The role of metabolites in the interplay between gut microbiota and cardiometabolic health, with a focus on short-chain fatty acids

Nogal MacHo, Ana

Awarding institution:
King's College London

The role of metabolites in the interplay between gut microbiota and cardiometabolic health, with a focus on short-chain fatty acids

Ana Nogal

Supervisor: Dr Cristina Menni
Prof Ana M. Valdes

Department of Twin Research Genetic Epidemiology
King's College London

This dissertation is submitted for the degree of
Doctor of Philosophy

April 2024

Abstract

Cardiometabolic diseases (CMD) are the most common cause of morbidity and mortality worldwide, representing a major public health challenge. Besides well-established genetic and environmental risk factors, circulating metabolites, especially gut bacteria-derived metabolites, play a crucial role in the development and progression of CMD. Among these metabolic products, short-chain fatty acids (SCFA), including acetate, propionate and butyrate, have gained attention for their potential role in regulating glucose and lipid metabolism, gut barrier integrity, blood pressure, and immune responses.

Two main hypotheses were proposed for this work: (i) specific metabolites contribute to the individual cardiometabolic risk and are useful biomarkers of prevalent and incident disease and (ii) gut microbial metabolites in serum and in stool, such as SCFAs, are important determinants of CMD and represent specific pathways to be targeted by gut microbiome interventions. These hypotheses were explored by employing different statistical and computational approaches in data from multiple population-based cohorts. These include TwinsUK, ZOE Personalised Responses to Dietary Composition Trial (PREDICT)-1, the Cooperative Health Research in the Region of Augsburg (KORA) and an acute trauma case-control study, and cohorts from the Consortium of METabolomics Studies (COMETS) with myocardial infarction (MI) information. The overarching aims of this thesis were to (i) identify circulating and faecal biomarkers of prevalent and incident CMD, and (ii) investigate the role of SCFAs in the interplay between the gut microbiota and CMD.

To achieve the first aim, in **Chapter 4**, I searched for circulating biomarkers of atherosclerotic cardiovascular disease risk (ASCVD). I identified a panel of 21 circulating metabolites cross-sectionally associated with ASCVD and longitudinally predictive of CVD mortality and morbidity independently of environmental and traditional risk factors. Then, in **Chapter 5**, I searched for circulating biomarkers of incident MI in the largest metabolome study of MI to date consisting of 6 intercontinental COMETS cohorts with diverse race backgrounds. I identified 56 biomarkers (of which 10 were novel) of incident MI. Finally, in **Chapter 6**, I explored the role of faecal metabolites and gut microbiota in prediabetes. I created a faecal metabolite signature that was cross-sectionally associated with impaired fasting glucose in two independent cohorts and was also predictive of incident type-2 diabetes. Although the signature consisted of xenobiotics and host-produced metabolites, it was strongly associated with the gut microbiota composition. For the second aim, in **Chapter 7**, I comprehensively assessed the host genetics and gut microbiota contribution to a panel of eight serum and stool SCFAs, examined their postprandial changes, and explored their links with chronic and acute inflammatory responses in three independent cohorts. I showed that SCFA levels present a heritable genetic component, and that the gut microbiome is strongly correlated with their faecal levels. Moreover, I reported significant changes in SCFA postprandial circulating levels. Furthermore, I found different correlation patterns with inflammatory markers depending on the type of inflammatory response. Lastly, in **Chapter 8**, I further analysed acetate, one of the major SCFAs, and explored the associations between its circulating levels, gut microbiota and visceral fat. I found specific gut bacterial genera associated with its levels, including *Coprococcus* and *Lachnoclostridium*, and identified the mediatory role of acetate in the association between gut microbiota and visceral fat.

In conclusion, these findings illustrate the breadth of the physiological relevance of metabolites, particularly SCFAs, on CMD, and highlight the importance of the gut microbiota in the pathogenesis of CMD not only by producing metabolic products but also by affecting intestinal absorption/excretion of host-produced metabolites. Future studies should determine causality and explore translational strategies that could modulate the identified metabolites by targeting the gut microbiota.

Acknowledgement

I would like to express my gratitude to my supervisor, Dr Cristina Menni and Prof Ana M. Valdes, for their unwavering support, guidance, and encouragement throughout my PhD. Their expertise, feedback, and mentorship have been invaluable in shaping my research and helping me grow as a researcher.

I am also deeply grateful to the members of my thesis committee, Dr Pirro Hysi, Dr Massimo Mangino, Prof Paul Franks, and Prof Jane Grove, for their constructive feedback, insightful comments, and valuable suggestions, which have significantly improved the quality of my work.

Moreover, I would like to thank my colleagues, especially Dr Ilias Attaye and Dr Bano Louca for letting me learn from them and for all the good and fun moments we have had together. I also want to acknowledge the Department of Twin Research & Genetic Epidemiology and Prof Tim Spector for hosting me during my PhD, the Chronic Disease Research Foundation which has funded my research, and all the collaborators and participants of the TwinsUK, ZOE PREDICT-1, KORA, COMETS and the acute trauma case-control cohorts. Without their support, this work would not have been possible.

Additionally, I would like to thank my examiners, Prof Tanya Monaghan and Dr Kostas Tsilidis. I am very grateful for having had the opportunity to receive their feedback and to discuss my thesis projects with them, which was a very enjoyable and insightful experience.

Lastly, I would like to give a big thanks to my parents Begoña and Jose Manuel and my sisters Esther, Maria and Lili. They have always been my main pillar of life, being a constant source of motivation and inspiration. Every day they have been there for me to listen, advise and support me. For sure, I could not be where I am without them. I am also very grateful for the advice and encouragement given by my friends Alba, Roberto and Romain.

Thank you all for your contributions to my academic and personal growth.

Publications

First author publications:

1. Nogal A.*, et al. (2023). A faecal metabolite signature of impaired fasting glucose: results from two independent population-based cohorts. ***Diabetes***, 72(12), 1870-1880. 2022 impact factor: 9.4.

2. Nogal A., et al. (2023). Genetic and gut microbiome determinants of SCFA circulating and fecal levels, postprandial responses and links to chronic and acute inflammation. ***Gut microbes***, 15(1), 2240050. 2022 impact factor: 9.4.

3. Nogal A.*, et al. (2023). Predictive metabolites for incident myocardial infarction: a two-step meta-analysis of Individual Patient Data from six cohorts comprising 7,897 individuals from the COnsortium of METabolomic Studies. ***Cardiovascular research***. 2022 impact factor: 14.2.

4. Nogal A., et al. (2023). Genetic and molecular basis of metabolic and nutritional cardiovascular regulation. In D. Kumar, A.A.M. Wilde & P.M. Elliott (Eds.), Genomic and Molecular Cardiovascular Medicine. ***Elsevier*** - *in press*.

5. Nogal A., et al. (2022). Incremental value of a panel of serum metabolites for predicting risk of atherosclerotic cardiovascular disease. ***Journal of the American Heart Association***, 11(4), e024590. 2022 impact factor: 6.1.

6. Nogal A., et al. (2021). Circulating levels of the short-chain fatty acid acetate mediate the effect of the gut microbiome on visceral fat. ***Frontiers in Microbiology***, 12, 711359. 2022 impact factor: 6.1.

7. Nogal A., et al. (2021). The role of short-chain fatty acids in the interplay between gut microbiota and diet in cardio-metabolic health. ***Gut microbes***, 13(1), 1897212. 2022 impact factor: 9.4.

Other publications:

1. Louca, P., [2 authors], Nogal A., [11 authors]. (2023). Plasma protein N-glycome composition associates with postprandial lipaemic response. ***BMC Medicine***, 21(1), 1-11. 2022 impact factor: 11.8.
2. Louca, P., Meijnikman, A.S., Nogal A., [26 authors]. (2023). The secondary bile acid isoursodeoxycholate correlates with post-prandial lipemia, inflammation, and appetite and changes post-bariatric surgery. ***Cell Reports Medicine***, 4(4). 2022 impact factor: 16.99.
3. Louca, P., Nogal A., [19 authors]. (2022). Cross-sectional blood metabolite markers of hypertension: a multicohort analysis of 44,306 individuals from the COnsortium of METabolomics Studies. ***Metabolites***, 12(7), 601. 2022 impact factor: 5.6.
4. Menni C., [4 authors], Nogal A., [18 authors]. (2022). Symptom prevalence, duration, and risk of hospital admission in individuals infected with SARS-CoV-2 during periods of omicron and delta variant dominance: a prospective observational study from the ZOE COVID Study. ***The Lancet***, 399(10335), 1618-1624. 2022 impact factor: 202.7.
5. Louca, P., Nogal A., [8 authors]. (2021). Body mass index mediates the effect of the DASH diet on hypertension: Common metabolites underlying the association. **Journal of Human Nutrition and Dietetics**, 35(1), 214-222. 2022 impact factor: 3.0.
6. Louca, P., Nogal A., [9 authors]. (2021). Gut microbiome diversity and composition is associated with hypertension in women. ***Journal of Hypertension***, 39(9), 1810. 2022 impact factor: 4.8.

7. Tettamanzi F., [2 authors], Nogal A., [10 authors]. (2021). A high protein diet is more effective in improving insulin resistance and glycemic variability compared to a Mediterranean diet-a cross-over controlled inpatient dietary study. ***Nutrients***, 13(12), 4380. 2022 impact factor: 6.7.

* *denotes equal contribution*

Table of contents

List of figures

List of tables

Abbreviations

ACC	American College of Cardiology
ADA	American Diabetes Association
AHA	American Heart Association
aHEI	Alternate healthy eating index
ANI	Average nucleotide identity
ARIC	Atherosclerosis Risk in Communities Study
ASCVD	Atherosclerotic cardiovascular disease
ASV	Amplicon sequencing variants
AUC	Area under the ROC curve
BCAA	Branched-chain amino acids
BMI	Body mass index
CE	Capillary electrophoresis
CMD	Cardiometabolic diseases
COMETS	Consortium of Metabolomics Studies
CRP	C-reactive protein
CVD	Cardiovascular diseases

DHA	Docosahexaenoic acid
DXA	Dual-energy X-ray absorptiometry
DZ	Dizygotic
EPA	Eicosapentaenoic acid
EPIC	European Prospective Investigation into Cancer and Nutrition
ET2DS	Edinburgh Type 2 Diabetes Study
FBA	Flux balance analyses
FDA	Food and Drug Administration
FDR	False discovery rate
FFAR	Free fatty acid receptor
FFQ	Food frequency questionnaires
FMT	Faecal microbiota transplantation
GC/LC-MS	Untargeted gas and liquid chromatography-mass spectrometry
GDM	GenoDiabMar
GLP-1	Glucagon-like peptide 1
GPR	G-protein coupled receptor
GWAS	Genome-wide association studies
HABC	Health, Aging and Body Composition
HbA1c	Haemoglobin A1c
HEI	Healthy eating index
HLD	High-density lipoprotein
HMDB	Human metabolon database

IEC	Intestinal epithelial cells
IFG	Impaired fasting glucose
IFN	Interferon
IGT	Impaired glucose tolerance
IL	Interleukin
IPA	Ingenuity Pathways Analysis
IPD	Individual Patient Data
K/HDAC	Lysine and histone deacetylas
KEGG	Kyoto Encyclopedia of Genes and Genomes
KORA	Cooperative Health Research in the Region of Augsburg
LC-MS/MS	Liquid chromatography with tandem mass spectrometry
LPS	Lipopolysaccharides
MAG	Metagenome-assembled genome
MetaCyc	Metabolic Pathway Database
MI	Myocardial infarction
MinPath	Minimal set of Pathways
MS	Mass spectrometry
MWAS	Metabolome-wide association study
MZ	Monozygotic
NCBI	National Center for Biotechnology Information
NF-$\varkappa\beta$	Nuclear factor kappa β
NMR	Nuclear magnetic resonance

NPY	Neuropeptide Y
Olfr	Olfactory receptor
ONS	Office for National Statistics
PCA	Principal component analysis
PEP	Phosphoenolpyruvate
PPARγ	Peroxisome proliferator-activated receptor γ
PREDICT	Personalised Responses to Dietary Composition Trial
PYY	Peptide YY
QC	Quality control
RF	Random Forest
SCFA	Short-chain fatty acid
SDG	Sustainable development goals
SD	Standard deviation
SE	Standard error
SGB	Species-level genome bins
SHAP	SHapley Additive exPlanations
STROBE	STrengthening the Reporting of OBservational studies in Epidemiology
T2D	Type-2 diabetes
TE	Treatment effect
TMAO	Trimethylamine N-oxide
TMA	Trimethylamine
TNF-α	Tumour necrosis factor-α

UHGG	Unified Human Gastrointestinal Genome
VAF	Variance accounted for
WHI	Women’s Health Initiative

Chapter 1

Introduction

In this introductory chapter, I provide an overview of the role of the human metabolome in cardiometabolic diseases, the current knowledge and main findings. As it is estimated that 90% of the gut microbial species are associated with 82% of the faecal metabolites [1], and nearly half of the circulating metabolites are associated with gut microbial species and/or metabolic pathways [1], I then discuss the role of the gut microbiota in cardiometabolic health. In particular, I focus on short-chain fatty acids, bacteria-derived metabolites with the potential of mitigating disease factors and/or preventing cardiometabolic diseases. I provide a comprehensive description of their metabolic production routes, their beneficial effects on cardiometabolic health and the involved mechanisms.

Parts of this chapter have been published in *Gut microbes* (Nogal *et al.*, 2021) and in the book chapter: Genetic and molecular basis of metabolic and nutritional cardiovascular regulation. Genomic and Molecular Cardiovascular Medicine. *Elsevier* (Nogal *et al.,* 2023 - *in press*).

Cardiometabolic diseases (CMD) are the most common cause of morbidity and mortality worldwide, representing a major public health challenge [3, 4].

There are many well-established genetic and environmental risk factors associated with CMD including hypertension, dyslipidaemia, smoking, and abdominal adiposity, among others [5]. However, emerging studies have proposed a pivotal role for circulating metabolites in the onset and progression of CMD, such as type-2 diabetes (T2D), atherosclerosis and obesity [6–8].

Metabolomics provides a snapshot of the individual's metabolic state at a particular time, enabling the identification of at-risk individuals before the disease process is well underway [9, 10]. Moreover, the gut microbiota contributes significantly to the generation of these metabolites, and this could potentially explain the observed correlations between gut microbiota composition and CMD [1].

Among gut bacteria-derived metabolites, short-chain fatty acids (SCFA), mainly acetate, propionate and butyrate, are gaining special attention in CMD [11] as they present regulatory functions in the lipid and glucose metabolism, anti-inflammatory and immune response and gut barrier integrity [12].

In this chapter, I will provide an overview of the associations between the human metabolome, the gut microbiome and CMD. I will then focus on SCFAs, discussing the metabolic routes involved in their production, the benefits they exert on cardiometabolic health, and the mechanisms underlying these effects.

1.1 Cardiometabolic diseases

CMD are defined as a combination of multiple derangements in the cardiovascular system and metabolic processes, leading to an increased risk of developing cardiovascular complications. These diseases are typically characterised by dyslipidaemia, hypertension, impaired glucose tolerance, insulin resistance and central adiposity [13].

Several prevalent conditions fall under CMD, including atherosclerosis, myocardial infarction (MI), T2D and obesity [14]. A summary of these along with other related concepts are provided below, as they will be the main cardiometabolic outcomes used in the following chapters.

Summary of the main cardiometabolic outcomes studied in this thesis.

Atherosclerosis is a condition characterised by the hardening and narrowing of arteries due to the development of atherosclerotic plaques within them. This leads to the reduction of blood flow, which can cause serious cardiovascular complications, including MI, stroke, and death [15]. Atherosclerotic cardiovascular disease (ASCVD) refers to a collection of diseases caused by atherosclerosis [16]. The ASCVD risk score is an assessment tool that helps healthcare providers to estimate an individual's 10-year cardiovascular disease (CVD) risk based on several factors, including age, sex, race, total cholesterol and high-density lipoprotein (HDL) levels, systolic blood pressure, use of blood-pressure-lowering medications, T2D status, and smoking status [16].

MI occurs when the blood supply to the myocardium is impeded by a coronary artery blockage. If the blockage is not removed within a few hours, the heart tissue begins to die due to a lack of oxygen [17]. It is estimated that MI causes the death of one person every 40 seconds in the US [18] and one hospital admission every five minutes in the UK (British Heart Foundation, 2021).

T2D is a metabolic disorder characterised by persistent hyperglycemia resulting from insulin resistance and impaired insulin secretion. Its onset is gradual, with people progressing through a state of **prediabetes** [19] defined as impaired levels of fasting glucose (IFG), and/or glucose tolerance (IGT), and/or elevated haemoglobin A1c (HbA1c) [20].

Obesity is a pathophysiological state defined by an excess accumulation of adipose tissue that adversely affects health status [21]. It is frequently defined using the body mass index (BMI) [22]. Nevertheless, the utility of BMI as an obesity metric is limited by its inability to differentiate between lean and adipose tissue or to delineate the distribution of adipose tissue within the body [22]. This is particularly noteworthy because adipose tissue is not a metabolically homogeneous entity. Specifically, **visceral fat**, the fat localised within the abdominal cavity in proximity to internal organs, exhibits more detrimental metabolic activity than subcutaneous adipose tissue, which is situated beneath the skin [23]. Consequently, the quantification of visceral fat is an essential component of a thorough obesity-associated risk assessment [24]. Indeed, visceral fat reduction may result in greater mitigation of obesity-associated morbidities, such as T2D, hypertension, and CVD, compared to interventions focusing solely on BMI [24].

Moreover, these CMD are often accompanied by chronic low-grade inflammation, which is characterised by the widespread release of pro-inflammatory mediators, such as cytokines and C-reactive protein (CRP) [25, 26].

The prevalence of CMD is gradually increasing worldwide, making it a significant public health burden [3, 4]. Indeed, it is estimated that 25% of the total population has CMD and approximately 30% of all the deaths are caused by CMD [27].

Predicting, preventing and treating CMD is highly complex due to their multifactorial aetiology. Despite significant research investments, classical risk and genetic factors lack sufficient power to accurately predict CMD [28]. For instance, genomics approaches such as genome-wide association studies (GWAS) have been able to explain only a small fraction of these diseases, providing limited contributions to mechanism-based intervention strategies [29]. On the other hand, high-throughput metabolomics has been shown to be a powerful tool to identify novel biomarkers of CMD risk as it allows for the probing of interactions between genetics and environmental factors, including diet, lifestyle and the gut microbiome, and has shown to be a powerful tool to identify novel CMD biomarkers [30, 31] (see **Section 1.2**).

1.2 Metabolomics

Metabolomics is a high-throughput technology able to simultaneously measure an extensive set of low-molecular-weight metabolites such as amino acids, lipids, carbohydrates, nucleotides and xenobiotics in various biological samples including stool, serum, saliva, and urine. Therefore, this technique enables to capture the global metabolic state of an individual at a given time [32].

The Human Metabolome Database (HMDB), which is the world's largest and most comprehensive metabolomic database, contains 217920 metabolite entries [33]. Presently, most metabolomics studies are based on two core techniques: mass spectrometry (MS) and nuclear magnetic resonance (NMR) [34]. However, neither method can capture all metabolites in a sample [34]. MS is an analytical technique that ionises chemical compounds to create charged molecules, which are then separated based on their mass-to-charge ratio [35]. To enhance analytical sensitivity and selectivity, MS is often combined with various separation techniques, including liquid chromatography (LC), gas chromatography (GC), and capillary electrophoresis (CE). Nonetheless, the preparation steps involved in MS are complex and may result in the loss of some metabolites [36]. On the other hand, NMR is a non-destructive and quantitative technique based on the interaction of atomic nuclei (e.g., ^{1}H, ^{13}C, or ^{31}P) with a magnetic field [37]. Although it does not require extensive sample preparation, it primarily detects the high-abundance metabolites (from 100 nmol/l to 1 mmol/l or higher) and requires a high amount of sample volume [37].

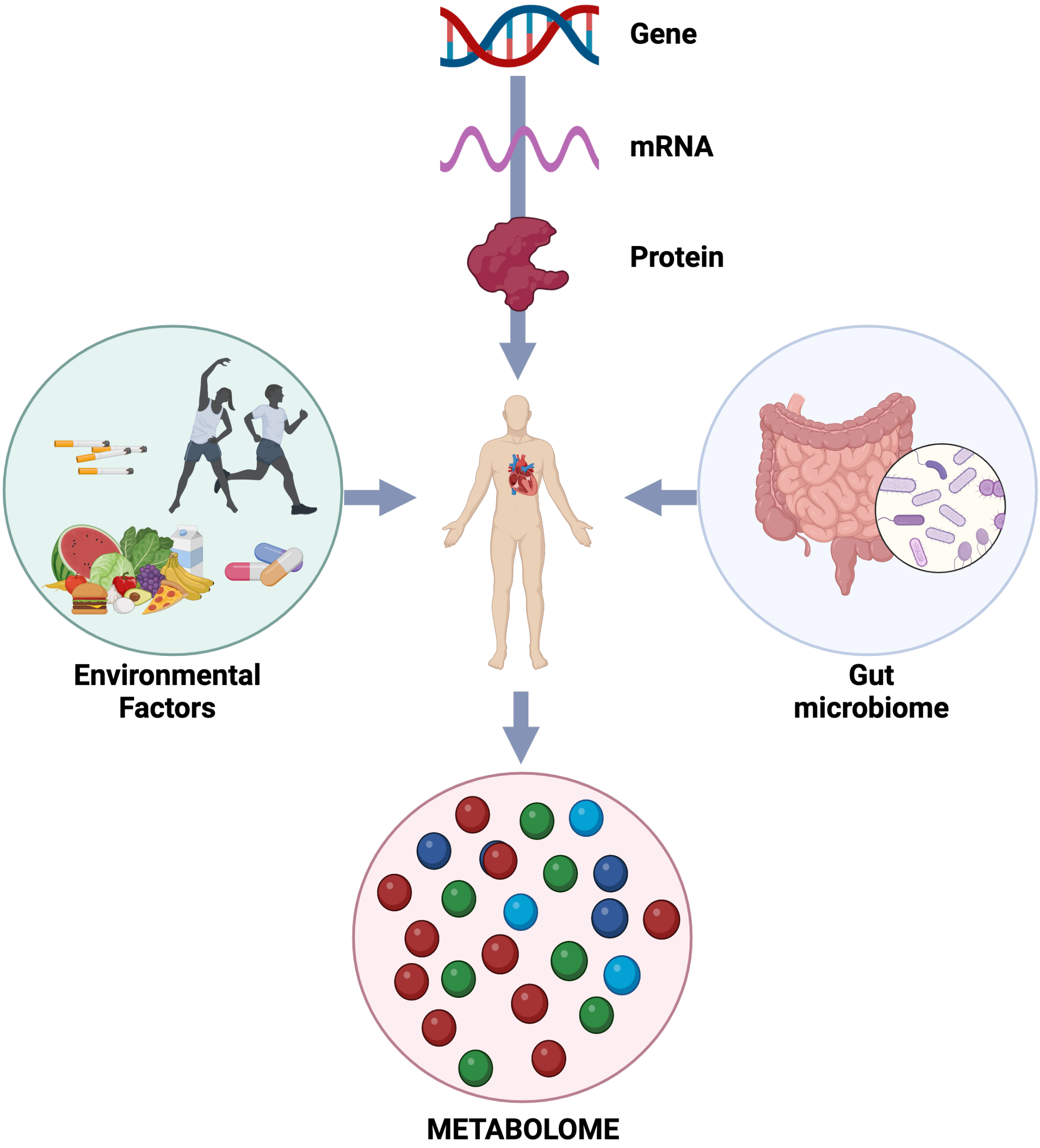


Fig. 1.1 Metabolic profiling as a tool for human diseases. Measurements of metabolite levels do not reflect only the genetic background and the gene expression profile of an individual, but also their lifestyle, dietary intake, medication usage and gut microbiome.

Metabolites, as the downstream products of genetic variations, transcriptional changes, and post-translational protein modifications, are closest to the phenotype. Consequently, the metabolome provides a comprehensive representation of all alterations and interactions among gene expression, protein expression, environmental factors, and the gut microbiome (**Figure 1.1**) [38, 39]. This unique position makes metabolomics a powerful, precise and noninvasive tool for identifying biomarkers useful in patient care, including for screening of asymptomatic individuals (screening biomarkers), diagnosing suspected diseases (diagnostic biomarkers), or monitoring disease progression (prognostic biomarkers). [40–43]. It has thus the potential to provide a more individualised and precision approach to healthcare, enabling more effective prediction, prevention, and treatment strategies for a range of diseases [44]. Furthermore, metabolomics provides an insightful blueprint for deepening our understanding of the pathophysiological mechanisms underlying diseases [45].

1.2.1 Metabolites in cardiometabolic diseases

To date, several metabolites have been linked to CMD and have been used as disease biomarkers [46]. These may exert beneficial or detrimental effects on cardiometabolic health through a variety of mechanisms [46]. Among the most well-known biomarkers of cardiometabolic health, elevated circulating levels of CRP and branched-chain amino acids (BCAA) have been largely associated with unfavourable cardiometabolic outcomes, while other amino acids, such as serine and glycine, and omega-3 fatty acids have been linked to protective effects on cardiometabolic health.

CRP is an acute-phase protein produced primarily by the liver in the presence of inflammatory cytokines, especially interleukin-6 (IL-6) [47]. CRP plays a crucial role in the innate immune system mediating the process of phagocytosis and activating the complement system [47]. In the context of CMD, CRP is widely recognised as a non-specific biomarker of systemic inflammation. Cross-sectional studies have consistently reported CRP levels to be higher in individuals with CMD, including T2D, obesity, atherosclerosis, stroke and MI [48–51]. Although a causal association between CRP levels and the risk of suffering some CVD, such as MI, coronary artery disease, heart failure, and atherosclerosis has not been shown [52, 53], a recent study by Kuppa and collaborators has reported genetically determined CRP to be causally associated with a higher risk of hypertensive heart disease in the European population using a two-sample Mendelian randomisation [53].

BCAA, namely valine, leucine and isoleucine, are essential amino acids and must be thus obtained through diet. They are mainly prevalent in protein-rich foods, including meat, dairy products, and legumes. BCAA have several roles in human metabolism, including protein synthesis and turnover, modulation of glucose homeostasis, and production of energy during prolonged physical activity [54]. However, metabolism dysregulation resulting in elevated levels has been observed in individuals with different forms of CMD, including T2D and obesity [46, 55].

On the other hand, certain amino acids like **glycine and serine** might potentially offer a protective effect against various CMD [56–60]. Indeed, elevated levels of circulating glycine and serine have been negatively associated with a wide range of CMD, such as MI, prediabetes, metabolic syndrome and atherosclerosis, among others [56–60].

Omega-3 fatty acids, particularly docosahexaenoic acid (DHA) and eicosapentaenoic acid (EPA), are polyunsaturated fatty acids that have been extensively studied for their cardioprotective effects [61, 62]. DHA and EPA exert potent anti-inflammatory and anti-atherosclerotic effects [61], and can also improve lipid profiles and endothelial function, and decrease platelet aggregation, which are crucial aspects in preventing the development and progression of CMD [63].

Although most epidemiological studies exploring the associations between the metabolome and CMD have used blood as a biological sample [64–66], metabolites measured in stool samples have been also linked to CMD. For instance, faecal levels of 2-phenylpropionate and hydrocinnamic acid were negatively associated with T2D [67], while higher faecal levels of palmitoylcarnitine were found in diabetic subjects [67]. Faecal levels of SCFAs have been also associated with different CMD, including obesity and CMD risk factors, among others [68, 69].

Importantly, some of the reported metabolite-CMD associations are causally supported by Mendelian randomisation studies, as previously discussed with CRP and hypertensive heart disease risk. The genetically determined BCAA has been associated with a higher risk of T2D [70], coronary artery disease [70, 71], and obesity [72]. In a systematic review summarising 2725 Mendelian randomisation associations between risk factors and coronary artery disease or stroke found that higher levels of genetically predicted omega 6 fatty acid were associated with coronary artery disease [73]. On the other hand, higher genetically predicted circulating EPA has been significantly associated with a lower risk of coronary artery disease and MI [74]. Bidirectional Mendelian randomisation revealed that the host-genetic-driven increase in gut production of the SCFA butyrate was associated with improved insulin response during an oral glucose tolerance test, while disruptions in the production or absorption of the SCFA propionate were causally linked to an increased risk of T2D [75].

Finally, a large proportion of the identified metabolites is produced by the gut microbiota, including SCFAs and indolepropionic acid [2, 76]. As discussed in **Section 1.3**, these may provide mechanistic insights to support correlations between the gut microbiome and CMD [1].

1.3 Gut microbiota

The human intestine contains approximately 10^{14} microorganisms, collectively known as the gut microbiota [77]. The collective genomes of these microorganisms inhabiting the gut constitute what is defined as the gut microbiome [78]. The bacteriome (bacteria) comprises a large proportion of the well-characterised gut microbiome, however, the microbiome also includes the virome - viruses that can infect both human cells and bacteria (phageome), the mycobiome - the collection of fungal species inhabiting the intestine, the archaeome (archaea), and eukaryotic parasites [79]. These organisms represent various kingdoms of life, highlighting the microbiota's transkingdom composition. The gene repertoire present in these microbes is 100-fold higher than the number of genes present in the human genome [80]. In a healthy gut microbiota, the most predominant phyla are Firmicutes and Bacteroidetes (90% of the population), followed by Actinobacteria and Verrucomicrobia [81], although inter-variability between individuals exists. Gut microbiota diversity, richness and composition vary depending on multiple determinants, either endogenous such as sex, transkingdom microbial interactions and host genotype [77, 82], or exogenous, such as diet, age, exercise, smoking, stress [83], and single and multiple drug exposures, including proton pump inhibitors, osmotic laxatives, alpha-glucosidase inhibitors and antibiotics [84].

Over the last years, advances in bioinformatics tools and next-generation sequencing have increased our knowledge of the relationship between microbiota organisms and humans [85], allowing us to discover the benefits and detriments of gut bacteria to human health. Bacteria-derived metabolites play important functions in the intestine (e.g., digestion, energy harvest and barrier integrity) [86] and even in other organs when they enter into the systemic circulation (e.g., glucose circulation in the pancreas, lipid metabolism in the liver and cognitive functions in the brain) [87]. When there is an intestinal microbial ecosystem balance (eubiosis), the gut microbiota plays important immunological, homeostatic and metabolic functions that maintain the human host health [88]. On the other hand, the imbalance of the gut microbiota, known as dysbiosis, and reduction of bacterial diversity can lead to a variety of metabolic abnormalities, such as inflammation and oxidative stress, impacting negatively the host pathophysiologic and physiology conditions [89].

1.3.1 Gut microbiota in cardiometabolic diseases

The role of the gut microbiota has recently been implicated in the development and progression of CMD [90, 91]. Many studies have shown alterations in the composition and function of gut microbiota in patients suffering from CMD. Obesity has been associated with an increased Firmicutes/Bacteroidetes ratio in animals and humans [92]. Moreover, gut dysbiosis can reduce gut barrier integrity, affecting glucose sensibility and absorption, leading to insulin resistance and T2D [93]. Another example is the lipopolysaccharides (LPS) present in the Gram-negative bacteria cell wall, which can trigger the immune system response and potentiate CVD pathogenesis [94, 95]. Although the primary focus of this thesis is the bacteriome, it is important to note that variations in the gut virome and mycobiome have been shown to be involved in CMD [96, 97]. For instance, the gut phageome has been shown to affect T2D not only by modulating gut bacteria composition but also through bacteria-independent mechanisms [96]. Another study reported that gut fungal composition was able to discriminate between obese and non-obese individuals, and metabolically "healthy" from "unhealthy" obesity, indicating that mycobiome deregulation might be linked to obesity [97].

Our body can functionally interact with gut microbial metabolic products [98], and cardiometabolic health is associated with these metabolic products. Trimethylamine (TMA) [66, 99–102], secondary bile acids [103–106] are examples of metabolic products that have been negatively associated with CMD, whereas SCFAs [107], anthocyanins [108] and indoleproprionic acid [109] might influence positively the host health.

TMA is metabolised from choline-containing compounds (e.g., choline, betaine and L-carnitine) present in the human diet by the gut microbiota [110]. Then, TMA enters the portal circulation, where is oxidised by liver enzymes to produce TMAO [111]. TMAO pathway has been associated with atherosclerosis and thrombosis promotion in mice [66, 99, 100], and with CMD in humans such as obesity, chronic kidney disease and T2D [101, 102, 112]. Detailed therapeutic potential and clinical prognostic of TMAO in CMD can be found in several reviews [11, 113, 114].

Gut microbiota is responsible for the generation of **unconjugated free bile acids and secondary bile acids** through deconjugation and dihydroxylation reactions [115]. Bile acids can act as signalling molecules involved in inflammation, host metabolism and energy expenditure, and thus, they might play a role in CMD [103–106]. The role of bile acids in metabolic disorders and CVD has been reviewed by [116, 117].

Anthocyanins are glycosyl-anthocyanidins present in plant vacuoles. Gut bacteria can degrade anthocyanins, generating protocatechuic acid and free anthocyanidins [118]. Protocatechuic acid can influence positively atherosclerosis and CVD, thanks to its anti-inflammatory and antioxidant properties [108].

Indoleproprionic acid is a compound synthesised from tryptophan by a reduced number of bacterial strains [76]. Circulating levels of indoleproprionic acid are negatively correlated with different metabolic syndrome parameters [109], and higher levels of this compound have been also associated with a lower risk of developing T2D [119].

