afore mentioned factors, alone or combined with genetic background, can explain the causes, severity or protection against some diseases.

1.3 Obesity

In order to determine how far an individual's body weight varies from what is normal or desirable for a person of the same height, the most commonly used method is the body mass index (BMI), which is calculated by dividing the body weight in kilograms by the square of the person's height in meters (kg/m^2) [4]. The BMI provides a straightforward, inexpensive and useful measure of overweight and obesity [5], however it offers no detailed insight into body fat distribution or composition [6]. The association between body composition and sex and age is well established. For instance, women tend to store more fat subcutaneously than in the form of visceral adipose tissue [7] and will tend to carry more body fat than men at the same BMI [8]. Other common anthropometric measures of adiposity are waist and hip measure-

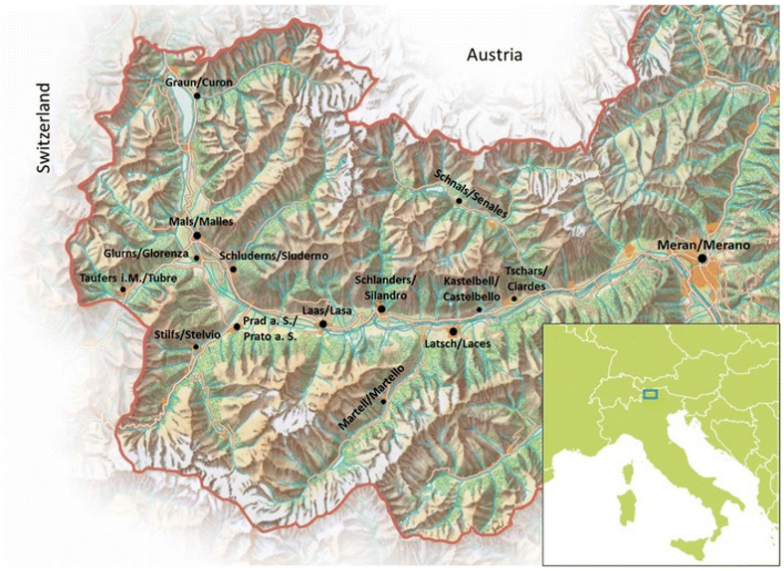

Figure 1.1: The Val Venosta region with all the municipalities involved in the CHRIS study

Reference source: From Pattaro et al. The Cooperative Health Research in South Tyrol (CHRIS) study: rationale, objectives, and preliminary results. J Transl Med. 2015;13: 348.

ments. The proportion of abdominal fat correlates with waist circumference and waist-to-hip ratio (WHR) [9, 10]. WHR is a simple and convenient way to estimate body fat distribution in epidemiological studies. WHR does not distinguish between accumulations of deep abdominal (visceral) fat and subcutaneous abdominal fat but is moderately associated with the amount of abdominal visceral adipose tissue [11]. Bio-electrical impedance analysis (BIA) can also establish body composition, with analysis of parameters including fat mass content by running a small electrical current through the body. Although BIA is simple and does not require special operator skills or substantial patient involvement [12], it has some limits in terms of va-

lidity, especially as regards the hydration factor [13]. There is a relatively high amount of extracellular water and total body water in obese individuals, which may lead to overestimation of fat-free mass and underestimation of fat mass [14].

Since people with excess visceral fat are at higher risk of obesity-related diseases such as hypertension, type 2-diabetes and insulin resistance than those with excess subcutaneous fat [15], more direct measures of fat distribution and composition are valuable. Dual-energy X-ray absorptiometry (DXA) provides estimates on fat and lean mass in the whole body or in specific body segments [16], and imaging methods such as magnetic resonance imaging (MRI) enable separate quantification of both visceral and subcutaneous adipose tissue [17]. These measures of body fat are more reliable but not easy to collect in population studies, as they require considerable resources as well as dedicated medical imaging facilities and staff. So, although BMI cannot differentiate between fat and muscle weight associated with fat, it provides the most useful population-level measure [18] and is an easy way to assess relative weight, also being used in the definition of obesity by the World Health Organization (WHO) [19]. The WHO classification of adult underweight, overweight, and obesity based on BMI (see Table 1.1) are the established reference standard.

The risk of unfavourable health outcomes for a slightly overweight individual gradually and seriously increases with further weight gain, i.e. the more overweight the individual is, the higher the likelihood of unfavourable health conditions [20, 21].

The effects of obesity can be felt in almost every part of the body, and

Table 1.1: WHO BMI classification

Classification	**BMI** (kg/m^2)
Underweight	< 18.50
Normal range	18.5 - 24.99
Overweight	25.00 - 29.99
Obese	≥ 30

the health effects related to obesity have been classified by WHO into four major categories: cardiovascular problems (including hypertension, stroke, and coronary heart disease); conditions associated with insulin resistance (such as diabetes mellitus); certain types of cancer (particularly hormonally related and large bowel cancer); and gallbladder disease [22, 23]. In addition to body weight, the location of body fat, the degree of weight gain during adulthood, and physical inactivity are factors that contribute to these adverse health consequences.

The current obesity epidemic is to a large extent the result of complex interplay between genetic, behavioural and environmental factors, including economic growth, modernisation, urbanisation and, most importantly, changes in lifestyle [24] with greater consumption of energy-dense and ultra-processed foods and at the same time, decreased physical activity [25].

Changing unhealthy lifestyles can reduce the chances of developing overweight or obesity [26] but risk factors such as a person's age (the risk of unhealthy weight gain increases with age), family history or genetics, ethnicity (overweight and obesity are highly prevalent in some ethnic groups), and gender cannot be changed.

Obesity is usually associated with metabolic disorders and cardiometabolic

diseases [27]. Nevertheless, not all obese people have metabolic complications, raising the question of whether metabolically healthy obese individuals are a unique subgroup of people with obesity [28]. Metabolically healthy obesity (MHO) has been characterised as the absence of components in metabolic syndrome, such as insulin resistance, hypertension and an unfavourable inflammatory profile [29]; while metabolically unhealthy obesity (MUO) has been associated with greater visceral abdominal fat, higher insulin levels and lower baseline high-density lipoprotein cholesterol [30]. The prevalence of MHO is currently in question [31, 32] and is even debatable [33]. MHO has recently been found to have prognostic value in predicting coronary heart disease, but a study shows that MHO individuals are still at higher risk of coronary heart disease, cerebrovascular disease, and heart failure than normal-weight metabolism-healthy individuals [33]. Weight loss intervention may not improve cardiovascular risk in MHO individuals [34], a finding with implications for obesity management.

Obesity, however, results from complex interactions between multiple genes and environmental factors such as physical activity, socioeconomic status, parent feeding behaviour, and diet. For this reason, obesity is mainly considered as a multi-factorial disease [35].

1.3.1 The genetics of obesity

Heritability studies provide strong evidence of a genetic contribution to obesity susceptibility [36], indeed fact obesity is influenced by hundreds of polymorphisms, each of which has a small impact [37].

The finding of common genetic variants associated with obesity and its

traits accelerated significantly thanks to Genome-Wide Association Study (GWAS) in 2007 [38]. GWAS explores associations between genetic variants across the genome, usually single nucleotide polymorphisms (SNPs), and phenotypes using DNA microarrays [1]. Therefore, GWAS examines the association between a wide range of SNPs across the whole genome without making assumptions about which genes or regions are likely to be associated with the phenotype.

The fat mass and obesity-associated gene (FTO) has been identified as an important locus harbouring common variants, with a clear impact on obesity predisposition and fat mass at population level [39]. Furthermore, FTO is highly expressed in the hypothalamus [40, 41] and controls appetite and energy expenditure. More specifically, hypothalamic neurons alter their own activities or those of their downstream targets in response to hormonal and nutrient signals [42]. Moreover, glucose metabolism in the hypothalamus plays a critical role in regulating food intake, energy expenditure, and carbohydrates and lipid metabolism [43, 44]. Researchers have found that in obese individuals, hypothalamic responses to glucose stimulation are lessened, supporting the importance of hypothalamic glucose sensing in regulating energy homeostasis [45].

1.3.2 Heritability of obesity

A population's heritability is the proportion of the variation in a given trait within a population that is not explained by the environment or random chance. Studies of large families, twins, and adoptions enable an assessment

[1] DNA microarrays commonly detect SNPs, to analyse the DNA of thousands of people.

of heritability and detect a strong genetic component in obesity (between 40 and 70 percent) [36]. Higher estimates of heritability for BMI between mother-child pairs than between father-child pairs suggest that the uterine environment of the mother may modify the genetic predisposition to obesity in offspring [46]. As a result of maternal weight gain during pregnancy and genetic factors, young children may develop obesity as adults [47]. Furthermore, genetic influences on obesity may also interact with sex and age. There are genetic factors that affect obesity differently in men and women [48]. An individual's heritability estimate increases from infancy to childhood [49], from childhood to preadolescence [50], and from preadolescence to adolescence [51], reaching a plateau during adolescence and adulthood. In late adulthood, however, it decreases slightly [52]. It is noteworthy that some forms of obesity can also display a somewhat variable phenotype [53]. Gene-gene interactions, mode of inheritance, and genetic heterogeneity can also be linked to obesity phenotypic variability [54], as well as interactions with environmental factors [55]. Indeed, lifestyle factors such as stopping smoking, sleep duration/quality, and psychosocial stress can significantly modify the impact of obesity predisposition [56].

1.3.3 Metabolomics and obesity

Omics research is currently exploring how genetic effects on phenotype are filtered through the metabolome, with metabolomics approaches playing a valuable role in bridging the gap between genotype and phenotype. This can be a powerful tool in informed decision-making and preventative healthcare because metabolites play a key role as modulators of biological processes [57].

For this reason, we also focused our attention on investigating the metabolomic profiles of obese and metabolically healthy/unhealthy obese participants in the CHRIS study. Characterisation of the metabolites associated with obesity and metabolically healthy/unhealthy obesity can provide an insight into the mechanisms that lead to this disease and its associated consequences, as metabolomics indeed provides a better understanding of the molecular events that control body weight [58]. Human serum metabolomics in particular can be used to discriminate the metabolic fingerprint of obese individuals who are metabolically unhealthy, compared to metabolically healthy individuals [59], as this fingerprint allows characterisation of the metabolism and exposure [60].

There are numerous metabolites in the human metabolome. This makes profiling them difficult due to the chemical diversity, abundance, and range of concentration, so different methods have been developed for metabolite profiling [61]. There are essentially two kinds of approaches: untargeted screens that are discovery-oriented and profile thousands of chemical features in a wide range of compounds for relative quantification, or targeted approaches focusing on absolute quantification of specific compounds, having the advantage of increased sensitivity and selectivity [62]. Some aspects of these methods differ, such as sample preparation, experimental precision, data interpretation, quantification level (relative or absolute), and metabolites detected [63]. The metabolic changes associated with obesity have been the subject of numerous reports and reviews both in humans and animals in the last decade [64], and most of these reviews report the results of targeted analyses based on LC-MS assay platforms (e.g., Biocrates, Metabolon) [65].

Furthermore, the identification of metabolites associated with obesity can provide an opportunity to gain new insight into metabolic profiling and pathophysiological mechanisms [66]. Branched-chain amino acids (BCAAs), including leucine (Leu), isoleucine (Ile), and valine (Val), are crucial for energy homeostasis, nutrition metabolism and gut health, immunity and disease [67]. These amino acids are associated with obesity, although changes in acylcarnitine and phospholipid levels have also been detected in obese subjects [60]. Metabolomics, especially when combined with multiple detection technologies or using a multi-omics approach, is a valuable tool for exploring obesity as well as metabolically healthy and unhealthy obesity. As a result of metabolomic analysis, we have increased our knowledge of various metabolic processes, such as how the metabolic network is affected by obesity. In metabolomics, however, dietary influences are a major issue, which we have taken into account by relating changes in biologically significant metabolites to obesity. Despite fasting before serum collection, the metabolic profile can be strongly affected by diet, particularly by energy metabolism and glucose utilization [60]. Metabolic analysis thus provides insightful information about the underlying processes in cells, disease, and consequently health. Studying the metabolome provides a better understanding of the underlying biology at all levels, from cellular to organismal level, by exploring the dynamics of cell response to internal and external perturbation. This information can be used to identify biomarkers and biological pathways that are active or inactive in states of disease or health [68, 69].

1.4 Data analysis strategies

Machine learning methods represent a powerful set of algorithms capable of characterising, adapting, learning, predicting and analyzing data [70] thus amplifying our understanding of obesity, metabolically healthy obesity and metabolically unhealthy obesity and our capacity to identify metabolites associated with it. In our study, the focus was on the identification of metabolites related to obesity and the prediction of obesity sub-types based on metabolomic signatures using logistic regression models and random forest classifiers. The random forest algorithm has been widely adopted as a biomarker detection tool in a wide range of studies related to metabolomics [71, 72] especially for estimates of variable importance for classification, due to resilience to high dimensionality and generally good performance in comparison to other approaches.

Prediction categories

Linear regression is the simplest and most widely used statistical technique for predictive modelling analysis. It is used to assess the relationship between a continuous dependent variable and one or more independent variables [73]. The overall goal of linear regression is to find the line that best fits the data points, to accurately predict output. If there are more than two independent variables in the model, it is then called multiple linear regression. The form of linear regression equation is exemplified by formula 1.1.

$$y = a_0 + a_1 x_1 + ... \tag{1.1}$$

where y is the dependent variable, x_1 is the independent variable, a_1 is the slope of the line, which can be positive or negative depending on the relationship between the dependent and independent variables, and a_0 is the y-intercept [74].

Logistic regression is a method used to predict a dependent variable, given a set of independent variables, so that the dependent variable is categorical (0/1) [75] and it functions in very similar way to linear regression. When analysing classification problems, this method is useful to determine whether a new sample fits best into a given category [74]. The form of the linear logistic equation is exemplified by formula 1.2, with output p between 0 and 1 for all values of independent variable x.

$$p = \frac{1}{1 + e^{-(a_0 + a_1 x)}} \tag{1.2}$$

Random forest is a classification algorithm consisting of many decision trees and it is used in classification and regression problems [76]. Decision trees seek to find the best split to subset the data, and they are typically trained through the Classification and Regression Tree (CART) algorithm [2]. Based on the predictions of decision trees, the random forest algorithm calculates the outcome chosen by the majority of decision trees avoiding the limitations of decision trees and minimising over-fitting by increasing accuracy [77].

The random forest algorithm is often used in regression as well as in classification because it can handle data sets that contain continuous variables,

[2]The CART algorithm split the nodes by searching for the best homogeneity for subnodes, with the help of the Gini Index criterion.

as in the case of regression, and categorical variables, as in the case of classification and it can handle non-linear parameters efficiently to provide better accuracy [78].

The ROC (Receiver Operating Characteristic) curve and the AUC (Area Under The Curve) [3] are used to assess model classification performance, while the mean decrease in Gini coefficient is used to measure feature importance[4]. Random forest allowed us to determine the most important predictors across the explanatory variables by generating 100 decision trees and ranking the variables by importance.

[3]AUC takes values from 0 to 1. An AUC of 0.5 suggests no discrimination between classes, 0.7 to 0.8 is considered acceptable, 0.8 to 0.9 is considered excellent, and more than 0.9 is considered outstanding [79].

[4]The higher the value of the mean decrease Gini score, the higher the importance of the variable in the model.

Chapter 2

Obesity in the CHRIS study

2.1 Introduction

The aim of this chapter is to investigate obesity in the CHRIS baseline of 13 388 participants (this number refers to the participant data analysed). The main focus was on evaluation of the quality of anthropometry data. In the medical literature, we found only one paper published in 2012 [80] in which weight was adjusted based on the average weight of the participants' clothing. Based on this work we wondered whether our data were affected by seasonality bias, due to the fact that CHRIS study participants were asked to take off only their outerwear and shoes to collect anthropometry data.

More specifically, we investigated whether BMI was associated with external temperature or not (in this case adjusting the BMI for seasonality). The purpose of data quality checks is to identify and fix bias or at least to minimise its impact on the results of a study [81].

Another aim of this chapter was to investigate familial predisposition to

obesity [82] in the CHRIS population, because this may be of great value for better understanding obesity aetiology, supporting approaches to predict outcomes and developing more effective preventive or therapeutic strategies.

2.2 Materials and methods

2.2.1 Anthropometry data

In the CHRIS study, four body composition measures were collected: body weight, body fat percentage, visceral fat and BMI derived from weight and height. Measurements were taken using an **Omron BF508** [1] monitor with an accuracy of 0.1 kg. All measurements were taken once by trained nurses or study assistants.

2.2.2 Effect of season on body weight

To verify if the seasonality of clothing weight affected BMI measurement, we plotted the BMI of each participant for every week in each year that the CHRIS study recruited participants, to investigate fluctuations in BMI, also evaluating average BMI and the average for participants over each week. Analysis of seasonality was carried out on CHRIS release III, made up of 10 517 participants, because at the time of analysis CHRIS follow-up had not been completed, so we analysed data from 2011 to 2017.

As outside temperatures could affect the weight of participants' clothes and consequently BMI measurement, a linear model was applied to investi-

[1] Body composition was measured using bioimpedance.

gate the significant effects of temperature on BMI.

Information regarding the temperature in Silandro [2] on the days when the participants visited the CHRIS centre was not available in our database, but could be downloaded from the province of Bolzano website [3], also providing the following information, recorded every 10 minutes: place where the temperature was taken; date on which the temperature was taken; time when the temperature was detected (from 00:00 am to 23:50 pm each day); outside temperature [4].

2.2.3 Obesity in the CHRIS study

We analysed BMI data from the 13 388 adults who participated in the CHRIS study. Women who were pregnant at the time of the interview, suspected to be pregnant, or who did not provide information on pregnancy status were excluded from the analysis. In total, we analysed 13 240 individuals.

In the CHRIS study, we classified the BMI of participants according to the WHO classification (see Table 2.1).

Some of the CHRIS study participants with obesity seemed to be protected from many of the adverse metabolic effects of excess body fat and were therefore considered to be *metabolically healthy*. Since metabolically healthy obesity does not have a universally accepted definition, its prevalence varies widely in different studies. It has also been debated whether metabolically healthy obesity (MHO) can be useful for prognosis or not,

[2] A small town with 6 256 inhabitants in the province of Bolzano.

[3] `http://dati.retecivica.bz.it/it/dataset/misure-meteo-e-idrografiche/resource/8cc47a38-1a93-47bf-871d-07b49dce56d0`

[4] Outside temperature was measured in Celsius.

Table 2.1: BMI classification in CHRIS according to WHO criteria

Classification	**BMI** (kg/m^2)	**Number of individuals**	**Percentage of individuals**
Underweight	< 18.50	207	1.56%
Normal range	18.5 - 24.99	6234	47.08%
Overweight	25.00 - 29.99	4556	34.41%
Obese	≥ 30	2243	16.94%

mainly because it is likely to shift gradually towards metabolically unhealthy obesity (MUO) [83]. We used the NCEP ATP III [84] definition to define metabolically healthy/unhealthy obesity because it is consistent with the data available from CHRIS participants, namely decreased HDL-cholesterol, diagnosis of cardiovascular disease, diagnosis for peripheral vascular diseases, high blood pressure, elevated triglyceride and elevated blood glucose. The definition of *metabolically healthy* or *unhealthy* is almost the same for males and females. The only difference concerns the high-density lipoprotein (HDL) cholesterol levels considered unhealthy, lower for males (less than 40mg/dl), and higher for females (50 mg/dl) (see Table 2.2).

We considered obesity and the presence of at least one of the six conditions in defining MUO. When defining MHO it is necessary to be careful about how to deal with missing values. Participants for whom we did not have any information about the criteria we used to define MHO were excluded from the analysis, so the people we considered to be MHO were those who had all variables in a healthy range and did not have any missing values.

Given the many serious health problems associated with obesity, studying the phenotype of healthy/unhealthy obesity may provide new insights into

Table 2.2: MUO individuals, defined as having a BMI $\geq$ 30 and $\geq$ 1 of the listed conditions. All individuals with a BMI $\geq$ 30 who were not MUO were considered to be MHO, but only if the listed conditions could be assessed (no missing data)

	MUO Criteria
Decreased HDL-cholesterol	(sex Male ***and*** (HDL $<$ 40 mg/dL ***or*** medication for blood lipids)) ***or*** (sex Female ***and*** (HDL $<$ 50 mg/dL ***or*** medication for blood lipids))
Diagnosis for cardiovascular diseases	stroke ***or*** transient ischemic attack ***or*** ischemic heart disease ***or*** atrial fibrillation
Diagnosis for peripheral vascular diseases	claudication or peripheral arterial disease ***or*** pulmonary embolus or blood clots in the lungs ***or*** treatment for blood clots in the legs called deep thrombosis or DVT
Blood pressure	SBP $\geq$ 140 mmHg ***or*** DBP $\geq$ 90 mmHg ***or*** medication for hypertension ***or*** medication for high blood pressure
Elevated TGs	triglycerides $\geq$ 150 mg/dL ***or*** lipid modifying agents
Elevated blood glucose	glucose $\geq$ 126 mg/dL ***or*** diabetes medication ***or*** diabetes mellitus ***or*** glycated haemoglobin $\geq$ 48 mmol/mol

the pathophysiology of obesity-related comorbidities and help identify obese individuals at risk [85].

2.2.4 Familial aggregation of obesity in the CHRIS study

To analyse the familial clustering of phenotypes in CHRIS families, we used an R/Bioconductor package, FamAgg package [86], to examine whether CHRIS participants with obesity or MHO/MUO aggregated within clusters of close relatives [5]. Of the 13 388 CHRIS baseline participants, we used 12 398 phenotyped individuals included in reconstructed families for the analysis. The families also included relatives of these CHRIS participants, with 24 173 individuals in total.

Pairwise relatedness was expressed as family-based kinship coefficient. According to the definition of obesity [19], we identified 2082 cases in the familial aggregation analysis. Similarly, 11 062 unaffected participants served as controls, and after removing 724 singletons and 22 opt-outs from the analysis in the control group, we ended up with 10 316 controls and 2 082 cases linked by 11 708 unphenotyped individuals.

The kinship sum can quantitatively identify individuals with a significantly large number of affected individuals related by common ancestry [86, 87]. The empirical p-value for affected cases was adjusted using the Benjamini-Hochberg test for multiple hypothesis testing and the false discovery rate was set to 0.05, while to examine familiarity we plotted the results using HaploPainter software v.1.043 [88]. The summary table provides an overview of the composition of families with significant familial aggregation of obesity in the CHRIS baseline. Family pedigree charts are commonly used to visualize

[5] Dr. Christian Weichenberger performed the familial aggregation analysis and collected the results.

the distribution of traits in families, but they can be problematic in terms of safeguarding the privacy of the participants and their relatives, especially if individuals are annotated with a visible trait like obesity. To describe families with familiar aggregation, instead of the standard pedigree charts, a summary table of each family has been provided to protect individual privacy.

2.2.5 GWAS of BMI

In most cases of human obesity, no single genetic cause can be identified [89]. To understand the genes and biological pathways that influence obesity in the CHRIS population study and to help expand our understanding of the biological processes underlying obesity, we performed a GWAS of BMI and compared the results with existing information in the literature to support candidate genes.

Genotyping data [6]

In genotyping, the DNA sequence, the so-called genotype, is determined at specific positions in the genome of an individual [90]. Sequence variations in population samples identified through genotyping can be used as markers in linkage and association studies to identify genes that may be relevant for certain traits or diseases. Depending on the number of samples, the number of genotypes to be tested, and the amount of sample material available, there are a variety of approaches to SNP genotyping.

The CHRIS 10K genotype data covers 10 518 individuals, and was first

[6] Genotype data was first collected by the CHRIS study team, while quality control and imputation were conducted by the Computational Genomics group led by Dr. Christian Fuchsberger.

collected with Illumina Human OmniExpressExon and Omni2.5Exome [91] chip arrays in two batches covering 612 000 variants.

Following quality control, the two genotype datasets were merged and imputed according to the TOPMed-r2 panel [92]. The TOPMed-r2 panel is based on 97 256 reference samples with high-quality whole-genome sequence data in GRCh38 and consists of 308 107 085 genetic variants distributed across the 22 autosomal chromosomes. Imputation was conducted using Michigan Imputation Server technology [93] which provides a web-based service for imputation that facilitates access to imputation reference panels [93]. GWAS of BMI was performed using EPACTS (Efficient and Parallelizable association Container Toolbox) [94] software.

2.3 Results

2.3.1 Effect of season on body weight

Average BMI was consistent over the weeks and years, indicating that the potential seasonal effects of clothing weight did not have a significant impact on measured weight, otherwise we would have recorded considerable differences, especially from winter to summer (see Figure 2.1).