SCFAs are the most well-studied gut bacteria-derived metabolites and they have been suggested as potential disease-mitigating factors and/or disease-preventing in CMD, including T2D, obesity and CVD, among others [75, 107].

Hence, CMD development might be modulated via specific beneficial bacteria-derived metabolites. SCFAs will be explained in detail in the below sections.

1.4 Short-chain fatty acids

Fatty acids are carboxylic acids with an aliphatic chain, which can be saturated or unsaturated [120]. Depending on the length of their aliphatic tails, fatty acids can be classified as short- (<6C), medium- (6-12C) or long- (>12C) chain fatty acids. SCFAs include formate (C1), acetate (C2), propionate (C3), butyrate (C4) and valerate (C5), and their chemical properties depend on the number of carbons [121].

SCFAs are produced by anaerobic gut bacteria through saccharolytic fermentation of complex resistant carbohydrates (e.g., fructooligosaccharides, sugar alcohols, resistant starch, inulin and polysaccharides from plant cell walls), which escape digestion and absorption in the small intestine [122]. As a result of the fermentative reactions, some gases, including hydrogen, methane and carbon dioxide are generated [123]. It is estimated that the fermentation of 50-60 g of carbohydrates per day yields the approximated production of 500-600 mmol of SCFAs in the gut [124]. Amino acids can be also fermented to produce SCFAs [125]. Although SCFAs are dependent on diet and bacteria present in the gut, there are specific foods containing SCFAs, for instance, vinegar, sourdough bread and some dairy products such as crème fraiche, butter and cheese [126].

The major SCFAs formed by the gut bacteria are acetate, propionate and butyrate which account for approximately 80% of all SCFAs and will be the focus of the following sections. In order to comprehensively understand the effect of these metabolites on human health, it is essential to consider the production site and the gradient along different cells and

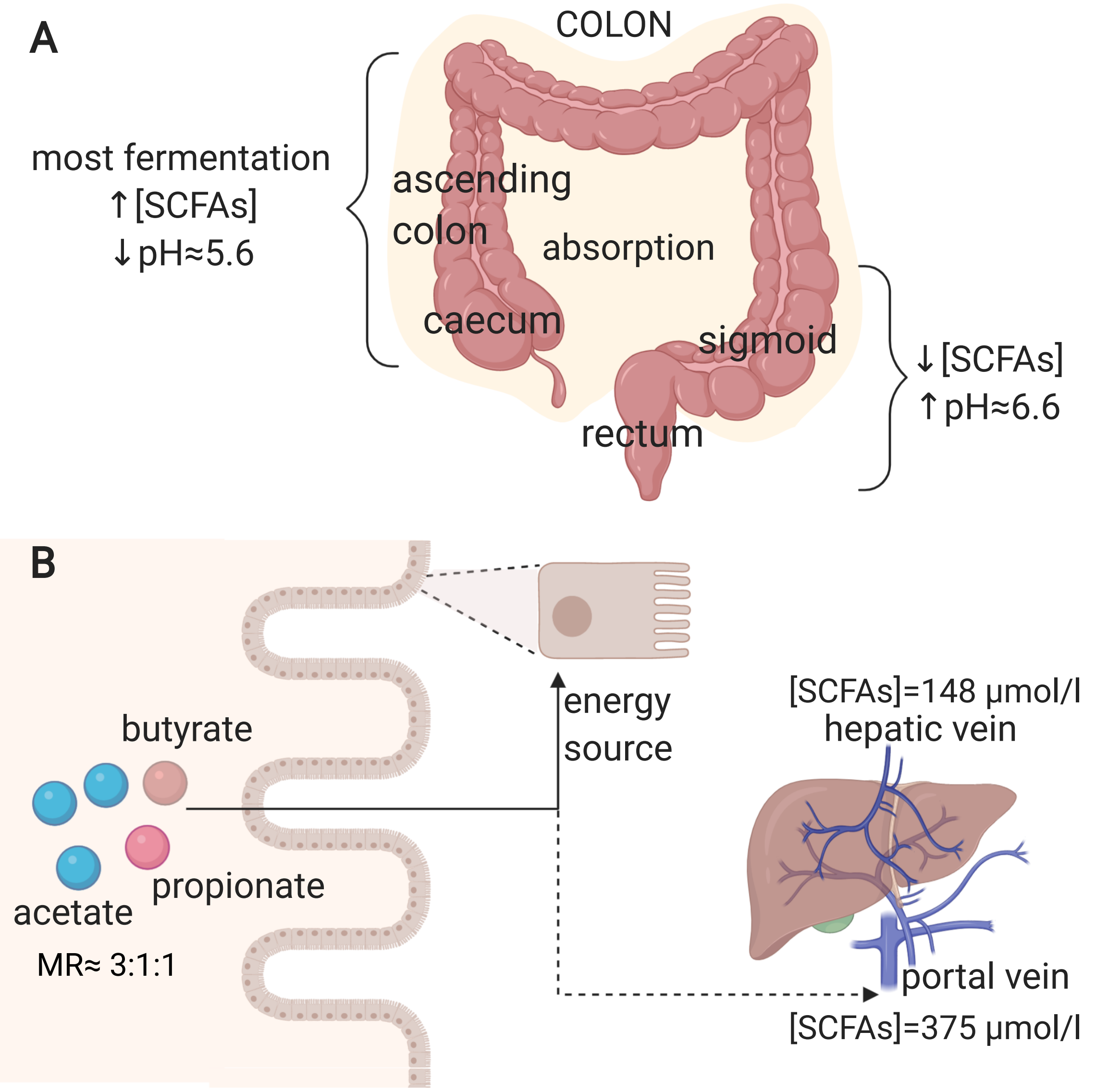


Fig. 1.2 Overview of the production and absorption sites, and transport of acetate, propionate and butyrate (SCFAs). (A) Most undigested carbohydrates are fermented in the caecum and ascending colon, whereas SCFA absorption takes place along the whole colon. A negative correlation between the SCFA concentrations and pH exists. The highest SCFA concentration levels are in the caecum and ascending colon, where the pH is approximately 5.6, whereas in the sigmoid and rectum, the pH is higher (approximately 6.6) and the SCFA concentrations are lower. **(B)** In the colon, acetate, propionate and butyrate are found in an approximate molar ratio of 3:1:1, respectively. Most SCFAs are utilised by colonocytes as an energy source. The SCFAs that are not used by these cells can be transported towards the hepatic portal vein, where the SCFA concentrations are 375 µmol/l, and the hepatic vein, where the SCFA concentrations are 39% of those found in portal blood. *Abbreviations:* MR, molar ratio; SCFAs, short-chain fatty acids.

tissues (**Figure 1.2**). Fermentation takes place in the large intestine, mainly at the right side, and the SCFA absorption occurs rapidly from the human colon [127]. Changes in pH vary depending on the SCFA concentration [128]. In the caecum, the pH is more acidic and the SCFA concentrations are higher than in the sigmoid/rectum, where the pH is higher (**Figure 1.2A**). In the colon and stool, butyrate, propionate and acetate are found in an approximate molar ratio of 20:20:60, respectively [129], although these values vary depending on the microbiota composition, SCFA substrates and gut transit time [130]. Additionally, a strong gradient from the gut lumen to the periphery exists, leading to different cell SCFA exposure [131]. Most SCFAs are utilised by colonocytes as an energy source [124]. The SCFAs that are not used by these cells can be transported towards the hepatic portal vein. Acetate, propionate and butyrate concentrations in portal blood (375 µmol/l) are almost 5 times greater than peripheral venous blood (79 µmol/l), suggesting that the gut is a principal SCFA source, whereas SCFA concentrations in the hepatic vein (148 µmol/l) are 39% of those found in portal blood [129] (**Figure 1.2B**).

1.4.1 SCFA production: metabolic routes and gut bacteria

The pathways involved in SCFA production have been recently described in detail [132]. In addition, metagenomic analyses have allowed the characterisation of the major SCFA-producing bacteria [133] (**Figure 1.3**).

1.4.1.1 Acetate formation

Acetate can be synthesised through two different pathways. Firstly, acetyl-CoA can be produced by decarboxylation of pyruvate, then, acetyl-CoA is hydrolysed to acetate by an acetyl-CoA hydrolase [134]. Most acetate is produced by enteric bacteria, including *Prevotella spp., Ruminococcus spp., Bifidobacterium spp., Bacteroides spp., Clostridium spp., Streptococcus spp., A. muciniphila* and *B. hydrogenotrophica*, using this pathway [135]. Secondly, the Wood-Ljungdahl pathway can be also used by acetogenic bacteria to form acetate from acetyl-CoA. Here, the reduction of a carbon dioxide generates carbon monoxide, which reacts with a coenzyme A molecule and a methyl group to produce acetyl-CoA. At the same time, acetyl-CoA is the substrate to obtain acetate [136].

1.4.1.2 Propionate formation

Although propionate-producers are distributed across several phyla, only a few bacterial genera are able to form propionate, and unlike acetate, the utilised propionate pathways are more conserved and substrate-specific [137].

Propionate can be synthesised through three different biochemical pathways, namely succinate, acrylate and propanediol pathway [137]. In the succinate pathway, the primitive electron transfer chain using phosphoenolpyruvate (PEP) can be utilised to generate propionate [138]. Specifically, PEP is carboxylated to oxalacetate, and then oxalacetate is sequentially converted into malate and fumarate. The latter accepts electrons from NADH using a fumarate reductase and a NADH dehydrogenase, which form a simple electron-transfer chain. The NADH dehydrogenase transports protons across the cell membrane. These protons are utilised for chemiosmotic ATP synthesis. Likewise, succinate is generated as a result of the fumarate reductase. When the carbon dioxide partial pressure is low, succinate is transformed to methylmalonate, which leads to propionate and carbon dioxide. The latter can be recycled for the PEP carboxylation, repeating the process. Bacteroidetes [139] and several Firmicutes belonging to the Negativicutes class [140] use this pathway for propionate formation. Besides, the acrylate pathway can be used to reduce lactate to propionate by a lactoyl-CoA dehydratase [134]. This pathway is only present in a very reduced number of gut bacteria, including *Coprococcus catus* [137]. Lastly, 1,2-propanediol can be formed from deoxy sugars such as rhamnose and fucose in the propanediol pathway. Likewise, 1,2-propanediol is sequentially converted into propionaldehyde and propionyl-CoA, which leads to propionate formation [141]. *Salmonella enterica serovar Typhimurium* [142] and *R. inulinivorans* [143] are bacteria utilising this pathway, just as *A. municiphilla* which appears to be the major propionate-producing species [144].

1.4.1.3 Butyrate formation

Butyrate production, like propionate, is more conserved and substrate-specific [137]. Resistant starch fermentation highly contributes to the formation of butyrate in the colon,

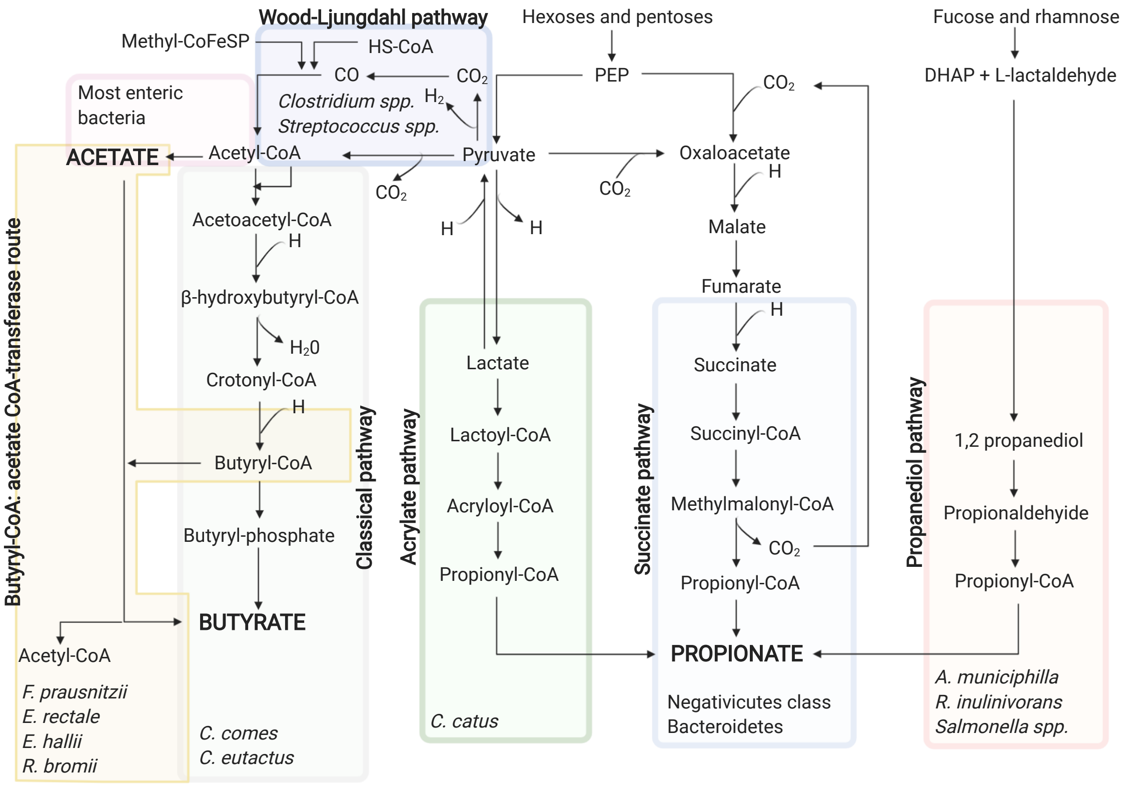


Fig. 1.3 SCFA biosynthesis pathways from the dietary carbohydrate fermentation and the major SCFA-producing bacteria for each pathway. Acetate can be formed by the Wood-Ljungdahl pathway and from pyruvate via acetyl-CoA. Acetyl-CoA can be also produced from lactate by lactate-utilising bacteria. 3 pathways exit for the propionate formation, namely acrylate, succinate and propanediol pathways. The two first use PEP and the latter utilises deoxy sugars such as rhamnose and fucose. Butyrate can be formed through the classical pathway from the condensation of two acetyl-CoA molecules or by the butyryl-CoA: acetate CoA-transferase route, in which butyryl-CoA is converted into butyrate and acetyl-CoA using exogenously derived acetate. *Abbreviations:* DHAP, dihydroxyacetone phosphate; PEP, phosphoenolpyruvate.

with *R. bromii* the main producer as its absence has been associated with a reduction in the resistant starch fermentation [145].

To form butyrate, first, two acetyl-CoA molecules must be condensed to obtain acetoacetyl-CoA, which is subsequently reduced to β-hydroxybutyryl-CoA, crotonyl-CoA and lastly to butyryl-CoA. In the case of lactate-utilising bacteria, acetyl-CoA can be produced from lactate [146]. From butyryl-CoA, butyrate can be synthesised following two different pathways. In the pathway referred to as classical, phosphotransbutyrylase and butyrate kinase enzymes are responsible for such a conversion [147]. In the second pathway, butyryl-CoA: acetate CoA-transferase converts butyryl-CoA into butyrate and acetyl-CoA using exogenously derived acetate. The latter pathway seems to be preferred by the human gut microbiota rather than the classical pathway [148], which is limited to some *Coprococcus* species [132]. *F. prausnitzii, E. rectale, E. hallii* and *R. bromii* present this pathway and appear to be the major butyrate producers [149].

1.4.2 Beneficial roles of SCFAs in cardiometabolic health and involved mechanisms

SCFAs act as signalling molecules on both the gut cells and other tissue cells. This is possible due to six receptors to which SCFAs can bind, triggering intracellular signalling cascades: free fatty acid receptor 3 (FFAR3 or GPR41), FFAR2 (also known as GRP43), G-protein coupled receptor 109a (GPR109a or HCAR2), olfactory receptor-78 (Olfr78 in mice or OR51E2 in humans), GPR42 and OR51E1, being the four first the most well-studied [98]. Olfr78 mainly binds acetate and propionate, leading to an increase of cyclic adenosine monophosphate (cAMP) and renin release [150] and is expressed in the vascular smooth muscle cells in the peripheral vasculature and renal afferent arteriole [151]. FFAR3, FFAR2 and GPR109a are expressed by different organs and cells: small intestine, colon, liver, spleen, heart, skeletal muscle, neurons, immune cells and adipose tissues [152]. Additionally, depending on the length of their aliphatic tails, the receptors present different affinities for SCFAs. FFAR2 prefers binding to acetate and propionate, whereas FFAR3 binds propionate, butyrate and acetate with a lower affinity [153], and GPR109a mainly binds butyrate [154]. Moreover, butyrate and propionate play an important role in transcriptional regulations and post-translational modifications, as they appear to strongly inhibit lysine and histone deacetylase (K/HDAC) activity [98, 155]. Such an inhibition leads to histone hyperacetylation, which turns into a higher accessibility of transcription factors to the promoter regions of different genes [156]. Likewise, butyrate is a ligand of two transcription factors: peroxisome proliferator-activated receptor γ (PPARγ) [157] and aryl hydrocarbon receptor [158]. Thanks to these and direct mechanisms, SCFAs can play beneficial roles in human health, such as improvement of gut barrier integrity, regulation of the energy intake and energy use, modulation of glucose and lipid metabolism and mediation of the immune system and anti-inflammatory response (**Figure 1.4**).

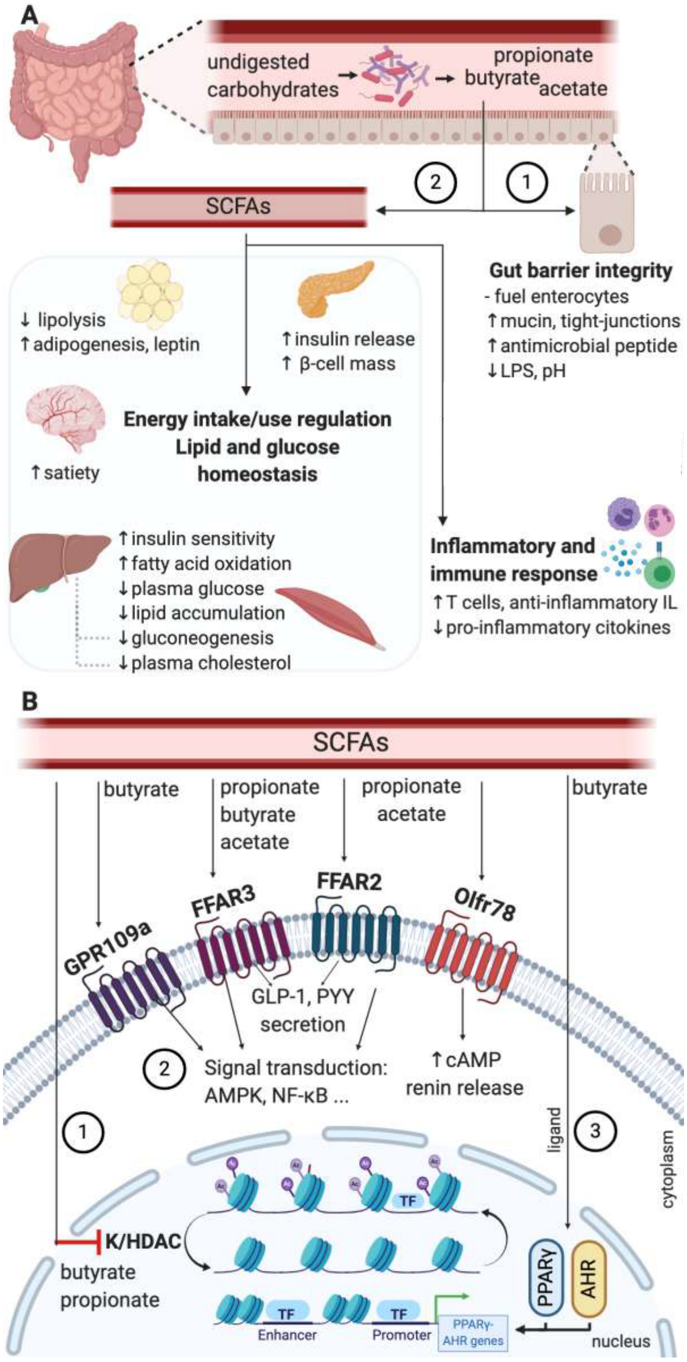


Fig. 1.4 Beneficial roles of SCFAs in cardiometabolic health and the indirect mechanisms involved. **(A)** Undigested carbohydrates reach the intestine, where they are fermented by the SCFA-producing bacteria generating acetate, propionate and butyrate.

SCFAs can act using two different mechanisms: 1) direct action on the enterocytes, maintaining the gut barrier integrity or 2) indirect action regulating the inflammatory and immune response, energy intake and use, and lipid and glucose homeostasis, through the mechanisms illustrated in (B). **(B)** 1) Inhibition of K/HDAC leads to histone hyperacetylation, which turns in higher accessibility of transcription factors to the promoter regions of different genes; 2) signalling transduction activation (in the small intestine, colon, liver, spleen, heart, skeletal muscle, neurons, immune cells and adipose tissues), and GLP-1 and PYY secretion (in intestinal enteroendocrine L-cells) caused by the binding of SCFAs to the G protein-coupled receptors, and increase of cAMP levels by the binding of propionate or acetate to the receptor Olfr78/OR51E2 (in vascular smooth muscle cells in the peripheral vasculature and renal afferent arteriole). GLP-1 and PYY enter into the systematic circulation exerting benefits in different tissues and cells; 3) butyrate working as the ligand of the AHR and PPARγ, leading to the expression of genes dependent on these two transcription factors. *Abbreviations*: AMPK, AMP-activated protein kinase; AHR, aryl hydrocarbon receptor; cAMP, cyclic adenosine monophosphate; FFAR, free fatty acid receptor; GLP-1, glucagon-like peptide-1; GPR109a, G-protein coupled receptor-109a; IL, interleukins; K/HDAC, lysine/histone deacetylase; LPS, lipopolysaccharides; NF-$\varkappa$B, nuclear factor kappa β; Olfr78, olfactory receptor-78; PYY: peptide YY; SCFA, short-chain fatty acid; TF, transcription factor.

1.4.2.1 Energy intake and energy use

SCFAs might present positive effects on body weight control by regulating energy intake and energy expenditure. Some insights have been obtained into the mechanisms by which SCFAs regulate appetite. A potential mechanism might be the stimulation of secretion of gut-derived satiety hormones, such as peptide YY (PYY) and glucagon-like peptide 1 (GLP-1), by SCFAs binding to the free fatty acid receptor FFAR2 and FFAR3 [159]. Both hormones, which are secreted by intestinal enteroendocrine L-cells [160], influence appetite by activating proopiomelanocortin (POMC) neurons in the hypothalamic arcuate nucleus, suppressing neuropeptide Y (NPY) and delaying or inhibiting gastric emptying [161–163]. The expression of genes encoding PYY is also regulated by receptor-independent pathways. Indeed, the inhibitory activity of HDAC by butyrate leads to an increased PYY expression in human L-cells [12]. Besides, a study using *in vivo* ^{11}C-acetate and PET-CT acetate demonstrated that acetate can cross the blood-brain barrier and is taken up by the hypothalamus, causing an appetite decrease and increase of γ-aminobutyric acid and lactate [164]. The secretion of leptin, which is often referred to as the “satiety hormone”, might be also stimulated by SCFAs, resulting in a decreased appetite [165, 166]. For instance, human adipocytes incubated with a high concentration of propionate appeared to increase leptin mRNA expression and leptin secretion [167].

1.4.2.2 Glucose homeostasis and insulin resistance

Several studies have suggested that SCFAs can improve glucose homeostasis *in vivo* by controlling blood glucose levels and increasing glucose uptake mediated by FFAR2 and FFAR3 activation [168–170]. Although the mechanisms are not completely clear, such effects might happen directly via an AMP-activated protein kinase (AMPK)-dependent co-regulated pathway or indirectly via the PPY and GLP-1 hormones. Indeed, Li et al., (2019) [171] have reported that butanoate can affect glucose metabolism through the up-regulation of AMPK-dependent gene expression. Another study has shown that propionate declines hepatic gluconeogenesis via the same mechanism [172]. Furthermore, apart from the previously commented functions of PYY and GLP-1, PYY can also contribute to glucose clearance in adipose tissue and muscle, and GLP-1 can increase insulin secretion and decrease glucagon secretion by the pancreas, regulating blood glucose levels [173]. At the same time, it seems that SCFAs can exert anti-diabetic effects in the host. Propionate presents benefits on pancreatic β-cell function *in vivo*, enhances glucose-stimulated insulin release via FFAR2 activation and increases β-cell mass [170]. Besides, the binding of SCFAs to FFAR2 receptor might ameliorate insulin resistance by promoting autophagy of skeletal muscle cells [174].

1.4.2.3 Gut barrier integrity

It is well-recognised that SCFAs are necessary substrates for colonic epithelium maintenance, with butyrate being the preferred oxidative fuel by colonocytes [175]. Butyrate can induce proliferation in normal colonocytes, but also terminal differentiation and apoptosis in neoplastic cells. This dual role is known as the “butyrate paradox” or “Warburg effect” [155, 176]. Additionally, intestinal epithelial cells (IEC) are connected by transmembrane proteins, namely, tight junctions, adherent junctions and desmosomes. SCFAs, in particular butyrate, seem to improve the epithelial barrier integrity by regulating the tight-junction integrity. Several *in vitro* and experimental animal studies have examined the impact of SCFAs on tight junctions. A study using differentiated IEC observed that butyrate improved the gut barrier integrity through the expression increase of the tight-junctions claudin-1 [177]. An increased expression of other tight-junctions, including claudin-7, ZO-1, ZO-2, occluding and junctional adhesion molecule A (JAMA), was associated with SCFA production in a mouse model (male 7-week-old ICR mice) study [178]. Butyrate also influences epithelial O2 consumption, contributing to the stabilisation of transcription factor hypoxia-inducible factor (HIF), which coordinates the gut barrier protection [179]. A proper gut barrier integrity is essential to avoid some pathogenic bacteria (e.g., *C. pneumoniae, H. pylori, A. actinomycetemcomitans* and *P. gingivalis*) entering into the bloodstream and reaching different tissues, in which they can promote CMD through immune system elicitation, host metabolic and inflammatory response regulation [180–182]. A correct modulation of the mucus layer thickness is also important for the epithelial barrier function. Butyrate can increase the production of MUC2, a predominant mucin glycoprotein secreted by goblet cells [183, 184]. Finally, SCFAs can promote antimicrobial peptide secretion by the IEC. For instance, SCFAs promote the RegIIIγ and defensins production by activating mTOR and STAT3, and thus regulating the epithelial barrier functions [185].

1.4.2.4 Immune function and anti-inflammatory response

SCFAs play a role in the immune system regulation. Of note, it has been shown that butyrate can inhibit HDAC and the activation of nuclear factor kappa β (NF-$\varkappa\beta$) in macrophages [186, 187]. Both HDAC and NF-$\varkappa$B, contribute to the immune and inflammatory response [188]. SCFAs are also involved in anti-inflammatory responses by up-regulating anti-inflammatory cytokines and down-regulating pro-inflammatory ones. For example, SCFAs binding to FFAR2 and GPR109A in IEC stimulates K+ efflux and hyperpolarisation, leading to the inflammasome-activating protein NLRP3 activation, and thus, inducing the IL-18 release, which helps in the maintenance of integrity, repair and intestinal homeostasis [189, 190]. Increased protein acetylation and production of TGF-β 1 in IEC by butyrate lead to a decrease of IL-8 production in IEC [191] and promotion of anti-inflammatory regulatory T cells (Treg) [192], respectively. In human mature dendritic cells, butyrate and propionate appear to reduce the release of pro-inflammatory chemokines, such as CXCL11, CXCL10, CXCL9, CCL5, CCL4 and CCL3, just as inhibiting the expression of LPS-induced cytokines, including IL-6 and IL-12p40 [193]. Apart from the cytokine production regulation, the luminal pH reduction by SCFA inhibits the growth of pathogenic bacteria [194]. Lastly, SCFAs, specifically butyrate, can contribute to host defence by inducing the antimicrobial protein cathelicidin IL-37 [195, 196] and increasing the levels of T regulatory cells in the gut [197].

1.4.2.5 Lipid metabolism

SCFAs can regulate lipolysis and adipogenesis. Acetate and propionate may inhibit endogenous lipolysis, whereas propionate can regulate extracellular lipolysis mediated by an increase of lipoprotein lipase expression, both cases resulting in a decrease of the circulating lipid plasma levels and body weight [166, 198]. As well, SCFAs might play an important role in adipogenic differentiation. Indeed, preadipocytes treated with propionate, and acetate promoted adipocyte differentiation, via overexpression of FFAR2 and PPARγ [199, 200]. Finally, a study involving 40 male Syrian Hamsters, exposed to either a high-cholesterol diet (control) or the same diet enriched with 0.5 mol of acetate, propionate, or butyrate over 6 weeks, reported that these SCFAs reduced plasma cholesterol levels by enhancing hepatic uptake of cholesterol from the blood [201]. Besides, propionate is a potent inhibitor of cholesterol synthesis [202].

Taking all this together, we can deduce that SCFAs can exert benefits in CMD, which are characterised by a deregulation of the glucose and lipid metabolism, inflammation response and/or gut barrier integrity. Indeed, several studies have demonstrated the benefits exerted by SCFAs in CMD. **Table 1.1** shows some of these studies.

Table 1.1 Summary of studies reporting beneficial effects of SCFAs in cardiometabolic health through different traits.

Phenotype	Trait	Study	Study Design	Mechanism/ SCFA-producing bacteria	Main Results
Obesity	Gut barrier integrity and energy usage	Kang (2017) [203]	Mice fed a high-fat diet supplemented with capsaicin	↑Ruminococcaceae and Lachnospiraceae	↓ metabolic endotoxemia and body weight gain
	Gut barrier integrity and inflammation	Cani (2007) [204]	High-fat-diet-fed mice using oligofructose or control	*↑Bifidobacterium spp.*	- ↓ endotoxemia and proinflammatory cytokines (plasma and adipose tissue) - Improvement: glucose tolerance, glucose-induced insulin secretion
	Inflammatory response and lipid metabolism	Schneeberger (2015) [196]	Diet-induced obesity male C57BL/6J mice	*Akkermansia muciniphila*	↓ inflammation, metabolic disorders, altered adipose. tissue metabolism
	Metabolic homeostasis	Dao (2016) [205]	49 overweight and obese adults with calorie restriction	*Akkermansia muciniphila*	Improvement: insulin sensitivity markers and metabolic status
	Metabolic homeostasis	Kimura (2013) [206]	GPR43-deficient and GPR43-overexpressing mice	SCFA-mediated activation of GPR43	↓ fat accumulation (adipose tissue) ↑ lipid and glucose metabolism (other tissues)
	Energy intake	Lin (2012) [207]	C57BL/6N male mice receiving SCFAs	FFAR3-independent	Butyrate and propionate: ↓food intake and body weight
	Insulin resistance	Gao (2009) [208]	Dietary-obese C57BL/6J mice fed with butyrate	Energy expenditure promotion and mitochondria function induction	↑ thermogenesis and fatty acid oxidation - Prevention of insulin resistance and obesity development
T2D	Metabolic homeostasis	Sakakibara (2006) [209]	Diabetic KK-A mice fed with acetic acid and control	AMPK in the liver	↑ gluconeogenesis and lipogenesis genes
	Metabolic homeostasis and blood pressure	Roshanravan (2017) [210]	60 patients with T2D receiving (A) sodium butyrate capsules, (B) inulin supplement powder, (C) inulin and sodium butyrate and (D) placebo	GLP-1	- Group A,B,C: ↓ diastolic blood pressure - Group C: ↓ fasting blood sugar and waist to hip ratio
Cardiovascular diseases	Inflammation	Bartolomaeus (2019) [211]	Wild-type NMRI or apolipoprotein E knockout–deficient mice receiving propionate or control	T-cell dependent	Both models: ↓ systemic inflammation, cardiac hypertrophy, fibrosis, vascular dysfunction and hypertension

Abbreviations: AMPK, AMP-activated protein kinase; CMH, cardiometabolic health; FFAR, free fatty acid receptor; GLP-1, glucagon-like peptide-1; GPR43, G-protein coupled receptor 43; HUVEC, human umbilical vein endothelial cells; IL, interleukins; HDAC, histone deacetylase; SCFA, short-chain fatty acid; OLETF, Otsuka Long-Evans Tokushima Fatty; PBMC, peripheral blood mononuclear cell; PYY, peptide YY; T2D, type-2 diabetes; VCAM-1, vascular cell adhesion molecule-1.

1.4.3 Potential detrimental effects of SCFAs in cardiometabolic health

As previously shown (**Section 1.4.2**), the majority of the current evidence supports the beneficial effect of SCFAs in cardiometabolic health. However, potentially negative effects of SCFAs on CMD have also been reported [168, 212–216].

Indeed, though SCFAs have been shown to promote satiety and reduce appetite (see **Section 1.4.2.1**), there is some controversial work indicating that SCFAs may contribute to weight gain. A systematic review including 7 human clinical studies with 246 obese and 198 healthy subjects reported obese individuals to have higher faecal abundances of acetate, propionate and butyrate compared to the non-obese participants [212]. Some animal studies have also suggested that excessive SCFA production might lead to obesity and weight gain by enhancing the capacity to extract calories from the diet [168].

While butyrate and propionate have been shown to improve insulin sensitivity (see **Section 1.4.2.2**), excessive levels of acetate have been linked to reduced insulin sensitivity in some studies. A study conducted using mice with T2D and obesity induced from a high-fat diet reported that elevated acetate production by pancreatic islets and systemic levels, acting through FFA2 and FFA3 receptors, impaired beta cell response to hyperglycemia. Insulin secretion and glucose tolerance were enhanced in Ffar2-/-, Ffar3-/- and double knockout mice, highlighting the role of acetate in inhibiting glucose-stimulated insulin release under diabetic states [213]. Moreover, a study conducted by Müller and colleagues found that circulating acetate levels were negatively associated with peripheral insulin sensitivity in prediabetic individuals with obesity [214]. However, it was suggested that this observation might be reflecting an altered endogenous acetate metabolism rather than variations in microbial-derived acetate production in individuals with metabolic impairments [214].

Finally, evidence exists for pro-inflammatory actions of SCFAs under some conditions. It is been suggested that the discrepancies with other studies, such as those discussed in **Section 1.4.2.4**, might stem from the ability of SCFAs to induce neutrophil migration, thereby amplifying inflammatory processes [215, 216]. Another explanation proposes that SCFAs may exert pro- or anti-inflammatory effects depending on the cell type in which they act [216, 217].

Chapter 2

Hypotheses, aims and outline

In this chapter, I define the hypotheses and aims, and I provide a thesis outline.

2.1 Hypotheses

1. Specific metabolites contribute to the individual cardiometabolic risk and are useful biomarkers of prevalent and incident disease.
2. Gut microbial metabolites in serum and in stool, such as SCFAs, are important determinants of CMD and represent specific pathways to be targeted by gut microbiome interventions.

2.2 Aims

The overarching aims of this thesis are to (i) identify circulating and faecal biomarkers of prevalent and incident CMD; and (ii) investigate the role of SCFAs in the interplay between the gut microbiota and CMD. This will be achieved by investigating the following tasks:

Aim 1: To identify circulating and faecal biomarkers of prevalent and incident CMD

Task 1: To identify a panel of serum metabolites associated with the ASCVD risk score and predictive of CVD mortality and morbidity independently of environmental and traditional risk factors, and their underlying mechanisms of action.

Task 2: To discover circulating biomarkers predictive of MI and their underlying mechanisms of action.

Task 3: To find a faecal metabolite signature associated with prediabetes and predictive of incident T2D, and its association with the gut microbiome.

Aim 2: To investigate the role of SCFAs in the interplay between the gut microbiota and CMD

Task 1: To explore the genetic component and the role of the gut microbiome on their circulating and faecal levels, their postprandial responses, and the influence of circulating SCFAs in inflammatory responses.

Task 2: To study the host-microbial cross-talk involving circulating acetate levels and its effect on visceral fat.

2.3 Outline

A graphical outline of this thesis is depicted in **Figure 2.1**. In **Chapter 3**, I describe the cohorts and methodology used. The first aim is addressed in **Chapters 4, 5** and **6**. Specifically, in **Chapter 4**, I search for a panel of circulating metabolites cross-sectionally associated with the ASCVD risk score and predictive of CVD mortality and morbidity independently of environmental and traditional risk factors. In **Chapter 5**, I then identify circulating metabolites predictive of incident MI in the largest metabolome-wide association study (MWAS) of MI to date, including 10 novel biomarkers, and I explore their underlying mechanisms of action. In **Chapter 6**, I finally search for a faecal metabolite signature of prediabetes and predictive of incident T2D in two independent cohorts. I then explore the gut microbiota contribution to the levels of the metabolites making up the signature, which indicates another mechanism of how the gut microbiome affects prediabetes. The second aim is explored through **Chapters 7** and **8**. In **Chapter 7**, I determine the contribution of the host genetics and gut microbiota composition to the circulating and faecal levels of eight SCFAs, their postprandial responses, and the influence of circulating SCFAs in inflammatory responses. In **Chapter 8**, I then further focus on one of the major SCFAs, namely acetate, and I examine the host-microbial cross-talk involving its circulating levels and its influence on visceral fat. Finally, in **Chapter 9**, I discuss the findings, their implications, limitations, strengths and potential future research directions.

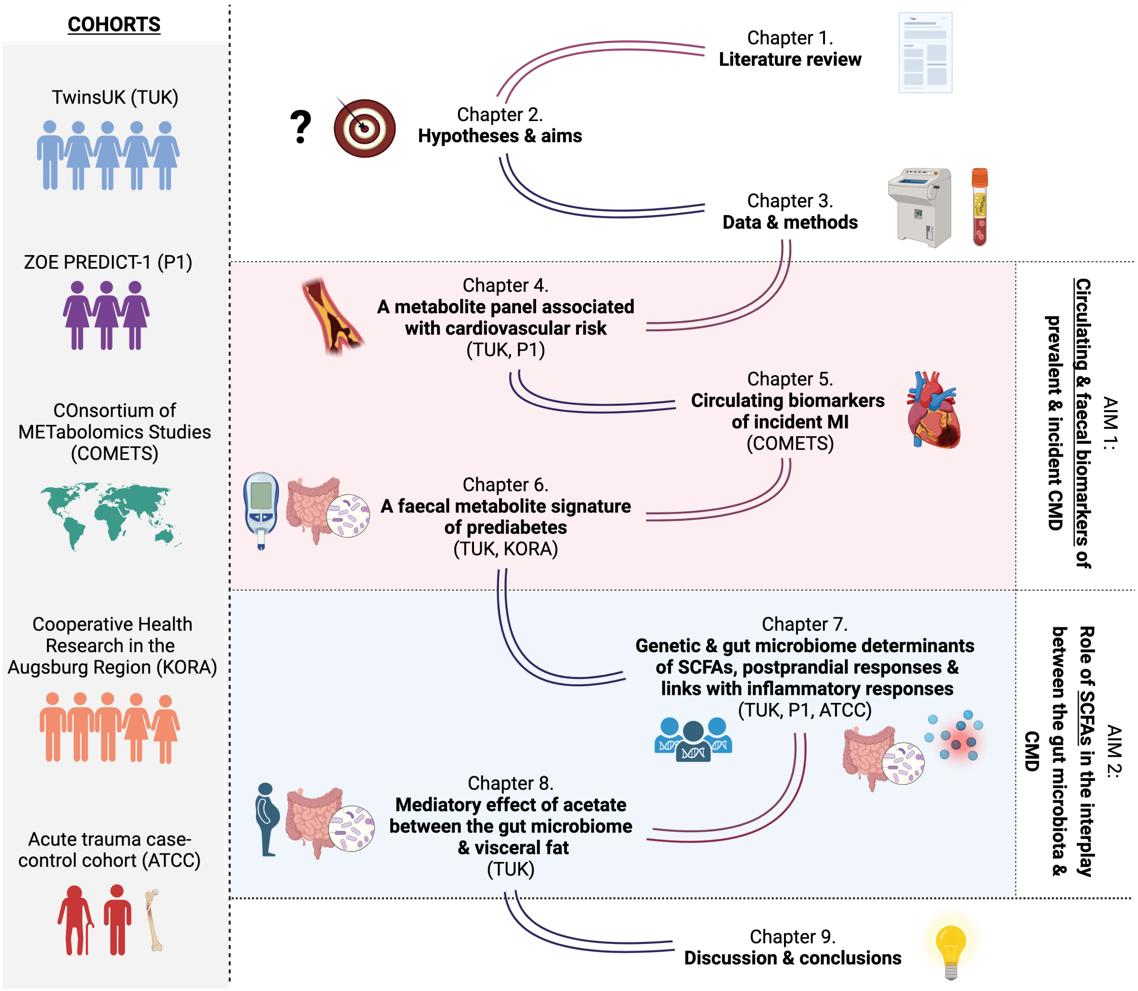

COHORTS
TwinsUK (TUK)
ZOE PREDICT-1 (P1)
COnsortium of METabolomics Studies (COMETS)
Cooperative Health Research in the Augsburg Region (KORA)
Acute trauma case-control cohort (ATCC)
Chapter 1. Literature review
?
Chapter 2. Hypotheses & aims
Chapter 3. Data & methods
Chapter 4. A metabolite panel associated with cardiovascular risk (TUK, P1)
Chapter 5. Circulating biomarkers of incident MI (COMETS)
Chapter 6. A faecal metabolite signature of prediabetes (TUK, KORA)
AIM 1: Circulating & faecal biomarkers of prevalent & incident CMD
Chapter 7. Genetic & gut microbiome determinants of SCFAs, postprandial responses & links with inflammatory responses (TUK, P1, ATCC)
Chapter 8. Mediatory effect of acetate between the gut microbiome & visceral fat (TUK)
AIM 2: Role of SCFAs in the interplay between the gut microbiota & CMD
Chapter 9. Discussion & conclusions

Fig. 2.1 Graphical outline of the thesis. The aims, summary and cohorts used for each chapter are indicated.

Chapter 3

Data and methods

In this chapter, I first describe the cohorts used in this thesis, along with their available clinical and molecular phenotypes. I then explain the different statistical analyses applied throughout the thesis.

The work presented in this thesis is *a posteriori* analysis of existing data from multiple cohorts. Phenotype data, metabolomics and gut microbiome profiling were already available. I calculated the ASCVD risk score for TwinsUK individuals and I was in charge of the quality control of the SCFA data. TwinsUK is the discovery cohort used for most analyses shown in this thesis. Data from three independent population-based cohorts, including ZOE PREDICT-1, KORA and the acute trauma case-control cohort, were used to replicate some key findings. MI and metabolomics data from six COMETS cohorts were also meta-analysed.

3.1 TwinsUK: Discovery cohort

TwinsUK is the largest adult twin registry worldwide, comprising over 14,000 individuals who volunteer to participate without any selection for specific traits or diseases [218]. This registry was developed to answer a range of health-related inquiries. Therefore, comprehensive cross-sectional and longitudinal clinical, biochemical, behavioural, dietary, and socioeconomic data have been collected, making TwinsUK the most clinically detailed cohort worldwide [218]. Twins generally attend full-day clinical visits at St. Thomas' Hospital in London (UK) for non-questionnaire-based data collection, although some twins submit samples via their General Practitioners (GPs). All twins provided informed written consent and the study was approved by St. Thomas' Hospital Research Ethics Committee (REC Ref: EC04/015).

Initially, the research inquiries for TwinsUK were geared toward middle-aged women [219], resulting in most participants being female (82%) and middle-aged (mean age=59 years). These values differ from the general UK population (50.6% female, mean age=40.4 years) [220]. However, on average, both TwinsUK and the UK are overweight, with an average BMI of 26.4 kg/m^2 and 27.6 kg/m^2, respectively. Additionally, TwinsUK is comparable to the general British population regarding lifestyle characteristics, such as smoking, exercise, and dietary habits [219].

Since TwinsUK is an ongoing cohort study, phenotype data is collected in batches on a rolling basis. During wave 3 of data collection, there was an average of 4 years between visits [221]. Demographic data, food frequency questionnaires (FFQ), and blood biochemistry are available for the full registry. Twins zygosity was determined by multiplex DNA fingerprinting (PE Applied Biosystems, Foster City, CA) [222], with the registry containing 51% monozygotic (MZ) and 49% dizygotic (DZ) twins [221, 222]. Omics data is also available. Data relevant to this thesis include metabolomics data assessed with Metabolon Inc. in both serum and stool, and with Nightingale Ltd. in serum, and gut microbiome composition data from the 16S rRNA gene and shotgun metagenome sequencing.

The following subsections describe the main phenotypes utilised throughout this thesis, and which have been assessed in TwinsUK through either questionnaires or trained research nurses.

3.1.1 Body mass index

Height and weight measurements were obtained from all study participants using a free-standing stadiometer and weighing scales at various times during clinical visits. BMI was subsequently calculated as weight divided by height squared (kg/m^2). BMI was used as a covariate in most models reported from **Chapters 4** to **8**.

3.1.2 Blood pressure and hypertension

Office blood pressure was measured by a trained research nurse using a digital blood pressure monitor with the patient, who was fasting, seat for more than 3 minutes. The cuff was placed on the participant's arm, approximately 2-3 cm above the elbow joint of the inner arm, and with the air tube lying over the brachial artery. The participant's arm was supported with the palm facing upward, and the cuff tab was placed at the same level as the heart. Three measurements were taken with an interval of 1 minute between each reading. The first measurement was discarded, whereas the average of the second and third measurements was calculated and recorded in mmHg.

Hypertension was defined following the European Society of Hypertension (ESH) guidelines [223]. A participant was suffering from hypertension whether presented systolic blood pressure >140 mmHg or diastolic blood pressure >90 mmHg or was taking hypertension-lowering medications or was diagnosed hypertensive by the doctor.

Models from sub-analyses conducted in **Chapters 5** and **6** were adjusted for systolic and diastolic blood pressure levels or hypertension. Moreover, in **Chapter 4**, these variables were included in the calculation of the estimated ASCVD risk score.

3.1.3 Cardiometabolic phenotypes and related conditions

Definitions of the cardiometabolic phenotypes and related conditions used as the main responses throughout this thesis are explained in detail below.

- **Estimated ASCVD risk score and CVD mortality:** The 10-year ASCVD risk score is a sex- and race-specific single multivariable risk assessment tool used to estimate the 10-year CVD risk of an individual based on age, sex, and traditional risk factors including HDL and total cholesterol, blood pressure levels and medications, smoking, and T2D [16]. The score was individually calculated following the developed formula by the American College of Cardiology/American Heart Association (ACC/AHA) [16] and used in **Chapter 4**. Details on how the traditional risk factors included in the score were measured are presented in **Sections 3.1.2** and **3.1.4**.

- **MI:** MI assessment was based on self-reported questionnaires, in which participants were asked whether they have suffered from heart attacks, and if so, when this occurred. Incident MI was the main outcome studied in **Chapter 5**.

- **Prediabetes and T2D:** Subjects were classified based on isolated fasting glucose following the American Diabetes Association (ADA) guidelines [224]. Individuals with T2D presented fasting glucose concentrations $\geq$7 mmol/L or their condition was confirmed by a physician's letter. Prediabetes status was based on whether an individual presented IFG, which was characterised by not taking diabetic medication and presenting fasting glucose concentrations >5.5 mmol/L and <7 mmol/L. Healthy participants did not have IFG and T2D, with fasting glucose concentrations >3.9 mmol/L and $\leq$5.5 mmol/L. Details on the glucose measurements are shown in **Section 3.1.4**. Prevalent prediabetes and incident T2D were the outcomes analysed in **Chapter 6**. Likewise, T2D was used as a covariate in different models conducted in **Chapters 4** and **5**.

- **Chronic and acute inflammation:** Chronic and acute inflammation was estimated based on the circulating levels of a set of pro- (tumour necrosis factor-α (TNF-α), GlycA, interferon-γ (IFN-γ) and IL-6) and anti- (IL-10) inflammatory cytokines. Information on how these were measured in TwinsUK is shown in **Section 3.1.4**. This information for the ZOE PREDICT-1 and acute trauma case-controls cohorts is indicated in **Sections 3.2.1.1** and **3.2.4.1**, respectively. These outcomes were examined in **Chapter 7**.

- **Visceral fat:** Visceral fat was determined using a Dual-energy X-ray absorptiometry (DXA) (Hologic QDR; Hologic Inc., Waltham, MA, USA). Briefly, participants were positioned in a supine position wearing only a gown. The DXA machine was calibrated following the manufacturer's suggestions. The scans were analysed using the QDR System Software v12.6. Regions of interest were defined manually by the same operator following the SOP derived from the manufacturer's guidelines. The lower and upper horizontal margins were placed just above the iliac crest and at half of the distance between the acromion and the iliac crest, respectively. The vertical margins were adjusted at the external body borders so that all the soft tissue was included. This DXA-based measurement has been validated against VF measured by CT scan and shown to be reliable and reproducible [225]. Visceral fat was used as the main outcome in the analyses performed in **Chapter 8**.

3.1.4 Clinical biochemistry

Clinical biochemistry measures including glucose levels and lipid profile (e.g., total cholesterol, HDL and triglyceride levels) were determined from the blood samples collected during the visits at fasting.

Moreover, for a subset of 82 individuals, cytokines were also determined from the blood samples. Specifically, IL-10, TNF-α, and IL-6 levels were measured using the bead-based high-sensitivity human cytokine kit (HSCYTO-60SK, Linco-Millipore) according to the manufacturer's instructions, while GlycA was measured using the high-throughput NMR metabolomic 2016 panel (Nightingale Ltd.).

Glucose levels were used in **Chapter 6** to determine the individuals suffering from IFG. Levels from different lipids were included as covariates in the models constructed in **Chapters 4**, **5** and **6**. Levels of the aforementioned cytokines were utilised in **Chapter 7**.

3.1.5 Dietary intake

Dietary intake was assessed using a 131-item paper-based FFQ, which was adapted from the European Prospective Investigation into Cancer and Nutrition (EPIC) FFQ [226]. The EPIC FFQ has been previously validated against plasma ascorbic acid levels and urinary biomarkers [227]. Over time, the FFQ was customised to align with the evolving dietary patterns of the general population in the UK and to explore novel research questions.

FFQs were processed using the FETA software, which is an open-source, cross-platform tool designed specifically for the EPIC FFQ in accordance with their guidelines [228]. The default nutritional database of FETA is based on McCance and Widdowson's The Composition of Foods (5^{th} edition) [229]. FETA generates a spreadsheet containing the daily intake of nutrients, food items, and energy. From these metrics, different dietary indices/scores that are widely used in nutritional epidemiological studies were performed [230]. Among them, details of the healthy eating index (HEI) [231] and the alternate healthy eating index (aHEI) [232] are provided below as these were included in **Chapters 4** and **6**, respectively.

- **HEI score:** A measure to determine overall diet quality and the quality of several dietary components based on the recommendations of the Dietary Guidelines for Americans [233]. It is a numerical score ranging from 0 to 100 points, where higher values indicate a better alignment with the US dietary recommendations [233].
- **aHEI score:** It is an alternative to the HEI score, where the focus is given to foods and nutrients associated with a decreasing risk of chronic diseases [234]. It is based on 11 components, including vegetables, alcohol, and red and processed meats. Each component is ranked from 0 to 10, with all the components creating a score ranging from 0 to 110. Higher scores suggest better dietary quality [234].

3.1.6 Metabolomics

The different technologies used for metabolomics profiling, namely liquid chromatography with tandem MS (LC-MS/MS) (Metabolon Inc.) and NMR spectrometry (Nightingale Ltd.), are explained below.

3.1.6.1 LC-MS/MS metabolomics (Metabolon Inc. platform)

Metabolomics profiling in 5091 serum and 4015 stool samples were quantified by Metabolon Inc. (Morrisville, USA) using the untargeted MS platform. The resultant metabolites measured in serum were analysed in **Chapters 4** and **5**, while the metabolites measured from stool samples were studied in **Chapter 6**.

Samples were collected during the clinical visits and shipped to Metabolon Inc. on ice. Upon arriving and as a means of quality control (QC), several recovery standards were added before the first step in the extraction process. Briefly, to remove protein, dissociate small molecules bound to proteins or trapped within the precipitated protein matrix, and to recover chemically diverse metabolites, proteins were precipitated in methanol and vigorously shaken for 2 minutes (Glen Mills GenoGrinder 2000), then centrifuged. The resulting extract was divided into five fractions; both aliquots (i) and (ii) were analysed using acidic positive ion conditions and chromatographically optimised for hydrophilic and hydrophobic compounds respectively, aliquot (iii) was analysed using basic negative ion optimised conditions using a dedicated separate dedicated C18 column, an aliquot (iv) was analysed using negative ionisation following elution from a hydrophilic interaction liquid chromatography column, while aliquot (v) was reserved as a back-up. Several controls were analysed in concert with experimental samples. (i) A pooled sample generated from a small volume of each experimental sample of interest served as a technical replicate throughout the platform run; (ii) extracted water samples served as process blanks; (iii) and a cocktail of standards, known not to interfere with measurements, spiked into every analysed sample facilitated instrument performance monitoring and aided chromatographic alignment. Instrument variability was determined by calculating the median relative standard deviation (RSD) for the standards that were added to each sample prior to injection into the mass spectrometers. Overall process variability was determined by calculating the median RSD for all endogenous metabolites (i.e., non-instrument standards) present in 100% or more of the pooled technical replicate samples. Experimental samples and controls were randomised across the platform run.

Metabolites were identified by comparison of the ion features in the experimental samples to a reference library of chemical standard entries that included retention time/index, molecular weight (m/z), and MS spectra. Identification of known chemical entities is based on comparison across all 3 features to metabolomic library entries of purified standards. More than 3300 commercially available purified standard compounds have been acquired and registered into the library, while additional mass spectral entries have been created for structurally unnamed biochemicals, which have been identified by their recurrent nature (both chromatographic and mass spectral). These compounds have the potential to be identified by the future acquisition of a matching purified standard or by classical structural analysis. Peaks were quantified using area-under-the-curve. The raw area counts for each metabolite in each sample were normalised to correct for variation resulting from instrument inter-day tuning differences by the median value for each run-day, therefore, setting the medians to 1.0 for each run. This preserved variation between samples but allowed metabolites of widely different raw peak areas to be compared on a similar graphical scale.

We included metabolites detected in at least 80% of study participants. Those with more than 20% were therefore excluded from all analyses presented in this thesis. Missing values were imputed using the minimum run-day measures under the assumption that missingness is not random and missing values are missing because concentrations for that particular metabolite are below the limit of detection. This approach offers several advantages. By assuming the lowest detectable concentration for imputed values, the risk of overestimating metabolite concentrations is minimised. Likewise, it helps to reduce run-to-run variability, leading to more consistent and comparable data. Data was then inverse normalised to counteract a non-normal distribution [235, 236].

3.1.6.1.1 Measures of SCFA levels from serum and stool samples

Profiling of a set of eight SCFA, namely acetate, propionate, butyrate, methylbutyrate, isobutyrate, valerate, isovalerate and hexanoate, was performed on 2507 serum and 2229 stool samples by Metabolon Inc. using LC-MS/MS. This panel is separately measured from the rest of the metabolites using the following explained approach. As a result, concentrations that can be interpreted are obtained. The measured levels in serum and stool were used in **Chapter 7**.

Serum and stool samples were spiked with stable labelled internal standards, homogenised and subjected to protein precipitation with an organic solvent. After centrifugation, an aliquot of the supernatant was derivatised. The reaction mixture was injected into an Agilent 1290/AB Sciex QTrap 5500 LC-MS/MS system equipped with a C18 reversed-phase UHPLC column. The mass spectrometer was operated in negative mode using electrospray ionization.

The peak area of the individual analyte product ions was measured against the peak area of the product ions of the corresponding internal standards. Quantitation was performed using a weighted linear least squares regression analysis generated from fortified calibration standards prepared immediately prior to each run.

LC-MS/MS raw data were collected and processed using AB SCIEX software Analyst 1.6.3 and processed using SCIEX OS-MQ software v1.7.

Sample analyses were carried out in a 96-well plate format containing two calibration curves. Accuracy was evaluated using the corresponding QC replicates in the sample runs. QCs met acceptance criteria at all levels for all analytes (QC acceptance criteria: At least 50% of QC samples at each concentration level per analyte should be within ±20.0% of the corresponding historical mean, and at least 2/3 of all QC samples per analyte should fall within ±20.0% of the corresponding historical mean).

3.1.6.2 NMR metabolomics (Nightingale Ltd. platform)

A targeted NMR spectroscopy platform was used to measure levels of acetate from serum (used in **Chapter 8**) by Nightingale Health Ltd. (Helsinki, Finland; previously known as Brainshake Ltd.). Briefly, samples were mixed with sodium phosphate buffer and subsequently transferred to SampleJet ^{1}H NMR tubes (Bruker, Billerica, MA, USA) using a PerkinElmer JANUS handler (Waltham, MA, USA). Samples were analysed on a Bruker AVANCE III (Bruker, Billerica, MA, USA) 500 MHz spectrometer for 5 min. Two control samples, one plasma sample and one mixture of two low-molecular-weight metabolites, were added to each 96-well plate for quality control. The initial data processing, including the Fourier transformations to NMR spectra and automated phasing, was done using the computers that control the spectrometers. The spectra were then automatically transferred to a centralised server that performed more automated spectral processing steps, including an overall signal check for missing/extra peaks, background control, baseline removal and spectral area-specific signal alignments. The spectral information of the actual sample also underwent various comparisons with the spectra of the 2 quality control samples; the data for which was also followed and compared consecutively. For those spectral areas that passed all the quality control steps, regression modelling was then performed to produce the quantified molecular data. A proprietary Bayesian algorithm [237] was used to quantify absolute concentrations of a predefined set of metabolic traits, including lipid constituent measures from lipoprotein subclasses. Moreover, the algorithm provided measures of average particle sizes for very-low-density lipoprotein, low-density lipoprotein, intermediate-density lipoprotein, and HDL as well as a semi-quantitative measure of albumin concentration. This platform has been extensively applied for biomarker profiling in epidemiological studies [62].

3.1.7 Gut microbiome composition

Gut microbiome composition was assessed in TwinsUK using both 16S rRNA gene and shotgun metagenomic sequencing, as described below.

3.1.7.1 16S rRNA gene sequencing

Pre-visit stool collection kits were mailed, and samples were brought or sent on ice to the clinical research facility where they were stored at -80°C. Samples were packed with dry ice and shipped to Cornell University (NY, USA). The gut microbiome composition was then determined there based on 16S rRNA gene sequencing as described elsewhere [238]. Briefly, DNA was isolated from each sample using the PowerSoil kit (MO BIO Laboratories, Carlsbad, CA, USA). The V4 variable region of the 16S rRNA sequence was then amplified and sequenced using a multiplexed approach on the Illumina MiSeq platform (Illumina, San Diego, CA, USA). 16S rRNA sequences were demultiplexed in QIIME 1 v1.8 [239]. Amplicon sequencing variants (ASV) were then generated using the 'DADA2' package [240] following the pipeline as described by Wells and colleagues [241]. The ASV were grouped into genera and the samples with less than 10000 reads were discarded. The indices of microbiome alpha-diversity quantified as Shannon, inverse Simpson, Gini Simpson diversity, CHAO1 and number of observed ASVs were calculated using the ’diversity’ function implemented in the ’microbiome’ package [242]. The calculated alpha-diversity indices and compositional data at the genus level were used in **Chapter 8**.

3.1.7.2 Shotgun metagenomic sequencing

Deep shotgun metagenomic sequencing in stool samples was performed as previously described [1] and as detailed below.

3.1.7.2.1 Faecal sample collection

TwinsUK participants collected stool samples at home in pre-labelled kits (containing 2 x 25ml tube or 1 x 25ml tube and 1 x 10ml Zymo buffer) posted to them prior to their clinic visit date and brought with them to the visit. Alternatively, samples could be posted to the clinic using blue Royal Mail safe boxes. In the laboratory, samples were homogenised, aliquoted into 4 bijou tubes, and stored at -80°C within 2 hours of receipt.

3.1.7.2.2 DNA extraction, library preparation and DNA sequencing

To isolate genomic DNA from faecal material, bijou tubes were removed from the freezer and ground with glass beads and 5-6ml distilled water (Spex Grinder, 10 seconds, 800 strokes per minute). The supernatant was centrifuged and ground further (5 minutes, 1000 strokes per minute) before 200-300µl of the sample was mixed with 10µl PK solution and 720µl of Lysis/Bind Master Mix. Proteins were degraded by the binding solution and subsequently extracted by the KingFisher Flex robot. DNA was washed in 2 steps by washing solutions and eluted in MagMax Core Elution Buffer in 100µl. Library preparation and sequencing were then performed by GenomeScan (Leiden, The Netherlands).

3.1.7.2.3 Metagenome quality control and preprocessing

Sequenced metagenomes were processed using the YAMP pipeline (v. 0.9.5.3) [243]. Briefly, identical reads, potentially generated by PCR amplification [244], were removed. Reads were filtered to remove adapters, known artefacts, phi X 174, and then quality trimmed (PhRED quality score<10). Reads that became too short after trimming (N<60 bp) were discarded. Singleton reads (i.e., reads whose mate has been discarded) were kept to retain as much information as possible. Contaminant reads belonging to the host genome (build: GRCh37) and low-quality samples (i.e., samples with <10M reads after QC) were removed.

3.1.7.2.4 Microbiome taxonomic profiling

The metagenomic analysis was conducted following the general guidelines [245] and based on the bioBakery computational environment [246, 247]. High-resolution taxonomic profiling of the metagenomes was performed using MetaPhlAn 4.beta.2 (default parameters) with the Jan21 database that comprises 26,970 species-level genome bins (SGB) [248]. The obtained compositional data at the SGB level was used in **Chapters 6** and **7**.

3.2 Replication cohorts

Results from each chapter were replicated in different independent cohorts, including ZOE PREDICT-1 (**Chapters 4** and **7**), KORA (**Chapters 5** and **6**), and the acute trauma case-control cohort (**Chapter 7**). Moreover, cohorts from COMETS with MI data were also analysed (**Chapter 5**). An overview of the cohorts used in each chapter and their respective included data is provided in **Table 3.1**.

3.2.1 ZOE Personalised Responses to Dietary Composition Trial (PREDICT)-1

The ZOE PREDICT-1 study is a single-arm nutritional intervention conducted between June 2018 and May 2019 [249]. The 60% of the participants were healthy subjects aged between 18 and 65 years recruited from the TwinsUK registry [219], and the remaining 40% were recruited from the general population using online advertising (115). All participants provided written informed consent and the study was approved by St. Thomas' Hospital Research Ethics Committee (IRAS 236407). The trial was registered on ClinicalTrials.gov (registration number: NCT03479866).

Table 3.1 An overview of the replication cohorts used in each chapter and their respective included OMICS data (metabolomics and metagenomics).

Cohort use	Cohort	OMICS data	Methodology	Chapter
Discovery	**TwinsUK**	Serum metabolomics	LC-MS/MS	5
		Faecal metabolomics	LC-MS/MS	6
		Serum and stool SCFAs	LC-MS/MS	7
		Serum acetate	NMR	8
		Gut microbiome	Shotgun metegenomic sequencing	6, 7
		Gut microbiome	16S rRNA sequencing - ASV	8
Replication	**COMETS**	Serum metabolomics	LC/GC-MS/MS, NMR	5
	ZOE PREDICT-1	Serum metabolomics	LC-MS/MS	4
		Serum and stool SCFAs	LC-MS/MS	7
		Gut microbiome	Shotgun metagenomic sequencing	7
	KORA	Serum metabolomics	LC-MS/MS	5, 6
	Acute trauma case-control cohort	Serum SCFAs	LC-MS/MS	7

Abbreviations: ASV, amplicon sequence variants; COMETS, COnsortium of METabolomics Studies; KORA, Cooperative Health Research in the Region Augsburg; NMR, nuclear magnetic resonance; LC/GC-MS/MS, liquid chromatography or gas chromatography with tandem mass spectrometry; PREDICT, Personalised Responses to Dietary Composition Trial; SCFA, short-chain fatty acid.

Participants attended a full-day clinical visit consisting of test meal challenges followed by a 13-day home-based phase, as previously described [249]. Briefly, within a tightly controlled clinical setting, participants consumed a meal consisting of breakfast muffins and a milkshake (890 kcal, 85.5g carbohydrate (38.4%), 52.7g fat (53.3%), 16.1g protein (7.2%), and 2.3g fibre at the 0-hour time point, following baseline blood draw). Blood samples were collected at 15, 30, 60, 120, 180, 240, 300, and 360 minutes post-meal.

ZOE PREDICT-1 was used as a replication cohort in **Chapters 4** and **7**. Although participants were recruited from the TwinsUK registry, in the different works presented through this thesis using ZOE PREDICT-1 as a replication cohort, ZOE PREDICT-1 and TwinsUK are completely independent and there is no overlap in participants.

The following subsections describe the main data from ZOE PREDICT-1 used to replicate the results from different chapters.

3.2.1.1 Clinical biochemistry

IL-6 was measured by Affinity Biomarkers Lab using a Sandwich Immunoassay by Meso Scale Diagnostics, while GlycA was measured using the high-throughput NMR metabolomic 2016 panel (Nightingale Ltd.). These inflammatory markers were utilised in **Chapter 7**.

3.2.1.2 Metabolomics

Metabolomics profiling in ZOE PREDICT-1 was conducted on a subset of 332 individuals using LC-MS/MS by Metabolon Inc. as previously described for TwinsUK in **Section 3.1.6.1**. This data was used in **Chapter 4** to replicate the main findings.

3.2.1.2.1 Measures of SCFA levels in serum and stool

SCFA levels were measured in 328 serum and stool samples by Metabolon Inc. as previously described in TwinsUK (see **Section 3.1.6.1.1**). Moreover, postprandial (30 min, 2h and 4h) serum samples were also collected after consuming the standardised meal previously described (see **Section 3.2.1**), and levels were measured following the same technique as in the fasting samples. For each SCFA, the peak and dip were calculated from their respective postprandial measurements. Specifically, the peak was defined as the maximum SCFA concentration in the 4 hours following the test meal challenge minus the fasting level, while the dip was defined as the fasting level minus the minimum SCFA concentration in the 4 hours following the test meal challenge. Circulating (fasting and postprandial) and faecal SCFA levels from ZOE PREDICT-1 were analysed in **Chapter 7**.

3.2.1.3 Gut microbiome composition

The gut microbiome composition was assessed from faecal samples using shotgun sequencing and used in **Chapter 7**. Specifically, faecal samples were collected as described in TwinsUK (see **Section 3.1.7.2.1**). The quality and yield after sample preparation were measured with the Fragment Analyzer system following the manufacturer's guidelines. The size of the resulting product was consistent with the expected size of approximately 500-700 bp. Libraries were sequenced for 300 bp paired-end reads using the Illumina NovaSeq6000 platform according to the manufacturer's protocols. 1.1 nM library was used for flow cell loading. NovaSeq control software NCS v1.5 was used. Image analysis, base calling, and the quality check were performed with the Illumina data analysis pipeline RTA3.3.5 and Bcl2fastq v2.20. Sequenced metagenomes were QCed using the preprocessing pipeline as implemented in Segata's Lab preprocessing. The microbiome taxonomic profiling was performed as described in TwinsUK (see **Section 3.1.7.2.4**).