The CHRIS study work programme started between 7:45 and 8:30 am, so our participants' clothing was influenced by the temperature outside. To simplify the data, we took into account the temperature in Silandro between 6:00 and 8:00 am, which was the approximate period during which the participants left home to go to the CHRIS study centre, and calculated the average. This information was used as an independent variable in the linear model. In

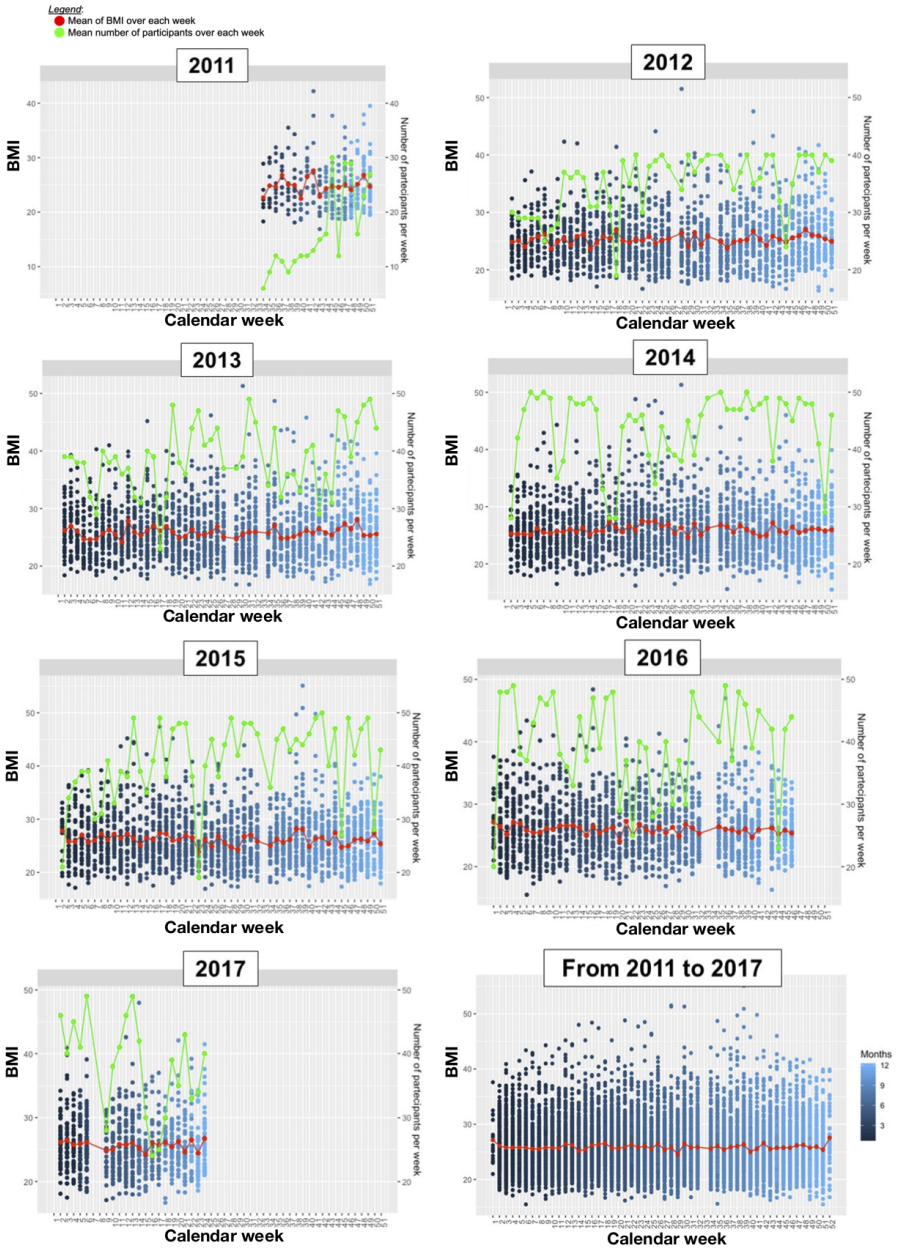

Figure 2.1: Weekly BMI for the years in which the CHRIS study was carried out (data are taken from CHRIS release III)

addition to temperature, we used sex and age as independent variables and BMI as a dependent variable, as exemplified by the following formula 2.1:

$$BodyMassIndex \sim meanTemperature + age + sex. \tag{2.1}$$

The significance of the relationships was estimated with a t-statistic, and a p-value below 0.05 was deemed significant. In our model the p-values for the intercept, sex and age were extremely small (p-value $<$ 2e-16***) therefore significant, so the outcome suggests that sex and age are strongly associated with BMI, while temperature is not. This model also underlines that age and sex are factors that should be considered when dealing with health, because they are key determinants significantly influencing health-care-seeking behaviour [95].

2.3.2 Obesity in the CHRIS study

We observed a conspicuous number of overweight and obese individuals (see Table 2.1). These numbers are in good agreement with previous European population reports [85], where the percentage of obesity varied between 6% and 20% [96].

The analysis was repeated, stratifying BMI by age group: 18-30; 30-40; 40-50; 50-60; $\geq$ 60 (see Figure 2.2). Overweight and obesity are prevalent in middle adulthood. With age, there is a greater tendency to weigh more due to a sedentary lifestyle, a decrease in muscle mass and the development of possible health conditions. In addition, changes in metabolism and hormone levels affect the rate and extent of fat accumulation [97]. This is the reason

why BMI gradually increases from the age of 40 [98]. The dataset we analysed consisted of 7 160 females and 6 080 males and average BMI was slightly higher for males: 26.38 versus 25.27 (see Figure 2.3).

The majority of participants who were within the normal range were female, while the majority of participants who were overweight were male. However, obesity prevailed in females (see Figure 2.4). These figures were to be expected because there are gender differences in the storage and metabolism of fat tissue. In women, the biological factor of menopause affects fat distribution, which may increase the risk or exacerbate the negative health effects of obesity [99].

Not all obesity is the same, and the CHRIS study has so far investigated the largest number of MHO/MUO individuals as compared to other published studies, highlighting its strength and potential relevance. We identified 1 645 participants who were metabolically unhealthy obese and 529 who were metabolically healthy obese. The age prevalence for the MUO peaks at age 60, while for the MHO it is 45 (see Figure 2.5). These data suggest that the MHO represent a transient phenotype [100], due to likely development of cardiometabolic abnormalities over time [101]. Early identification of the MHO group is an opportunity for primary intervention.

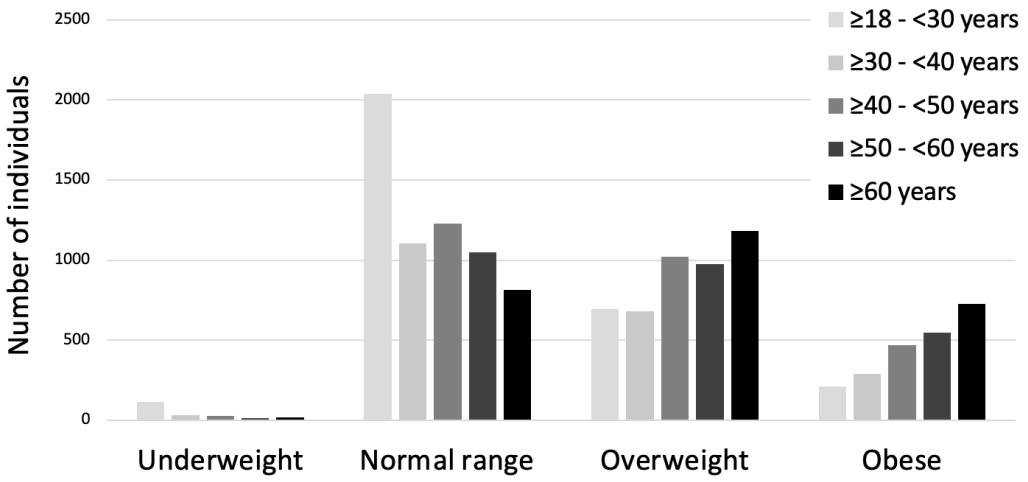

Figure 2.2: Stratified BMI WHO categories according to age group

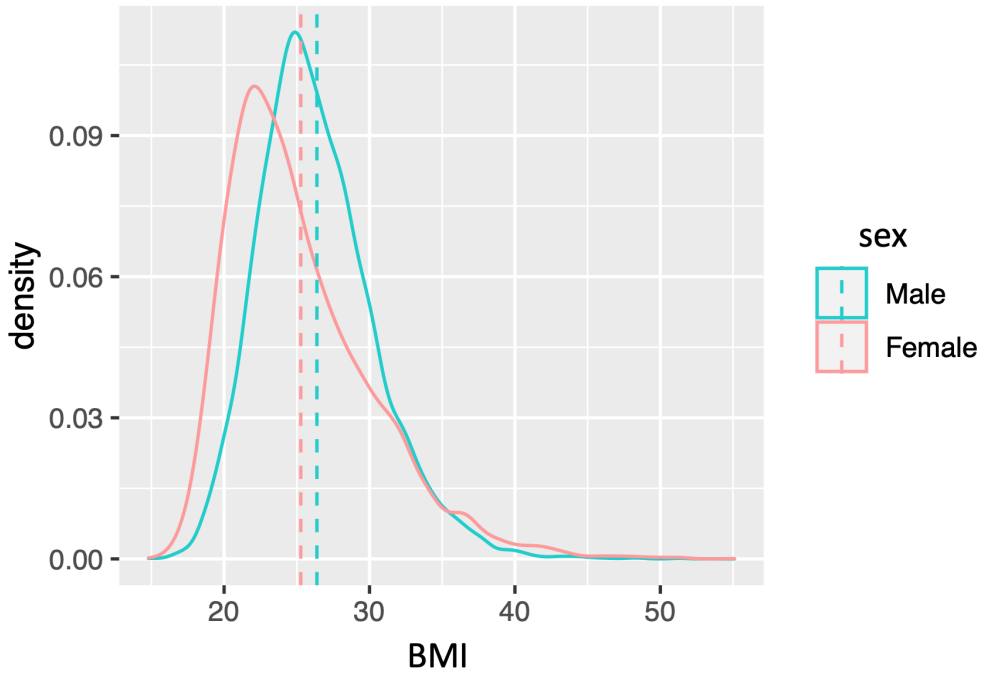

Figure 2.3: BMI distribution according to sex. The dashed line shows the average BMI in males and females

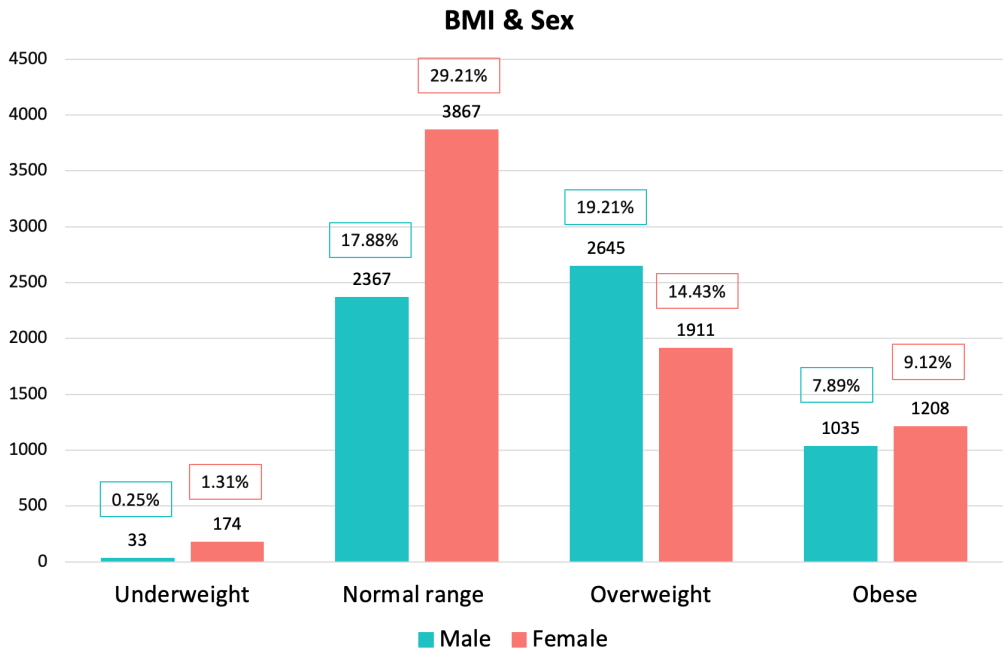

Figure 2.4: Stratified BMI WHO categories according to the sex group

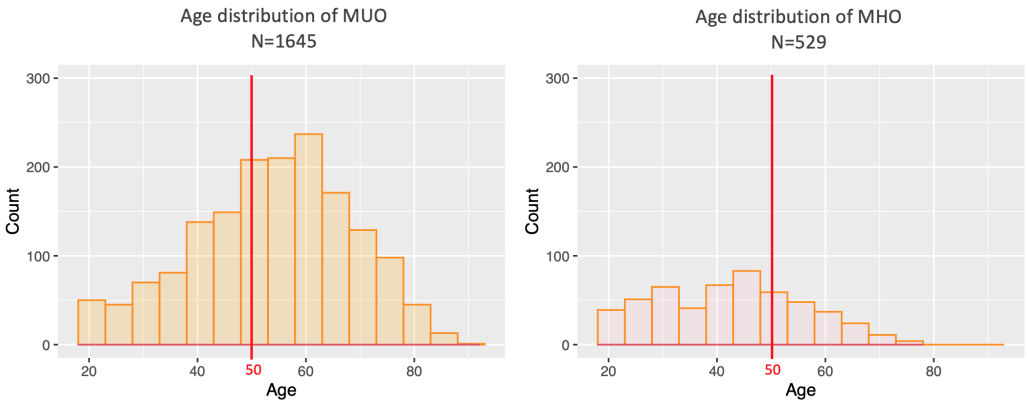

Figure 2.5: Age distribution for the metabolically unhealthy and metabolically healthy obese

2.3.3 Familial aggregation of obesity in the CHRIS study

The obesity trait is strongly influenced by genetics, so, to understand how it is distributed in families we carried out famAgg analysis to see how the trait (obesity, metabolically healthy obesity and metabolically unhealthy obesity) clusters in families. The summary table for each family includes the following information: the *family* column refers to a family identifier; *age* refers to the age (binned in decades) of the individual giving rise to a significant hit in the kinship sum test, this person being thought of as a central element connecting a significant number of affected relatives and termed a *significant hit*; *nped* is the number of individuals in the family and *nphe* gives the number of phenotyped individuals in the family. The remaining columns provide the total number of individuals affected and phenotyped (in parentheses) with different degrees of relationship to the significant hit: first, second, third,

and higher than third-degree. For each degree, there is a pair of numbers N (M), where N is the number of affected individuals and M is the number of phenotyped individuals with a certain degree of relationship to the significant hit.

We observed six families with significant familial aggregation of obesity in the CHRIS baseline (see Table 2.3). In the 1ob family we observed 9 obese individuals related by common ancestry to the significant hit. Of these 10 obese individuals (including the significant hit), six belonged to a more extreme obesity class II (BMI $\geq 35 kg/m^2$), the obese individuals also tending to be older, with eight over 40 years of age. It was also noticeable that out of all 13 younger individuals (age < 40) related by ancestry to the significant hit, only two were obese but seven were overweight, indicating that the tendency for obesity was also apparent at younger ages. Noticeably, in both the 1ob and 4ob families the distribution of obesity fitted an autosomal dominant mode of inheritance, indicating potential monogenic forms of obesity.

We also observed three families with significant familial aggregation of MUO (see Table 2.4). In the 1muo family we observed a wide range in the age of affected individuals (from 34 to 75 years old). Interestingly all the obese individuals in the family were MUO with a consistent obesity phenotype. In the 2muo family we observed in total 10 MUO individuals, all related by a common ancestor, but we also observed an MHO phenotype in one additional young person (28 years old). Obesity in the 3muo family, where we observed 10 MUO individuals, together with three additional MHO young people, was not as uniform. There was an overlap between these families and the familial aggregation results for obesity, with the 1muo, 2muo and 3muo

families matching 1ob, 4ob and 6ob respectively. Noticeably, in these three MUO families, the pattern of MUO fitted an autosomal dominant mode of inheritance. Familial aggregation of MHO was also observed in two families (see Table 2.5). In the 1mho family, both younger and older individuals were affected by MHO (18 to 61 years old). In the main part of the family we had six MUO individuals, all with a first, second or third-degree relationship to the significant hit, obesity also being relatively uniform, with only one additional MUO in this part of the family. This family overlapped with the 5ob family. The second 2mho family is more diverse, as we also observed MUO in older individuals.

Taking into account the familial aggregation results for MUO and MHO, in some families a stable MUO phenotype predominated for both younger and older individuals (1muo and 2muo), while in other families we detected a transient MHO phenotype with a mixture of older MUO and younger MHO (3muo and 2mho), and in one case a more stable MHO obesity phenotype that predominated over different age groups (1mho subset).

Table 2.3: Significant familial aggregation of the obesity trait

family	age	nped	nphe	first	second	third	higher
1ob	60s	46	31	6(7)	2(9)	1(10)	0(0)
2ob	60s	51	32	6(9)	1(8)	3(5)	0(4)
3ob	60s	93	70	4(5)	3(9)	5(11)	3(29)
4ob	70+	41	31	4(4)	5(10)	1(9)	0(0)
5ob	40s	95	63	3(4)	3(4)	6(17)	9(25)
6ob	70+	132	88	4(5)	2(9)	5(16)	2(40)

Table 2.4: Significant familial aggregation results of the metabolically unhealthy obesity

family	age	nped	nphe	first	second	third	higher
1muo	60s	46	31	6(7)	2(9)	1(10)	0(0)
2muo	70+	41	31	4(4)	4(10)	1(9)	0(0)
3muo	70+	132	87	4(5)	0(9)	4(16)	2(39)

Table 2.5: Significant familial aggregation results of the metabolically healthy obesity

family	age	nped	nphe	first	second	third	higher
1mho	40s	55	38	3(5)	1(8)	1(1)	1(15)
2mho	40s	130	93	2(5)	1(7)	0(15)	3(49)

2.3.4 GWAS of BMI

The SNPs most strongly associated with BMI (p-value $< 5 \times 10^{-8}$) (see Figure 2.6) were mapped to chromosome 16 where the FTO gene is also allocated. FTO has previously been identified as an important locus harbouring common variants with a clear impact on obesity predisposition and fat mass at population level [39]. To facilitate visualisation of the results we used LocusZoom [102], which also gives useful information about the locus, such as the location and orientation of the genes it includes, and the location of markers associated with other traits near the SNPs of interest (see Figure 2.7).

We investigated whether the results replicated findings from previous studies, or whether there were any novel associations. For this purpose, we used the SWISS v1.1.1 package [103] to look up the locus in the EBI GWAS catalogue [104] with LD clumping set to r2 > 0.1, and distance-based clumping set at 1Mb. Statistical significance was assessed at the p-value threshold of 5×10^{-8}.

Our results are consistent with previous reports and we did not find any new significant associations. This is probably due to the relatively small size of our study (10 518) compared to the largest BMI GWAS of 700 000 samples [105] reported in the literature.

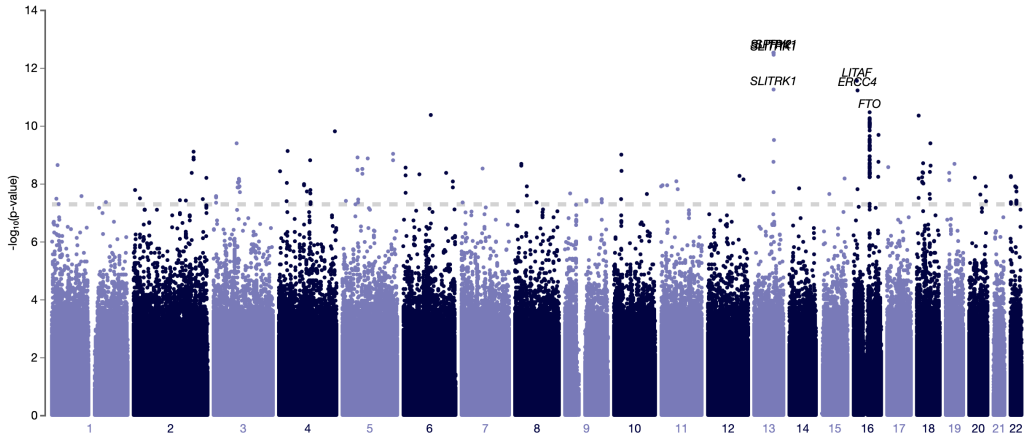

Figure 2.6: Manhattan plot for GWAS analysis of BMI in the CHRIS population. The x-axis represents the genomic coordinates from each region of the genome tested, organised by chromosome; the y-axis represents the p-value for each individual nucleotide polymorphism

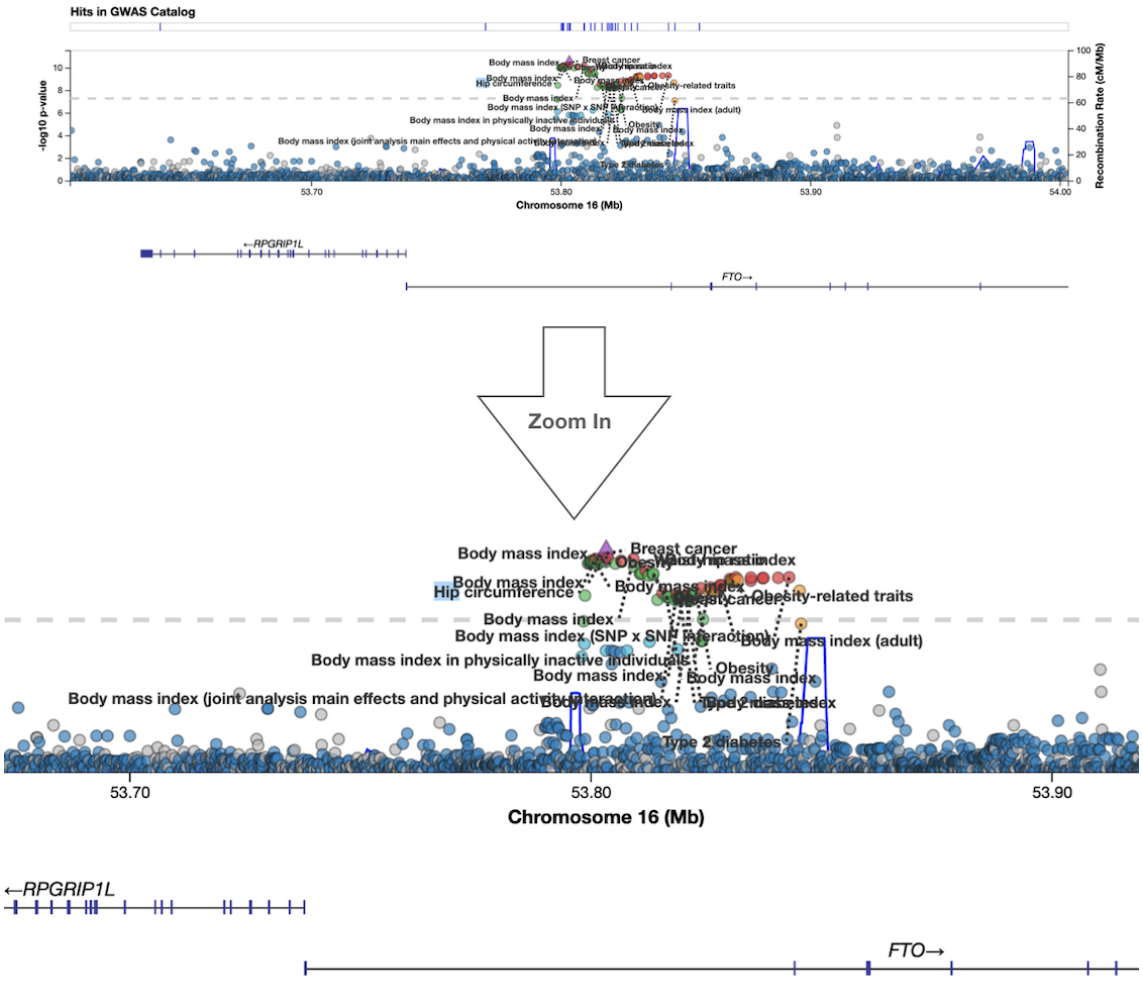

Figure 2.7: LocusZoom plots showing the association (left y-axis; log10-transformed p-values) with the BMI trait. The colour of the data points represents the linkage disequilibrium between each SNP and the top SNPs. The positions of genes are shown below the plot

2.4 Conclusion

In this chapter we characterised obesity in the CHRIS population according to WHO criteria, with 17% of the population studied classified as obese. Obesity is not only a problem of individuals but also a generalised problem for society and populations. This implies a need for population-based changes in lifestyle for better prevention.

Obese people who do not have metabolic complications are defined as metabolically healthy obese. This should not be considered a safe condition not requirung obesity treatment, but should rather guide decision-making for personalised and risk-stratified obesity treatment. For this scope, clearer and more consistent diagnostic criteria would have obvious clinical implications, such as improvement of risk stratification and more appropriate, cost-effective forms of treatment.

It is not only the study of the obesity trait itself, but also analysis of family aggregation that can be of great value for better understanding obesity aetiology and for supporting approaches to predict outcomes and develop more effective preventive or therapeutic strategies. It will be of great interest to further explore potential genetic and lifestyle determinants (including diet and physical activity) for stable MUO, transient MHO and stable MHO in the future.

Furthermore, GWAS pinpoints common BMI alleles that occur in over 5% of the population. The role of FTO risk alleles in obesity is significantly intensified by reduced physical activity and high-calorie diets.

Chapter 3

Genome-Wide Polygenic Score

3.1 Introduction

The aim of this chapter is to evaluate the susceptibility to obesity and predict BMI in the CHRIS population study, based on the individual's genotype. To achieve this goal we have used the genome-wide polygenic score (GPS), a calculation that can be used to assess the genetic predisposition for a certain complex condition or disease. When many genes and hundreds or thousands of variants with small effects are involved, combining their effects can help determine the predisposition for underlying conditions or disease [106]. Variants identified by genome-wide association studies (GWAS) can therefore be combined, in the simplest form in proportion to the effect size, to provide a GPS. The GPS gives information about where an individual falls within a population, for example within the top 10%. Identification of individuals at high risk of developing disease brings multiple benefits to preventive approaches, facilitating earlier lifestyle changes that bring positive health outcomes and

delay disease progression [106, 107]. With this scope, the GPS can be used as a tool to assess the risk of complex disease [108], but it is fundamental to be aware of problems related to portability and interpretation. Individuals with one genetic ancestry may have more predictive power than those with another ancestry [109]. So, there is no good "portability" of the score across populations. The reason for this is that different populations display different patterns of linkage disequilibrium and allele frequencies. The second problem is related to interpretation of the results, due to the fact that the GPS requires extensive genotype data to have good prediction performance [110], and furthermore it does not take into account the fact that the same variants can have a different overall effect in different environments [111].

Despite the limitations of the GPS, the complexity of polygenic phenotypes makes it challenging to study each individual locus so GPSs have become widely used in research on polygenic traits as a tool for representing the genetic burden across the genome [112].

3.2 Materials and methods

The GPS provides an estimate of genetic impact on phenotypes. It is the weighted sum of alleles that may contribute to phenotype, where x is the genotype count and β is the effect size, as summarised in formula 3.1.

$$GPS = \sum \beta x \tag{3.1}$$

In general, the weight is derived from effect estimates from previous GWAS studies. These effect estimates or GWAS data are described as base data.

The raw genotypes used to make the prediction are known as the target data. While the GPS model may not represent causal SNPs, it can represent the genetic burden of an individual by using SNPs that are included in the model.

Even though the GPS is a straight-forward concept, a few details must be kept in mind:

- When the base and target data overlap, overfitting occurs [113]. The model can underperform or be biased if the base and target populations have different allele frequencies, population structures, and environmental factors. If there is a multi-ethnic population, great caution is required to adjust possible confounding factors [114].

- The SNPs in linkage disequilibrium (LD) [1] will cause redundant signals in the GPS. The clumping method is the most widely used for preventing redundant signals, with SNPs selected according to LD and p-value. In this way, the correlation between the remaining SNPs is reduced while the strongest statistical evidence is retained [115].

- In addition, the correction of missing SNPs needs to be carried out excluding samples with too many missing SNPs. As indicated by formula 3.2, when calculating the GPS, the raw $\sum \beta x$ should be divided by the number of SNPs N (excluding all the missing SNPs) used in the GPS.

[1] LD refers to the association between alleles at different loci.

$$GPS = \frac{1}{N} \sum \beta x \tag{3.2}$$

The GPS is the average genetic burden for non-missing SNPs.

P-value thresholding is the most common way to optimise the GPS. In general, the smaller the p-value, the more likely the association between SNP and trait and the more likely their inclusion will contribute to GPS predictive ability. Those SNPs with p-values within the threshold deemed to have valid estimates of effect size are included; all others are excluded. With the p-value thresholding method, different p-value thresholds are used in order to gain the best signal-to-noise ratio. On conclusion, the GPS result with the highest signal-to-noise ratio is considered the best of all the iterations. Currently, the most popular approach for calculating the GPS is to use p-value thresholding on clumped SNPs (clumping + p-value thresholding) [116].

3.2.1 PRSice2

Our analysis was carried out on large-scale data using PRSice-2 [117], an efficient and scalable program for automation and simplification of GPS analysis of large-scale data. In PRSice-2, genotyped and imputed data are handled for empirical association p-values free of overfitting, inheritance models are supported, and multiple continuous and binary traits can be evaluated simultaneously [117]. Its predictive power is comparable to that of PRSice and other GPS alternatives, LDpred [118] and lassosum [119], while being dramatically faster and more memory-efficient [117].

Input data:

Base data must be provided as a whitespace delimited file containing association analysis results for SNPs on the base phenotype [120]. Columns with the following header names are essential and must be present: SNP name, effect allele, effect size estimate as an odds ratio (binary phenotype) or continuous beta effect (continuous phenotype) and p-value for the association. Other fields can be included, such as the chromosome number, base pair position, reference allele and standard error [120]. As base data we used the summary statistics generated in the Yengo et al. study [105], which includes summary statistics from previous GWAS in the Genetic Investigation of ANthropometric Traits (GIANT) consortium [121, 122] studies with new GWAS of height and BMI in $\sim 450\,000$ participants in the UK Biobank (UKB). In total, the sample size included $\sim 700\,000$ participants with European ancestry. Combined GWAS meta-analysis substantially increases the number of GWAS signals associated with BMI. These data correspond to the most powerful GWAS results available on the trait studied. The following checks were made on the base data, according to the QC checklist present in the GPS guideline paper [114]: heritability check - GWAS heritability data should be greater than 0.05 ($h^2_{snp} > 0.05$); allele check - effect allele and non effect allele to discover the correct direction; genome check - base and target data must be on the same genome build; checking of duplicated and ambiguous SNPs.

Target data refer to imputed genotyping data from the CHRIS III study, involving 10 758 participants [2]. Since the base data come from genome build hg19 [3], we used the HRC imputation panel for the target data. In this way the base and target SNP data had genomic positions assigned on the same genome build.

Some of the individuals involved in the CHRIS study were present in summary statistics in the Yengo et al. study [105], because a subsection of the CHRIS study (overlapping with the MICROS study [123]) contributed to the Locke et al. study [122], a study also taken into account by Yengo et al. [105] (Locke et al. study + UKBiobank data). An overlap in base and target data samples could substantially inflate the association between the GPS and the trait tested in the target data, and therefore had to be eliminated [114]. Ideally, the base data must be cleared of overlapping samples and the GWAS calculated again [114]. Since this was not possible because we had no access to the public data resource, we decided to eliminate sample overlap in the target data. We therefore removed the 704 individuals in the target data who were also present in the base data, ending up with 10 054 individuals in the target data. We also needed BMI phenotype information for these 10 054 individuals, and considering that we might want to account for age and sex in the GPS, we incorporated both into the analysis.

[2]The target data set must be supplied in PLINK binary format, with the following extensions: *.bed* - BED (Browser Extensible Data) is a text file used to store genomic regions as coordinates and associated annotations; *.bim* - BIM is an extended variant information file accompanying a .bed binary genotype table; *.fam* - FAM gives the ID, sex and phenotype information for each individual.

[3]Homo sapiens (human) genome assembly GRCh37 (hg19) from the Genome Reference Consortium.

3.3 Results

After running PRSice2, we were able to explore the predictive power of the GPS on BMI for 10 054 individuals in our independent test set, evaluating whether the GPS could predict obesity in a test dataset of participants in the CHRIS study, where the mean BMI was 25.79 kg/m^2, and 17% met the criteria for obesity. The GPS approximated normal distribution in the population, the cut-off for the predictive test being expressed in the vertical lines (see Figure 3.1). We considered the top 10% of the distribution to be carriers of a high GPS with a strong probability of becoming obese or severely obese (see Figure 3.1).

We next stratified the population according to GPS decile. The average BMI was 28.3 kg/m^2 for those in the top decile of the GPS and 23.74 kg/m^2 for those in the bottom decile, a difference of 4.6 kg/m^2. This means we can observe a clear tendency to lower and higher BMI values at the two extremes of the GPS (see Figure 3.2). At the top of the distribution, we can see many overweight and obese participants, but also a conspicuous number of individuals whose weight is in the normal range (see Figure 3.1). In the lower part of the distribution, we can instead see a majority of participants of normal weight and only a small fraction of overweight or obese individuals (see Figure 3.1). Despite this, it is important to point out that a low-risk score does not mean there is no risk of becoming overweight or obese.

To assess whether participants with a high GPS were also at a higher risk of developing certain cardiometabolic diseases, we tested for a potential association between a high GPS and both diabetes mellitus and coronary artery disease. To do this we used logistic regression, taking the diseases as

dependent variable and age, sex and binarized GPS (1 for the top decile of the polygenic score, 0 for the rest) as independent variables. A high polygenic score was associated with the cardiovascular diseases under consideration (p-value 0.00839 for diabetes mellitus and 0.0296 for coronary artery disease). Susceptibility to obesity was thus also associated with an increased risk of debilitating diseases, which in turn decreases the quality and length of life.

We compared the results obtained thus far with those presented in the Khera et al. study [124], which quantified and tested susceptibility to obesity for $\sim$ 300 000 individuals (see Table 3.1). The study design was the same but there were some differences: target data (the Kera et al. study [124] had a larger number of variants); base data, we used the most recent summary statistics present when the analysis was done [105]; there were fewer common variants in the base and target data in the CHRIS study than in the Kera et al. study [124] because the CHRIS study had fewer individuals; the average age of participants was 57y in the Khera et al. study [124] while it was 46y in CHRIS study. This difference in age also explains why mean BMI was lower in the CHRIS study than in Khera et al. study [124]. In our study, to calculate the GPS we decided to use PRSice2 [117] and not LDpred [118] because it is more efficient and scalable. The correlation score in the Khera et al. study [124] explains 29% of trait variance, while in the CHRIS study it was only 18%.

Despite the limited sample sizes used for our study our results were promising and helped us to understand individuals' susceptibility to obesity.

Table 3.1: Comparison of the Khera et al. and CHRIS studies

	CELL – Khera et al.	**CHRIS study (release 3)**
Target data	$\sim$ 300 000 individual-level genotype-phenotype data	$\sim$ 10 000 individual-level genotype-phenotype data
Base data	2 200 302 genetic variants on BMI (Locke et al., 2015)	2 336 270 genetic variants on BMI (Yengo et al., 2018)
Base data ethnicity	European	European
Common variants (Base data and Target data)	2 100 000	953 852
Participants mean age	57y	46y
Mean BMI	27.4	25.8
GPS algorithm	LDpred (Vilhjàlmsson et al., 2015)	PRSice2 (Choi et al., 2019)
Correlation score	0.292	0.182
Stratified population according to GPS decile	Average BMI was 30.0 kg/m^2 for those in the top decile of the GPS and 25.2 kg/m^2 for those in the bottom decile	Average BMI was 28.30 kg/m^2 for those in the top decile of the GPS and 23.7 kg/m^2 for those in the bottom decile

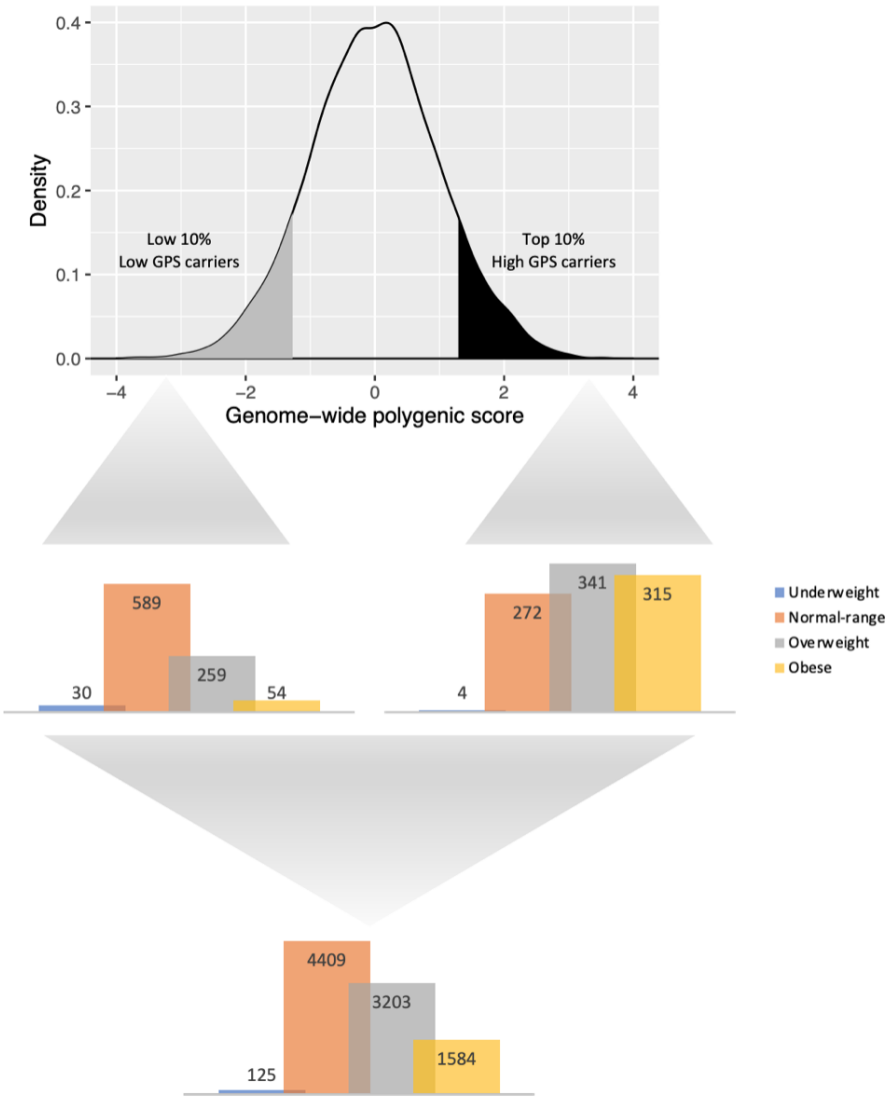

Figure 3.1: We considered the top 10% of distribution to be 'carriers' of a high genome-wide polygenic score (GPS). The x-axis represents the polygenic score, with values scaled to a mean of 0 and standard deviation 1 to facilitate interpretation. The histograms represent the distribution of BMI with a low GPS, high GPS, and across the GPS

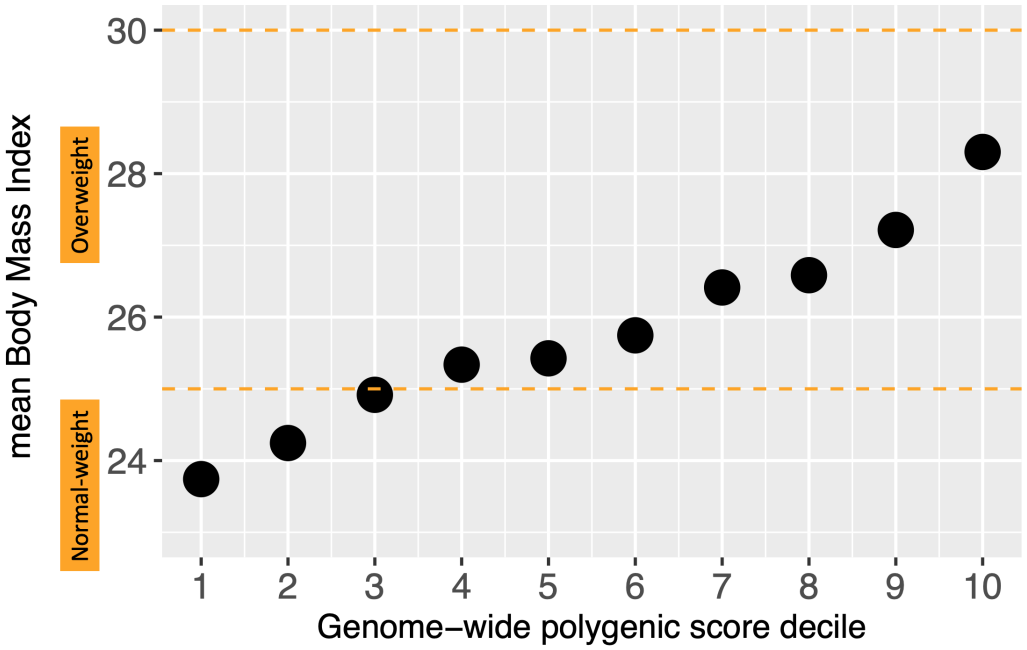

Figure 3.2: We stratified CHRIS participants according to the GPS decile. The x-axis represents the GPS decile and the y-axis mean BMI for each decile

3.4 Conclusion

We derived and validated a GPS for predicting polygenic susceptibility to obesity based on information from $\sim$ 1 million common genetic variations and then tested the polygenic score obtained from participants in the CHRIS study. GPS analysis of the CHRIS study did not provide any new information, but strengthened previously known associations, and also helped to assess which individuals have a high risk of becoming obese in our specific population study. This analysis could have important potential implications for clinical medicine, because it could allow us to identify individuals at risk of developing obesity before the condition manifests itself. Unfortunately, in the CHRIS study, we do not have a follow-up study so we cannot carry out any prevention, thus we have made predictions but we do not know if the predictions will be correct or not. In the future, it could also be interesting to analyse individuals with a low polygenic score, because it is possible that they are more likely to have monogenic causes of disease.

Chapter 4

Metabolomic profiling of obesity

4.1 Introduction

In this chapter, we investigate the metabolomic profiles of obese and metabolically healthy/unhealthy obese participants in the CHRIS study. Metabolomics is an appropriate approach for uncovering the differences in metabolism between each group of individuals [125], while the characterisation of metabolites associated with obesity and metabolically healthy/unhealthy obesity can provide an insight into the mechanisms leading to this disease and its associated consequences [126]. Nevertheless, characterisation of the metabolic mechanisms underlying health disparity in metabolic diseases is still lacking and we wish to know more, especially as regards our population study.

Performing longitudinal studies of large cohorts, as in the CHRIS study, allows accurate identification and quantification of the different metabo-

lites in serum biosamples, using a targeted quantitative approach such as that provided by Biocrates. The Biocrates AbsoluteIDQ® p180 kit [127] (Biocrates Life Sciences AG, Innsbruck, Austria) was used to collect targeted metabolomics data on serum samples in the CHRIS baseline [1] [128]. The kit covers multiple compound classes involved in various central metabolic processes: 21 amino acids, 21 biogenic amines, 1 monosaccharide, 40 acylcarnitines, 90 glycerophospholipids (76 phosphatidylcholines and 14 lysophosphatidylcholines), and 15 sphingolipids. It has already been applied to many other studies of human serum such as EPIC (European Prospective Investigation into Cancer and Nutrition) [129, 130] and KORA (Cooperative Health Research in the Region of Augsburg) [131].

The use of metabolomics data in the CHRIS study allowed us to achieve three goals: to replicate previous analysis performed with different cohorts (as a sort of validation study) but with a larger data set; to better characterise the CHRIS metabolome; and to derive predictive models to discriminate obesity subtypes.

4.2 Materials and methods

The general workflow characterising the acquisition and use of metabolic data is complex and requires different skills for data acquisition and data analysis. We summarise all the basic steps in Figure 4.1 and describe sample preparation and data acquisition in Appendix A.

After data acquisition, Biocrates-based targeted metabolomics data from

[1] Serum collected after fasting (in the vast majority of cases).

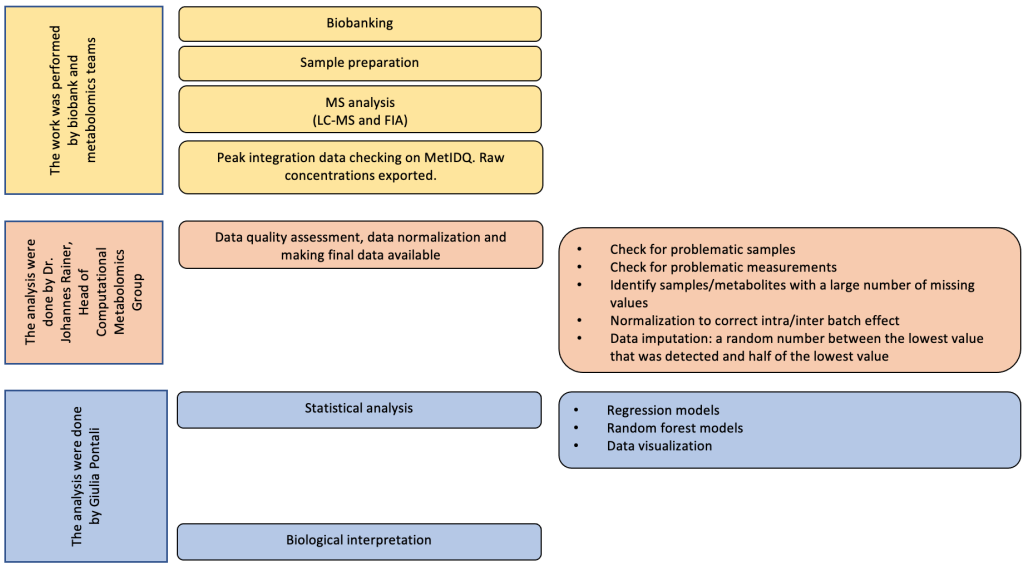

Figure 4.1: Main steps in obtaining metabolomics data.
LC-MS: liquid chromatography-mass spectrometry; FIA: flow injection analysis

the CHRIS study was processed with the BioCHRIStes R package [2] establishing a connection to a BioCHRIStes database containing Biocrates targeted metabolomics data for the CHRIS population study. Data for the $\sim$ 7 000 CHRIS participants were bundled in the biochristes7500 R package, providing metabolite concentrations for CHRIS study participants. The data were then normalised based on quality control (QC) samples to remove any between-batch differences. As a result of normalisation, the average coefficient of variation (CV) across QC samples was reduced, without affecting the CV of study samples [128]. To perform data analysis, we should have a data matrix without missing information. For this reason, we assumed that most

[2]BioCHRIStes is an R package written by Dr. Johannes Reiner. Use of this package is exclusive to EURAC.

of the missing values (6 634 out of 1 208 550) were sign that the metabolites were below the detection limit. In these cases, the missing values were replaced with random numbers taken from the uniform distribution, from half of the smallest measured value to the smallest measured value for that analyte. Uniform distribution was defined with two parameters (minimum and maximum) and all values between them had an equal probability of being selected. Then, we conducted all analyses on log2-transformed metabolite concentrations [128].

4.2.1 Clustering of analytes

After data quality checking and normalisation we had 175 metabolites (see Appendix B) quantified in 6 872 CHRIS participants, 1 087 of whom obese: 775 metabolically unhealthy obese; 281 metabolically healthy obese. Before identifying markers that discriminate obesity subgroups, we investigated the correlation between the metabolites using the Pearson correlation coefficient. If the degree of correlation between two or more independent metabolites is high (between + 0.4 and + 1), there is a multicollinearity problem reducing the precision of the estimated coefficients, which weakens the statistical power of the regression model (see Figure 4.2). So, it is always worth paying a little more attention during exploratory data analysis to uncover and address multicollinearity.

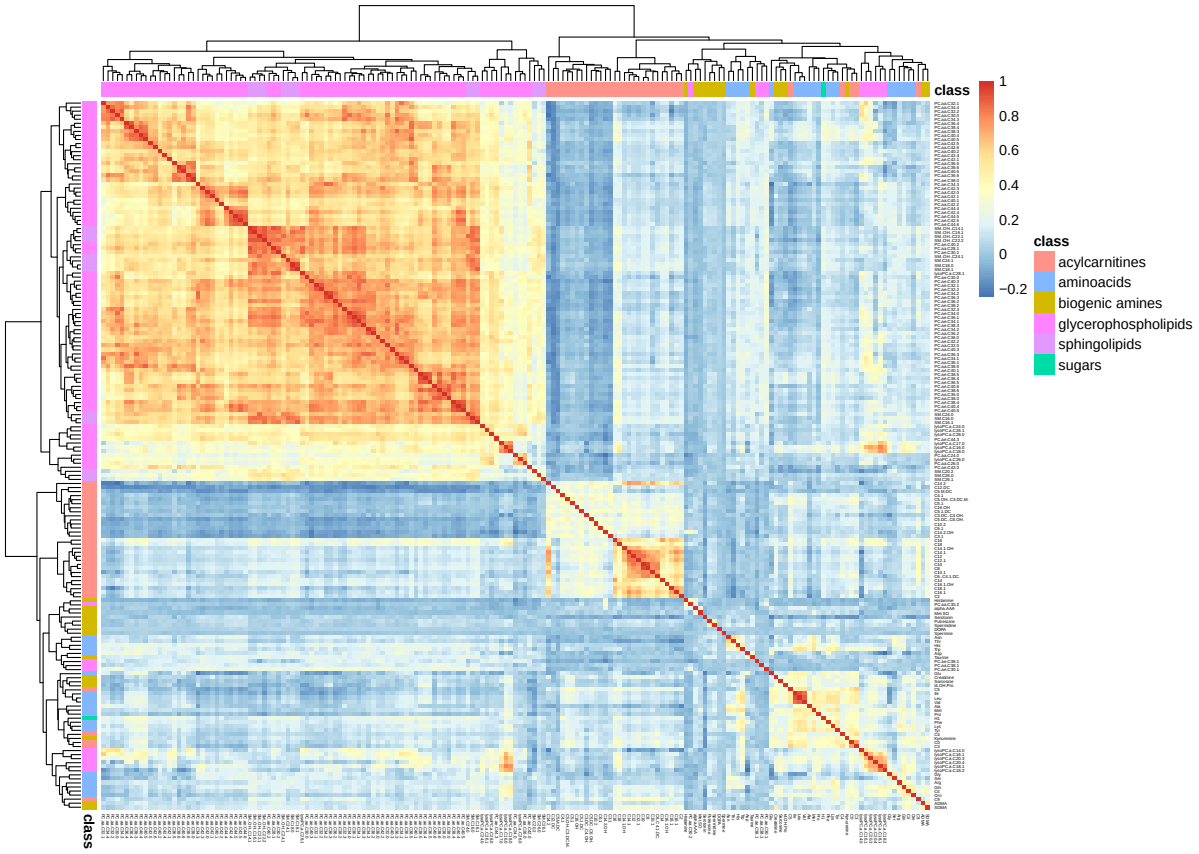

Figure 4.2: Hierarchical clustering of correlated/anti-correlated metabolites. Correlation matrix of the 175 metabolites studied. Cross-sectional correlation colours represent Pearson correlation coefficients

The most correlated metabolites came from glycerophospholipid and sphingolipid classes, due to their similarities, as both classes indeed contain fatty acids [132]. Furthermore, branched-chain amino acids (BCAAs), valine, leucine and isoleucine clustered together. These are essential amino acids with protein anabolic properties carrying a range of signalling functions, especially by activating the mTOR (mammalian target of rapamycin) [133] signalling pathway which is essential for initiation of protein synthesis [134].

Several acylcarnitines clustered together and they are anti-correlated with the major lipid cluster. This was expected because an essential role of acylcarnitines is to regulate the balance of intracellular sugar and lipid metabolism [135]. As a key component of cell metabolism, acylcarnitines play an important role in cell physiological activities [136].

4.2.2 Statistical analysis

To characterise the metabolites associated with obesity and metabolically healthy/unhealthy obesity, logistic regression models were fitted separately for each metabolite using obesity status (obese *vs* not obese, metabolically unhealthy obese *vs* not metabolically unhealthy obese, metabolically healthy obese *vs* not metabolically healthy obese, and metabolically unhealthy obese *vs* metabolically healthy obese) as a dependent variable and the concentration of each individual metabolite, sex, age, and fasting status[3] as independent variables.

By using the Bonferroni correction method, p-values from all analyses were adjusted for multiple testing. Then, to obtain comparable coefficients, prior to logistic modelling all the variables were standardised to zero mean and unit variance[4]. In autoscaled data, a difference of 1 corresponds to a standard deviation of 1 and is called *effect size*.

Our next step was to define the criteria for selection of significant metabolites. We evaluated whether we could combine the criteria for statistical sig-

[3]Binary fasting status was retrieved from self-reported fasting information, with 6415 participants declaring that they had not eaten within 12 h of blood sampling while 455 had eaten (2 participants were missing).

[4]Note that this does not influence the linear models since they are affine equivariant.

nificance (the p-value) with another criterion that would allow us to establish the metabolites most strongly associated with obesity status. So we required the difference in the abundances of obesity subtypes to be at least twice as large as the coefficient of variation for quality control samples. QC CHRIS pool samples were evaluated for the coefficient of variation, representing the technical variability observed for a particular metabolite. This method for assessing which metabolites are associated with the trait of interest differs from the methods presented in the literature, as it is more rigorous and allows us to describe the data in a more solid way and with fewer false positives.