3.2.1.4 Postprandial metrics

Besides the postprandial SCFA measurements (see **Section 3.2.1.2.1**), measurements of postprandial lipaemic and glycaemic parameters (the 2-h glucose iAUC, rise in triglyceride at 6h postprandially, rise in insulin at 2h postprandially and rise in C-peptide at 2h postprandially) and circulating cytokines (the highest concentration of GlycA and IL-6 within 6h postprandially) were also available. These were used in **Chapter 7**.

3.2.2 Consortium of METabolomics Studies (COMETS)

COMETS is a partnership among 47 worldwide cohorts (**Figure 3.1**) aiming to facilitate large-scale collaborative research on the human metabolome and its relationship with a range of different diseases, such as CVD, hypertension and T2D. It includes metabolic data and information about different outcomes from more than 136,000 individuals [250].

Specifically, the population-based cohorts from the United States and Europe, namely, the Atherosclerosis Risk in Communities (ARIC) study, Edinburgh Type 2 Diabetes Study (ET2DS), GenoDiabMar (GDM), Health, Aging and Body Composition (HABC), and the Women's Health Initiative (WHI), TwinsUK and KORA were used and meta-analysed in **Chapter 5**. A brief description of these cohorts is presented below and in **Table 5.1**. Ethical approval for each study was obtained by the ethical research boards pertaining to each study and was also granted by the COMETS steering committee.

- ARIC: Prospective cohort recruited from 4 U.S communities to investigate the aetiology of atherosclerosis and its clinical outcomes [251].

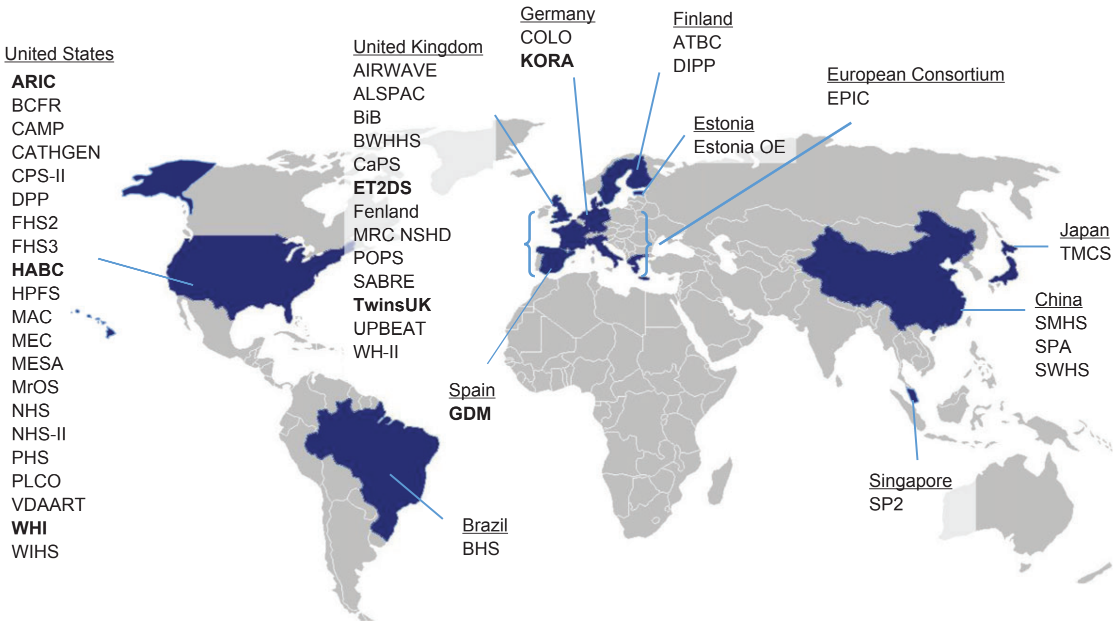


Fig. 3.1 Geographical locations of the studies participating in Consortium of METabolomics Studies (COMETS). The cohorts used in Chapter 5 are indicated in bold. Figure adapted from Yu *et al.*, (2019).

- ET2DS: Longitudinal cohort of older men and women based in Lothian, Scotland, designed to investigate the role of risk factors for vascular complications of T2D [252].
- GDM: Prospective study that aims to provide data on demographic, biochemical, and clinical changes in type-2 diabetic patients attending real medical outpatient consultations [253].
- HABC: Prospective cohort focused on risk factors for the decline of function in initially well-functioning older persons, particularly change in body composition with age [254].
- KORA: A population-based adult cohort that consists of interviews, medical and laboratory examinations, biological sample collection and multiple OMICs data generation and management [255].
- TwinsUK: The largest most clinically characterised adult twin registry in the UK, recruited as volunteers without selecting for particular diseases or traits [218].

- WHI: A large and complex clinical investigation of strategies for the prevention and control of some of the most common causes of morbidity and mortality among postmenopausal women, including cancer, CVD and osteoporotic fractures [256, 257].

The following subsections describe the main data utilised in COMETS.

3.2.2.1 Metabolomics

A summary of the metabolomics methodology used for each cohort is depicted in **Table 5.1**. Serum samples from ARIC, ET2DS, GDM, KORA, and TwinsUK, and samples of EDTA plasma from HABC, and WHI were held at -80°C [250]. Serum metabolites were detected and quantified in ARIC, KORA, and TwinsUK at Metabolon Inc. using untargeted gas and liquid chromatography-mass spectrometry (GC/LC-MS) methods, in ET2Ds and GDM at Nightingale Health using a NMR method. EDTA plasma metabolites were detected and quantified in HABC and WHI at the Broad Institute using LC-MS. Metabolites were harmonised across platforms by manual curation by matching chemical structure, and HMDB and Kyoto Encyclopedia of Genes and Genomes (KEGG) identifiers.

3.2.2.2 MI definition

All the cohorts presented longitudinal MI data, except for KORA which only had cross-sectional MI data. Specific information about how each cohort defined MI is shown in **Supplementary Text 5.1**. Briefly, MI was assessed based on one or more of the following:

- Diagnosed by a doctor (based on clinical evidence such as chest pain, electrocardiogram, and cardiac enzymes).
- Self-reported questionnaires.
- Hospital/GP records.
- Death certificates including the adjudication.

3.2.3 The Cooperative Health Research in the Region of Augsburg (KORA)

The KORA study is a population-based cohort initiated as part of the World Health Organization Multinational Monitoring of Trends and Determinants in Cardiovascular Diseases (MONICA) project since 1984 [255].

In **Chapter 6**, the KORA FF4 study (2013–2014), which is the second follow-up of KORA S4 (1999–2001), was used to replicate results. Samples were collected in the morning between 8:00 A.M. and 10:30 A.M. after at least 8 h of fasting. Ethical approval was obtained from the Bavarian Medical Association Ethics Committee (Bayerische Landesärzte-kammer) and the Bavarian commissioner for data protection and privacy (Bayerischer Datenschutzbeauftragter).

The following subsections describe the main data from KORA used in **Chapter 6**. Of note, KORA was also used as part of COMETS (sub-analysis from **Chapter 5**), and its respective data description is indicated in **Section 3.2.2**.

3.2.3.1 Metabolomics

A LC-MS/MS (Metabolon Inc.) technique was applied for the faecal metabolite profiling (a different version of the platform used in TwinsUK).

3.2.3.2 Prediabetes

Healthy subjects and individuals with IFG were assigned based on the same criteria as in TwinsUK (see **Section 3.1.3**).

3.2.4 Acute trauma case-control cohort

The acute trauma case-control cohort was used as a replication cohort in **Chapter 7**. Patients were all recruited at Queens Medical Hospital part of the Nottingham University Hospital's (NUH) NHS Trust.

- **Rib fracture cohort (OPERA):** Participants needed to be adults (≥16 years) presenting multiple (≥3) rib fractures suitable for surgical repair and having, as per British Orthopaedic Association Audit Standards For Trauma (BOAST-15) Standard 8, indications for fixation as clinical flail chest, respiratory difficulty requiring respiratory support or uncontrollable pain using standard modalities, and was a surgical candidate. Patients were excluded whether (i) they presented a head or thoracic injury requiring emergency intervention, or (ii) could not be operated on within 72 hours as unfit for surgery, or (iii) presented significant thoracic injury requiring surgery where conservative management would be inappropriate. This cohort was collected as part of The Operative Rib Fixation (ORiF) Study (REC Reference: 18/SC/066, IRAS 248460, IRSCTN 10777575).
- **Hip fracture cohort (FEMUR):** Participants needed to (i) be over the age of 65 years (no upper age limit), (ii) with a Rockwood frailty score ≥4, and (iii) have a fractured hip sustained following a fall that required surgery. Moreover, they needed to (iv) have a good understanding of the spoken and written English language, (v) the ability to give informed consent or to provide assent and (vi) the availability of a legally acceptable surrogate to provide consent. Patients were excluded whether (i) fell and sustained the hip fracture more than 12 hours prior to hospitalisation, or (ii) had fallen and sustained a hip fracture whilst in-patient, or (iii) their surgery had to be delayed to 96 hours or more after the fall. This cohort was collected under the Functioning of Elder Muscle; Understanding Recovery (FEMUR) study (REC approval: 20/LO/0841 clinicaltrials.gov registration NCT04764617).
- **Control cohort with measured SCFAs and cytokines:** Healthy students from the School of Medicine at the University of Nottingham or healthcare workers, who had circulating levels of cytokines and SCFA measured. The control individuals were collected under REC ref FMHS 302-0621 by the internal review board of the University of Nottingham School of Medicine.

The following subsections describe the main data utilised in the acute trauma case-control cohorts. The patients with hip or rib fractures were characterised by having the serum samples taken at the time of the patient going into anaesthesia ahead of entering the operating theatre.

3.2.4.1 Circulating cytokines

The pro-inflammatory markers TNF-α, IFN-γ, GlycA and IL-6, and the anti-inflammatory marker IL-10 were measured by Affinity Biomarkers Labs using enzyme-linked immunosorbent assay (ELISA) technique.

3.2.4.2 Circulating SCFAs

Circulating levels of the 8 SCFAs were measured by Metabolon Inc. following the same methodology as described for TwinsUK and ZOE PREDICT-1 in **Sections 3.1.6.1.1** and **3.2.1.2.1**, respectively.

3.3 Statistical analyses

An illustrative overview of the statistical analyses used throughout this thesis is presented in **Figure 3.2**. Statistical analyses and QCs were conducted using R version 1.3.1093 [258]. If not indicated otherwise, all the functions and packages mentioned below are implemented in R. Before running the analyses, the distribution of the continuous variables was checked. If they were not following a normal distribution, different normalisation approaches, including log transformation (when the distribution was left-skewed) and quantile normalization on rank-transformed values, were used to obtain the desired distribution. Likewise, outliers, defined as values 4 standard deviations from the mean, were excluded. P-values were adjusted for multiple testing using the Benjamini and Hochberg method (false discovery rate (FDR) <0.05) [259].

3.3.1 Regression models

Linear regression models are useful and widely used for univariate statistical analyses [64, 109]. Specifically in this thesis, the following regression models were applied throughout

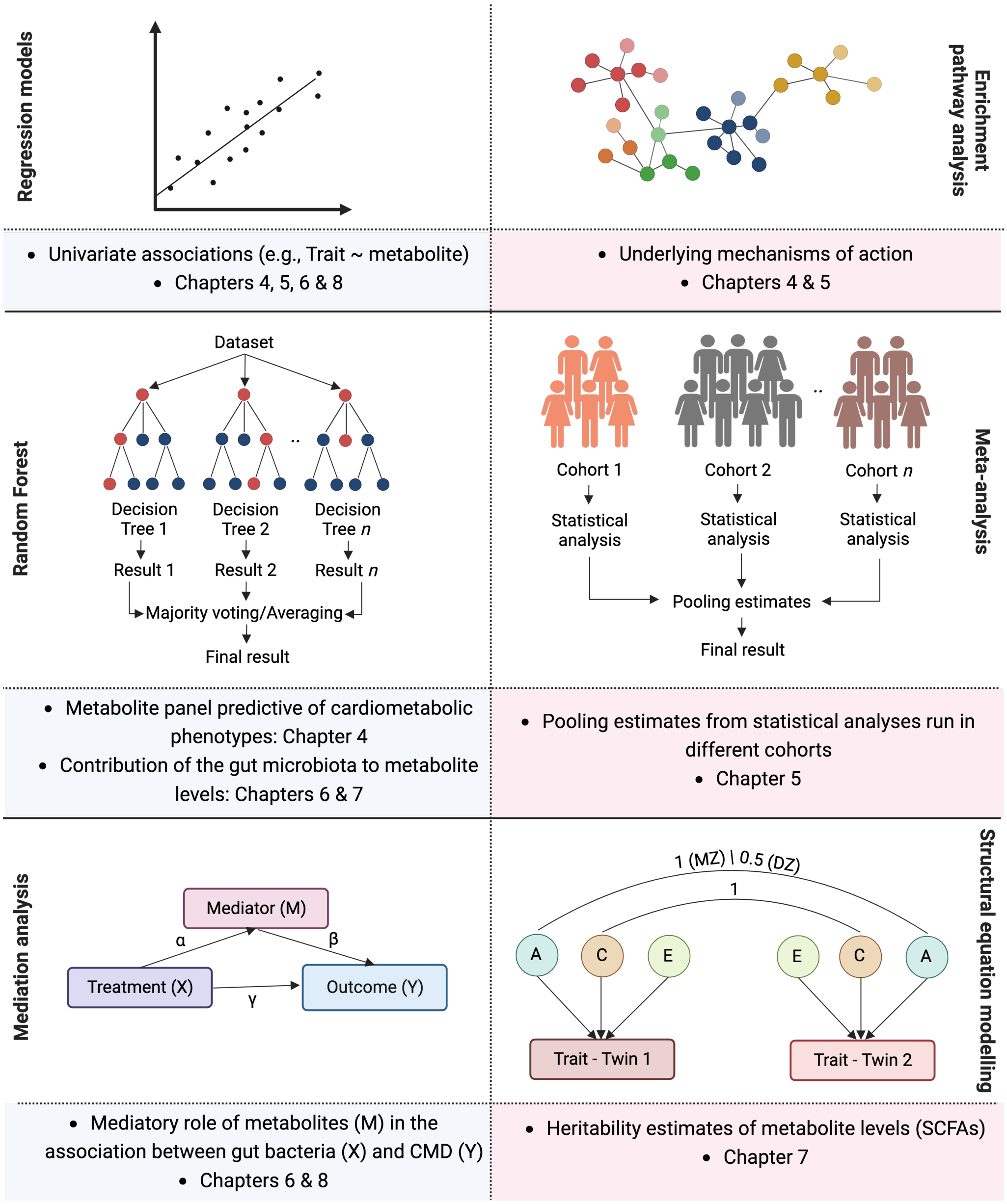

Regression models
Univariate associations (e.g., Trait ~ metabolite)
Chapters 4, 5, 6 & 8
Enrichment pathway analysis
Underlying mechanisms of action
Chapters 4 & 5
Random Forest
Dataset
Decision Tree 1
Decision Tree 2
Decision Tree n
Result 1
Result 2
Result n
Majority voting/Averaging
Final result
Metabolite panel predictive of cardiometabolic phenotypes: Chapter 4
Contribution of the gut microbiota to metabolite levels: Chapters 6 & 7
Meta-analysis
Cohort 1
Cohort 2
Cohort n
Statistical analysis
Statistical analysis
Statistical analysis
Pooling estimates
Final result
Pooling estimates from statistical analyses run in different cohorts
Chapter 5
Mediation analysis
Mediator (M)
α
β
Treatment (X)
Outcome (Y)
γ
Mediatory role of metabolites (M) in the association between gut bacteria (X) and CMD (Y)
Chapters 6 & 8
Structural equation modelling
1 (MZ) | 0.5 (DZ)
1
A
C
E
E
C
A
Trait - Twin 1
Trait - Twin 2
Heritability estimates of metabolite levels (SCFAs)
Chapter 7

Fig. 3.2 An illustrative overview of the main statistical analyses used throughout this thesis. The chapters in which these were used are indicated.

the different chapters. For all of them, it was checked that the residuals followed a normal distribution, and the predictor of interest was Z-scaled to have mean=0 and SD=1.

- **Multiple linear regression models:** This type of model was applied in **Chapters 5** and **6** to identify the circulating and faecal metabolites cross-sectionally associated with MI and IFG, respectively, after controlling for covariates. For that, the ’lm’ function implemented in the ’stats’ package [260] was used. They were also used in

Chapter 4 to obtain the residuals from models where the levels of the metabolites of interest were the response and different covariates were the predictors. These residuals were then used as predictors of other types of models (e.g., RF), and thus further adjustment was not required.

- **Cox Proportional Hazard models:** This statistical model evaluates how various factors influence the rate at which a specific event occurs at a given time point. Therefore, the time interval between these factors and the event under study is considered. Specifically, Cox Proportional Hazard models were used in **Chapters 5** and **6** to identify metabolites predictive of incident MI and T2D, respectively, after adjusting for covariates. For that, the 'coxph' function from the 'survival' package [261] was used.

- **Linear mixed models:** As TwinsUK and ZOE PREDICT-1 consist of twins, linear mixed models were used as they account for family structure (included as random effects) along with other covariates (included as fixed effects). For that, the 'lmer' function implemented in the 'lmerTest' package [262] was utilised. Linear mixed models were used in **Chapter 8** to test the associations between circulating acetate levels and (i) indices of alpha-diversity, (ii) gut bacterial genera abundances, and (iii) visceral fat.

3.3.2 Meta-analysis

Association analyses are often performed in several independent cohorts to obtain larger sample sizes and validate results independently of study effects. Meta-analysis can be conducted to combine the results (effect estimators and p-values) from different cohorts.

There are two main different types of meta-analysis, namely fixed-effect and random-effect meta-analysis. A fixed-effect meta-analysis assumes that all included studies present the same underlying effect estimator, while a random-effect meta-analysis assumes both within-study and between-study variation, incorporating an additional level of variability [263].

Specifically, fixed-effect inverse-variance meta-analyses (using the ’metagen’ function from the ‘meta’ package [264]) were used to combine the results from the MI-serum metabolite associations obtained for each COMETS cohort included in **Chapter 5**, and from the IFG-faecal metabolite associations obtained for TwinsUK and KORA in **Chapter 6**. Heterogeneity between studies and percentage of variability of between-study heterogeneity not due to the sampling error were computed using Cochran’s Q test and I^2 index, respectively. Han-Eskin random-effect meta-analyses were also run to ensure the results were consistent with the fixed-effect results. Han-Eskin random-effect meta-analysis, which is largely applied in GWAS, synthesises data from multiple independent studies while accounting for both within-study and between-study variances [265]. It is able to detect subtle effect sizes by attributing part of the observed variability to random effects, offering enhanced power compared to conventional random-effects models [265]. Meta-analyses were undertaken and reported according to the STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) guidelines (**Supplementary Text 5.2**).

3.3.3 Enrichment pathway analysis

Enrichment pathway analyses were run in **Chapters 4** and **5**, however, in each chapter a different platform was used for it.

In **Chapter 4**, Ingenuity Pathways Analysis (IPA) from QIAGEN Inc. [266] was used to explore the pathways in which the identified metabolite panel associated with the ASCVD risk score were involved. IPA uses the Ingenuity Knowledge Base, which is a repository of curated biological interactions and functional annotations, for pathway analysis and interpretation. Specifically, a right-tailed Fisher’s exact test is used to calculate the p-value determining the probability ($\alpha = 0.05$) that the observed distribution of the input metabolites in a pathway is not explained by chance.

In **Chapter 5**, MetaboAnalyst 5.0 [267] was used to identify the metabolomic pathways enriched for the identified MI-associated metabolites. Within this platform, over-representation analysis was performed using a hypergeometric test to identify groups of compounds that are represented more than expected in each pathway by chance, and pathway topology analysis was performed based on relative betweenness centrality focusing on the entire metabolomic network.

3.3.4 Machine learning: Random Forest

Random Forest (RF) is a tree-based algorithm able to integrate many predictors to build powerful prediction models. As discussed below, for different analyses RF models were preferred over other statistical approaches (e.g., elastic net regression) as they provide some advantages over these [268]. Firstly, RF can model non-linear relationships, and thus, complex relationships between the predictors and the response can be captured [269]. Moreover, predictors of different natures (e.g., numerical or categorical variables) can be combined in the models independently of their distributions [269]. RF can also deal with inter-correlation between predictors as the prediction is based on multiple decision trees including different subsets of predictors [269]. Finally, it can be also used as a feature reduction method as it identifies the most important predictors [269]. RF models were constructed in **Chapters 4**, **6** and **7**, however, the applied methodology slightly differs between them based on their purpose within the work.

In **Chapter 4**, RF models were used to identify a panel of metabolites cross-sectionally associated with the estimated ASCVD, and which further improved the prediction of CVD mortality and morbidity over and above conventional risk factors. Data was split into training and test sets (80:20). Hyperparameters (number of trees and parameters chosen for each split) for the RF classifiers and regressors were tuned using the adaptive resampling search and 5-fold cross-validation. The effect direction of the predictors on the response was examined using the SHapley Additive exPlanations (SHAP) plot [270]. The performance metrics, which were calculated in the test set, were the area under the ROC curve (AUC) for the RF classifiers and R^2 (variance of the response explained by the predictors) for the RF regressors.

In **Chapters 6** and **7**, RF models were used to determine to what extent the gut microbiota composition was associated with the levels of SCFAs and faecal metabolites making up the prediabetes signature, respectively. This algorithm was selected for this goal as it has been repeatedly shown to be particularly suitable and robust to the statistical challenges inherent to microbiome abundance data [271]. Specifically, RF regression (1000 trees and a third of features number as a number of variables randomly sampled as candidates at each split) and classification models (1000 trees and the square root of features number as a number of variables randomly sampled as candidates at each split) with compositional data using 5-fold cross-validation were built. For the classifiers, the continuous response was converted into two classes based on the top and bottom quartiles. To avoid overfitting due to the twin nature of the data used and their shared factors, any twin was removed from the training fold if their twin was present in the test fold. The performance was calculated using the average of the obtained Spearman's correlations (between the observed metabolite levels and the levels predicted by the model - denoted as *rho*) over the 5 folds used as a test set for the regressors, and the average of the obtained AUC values over the testing folds for the classifiers.

For both cases, before running the models, predictors with variance zero or near zero were excluded using the 'nearZeroVar' function implemented in the 'caret' package [272]. The models were created using the 'randomForest' function from the 'randomForest' package [273], and the predictors were ranked based on the node purity. Moreover, as an additional control, it was verified that when randomly swapping the target labels (RF classifiers) or values (RF regressors), the performances were reflecting a random prediction curve (AUC very close to 0.5) and a non-significant Spearman's correlation between the real and predicted values (*rho* close to 0) or a R^2 close to 0.

The main difference between the above approaches is how the hyperparameters were chosen. For the first instance, hyperparameters (e.g., tree number and a number of variables randomly sampled as candidates at each split) were tuned to obtain the set of metabolites providing the best performance in the prediction of the ASCVD risk score, as the performance significantly varied depending on the chosen values. For the second instance, the hyperparameters are set before training the models, and thus, hyperparameter tuning is not applied. The rationale behind this is that the set values have been shown to provide good performance levels when working with gut microbiome compositional data [271, 274]

3.3.5 Mediation analyses

Mediation analysis is a statistical method used to explore the mechanisms that underlie a relationship between an independent variable and a dependent variable via the inclusion of a third explanatory variable, known as a mediator. In a mediation model, the effect of the independent variable on the dependent variable is decomposed into the direct effect (effect not transmitted through the mediator) and the indirect effect (effect transmitted through the mediator). If the indirect effect is significant, this indicates the presence of a mediation effect. This analysis is utilised to gain insights into the network of causal relations.

The key assumptions for mediation analysis are the following:

- Temporal precedence: The independent variable and the mediator must precede the dependent variable in time. Violation of this assumption might lead to erroneous conclusions about causal relationships.

- No confounding variables: No other variable should affect both the mediator and the outcome simultaneously. If the analysis is not adjusted for existing confounding variables, the estimates of the mediation effect may suffer from bias.

- Linearity: The relationships among the variables should be linear. Otherwise, this could lead to an inaccurate estimation of mediation effects.

- Relationship among variables: The independent variable must be related to the dependent variable, and the independent variable must also be related to the mediator. Similarly, the mediator must have a significant relationship with the dependent variable after controlling for the independent variable. Without these relationships, a mediation effect does not exist.

A formal mediation analysis was used in **Chapters 6** and **8** to examine the mediatory role of (i) the faecal metabolites making up the prediabetes signature in the relationship between the gut microbiome and prediabetes, and (ii) circulating acetate in the relationship between the gut microbiome and visceral fat, respectively. For that, the function ’mediate’ implemented in the package ‘mediation’ with 1000 non-parametric bootstrap samples was used [275]. The variance accounted for (VAF) score, which represents the ratio of indirect-to-total effect and determines the proportion of the variance explained by the mediation process, was used to determine the significance of the mediation effect.

3.3.6 Structural equation modelling

To estimate the heritability of the levels of a given metabolite (either serum or stool), a classical twin model was applied and compared the degree of similarity among MZ twins, who share 100% of their genetic make-up, and DZ twins, who share on average 50% of their segregating genes. Under the equal environment assumption, the variance of the trait/phenotype is explained by three latent parameters: additive genetic variance (A), shared (familial) environmental variance (C) and individual-specific environmental variance/error (E) [276]. Additive genetic influences are indicated when MZ twins are more similar than DZ twins. The shared environmental component estimates the contribution of the family environment, which is assumed to be equal in both MZ and DZ twin pairs [276]. The environmental component does not contribute to twin similarity, it rather estimates the effects that apply only to each individual and includes measurement error. Any greater similarity between MZ twins than DZ twins is attributed to a greater sharing of genetic influences. Specifically, structural equation models, which use the observed covariates from both MZ and DZ pairs to establish a causal relationship between them and the latent parameters, were built using the ’twinlm’ function implemented in the ’METs’ package [277]. This analysis was used in **Chapter 7** to estimate the heritability of the SCFA levels in serum and stool.

3.3.7 Genomic characterisation of gut bacterial genera

In **Chapter 8**, the identified acetate-associated gut genera (*Lachnoclostridium* and *Coprococcus*) were genomically characterised. To do so, *Lachnoclostridium* and *Coprococcus* metagenomes were retrieved as explained below and the different analyses described in the following subsections were conducted.

3.3.7.1 Metagenome retrieval and preliminary filtering

First, the metagenomes belonging to these genera and their corresponding metadata were retrieved from the UHGG catalog and RefSeq dataset (January, 2021) [278]. RefSeq genomes derived from metagenomes and not sampled from human faeces, stool or the gastrointestinal tract were removed. Inconsistencies related to the variable country were corrected and the missing sample accessions were added. Genomes from sample identifiers not found in the National Center for Biotechnology Information (NCBI) [279] were discarded. The two datasets were merged and then filtered by completeness, contamination and contig number (>90%, <3%, and <400 for *Lachnoclostridium* and >95%, <1%, and <300 for *Coprococcus*, respectively). The thresholds in *Lachnoclostridium* were less strict due to the scarcity of genomes presenting higher standards. Duplicated genomes were discarded, keeping the one with the highest N50 value. Finally, genomes from uncharacterised species or misclassified species were renamed based on the cluster given by fastANI classification (see **Section 3.3.7.3**).

3.3.7.2 Quality assessment of genome assemblies and genome annotation

Completeness and contamination were estimated with CheckM version 1.1.3 [280] using the 'lineage_wf' workflow. QUAST version 5.0.2 [281] was run to retrieve the total length, GC-content, contig number and N50. Genome annotation was performed using Prokka version 1.12 [282] using the default parameters.

3.3.7.3 Average nucleotide identity-based taxonomic classification

FastANI version 1.32 [283] was separately run on *Lachnoclostridium* and *Coprococcus* genomes to calculate the average nucleotide identity (ANI) between all pairs of sequences. Results were filtered by the alignment fraction (>0.4), and symmetric pairwise ANI dissimilarities (100-95, ANI = 95%) were calculated from the ANI values to construct a dendrogram for each genus using the single linkage hierarchical clustering method ('hclust' function from the 'stats' package [260]). Two networks analyses based on the information given by the dendrograms were conducted using the 'layoutwithdrl' layout implemented in the ’igraph’ package [284] with an expansion and simmer attraction of 0, and an innit, liquid and crunch temperature of 100, 50, and 50, respectively.

3.3.7.4 Phylogeny inference at the genus level

Evolutionary relationships among the *Coprococcus* and *Lachnoclostridium* species were inferred using ezTree version 0.1 [285]. For each species, up to three genomes (depending on the number of available genomes) sequenced from isolates were used as input. If genomes sequenced from isolates were not available, then the metagenome-assembled genomes (MAGs) with the highest completeness percentage were selected.

3.3.7.5 Prediction of functional capabilities

Metabolic Pathway Database (MetaCyc) [286] and KEGG [287] information for each genome was retrieved using the enzyme commission numbers from the gff files generated by Prokka and MinPath (Minimal set of Pathways) [288] (**Supplementary Table 8.5**). The retrieved information was utilised to construct heatmaps (’Heatmap’ function implemented in the ’ComplexHeatmap’ package [289]) showing the genome percentage of each species with a given pathway. For KEGG data, only the highly different pathways between species were selected (for a given pathway, at least one species has a percentage <5% and another species has a percentage >80%). Moreover, a principal component analysis (PCA) was performed using the presence/absence matrix with the MetaCyc biosynthesis/degradation pathways using the ’prcomp’ function within the ’stats’ package [260].

Chapter 4

A metabolites panel associated with cardiovascular risk

The ASCVD risk score is a tool used to estimate the 10-year CVD risk of an individual based on traditional risk factors. Although these factors considerably contribute to disease risk, they might not show enough predictive power to identify at-risk individuals before the disease's onset.

As individual circulating metabolites have been associated with cardiovascular traits, in this chapter, I search for a panel of serum metabolites associated with the ASCVD risk score and predictive of CVD mortality and morbidity independently of environmental and traditional risk factors. I also explore the pathways in which these metabolites are involved to better understand their mechanisms of action.

The obtained findings shed light on a panel of serum metabolites that has the potential to be used for early prediction and treatment of CVD.

Coauthor Dr Panayiotis Louca ran the Ingenuity Pathway Analysis. I calculated the ASCVD risk score, performed the statistical analyses and wrote the original draft of the manuscript.

This chapter has been published as a research letter in *Journal of the American Heart Association* (Nogal et al., 2022).

Journal of the American Heart Association

RESEARCH LETTER

Incremental Value of a Panel of Serum Metabolites for Predicting Risk of Atherosclerotic Cardiovascular Disease

Ana Nogal, MSc; Panayiotis Louca, MSc; Tran Quoc Bao Tran, MBBS; Ruth C. Bowyer, PhD; Paraskevi Christofidou, PhD; Claire J. Steves, PhD; Sarah E. Berry, PhD; Kari Wong, PhD; Jonathan Wolf, MA; Paul W. Franks, PhD; Massimo Mangino, PhD; Tim D. Spector, MD; Ana M. Valdes, PhD*; Sandosh Padmanabhan, PhD*; Cristina Menni, PhD*

Cardiovascular diseases (CVDs) are the leading causes of mortality and morbidity worldwide, accounting for 17.3 million deaths per year.[1] The American College of Cardiology/American Heart Association 10-year atherosclerotic CVD risk score is a sex- and race-specific single multivariable risk assessment tool used to estimate the 10-year CVD risk of an individual based on age, sex, and traditional risk factors (TRFs), including high-density lipoprotein and total cholesterol, blood pressure, blood pressure medications, smoking, and type 2 diabetes.[1] These factors contribute considerably to disease risk, although they may not identify at-risk individuals before disease onset.[2,3] Previous studies found circulating metabolites predictive of cardiovascular traits, mostly using linear approaches and a limited number of metabolites.[3–5]

By combining the effects of a larger number of individual biomarkers, TRFs, and environmental variables, we applied a machine learning technique to identify a metabolite panel cross-sectionally associated with estimated atherosclerotic CVD (eASCVD) risk and longitudinally predictive of CVD mortality and morbidity in a population-based cohort with independent replication, to gain further insights into the metabolic pathways underlying CVD risk.

The data used in this study are held by the Department of Twins Research at King's College London. The data can be released to bona fide researchers using our normal procedures overseen by the Wellcome Trust and its guidelines as part of our core funding (https://twinsuk.ac.uk/resources-for-researchers/access-our-data/). The scripts in R and all the necessary information to replicate the findings reported in this article are publicly available at https://github.com/ananogal1/ASCVD-metabolite-panel.

The flowchart of the study design is depicted in the Figure (A). We included women from TwinsUK[1] with fasting serum metabolomic profiling (533 metabolites; Metabolon) along with eASCVD,[1] TRFs, diet (healthy eating index),[1] menopause status, and physical activity at 2 time points 6 years apart (SD=2) (Figure [B]). Individuals with prevalent CVD were excluded. TwinsUK provided informed written consent, and the study was approved by the St. Thomas' Hospital Research Ethics Committee (REC Ref: EC04/015).