We also ran the random forest algorithm. We proceeded in the following way for the entire data set: we randomly chose data for the test and the remainder for a training set, then we built a random forest model on the training set and lastly we estimated performance with the test. We repeated the procedure 10 times (10 fold cross-validation). Cross validation ensures all samples will appear in the training and test sets, so 100% of them are used at some point for training and testing. To evaluate the overall performance of our models, we averaged 10-fold validations.

4.3 Results

The logistic regression and random forest results were analysed, thus investigating how many and which metabolites discriminated between obese and not obese, metabolically unhealthy obese and not metabolically unhealthy obese, metabolically healthy obese and not metabolically healthy obese, and metabolically healthy obese and metabolically unhealthy obese

for each model.

4.3.1 Logistic regression

Model 1: Discriminates obese from not obese using binary obesity status as a dependent variable and one of the 175 metabolites at a time, sex, age and fasting status a independent variables. The model identified 91 out of 175 significant metabolites (see Figure 4.3). Among the metabolites with a stronger effect, we found glutamate as well as the branched-chain amino acids (BCAAs) group of three essential amino acids (leucine, isoleucine and valine) but also lysophosphatidylcholine, phenylalanine and tyrosine. Several independent studies have linked BCAAs to the well-known consequences of obesity: insulin resistance and diabetes [137]. Aromatic amino acids especially tyrosine and phenylalanine were strongly related to BMI and insulin resistance [138]. Lysophosphatidylcholine acyl C18:2, lysophosphatidylcholine acyl C18:1 and lysophosphatidylcholine acyl C17:0 were negatively correlated with BMI, in agreement with several other studies [139] while phosphatidylcholine acyl-alkyl C38:3 was also positively associated with BMI. These results are in line with the findings published in 2021 by the University of Milan's Department of Clinical Sciences and Community Health [140].

Model 2: Discriminates metabolically unhealthy obese from not metabolically unhealthy obese using binary metabolically unhealthy obesity status as a dependent variable and one of the 175 metabolites at a time, sex, age, and fasting status as independent variables. The model identified 91 out of 175 significant metabolites (see Figure 4.4). We found glutamate, BCAAs, phenylalanine and tyrosine. These results were similar to those obtained in

the previous model (see Figure 4.3). Similar results were expected as the majority of obese people were metabolically unhealthy obese (775 out of 1 087 obese people were metabolically unhealthy obese).

Model 3: Discriminates metabolically healthy obese from not metabolically healthy obese using binary metabolically healthy obesity status as a dependent variable and one of the 175 metabolites at a time, sex, age, and fasting status as independent variable. The model identified 33 out of 175 significant metabolites (see Figure 4.5). Lysophosphatidylcholines were negatively correlated with healthy obesity. These results reflect the biology because the plasma lipidome can be altered in obese unhealthy individuals due to several diseases such as type 2 diabetes and coronary artery disease [141].

Model 4: Discriminates metabolically unhealthy obese from metabolically healthy obese using binary metabolically healthy/unhealthy obesity status as a dependent variable and one of the 175 metabolites at a time, sex, age, and fasting status as independent variables. This model did not take into account the metabolic data for 6 872 participants but only the data of metabolically healthy and unhealthy obese participants. We therefore had 1 056 participants in total, of whom 775 were metabolically unhealthy obese and 281 metabolically healthy. The model identified 17 out of 175 significant metabolites. Looking at the seventeen metabolites (see Figure 4.6) we identified significant associations with the essential branched-chain amino acids isoleucine, leucine and valine which had previously been associated with obesity [140] and metabolic health [142, 143]. In agreement with these previous studies, we observed that the metabolically unhealthy obese phenotype is associated with higher levels of branched-chain amino acids. Several sphingomyelines

are also significantly associated, and sphingolipid metabolism is known to be affected in obese subjects [144]. We also saw an effect of propenoylcarnitine, a short chain acylcarnitine. Another short chain acylcarnitine (propionylcarnitine) had already been associated with metabolically unhealthy central obesity in a study with a small sample size [145].

In each model, the most significant metabolites included glutamate, which is a by-product of branched-chain-amino-acid (BCAA) catabolism and was shown to increase in obese individuals [146], BCAAs, which play a critical role in the regulation of energy homeostasis, nutrition metabolism, gut health, immunity and disease, and contribute to impaired glucose metabolism, especially in obese individuals [133], lysophosphatidylcholines, which are significantly negatively associated with obesity and related abnormal metabolism [141] and sphingolipids, involved in obesity-mediated inflammation and cardiovascular diseases [147].

Thanks to these four models we also found metabolites that were unique to the obese, metabolically unhealthy obese, and metabolically healthy obese (see Figure 4.7).

4.3.2 Random forest

In this section, in addition to obesity status/health, age, sex and fasting status, we also used visceral fat [5] as a predictor. Visceral fat is reported as a characteristic of the MUO phenotype [148] and it would be of interest to know if it is relevant for classification.

Model 1: Discriminates obese from not obese. The results are shown

[5]Visceral fat is the fat that surrounds abdominal organs.

in Figure 4.8. The model was able to discriminate between obese individuals and not obese individuals, and it is interesting to note that the most important predictors were glutamate, BCAAs, aromatic amino acids and lysophosphatidylcholine, those present in *logistic regression model 1*. Furthermore, age as a predictor was associated with increased obesity, being a major contributor to insulin resistance and metabolic syndrome [149]. The addition of visceral fat as a predictor (see Figure 4.9) increased the performance of the model and was the most important variable. A summary of prediction results for the classification model is presented in Appendix C.

Model 2: Discriminates metabolically unhealthy obese from not metabolically unhealthy obese. The results are shown in Figure 4.10 and were comparable with those obtained in the previous model (see Figure 4.8), because, as already explained in *logistic regression model 2* above, the majority of obese people were metabolically unhealthy obese (775 out of 1 087 obese people were metabolically unhealthy obese). The model was able to discriminate between metabolically unhealthy obese individuals and not metabolically unhealthy obese individuals, and the addition of visceral fat improved the model (see Figure 4.11). A summary of prediction results for the classification model is presented in Appendix C.

Model 3: Discriminates metabolically healthy obese from not metabolically healthy obese. The results are shown in Figure 4.12. The performance of the model was lower than the two models above because the dataset used was unbalanced: 281 metabolically healthy individuals vs 6591 not metabolically healthy individuals. In this model, visceral fat again improved the performance of the model (see Figure 4.13). A summary of prediction results

for the classification model is presented in Appendix C.

Model 4: Discriminates metabolically unhealthy obese from metabolically healthy obese. The results are shown in Figure 4.14. Including visceral fat (see Figure 4.15) did not improve model performance. Despite this, in both models (with and without visceral fat), the most important variable was age, confirming the data presented in *chapter 2*, namely MHO is a trait that can be transient with increasing age [100] due to the likely development of cardiometabolic abnormalities over time [101].

Of the 17 significant compounds in *logistic regression model 4* (see Figure 4.6), eight are among the top 20 important features in random forest model 4 (see Figure 4.15). These include glutamate, the branched-chain amino acid and isoleucine, which are associated with higher levels in MUO. Among these overlapping eight compounds we also observed three phosphatidylcholines (PC aa C32:1, PC ae C34:3 and PC ae C34:2) and a sphingomyelin (SM (OH) C22:2). A tendency for MHO to be associated with higher levels of phospholipids composed of more unsaturated fatty acids (PC ae C34:3, PC ae C34:2, SM (OH) C22:2), and lower levels of phospholipids with more saturated fatty acids (PC aa C32:1) was noticeable in relation to the MUO phenotype. In the future it would be of interest to investigate the potential relationship between saturated, monounsaturated and polyunsaturated fatty acid dietary intake and MUO/MHO phenotypes in the CHRIS study, and further clarify the role of dietary fat composition in metabolic health [150].

A summary of prediction results for the classification model is presented in Appendix C.

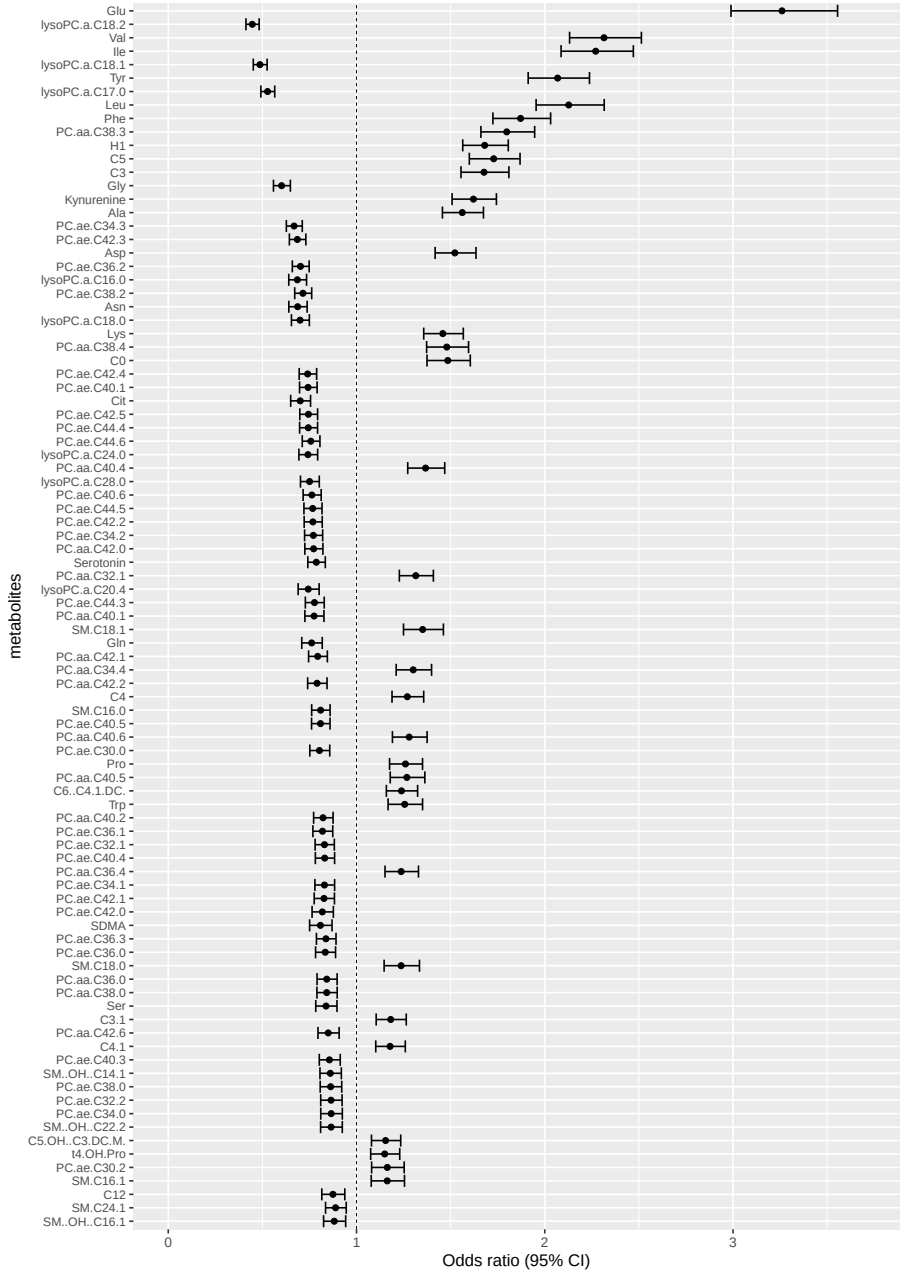

Figure 4.3: Significant metabolites associated with obese - not obese. The odds ratio is the standardised effect size of the log-transformed units of each metabolite

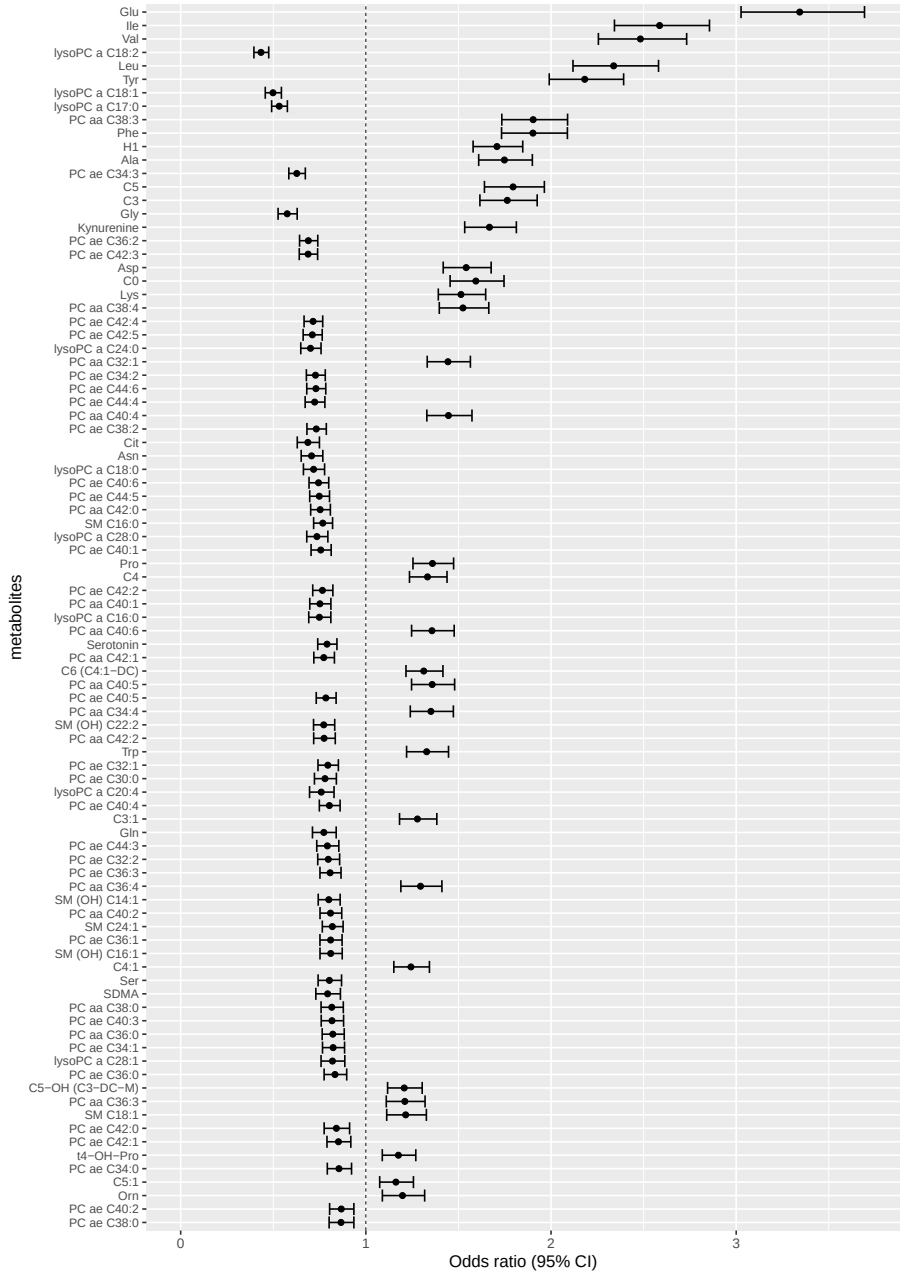

Figure 4.4: Significant metabolites associated with metabolically unhealthy - not metabolically unhealthy obese. The odds ratio is the standardised effect size of the log-transformed units of each metabolite

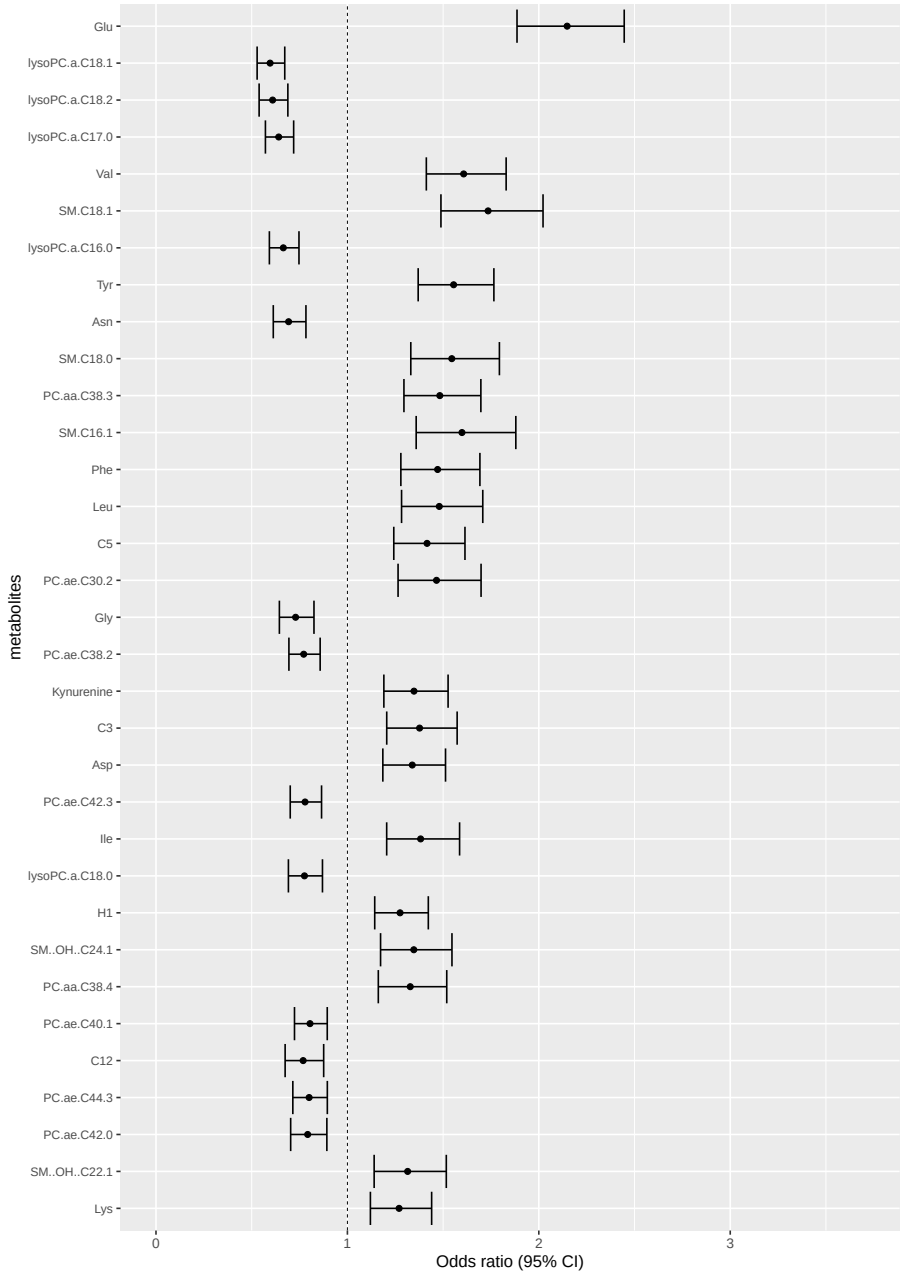

Figure 4.5: Significant metabolites associated with metabolically healthy - not metabolically healthy obesity. The odds ratio is the standardised effect size of the log-transformed units of each metabolite

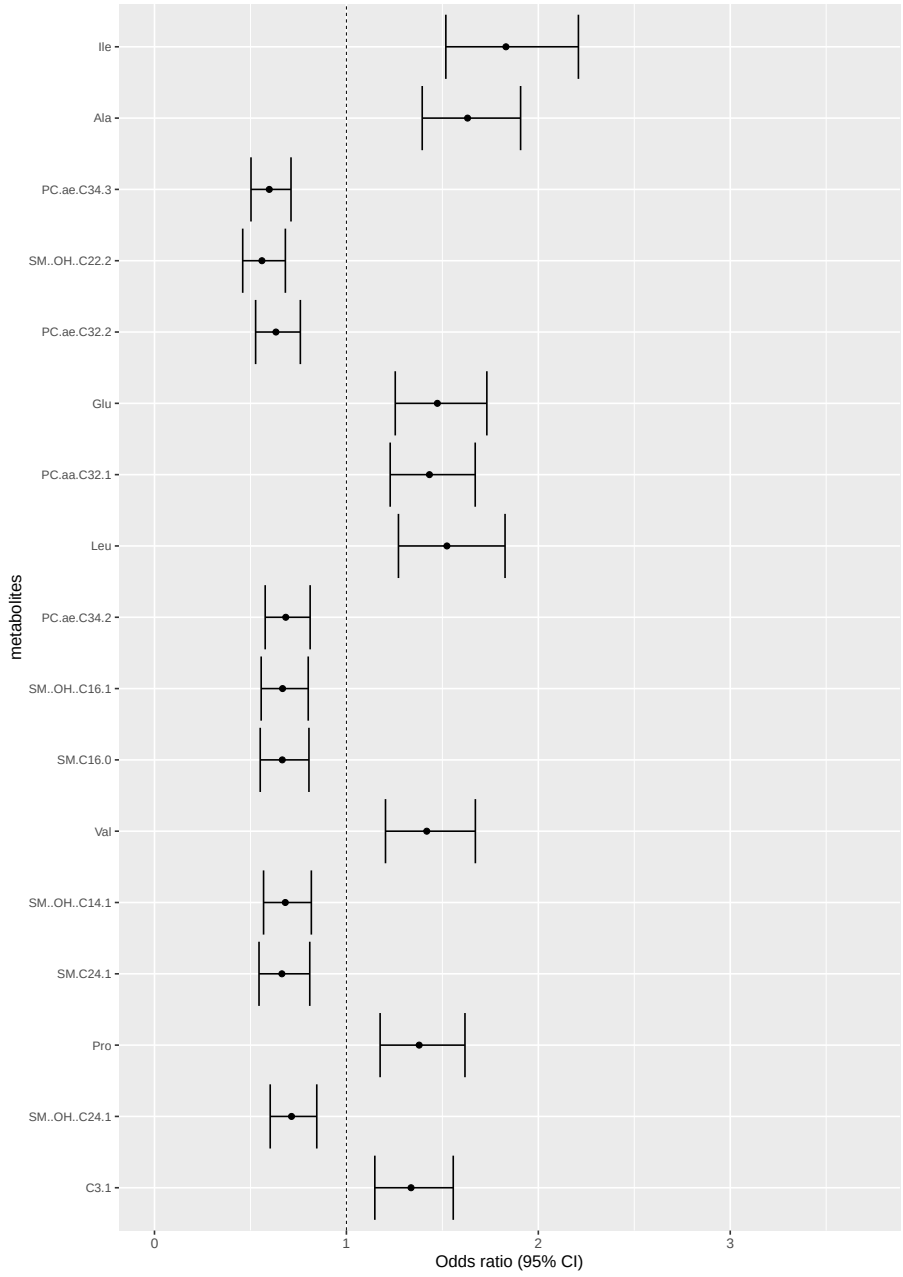

Figure 4.6: Significant metabolites associated with metabolically unhealthy - metabolically healthy obesity. The odds ratio is the standardised effect size of the log-transformed units of each metabolite

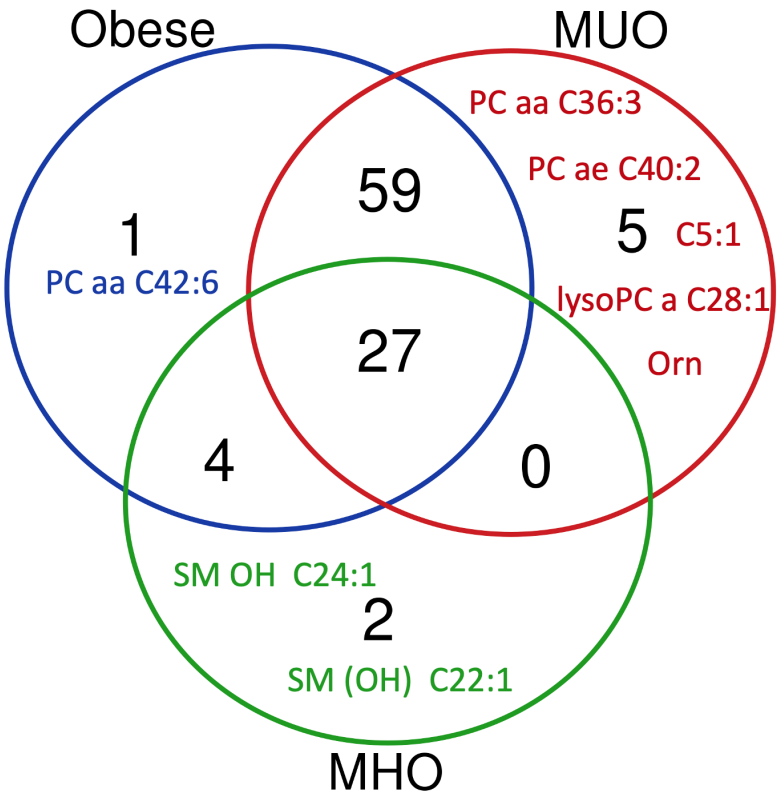

Figure 4.7: The Venn Diagram shows the number of metabolites shared and not shared by obese, MUO and MHO