Metabolites were inverse normalized, and missing values imputed using minimum run-day measures. For each metabolite, we calculated residuals by running linear regressions adjusting for age, body mass index, menopause status, diet, and physical activity. To identify a metabolite panel associated with eASCVD, we built random forest models on the residuals at each time point, splitting the data set into training and test sets (80:20). We tuned hyperparameters

Key Words: atherosclerosis ■ biomarkers ■ cardiovascular disease risk ■ machine learning ■ serum metabolites

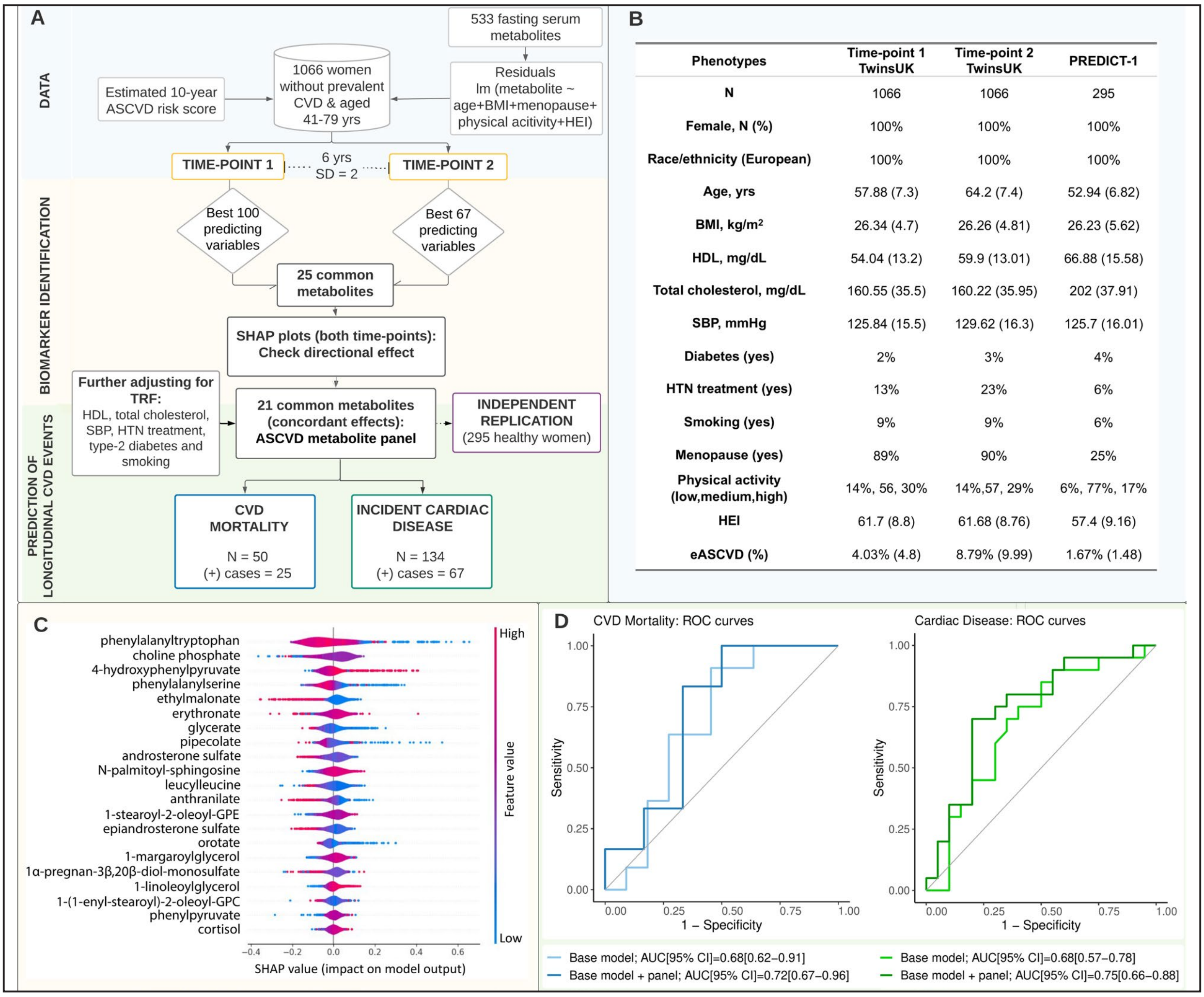
Correspondence to: Cristina Menni, PhD, Department of Twin Research, King's College London, St Thomas' Hospital Campus, Westminster Bridge Road, London SE1 7EH, United Kingdom. E-mail: cristina.menni@kcl.ac.uk

*A. M. Valdes, S. Padmanabhan, and C. Menni contributed equally.

For Sources of Funding and Disclosures, see page 3.

JAHA is available at: www.ahajournals.org/journal/jaha

using the adaptive resampling search and used 5-fold cross-validation and node purity to select the optimal predictors' number. We identified common predictors between the 2 time points and examined the effect on model prediction using the Shapley additive explanations plot. Common metabolites with concordant effects at both time points were included in the eASCVD metabolites panel. Results were replicated in 295 women from PREDICT-1 (Personalised Responses to Dietary Composition Trial).[1] We further tested the incremental area under the curve (AUC) value of the eASCVD metabolites panel in predicting incident cardiac disease (including congestive heart disease, angina, atrial fibrillation, and coronary heart disease) and CVD mortality (through record linkage with the Office for National Statistics [ONS]) in independent sets of 50 to



Phenotypes	Time-point 1 TwinsUK	Time-point 2 TwinsUK	PREDICT-1
N	1066	1066	295
Female, N (%)	100%	100%	100%
Race/ethnicity (European)	100%	100%	100%
Age, yrs	57.88 (7.3)	64.2 (7.4)	52.94 (6.82)
BMI, kg/m^2	26.34 (4.7)	26.26 (4.81)	26.23 (5.62)
HDL, mg/dL	54.04 (13.2)	59.9 (13.01)	66.88 (15.58)
Total cholesterol, mg/dL	160.55 (35.5)	160.22 (35.95)	202 (37.91)
SBP, mmHg	125.84 (15.5)	129.62 (16.3)	125.7 (16.01)
Diabetes (yes)	2%	3%	4%
HTN treatment (yes)	13%	23%	6%
Smoking (yes)	9%	9%	6%
Menopause (yes)	89%	90%	25%
Physical activity (low,medium,high)	14%, 56, 30%	14%,57, 29%	6%, 77%, 17%
HEI	61.7 (8.8)	61.68 (8.76)	57.4 (9.16)
eASCVD (%)	4.03% (4.8)	8.79% (9.99)	1.67% (1.48)

Figure. Serum metabolites associated to atherosclerotic cardiovascular disease: flowchart, data, and main results.
A, Flowchart of the study design. "N" indicates the number of individuals included to build the random forest (RF) classifiers, whereas "(+) cases" refers to the number of individuals suffering from a specific cardiovascular disease (CVD) phenotype. **B**, Demographic characteristics of the study samples PREDICT-1 (TwinsUK and Personalised Responses to Dietary Composition Trial). Demographic characteristics by outcome (ie, incident cardiac disease and CVD mortality) are provided on GitHub. **C**, Directional effect of each single metabolite from the estimated atherosclerotic cardiovascular disease (eASCVD) risk panel on the model predictions using a Shapley additive explanations (SHAP) plot. The SHAP values (*x* axis) quantify the magnitude and direction (positive or negative using the feature values) of each metabolite on the target variable (ASCVD). Each point represents a feature instance, whereas the color indicates the feature value (high=red, low=blue). **D**, Area under the curve (AUC) values and receiver operating characteristic (ROC) curves obtained for RF classifiers built on (1) the base model including environmental and traditional risk factors and (2) the base model plus the eASCVD metabolites panel. Each ROC curve represents the performance of the RF classifiers in predicting each CVD event (CVD mortality and incident cardiac disease) at different classification thresholds (range = 0–1). The AUC is computed for each curve and used as a model performance metric. ASCVD indicates atherosclerotic cardiovascular disease; BMI, body mass index; GPC, glycerophosphocholine; GPE, glycerophosphoethanolamine; HEI, health eating index; HDL, high-density lipoprotein; HTN, hypertension; lm, linear models; SBP, systolic blood pressure; and TRF, traditional risk factors.

134 individuals (follow-up, 5.6 years [SD, 2.2 years]). Finally, we explored the pathways in which the identified metabolites were involved using Ingenuity Pathway Analysis (QIAGEN; Fisher exact test, false discovery rate [Benjamini-Hochberg] <0.05).

The random forest models on residuals in 1066 TwinsUK women adjusted for age, body mass index, menopause, physical activity, and diet identified 100 and 67 predictors of eASCVD at time point 1 and 2, respectively, of which 25 were overlapping. Of these, 21 had concordant effects at both time points and were included in the eASCVD metabolites panel. After adjusting for family, the panel explained 12.7% of the variance in eASCVD in the test set and 13.6% in PREDICT-1. When further adjusting for TRFs, the panel explained 9.3% in the test set and 8.5% in PREDICT-1. Among the metabolites identified, 9 were positively associated with eASCVD, whereas 12 were negatively associated (Figure [C]). The peptide phenylalanyltryptophan, the lipid choline phosphate, and the amino acid 4-hydroxyphenylpyruvate were the most important contributors (Figure [C]). The incremental predictive value of the eASCVD-metabolites panel over environmental and TRFs improved prediction of incident cardiac disease by 7% (AUC from 0.68 [95% CI, 0.57–0.78] to 0.75 [95% CI, 0.66–0.88]) and CVD mortality by 4% (AUC from 0.68 [95% CI, 0.62–0.91] to 0.72 [95% CI, 0.67–0.96]) (Figure [D]). Finally, pathway enrichment analysis highlighted the involvement (false discovery rate range = 0.01–0.02) of the metabolites positively associated with eASCVD in the biosynthesis of 4-hydroxyphenylpyruvate, choline, phosphatidylcholine and glucocorticoids, sphingomyelin metabolism, tyrosine degradation, and phospholipases. Moreover, the panel was enriched (false discovery rate range = 0.001–0.04) in metabolites related to cardiac inflammation, dysfunction damage, and infarction.

Here, we report for the first time a panel of serum metabolites correlated with eASCVD explaining 9.3% of the variance not already explained by environmental and TRFs. The panel further improved prediction of incident cardiac disease and CVD mortality over and above conventional risk factors, thereby generating new research avenues. Metabolites positively associated with eASCVD are enriched in pathways previously linked with atherosclerotic CVD.[2] The sphingomyelin:-phosphatidylcholine ratio, choline and glucocorticoids biosynthesis, tyrosine degradation, and phospholipases have been shown to increase the CVD risk and/or mortality risk.[2,4,5] Therefore, this study sheds light into the metabolites behind these pathways.

Limitations include the homogeneous ethnicity and women-only composition of the samples, the lack of longitudinal data in PREDICT-1, and the limited number of CVD events. However, we benefit from cross-sectional ASCVD data, independent data sets to test the panel predictive power, and independent replication. Our results illustrate how metabolic profiling along with machine learning might identify novel biomarkers implicated in CVD, which are crucial for early diagnosis and treatment.

ARTICLE INFORMATION

Received November 11, 2021; accepted December 7, 2021.

Affiliations

Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom (A.N., P.L., R.C.B., P.C., C.J.S., M.M., T.D.S., A.M.V., C.M.); Institute of Cardiovascular & Medical Sciences, University of Glasgow, Glasgow, United Kingdom (T.Q.T., S.P.); Department of Nutritional Sciences, King's College London, London, United Kingdom (S.E.B.); Metabolon, Morrisville, NC (K.W.); Zoe Limited, London, United Kingdom (J.W.); Department of Clinical Sciences, Lund University Diabetes Center, Lund University, Malmö, Sweden (P.W.F.); Harvard T. H. Chan School of Public Health, Boston, MA (P.W.F.); NIHR Biomedical Research Centre at Guy's and St. Thomas' Foundation Trust, London, United Kingdom (M.M.); and Injury, Recovery and Inflammation Sciences Clinical Sciences Building, Nottingham City Hospital, University of Nottingham, Nottingham, United Kingdom (A.M.V.).

Acknowledgments

The authors thank all of the participants of TwinsUK for contributing and supporting our research.

Sources of Funding

This research was funded by the Chronic Disease Research Foundation and in part by the Wellcome Trust (grant number: 212904/Z/18/Z). For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) public copyright license to any author-accepted article version arising from this submission. TwinsUK receives funding from the Wellcome Trust, the European Commission H2020 grants SYSCID (contract number 733100); the National Institute for Health Research (NIHR) Clinical Research Facility and the Biomedical Research Centre based at Guy's and St. Thomas' National Health Service (NHS) Foundation Trust in partnership with King's College London, the Chronic Disease Research Foundation, the UK Research and Innovation (UKRI) Medical Research Council (MRC)/British Heart Foundation, Ancestry and Biological Informative Markers for Stratification of Hypertension (AIM-HY) (MR/M016560/1), and Zoe Limited. Dr Mangino is supported by the NIHR Clinical Research Facility and the Biomedical Research Centre based at Guy's and St. Thomas' NHS Foundation Trust in partnership with King's College London. Dr Menni is funded by the Chronic Disease Research Foundation and by the UKRI MRC AIM-HY project grant (MR/M016560/1). A. Nogal and P. Louca are funded by the Chronic Disease Research Foundation. Dr Valdes is supported by the National Institute for Health Research Nottingham Biomedical Research Centre. Dr Christofidou is funded by the European Union (H2020 contract number 733100). Dr Franks receives support from the Swedish Research Council, Swedish Heart-Lung Foundation and the Swedish Foundation for Strategic Research (LUDC-IRC 15-0067). Dr Padmanabhan is funded by the UKRI MRC AIM-HY project grant (MR/M016560/1); the British Heart Foundation (BHF) (CS/16/1/31878) and BHF Centre of Excellence (RE/18/6/34217). The PREDICT cohort is supported by Zoe Limited.

Disclosures

Drs Valdes, Franks, and Berry are consultants to Zoe Limited, and J. Wolf and Dr Spector are cofounders of Zoe Limited. Dr Wong is an employee of Metabolon Inc. All other authors declare no competing financial interests.

REFERENCES

1. Berry SE, Valdes AM, Drew DA, Asnicar F, Mazidi M, Wolf J, Capdevila J, Hadjigeorgiou G, Davies R, Al Khatib H, et al. Human postprandial responses to food and potential for precision nutrition. *Nat Med.* 2020;26:964–973. doi: 10.1038/s41591-020-0934-0

2. Ussher JR, Elmariah S, Gerszten RE, Dyck JR. The emerging role of metabolomics in the diagnosis and prognosis of cardiovascular disease. *J Am Coll Cardiol*. 2016;68:2850–2870. doi:10.1016/j.jacc.2016.09.972
3. Wang Z, Zhu C, Nambi V, Morrison AC, Folsom AR, Ballantyne CM, Boerwinkle E, Yu B. Metabolomic pattern predicts incident coronary heart disease: findings from the atherosclerosis risk in communities study. *Arterioscler Thromb Vasc Biol*. 2019;39:1475–1482. doi: 10.1161/ATVBAHA.118.312236
4. Murthy VL, Reis JP, Pico AR, Kitchen R, Lima JAC, Lloyd-Jones D, Allen NB, Carnethon M, Lewis GD, Nayor M, et al. Comprehensive metabolic phenotyping refines cardiovascular risk in young adults. *Circulation*. 2020;142:2110–2127. doi: 10.1161/CIRCULATIONAHA.120.047689
5. Cavus E, Karakas M, Ojeda FM, Kontto J, Veronesi G, Ferrario MM, Linneberg A, Jørgensen T, Meisinger C, Thorand B, et al. Association of circulating metabolites with risk of coronary heart disease in a European population: results from the Biomarkers for Cardiovascular Risk Assessment in Europe (BiomarCaRE) Consortium. *JAMA Cardiol*. 2019;4:1270–1279. doi: 10.1001/jamacardio.2019.4130

Supplementary material

Supplementary Table 4.1 Demographics characteristics of the different cohorts used.

Phenotypes	Time-point 1 TwinsUK	Time-point 2 TwinsUK	PREDICT-1	Cardiac disease TwinsUK	CVD mortality TwinsUK
N	1066	1066	295	134	50
Female, N (%)	100%	100%	100%	100%	100%
European race/ethnicity (%)	100%	100%	100%	100%	100%
Age, yrs	57.88 (7.3)	64.2 (7.4)	52.94 (6.82)	66.42 (8.46)	68.49 (8.65)
BMI, kg/m^2	26.34 (4.7)	26.26 (4.81)	26.23 (5.62)	26.56 (4.67)	26.94 (4.68)
HDL, mg/dL	54.04 (13.2)	59.9 (13.01)	66.88 (15.58)	56.26 (13.14)	54.61 (11.99)
Total cholesterol, mg/dL	160.55 (35.5)	160.22 (35.95)	202 (37.91)	143.98 (35.18)	146.85 (34.5)
SBP, mmHg	125.84 (15.5)	129.62 (16.3)	125.7 (16.01)	129.78 (17.06)	134.59 (20.41)
Type-2 diabetes (yes)	2%	3%	4%	5%	2%
HTN treatment (yes)	13%	23%	6%	37%	50%
Smoking (yes)	9%	9%	6%	5%	14%
Menopause (yes)	89%	90%	25%	90%	92%
Physical activity (low,medium,high)	14%, 56%, 30%	14%,57%, 29%	6%, 77%, 17%	19%, 52%, 29%	12%, 62%, 26%
HEI	61.7 (8.8)	61.68 (8.76)	57.4 (9.16)	62.37 (8.65)	59.96 (8.77)
eASCVD (%)	4.03% (4.8)	8.79% (9.99)	1.67% (1.48)	-	-

Abbreviations: BMI, body mass index; CVD, cardiovascular disease; eASCVD, estimated atherosclerotic cardiovascular disease; HDL, high-density lipoprotein; HEI, health eating index; HTN, hypertension; PREDICT-1, Personalized Responses to Dietary Composition Trial-1; SBP, systolic blood pressure.

Supplementary Table 4.2 Metabolite IDs used to run the pathway enrichment analysis with Ingenuity Pathway Analysis (IPA) (QIAGEN Inc.).

Metabolite name	Subpathway	Superpathway	HMDB	CAS number	Metabolite ID	PubChem ID
1-(1-enyl-stearoyl)-2-oleoyl-GPC (P-18:0/18:1)	phospholipid	Lipid	HMDB0011243			
1-linoleoylglycerol (18:2)	Monoacylglycerol	Lipid		2277-28-3		
1-margaroylglycerol (17:0)	Monoacylglycerol	Lipid			ME273084	
1-stearoyl-2-oleoyl-GPE (18:0/18:1)	Phospholipid Metabolism	Lipid	HMDB08993			
4-hydroxyphenylpyruvate	Phenylalanine and Tyrosine Metabolism	Amino Acid	HMDB00707			
5alpha-pregnan-3beta, 20beta-diol monosulfate (1)	Steroid	Lipid				
androsterone sulfate	Steroid	Lipid	HMDB02759			
anthranilate	Tryptophan Metabolism	Amino Acid	HMDB01123			
choline phosphate	Phospholipid Metabolism	Lipid	HMDB01565			
cortisol	Steroid	Lipid	HMDB0000063			
epiandrosterone sulfate	Steroid	Lipid	HMDB00365			
erythronate*	Aminosugar Metabolism	Carbohydrate	HMDB00613			
ethylmalonate	Leucine, Isoleucine and Valine Metabolism	Amino Acid	HMDB00622			
glycerate	Glycolysis, Gluconeogenesis, and Pyruvate Metabolism	Carbohydrate	HMDB00139			
leucylleucine	Dipeptide	Peptide	HMDB28933			
N-palmitoyl-sphingosine (d18:1/16:0)	Sphingolipid Metabolism	Lipid	HMDB04949			
orotate	Pyrimidine Metabolism, Orotate containing	Nucleotide	HMDB00226			
phenylalanylserine	Dipeptide	Peptide				193508
phenylalanyltryptophan	Dipeptide	Peptide				134906
phenylpyruvate	Phenylalanine and Tyrosine Metabolism	Amino Acid	HMDB00205			
pipecolate	Lysine Metabolism	Amino Acid	HMDB00070			

Abbreviations: CAS, Chemical Abstracts Service; HMDB, Human Metabolome Database.

Chapter 5

Circulating biomarkers of incident myocardial infarction

MI is one of the main causes of CVD. Previous studies have reported metabolites associated with MI. However, these are limited by the participant number and/or the demographic diversity, hampering the identification of wide-spectrum biomarkers of MI.

In this chapter, I search for circulating biomarkers predictive of incident MI and explore the potential underlying mechanisms of action in the largest metabolome study of MI to date, which consists of 6 intercontinental COMETS cohorts with diverse race/ethnic backgrounds.

The obtained findings shed light on novel metabolic preventive biomarkers of MI and the involved pathways. They might help to identify high-risk individuals before the disease onset and pave the way towards the development of novel preventative strategies.

This work is part of the large international metabolomics consortium (COMETS) with many people and analysts involved. Collaborator Dr Domagoj Kifer wrote the pipeline to run the association analysis at each contributing cohort. I coordinated the different involved cohorts and ran the association analyses on the TwinsUK and GDM cohorts. With co-author Ms Taryn Alkis, I performed the fixed- and random-effect meta-analyses and the enrichment pathway analysis. Along with co-authors Ms Taryn Alkis and Ms Yura Lee, I wrote the first draft of the manuscript.

This chapter has been published in *Cardiovascular Research* (Nogal et al., 2023). An extension of the discussion, which is not included in the published manuscript, can be found in **Appendix B**.

ESC
European Society
of Cardiology

Cardiovascular Research (2023) **00**, 1–12
https://doi.org/10.1093/cvr/cvad147

Predictive metabolites for incident myocardial infarction: a two-step meta-analysis of individual patient data from six cohorts comprising 7897 individuals from the COnsortium of METabolomics Studies

Ana Nogal[1†], Taryn Alkis[2†], Yura Lee[2†], Domagoj Kifer[3], Jie Hu[4], Rachel A. Murphy [5,6], Zhe Huang [7], Rui Wang-Sattler [8], Gabi Kastenmüler[9], Birgit Linkohr[10], Clara Barrios[11], Marta Crespo[11], Christian Gieger[8], Annette Peters [10], Jackie Price[7], Kathryn M. Rexrode[4], Bing Yu[2‡*], and Cristina Menni [1‡*]

[1]Department of Twin Research, King's College London, St Thomas' Hospital Campus, Westminster Bridge Road, SE1 7EH London, UK; [2]Department of Epidemiology, Human Genetics and Environmental Sciences, University of Texas Health Science Center at Houston School of Public Health, 1200 Pressler St, Suite E407, Houston, 77030 TX, USA; [3]Faculty of Pharmacy and Biochemistry, University of Zagreb, Zagreb, Croatia; [4]Division of Women's Health, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA; [5]Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada; [6]Cancer Control Research, BC Cancer, Vancouver, BC, Canada; [7]Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK; [8]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany; [9]Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München, Neuherberg, Germany; [10]Institute of Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany; and [11]Department of Nephrology, Hospital del Mar, Institut Hospital del Mar d´Investigacions Mediques, Barcelona, Spain

Received 6 March 2023; revised 28 June 2023; accepted 18 July 2023; online publish-ahead-of-print 14 September 2023

Time of primary review: 48 days

Aims

Myocardial infarction (MI) is a major cause of death and disability worldwide. Most metabolomics studies investigating metabolites predicting MI are limited by the participant number and/or the demographic diversity. We sought to identify biomarkers of incident MI in the COnsortium of METabolomics Studies.

Methods and results

We included 7897 individuals aged on average 66 years from six intercontinental cohorts with blood metabolomic profiling ($n =$ 1428 metabolites, of which 168 were present in at least three cohorts with over 80% prevalence) and MI information (1373 cases). We performed a two-stage individual patient data meta-analysis. We first assessed the associations between circulating metabolites and incident MI for each cohort adjusting for traditional risk factors and then performed a fixed effect inverse variance meta-analysis to pull the results together. Finally, we conducted a pathway enrichment analysis to identify potential pathways linked to MI. On meta-analysis, 56 metabolites including 21 lipids and 17 amino acids were associated with incident MI after adjusting for multiple testing (false discovery rate < 0.05), and 10 were novel. The largest increased risk was observed for the carbohydrate mannitol/sorbitol {hazard ratio [HR] [95% confidence interval (CI)] = 1.40 [1.26–1.56], $P < 0.001$}, whereas the largest decrease in risk was found for glutamine [HR (95% CI) = 0.74 (0.67–0.82), $P < 0.001$]. Moreover, the identified metabolites were significantly enriched (corrected $P < 0.05$) in pathways previously linked with cardiovascular diseases, including aminoacyl-tRNA biosynthesis.

Conclusions

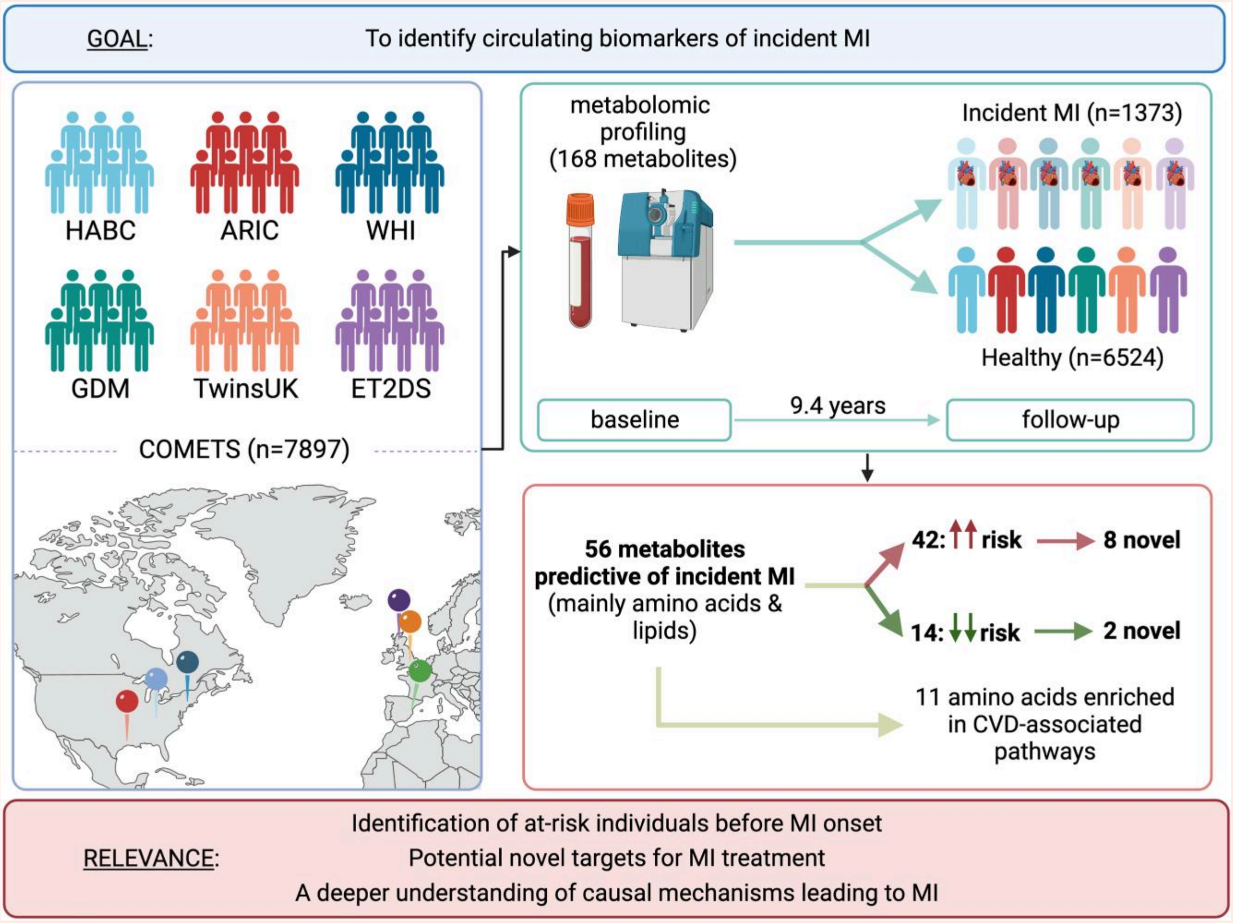
In the most comprehensive metabolomic study of incident MI to date, 10 novel metabolites were associated with MI. Metabolite profiles might help to identify high-risk individuals before disease onset. Further research is needed to fully understand the mechanisms of action and elaborate pathway findings.

* Corresponding author. Tel: +44 0207 188 7188 (ext. 52594), Email: cristina.menni@kcl.ac.uk (C.M.); Tel: +1 713 500 9285, Email: bing.yu@uth.tmc.edu (B.Y.)

† The first two authors contributed equally to the study.

‡ The last two authors contributed equally to the study.

Graphical Abstract


GOAL: To identify circulating biomarkers of incident MI
HABC
ARIC
WHI
GDM
TwinsUK
ET2DS
COMETS (n=7897)
metabolomic profiling (168 metabolites)
Incident MI (n=1373)
Healthy (n=6524)
baseline
9.4 years
follow-up
56 metabolites predictive of incident MI (mainly amino acids & lipids)
42: ↑↑ risk → 8 novel
14: ↓↓ risk → 2 novel
11 amino acids enriched in CVD-associated pathways
RELEVANCE:
Identification of at-risk individuals before MI onset
Potential novel targets for MI treatment
A deeper understanding of causal mechanisms leading to MI

Keywords Myocardial infarction • Metabolomics • Biomarkers • Two-step individual patient data meta-analysis • Amino acids

1. Introduction

Cardiovascular diseases (CVD) are a huge public health burden accounting for 32% of all global deaths in 2019.[1] Myocardial infarction (MI) is one of the main causes of CVD, causing the death of one person every 40 s in the USA[2] and one hospital admission every 5 min in the UK.[3]

Besides the well-established risk factors associated with MI, such as obesity, diabetes, hypertension, and smoking,[4] many studies suggest that circulating metabolites might play an important role in MI development.[5,6] For instance, glycine has been recognized as a protective biomarker of cardiac diseases, especially coronary heart disease,[7] whereas trimethylamine *N*-oxide (TMAO) has been associated with MI by accelerating atherosclerosis.[5,6]

Metabolomics enables the comprehensive characterization of small-weight molecules, such as carbohydrates, amino acids, lipids, nucleotides, and peptides,[8–10] providing a snapshot of the individual's metabolic state at a particular time. Thus, metabolites might enable the identification of at-risk individuals before the disease process is well underway.[11,12]

Advances in this field have allowed the detection of metabolites whose deregulation may be involved in the onset and development of complex diseases including CVD,[13,14] cancer,[15] and autoimmune diseases.[16] Nonetheless, most metabolomic studies are limited by the number of participants and/or the demographic diversity, affecting the statistical power of the results and hampering the discovery of potential universal biomarkers.[13,17] To address these issues, the COnsortium of METabolomics Studies (COMETS) was established in 2014, aggregating metabolic data from 47 cohorts from around the world.[17]

By using individual patient data (IPD) from six COMETS cohorts with MI and metabolomic data, we aimed to identify biomarkers associated with incident MI in 7897 participants. We further explored the pathways in which these metabolites might be involved to better understand their mechanisms of action.

2. Methods

2.1 Study populations

For the primary analysis of metabolites associated with incident MI, we included participants from six population-based cohorts from the USA and Europe, namely, the Atherosclerosis Risk in Communities (ARIC) study, Edinburgh Type 2 Diabetes Study (ET2DS), GenoDiabMar (GDM), Health, Aging and Body Composition (HABC), TwinsUK, and the Women's Health Initiative (WHI). Secondary analyses of metabolites associated with prevalent MI included participants from ARIC, ET2DS, GDM, HABC, TwinsUK, and Cooperative Health Research in the Region of Augsburg (KORA). Participants with available metabolomic data, covariates, and incident and prevalent MI data were included. Other COMETS cohorts could not be included in this study as they were lacking MI assessment and/or the metabolomic profile had not been performed by Metabolon Inc., the Broad Institute, or Nightingale Health. A flowchart of the study design is presented in *Figure 1*.

A brief description of the included COMETS cohorts is presented below and in *Table 1*.