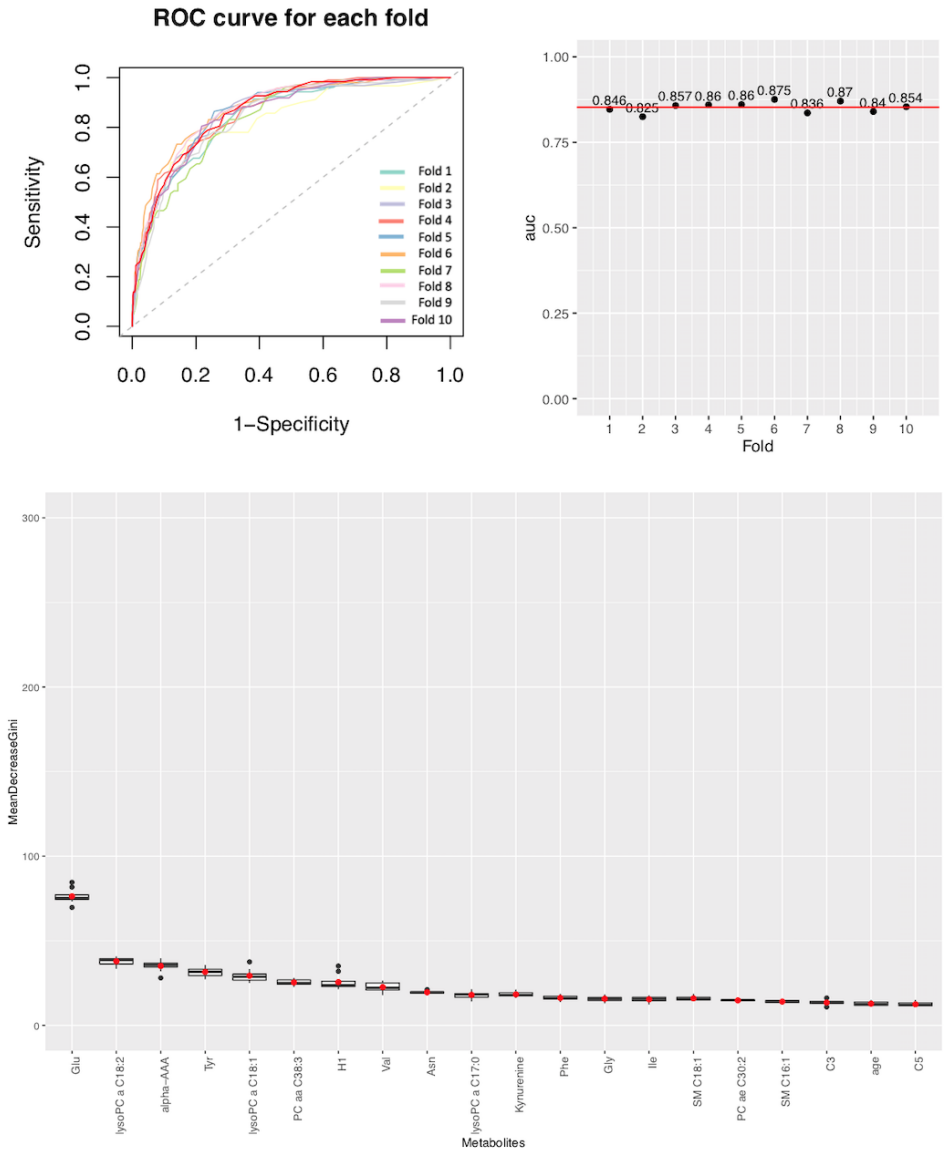

Figure 4.8: Discriminating obese status. The box plot represents the mean decrease Gini score for the top 20 variables. The red line in the AUC and the red dots on the mean decrease Gini represent the mean

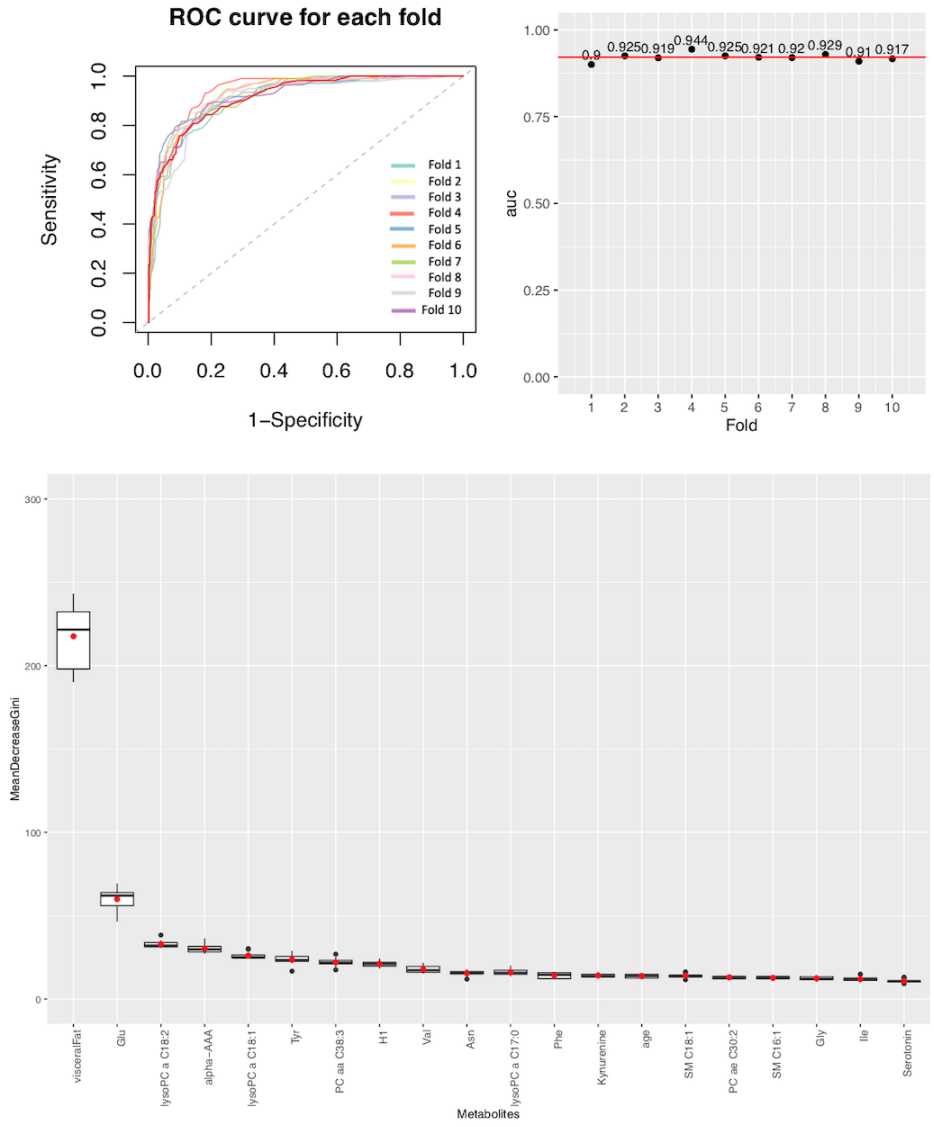

Figure 4.9: Discriminating obese status with the addition of visceral fat. The box plot represents the mean decrese Gini score for the top 20 variables. The red line in the AUC and the red dots on the mean decrease Gini represent the mean

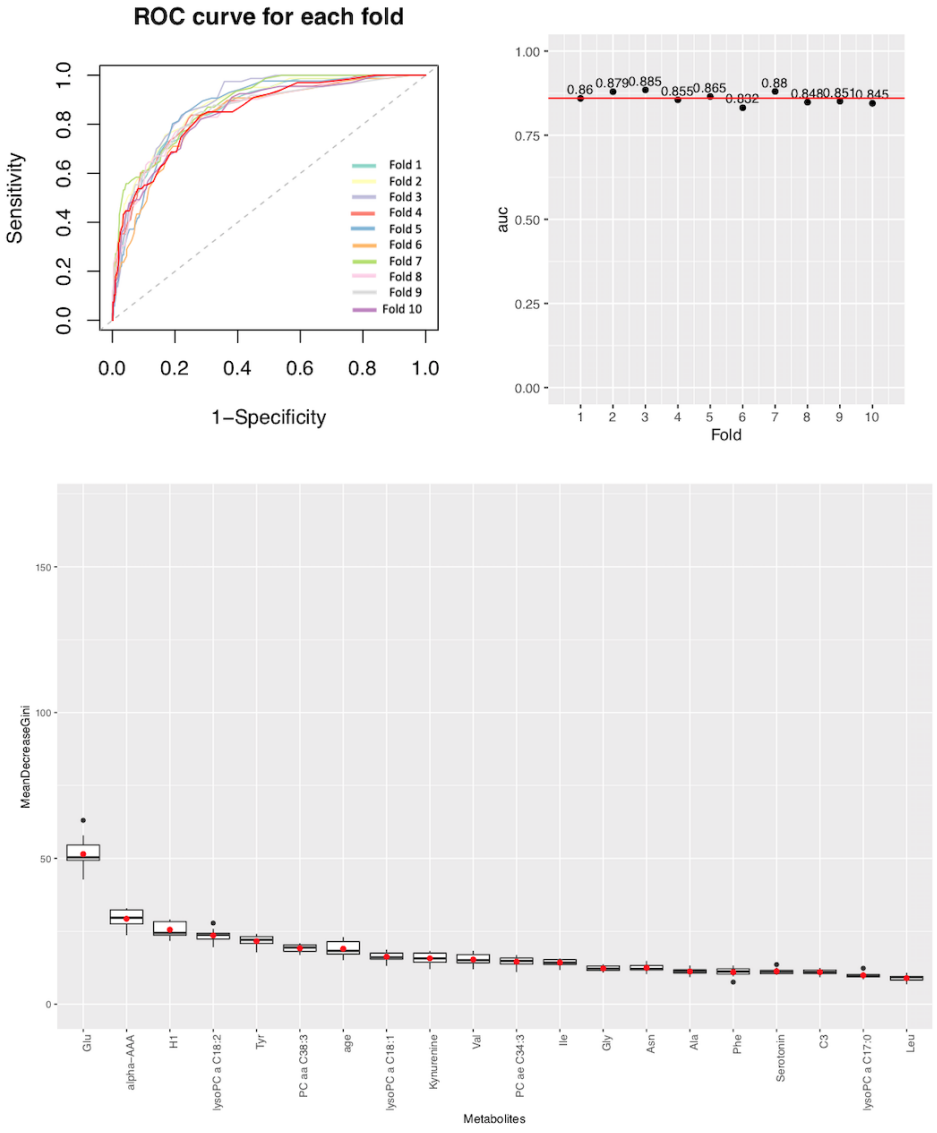

Figure 4.10: Discriminating metabolic unhealthy obese status. The box plot represents the mean decrese Gini score for the top 20 variables. The red line in the AUC and the red dots on the mean decrease Gini, represent the mean

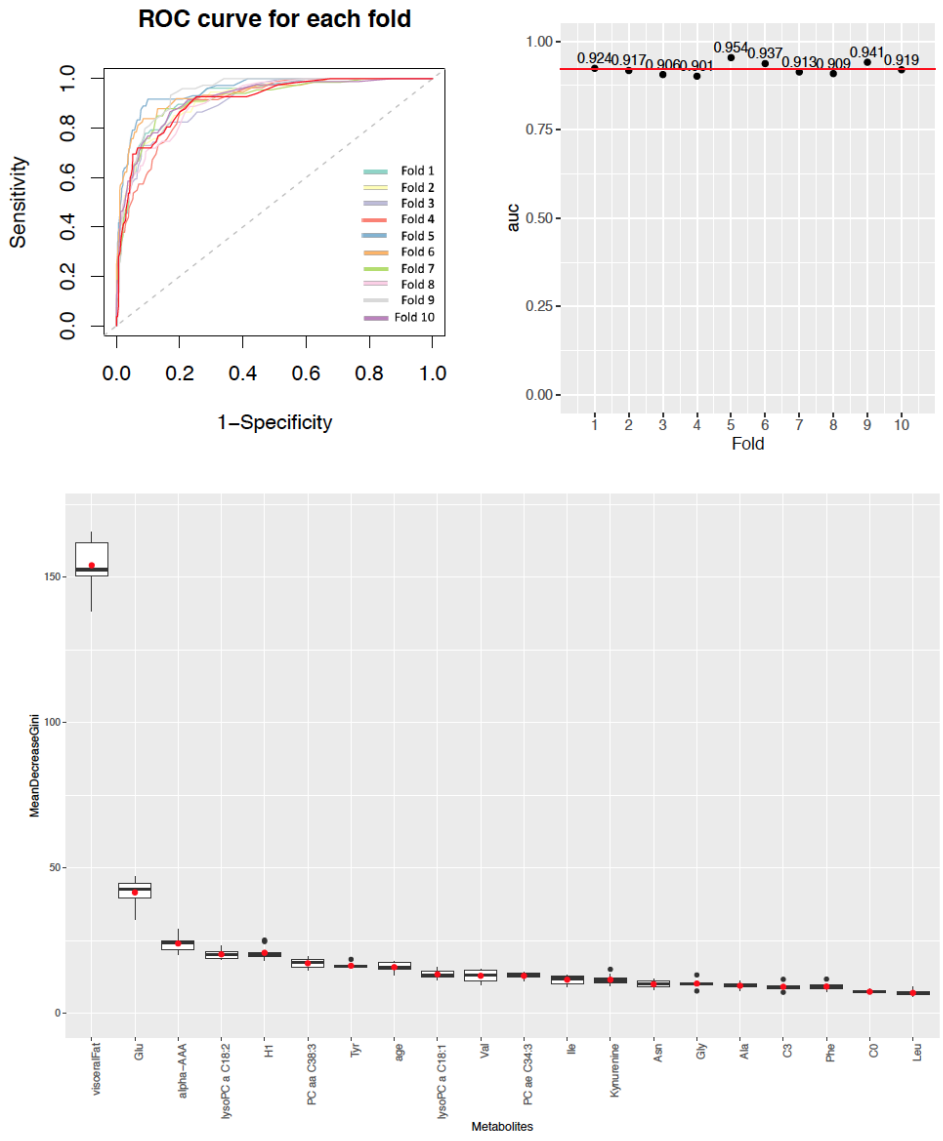

Figure 4.11: Discriminating metabolic unhealthy obese status with the adding of visceral fat. The box plot represents the mean decrese Gini score for the top 20 variables. The red line in the AUC and the red dots on the mean decrease Gini, represent the mean

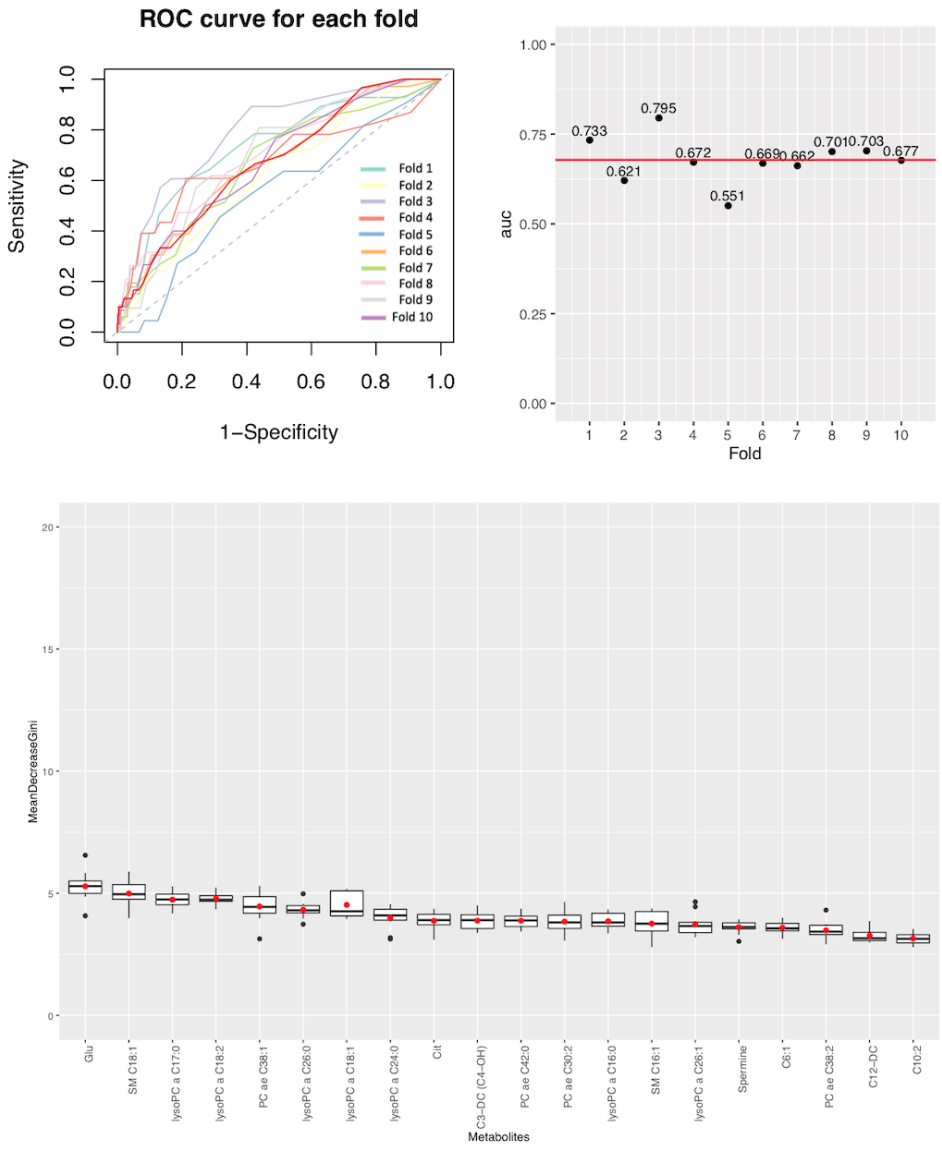

Figure 4.12: Discriminating metabolic healthy obese status. The box plot represents the mean decrease Gini score for the top 20 variables. The red line in the AUC and the red dots on the mean decrease Gini represent the mean

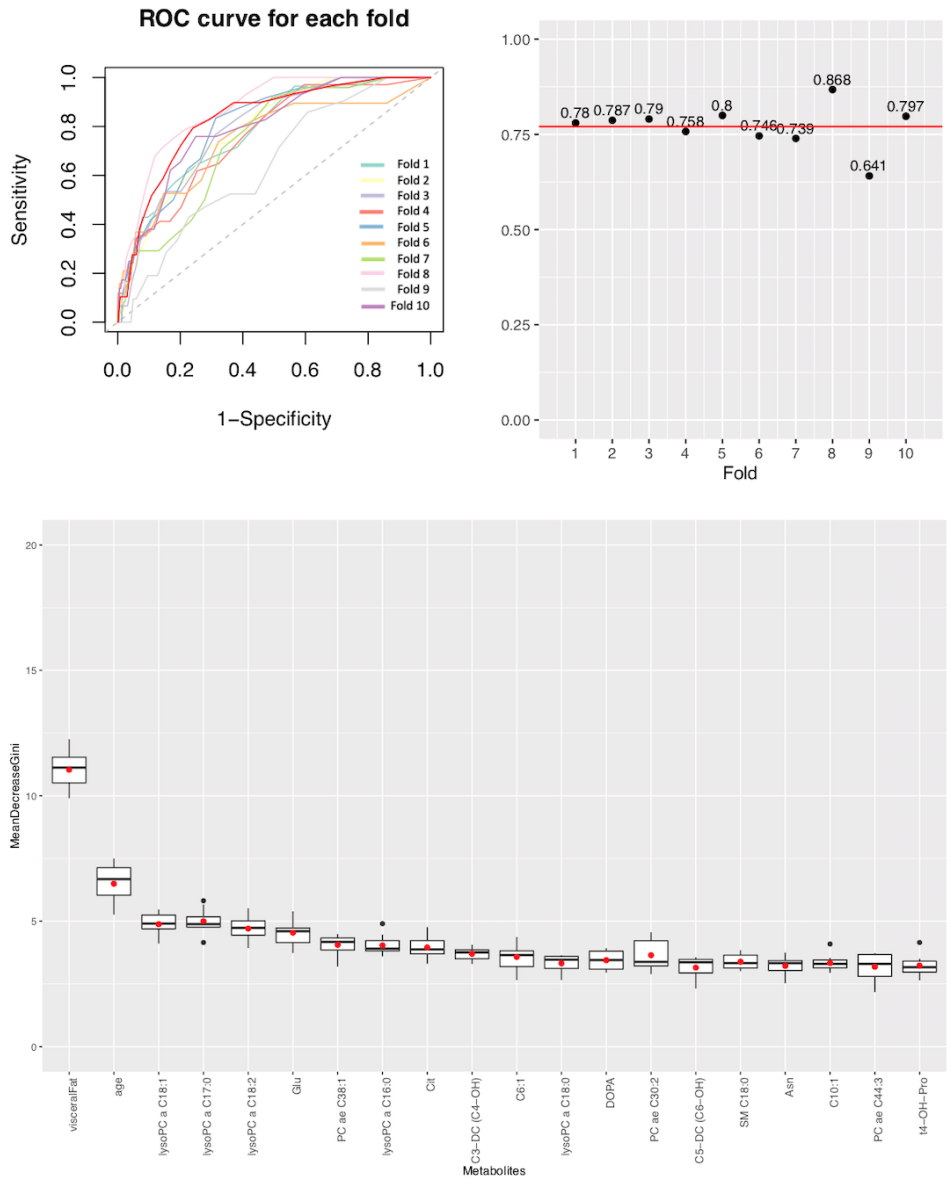

Figure 4.13: Discriminating metabolic healthy obese status with the addition of visceral fat. The box plot represents the mean decrease Gini score for the top 20 variables. The red line in the AUC and the red dots on the mean decrease Gini represent the mean

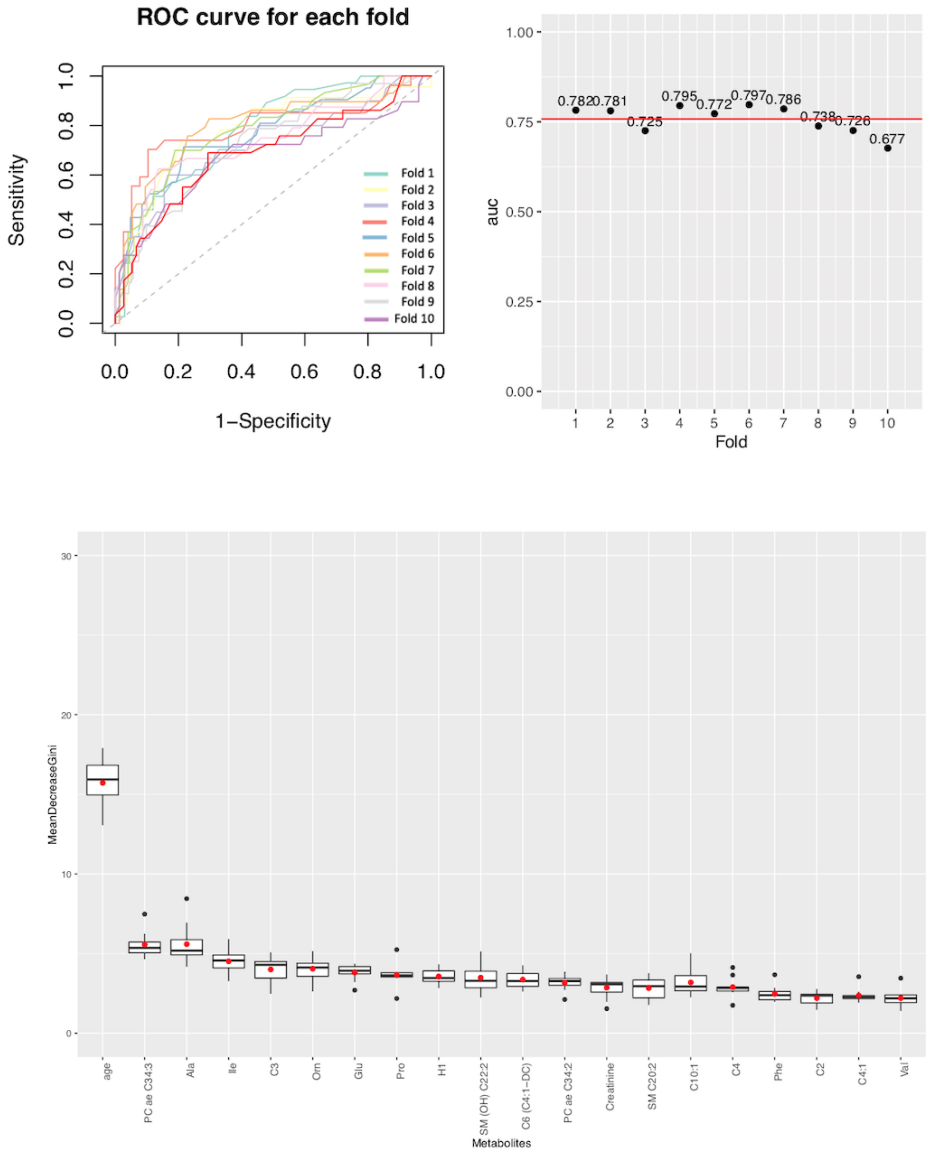

Figure 4.14: Discriminating metabolic unhealthy - healthy obese status. The box plot represents the mean decrease Gini score for the top 20 variables. The red line in the AUC and the red dots on the mean decrease Gini represent the mean

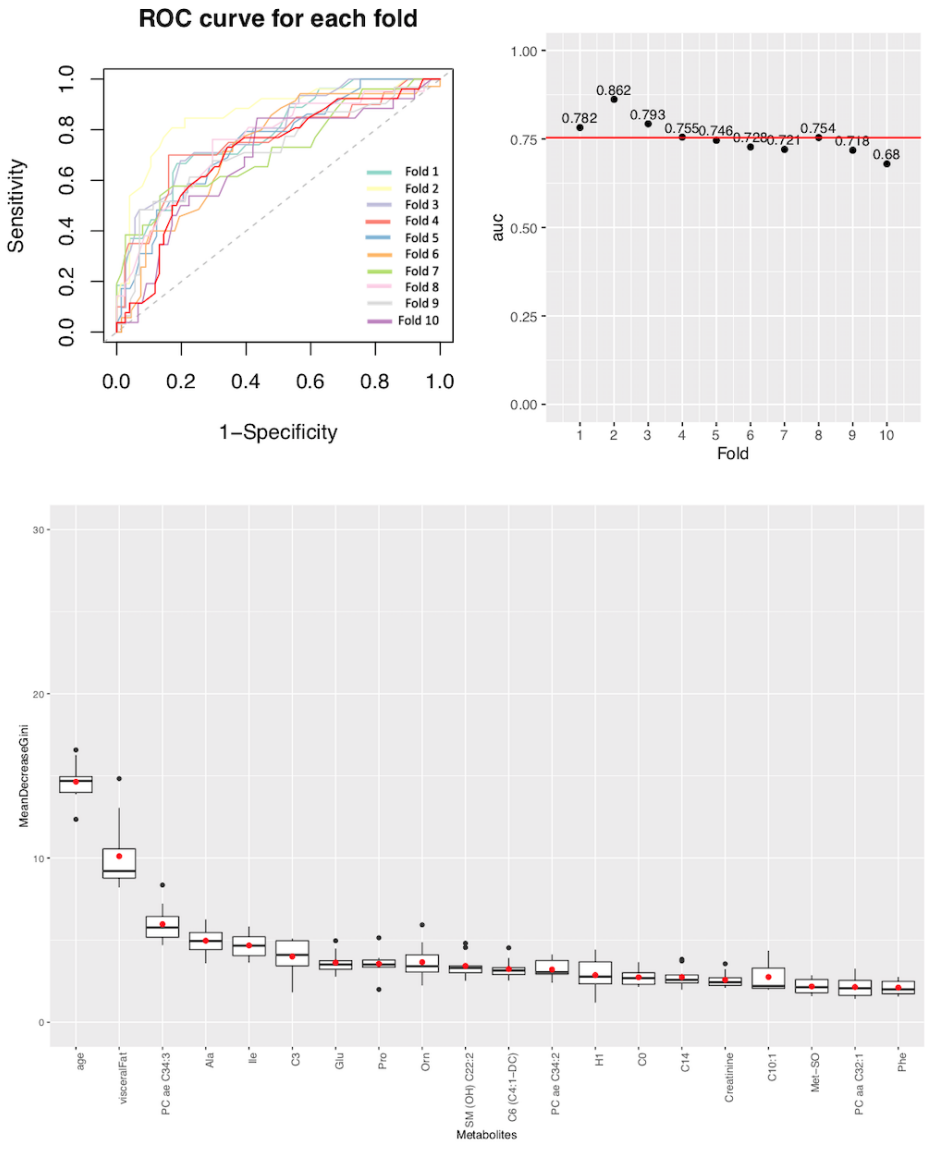

Figure 4.15: Discriminating metabolic unhealthy - healthy obese status with the addition of visceral fat. The box plot represents the mean decrease Gini score for the top 20 variables. The red line in the AUC and the red dots on the mean decrease Gini represent the mean

4.4 Conclusion

The results of our analysis underline the relationships between obesity status and obesity health, showing signs of alteration in the regulation of energy homeostasis, nutrition metabolism, and disease thanks to the significant metabolites from each model: BCAAs (some of these biomarkers are often associated with metabolically unfavourable markers, especially lipid metabolism, and may be useful for distinguishing obese individuals), lysophosphatidylcholines (markers linked to cardiovascular disease) and sphingolipids (fat metabolism and obesity-related complications are affected by sphingolipids).