- ARIC: Prospective cohort recruited from four US communities to investigate the aetiology of atherosclerosis and its clinical outcomes.[18]
- ET2DS: Longitudinal cohort of older men and women based in Lothian, Scotland, designed to investigate the role of risk factors for vascular complications of type 2 diabetes.[19]
- GDM: Prospective study that aims to provide data on demographic, biochemical, and clinical changes in type 2 diabetic patients attending real medical outpatient consultations.[20]
- HABC: Prospective cohort focused on risk factors for the decline of function in initially well-functioning older persons, particularly change in body composition with age.[21]

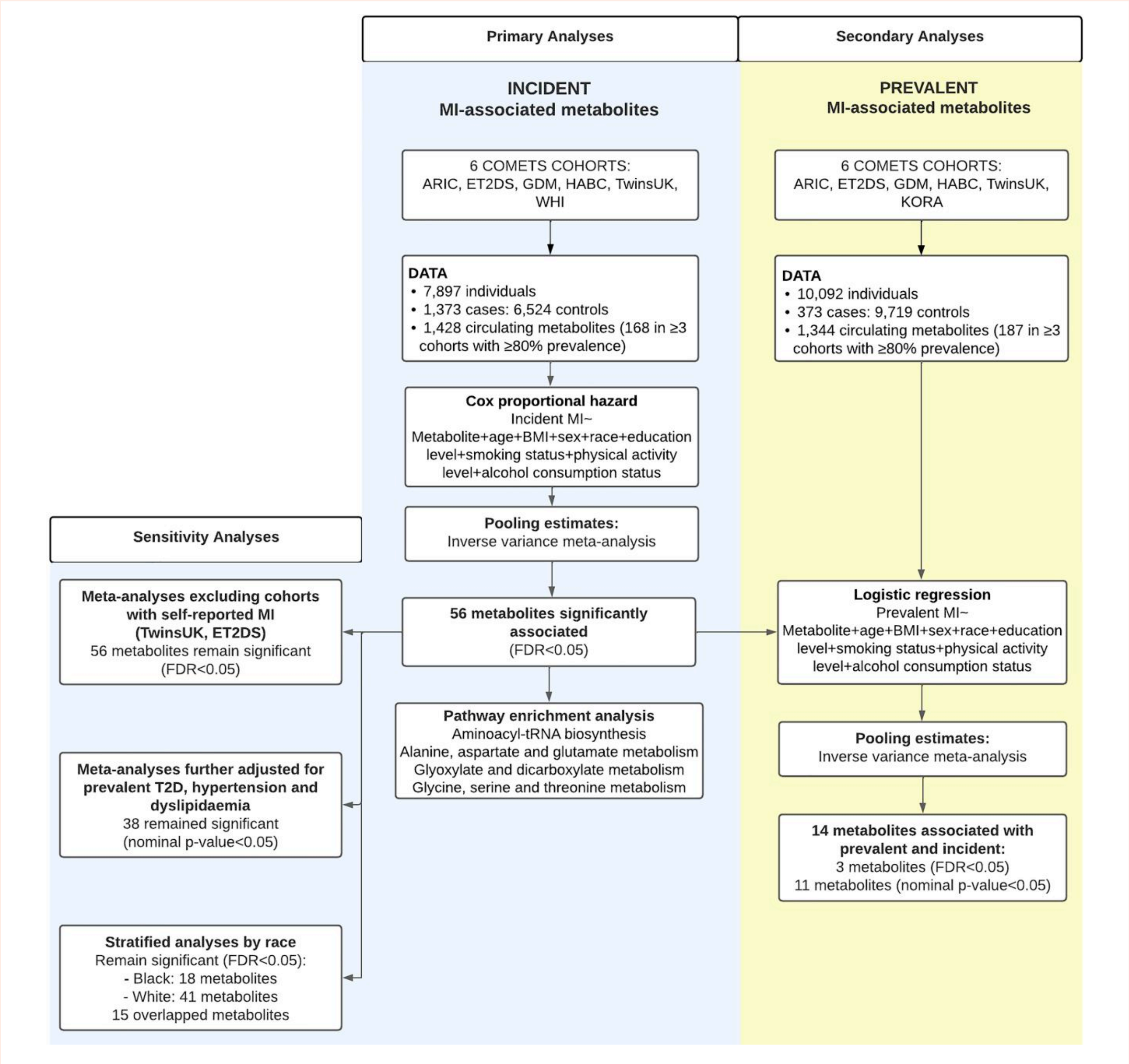

Primary Analyses
Secondary Analyses
INCIDENT
MI-associated metabolites
PREVALENT
MI-associated metabolites
6 COMETS COHORTS:
ARIC, ET2DS, GDM, HABC, TwinsUK, WHI
6 COMETS COHORTS:
ARIC, ET2DS, GDM, HABC, TwinsUK, KORA
DATA
• 7,897 individuals
• 1,373 cases: 6,524 controls
• 1,428 circulating metabolites (168 in ≥3 cohorts with ≥80% prevalence)
DATA
• 10,092 individuals
• 373 cases: 9,719 controls
• 1,344 circulating metabolites (187 in ≥3 cohorts with ≥80% prevalence)
Cox proportional hazard
Incident MI~
Metabolite+age+BMI+sex+race+education level+smoking status+physical activity level+alcohol consumption status
Sensitivity Analyses
Pooling estimates:
Inverse variance meta-analysis
Meta-analyses excluding cohorts with self-reported MI (TwinsUK, ET2DS)
56 metabolites remain significant (FDR<0.05)
56 metabolites significantly associated
(FDR<0.05)
Logistic regression
Prevalent MI~
Metabolite+age+BMI+sex+race+education level+smoking status+physical activity level+alcohol consumption status
Pathway enrichment analysis
Aminoacyl-tRNA biosynthesis
Alanine, aspartate and glutamate metabolism
Glyoxylate and dicarboxylate metabolism
Glycine, serine and threonine metabolism
Pooling estimates:
Inverse variance meta-analysis
Meta-analyses further adjusted for prevalent T2D, hypertension and dyslipidaemia
38 remained significant (nominal p-value<0.05)
14 metabolites associated with prevalent and incident:
3 metabolites (FDR<0.05)
11 metabolites (nominal p-value<0.05)
Stratified analyses by race
Remain significant (FDR<0.05):
- Black: 18 metabolites
- White: 41 metabolites
15 overlapped metabolites

Figure 1 Flowchart overview containing the available data, steps conducted, and main results. ARIC, Atherosclerosis Risk in Communities; BMI, body mass index; ET2DS, Edinburgh Type 2 Diabetes Study; FRD, false discovery rate; GDM, GenoDiabMar; HABC, Health, Aging and Body Composition; KORA, Cooperative Health Research in the Region of Augsburg; WHI, Women's Health Initiative.

- KORA: A population-based adult cohort that consists of interviews, medical and laboratory examinations, biological sample collection, and multiple omic data generation and management.[25]
- TwinsUK: The largest most clinically characterized adult twin registry in the UK, recruited as volunteers without selecting for particular diseases or traits.[23]
- WHI: A large and complex clinical investigation of strategies for the prevention and control of some of the most common causes of morbidity and mortality among postmenopausal women, including cancer, CVD, and osteoporotic fractures.[13,24]

2.2 Metabolomics

A summary of the metabolomics methodology used for each cohort is depicted in *Table 1*. Serum samples from ARIC, ET2DS, GDM, KORA, and TwinsUK and samples of ethylenediaminetetraacetic acid (EDTA) plasma from HABC, TwinsUK, and WHI were held at −80°C.[17] Serum metabolites were detected and quantified in ARIC, KORA, and TwinsUK at Metabolon Inc. using untargeted gas chromatography/liquid chromatography-mass spectrometry (GC/LC-MS) methods, in ET2DS and GDM at Nightingale Health using a nuclear magnetic resonance (NMR) method. EDTA plasma metabolites were detected and quantified in HABC and WHI at the Broad Institute using LC-MS. Metabolites were harmonized across platforms by manual curation by matching chemical structure, and the Human Metabolon Database and Kyoto Encyclopedia of Genes and Genomes (KEGG) identifiers. A total of 1442 unique named and known metabolites were measured across seven participating studies. For the primary analysis, we included 1428 metabolites, from which 168 were present in at least three studies and detected in at least 80% of participants from each cohort. For the secondary analysis, measurements of 1344 metabolites were available (from which 187 were present in at least three studies and detected in at least 80% of participants from each cohort). In this study, our focus is to explore the metabolites significantly associated with incident MI and the pathways in which are enriched. The prevalent analysis aimed to explore the overlap of metabolites associated with incident and prevalent MI.

2.3 Assessment of MI and co-variables

Specific information about how each cohort defined MI is shown in Supplementary material online, *Text S1*. In summary, MI was assessed based on one or more of the following:

- Diagnosed by a doctor (based on clinical evidence such as chest pain, electrocardiogram, and cardiac enzymes).
- Self-reported questionnaires.
- Hospital/GP records.
- Death certificates including the adjudication.

On the other hand, co-variables used to adjust the models were described identically across the cohorts. How these were defined is indicated in Supplementary material online, *Text S1*.

2.4 Statistical analysis

We conducted a two-step IPD meta-analysis. In the first step, we performed analyses separately by study cohort. Outliers defined as values four standard deviations (SDs) from the mean were excluded. To obtain normal distributions, metabolite measures were transformed to rankits by performing quantile normalization on rank-transformed raw metabolite values. Power calculation was performed using the ‘dmetar’ package implemented in R. For each metabolite included in the primary analysis, Cox proportional hazard models for incident MI were fit adjusting for age, sex, race/ethnicity, body mass index (BMI), education level, smoking status, physical activity level, and alcohol consumption status, all at the baseline visit. In the second step, we meta-analysed the results from each cohort using fixed effect inverse variance meta-analyses (using the package ‘meta’ in R) for metabolites present in three or more studies. Heterogeneity between studies and percentage of variability of between-study heterogeneity not due to the sampling error were computed using Cochran’s Q test and I^2 index, respectively.

Sensitivity analyses were conducted by (i) running Han–Eskin random effect meta-analyses[26]; (ii) further adjusting for prevalent type 2 diabetes, prevalent hypertension, and prevalent dyslipidaemia; (iii) excluding cohorts where MI was assessed through self-reported questionnaires (e.g. TwinsUK and ET2DS); and (iv) stratifying by race (White individuals and Black individuals).

Secondary analyses were conducted to assess the associations between metabolites and prevalent MI using two-step IPD meta-analysis. Logistic regression models were first run in each cohort on rankit transformed metabolite measures adjusting for the same covariates, and then a fixed effect inverse variance meta-analysis was performed.

We adjusted for multiple testing using Benjamini and Hochberg[27] false discovery rate (FDR <0.05). If not indicated otherwise, all reported *P*-values are FDR-adjusted. Analyses were undertaken and reported according to the STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) guidelines (see Supplementary material online, *Text S2*). We define that a metabolite is novel when, to our knowledge, such a metabolite has never been associated with any cardiac disease before.

2.5 Metabolomic pathway analysis

To explore the metabolomic pathways enriched for MI-related metabolites, we used MetaboAnalyst 5.0.[28] Over-representation analysis was performed using a hyper-geometric test to identify groups of compounds that are represented more than expected in each pathway by chance, and pathway topology analysis was performed based on relative betweenness centrality focusing on our entire metabolomic network. Metabolites significantly associated with incident MI (FDR < 0.05) were mapped to the *Homo sapiens* KEGG pathways. Metabolomic pathways with FDR < 0.05 were considered statistically significant.

2.6 Ethical approval

Approval was granted by the COMETS steering committee. Ethical approval for each study was obtained by the ethical research boards pertaining to each study.

3. Results

The descriptive characteristics of the study participants are shown in *Table 2*. We included 7897 individuals [average age = 66 years (SD = 7.1)] with blood metabolomic profiling (n = 1428 metabolites) and incident MI assessment from six cohorts including ARIC, ET2DS, GDM, HABC, TwinsUK, and WHI. All included participants were free from MI at baseline. There were 1373 incident MI cases across the six cohorts [average follow-up time = 9.4 years (SD = 7.1); average follow-up time per cohort is presented in *Table 2*]. For the secondary analysis, we included 373 prevalent MI cases and 9719 prevalent MI controls from the ARIC, ET2DS, GDM, HABC, TwinsUK, and KORA cohorts (descriptive characteristics are shown in *Table 2*).

3.1 Metabolites associated with incident MI

For our primary analysis including 1373 incident MI cases and 6524 controls, assuming a modest effect size of 0.12 [corresponding to hazard ratio (HR) = 1.127 or HR = 0.887], our study has over 90% power for a given metabolite adjusting for multiple testing ($P < 3.5 * 10^{-5}$). We meta-analysed 1428 metabolites, of which 168 were present in at least 80% of the participants from at least three studies. In total, 56 metabolites were significantly associated with incident MI after adjusting for multiple testing (FDR < 0.05) (*Figure 1*; see Supplementary material online, *Table S1*). Out of the 56 metabolites, 42 had a direct association, and 14 had an inverse association with incident MI (*Figure 2*). Moreover, 21 were lipids, primarily lysophospholipids (n = 5), long-chain polyunsaturated fatty acids (n = 3), phosphatidylethanolamine (n = 2), and products of the

Table 1 Location and analytical information about the cohorts comprising COMETS

Cohort Name	Name abbreviation	Continent	Platform	Analytical technology	Targeted/ untargeted	Description
Atherosclerosis Risk in Communities Study	ARIC	North America	Metabolon	GC/LC-MS	Untargeted	Prospective cohort recruited from four US communities to investigate the aetiology of atherosclerosis and its clinical outcomes[18]
Edinburgh Type 2 Diabetes Study	ET2DS	Europe	Nightingale	NMR	Targeted	Longitudinal cohort of older men and women based in Lothian, Scotland, designed to investigate the role of risk factors for vascular complications of type 2 diabetes[19]
GenoDiabMar	GDM	Europe	Nightingale	NMR	Targeted	Prospective study that aims to provide data on demographic, biochemical, and clinical changes in type 2 diabetic patients attending real medical outpatient consultations[20]
Health, Aging and Body Composition	HABC	North America	Broad Institute	LC-MS	Untargeted	Interdisciplinary cohort focused on risk factors for the decline of function in initially well-functioning older persons, particularly change in body composition with age[21]
Cooperative Health Research in the Region of Augsburg	KORA	Europe	Metabolon	GC/LC-MS	Untargeted	A population-based adult cohort and initiated as part of the World Health Organization Multinational Monitoring of Trends and Determinants in Cardiovascular Diseases (MONICA) project since 1984[22]
TwinsUK	TwinsUK	Europe	Metabolon	GC/LC-MS	Untargeted	The largest most clinically characterized adult twin registry in the UK, recruited as volunteers without selecting for particular diseases or traits[23,25]
Women's Health Initiative	WHI	North America	Broad Institute	LC-MS	Untargeted	A large and complex clinical investigation of strategies for the prevention of some of the most common causes of morbidity and mortality among postmenopausal women, including cancer, cardiovascular disease, and osteoporotic fractures.[13,24]

primary bile acid metabolism ($n = 2$), and 17 were amino acids including products of tryptophan metabolism ($n = 4$), glycine, serine, and threonine ($n = 4$) and glutamate metabolism ($n = 2$). There were also 4 nucleotides, 4 carbohydrates, 3 xenobiotics, 3 energy-producing metabolites, 3 co-factors/vitamins, and 1 peptide (*Figure* 2). Out of the 21 associated lipids, 3-methyladipate and 1-palmitoyl-2-linoleoyl-glycerol (16:0/18:2) were associated with a higher risk with HR estimates ranging from 1.28 [95% confidence interval (CI) = 1.13–1.44, $P < 0.001$] to 1.21 (95% CI = 1.08–1.35, $P = 4.29 \times 10^{-3}$), respectively (*Figure* 2). Among the amino acids, 4-hydroxyphenylacetate and cystathionine had the largest increase in risk presenting HR estimates of 1.24 (95% CI = 1.11–1.38, $P = 1.11 \times 10^{-3}$) and 1.2 (95% CI = 1.07–1.35, $P = 7.58 \times 10^{-3}$), respectively (*Figure* 2). Likewise, overall, the highest increase of risk was observed for the carbohydrates mannitol/sorbitol [HR (95% CI) = 1.40 (1.26–1.56), $P < 0.001$] and glucuronate [HR (95% CI) = 1.37 (1.26–1.5), $P < 0.001$], whereas the metabolites associated with reduced risk of incident MI included the amino acid glutamine [HR (95% CI) = 0.74 (0.67–0.82), $P < 0.001$], the nucleotide uridine [HR (95% CI) = 0.82 (0.76–0.88), $P < 0.001$], and the co-factor 1-methylnicotinamide [HR (95% CI) = 0.84 (0.76–0.94), $P = 7.37 \times 10^{-3}$], among others (*Figure* 2). The list of metabolites previously associated with any cardiac diseases and the super- and sub-pathways for incident MI-associated metabolites are presented in Supplementary material online, *Table S2*.

Of note, the obtained heterogeneity estimated for the associated metabolites was only significant (Q $P < 0.05$) for seven metabolites with also I^2 values indicating considerable variability of between-study heterogeneity ($I^2 > 70\%$).[29] However, most identified metabolites presented not relevant or moderate between-study heterogeneity ($I^2 < 60\%$).[29]

3.2 Sensitivity analyses

Results were consistent when running Han–Eskin random effect inverse variance meta-analyses[26] (see Supplementary material online, *Table S3*). Results were also consistent when the meta-analysis was performed excluding cohorts in which MI was assessed by self-reported questionnaires (i.e. TwinsUK and ET2DS) (see Supplementary material online, *Table S4*). When we further adjusted for prevalent type 2 diabetes, hypertension, and dyslipidaemia, 38 metabolites remained associated (see Supplementary material online, *Table S5*). Interestingly, the metabolites that did not reach the significance level after adjustment for co-morbidities have been previously linked with those commodities (see Supplementary material online, *Table S2*). Finally, we investigated whether there were demographic differences in the associations between the identified metabolites and MI by conducting a meta-analysis stratified by race. Out of the 56 metabolites, 41 remained significantly associated in White individuals, whereas 18 were significantly associated in Black individuals, with 3 of them, namely, dimethylglycine, glycine, and glycoursodeoxycholate, presenting a significant association only in individuals with an African ancestry (see Supplementary material online, *Table S6*).

*Typo: "comorbidities" instead of "commodities".

Table 2 Descriptive characteristics at baseline of the participants from the COMETS cohorts containing incident and/or prevalent myocardial infarction data

Cohort (Metabolite number)	MI type	Subsets	Sample Size, *N*	Women, %	Baseline Age, years	Follow-up Age, years	BMI, kg/m^2	Race, %	Follow-up Time, years
ARIC (*n* = 311)	Incident	All participants	3776	61	53 (5.7)	76 (8.7)	28.8 (5.9)	38% White, 62% Black	22.8 (8.4)
		MI cases	442	55	55 (5.7)	70 (9)	29.3 (5.4)	41% White, 59% Black	15.5 (8)
		Controls	3334	62	53 (5.7)	77 (8.3)	28.7 (5.9)	38% White, 62% Black	23.8 (7.8)
	Prevalent	All participants	3395	62	53 (5.8)	–	28.7 (5.9)	38% White, 62% Black	–
		MI cases	54	33	57 (5.7)	–	29.1 (5.1)	56% White, 44% Black	–
		Controls	3341	62	53 (5.7)	–	28.7 (5.9)	38% White, 62% Black	–
ET2DS (*n* = 208)	Incident	All participants	909	53	68 (4.2)	77 (4.6)	31.4 (5.8)	98% White, 2% non-White[b]	9.5 (2.8)
		MI cases	66	47	69 (3.8)	75 (4.9)	31.2 (5.4)	98% White, 2% non-White[b]	5.9 (3.1)
		Controls	843	53	68 (4.2)	77 (4.6)	31.4 (5.8)	98% White, 2% non-White[b]	9.8 (2.6)
	Prevalent	All participants	992	49	68 (4.2)	–	31.4 (5.7)	98% White, 2% non-White[b]	–
		MI Cases	147	22	69 (4.1)	–	31.3 (5.2)	98% White, 2% non-White[b]	–
		Controls	845	53	68 (4.2)	–	31.5 (5.8)	98% White, 2% non-White[b]	–
GDM (*n* = 210)	Incident	All participants	477	41	69 (9.3)	73 (9.1)	30.3 (5.2)	100% White	4.4 (1.3)
		MI cases	42	33	70 (8.4)	73 (8.2)	30.1 (5.3)	100% White	2.3 (1.6)
		Controls	435	42	69 (9.4)	74 (9.2)	30.4 (5.2)	100% White	4.5 (1.2)
	Prevalent	All participants	468	41	69 (9.4)	–	30.4 (5.2)	100% White	–
		MI cases	33	33	71 (10.1)	–	31.5 (4.8)	100% White	–
		Controls	435	42	69 (9.4)	–	30.4 (5.2)	100% White	–
HABC (*n* = 350)	Incident	All participants	236	0	75 (2.8)	83 (4.7)	27.0 (4.5)	100% Black	10.6 (5)
		MI cases	25	0	75 (2.8)	83 (4.7)	27.0 (4.6)	100% Black	7.6 (3.7)
		Controls	211	0	75 (2.9)	81 (4.4)	26.8 (3.2)	100% Black	10.6 (5.1)
	Prevalent	All participants	1764	0	75 (2.8)	–	27.0 (4.5)	100% Black	–
		MI cases	63	0	75 (2.8)	–	27.6 (4.6)	100% Black	–
		Controls	1701	0	74 (2.8)	–	26.8 (4.4)	100% Black	–
TwinsUK (*n* = 591)	Incident	All participants	911	97	65 (8)	70 (7.7)	26.1 (4.8)	100% White	3.9 (2.9)
		MI cases	5	80	74 (5.2)	77 (5.2)	31.7 (9.6)	100% White	2.6 (0.1)
		Controls	906	97	66 (8)	70 (7.7)	26.1 (4.7)	100% White	3.9 (2.9)
	Prevalent	All participants	1708	97	65 (8.6)	–	26.3 (4.8)	100% White	–
		MI cases	13	77	71 (5.8)	–	28.5 (5.8)	100% White	–
		Controls	1695	97	65 (8.6)	–	26.3 (4.8)	100% White	–
WHI (*n* = 414)	Incident	All Participants	1588	100	67 (6.9)	72 (7.5)	28.4 (6.1)	77% White, 23% non-White[a]	5.1 (3.3)
		MI cases	793	100	67 (7.0)	72 (7.5)	29.0 (6.3)	77% White, 23% non-White[a]	5.1 (3.3)
		Controls	795	100	67 (6.9)	72 (7.4)	27.9 (5.9)	77% White, 23% non-White[a]	5.1 (3.3)
KORA (*n* = 353)	Prevalent	All participants	1765	52	61 (8.8)	–	28.2 (4.8)	100% White	–
		MI cases	63	22	67 (6.6)	–	30.7 (5.1)	100% White	–
		Controls	1702	53	61 (8.8)	–	28.1 (4.8)	100% White	–
Total (unique: *n* = 1428)	Incident	All participants	7897	59	66 (7.1)	75 (4.6)	28.7 (2)	69% White, 29% Black, 2% Others	9.4 (7.1)

Continued

Table 2 Continued

Cohort (Metabolite number)	MI type	Subsets	Sample Size, *N*	Women, %	Baseline Age, years	Follow-up Age, years	BMI, kg/m^2	Race, %	Follow-up Time, years
		MI cases	1373	53	68 (7.3)	75 (4.6)	29.7 (1.7)	69% White, 29% Black, 2% Others	6.5 (4.8)
		Controls	6524	59	66 (7.2)	75 (4)	28.6 (2.1)	69% White, 29% Black, 2% Others	9.6 (7.5)
	Prevalent	All participants	10 092	50	65 (7.4)	–	28.7 (2)	73% White, 27% Black[c]	–
		MI cases	373	31	68 (6)	–	29.8 (1.6)	76% White, 24% Black[c]	–
		Controls	9719	51	65 (7.2)	–	28.6 (2)	73% White, 27% Black[c]	–

[a] In the WHI, non-White included: 14% Black or African-American, 3% Hispanic/Latino, 2% Asian or Pacific Islander, and 4% others.

[b] In the ET2DS, non-White included for prevalent: 1.1% Asian, 0.2% Black, 0.1% White and Black Caribbean, and 0.1% White and Asian. Specifically, 98.4% are White. Non-White included for incident: 1.1% Asian, 0.2% Black, 0.1% White and Black Caribbean, and 0.1% White and Asian. Specifically, 98.5% are White.

[c] It presents <1% of other ethnicities (non-White and non-Black).

3.3 Metabolites associated with prevalent MI

As a secondary analysis, we further investigated whether the 56 metabolites associated with incident MI were also correlated with prevalent MI (*Figure 1*). On meta-analyses, 11 metabolites, including tryptophan, malate, allantoin, and 1-linoleoyl-GPC (18:2), were nominally associated with prevalent MI with concordant directional effects in both incident and prevalent analyses, and three [xenobiotic 2-hydroxyhippurate (salicylurate), lactate, and glucoronate] were associated after correcting for multiple testing [2-hydroxyhippurate: odds ratio (OR) (95% CI) = 1.9 (1.5–2.42), $P < 0.001$; lactate: OR (95% CI) = 1.36 (1.2–1.54), $P < 0.001$; and glucuronate: OR (95% CI) = 1.51 (1.19–1.93), $P = 0.03$] (see Supplementary material online, *Table S7*).

3.4 Pathways behind the metabolites associated with incident MI

To identify the potential biological pathways involved in incident MI, we assessed the enriched pathways for the 56 metabolites (*Figure 1*). These metabolites included 41 pathways, 12 of which had a significant nominal *P*-value, including the citrate cycle [trichloroacetic acid (TCA) cycle] (nominal $P =$ 0.016) and the primary bile acid biosynthesis (nominal $P = 0.024$) (see Supplementary material online, *Table S8*). Of these 12, 4 pathways were significantly enriched (FDR < 0.05), namely, aminoacyl-tRNA biosynthesis ($P <$ 0.001), alanine, aspartate, and glutamate metabolism ($P = 0.018$), glyoxylate and dicarboxylate metabolism ($P = 0.02$), and glycine, serine, and threonine metabolism ($P = 0.02$) (*Figure 3*). Specifically, 9 amino acids were involved in the 1st pathway, 3 amino acids and the energy-producing metabolites fumarate and succinate in the 2nd pathway, 4 amino acids and the energy-producing metabolite malate in the 3rd pathway, and 5 amino acids in the 4th pathway (see Supplementary material online, *Table S8*). There were 14 unique metabolites involved in these four pathways. Glycine and serine are intermediates/products of aminoacyl-tRNA biosynthesis; glycine, serine, and threonine metabolism; and glyoxylate and dicarboxylate metabolism, whereas glutamine and glutamate are present in all the pathways but the glycine, serine, and threonine metabolism.

4. Discussion

In this comprehensive study investigating biomarkers of incident MI by leveraging IPD from six intercontinental cohorts with 7897 participants from diverse race/ethnic backgrounds, we identified 56 metabolites, mainly lipids and amino acids, significantly associated with incident MI. We report 10 novel biomarkers of incident MI, including 8 lipids (3 lysophospholipids, 1 phosphatidylethanolamine, 1 diacylglycerol, 1 intermediate of the primary bile acid metabolism, 1 dicarboxylate fatty acid, and 1 glycerolipid), 1 xenobiotic (involved in xanthine metabolism), and 1 nucleotide (involved in purine metabolism). Of these, 6 have underlying mechanisms of action leading to MI onset which are independent of hypertension, type 2 diabetes, and dyslipidaemia, known as risk factors for MI.[14,30–32] We also confirm previous associations, including the protective association of nonessential amino acids (e.g. glutamine, glycine, and serine),[7,33] and the detrimental effect of the well-known branched-chain amino acid isoleucine on cardiac diseases,[34] thus demonstrating the robustness of our approach. Our stratified analyses revealed that dimethylglycine, glycine, and glycoursodeoxycholate were associated with incident MI only in Black individuals, highlighting the role of ethnicity in the aetiology of MI. We also show that the metabolites that might lead to the MI onset differ from the metabolites deregulated once the disease is well established, highlighting the importance of survival analyses to identify preventive biomarkers. Finally, we report the pathways in which the identified amino acids are enriched, shedding light on the mechanisms by which these metabolites may be implicated in MI onset. Of note, most of the identified metabolites are lipids, and enrichment of lipid metabolism pathways was observed, but these did not attain statistical significance due to the involvement of many metabolites and thus the need for a large overlap with the lipid-associated MI to be considered significant. This complexity underscores the intricate nature of lipid metabolism pathways, and the multiple roles lipids play in the onset of MI.

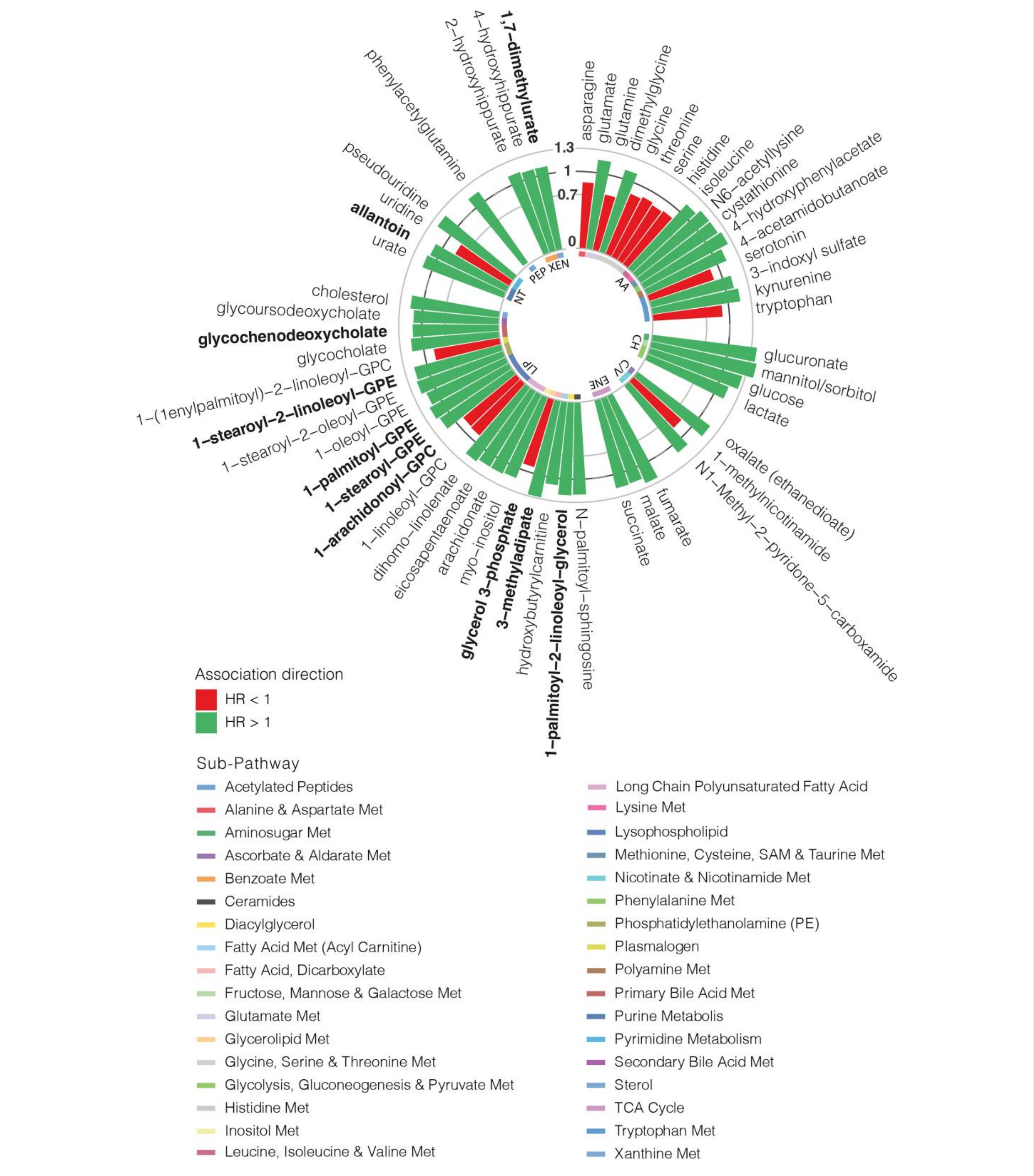


Figure 2 Metabolites significantly associated with incident myocardial infarction. The bar height represents the hazard ratio (HR) value. Novel metabolites are highlighted in bold. Each metabolite super-pathway and sub-pathway is also indicated. AA, amino acid; CH, carbohydrate; C/V, co-factors/vitamins; ENE, energy; LIP, lipid; Met, metabolite; NT, nucleotide; XEN, xenobiotic.

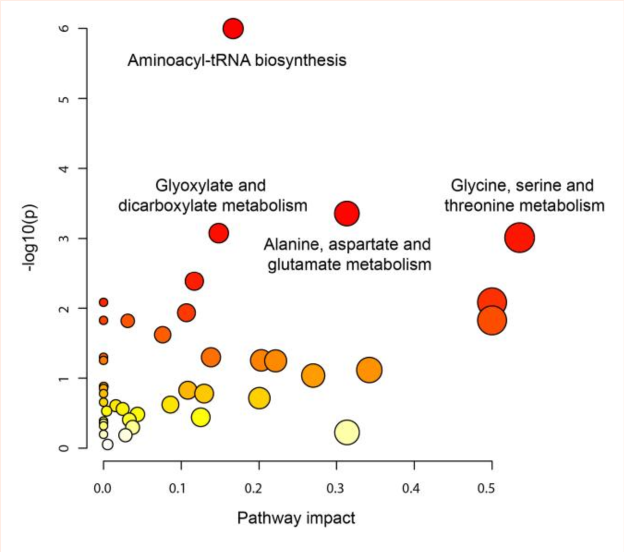


Figure 3 Enrichment pathway analysis results indicating the significant pathways (FDR < 0.05) among the identified metabolites associated with incident myocardial infarction.

4.1 Lysophospholipids

Among the lipids, lysophospholipids represent the largest subgroup found to be associated with incident MI. Specifically, we identified 5 metabolites belonging to this sub-pathway, with 3 of them, namely, 1-oleoyl-GPE (18:1), 1-palmitoyl-GPE (16:0), and 1-stearoyl-GPE (18:0), associated with an increased risk of MI and two of them, 1-linoleoyl-GPC (18:2) and 1-arachidonoyl-GPC (20:4), associated with decreased risk of MI. Of these, 1-palmitoyl-GPE (16:0), 1-arachidonoyl-GPC (20:4), and 1-stearoyl-GPE (18:0) are novel biomarkers of MI. Lysophospholipids are a group of bioactive molecules with diverse biological roles, including activation of specific G-protein-coupled receptors, and have been associated with atherosclerosis, coronary heart disease, and hypertension.[35] Nonetheless, their effects on CVD are controversial as both beneficial and detrimental effects have been reported. For instance, they might possess cardioprotective effects, but, also, they might stimulate platelet aggression, enhancing ischaemia in MI.[35] This fact along with the opposing results found between these metabolites and MI might indicate that lysophospholipids' function might vary depending on their subclasses.

4.2 Intermediates of bile acid metabolism

Here, we report for the first time that incident MI cases have higher circulating levels of the secondary bile acid glycochenodeoxycholate compared to controls. Bile acids can act as signalling molecules involved in inflammatory processes and host metabolism.[36] Several CVD metabolomics studies have highlighted the negative role of bile acids on CVD morbidity/mortality.[37,38] Glycochenodeoxycholate is a bile acid-lycine conjugate produced by the gut microbiota.[39] Studies have reported glycochenodeoxycholate is toxic and can induce hepatocyte apoptosis, which might lead to liver disease.[40] Likewise, liver and cardiac diseases co-exist through complex cardio hepatic interactions.[41] Our results may suggest that high levels of this bile acid can have detrimental effects on MI by causing alterations in the liver, and the gut microbiota might be targeted to modulate its levels.

4.3 Nucleotide metabolism intermediates

We are the first to report the association between allantoin and MI. Allantoin is involved in purine metabolism and is formed from the oxidation of urate by various reactive oxygen species.[42] Allantoin has been reported as a potential marker of oxidative stress in humans,[42] possibly explaining the observed positive association with MI. Moreover, we show the associations of pseudouridine and uridine, intermediates of the pyrimidine metabolism, and also urate, involved in the purine metabolism, with incident MI. This confirms previous findings and points out the important role of the nucleotide metabolism intermediates in cardiovascular risk.[38] For instance, hyperuricaemia has been shown to be strongly positively associated with carotid and coronary vascular disease and stroke.[43]

4.4 Co-factors involved in the nicotinate and nicotinamide metabolism

We identified 3 co-factors associated with incident MI, from which 1-methylnicotinamide and N1-methyl-2-pyridone-5-carboxamide were intermediates of the nicotinate and nicotinamide metabolism. 1-Methylnicotinamide presented an important protective effect in MI, which is concordant with their shown antithrombotic action in rats.[44] On the contrary, N1-methyl-2-pyridone-5-carboxamide was negatively associated with MI, and to our knowledge, no studies have previously reported such an association with incident MI. Nonetheless, Surendran and colleagues[45] stated changes in its plasma levels during myocardial ischaemia-reperfusion injury. N1-Methyl-2-pyridone-5-carboxamide has been reported as a uremic toxin.[46] These are organic compounds that accumulate in the bloodstream, as they cannot be eliminated from the body, reaching diverse organs, including the heart,[47] and they are a risk factor for the progression of chronic kidney disease. Likewise, patients with chronic kidney disease have an increased risk for CVD, for instance, these molecules can lead to vascular damage by enhancing the expression of cytokines and pro-inflammatory molecules.[47]

4.5 Amino acids

Pathway enrichment analysis revealed that 11 incident MI-associated amino acids are enriched in pathways previously associated with CVD. Firstly, the aminoacyl-tRNA biosynthesis pathway has been reported to be closely related to angiogenesis and cardiomyopathy.[48] Likewise, the glyoxylate and dicarboxylate metabolism is another commonly disturbed pathway found in different CVD.[49] Eventually, the metabolism of glycine, serine, and threonine has been linked with benefits in atherosclerosis,[50] being concordant with the found negative associations of glycine, serine, and threonine with incident MI. Of note, these pathways share most of the included metabolites and are characterized for being sensitive to the amino acids availability,[48] suggesting that deregulation of the matched amino acids might lead to different cardiovascular complications, including MI, and emphasizes the importance of a balanced amino acid profile.

Our study has some limitations. Firstly, the number of healthy participants is 5.7-fold larger than the number of incident MI cases, although we have been able to identify 56 metabolites whose levels significantly differ between MI cases and controls. Secondly, the clinical definition of MI varies in each cohort depending on the protocol for data collection. This may introduce a procedural bias. However, when we ran a sensitivity analysis by excluding cohorts where MI was assessed by self-reported questionnaires, the results remained consistent. Thirdly, metabolomics profiling was conducted using different metabolomic platforms, raising some caveats: (i) a different, somehow overlapping, set of metabolites was measured by each platform, and we are only including metabolites present in at least three cohorts; (ii) we quantile normalized metabolites to meta-analyse results across studies using different metabolomic platforms. However, ranks do not have practical significance and could be influenced by the sample size; (iii) metabolite sampling and detection times could not be unified as each cohort applies used a different metabolomics methodology. Fourth, though metabolite concentrations might be influenced by medications (e.g. statins),[51] we were unable to adjust for drug usage as the data were not available across the studies. Statins are the main therapy for the worldwide prevention of CVD, including MI.[52,53] They inhibit the rate-limiting step in cholesterol synthesis, thereby lowering serum cholesterol levels and reducing MI risk.[54] Statins can also reduce MI risk via cholesterol-independent mechanisms, for instance, by inhibiting the isoprenoid synthesis.[55] Hence, statin usage and adherence could be confounding our results, and this should be

*Typo: "information bias" instead of "procedural bias".

addressed in future studies. Fifth, our study sample was predominantly White, and some MI-associated metabolites might have not reached the significance level in Black individuals due to lack of power. Future studies should further investigate race–metabolite interactions[56] to better understand the role of race in the metabolite–MI association. Finally, it is important to note that these results do not necessarily imply causality.

Notwithstanding the above limitations, our study benefits from a two-step meta-analysis using IPD, which has been recognized as a 'gold standard' to evidence synthesis,[57] and a high number of participants, which increases the power of our statistical analyses and minimizes the chances of obtaining false positives. Also, sensitivity analyses were run stratifying by race, allowing us to investigate the influence of demographic diversity in the identified associations. Furthermore, measurements of a wide range of metabolites, belonging to different pathways and sub-pathways, were available for each cohort allowing us to obtain a wide picture of the role played by metabolomics in MI. Different platforms were used for the metabolite measurements, reducing the inclusion of measurement errors or misidentified metabolites given by a certain platform. Moreover, despite using distinct platforms and manners to define MI, the significance of the identified metabolites was concordant across cohorts. Finally, the prospective nature of the current study permitted us to investigate how distinct metabolomic profiles are associated with incident MI.

In conclusion, these findings shed light on novel metabolic preventive biomarkers of MI and the involved pathways and might help to identify high-risk individuals before the disease onset and pave the way towards the development of novel preventative strategies. Nonetheless, more research needs to be conducted to confirm the identified metabolites as biomarkers and to fully understand underlying the mechanisms of action.

Supplementary material

Supplementary material is available at *Cardiovascular Research* online.

Acknowledgements

We thank all the participants of the contributing cohorts for their time and effort and supporting our research. We also would like to thank Tamara Harris for her contributions.

Conflict of interest: R.A.M. is a consultant for Pharmavite for work unrelated to this study. All other authors declare no competing financial interests.

Funding

This research was funded in whole, or in part, by the Wellcome Trust (WT212904/Z/18/Z) and by the UKRI Medical Research Council (MRC)/ British Heart Foundation Ancestry and Biological Informative Markers for Stratification of Hypertension (AIM-HY; MR/M016560/1). For the purpose of open access, the authors have applied a CC BY public copyright to any author-accepted manuscript version arising from this submission. TwinsUK receives funding from the Wellcome Trust, the European Commission H2020 grants SYSCID (contract #733100), the National Institute for Health Research (NIHR) Clinical Research Facility and the Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London, the Chronic Disease Research Foundation, the UKRI Medical Research Council (MRC)/British Heart Foundation Ancestry and Biological Informative Markers for Stratification of Hypertension (AIM-HY; MR/ M016560/1), and Zoe Limited. C.M. and A.N. are funded by the Chronic Disease Research Foundation. C.M. is also funded by the MRC AIM-HY grant. The Atherosclerosis Risk in Communities (ARIC) study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services, under Contract nos. (75N92022D00001, 75N92022D00002, 75N92022D00003, 75N92022D00004, and 75N92022D00005). The authors thank the staff and participants of the ARIC study for their important contributions). B.Y. was in part supported by R01HL168683. Metabolomics measurements were sponsored by the National Human Genome Research Institute (3U01HG004402-02S1). The ET2DS was funded by the Medical Research Council (UK) (Project Grant G0500877) and the Chief Scientist Office of Scotland (Program Support Grand CZQ/1/38). C.B. was funded by the grant FIS-FEDER-ISCIII PI16/00620 (Ext 2021) and the Strategic Plan for Research and Innovation in Health, CatSalut, PERIS STL008 (2019–2021), and RICORS RD21/0005, to develop clinical and epidemiological studies mainly focused on diabetes and its associations with new biomarkers. HABC was supported in part by the Intramural Research Program of the National Institutes of Health, National Institute on Aging (NIA); contracts: N01-AG-6-2101, N01-AG-6-2103, and N01-AG-6-2106; NIA grant: R01-AG028050, and NINR grant R01-NR012459; and the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR000454. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Dr Murphy is supported by the Michael Smith Foundation for Health Research (grant #17644). The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through 75N92021D00001, 75N92021D00002, 75N92021D00003, 75N92021D00004, and 75N92021D00005. The authors thank the WHI investigators and staff for their dedication and the study participants for making the program possible. A full listing of WHI investigators can be found at https://www-whi-org.s3.us-west-2.amazonaws.com/wp-content/uploads/WHI-Investigator-Long-List.pdf.

Data availability

The phenotypic data used from the Atherosclerosis Risk in Communities (ARIC) Cohort are assessed via dbGaP (Study Accession: phs000280.v8.p2) or BioLINCC (https://biolincc.nhlbi.nih.gov/studies/aric/). The ARIC metabolomic data can be requested through the study's Data Coordinating Center upon an approved manuscript proposal and Data and Materials Distribution Agreement (DMDA). ET2DS can only share with bonafide researchers under managed access and when local resources are available for historical data management. GDM data available upon reasonable request from the author CB due to patient's privacy/ethical restrictions. HABC can only share with approved investigators under managed access. The KORA FF4 datasets are available upon application through the KORA-PASST (Project application self-service tool, https://www.helmholtz-munich.de/epi/research/cohorts/kora-cohort/data-use-and-access-viakorapasst/index.html.) The TwinsUK data are held by the Department of Twin Research at King's College London. The data can be released to bona fide researchers using our normal procedures overseen by the Wellcome Trust and its guidelines as part of our core funding (https://twinsuk.ac.uk/resources-for-researchers/access-our-data/). WHI data is publicly available in DbGAP.

References

1. WHO. 2021. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
2. Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, Chamberlain AM, Cheng S, Delling FN, Elkind MSV, Evenson KR, Ferguson JF, Gupta DK, Khan SS, Kissela BM, Knutson KL, Lee CD, Lewis TT, Liu J, Loop MS, Lutsey PL, Ma J, Mackey J, Martin SS, Matchar DB, Mussolino ME, Navaneethan SD, Perak AM, Roth GA, Samad Z, Satou GM, Schroeder EB, Shah SH, Shay CM, Stokes A, VanWagner LB, Wang N-Y, Tsao CW. Heart disease and stroke statistics—2021 update: a report from the American Heart Association. *Circulation* 2021;**143**:e254–e743.
3. British Heart Foundation. 2021. https://www.bhf.org.uk/what-we-do/annual-report-2021
4. Yusuf S, Hawken S, Ôunpuu S, Dans T, Avezum A, Lanas F, McQueen M, Budaj A, Pais P, Varigos J, Lisheng L. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet* 2004;**364**:937–952.
5. Koeth RA, Wang Z, Levison BS, Buffa JA, Org E, Sheehy BT, Britt EB, Fu X, Wu Y, Li L, Smith JD, DiDonato JA, Chen J, Li H, Wu GD, Lewis JD, Warrier M, Brown JM, Krauss RM, Tang WHW, Bushman FD, Lusis AJ, Hazen SL. Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat Med* 2013;**19**:576–585.
6. Tang WW, Wang Z, Levison BS, Koeth RA, Britt EB, Fu X, Wu Y, Hazen SL. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N Engl J Med* 2013;**368**:1575–1584.

7. Ding Y, Svingen GF, Pedersen ER, Gregory JF, Ueland PM, Tell GS, NygAard OK. Plasma glycine and risk of acute myocardial infarction in patients with suspected stable angina pectoris. *J Am Heart Assoc* 2015;**5**:e002621.

8. McKirnan MD, Ichikawa Y, Zhang Z, Zemljic-Harpf AE, Fan S, Barupal DK, Patel HH, Hammond HK, Roth DM. Metabolomic analysis of serum and myocardium in compensated heart failure after myocardial infarction. *Life Sci* 2019;**221**:212–223.

9. Hunter WG, Kelly JP, McGarrah RW, Kraus WE, Shah SH. Metabolic dysfunction in heart failure: diagnostic, prognostic, and pathophysiologic insights from metabolomic profiling. *Curr Heart Fail Rep* 2016;**13**:119–131.

10. Shah SH, Hunter WG. *Realizing the potential of metabolomics in heart failure: signposts on the path to clinical utility*. Washington. DC: American College of Cardiology Foundation; 2017. p833–836.

11. Shah SH, Bain JR, Muehlbauer MJ, Stevens RD, Crosslin DR, Haynes C, Dungan J, Newby LK, Hauser ER, Ginsburg GS, Newgard CB, Kraus WE. Association of a peripheral blood metabolic profile with coronary artery disease and risk of subsequent cardiovascular events. *Circ Cardiovasc Genet* 2010;**3**:207–214.

12. Cheng S, Shah SH, Corwin EJ, Fiehn O, Fitzgerald RL, Gerszten RE, Illig T, Rhee EP, Srinivas PR, Wang TJ, Jain M. Potential impact and study considerations of metabolomics in cardiovascular health and disease: a scientific statement from the American Heart Association. *Circ Cardiovasc Genet* 2017;**10**:e000032.

13. Paynter NP, Balasubramanian R, Giulianini F, Wang DD, Tinker LF, Gopal S, Deik AA, Bullock K, Pierce KA, Scott J, Martínez-González MA, Estruch R, Manson JE, Cook NR, Albert CM, Clish CB, Rexrode KM. Metabolic predictors of incident coronary heart disease in women. *Circulation* 2018;**137**:841–853.

14. Nogal A, Louca P, Tran TQB, Bowyer RC, Christofidou P, Steves CJ, Berry SE, Wong K, Wolf J, Franks PW, Mangino M, Spector TD, Valdes AM, Padmanabhan S, Menni C. Incremental value of a panel of serum metabolites for predicting risk of atherosclerotic cardiovascular disease. *J Am Heart Assoc* 2022;**11**:e024590.

15. Schmidt DR, Patel R, Kirsch DG, Lewis CA, Vander Heiden MG, Locasale JW. Metabolomics in cancer research and emerging applications in clinical oncology. *CA Cancer J Clin* 2021;**71**: 333–358.

16. Cicalini I, Rossi C, Pieragostino D, Agnifili L, Mastropasqua L, di Ioia M, De Luca G, Onofrj M, Federici L, Del Boccio P. Integrated lipidomics and metabolomics analysis of tears in multiple sclerosis: an insight into diagnostic potential of lacrimal fluid. *Int J Mol Sci* 2019; **20**:1265.

17. Yu B, Zanetti KA, Temprosa M, Albanes D, Appel N, Barrera CB, Ben-Shlomo Y, Boerwinkle E, Casas JP, Clish C, Dale C, Dehghan A, Derkach A, Eliassen AH, Elliott P, Fahy E, Gieger C, Gunter MJ, Harada S, Harris T, Herr DR, Herrington D, Hirschhorn JN, Hoover E, Hsing AW, Johansson M, Kelly RS, Khoo CM, Kivimäki M, Kristal BS, Langenberg C, Lasky-Su J, Lawlor DA, Lotta LA, Mangino M, Le Marchand L, Mathé E, Matthews CE, Menni C, Mucci LA, Murphy R, Oresic M, Orwoll E, Ose J, Pereira AC, Playdon MC, Poston L, Price J, Qi Q, Rexrode K, Risch A, Sampson J, Seow WJ, Sesso HD, Shah SH, Shu X-O, Smith GCS, Sovio U, Stevens VL, Stolzenberg-Solomon R, Takebayashi T, Tillin T, Travis R, Tzoulaki I, Ulrich CM, Vasan RS, Verma M, Wang Y, Wareham NJ, Wong A, Younes N, Zhao H, Zheng W, Moore SC. The Consortium of Metabolomics Studies (COMETS): metabolomics in 47 prospective cohort studies. *Am J Epidemiol* 2019;**188**:991–1012.

18. Wright JD, Folsom AR, Coresh J, Sharrett AR, Couper D, Wagenknecht LE, Mosley TH Jr, Ballantyne CM, Boerwinkle EA, Rosamond WD, Heiss G. The ARIC (Atherosclerosis Risk in Communities) study: JACC focus seminar 3/8. *J Am Coll Cardiol* 2021;**77**:2939–2959.

19. Price JF, Reynolds RM, Mitchell RJ, Williamson RM, Fowkes FGR, Deary IJ, Lee AJ, Frier BM, Hayes PC, Strachan MW. The Edinburgh Type 2 Diabetes Study: study protocol. *BMC Endocr Disord* 2008;**8**:1–10.

20. Sierra A, Otero S, Rodríguez E, Faura A, Vera M, Riera M, Palau V, Durán X, Costa-Garrido A, Sans L, Márquez E, Poposki V, Franch-Nadal J, Mundet X, Oliveras A, Crespo M, Pascual J, Barrios C. The GenoDiabMar registry: A collaborative research platform of type 2 diabetes patients. *J Clin Med* 2022;**11**:1431.

21. Santanasto AJ, Goodpaster BH, Kritchevsky SB, Miljkovic I, Satterfield S, Schwartz AV, Cummings SR, Boudreau RM, Harris TB, Newman AB. Body composition remodeling and mortality: the health aging and body composition study. *J Gerontol Ser A Biomed Sci Med Sci* 2017;**72**:513–519.

22. Suhre K, Shin S-Y, Petersen A-K, Mohney RP, Meredith D, Wägele B, Altmaier E, Deloukas P, Erdmann J, Grundberg E, Hammond CJ, de Angelis MH, Kastenmüller G, Köttgen A, Kronenberg F, Mangino M, Meisinger C, Meitinger T, Mewes H-W, Milburn MV, Prehn C, Raffler J, Ried JS, Römisch-Margl W, Samani NJ, Small KS, -Erich Wichmann H., Zhai G, Illig T, Spector TD, Adamski J, Soranzo N, Gieger C. Human metabolic individuality in biomedical and pharmaceutical research. *Nature* 2011;**477**:54–60.

23. Verdi S, Abbasian G, Bowyer RC, Lachance G, Yarand D, Christofidou P, Mangino M, Menni C, Bell JT, Falchi M, Small KS, Williams FMK, Hammond CJ, Hart DJ, Spector TD, Steves CJ. TwinsUK: the UK adult twin registry update. *Twin Res Hum Genet* 2019;**22**:523–529.

24. WsHIS G. Design of the Women's Health Initiative clinical trial and observation study. *Control Clin Trials* 1998;**19**:61–109.

25. Han S, Huang J, Foppiano F, Prehn C, Adamski J, Suhre K, Li Y, Matullo G, Schliess F, Gieger C, Peters A, Wang-Sattler R. TIGER: technical variation elimination for metabolomics data using ensemble learning architecture. *Brief Bioinformatics* 2022;**23**:bbab535.

26. Han B, Eskin E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am J Hum Genet* 2011;**88**:586–598.

27. Thissen D, Steinberg L, Kuang D. Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *J Educ Behav Stat* 2002;**27**:77–83.

28. Pang Z, Chong J, Zhou G, de Lima Morais DA, Chang L, Barrette M, Gauthier C, Jacques P-É, Li S, Xia J. Metaboanalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res* 2021;**49**:W388–W396.

29. Cumpston M, Li T, Page MJ, Thomas J. Updated guidance for trusted systematic reviews: a new edition of the Cochrane Handbook for Systematic Reviews of Interventions. *The Cochrane database of systematic reviews* 2019.

30. Menni C, Graham D, Kastenmüller G, Alharbi NH, Alsanosi SM, McBride M, Mangino M, Titcombe P, Shin S-Y, Psatha M, Geisendorfer T, Huber A, Peters A, Wang-Sattler R, Xu T, Brosnan MJ, Trimmer J, Reichel C, Mohney RP, Soranzo N, Edwards MH, Cooper C, Church AC, Suhre K, Gieger C, Dominiczak AF, Spector TD, Padmanabhan S, Valdes AM. Metabolomic identification of a novel pathway of blood pressure regulation involving hexadecanedioate. *Hypertension* 2015;**66**:422–429.

31. Menni C, Fauman E, Erte I, Perry JR, Kastenmüller G, Shin S-Y, Petersen A-K, Hyde C, Psatha M, Ward KJ, Yuan W, Milburn M, Palmer CNA, Frayling TM, Trimmer J, Bell JT, Gieger C, Mohney RP, Brosnan MJ, Suhre K, Soranzo N, Spector TD. Biomarkers for type 2 diabetes and impaired fasting glucose using a nontargeted metabolomics approach. *Diabetes* 2013;**62**: 4270–4276.

32. Menni C, Migaud M, Glastonbury CA, Beaumont M, Nikolaou A, Small KS, Brosnan MJ, Mohney RP, Spector TD, Valdes AM. Metabolomic profiling to dissect the role of visceral fat in cardiometabolic health. *Obesity* 2016;**24**:1380–1388.

33. Chen J, Zhang S, Wu J, Wu S, Xu G, Wei D. Essential role of nonessential amino acid glutamine in atherosclerotic cardiovascular disease. *DNA Cell Biol* 2020;**39**:8–15.

34. Ruiz-Canela M, Toledo E, Clish CB, Hruby A, Liang L, Salas-Salvado J, Razquin C, Corella D, Estruch R, Ros E, Fitó M, Gómez-Gracia E, Arós F, Fiol M, Lapetra J, Serra-Majem L, Martínez-González MA, Hu FB. Plasma branched-chain amino acids and incident cardiovascular disease in the PREDIMED trial. *Clin Chem* 2016;**62**:582–592.

35. Li Y-F, Li R-S, Samuel SB, Cueto R, Li X-Y, Wang H, Yang X-F. Lysophospholipids and their G protein-coupled receptors in atherosclerosis. *Front Biosci (Landmark edition)* 2016;**21**: 70–88.

36. Khurana S, Raufman JP, Pallone TL. Bile acids regulate cardiovascular function. *Clin Transl Sci* 2011;**4**:210–218.

37. Liu L, Su J, Li R, Luo F. Changes in intestinal flora structure and metabolites are associated with myocardial fibrosis in patients with persistent atrial fibrillation. *Front Nutr* 2021;**8**: 702085.

38. Cruz DE, Tahir UA, Hu J, Ngo D, Chen Z-Z, Robbins JM, Katz D, Balasubramanian R, Peterson B, Deng S. Metabolomic analysis of coronary heart disease in an African American cohort from the Jackson Heart Study. *JAMA Cardiol* 2022;**7**:184–194.

39. Ridlon JM, Kang D-J, Hylemon PB. Bile salt biotransformations by human intestinal bacteria. *J Lipid Res* 2006;**47**:241–259.

40. Higuchi H, Bronk SF, Takikawa Y, Werneburg N, Takimoto R, El-Deiry W, Gores GJ. The bile acid glycochenodeoxycholate induces trail-receptor 2/DR5 expression and apoptosis. *J Biol Chem* 2001;**276**:38610–38618.

41. Xanthopoulos A, Starling RC, Kitai T, Triposkiadis F. Heart failure and liver disease: cardiohepatic interactions. *JACC Heart Fail* 2019;**7**:87–97.

42. Kand'ár R, Žáková P. Allantoin as a marker of oxidative stress in human erythrocytes. *Clin Chem Lab Med* 2008;**46**:1270–1274.

43. Bos MJ, Koudstaal PJ, Hofman A, Witteman JC. Breteler MM. Uric acid is a risk factor for myocardial infarction and stroke: the Rotterdam study. *Stroke* 2006;**37**:1503–1507.

44. Chlopicki S, Swies J, Mogielnicki A, Buczko W, Bartus M, Lomnicka M, Adamus J, Gebicki J. 1-Methylnicotinamide (MNA), a primary metabolite of nicotinamide, exerts anti-thrombotic activity mediated by a cyclooxygenase-2/prostacyclin pathway. *Br J Pharmacol* 2007;**152**: 230–239.

45. Surendran A, Aliani M, Ravandi A. Metabolomic characterization of myocardial ischemia-reperfusion injury in ST-segment elevation myocardial infarction patients undergoing percutaneous coronary intervention. *Sci Rep* 2019;**9**:1–13.

46. Rutkowski B, Slominska E, Szolkiewicz M, Smolenski RT, Striley C, Rutkowski P, Swierczynski J. N-methyl-2-pyridone-5-carboxamide: a novel uremic toxin? *Kidney Int* 2003;**63**:S19–S21.

47. Falconi CA, da Cruz Junho CV, Fogaça-Ruiz F, Vernier ICS, Da Cunha RS, Stinghen AEM, Carneiro-Ramos MS. Uremic toxins: an alarming danger concerning the cardiovascular system. *Front Physiol* 2021;**12**:686249.

48. Zou Y, Yang Y, Fu X, He X, Liu M, Zong T, Li X, Aung LH, Wang Z, Yu T. The regulatory roles of aminoacyl-tRNA synthetase in cardiovascular disease. *Mol Ther Nucleic Acids* 2021;**25**:372–387.

49. Amin AM. The metabolic signatures of cardiometabolic diseases: does the shared metabotype offer new therapeutic targets? *Lifestyle Med* 2021;**2**:e25.

50. Zaric BL, Radovanovic JN, Gluvic Z, Stewart AJ, Essack M, Motwalli O, Gojobori T, Isenovic ER. Atherosclerosis linked to aberrant amino acid metabolism and immunosuppressive amino acid catabolizing enzymes. *Front Immunol* 2020;**11**:2341.

51. Jarmusch AK, Vrbanac A, Momper JD, Ma JD, Alhaja M, Liyanage M, Knight R, Dorrestein PC, Tsunoda SM. Enhanced characterization of drug metabolism and the influence of the intestinal microbiome: a pharmacokinetic, microbiome, and untargeted metabolomics study. *Clin Transl Sci* 2020;**13**:972–984.

52. Han X, Zhang Y, Yin L, Zhang L, Wang Y, Zhang H, Li B. Statin in the treatment of patients with myocardial infarction: a meta-analysis. *Medicine (Baltimore)* 2018;**97**:e0167.

53. De Vera MA, Bhole V, Burns LC, Lacaille D. Impact of statin adherence on cardiovascular disease and mortality outcomes: a systematic review. *Br J Clin Pharmacol* 2014;**78**:684–698.
54. Schwartz GG, Olsson AG, Ezekowitz MD, Ganz P, Oliver MF, Waters D, Zeiher A, Chaitman BR, Leslie S, Stern T. Effects of atorvastatin on early recurrent ischemic events in acute coronary syndromes: the MIRACL study: a randomized controlled trial. *JAMA* 2001;**285**:1711–1718.
55. Wang C-Y, Liu P-Y, Liao JK. Pleiotropic effects of statin therapy: molecular mechanisms and clinical results. *Trends Mol Med* 2008;**14**:37–44.
56. Walker R, Stewart L, Simmonds M. Estimating interactions in individual participant data meta-analysis: a comparison of methods in practice. *Syst Rev* 2022;**11**:211.
57. Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Eval Health Prof* 2002;**25**:76–97.

Translational perspective

In the largest meta-analyses covering six international cohorts, we identify 10 novel and 46 known metabolites associated with incident MI that can be used to identify at-risk individuals before disease onset. Our results improve our understanding of the molecular changes that take place in MI development and provide potential novel targets for clinical prediction and a deeper understanding of causal mechanisms.

Supplementary material

Supplementary Table 5.1 Metabolites significantly associated (meta-analysis FDR<0.05) with incident MI. TE and SE refer to estimated overall treatment effect and standard error, respectively.

Large table. Access to the table is given in the attached OneDrive. A list of the entire OneDrive content is listed in ***Appendix A***.

Supplementary Table 5.2 Literature references for the metabolites previously associated with any cardiac diseases, and the super- and sub-pathways for metabolites associated with incident MI. For the metabolites that did not remain significant after further adjusting the meta-analyses for prevalent hypertension, dyslipidaemia and type-2 diabetes, references showing their associations with any of these 3 conditions are indicated. *Large table. Access to the table is given in the attached OneDrive. A list of the entire OneDrive content is listed in* ***Appendix A****.*

Supplementary Table 5.3 Results of the random effect inverse-variance meta-analysis performed in the MI-associated metabolites (meta-analysis FDR<0.05) based on the results from the fixed effect inverse-variance meta-analysis.TE and SE refer to estimated overall treatment effect and standard error, respectively.

HMDB	Metabolite	Number of cohorts	TE Random	SE Random	P-value Random
HMDB00020	4-hydroxyphenylacetate	3	0.26	0.11	1.94E-04
HMDB00067	cholesterol	5	0.15	0.03	2.13E-05
HMDB00092	dimethylglycine	4	0.11	0.04	3.37E-03
HMDB00099	cystathionine	3	0.15	0.17	2.24E-03
HMDB00122	glucose	6	0.23	0.04	1.02E-09
HMDB00123	glycine	5	-0.05	0.09	2.42E-06
HMDB00126	glycerol 3-phosphate	4	-0.08	0.06	2.12E-02
HMDB00127	glucuronate	4	0.32	0.05	4.05E-12
HMDB00134	fumarate	3	0.23	0.12	3.63E-04
HMDB00138	glycocholate	3	0.11	0.17	7.05E-03
HMDB00148	glutamate	4	0.16	0.18	1.74E-10
HMDB00156	malate	4	0.15	0.04	1.00E-04
HMDB00167	threonine	4	-0.13	0.04	5.16E-04
HMDB00168	asparagine	4	-0.04	0.17	3.09E-08
HMDB00172	isoleucine	6	0.12	0.04	9.86E-04
HMDB00177	histidine	6	-0.14	0.03	1.02E-04
HMDB00187	serine	4	-0.19	0.06	9.65E-06
HMDB00190	lactate	6	0.11	0.07	1.72E-04
HMDB00206	N6-acetyllysine	3	0.18	0.04	1.61E-06
HMDB00211	myo-inositol	4	0.12	0.04	1.94E-03
HMDB00247	mannitol/sorbitol	3	0.34	0.06	1.76E-09
HMDB00254	succinate	4	0.16	0.04	1.77E-05
HMDB00259	serotonin	4	-0.12	0.04	1.62E-03
HMDB00289	urate	4	0.12	0.04	1.19E-03
HMDB00296	uridine	4	-0.20	0.04	6.34E-08
HMDB00462	allantoin	4	0.11	0.07	1.29E-02
HMDB00555	3-methyladipate	3	0.24	0.06	9.37E-05
HMDB00637	glycochenodeoxycholate	3	0.14	0.09	3.22E-03
HMDB00641	glutamine	4	-0.13	0.17	2.88E-10
HMDB00682	3-indoxyl sulfate	4	0.11	0.04	3.16E-03
HMDB00684	kynurenine	4	0.15	0.04	1.97E-04
HMDB00699	1-methylnicotinamide	3	-0.17	0.05	2.12E-03
HMDB00708	glycoursodeoxycholate	3	0.10	0.06	8.07E-03
HMDB00767	pseudouridine	3	0.19	0.04	2.02E-06
HMDB00840	2-hydroxyhippurate (salicylurate)	4	0.16	0.08	4.74E-03
HMDB00929	tryptophan	4	-0.09	0.04	1.54E-02
HMDB01043	arachidonate (20:4n6)	3	0.12	0.07	2.48E-04
HMDB01999	eicosapentaenoate (EPA; 20:5n3)	3	0.14	0.04	1.74E-04
HMDB02329	oxalate (ethanedioate)	3	0.15	0.14	5.91E-03
HMDB02925	dihomo-linolenate (20:3n3 or n6)	3	0.07	0.14	1.75E-05
HMDB03681	4-acetamidobutanoate	3	0.12	0.04	1.67E-03
HMDB04193	N1-Methyl-2-pyridone-5-carboxamide	3	0.11	0.04	2.49E-03
HMDB04949	N-palmitoyl-sphingosine (d18:1/16:0)	3	0.18	0.05	8.36E-04
HMDB06344	phenylacetylglutamine	3	0.12	0.04	2.80E-03
HMDB07103	1-palmitoyl-2-linoleoyl-glycerol (16:0/18:2)	3	0.19	0.06	9.77E-04
HMDB08993	1-stearoyl-2-oleoyl-GPE (18:0/18:1)	3	0.35	0.24	1.31E-03
HMDB08994	1-stearoyl-2-linoleoyl-GPE (18:0/18:2)	3	0.32	0.21	5.53E-03
HMDB10386	1-linoleoyl-GPC (18:2)	4	-0.10	0.06	5.62E-03
HMDB10395	1-arachidonoyl-GPC (20:4)	4	-0.12	0.06	1.01E-03
HMDB11103	1.7-dimethylurate	3	0.10	0.04	1.95E-02
HMDB11130	1-stearoyl-GPE (18:0)	4	0.13	0.04	2.17E-04
HMDB11211	1-(1-enyl-palmitoyl)-2-linoleoyl-GPC (P-16:0/18:2)	3	-0.14	0.06	1.44E-02
HMDB11503	1-palmitoyl-GPE (16:0)	4	0.17	0.05	7.34E-06
HMDB11506	1-oleoyl-GPE (18:1)	4	0.12	0.06	1.28E-03
HMDB13127	hydroxybutyrylcarnitine	4	0.08	0.05	1.94E-02
HMDB13678	4-hydroxyhippurate	3	0.10	0.04	1.68E-02

Supplementary Table 5.4 Meta-analysis results from the 56 metabolites significantly associated with incident MI when the analyses were run excluding the cohorts in which MI was assessed by self-reported questionnaires (TwinsUK and ET2DS). TE and SE refer to estimated overall treatment effect and standard error, respectively.