Based on metabolomics, the CHRIS dataset allowed us to perform the largest study of metabolically healthy and metabolically unhealthy obese and this underlines the strength of the study. Additionally, we were able to depict independent metabolite patterns in metabolically healthy and metabolically unhealthy obese participants. However, our study, had some limitations. The models were not adjusted for medication, smoking and alcohol intake. Future work should explore their potential effect on the metabolomics profiles, bearing in mind that while the medication data in the CHRIS baseline is extensive and relatively reliable, smoking and alcohol intake are self reported and subject to strong bias.

We discovered which metabolites are associated with obesity and the different type of obesity, but we did not test causality. In the future, it would be interesting to study the cause and effect between metabolites and obesity traits to discover if a change in one metabolite causes a change in the obesity trait or vice-versa, with established approaches such as Mendelian

randomization.

This study therefore evaluated the metabolomic profile of the CHRIS cohort, composed of obese and metabolically healthy/unhealthy obese subjects, using a validated targeted approach.

Chapter 5

Conclusions

This PhD thesis gives a detailed overview of obesity in the CHRIS population. Thanks to the data available I performed a broad analysis, covering genetics, family history and metabolomics. I relied on the CHRIS study data to develop models for stratifying obesity subtypes. More specifically I implemented logistic regression models to explore associations and I trained and tested classifiers for predicting MUO and MHO based on metabolomic signatures and identified relevant metabolites for obesity stratification. In agreement with previous studies [140, 142, 143] I report significant associations between metabolically unhealthy obesity and higher levels of essential branched-chain amino acids (BCAA), including isoleucine, which is also among the most important features for classification with random forest models. These results confirm the association between isoleucine and metabolic health in obese individuals, based on the largest sample size to date. It has been reported that genes related to BCAA catabolism are more down-regulated in the adipose tissue of MUO individuals in comparison to

MHO [151], resulting in the higher circulating levels of these amino acids in MUO. These higher BCAA levels might impact metabolic health by activating mTOR signalling, causing insulin resistance and impaired glucose metabolism [152], while isoleucine has recently been reported to partially mediate the relationship between visceral adipose tissue volume and cardiometabolic risk [142].

Obesity is a complex trait; future efforts should extend the genomics and metabolomics models to integrate new proteomics datasets available only recently, as well as life-style factors like diet and physical activity that have also been collected within the CHRIS study. As regards this last point, it would be of interest to investigate the role of diet on metabolic unhealthy obesity and to what extent this effect is mediated by the metabolic signatures of MUO identified within this project. More specifically, adherence to specific diets can be quantified to characterize dietary patterns and investigate their association with MUO. Mediation analysis can then be performed to investigate to what extent circulating levels of isoleucine mediate the association between these diet patterns and MUO. The work would greatly benefit once the CHRIS study metabolomics collection is extended to better overlap with the GA2LEN food frequency questionnaire, increasing the dataset size for such analysis. In addition, the ongoing CHRIS follow-up study is collecting updated health data on the CHRIS baseline participants. The new information will allow the development of predictive models for health outcomes, and better characterise and discriminate stable and transient metabolically healthy obese phenotypes, a key issue that requires further investigation.

Appendix A

Sample preparation and data acquisition

- Serum samples were prepared using the AbsoluteIDQ® p180 kit from Biocrates;
- Samples were handled according to a protocol provided by Biocrates;
- Sample aliquots were then transferred into two distinct 96-well plates and filled with LC-MS and FIA-specific dilution solutions;
- Sciex Analyst 1.7 software was used for instrument control and data collection;
- Each sample was analysed with absolute quantification. LC-MS/MS technology was used for amino acids and biogenic amines, while FIA-MS/MS was used for cylcarnitines, glycerophospholipids, sphingolipids and 1 sum of hexoses.

Appendix B

List of Metabolites

Analyte Name	**Analyte Class**	**Biochemical Name**
Ala	Amino acids	Alanine
Arg	Amino acids	Arginine
Asn	Amino acids	Asparagine
Asp	Amino acids	Aspartate
Cit	Amino acids	Citrulline
Gln	Amino acids	Glutamine
Glu	Amino acids	Glutamate
Gly	Amino acids	Glycine
His	Amino acids	Histidine
Ile	Amino acids	Isoleucine
Leu	Amino acids	Leucine
Lys	Amino acids	Lysine

Analyte Name	Analyte Class	Biochemical Name
Met	Amino acids	Methionine
Orn	Amino acids	Ornithine
Phe	Amino acids	Phenylalanine
Pro	Amino acids	Proline
Ser	Amino acids	Serine
Thr	Amino acids	Threonine
Trp	Amino acids	Tryptophan
Tyr	Amino acids	Tyrosine
Val	Amino acids	Valine
Ac-Orn	Biogenic amines	Acetylornithine
ADMA	Biogenic amines	Asymmetric dimethylarginine
alpha-AAA	Biogenic amines	Alpha-aminoadipic acid
c4-OH-Pro	Biogenic amines	Cis-4-hydroxyproline
Carnosine	Biogenic amines	Carnosine
Creatinine	Biogenic amines	Creatinine
DOPA	Biogenic amines	Dihydroxyphenylalanine
Dopamine	Biogenic amines	Dopamine
Histamine	Biogenic amines	Histamine
Kynurenine	Biogenic amines	Kynurenine
Met-SO	Biogenic amines	Methioninesulfoxide
Nitro-Tyr	Biogenic amines	Nitrotyrosine
PEA	Biogenic amines	Phenylethylamine

Analyte Name	Analyte Class	Biochemical Name
Sarcosine	Biogenic amines	Sarcosine
Serotonin	Biogenic amines	Serotonin
Spermidine	Biogenic amines	Spermidine
Spermine	Biogenic amines	Spermine
t4-OH-Pro	Biogenic amines	Trans-4-hydroxyproline
Taurine	Biogenic amines	Taurine
SDMA	Biogenic amines	Symmetric dimethylarginine
Putrescine	Biogenic amines	Putrescine
C0	Acylcarnitines	Carnitine
C10	Acylcarnitines	Decanoylcarnitine
C10:1	Acylcarnitines	Decenoylcarnitine
C10:2	Acylcarnitines	Decadienylcarnitine
C12	Acylcarnitines	Dodecanoylcarnitine
C12-DC	Acylcarnitines	Dodecanedioylcarnitine
C12:1	Acylcarnitines	Dodecenoylcarnitine
C14	Acylcarnitines	Tetradecanoylcarnitine
C14:1	Acylcarnitines	Tetradecanoylcarnitine
C14:1-OH	Acylcarnitines	Hydroxytetradecenoylcarnitine
C14:2	Acylcarnitines	Tetradecadienylcarnitine
C14:2-OH	Acylcarnitines	Hydroxytetradecadienylcarnitine
C16	Acylcarnitines	Hexadecanoylcarnitine
C16-OH	Acylcarnitines	Hydroxyhexadecanoylcarnitine

Analyte Name	Analyte Class	Biochemical Name
C16:1	Acylcarnitines	Hexadecenoylcarnitine
C16:1-OH	Acylcarnitines	Hydroxyhexadecenoylcarnitine
C16:2	Acylcarnitines	Hexadecadienylcarnitine
C16:2-OH	Acylcarnitines	Hydroxyhexadecadienylcarnitine
C18	Acylcarnitines	Octadecanoylcarnitine
C18:1	Acylcarnitines	Octadecenoylcarnitine
C18:1-OH	Acylcarnitines	Hydroxyoctadecenoylcarnitine
C18:2	Acylcarnitines	Octadecadienylcarnitine
C2	Acylcarnitines	Acetylcarnitine
C3	Acylcarnitines	Propionylcarnitine
C3-OH	Acylcarnitines	Hydroxypropionylcarnitine
C3:1	Acylcarnitines	Propenoylcarnitine
C4	Acylcarnitines	Butyrylcarnitine
C3-DC (C4-OH)	Acylcarnitines	Hydroxybutyrylcarnitine
C4:1	Acylcarnitines	Butenylcarnitine
C5	Acylcarnitines	Valerylcarnitine
C5-M-DC	Acylcarnitines	Methylglutarylcarnitine
C5-OH (C3-DC-M)	Acylcarnitines	Hydroxyvalerylcarnitine
C5:1	Acylcarnitines	Tiglylcarnitine
C5:1-DC	Acylcarnitines	Glutaconylcarnitine
C6 (C4:1-DC)	Acylcarnitines	Hexanoylcarnitine
C5-DC (C6-OH)	Acylcarnitines	Glutarylcarnitine

Analyte Name	Analyte Class	Biochemical Name
C6:1	Acylcarnitines	Hexenoylcarnitine
C7-DC	Acylcarnitines	Pimelylcarnitine
C8	Acylcarnitines	Octanoylcarnitine
C9	Acylcarnitines	Nonaylcarnitine
lysoPC a C14:0	Glycerophospholipids	Lysophosphatidylcholine-acyl-C14:0
lysoPC a C16:0	Glycerophospholipids	Lysophosphatidylcholine-acyl-C16:0
lysoPC a C16:1	Glycerophospholipids	Lysophosphatidylcholine-acyl-C16:1
lysoPC a C17:0	Glycerophospholipids	Lysophosphatidylcholine-acyl-C17:0
lysoPC a C18:0	Glycerophospholipids	Lysophosphatidylcholine-acyl-C18:0
lysoPC a C18:1	Glycerophospholipids	Lysophosphatidylcholine-acyl-C18:1
lysoPC a C18:2	Glycerophospholipids	Lysophosphatidylcholine-acyl-C18:2
lysoPC a C20:3	Glycerophospholipids	Lysophosphatidylcholine-acyl-C20:3
lysoPC a C20:4	Glycerophospholipids	Lysophosphatidylcholine-acyl-C20:4
lysoPC a C24:0	Glycerophospholipids	Lysophosphatidylcholine-acyl-C24:0
lysoPC a C26:0	Glycerophospholipids	Lysophosphatidylcholine-acyl-C26:0
lysoPC a C26:1	Glycerophospholipids	Lysophosphatidylcholine-acyl-C26:1
lysoPC a C28:0	Glycerophospholipids	Lysophosphatidylcholine-acyl-C28:0
lysoPC a C28:1	Glycerophospholipids	Lysophosphatidylcholine-acyl-C28:1
PC aa C24:0	Glycerophospholipids	Phosphatidylcholine-diacyl-C24:0
PC aa C26:0	Glycerophospholipids	Phosphatidylcholine-diacyl-C26:0
PC aa C28:1	Glycerophospholipids	Phosphatidylcholine-diacyl-C28:1
PC aa C30:0	Glycerophospholipids	Phosphatidylcholine-diacyl-C30:0

Analyte Name	Analyte Class	Biochemical Name
PC aa C30:2	Glycerophospholipids	Phosphatidylcholine-diacyl-C30:2
PC aa C32:0	Glycerophospholipids	Phosphatidylcholine-diacyl-C32:0
PC aa C32:1	Glycerophospholipids	Phosphatidylcholine-diacyl-C32:1
PC aa C32:2	Glycerophospholipids	Phosphatidylcholine-diacyl-C32:2
PC aa C32:3	Glycerophospholipids	Phosphatidylcholine-diacyl-C32:3
PC aa C34:1	Glycerophospholipids	Phosphatidylcholine-diacyl-C34:1
PC aa C34:2	Glycerophospholipids	Phosphatidylcholine-diacyl-C34:2
PC aa C34:3	Glycerophospholipids	Phosphatidylcholine-diacyl-C34:3
PC aa C34:4	Glycerophospholipids	Phosphatidylcholine-diacyl-C34:4
PC aa C36:0	Glycerophospholipids	Phosphatidylcholine-diacyl-C36:0
PC aa C36:1	Glycerophospholipids	Phosphatidylcholine-diacyl-C36:1
PC aa C36:2	Glycerophospholipids	Phosphatidylcholine-diacyl-C36:2
PC aa C36:3	Glycerophospholipids	Phosphatidylcholine-diacyl-C36:3
PC aa C36:4	Glycerophospholipids	Phosphatidylcholine-diacyl-C36:4
PC aa C36:5	Glycerophospholipids	Phosphatidylcholine-diacyl-C36:5
PC aa C36:6	Glycerophospholipids	Phosphatidylcholine-diacyl-C36:6
PC aa C38:0	Glycerophospholipids	Phosphatidylcholine-diacyl-C38:0
PC aa C38:1	Glycerophospholipids	Phosphatidylcholine-diacyl-C38:1
PC aa C38:3	Glycerophospholipids	Phosphatidylcholine-diacyl-C38:3
PC aa C38:4	Glycerophospholipids	Phosphatidylcholine-diacyl-C38:4
PC aa C38:5	Glycerophospholipids	Phosphatidylcholine-diacyl-C38:5
PC aa C38:6	Glycerophospholipids	Phosphatidylcholine-diacyl-C38:6

Analyte Name	Analyte Class	Biochemical Name
PC aa C40:1	Glycerophospholipids	Phosphatidylcholine-diacyl-C40:1
PC aa C40:2	Glycerophospholipids	Phosphatidylcholine-diacyl-C40:2
PC aa C40:3	Glycerophospholipids	Phosphatidylcholine-diacyl-C40:3
PC aa C40:4	Glycerophospholipids	Phosphatidylcholine-diacyl-C40:4
PC aa C40:5	Glycerophospholipids	Phosphatidylcholine-diacyl-C40:5
PC aa C40:6	Glycerophospholipids	Phosphatidylcholine-diacyl-C40:6
PC aa C42:0	Glycerophospholipids	Phosphatidylcholine-diacyl-C42:0
PC aa C42:1	Glycerophospholipids	Phosphatidylcholine-diacyl-C42:1
PC aa C42:2	Glycerophospholipids	Phosphatidylcholine-diacyl-C42:2
PC aa C42:4	Glycerophospholipids	Phosphatidylcholine-diacyl-C42:4
PC aa C42:5	Glycerophospholipids	Phosphatidylcholine-diacyl-C42:5
PC aa C42:6	Glycerophospholipids	Phosphatidylcholine-diacyl-C42:6
PC ae C30:0	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C30:0
PC ae C30:1	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C30:1
PC ae C30:2	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C30:2
PC ae C32:1	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C32:1
PC ae C32:2	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C32:2
PC ae C34:0	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C34:0
PC ae C34:1	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C34:1
PC ae C34:2	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C34:2
PC ae C34:3	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C34:3
PC ae C36:0	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C36:0

Analyte Name	Analyte Class	Biochemical Name
PC ae C36:1	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C36:1
PC ae C36:2	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C36:2
PC ae C36:3	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C36:3
PC ae C36:4	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C36:4
PC ae C36:5	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C36:5
PC ae C38:0	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C38:0
PC ae C38:1	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C38:1
PC ae C38:2	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C38:2
PC ae C38:3	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C38:3
PC ae C38:4	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C38:4
PC ae C38:5	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C38:5
PC ae C38:6	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C38:6
PC ae C40:1	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C40:1
PC ae C40:2	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C40:2
PC ae C40:3	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C40:3
PC ae C40:4	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C40:4
PC ae C40:5	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C40:5
PC ae C40:6	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C40:6
PC ae C42:0	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C42:0
PC ae C42:1	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C42:1
PC ae C42:2	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C42:2
PC ae C42:3	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C42:3

Analyte Name	Analyte Class	Biochemical Name
PC ae C42:4	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C42:4
PC ae C42:5	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C42:5
PC ae C44:3	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C44:3
PC ae C44:4	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C44:4
PC ae C44:5	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C44:5
PC ae C44:6	Glycerophospholipids	Phosphatidylcholine-acyl-alkyl-C44:6
SM (OH) C14:1	Sphingolipids	Hydroxysphingomyeline-C14:1
SM (OH) C16:1	Sphingolipids	Hydroxysphingomyeline-C16:1
SM (OH) C22:1	Sphingolipids	Hydroxysphingomyeline-C22:1
SM (OH) C22:2	Sphingolipids	Hydroxysphingomyeline-C22:2
SM (OH) C24:1	Sphingolipids	Hydroxysphingomyeline-C24:1
SM C16:0	Sphingolipids	Sphingomyeline-C16:0
SM C16:1	Sphingolipids	Sphingomyeline-C16:1
SM C18:0	Sphingolipids	Sphingomyeline-C18:0
SM C18:1	Sphingolipids	Sphingomyeline-C18:1
SM C20:2	Sphingolipids	Sphingomyeline-C20:2
SM C22:3	Sphingolipids	Sphingomyeline-C22:3
SM C24:0	Sphingolipids	Sphingomyeline-C24:0
SM C24:1	Sphingolipids	Sphingomyeline-C24:1
SM C26:0	Sphingolipids	Sphingomyeline-C26:0
SM C26:1	Sphingolipids	Sphingomyeline-C26:1
H1	Sugars	Hexose

Appendix C

Confusion matrix

We had a confusion matrix for each of the ten cross-validation runs. We summarised the ten runs in a single line by taking the medians of the true positive (tp), false positive (fp), true negative (tn) and false negative (fn).

Comparisons	**tp**	**fp**	**tn**	**fn**
Obese vs not obese	8.5	138	425.5	20
Obese vs not obese **with visceral fat**	85	88	482.5	14.5
Metabolically unhealthy obese vs not metabolically unhealthy obese	63.5	158.5	437	12
Metabolically unhealthy obese vs not metabolically unhealthy obese **with visceral fat**	64.5	101.5	487.5	8.5
Metabolically healthy obese vs not metabolically healthy obese	19.5	245	393	9
Metabolically healthy obese vs not metabolically healthy obese **with visceral fat**	22	204.5	437.5	4.5
Metabolically unhealthy obese vs metabolically healthy obese	19.5	18.5	56	8
Metabolically unhealthy obese vs metabolically healthy obese **with visceral fat**	22.5	24.5	49	6.5

Bibliography

[1] A. Hruby and F. B. Hu, "The epidemiology of obesity: A big picture," *Pharmacoeconomics*, vol. 33, no. 7, pp. 673–689, Jul. 2015.

[2] A. Mastrangelo, E. G. Armitage, A. García, and C. Barbas, "Metabolomics as a tool for drug discovery and personalised medicine. a review," *Curr. Top. Med. Chem.*, vol. 14, no. 23, pp. 2627–2636, 2014.

[3] C. Pattaro, M. Gögele, D. Mascalzoni, R. Melotti, C. Schwienbacher, A. De Grandi, L. Foco, Y. D'Elia, B. Linder, C. Fuchsberger, C. Minelli, C. Egger, L. S. Kofink, S. Zanigni, T. Schäfer, M. F. Facheris, S. V. Smárason, A. Rossini, A. A. Hicks, H. Weiss, and P. P. Pramstaller, "The cooperative health research in south tyrol (CHRIS) study: rationale, objectives, and preliminary results," *J. Transl. Med.*, vol. 13, no. 1, p. 348, Nov. 2015.

[4] G. Eknoyan, "Adolphe quetelet (1796-1874)–the average man and indices of obesity," pp. 47–51, Jan. 2008.

[5] C. M. Rotella and I. Dicembrini, "Measurement of body composition as a surrogate evaluation of energy balance in obese patients," *World*

J. Methodol., vol. 5, no. 1, pp. 1–9, Mar. 2015.

[6] L. J. Aronne and K. R. Segal, "Adiposity and fat distribution outcome measures: assessment and clinical implications," *Obes. Res.*, vol. 10 Suppl 1, no. S11, pp. 14S–21S, Nov. 2002.

[7] B. M. Herrera and C. M. Lindgren, "The genetics of obesity," *Curr. Diab. Rep.*, vol. 10, no. 6, pp. 498–505, Dec. 2010.

[8] K. Karastergiou, S. R. Smith, A. S. Greenberg, and S. K. Fried, "Sex differences in human adipose tissues - the biology of pear shape," *Biol. Sex Differ.*, vol. 3, no. 1, p. 13, May 2012.

[9] R. W. Taylor, I. E. Jones, S. M. Williams, and A. Goulding, "Evaluation of waist circumference, waist-to-hip ratio, and the conicity index as screening tools for high trunk fat mass, as measured by dual-energy x-ray absorptiometry, in children aged 3-19 y," *Am. J. Clin. Nutr.*, vol. 72, no. 2, pp. 490–495, Aug. 2000.

[10] M. Dalton, A. J. Cameron, P. Z. Zimmet, J. E. Shaw, D. Jolley, D. W. Dunstan, T. A. Welborn, and AusDiab Steering Committee, "Waist circumference, waist-hip ratio and body mass index and their correlation with cardiovascular disease risk factors in australian adults," *J. Intern. Med.*, vol. 254, no. 6, pp. 555–563, Dec. 2003.

[11] B. L. Wajchenberg, "Subcutaneous and visceral adipose tissue: their relation to the metabolic syndrome," *Endocr. Rev.*, vol. 21, no. 6, pp. 697–738, Dec. 2000.

[12] T. Lemos and D. Gallagher, "Current body composition measurement techniques," *Curr. Opin. Endocrinol. Diabetes Obes.*, vol. 24, no. 5, pp. 310–314, Oct. 2017.

[13] P. Deurenberg, "Limitations of the bioelectrical impedance method for the assessment of body fat in severe obesity," *Am. J. Clin. Nutr.*, vol. 64, no. 3 Suppl, pp. 449S–452S, Sep. 1996.

[14] L. Z. Coppini, D. L. Waitzberg, and A. C. L. Campos, "Limitations and validation of bioelectrical impedance analysis in morbidly obese patients," *Curr. Opin. Clin. Nutr. Metab. Care*, vol. 8, no. 3, pp. 329–332, May 2005.

[15] M. M. Ibrahim, *Obes. Rev.*, vol. 11, no. 1, pp. 11–18, Jan. 2010.

[16] C. Messina, C. G. Monaco, F. M. Ulivieri, F. Sardanelli, and L. M. Sconfienza, "Dual-energy x-ray absorptiometry body composition in patients with secondary osteoporosis," *Eur. J. Radiol.*, vol. 85, no. 8, pp. 1493–1498, Aug. 2016.

[17] L. C. Pescatori, E. Savarino, G. Mauri, E. Silvestri, M. Cariati, F. Sardanelli, and L. M. Sconfienza, "Quantification of visceral adipose tissue by computed tomography and magnetic resonance imaging: reproducibility and accuracy," *Radiol. Bras.*, vol. 52, no. 1, pp. 1–6, Jan. 2019.

[18] M. A. Green, "Do we need to think beyond BMI for estimating population-level health risks?" *J. Public Health (Oxf.)*, vol. 38, no. 1, pp. 192–193, Mar. 2016.

[19] "Obesity: preventing and managing the global epidemic. report of a WHO consultation," *World Health Organ. Tech. Rep. Ser.*, vol. 894, pp. i–xii, 1–253, 2000.

[20] K. M. Flegal, B. K. Kit, H. Orpana, and B. I. Graubard, "Association of all-cause mortality with overweight and obesity using standard body mass index categories: a systematic review and meta-analysis," *JAMA*, vol. 309, no. 1, pp. 71–82, Jan. 2013.

[21] A. De Lorenzo, S. Gratteri, P. Gualtieri, A. Cammarano, P. Bertucci, and L. Di Renzo, "Why primary obesity is a disease?" *J. Transl. Med.*, vol. 17, no. 1, p. 169, May 2019.

[22] "Guidelines (2013) for managing overweight and obesity in adults. preface to the expert panel report (comprehensive version which includes systematic evidence review, evidence statements, and recommendations)," *Obesity (Silver Spring)*, vol. 22 Suppl 2, no. S2, p. S40, Jul. 2014.

[23] American College of Cardiology/American Heart Association Task Force on Practice Guidelines, Obesity Expert Panel, 2013, *Executive summary: Guidelines (2013) for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the Obesity Society published by the Obesity Society and American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Based on a systematic review from the The Obesity Expert Panel, 2013.* Wiley, Jul. 2014, vol. 22 Suppl 2, no. S2.

[24] B. M. Herrera and C. M. Lindgren, "Gene–lifestyle interactions in obesity," *Curr. Diab. Rep.*, vol. 10, no. 6, pp. 498–505, Dec. 2010.

[25] A. Marti, M. J. Moreno-Aliaga, J. Hebebrand, and J. A. Martínez, "Genes, lifestyles and obesity," *Int. J. Obes. Relat. Metab. Disord.*, vol. 28 Suppl 3, no. S3, pp. S29–36, Nov. 2004.

[26] W. H. Organization *et al.*, "Overweight and obesity," 2020.

[27] G. A. Head, "Cardiovascular and metabolic consequences of obesity," *Front. Physiol.*, vol. 6, p. 32, Feb. 2015.

[28] G. I. Smith, B. Mittendorfer, and S. Klein, "Metabolically healthy obesity: facts and fantasies," *J. Clin. Invest.*, vol. 129, no. 10, pp. 3978–3989, Oct. 2019.

[29] M. H. Seo and E. J. Rhee, "Metabolic and cardiovascular implications of a metabolically healthy obesity phenotype," *Endocrinol. Metab. (Seoul)*, vol. 29, no. 4, pp. 427–434, Dec. 2014.

[30] Y.-C. Hwang, T. Hayashi, W. Y. Fujimoto, S. E. Kahn, D. L. Leonetti, M. J. McNeely, and E. J. Boyko, "Visceral abdominal fat accumulation predicts the conversion of metabolically healthy obese subjects to an unhealthy phenotype," *Int. J. Obes. (Lond)*, vol. 39, no. 9, pp. 1365–1370, Sep. 2015.

[31] C. M. Phillips, "Metabolically healthy obesity: definitions, determinants and clinical implications," *Rev. Endocr. Metab. Disord.*, vol. 14, no. 3, pp. 219–227, Sep. 2013.

[32] M. Blüher, "Are metabolically healthy obese individuals really healthy?" *Eur. J. Endocrinol.*, vol. 171, no. 6, pp. R209–19, Dec. 2014.

[33] R. Caleyachetty, G. N. Thomas, K. A. Toulis, N. Mohammed, K. M. Gokhale, K. Balachandran, and K. Nirantharakumar, "Metabolically healthy obese and incident cardiovascular disease events among 3.5 million men and women," *J. Am. Coll. Cardiol.*, vol. 70, no. 12, pp. 1429–1437, Sep. 2017.

[34] K. Kantartzis, J. Machann, F. Schick, K. Rittig, F. Machicao, A. Fritsche, H.-U. Häring, and N. Stefan, "Effects of a lifestyle intervention in metabolically benign and malign obesity," *Diabetologia*, vol. 54, no. 4, pp. 864–868, Apr. 2011.

[35] J. Baxter, P. R. Armijo, L. Flores, C. Krause, S. Samreen, and T. Tanner, "Updates on monogenic obesity in a multifactorial disease," *Obes. Surg.*, vol. 29, no. 12, pp. 4077–4083, Dec. 2019.

[36] C. E. Elks, M. den Hoed, J. H. Zhao, S. J. Sharp, N. J. Wareham, R. J. F. Loos, and K. K. Ong, "Variability in the heritability of body mass index: a systematic review and meta-regression," *Front. Endocrinol. (Lausanne)*, vol. 3, p. 29, Feb. 2012.

[37] R. J. F. Loos and G. S. H. Yeo, "The genetics of obesity: from discovery to biology," *Nat. Rev. Genet.*, vol. 23, no. 2, pp. 120–133, Feb. 2022.

[38] T. M. Frayling, N. J. Timpson, M. N. Weedon, E. Zeggini, R. M. Freathy, C. M. Lindgren, J. R. B. Perry, K. S. Elliott, H. Lango, N. W. Rayner, B. Shields, L. W. Harries, J. C. Barrett, S. Ellard,

C. J. Groves, B. Knight, A.-M. Patch, A. R. Ness, S. Ebrahim, D. A. Lawlor, S. M. Ring, Y. Ben-Shlomo, M.-R. Jarvelin, U. Sovio, A. J. Bennett, D. Melzer, L. Ferrucci, R. J. F. Loos, I. Barroso, N. J. Wareham, F. Karpe, K. R. Owen, L. R. Cardon, M. Walker, G. A. Hitman, C. N. A. Palmer, A. S. F. Doney, A. D. Morris, G. D. Smith, A. T. Hattersley, and M. I. McCarthy, "A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity," *Science*, vol. 316, no. 5826, pp. 889–894, May 2007.

[39] K. A. Fawcett and I. Barroso, "The genetics of obesity: FTO leads the way," *Trends Genet.*, vol. 26, no. 6, pp. 266–274, Jun. 2010.

[40] G. Stratigopoulos, S. L. Padilla, C. A. LeDuc, E. Watson, A. T. Hattersley, M. I. McCarthy, L. M. Zeltser, W. K. Chung, and R. L. Leibel, "Regulation of Fto/Ftm gene expression in mice and humans," *Am. J. Physiol. Regul. Integr. Comp. Physiol.*, vol. 294, no. 4, pp. R1185–96, Apr. 2008.

[41] R. Fredriksson, M. Hägglund, P. K. Olszewski, O. Stephansson, J. A. Jacobsson, A. M. Olszewska, A. S. Levine, J. Lindblom, and H. B. Schiöth, "The obesity gene, FTO, is of ancient origin, up-regulated during food deprivation and expressed in neurons of feeding-related nuclei of the brain," *Endocrinology*, vol. 149, no. 5, pp. 2062–2071, May 2008.

[42] N. J. Poritsanos, P. S. Lew, J. Fischer, C. V. Mobbs, J. I. Nagy, D. Wong, U. Rüther, and T. M. Mizuno, "Impaired hypothalamic fto

expression in response to fasting and glucose in obese mice," *Nutr. Diabetes*, vol. 1, no. 10, p. e19, Oct. 2011.

[43] T. K. T. Lam, R. Gutierrez-Juarez, A. Pocai, S. Bhanot, P. Tso, G. J. Schwartz, and L. Rossetti, "Brain glucose metabolism controls the hepatic secretion of triglyceride-rich lipoproteins," *Nat. Med.*, vol. 13, no. 2, pp. 171–180, Feb. 2007.

[44] T. K. T. Lam, R. Gutierrez-Juarez, A. Pocai, and L. Rossetti, "Regulation of blood glucose by hypothalamic pyruvate metabolism," *Science*, vol. 309, no. 5736, pp. 943–947, Aug. 2005.

[45] Z. Song, B. E. Levin, J. J. McArdle, N. Bakhos, and V. H. Routh, "Convergence of pre- and postsynaptic influences on glucosensing neurons in the ventromedial hypothalamic nucleus," *Diabetes*, vol. 50, no. 12, pp. 2673–2681, Dec. 2001.

[46] C. M. Murrin, G. E. Kelly, R. E. Tremblay, and C. C. Kelleher, "Body mass index and height over three generations: evidence from the lifeways cross-generational cohort study," *BMC Public Health*, vol. 12, no. 1, p. 81, Jan. 2012.

[47] D. A. Lawlor, P. Lichtenstein, A. Fraser, and N. Långström, "Does maternal weight gain in pregnancy have long-term effects on offspring adiposity? a sibling study in a prospective cohort of 146,894 men from 136,050 families," *Am. J. Clin. Nutr.*, vol. 94, no. 1, pp. 142–148, Jul. 2011.

[48] K. Schousboe, G. Willemsen, K. O. Kyvik, J. Mortensen, D. I. Boomsma, B. K. Cornes, C. J. Davis, C. Fagnani, J. Hjelmborg, J. Kaprio, M. De Lange, M. Luciano, N. G. Martin, N. Pedersen, K. H. Pietiläinen, A. Rissanen, S. Saarni, T. I. A. Sørensen, G. C. M. Van Baal, and J. R. Harris, "Sex differences in heritability of BMI: a comparative study of results from twin studies in eight countries," *Twin Res.*, vol. 6, no. 5, pp. 409–421, Oct. 2003.

[49] E. W. Demerath, A. C. Choh, S. A. Czerwinski, M. Lee, S. S. Sun, W. C. Chumlea, D. Duren, R. J. Sherwood, J. Blangero, B. Towne, and R. M. Siervogel, "Genetic and environmental influences on infant weight and weight change: the fels longitudinal study," *Am. J. Hum. Biol.*, vol. 19, no. 5, pp. 692–702, Sep. 2007.

[50] C. M. A. Haworth, S. Carnell, E. L. Meaburn, O. S. P. Davis, R. Plomin, and J. Wardle, "Increasing heritability of BMI and stronger associations with the FTO gene over childhood," *Obesity (Silver Spring)*, vol. 16, no. 12, pp. 2663–2668, Dec. 2008.

[51] H.-R. Lajunen, J. Kaprio, A. Keski-Rahkonen, R. J. Rose, L. Pulkkinen, A. Rissanen, and K. Silventoinen, "Genetic and environmental effects on body mass index during adolescence: a prospective study among finnish twins," *Int. J. Obes. (Lond)*, vol. 33, no. 5, pp. 559–567, May 2009.

[52] C. Nan, B. Guo, C. Warner, T. Fowler, T. Barrett, D. Boomsma, T. Nelson, K. Whitfield, G. Beunen, M. Thomis, H. H. Maes, C. Derom, J. Ordoñana, J. Deeks, and M. Zeegers, "Heritability of body mass

index in pre-adolescence, young adulthood and late adulthood," *Eur. J. Epidemiol.*, vol. 27, no. 4, pp. 247–253, Apr. 2012.

[53] C. Vaisse, K. Clement, E. Durand, S. Hercberg, B. Guy-Grand, and P. Froguel, "Melanocortin-4 receptor mutations are a frequent and heterogeneous cause of morbid obesity," *J. Clin. Invest.*, vol. 106, no. 2, pp. 253–262, Jul. 2000.

[54] I. S. Farooqi, J. M. Keogh, G. S. H. Yeo, E. J. Lank, T. Cheetham, and S. O'Rahilly, "Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene," *N. Engl. J. Med.*, vol. 348, no. 12, pp. 1085–1095, Mar. 2003.

[55] P. W. Franks, E. Pearson, and J. C. Florez, "Gene-environment and gene-treatment interactions in type 2 diabetes: progress, pitfalls, and prospects," *Diabetes Care*, vol. 36, no. 5, pp. 1413–1421, May 2013.

[56] T.-C. Yang, S. A. Matthews, and V. Y.-J. Chen, "Stochastic variability in stress, sleep duration, and sleep quality across the distribution of body mass index: insights from quantile regression," *Int. J. Behav. Med.*, vol. 21, no. 2, pp. 282–291, Apr. 2014.

[57] M. M. Rinschen, J. Ivanisevic, M. Giera, and G. Siuzdak, "Identification of bioactive metabolites using activity metabolomics," *Nat. Rev. Mol. Cell Biol.*, vol. 20, no. 6, pp. 353–367, Jun. 2019.

[58] M. Pardo, A. Roca-Rivada, L. M. Seoane, and F. F. Casanueva, "Obesidomics: contribution of adipose tissue secretome analysis to obesity research," *Endocrine*, vol. 41, no. 3, pp. 374–383, Jun. 2012.

[59] F. Badoud, K. P. Lam, M. Perreault, M. A. Zulyniak, P. Britz-McKibbin, and D. M. Mutch, "Metabolomics reveals metabolically healthy and unhealthy obese individuals differ in their response to a caloric challenge," *PLoS One*, vol. 10, no. 8, p. e0134613, Aug. 2015.

[60] B. Xie, M. J. Waters, and H. J. Schirra, "Investigating potential mechanisms of obesity by metabolomics," *J. Biomed. Biotechnol.*, vol. 2012, p. 805683, May 2012.

[61] J. F. Xiao, B. Zhou, and H. W. Ressom, "Metabolite identification and quantitation in LC-MS/MS-based metabolomics," *Trends Analyt. Chem.*, vol. 32, pp. 1–14, Feb. 2012.

[62] R. W. McGarrah, S. B. Crown, G.-F. Zhang, S. H. Shah, and C. B. Newgard, "Cardiovascular metabolomics," *Circ. Res.*, vol. 122, no. 9, pp. 1238–1258, Apr. 2018.

[63] O. D. Rangel-Huerta and A. Gil, "Nutrimetabolomics: An update on analytical approaches to investigate the role of plant-based foods and their bioactive compounds in non-communicable chronic diseases," *Int. J. Mol. Sci.*, vol. 17, no. 12, p. 2072, Dec. 2016.

[64] O. D. Rangel-Huerta, B. Pastor-Villaescusa, and A. Gil, "Are we close to defining a metabolomic signature of human obesity? a systematic review of metabolomics studies," *Metabolomics*, vol. 15, no. 6, p. 93, Jun. 2019.

[65] S. Monnerie, B. Comte, D. Ziegler, J. A. Morais, E. Pujos-Guillot, and P. Gaudreau, "Metabolomic and lipidomic signatures of metabolic

syndrome and its physiological components in adults: A systematic review," *Sci. Rep.*, vol. 10, no. 1, p. 669, Jan. 2020.

[66] A. Zhang, H. Sun, and X. Wang, "Power of metabolomics in biomarker discovery and mining mechanisms of obesity," *Obes. Rev.*, vol. 14, no. 4, pp. 344–349, Apr. 2013.

[67] C. Nie, T. He, W. Zhang, G. Zhang, and X. Ma, "Branched chain amino acids: Beyond nutrition metabolism," *Int. J. Mol. Sci.*, vol. 19, no. 4, Mar. 2018.

[68] G. J. Patti, O. Yanes, and G. Siuzdak, "Innovation: Metabolomics: the apogee of the omics trilogy," *Nat. Rev. Mol. Cell Biol.*, vol. 13, no. 4, pp. 263–269, Mar. 2012.

[69] C. B. Clish, "Metabolomics: an emerging but powerful tool for precision medicine," *Cold Spring Harb. Mol. Case Stud.*, vol. 1, no. 1, p. a000588, Oct. 2015.

[70] M. W. Libbrecht and W. S. Noble, "Machine learning applications in genetics and genomics," *Nat. Rev. Genet.*, vol. 16, no. 6, pp. 321–332, Jun. 2015.

[71] T. Chen, Y. Cao, Y. Zhang, J. Liu, Y. Bao, C. Wang, W. Jia, and A. Zhao, "Random forest in clinical metabolomics for phenotypic discrimination and biomarker selection," *Evid. Based. Complement. Alternat. Med.*, vol. 2013, p. 298183, Feb. 2013.

[72] I. M. Scott, W. Lin, M. Liakata, J. E. Wood, C. P. Vermeer, D. Allaway, J. L. Ward, J. Draper, M. H. Beale, D. I. Corol, J. M. Baker, and R. D.

King, "Merits of random forests emerge in evaluation of chemometric classifiers by external validation," *Anal. Chim. Acta*, vol. 801, pp. 22–33, Nov. 2013.

[73] P. Schober and T. R. Vetter, "Linear regression in medical research," *Anesth. Analg.*, vol. 132, no. 1, pp. 108–109, Jan. 2021.

[74] R. Bender and U. Grouven, "Ordinal logistic regression in medical research," *J. R. Coll. Physicians Lond.*, vol. 31, no. 5, pp. 546–551, Sep. 1997.

[75] J. C. Stoltzfus, "Logistic regression: a brief primer," *Acad. Emerg. Med.*, vol. 18, no. 10, pp. 1099–1104, Oct. 2011.

[76] S. J. Rigatti, "Random forest," *J. Insur. Med.*, vol. 47, no. 1, pp. 31–39, 2017.

[77] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 1, p. 281, Dec. 2019.

[78] L.-L. Zhao, X.-J. Qiu, W.-B. Wang, R.-M. Li, and D.-S. Wang, "NMR metabolomics and random forests models to identify potential plasma biomarkers of blood stasis syndrome with coronary heart disease patients," *Front. Physiol.*, vol. 10, p. 1109, Sep. 2019.

[79] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*. John Wiley & Sons, 2013, vol. 398.

[80] L. D. Whigham, D. A. Schoeller, L. K. Johnson, and R. L. Atkinson, "Effect of clothing weight on body weight," *Int. J. Obes. (Lond)*, vol. 37, no. 1, pp. 160–161, Jan. 2013.

[81] J. Van den Broeck, S. A. Cunningham, R. Eeckels, and K. Herbst, "Data cleaning: detecting, diagnosing, and editing data abnormalities," *PLoS Med.*, vol. 2, no. 10, p. e267, Oct. 2005.

[82] L. A. Nielsen, T. R. H. Nielsen, and J.-C. Holm, "The impact of familial predisposition to obesity and cardiovascular disease on childhood obesity," *Obes. Facts*, vol. 8, no. 5, pp. 319–328, Oct. 2015.

[83] N. Stefan, H.-U. Häring, F. B. Hu, and M. B. Schulze, "Metabolically healthy obesity: epidemiology, mechanisms, and clinical implications," *Lancet Diabetes Endocrinol.*, vol. 1, no. 2, pp. 152–162, Oct. 2013.

[84] Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults, "Executive summary of the third report of the national cholesterol education program (ncep) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (adult treatment panel iii)," *JAMA*, vol. 285, no. 19, pp. 2486–2497, May 2001.

[85] J. V. van Vliet-Ostaptchouk, M.-L. Nuotio, S. N. Slagter, D. Doiron, K. Fischer, L. Foco, A. Gaye, M. Gögele, M. Heier, T. Hiekkalinna, A. Joensuu, C. Newby, C. Pang, E. Partinen, E. Reischl, C. Schwienbacher, M.-L. Tammesoo, M. A. Swertz, P. Burton, V. Ferretti, I. Fortier, L. Giepmans, J. R. Harris, H. L. Hillege, J. Holmen,

A. Jula, J. E. Kootstra-Ros, K. Kvaløy, T. L. Holmen, S. Männistö, A. Metspalu, K. Midthjell, M. J. Murtagh, A. Peters, P. P. Pramstaller, T. Saaristo, V. Salomaa, R. P. Stolk, M. Uusitupa, P. van der Harst, M. M. van der Klauw, M. Waldenberger, M. Perola, and B. H. Wolffenbuttel, "The prevalence of metabolic syndrome and metabolically healthy obesity in europe: a collaborative analysis of ten large cohort studies," *BMC Endocr. Disord.*, vol. 14, no. 1, p. 9, Feb. 2014.

[86] J. Rainer, D. Taliun, Y. D'Elia, C. Pattaro, F. S. Domingues, and C. X. Weichenberger, "FamAgg: an R package to evaluate familial aggregation of traits in large pedigrees," *Bioinformatics*, vol. 32, no. 10, pp. 1583–1585, May 2016.

[87] C. X. Weichenberger, J. Rainer, C. Pattaro, P. P. Pramstaller, and F. S. Domingues, "Comparative assessment of different familial aggregation methods in the context of large and unstructured pedigrees," *Bioinformatics*, vol. 35, no. 1, pp. 69–76, Jan. 2019.

[88] H. Thiele and P. Nürnberg, "HaploPainter: a tool for drawing pedigrees with complex haplotypes," *Bioinformatics*, vol. 21, no. 8, pp. 1730–1732, Apr. 2005.

[89] R. K. Singh, P. Kumar, and K. Mahalingam, "Molecular genetics of human obesity: A comprehensive review," *C. R. Biol.*, vol. 340, no. 2, pp. 87–108, Feb. 2017.

[90] https://eu.idtdna.com/pages/applications/genotyping#:~:text=Genotyping%20is%20the%20process%20of,to%20specific%20traits%20or%20disease.

[91] https://emea.illumina.com/products/by-type/microarray-kits/infinium-omni25-exome-8.html.

[92] D. Taliun, D. N. Harris, M. D. Kessler, J. Carlson, Z. A. Szpiech, R. Torres, S. A. G. Taliun, A. Corvelo, S. M. Gogarten, H. M. Kang, A. N. Pitsillides, J. LeFaive, S.-B. Lee, X. Tian, B. L. Browning, S. Das, A.-K. Emde, W. E. Clarke, D. P. Loesch, A. C. Shetty, T. W. Blackwell, A. V. Smith, Q. Wong, X. Liu, M. P. Conomos, D. M. Bobo, F. Aguet, C. Albert, A. Alonso, K. G. Ardlie, D. E. Arking, S. Aslibekyan, P. L. Auer, J. Barnard, R. G. Barr, L. Barwick, L. C. Becker, R. L. Beer, E. J. Benjamin, L. F. Bielak, J. Blangero, M. Boehnke, D. W. Bowden, J. A. Brody, E. G. Burchard, B. E. Cade, J. F. Casella, B. Chalazan, D. I. Chasman, Y.-D. I. Chen, M. H. Cho, S. H. Choi, M. K. Chung, C. B. Clish, A. Correa, J. E. Curran, B. Custer, D. Darbar, M. Daya, M. de Andrade, D. L. DeMeo, S. K. Dutcher, P. T. Ellinor, L. S. Emery, C. Eng, D. Fatkin, T. Fingerlin, L. Forer, M. Fornage, N. Franceschini, C. Fuchsberger, S. M. Fullerton, S. Germer, M. T. Gladwin, D. J. Gottlieb, X. Guo, M. E. Hall, J. He, N. L. Heard-Costa, S. R. Heckbert, M. R. Irvin, J. M. Johnsen, A. D. Johnson, R. Kaplan, S. L. R. Kardia, T. Kelly, S. Kelly, E. E. Kenny, D. P. Kiel, R. Klemmer, B. A. Konkle, C. Kooperberg, A. Köttgen, L. A. Lange, J. Lasky-Su, D. Levy, X. Lin, K.-H. Lin, C. Liu, R. J. F. Loos, L. Garman, R. Gerszten, S. A. Lubitz,

K. L. Lunetta, A. C. Y. Mak, A. Manichaikul, A. K. Manning, R. A. Mathias, D. D. McManus, S. T. McGarvey, J. B. Meigs, D. A. Meyers, J. L. Mikulla, M. A. Minear, B. D. Mitchell, S. Mohanty, M. E. Montasser, C. Montgomery, A. C. Morrison, J. M. Murabito, A. Natale, P. Natarajan, S. C. Nelson, K. E. North, J. R. O'Connell, N. D. Palmer, N. Pankratz, G. M. Peloso, P. A. Peyser, J. Pleiness, W. S. Post, B. M. Psaty, D. C. Rao, S. Redline, A. P. Reiner, D. Roden, J. I. Rotter, I. Ruczinski, C. Sarnowski, S. Schoenherr, D. A. Schwartz, J.-S. Seo, S. Seshadri, V. A. Sheehan, W. H. Sheu, M. B. Shoemaker, N. L. Smith, J. A. Smith, N. Sotoodehnia, A. M. Stilp, W. Tang, K. D. Taylor, M. Telen, T. A. Thornton, R. P. Tracy, D. J. Van Den Berg, R. S. Vasan, K. A. Viaud-Martinez, S. Vrieze, D. E. Weeks, B. S. Weir, S. T. Weiss, L.-C. Weng, C. J. Willer, Y. Zhang, X. Zhao, D. K. Arnett, A. E. Ashley-Koch, K. C. Barnes, E. Boerwinkle, S. Gabriel, R. Gibbs, K. M. Rice, S. S. Rich, E. K. Silverman, P. Qasba, W. Gan, NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, G. J. Papanicolaou, D. A. Nickerson, S. R. Browning, M. C. Zody, S. Zöllner, J. G. Wilson, L. A. Cupples, C. C. Laurie, C. E. Jaquish, R. D. Hernandez, T. D. O'Connor, and G. R. Abecasis, "Sequencing of 53,831 diverse genomes from the NHLBI TOPMed program," *Nature*, vol. 590, no. 7845, pp. 290–299, Feb. 2021.

[93] S. Das, L. Forer, S. Schönherr, C. Sidore, A. E. Locke, A. Kwong, S. I. Vrieze, E. Y. Chew, S. Levy, M. McGue, D. Schlessinger, D. Stambolian, P.-R. Loh, W. G. Iacono, A. Swaroop, L. J. Scott, F. Cucca, F. Kronenberg, M. Boehnke, G. R. Abecasis, and C. Fuchsberger,

"Next-generation genotype imputation service and methods," *Nat. Genet.*, vol. 48, no. 10, pp. 1284–1287, Oct. 2016.

[94] https://genome.sph.umich.edu/wiki/EPACTS.

[95] B. Babitsch, D. Gohl, and T. von Lengerke, "Re-revisiting andersen's behavioral model of health services use: a systematic review of studies from 1998-2011," *Psychosoc. Med.*, vol. 9, p. Doc11, Oct. 2012.

[96] B. A. Rabin, T. K. Boehmer, and R. C. Brownson, "Cross-national comparison of environmental and policy correlates of obesity in europe," *Eur. J. Public Health*, vol. 17, no. 1, pp. 53–61, Feb. 2007.

[97] M. W. Pataky, W. F. Young, and K. S. Nair, "Hormonal and metabolic changes of aging and the influence of lifestyle modifications," vol. 96, no. 3. Elsevier BV, Mar. 2021, pp. 788–814.

[98] M. Wang, Y. Yi, B. Roebothan, J. Colbourne, V. Maddalena, P. P. Wang, and G. Sun, "Body mass index trajectories among middle-aged and elderly canadians and associated health outcomes," *J. Environ. Public Health*, vol. 2016, p. 7014857, Jan. 2016.

[99] R. Kanter and B. Caballero, "Global gender disparities in obesity: a review," *Adv. Nutr.*, vol. 3, no. 4, pp. 491–498, Jul. 2012.

[100] F. Soriguer, C. Gutiérrez-Repiso, E. Rubio-Martín, E. García-Fuentes, M. C. Almaraz, N. Colomo, I. Esteva de Antonio, M. S. R. de Adana, F. J. Chaves, S. Morcillo, S. Valdés, and G. Rojo-Martínez, "Metabolically healthy but obese, a matter of time? findings from the prospective

pizarra study," *J. Clin. Endocrinol. Metab.*, vol. 98, no. 6, pp. 2318–2325, Jun. 2013.

[101] S. L. Appleton, C. J. Seaborn, R. Visvanathan, C. L. Hill, T. K. Gill, A. W. Taylor, R. J. Adams, and North West Adelaide Health Study Team, "Diabetes and cardiovascular disease outcomes in the metabolically healthy obese phenotype: a cohort study," *Diabetes Care*, vol. 36, no. 8, pp. 2388–2394, Aug. 2013.

[102] R. J. Pruim, R. P. Welch, S. Sanna, T. M. Teslovich, P. S. Chines, T. P. Gliedt, M. Boehnke, G. R. Abecasis, and C. J. Willer, "LocusZoom: regional visualization of genome-wide association scan results," *Bioinformatics*, vol. 26, no. 18, pp. 2336–2337, Sep. 2010.

[103] https://github.com/statgen/swiss.

[104] A. Buniello, J. A. L. MacArthur, M. Cerezo, L. W. Harris, J. Hayhurst, C. Malangone, A. McMahon, J. Morales, E. Mountjoy, E. Sollis, D. Suveges, O. Vrousgou, P. L. Whetzel, R. Amode, J. A. Guillen, H. S. Riat, S. J. Trevanion, P. Hall, H. Junkins, P. Flicek, T. Burdett, L. A. Hindorff, F. Cunningham, and H. Parkinson, "The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D1005–D1012, Jan. 2019.

[105] L. Yengo, J. Sidorenko, K. E. Kemper, Z. Zheng, A. R. Wood, M. N. Weedon, T. M. Frayling, J. Hirschhorn, J. Yang, P. M. Visscher, and GIANT Consortium, "Meta-analysis of genome-wide association stud-

ies for height and body mass index in 700000 individuals of european ancestry," *Hum. Mol. Genet.*, vol. 27, no. 20, pp. 3641–3649, Oct. 2018.

[106] A. V. Khera, M. Chaffin, K. G. Aragam, M. E. Haas, C. Roselli, S. H. Choi, P. Natarajan, E. S. Lander, S. A. Lubitz, P. T. Ellinor, and S. Kathiresan, "Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations," *Nat. Genet.*, vol. 50, no. 9, pp. 1219–1224, Sep. 2018.

[107] S. Chowdhury, T. Dent, N. Pashayan, A. Hall, G. Lyratzopoulos, N. Hallowell, P. Hall, P. Pharoah, and H. Burton, "Incorporating genomics into breast and prostate cancer screening: assessing the implications," *Genet. Med.*, vol. 15, no. 6, pp. 423–432, Jun. 2013.

[108] J. L. Slunecka, M. D. van der Zee, J. J. Beck, B. N. Johnson, C. T. Finnicum, R. Pool, J.-J. Hottenga, E. J. C. de Geus, and E. A. Ehli, "Implementation and implications for polygenic risk scores in healthcare," *Hum. Genomics*, vol. 15, no. 1, p. 46, Jul. 2021.

[109] A. R. Martin, M. Kanai, Y. Kamatani, Y. Okada, B. M. Neale, and M. J. Daly, "Clinical use of current polygenic risk scores may exacerbate health disparities," *Nat. Genet.*, vol. 51, no. 4, pp. 584–591, Apr. 2019.

[110] F. Dudbridge, "Power and predictive accuracy of polygenic risk scores," *PLoS Genet.*, vol. 9, no. 3, p. e1003348, Mar. 2013.

[111] A. C. F. Lewis and R. C. Green, "Polygenic risk scores in the clinic: new perspectives needed on familiar ethical issues," *Genome Med.*, vol. 13, no. 1, p. 14, Jan. 2021.

[112] D. J. M. Crouch and W. F. Bodmer, "Polygenic inheritance, GWAS, polygenic risk scores, and the search for functional variants," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 117, no. 32, pp. 18 924–18 933, Aug. 2020.

[113] N. R. Wray, J. Yang, B. J. Hayes, A. L. Price, M. E. Goddard, and P. M. Visscher, "Pitfalls of predicting complex traits from SNPs," *Nat. Rev. Genet.*, vol. 14, no. 7, pp. 507–515, Jul. 2013.

[114] S. W. Choi, T. S.-H. Mak, and P. F. O'Reilly, "Tutorial: a guide to performing polygenic risk score analyses," *Nat. Protoc.*, vol. 15, no. 9, pp. 2759–2772, Sep. 2020.

[115] A. T. Marees, H. de Kluiver, S. Stringer, F. Vorspan, E. Curis, C. Marie-Claire, and E. M. Derks, "A tutorial on conducting genome-wide association studies: Quality control and statistical analysis," *Int. J. Methods Psychiatr. Res.*, vol. 27, no. 2, p. e1608, Jun. 2018.

[116] F. Privé, B. J. Vilhjálmsson, H. Aschard, and M. G. B. Blum, "Making the most of clumping and thresholding for polygenic scores," *Am. J. Hum. Genet.*, vol. 105, no. 6, pp. 1213–1221, Dec. 2019.

[117] S. W. Choi and P. F. O'Reilly, "PRSice-2: Polygenic risk score software for biobank-scale data," *Gigascience*, vol. 8, no. 7, Jul. 2019.

[118] B. J. Vilhjálmsson, J. Yang, H. K. Finucane, A. Gusev, S. Lindström, S. Ripke, G. Genovese, P.-R. Loh, G. Bhatia, R. Do, T. Hayeck, H.-H. Won, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Discovery, Biology, and Risk of Inherited Variants in Breast

Cancer (DRIVE) study, S. Kathiresan, M. Pato, C. Pato, R. Tamimi, E. Stahl, N. Zaitlen, B. Pasaniuc, G. Belbin, E. E. Kenny, M. H. Schierup, P. De Jager, N. A. Patsopoulos, S. McCarroll, M. Daly, S. Purcell, D. Chasman, B. Neale, M. Goddard, P. M. Visscher, P. Kraft, N. Patterson, and A. L. Price, "Modeling linkage disequilibrium increases accuracy of polygenic risk scores," *Am. J. Hum. Genet.*, vol. 97, no. 4, pp. 576–592, Oct. 2015.

[119] T. S. H. Mak, R. M. Porsch, S. W. Choi, X. Zhou, and P. C. Sham, "Polygenic scores via penalized regression on summary statistics," *Genet. Epidemiol.*, vol. 41, no. 6, pp. 469–480, Sep. 2017.

[120] https://www.prsice.info/step_by_step/.

[121] https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files.

[122] A. E. Locke, B. Kahali, S. I. Berndt, A. E. Justice, T. H. Pers, F. R. Day, C. Powell, S. Vedantam, M. L. Buchkovich, J. Yang, D. C. Croteau-Chonka, T. Esko, T. Fall, T. Ferreira, S. Gustafsson, Z. Kutalik, J. Luan, R. Mägi, J. C. Randall, T. W. Winkler, A. R. Wood, T. Workalemahu, J. D. Faul, J. A. Smith, J. H. Zhao, W. Zhao, J. Chen, R. Fehrmann, Å. K. Hedman, J. Karjalainen, E. M. Schmidt, D. Absher, N. Amin, D. Anderson, M. Beekman, J. L. Bolton, J. L. Bragg-Gresham, S. Buyske, A. Demirkan, G. Deng, G. B. Ehret, B. Feenstra, M. F. Feitosa, K. Fischer, A. Goel, J. Gong, A. U. Jackson, S. Kanoni, M. E. Kleber, K. Kristiansson, U. Lim, V. Lotay, M. Mangino, I. M. Leach, C. Medina-Gomez, S. E. Medland, M. A. Nalls, C. D. Palmer,

D. Pasko, S. Pechlivanis, M. J. Peters, I. Prokopenko, D. Shungin, A. Stančáková, R. J. Strawbridge, Y. J. Sung, T. Tanaka, A. Teumer, S. Trompet, S. W. van der Laan, J. van Setten, J. V. Van Vliet-Ostaptchouk, Z. Wang, L. Yengo, W. Zhang, A. Isaacs, E. Albrecht, J. Ärnlöv, G. M. Arscott, A. P. Attwood, S. Bandinelli, A. Barrett, I. N. Bas, C. Bellis, A. J. Bennett, C. Berne, R. Blagieva, M. Blüher, S. Böhringer, L. L. Bonnycastle, Y. Böttcher, H. A. Boyd, M. Bruinenberg, I. H. Caspersen, Y.-D. I. Chen, R. Clarke, E. W. Daw, A. J. M. de Craen, G. Delgado, M. Dimitriou, A. S. F. Doney, N. Eklund, K. Estrada, E. Eury, L. Folkersen, R. M. Fraser, M. E. Garcia, F. Geller, V. Giedraitis, B. Gigante, A. S. Go, A. Golay, A. H. Goodall, S. D. Gordon, M. Gorski, H.-J. Grabe, H. Grallert, T. B. Grammer, J. Gräßler, H. Grönberg, C. J. Groves, G. Gusto, J. Haessler, P. Hall, T. Haller, G. Hallmans, C. A. Hartman, M. Hassinen, C. Hayward, N. L. Heard-Costa, Q. Helmer, C. Hengstenberg, O. Holmen, J.-J. Hottenga, A. L. James, J. M. Jeff, Å. Johansson, J. Jolley, T. Juliusdottir, L. Kinnunen, W. Koenig, M. Koskenvuo, W. Kratzer, J. Laitinen, C. Lamina, K. Leander, N. R. Lee, P. Lichtner, L. Lind, J. Lindström, K. S. Lo, S. Lobbens, R. Lorbeer, Y. Lu, F. Mach, P. K. E. Magnusson, A. Mahajan, W. L. McArdle, S. McLachlan, C. Menni, S. Merger, E. Mihailov, L. Milani, A. Moayyeri, K. L. Monda, M. A. Morken, A. Mulas, G. Müller, M. Müller-Nurasyid, A. W. Musk, R. Nagaraja, M. M. Nöthen, I. M. Nolte, S. Pilz, N. W. Rayner, F. Renstrom, R. Rettig, J. S. Ried, S. Ripke, N. R. Robertson, L. M. Rose, S. Sanna, H. Scharnagl, S. Scholtens, F. R. Schumacher, W. R. Scott, T. Seuffer-

lein, J. Shi, A. V. Smith, J. Smolonska, A. V. Stanton, V. Steinthorsdottir, K. Stirrups, H. M. Stringham, J. Sundström, M. A. Swertz, A. J. Swift, A.-C. Syvänen, S.-T. Tan, B. O. Tayo, B. Thorand, G. Thorleifsson, J. P. Tyrer, H.-W. Uh, L. Vandenput, F. C. Verhulst, S. H. Vermeulen, N. Verweij, J. M. Vonk, L. L. Waite, H. R. Warren, D. Waterworth, M. N. Weedon, L. R. Wilkens, C. Willenborg, T. Wilsgaard, M. K. Wojczynski, A. Wong, A. F. Wright, Q. Zhang, LifeLines Cohort Study, E. P. Brennan, M. Choi, Z. Dastani, A. W. Drong, P. Eriksson, A. Franco-Cereceda, J. R. Gådin, A. G. Gharavi, M. E. Goddard, R. E. Handsaker, J. Huang, F. Karpe, S. Kathiresan, S. Keildson, K. Kiryluk, M. Kubo, J.-Y. Lee, L. Liang, R. P. Lifton, B. Ma, S. A. McCarroll, A. J. McKnight, J. L. Min, M. F. Moffatt, G. W. Montgomery, J. M. Murabito, G. Nicholson, D. R. Nyholt, Y. Okada, J. R. B. Perry, R. Dorajoo, E. Reinmaa, R. M. Salem, N. Sandholm, R. A. Scott, L. Stolk, A. Takahashi, T. Tanaka, F. M. van 't Hooft, A. A. E. Vinkhuyzen, H.-J. Westra, W. Zheng, K. T. Zondervan, ADIPOGen Consortium, AGEN-BMI Working Group, CARDIOGRAMplusC4D Consortium, CKDGen Consortium, GLGC, ICBP, MAGIC Investigators, MuTHER Consortium, MIGen Consortium, PAGE Consortium, ReproGen Consortium, GENIE Consortium, International Endogene Consortium, A. C. Heath, D. Arveiler, S. J. L. Bakker, J. Beilby, R. N. Bergman, J. Blangero, P. Bovet, H. Campbell, M. J. Caulfield, G. Cesana, A. Chakravarti, D. I. Chasman, P. S. Chines, F. S. Collins, D. C. Crawford, L. A. Cupples, D. Cusi, J. Danesh, U. de Faire, H. M. den Ruijter, A. F. Dominiczak, R. Erbel, J. Erdmann, J. G. Eriks-

son, M. Farrall, S. B. Felix, E. Ferrannini, J. Ferrières, I. Ford, N. G. Forouhi, T. Forrester, O. H. Franco, R. T. Gansevoort, P. V. Gejman, C. Gieger, O. Gottesman, V. Gudnason, U. Gyllensten, A. S. Hall, T. B. Harris, A. T. Hattersley, A. A. Hicks, L. A. Hindorff, A. D. Hingorani, A. Hofman, G. Homuth, G. K. Hovingh, S. E. Humphries, S. C. Hunt, E. Hyppönen, T. Illig, K. B. Jacobs, M.-R. Jarvelin, K.-H. Jöckel, B. Johansen, P. Jousilahti, J. W. Jukema, A. M. Jula, J. Kaprio, J. J. P. Kastelein, S. M. Keinanen-Kiukaanniemi, L. A. Kiemeney, P. Knekt, J. S. Kooner, C. Kooperberg, P. Kovacs, A. T. Kraja, M. Kumari, J. Kuusisto, T. A. Lakka, C. Langenberg, L. L. Marchand, T. Lehtimäki, V. Lyssenko, S. Männistö, A. Marette, T. C. Matise, C. A. McKenzie, B. McKnight, F. L. Moll, A. D. Morris, A. P. Morris, J. C. Murray, M. Nelis, C. Ohlsson, A. J. Oldehinkel, K. K. Ong, P. A. F. Madden, G. Pasterkamp, J. F. Peden, A. Peters, D. S. Postma, P. P. Pramstaller, J. F. Price, L. Qi, O. T. Raitakari, T. Rankinen, D. C. Rao, T. K. Rice, P. M. Ridker, J. D. Rioux, M. D. Ritchie, I. Rudan, V. Salomaa, N. J. Samani, J. Saramies, M. A. Sarzynski, H. Schunkert, P. E. H. Schwarz, P. Sever, A. R. Shuldiner, J. Sinisalo, R. P. Stolk, K. Strauch, A. Tönjes, D.-A. Trégouët, A. Tremblay, E. Tremoli, J. Virtamo, M.-C. Vohl, U. Völker, G. Waeber, G. Willemsen, J. C. Witteman, M. C. Zillikens, L. S. Adair, P. Amouyel, F. W. Asselbergs, T. L. Assimes, M. Bochud, B. O. Boehm, E. Boerwinkle, S. R. Bornstein, E. P. Bottinger, C. Bouchard, S. Cauchi, J. C. Chambers, S. J. Chanock, R. S. Cooper, P. I. W. de Bakker, G. Dedoussis, L. Ferrucci, P. W. Franks, P. Froguel, L. C. Groop, C. A.

Haiman, A. Hamsten, J. Hui, D. J. Hunter, K. Hveem, R. C. Kaplan, M. Kivimaki, D. Kuh, M. Laakso, Y. Liu, N. G. Martin, W. März, M. Melbye, A. Metspalu, S. Moebus, P. B. Munroe, I. Njølstad, B. A. Oostra, C. N. A. Palmer, N. L. Pedersen, M. Perola, L. Pérusse, U. Peters, C. Power, T. Quertermous, R. Rauramaa, F. Rivadeneira, T. E. Saaristo, D. Saleheen, N. Sattar, E. E. Schadt, D. Schlessinger, P. E. Slagboom, H. Snieder, T. D. Spector, U. Thorsteinsdottir, M. Stumvoll, J. Tuomilehto, A. G. Uitterlinden, M. Uusitupa, P. van der Harst, M. Walker, H. Wallaschofski, N. J. Wareham, H. Watkins, D. R. Weir, H.-E. Wichmann, J. F. Wilson, P. Zanen, I. B. Borecki, P. Deloukas, C. S. Fox, I. M. Heid, J. R. O'Connell, D. P. Strachan, K. Stefansson, C. M. van Duijn, G. R. Abecasis, L. Franke, T. M. Frayling, M. I. McCarthy, P. M. Visscher, A. Scherag, C. J. Willer, M. Boehnke, K. L. Mohlke, C. M. Lindgren, J. S. Beckmann, I. Barroso, K. E. North, E. Ingelsson, J. N. Hirschhorn, R. J. F. Loos, and E. K. Speliotes, "Genetic studies of body mass index yield new insights for obesity biology," *Nature*, vol. 518, no. 7538, pp. 197–206, Feb. 2015.

[123] C. Pattaro, F. Marroni, A. Riegler, D. Mascalzoni, I. Pichler, C. B. Volpato, U. Dal Cero, A. De Grandi, C. Egger, A. Eisendle, C. Fuchsberger, M. Gögele, S. Pedrotti, G. K. Pinggera, S. A. Stefanov, F. D. Vogl, C. J. Wiedermann, T. Meitinger, and P. P. Pramstaller, "The genetic study of three population microisolates in south tyrol (MICROS): study design and epidemiological perspectives," *BMC Med. Genet.*, vol. 8, no. 1, p. 29, Jun. 2007.

[124] A. V. Khera, M. Chaffin, K. H. Wade, S. Zahid, J. Brancale, R. Xia, M. Distefano, O. Senol-Cosar, M. E. Haas, A. Bick, K. G. Aragam, E. S. Lander, G. D. Smith, H. Mason-Suares, M. Fornage, M. Lebo, N. J. Timpson, L. M. Kaplan, and S. Kathiresan, "Polygenic prediction of weight and obesity trajectories from birth to adulthood," *Cell*, vol. 177, no. 3, pp. 587–596.e9, Apr. 2019.

[125] B. F. Nobakht, M Gh, A. Arefi Oskouie, M. Rezaei-Tavirani, R. Aliannejad, S. Taheri, F. Fathi, and M. Taghi Naseri, "NMR spectroscopy-based metabolomic study of serum in sulfur mustard exposed patients with lung disease," *Biomarkers*, vol. 22, no. 5, pp. 413–419, Jul. 2017.

[126] H.-H. Chen, Y. J. Tseng, S.-Y. Wang, Y.-S. Tsai, C.-S. Chang, T.-C. Kuo, W.-J. Yao, C.-C. Shieh, C.-H. Wu, and P.-H. Kuo, "The metabolome profiling and pathway analysis in metabolic healthy and abnormal obesity," *Int. J. Obes. (Lond)*, vol. 39, no. 8, pp. 1241–1248, Aug. 2015.

[127] https://biocrates.com/absoluteidq-p180-kit/.

[128] V. Verri Hernandes, N. Dordevic, E. M. Hantikainen, B. B. Sigurdsson, S. V. Smárason, V. Garcia-Larsen, M. Gögele, G. Caprioli, I. Bozzolan, P. P. Pramstaller, and J. Rainer, "Age, sex, body mass index, diet and menopause related metabolites in a large homogeneous alpine cohort," *Metabolites*, vol. 12, no. 3, p. 205, Feb. 2022.

[129] T. Kühn, A. Floegel, D. Sookthai, T. Johnson, U. Rolle-Kampczyk, W. Otto, M. von Bergen, H. Boeing, and R. Kaaks, "Higher plasma

levels of lysophosphatidylcholine 18:0 are related to a lower risk of common cancers in a prospective metabolomics study," *BMC Med.*, vol. 14, no. 1, p. 13, Jan. 2016.

[130] M. Carayol, I. Licaj, D. Achaintre, C. Sacerdote, P. Vineis, T. J. Key, N. C. Onland Moret, A. Scalbert, S. Rinaldi, and P. Ferrari, "Reliability of serum metabolites over a two-year period: A targeted metabolomic approach in fasting and non-fasting samples from EPIC," *PLoS One*, vol. 10, no. 8, p. e0135437, Aug. 2015.

[131] C. Jourdan, A.-K. Petersen, C. Gieger, A. Döring, T. Illig, R. Wang-Sattler, C. Meisinger, A. Peters, J. Adamski, C. Prehn, K. Suhre, E. Altmaier, G. Kastenmüller, W. Römisch-Margl, F. J. Theis, J. Krumsiek, H.-E. Wichmann, and J. Linseisen, "Body fat free mass is associated with the serum metabolite profile in a population-based study," *PLoS One*, vol. 7, no. 6, p. e40009, Jun. 2012.

[132] D. Casares, P. V. Escribá, and C. A. Rosselló, "Membrane lipid composition: Effect on membrane and organelle structure, function and compartmentalization and therapeutic avenues," *Int. J. Mol. Sci.*, vol. 20, no. 9, p. 2167, May 2019.

[133] M. Holeček, "Branched-chain amino acids in health and disease: metabolism, alterations in blood plasma, and as supplements," *Nutr. Metab. (Lond.)*, vol. 15, p. 33, May 2018.

[134] M. Laplante and D. M. Sabatini, "mTOR signaling in growth control and disease," *Cell*, vol. 149, no. 2, pp. 274–293, Apr. 2012.

[135] S. Li, D. Gao, and Y. Jiang, "Function, detection and alteration of acylcarnitine metabolism in hepatocellular carcinoma," *Metabolites*, vol. 9, no. 2, p. 36, Feb. 2019.

[136] B. Batchuluun, D. Al Rijjal, K. J. Prentice, J. A. Eversley, E. Burdett, H. Mohan, A. Bhattacharjee, E. P. Gunderson, Y. Liu, and M. B. Wheeler, "Elevated medium-chain acylcarnitines are associated with gestational diabetes mellitus and early progression to type 2 diabetes and induce pancreatic β-cell dysfunction," *Diabetes*, vol. 67, no. 5, pp. 885–897, May 2018.

[137] X. Zhao, Q. Han, Y. Liu, C. Sun, X. Gang, and G. Wang, "The relationship between branched-chain amino acid related metabolomic signature and insulin resistance: A systematic review," *J. Diabetes Res.*, vol. 2016, p. 2794591, Aug. 2016.

[138] D. M. Libert, A. S. Nowacki, and M. R. Natowicz, "Metabolomic analysis of obesity, metabolic syndrome, and type 2 diabetes: amino acid and acylcarnitine levels change along a spectrum of metabolic wellness," *PeerJ*, vol. 6, p. e5410, Aug. 2018.

[139] M. Carayol, M. F. Leitzmann, P. Ferrari, R. Zamora-Ros, D. Achaintre, M. Stepien, J. A. Schmidt, R. C. Travis, K. Overvad, A. Tjønneland, L. Hansen, R. Kaaks, T. Kühn, H. Boeing, U. Bachlechner, A. Trichopoulou, C. Bamia, D. Palli, C. Agnoli, R. Tumino, P. Vineis, S. Panico, J. R. Quirós, E. Sánchez-Cantalejo, J. M. Huerta, E. Ardanaz, L. Arriola, A. Agudo, J. Nilsson, O. Melander, B. Bueno-de Mesquita,

P. H. Peeters, N. Wareham, K.-T. Khaw, M. Jenab, T. J. Key, A. Scalbert, and S. Rinaldi, "Blood metabolic signatures of body mass index: A targeted metabolomics study in the epic cohort," *J. Proteome Res.*, vol. 16, no. 9, pp. 3137–3146, Sep. 2017.

[140] G. Frigerio, C. Favero, D. Savino, R. Mercadante, B. Albetti, L. Dioni, L. Vigna, V. Bollati, A. C. Pesatori, and S. Fustinoni, "Plasma metabolomic profiling in 1391 subjects with overweight and obesity from the sphere study," *Metabolites*, vol. 11, no. 4, p. 194, Mar. 2021.

[141] S. Heimerl, M. Fischer, A. Baessler, G. Liebisch, A. Sigruener, S. Wallner, and G. Schmitz, "Alterations of plasma lysophosphatidylcholine species in obesity and weight loss," *PLoS One*, vol. 9, no. 10, p. e111348, Oct. 2014.

[142] X. Orozco-Ruiz, A. Anesi, F. Mattivi, and M. M. B. Breteler, "Branched-chain and aromatic amino acids related to visceral adipose tissue impact metabolic health risk markers," *J. Clin. Endocrinol. Metab.*, vol. 107, no. 7, pp. e2896–e2905, Jun. 2022.

[143] B. C. Batch, S. H. Shah, C. B. Newgard, C. B. Turer, C. Haynes, J. R. Bain, M. Muehlbauer, M. J. Patel, R. D. Stevens, L. J. Appel, L. K. Newby, and L. P. Svetkey, "Branched chain amino acids are novel biomarkers for discrimination of metabolic wellness," *Metabolism*, vol. 62, no. 7, pp. 961–969, Jul. 2013.

[144] E. Torretta, P. Barbacini, N. M. Al-Daghri, and C. Gelfi, "Sphingolipids in obesity and correlated co-morbidities: The contribution of

gender, age and environment," *Int. J. Mol. Sci.*, vol. 20, no. 23, p. 5901, Nov. 2019.

[145] X. Gao, W. Zhang, Y. Wang, P. Pedram, F. Cahill, G. Zhai, E. Randell, W. Gulliver, and G. Sun, "Serum metabolic biomarkers distinguish metabolically healthy peripherally obese from unhealthy centrally obese individuals," *Nutr. Metab. (Lond.)*, vol. 13, no. 1, p. 33, May 2016.

[146] M. M. Boulet, G. Chevrier, T. Grenier-Larouche, M. Pelletier, M. Nadeau, J. Scarpa, C. Prehn, A. Marette, J. Adamski, and A. Tchernof, "Alterations of plasma metabolite profiles related to adipose tissue distribution and cardiometabolic risk," *Am. J. Physiol. Endocrinol. Metab.*, vol. 309, no. 8, pp. E736–46, Oct. 2015.

[147] S.-C. Kang, B.-R. Kim, S.-Y. Lee, and T.-S. Park, "Sphingolipid metabolism and obesity-induced inflammation," *Front. Endocrinol. (Lausanne)*, vol. 4, p. 67, Jun. 2013.

[148] G. H. Goossens, "The metabolic phenotype in obesity: Fat mass, body fat distribution, and adipose tissue function," *Obes. Facts*, vol. 10, no. 3, pp. 207–215, Jun. 2017.

[149] M. Jura and L. P. Kozak, "Obesity and related consequences to ageing," *Age (Dordr.)*, vol. 38, no. 1, p. 23, Feb. 2016.

[150] A. Julibert, M. D. M. Bibiloni, and J. A. Tur, "Dietary fat intake and metabolic syndrome in adults: A systematic review," *Nutr. Metab. Cardiovasc. Dis.*, vol. 29, no. 9, pp. 887–905, Sep. 2019.

[151] F. Badoud, K. P. Lam, A. DiBattista, M. Perreault, M. A. Zulyniak, B. Cattrysse, S. Stephenson, P. Britz-McKibbin, and D. M. Mutch, "Serum and adipose tissue amino acid homeostasis in the metabolically healthy obese," *J. Proteome Res.*, vol. 13, no. 7, pp. 3455–3466, Jul. 2014.

[152] M. A. B. Siddik and A. C. Shin, "Recent progress on branched-chain amino acids in obesity, diabetes, and beyond," *Endocrinol. Metab. (Seoul)*, vol. 34, no. 3, pp. 234–246, Sep. 2019.