Large table. Access to the table is given in the attached OneDrive. A list of the entire OneDrive content is listed in ***Appendix A****.*

Supplementary Table 5.5 Meta-analysis results from the 56 metabolites significantly associated with incident MI when the models were further adjusted for prevalent hypertension, prevalent type-2 diabetes and prevalent dyslipidemia. Significant associations are marked in red. TE and SE refer to estimated overall treatment effect and standard error, respectively.

Large table. Access to the table is given in the attached OneDrive. A list of the entire OneDrive content is listed in ***Appendix A****.*

Supplementary Table 5.6 Meta-analysis results from the 56 metabolites significantly associated with incident MI when the models were stratified by race (White individuals and Black individuals). Significant associations are marked in red. TE and SE refer to estimated overall treatment effect and standard error, respectively.

Large table. Access to the table is given in the attached OneDrive. A list of the entire OneDrive content is listed in ***Appendix A****.*

Supplementary Table 5.7 Metabolites associated (meta-analysis nominal p-value<0.05) with prevalent MI, and that are also significantly associated with incident MI (meta-analysis FDR<0.05). TE and SE refer to estimated overall treatment effect and standard error, respectively.

Large table. Access to the table is given in the attached OneDrive. A list of the entire OneDrive content is listed in ***Appendix A****.*

Supplementary Table 5.8 Enrichment pathway analysis results showing all the identified pathways. 'Total' indicates the number of metabolites that are involved in each pathway, whereas 'Hits' indicates the number of metabolites associated with incident MI that are present in each pathway.

Pathway Name	Hits	Total	P-value	-log(p-value)	FDR	Impact	Matched Metabolites
Aminoacyl-tRNA biosynthesis	9	48	1.00E-06	6.00E+00	8.44E-05	0.17	Glycine, Histidine, Glutamine, Serine, Asparagine, Glutamate, Isoleucine, Threonine, Tryptophan
Alanine, aspartate and glutamate metabolism	5	28	4.41E-04	3.36E+00	1.85E-02	0.31	Glutamine, Aspargine, Glutamate, Fumarate, Succinate
Glyoxylate and dicarboxylate metabolism	5	32	8.40E-04	3.08E+00	2.04E-02	0.15	Glycine, Glutamine, Serine, Glutamate, Malate
Glycine, serine and threonine metabolism	5	33	9.73E-04	3.01E+00	2.04E-02	0.54	Glycine, Serine, Threonine, Dimethylglycine, Cystathionine
Arginine biosynthesis	3	14	4.11E-03	2.39E+00	6.90E-02	0.12	-
Nitrogen metabolism	2	6	8.25E-03	2.08E+00	9.90E-02	0	-
D-Glutamine and D-glutamate metabolism	2	6	8.25E-03	2.08E+00	9.90E-02	0.5	-
Citrate cycle (TCA cycle)	3	20	1.16E-02	1.93E+00	1.16E-01	0.11	-
Valine, leucine and isoleucine biosynthesis	2	8	1.49E-02	1.83E+00	1.16E-01	0	-
Ascorbate and aldarate metabolism	2	8	1.49E-02	1.83E+00	1.16E-01	0.5	-
Pyruvate metabolism	3	22	1.52E-02	1.82E+00	1.16E-01	0.03	-
Primary bile acid biosynthesis	4	46	2.40E-02	1.62E+00	1.68E-01	0.08	-
Butanoate metabolism	2	15	5.03E-02	1.30E+00	2.80E-01	0	-
Nicotinate and nicotinamide metabolism	2	15	5.03E-02	1.30E+00	2.80E-01	0.14	-
Biosynthesis of unsaturated fatty acids	3	36	5.56E-02	1.25E+00	2.80E-01	0	-
Glycerophospholipid metabolism	3	36	5.56E-02	1.25E+00	2.80E-01	0.2	-
Histidine metabolism	2	16	5.66E-02	1.25E+00	2.80E-01	0.22	-
Tryptophan metabolism	3	41	7.64E-02	1.12E+00	3.57E-01	0.34	-
Sphingolipid metabolism	2	21	9.18E-02	1.04E+00	4.06E-01	0.27	-
Glycolysis / Gluconeogenesis	2	26	1.32E-01	8.80E-01	5.53E-01	0	-
Galactose metabolism	2	27	1.40E-01	8.54E-01	5.60E-01	0	-
Glutathione metabolism	2	28	1.49E-01	8.28E-01	5.68E-01	0.11	-
Porphyrin and chlorophyll metabolism	2	30	1.66E-01	7.80E-01	5.81E-01	0	-
Inositol phosphate metabolism	2	30	1.66E-01	7.80E-01	5.81E-01	0.13	-
Cysteine and methionine metabolism	2	33	1.93E-01	7.15E-01	6.48E-01	0.2	-
Caffeine metabolism	1	10	2.20E-01	6.57E-01	7.12E-01	0	-
Arginine and proline metabolism	2	38	2.38E-01	6.23E-01	7.41E-01	0.09	-
Pyrimidine metabolism	2	39	2.48E-01	6.06E-01	7.43E-01	0.02	-
Tyrosine metabolism	2	42	2.75E-01	5.60E-01	7.97E-01	0.02	-
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	1	14	2.95E-01	5.31E-01	8.25E-01	0	-
Glycerolipid metabolism	1	16	3.29E-01	4.83E-01	8.92E-01	0.04	-
Pentose and glucuronate interconversions	1	18	3.62E-01	4.41E-01	9.50E-01	0.12	-
Fructose and mannose metabolism	1	20	3.93E-01	4.05E-01	1.00E+00	0.03	-
Beta-Alanine metabolism	1	21	4.08E-01	3.89E-01	1.00E+00	0	-
Propanoate metabolism	1	23	4.37E-01	3.59E-01	1.00E+00	0	-
Purine metabolism	2	65	4.79E-01	3.19E-01	1.00E+00	0	-
Phosphatidylinositol signaling system	1	28	5.04E-01	2.98E-01	1.00E+00	0.04	-
Arachidonic acid metabolism	1	36	5.95E-01	2.25E-01	1.00E+00	0.31	-
Valine, leucine and isoleucine degradation	1	40	6.34E-01	1.98E-01	1.00E+00	0	-
Steroid biosynthesis	1	42	6.52E-01	1.85E-01	1.00E+00	0.03	-
Steroid hormone biosynthesis	1	85	8.86E-01	5.27E-02	1.00E+00	0.01	-

Supplementary Text 5.1 Definition of MI by each COMETS cohort, and definition of the covariables used to adjust the statistical models.

MI was defined by each cohort as follows:

- ARIC: MI was assessed by hospital records and echocardiograms at follow-up visits, and by participants' medical histories and electrocardiograms administered at baseline visits.
- ET2DS: International Classification of Diseases (ICD) codes 121-123 were used for hospitalization and death records. Chest pain questionnaires, general practitioner records, and self-report questionnaires were also used to assist in identification.
- GDM: The patient's electronic medical records were reviewed and consider MI whether MI diagnosis was given by a cardiologist based on clinical evidence and complementary tests.
- HABC: Each participant was contacted every 6 months to query hospitalizations or major outpatient procedures. If participants could not be reached, information was ascertained from proxies. Records from all overnight hospitalizations were obtained and reviewed for incident MI. MI diagnoses were adjudicated by physicians at clinical sites from hospitalization and/or death records. The underlying and contributing causes of death were obtained from death certificates and including the adjudication.
- TwinsUK: MI was self-reported in questionnaires.
- WHI: In the WHI, the endpoint included MI or death due to CHD, and all events were confirmed by medical records and adjudication by a physician.
- KORA: MI were identified via the KORA Augsburg coronary event registry or through questionnaires for participants residing outside the study area 59.

MI prevalence and incidence, and the covariables were coded identically across the cohorts as indicated below:

- Prevalent MI: A given person had already suffered from MI at the time of the blood sample/metabolomics profiling (Prevalent MI = 1; dichotomous variable).
- Incident MI: A given person suffered from MI after the blood sample/metabolomics profiling (Incident MI = 1; dichotomous variable).
- Age baseline: age (years) of a subject at baseline timepoint.
- Age follow-up: age (years) of a subject at follow-up timepoint.
- Gender: dichotomous variable (1 or 0):
  - 1 = male
  - 0 = female
- BMI: body mass index (kg/m^2) of a subject at baseline timepoint.
- Smoking status: categorical variable with 3 levels:
  - 0 = never smoker
  - 1 = former smoker
  - 2 = current smoker
- Race: categorical variable indicating the ancestry:
  - 0 = White/European ancestry
  - 1 = Non-European ancestry
- Education level: categorical variable with 4 levels:
  - 0 = did not complete high school
  - 1 = completed high school
  - 2 = post-high school training/ some college
  - 3 = completed college
- Alcohol consumption: categorical variable with 4 levels:
  - 0 = zero alcohol intake
  - 1 = <0, 15] g/day

  - 2 = <15, 30] g/day
  - 3 = >30 g/day

- Physical activity level: categorical variable with 3 levels:

  - 0 = low
  - 1 = medium or missing
  - 2 = high

- Prevalent type-2 diabetes: dichotomous variable (1 or 0):

  - 0 = Type-2 diabetes had not been diagnosed by the time of the blood sample/metabolomics profiling (baseline)
  - 1 = Type-2 diabetes had been diagnosed by the time of the blood sample/metabolomics profiling (baseline)

- Prevalent hypertension (defined as systolic blood pressure>140 mmHg or diastolic blood pressure >90 mmHg or taking hypertension-lowering medications or diagnosed by the doctor as having hypertension at baseline): dichotomous variable (1 or 0):

  - 0 = Hypertension had not been diagnosed by the time of the blood sample/metabolomics profiling (baseline)
  - 1 = Hypertension had been diagnosed by the time of the blood sample/metabolomics profiling (baseline)

- Prevalent dyslipidaemia (defined as high levels of total cholesterol (>240 mg/dL) or high levels of triglycerides ($\geq$500 mg/dL) or low levels of HDL cholesterol ($\leq$40 mg/dL): dichotomous variable (1 or 0):

  - 0 = Dyslipidaemia had not been diagnosed by the time of the blood sample/metabolomics profiling (baseline)
  - 1 = Dyslipidaemia had been diagnosed by the time of the blood sample/metabolomics profiling (baseline)

Supplementary Text 5.2 STROBE Statement—Checklist of items that should be included in reports of cohort studies

	Item No	Recommendation
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported
Objectives	3	State specific objectives, including any prespecified hypotheses
Methods		
Study design	4	Present key elements of study design early in the paper
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants	6	(a) Give the eligibility criteria and the sources and methods of selection of participants. Describe methods of follow-up
		(b) For matched studies, give matching criteria and number of exposed and unexposed
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable
Data sources/measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group
Bias	9	Describe any efforts to address potential sources of bias
Study size	10	Explain how the study size was arrived at
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding
		(b) Describe any methods used to examine subgroups and interactions
		(c) Explain how missing data were addressed
		(d) If applicable, explain how loss to follow-up was addressed
		(e) Describe any sensitivity analyses
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed
		(b) Give reasons for non-participation at each stage
		(c) Consider use of a flow diagram
Descriptive data	14*	(a) Give characteristics of study participants (eg, demographic, clinical, social) and information on exposures and potential confounders
		(b) Indicate number of participants with missing data for each variable of interest
		(c) Summarise follow-up time (eg, average and total amount)
Outcome data	15*	Report numbers of outcome events or summary measures over time
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included
		(b) Report category boundaries when continuous variables were categorised
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses
Discussion		
Key results	18	Summarise key results with reference to study objectives
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence
Generalisability	21	Discuss the generalisability (external validity) of the study results
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.

Chapter 6

A faecal metabolite signature of prediabetes

10.5% of the total population is affected by T2D. Before the disease onset, individuals suffer from prediabetes. Prediabetes is a reversible metabolic condition associated with the gut microbiome, however, mechanisms remain elusive. Faecal metabolites provide a functional readout of the gut microbiome, offering a novel framework for investigating the impact of the gut microbiome on human health.

In this chapter, I search for a faecal metabolite signature associated with prediabetes and predictive of incident T2D. Moreover, I analyse the association between the identified signature and the gut microbiome composition to gain insights into the underlying mechanisms of action.

The obtained results illustrate a novel mechanism through which the gut microbiome impacts prediabetes. Specifically, the gut microbiome might influence prediabetes by affecting intestinal absorption or excretion of host compounds and xenobiotics.

QC of the faecal metabolites was conducted by Ms Francesca Tettamanzi. Dr Colette Christiansen cleaned the diabetic data. Andrei-Florin Balean processed the stool samples for the gut microbiome sequencing, and Dr Alessia Visconti generated and performed the quality control of the shotgun metagenome data. I conducted the statistical analyses on TwinsUK and Ms Qiuling Dong replicated the results in KORA. Finally, I wrote the original draft of the manuscript.

This chapter has been published in *Diabetes* (Nogal et al., 2023). An extension of the discussion, which is not included in the published manuscript, can be found in **Appendix B**.

GENETICS/GENOMES/PROTEOMICS/METABOLOMICS

A Fecal Metabolite Signature of Impaired Fasting Glucose: Results From Two Independent Population-Based Cohorts

Ana Nogal,[1] Francesca Tettamanzi,[1,2] Qiuling Dong,[3] Panayiotis Louca,[1] Alessia Visconti,[1] Colette Christiansen,[1,4] Taylor Breuninger,[5] Jakob Linseisen,[5,6,7] Harald Grallert,[3,8] Nina Wawro,[5,9] Francesco Asnicar,[10] Kari Wong,[11] Andrei-Florin Baleanu,[1] Gregory A. Michelotti,[11] Nicola Segata,[10] Mario Falchi,[1] Annette Peters,[8,9,12] Paul W. Franks,[13,14] Vincenzo Bagnardi,[15] Tim D. Spector,[1] Jordana T. Bell,[1] Christian Gieger,[3,8] Ana M. Valdes,[16] and Cristina Menni[1]

Diabetes 2023;72:1870–1880 | https://doi.org/10.2337/db23-0170

Prediabetes is a metabolic condition associated with gut microbiome composition, although mechanisms remain elusive. We searched for fecal metabolites, a readout of gut microbiome function, associated with impaired fasting glucose (IFG) in 142 individuals with IFG and 1,105 healthy individuals from the UK Adult Twin Registry (TwinsUK). We used the Cooperative Health Research in the Region of Augsburg (KORA) cohort (318 IFG individuals, 689 healthy individuals) to replicate our findings. We linearly combined eight IFG-positively associated metabolites (1-methylxanthine, nicotinate, glucuronate, uridine, cholesterol, serine, caffeine, and protoporphyrin IX) into an IFG-metabolite score, which was significantly associated with higher odds ratios (ORs) for IFG (TwinsUK: OR 3.9 [95% CI 3.02–5.02], *P* < 0.0001, KORA: OR 1.3 [95% CI 1.16–1.52], *P* < 0.0001) and incident type 2 diabetes (T2D; TwinsUK: hazard ratio 4 [95% CI 1.97–8], *P* = 0.0002). Although these are host-produced metabolites, we found that the gut microbiome is strongly associated with their fecal levels (area under the curve >70%). Abundances of *Faecalibacillus intestinalis*, *Dorea formicigenerans*, *Ruminococcus torques*, and *Dorea sp. AF24-7LB* were positively associated with IFG, and such associations were partially mediated by 1-methylxanthine and nicotinate (variance accounted for mean 14.4% [SD 5.1], *P* < 0.05). Our results suggest that the gut microbiome is linked to prediabetes not only via the production of microbial metabolites but also by affecting intestinal absorption/excretion of host-produced metabolites and xenobiotics, which are correlated with the risk of IFG. Fecal metabolites enable modeling of another mechanism of gut microbiome effect on prediabetes and T2D onset.

ARTICLE HIGHLIGHTS

- Prediabetes is a metabolic condition associated with gut microbiome composition, although mechanisms remain elusive.
- We investigated whether there is a fecal metabolite signature of impaired fasting glucose (IFG) and the possible underlying mechanisms of action.
- We identified a fecal metabolite signature of IFG associated with prevalent IFG in two independent cohorts and incident type 2 diabetes in a subanalysis. Although the signature consists of metabolites of nonmicrobial origin, it is strongly correlated with gut microbiome composition.
- Fecal metabolites enable modeling of another mechanism of gut microbiome effect on prediabetes by affecting intestinal absorption or excretion of host compounds and xenobiotics.

Type 2 diabetes (T2D) is a leading cause of mortality and morbidity (1), affecting >536.6 million people (10.5% of the total population) worldwide (2), thus representing a huge public health burden (1). The causation of T2D is multifactorial, influenced by host genetics and environmental

[1] Department of Twin Research, King's College London, St Thomas' Hospital Campus, London, U.K.

[2] Humanitas Clinical and Research Centre, IRCCS, Rozzano (Milan), Italy

[3] Institute of Epidemiology, Helmholtz Zentrum München, Research Unit of Molecular Epidemiology, German Research Center for Environmental Health (GmbH), Neuherberg, Germany

[4] School of Mathematics and Statistics, The Open University, Milton Keynes, U.K.

[5] Epidemiology, University Hospital Augsburg, University of Augsburg, Augsburg, Germany

[6] ZIEL-Institute for Food & Health, Technische Universität München, Freising, Germany

[7] Institute for Medical Information Processing, Biometry, and Epidemiology, Medical Faculty, Ludwig-Maximilian University Munich, Munich, Germany

[8] German Center for Diabetes Research (DZD), Neuherberg, Germany

[9] Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health (GmbH), Neuherberg, Germany

[10] Department of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, Trento, Italy

factors, including diet, obesity, inactivity, and smoking, and the interaction between these factors (3). Furthermore, its onset is gradual, with people progressing through a state of prediabetes (4), and is defined as impaired levels of fasting glucose (IFG), and/or glucose intolerance, and/or elevated hemoglobin A_{1c} (HbA_{1c}) (5).

Over the past decade, T2D and prediabetes have been linked by us and others (6–8) to changes in the gut microbiota, and we have recently demonstrated that T2D development is preceded by an alteration in gut microbiota composition (7). A critical challenge in human microbiome research, however, is to characterize and quantify metabolic activity across the full microbial ecosystem (9). The gut microbiome is highly variable, and different bacterial types may have similar metabolic effects on the host. Microbial metabolites are now widely seen as key mediators of the effects of gut microbiome composition on human physiology (10). Fecal metabolites provide a functional readout of the gut microbiome (11,12) and are a novel tool to explore links between gut microbiome composition and activity, host phenotypes, and heritable complex traits, thus improving our understanding of the impact that the gut microbiome can have on its host (11). As the gut microbiome is modifiable with nutritional and lifestyle interventions (13), it is of utmost importance to identify alterations in the fecal metabolites abundances, which reflect metabolic activity perturbations of the human gut microbial ecosystem that might lead to T2D onset.

In the first fecal metabolomics study of prediabetes to date, we aim to identify a fecal metabolite signature of this condition in two independent cohorts to shed light on mechanisms of action underlying T2D onset and development. Addressing this challenge also has long-term implications for future studies into therapies and lifestyle interventions that alter microbial metabolic activity to improve human health.

RESEARCH DESIGN AND METHODS

A flowchart of the study design with the main results is presented in Fig. 1.

Discovery Cohort

We analyzed data from 1,247 nonrelated individuals from UK Adult Twin Registry (TwinsUK) (14), for whom concurrent nontargeted fecal metabolomic profiling (526 metabolites at fasting) and glucose/diabetes information were available (cross-sectional design). Concurrent metagenome sequencing (as a measure of the gut microbiome composition) was also available for a subset of 342 individuals. Subjects were classified into three groups following the American Diabetes Association criteria based on isolated fasting glucose levels (15) at the time of the initial sampling and at subsequent visits (on average, 3.5 [SD 2.0] visits, 4.6 [SD 2.7] years apart): individuals with T2D (fasting glucose ≥7 mmol/L or physician's letter confirming diabetes diagnosis), individuals with IFG (fasting glucose >5.5 to <7 mmol/L, not on diabetes medication), and subjects without IFG and T2D (fasting glucose >3.9 to ≤5.5 mmol/L) (see Table 1). We refer to "healthy individuals" to indicate individuals without IFG and/or T2D.

Only one twin per twin pair was included in the analyses to eliminate potential bias through correlated error, which might inflate effect estimates.

In a small subanalysis, we included individuals with incident T2D (average follow-up time 2.1 [SD 1.3] years) and an independent subset of healthy individuals who remained healthy during follow-up.

All twins provided informed written consent and the study was approved by St Thomas' Hospital Research Ethics Committee (REC Ref: EC04/015).

Replication Cohort

The Cooperative Health Research in the Region of Augsburg (KORA) study is a population-based cohort study. The KORA FF4 study (2013–2014) is the second follow-up of KORA S4 (1999–2001). The 1,007 samples included in the study were collected in the morning between 8:00 A.M. and 10:30 A.M. after at least 8 h of fasting. Metabolon untargeted liquid chromatography/mass spectrometry (MS)-based techniques were applied to measure the metabolites in the KORA cohort (a different version of the platform used in TwinsUK). Healthy individuals and IFG individuals were assigned based on the same criteria as in TwinsUK (described in the above section and in Table 1).

Fecal Metabolomics Profiling

Metabolomics profiling was conducted using ultrahigh-performance liquid chromatography-tandem MS (MS/MS) by the metabolomics provider Metabolon Inc. (Morrisville,

[11] Metabolon, Morrisville, NC
[12] Munich Heart Alliance, German Center for Cardiovascular Research (DZHK e.V., Partner-Site Munich), Munich, Germany
[13] Lund University Diabetes Center, Lund University, Malmö, Sweden
[14] Department of Clinical Sciences, Lund University, Malmö, Sweden
[15] Department of Statistics and Quantitative Methods, University of Milan-Bicocca, Milan, Italy
[16] Academic Rheumatology Clinical Sciences Building, Nottingham City Hospital, University of Nottingham, U.K.

Corresponding authors: Cristina Menni, cristina.menni@kcl.ac.uk, and Ana M. Valdes, ana.valdes@nottingham.ac.uk

Received 28 February 2023 and accepted 30 August 2023

This article contains supplementary material online at https://doi.org/10.2337/figshare.24114387.

A.N. and F.T. contributed equally.

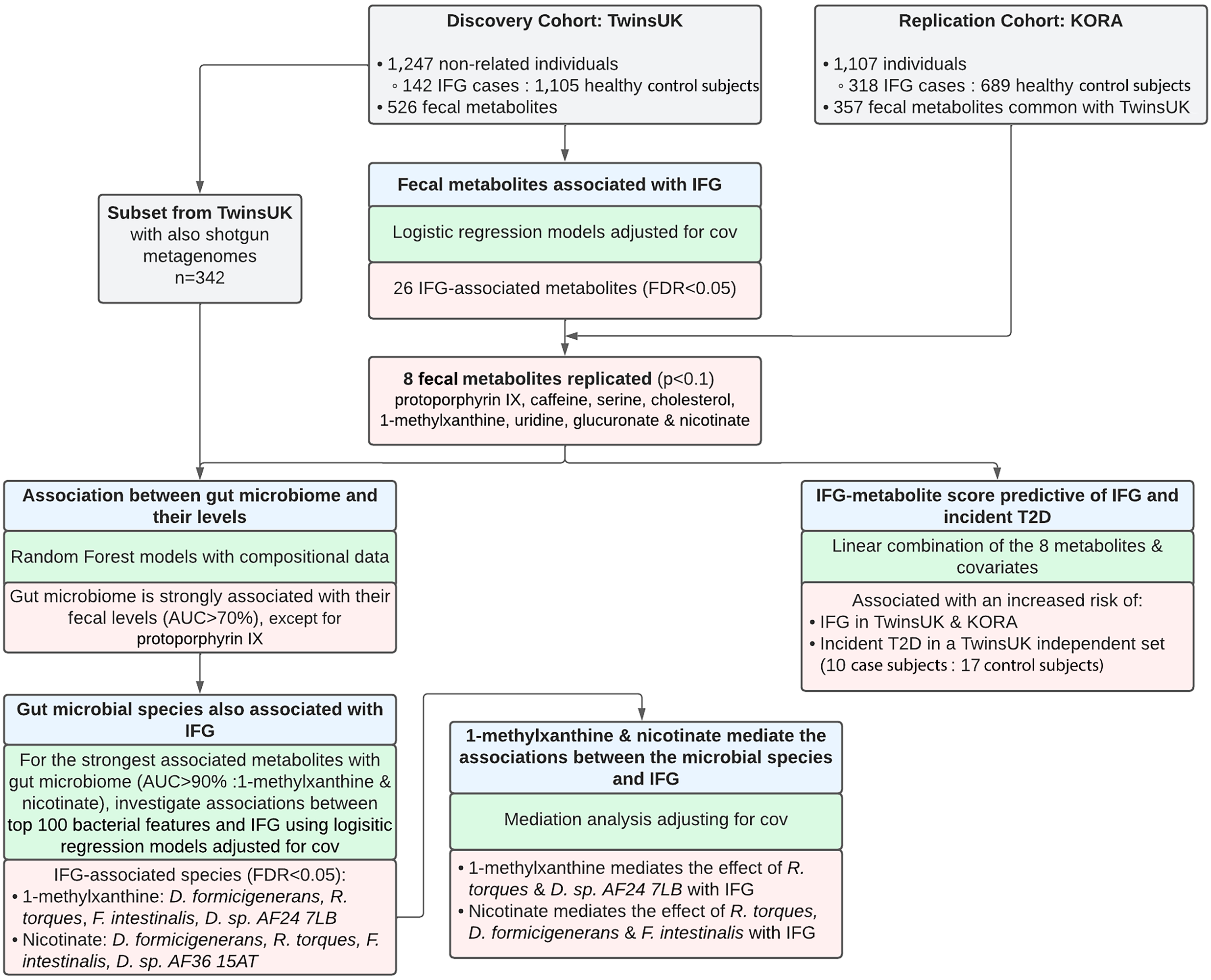

Discovery Cohort: TwinsUK
• 1,247 non-related individuals
◦ 142 IFG cases : 1,105 healthy control subjects
• 526 fecal metabolites
Replication Cohort: KORA
• 1,107 individuals
◦ 318 IFG cases : 689 healthy control subjects
• 357 fecal metabolites common with TwinsUK
Subset from TwinsUK with also shotgun metagenomes n=342
Fecal metabolites associated with IFG
Logistic regression models adjusted for cov
26 IFG-associated metabolites (FDR<0.05)
8 fecal metabolites replicated (p<0.1)
protoporphyrin IX, caffeine, serine, cholesterol, 1-methylxanthine, uridine, glucuronate & nicotinate
Association between gut microbiome and their levels
Random Forest models with compositional data
Gut microbiome is strongly associated with their fecal levels (AUC>70%), except for protoporphyrin IX
IFG-metabolite score predictive of IFG and incident T2D
Linear combination of the 8 metabolites & covariates
Associated with an increased risk of:
• IFG in TwinsUK & KORA
• Incident T2D in a TwinsUK independent set (10 case subjects : 17 control subjects)
Gut microbial species also associated with IFG
For the strongest associated metabolites with gut microbiome (AUC>90% :1-methylxanthine & nicotinate), investigate associations between top 100 bacterial features and IFG using logisitic regression models adjusted for cov
IFG-associated species (FDR<0.05):
• 1-methylxanthine: D. formicigenerans, R. torques, F. intestinalis, D. sp. AF24 7LB
• Nicotinate: D. formicigenerans, R. torques, F. intestinalis, D. sp. AF36 15AT
1-methylxanthine & nicotinate mediate the associations between the microbial species and IFG
Mediation analysis adjusting for cov
• 1-methylxanthine mediates the effect of R. torques & D. sp. AF24 7LB with IFG
• Nicotinate mediates the effect of R. torques, D. formicigenerans & F. intestinalis with IFG

Figure 1—Flowchart of the study design with the main results. Data, aims, methods, and results are shown in gray, blue, green, and pink squares, respectively. Mediation analyses were also performed for the metabolites making up the score that was predicted by the gut microbiome composition with an AUC >70%. Cov, covariates (age, BMI, and sex).

NC) on fecal samples from participants in the TwinsUK and KORA cohorts (Supplementary Material). The metabolomic data set measured by Metabolon includes 526 known metabolites for TwinsUK belonging to the following broad categories—amino acids, peptides, carbohydrates, energy intermediates, lipids, nucleotides, cofactors and vitamins, and xenobiotics—of which 357 were also measured in KORA. These include metabolites of established microbial origin (16). A complete list of the included metabolites with their superpathways, subpathways, Kyoto Encyclopedia of Genes and Genomes and Human Metabolome Database identifiers are reported in Supplementary Table 1. We imputed to the day minimum metabolites with <20% missing.

Metagenomic Assessment

Gut microbiota composition was generated from fecal shotgun metagenomes for a subset of the discovery cohort. DNA extraction, library preparation, and sequencing were conducted as detailed in Visconti et al. (11). For details see the Supplementary Material. Of note, gut microbiota composition is described by species-level genome bins (SGBs), which is the best proxy to define microbial species (17).

Statistical Analysis

Statistical analyses were conducted using R 4.2.2 software. To identify a fecal metabolite signature of prediabetes, we ran logistic regressions adjusting for age, BMI, sex, and multiple testing using the Benjamini and Hochberg method (18) (false discovery rate [FDR] <0.05). We then checked whether the metabolites significantly associated with IFG in the discovery set were also replicated in KORA ($P < 0.1$). We used a less stringent threshold for KORA because of the winner's curse (the effect sizes of the most strongly associated variables within a cohort-specific analysis are inflated) (19). Results were meta-analyzed using inverse-variance random-effect meta-analysis. We then created the IFG-metabolite score

Table 1—Descriptive characteristics of the study populations

	Discovery cohort: TwinsUK						Replication cohort: KORA		
	Prevalent IFG (n = 1,247)			Incident T2D (n = 27)			Prevalent IFG (n = 1,007)		
	Healthy individuals	IFG individuals	Differences between groups (*P* value)	Healthy individuals	T2D individuals	Differences between groups (*P* value)	Healthy individuals	IFG individuals	Differences between groups (*P* value)
ADA definition (15), fasting glucose, mmol/L	≤5.5	>5.5 and <7	—	≤5.5	≥7	—	≤5.5	>5.5 and <7	—
No.	1,105	142	—	17	10	—	689	318	—
Females, %	88.8	79	0.003	94.1	90	1	58.1	35.5	
Age, years	56.6 (14.9)	67.1 (10)	<0.0001	66.5 (6.6)	65 (7.7)	0.7	55.2 (10.9)	59.8 (10.8)	<0.0001
BMI, kg/m^2	25.2 (4.6)	28.5 (5.1)	<0.0001	25.3 (3.3)	35.1 (6.7)	0.0004	26.1 (4.1)	28.4 (4.5)	<0.0001
Circulating fasting glucose, mmol/L	4.5 (0.3)	5.9 (0.4)	<0.0001	3.8 (0.3)	4.6 (1.5)	0.06	5.1 (0.3)	5.9 (0.3)	<0.0001
SBP, mmHg	125 (13.6)	134 (17)	<0.0001	132 (21.5)	133 (9.5)	0.8	114.5 (15.9)	123.1 (15.3)	<0.0001
DBP, mmHg	74.7 (8.1)	77.9 (10.3)	<0.0001	73.8 (11.9)	81.7 (8.1)	0.05	72.3 (8.9)	76 (9.7)	<0.0001
Circulating HDL, mmol/L	1.8 (1.2)	1.6 (1)	0.003	1.7 (0.4)	1.3 (0.2)	0.004	1.8 (0.5)	1.6 (0.5)	<0.0001
Circulating total cholesterol, mmol/L	4.1 (0.5)	4.1 (0.7)	0.73	4.7 (1.2)	3.6 (0.8)	0.02	5.6 (1)	5.7 (1)	0.008
Circulating triglycerides, mmol/L	1 (1)	1.6 (2.7)	0.0003	1 (0.3)	1.3 (0.4)	0.01	1.2 (0.7)	1.4 (0.9)	<0.0001
aHEI	70.5 (6.4)	70.1 (6.5)	0.49	72.8 (9.9)	71.4 (6.2)	0.68	NA	NA	NA
Current smoker, *n*	No: 1,060 Yes: 45	No: 139 Yes: 3	0.36	No: 17	No: 10	—	No: 346 Yes: 343	No: 139 Yes: 179	0.9
Activity levels, *n*	Low: 100 Moderate: 802 High: 203	Low: 13 Moderate: 102 High: 27	0.98	Low: 3 Moderate: 11 High: 3	Low: 2 Moderate: 5 High: 3	0.71	Inactive: 225 Active: 464	Inactive: 133 Active: 185	0.006

Continuous variables are presented as mean (SD). Measures are shown at baseline. “Healthy individuals” refers to individuals with no IFG or T2D. There was no overlap between the healthy subjects from the IFG and incident T2D data sets. The *P* values are from a Wilcoxon test/*t* test (continuous variable) or χ^2 test (categorical variable), calculated to check whether differences existed between the different subject groups for the described parameters. ADA, American Diabetes Association; DBP, diastolic blood pressure; NA, not available; SBP, systolic blood pressure.

by linearly combining the replicated metabolites along with covariates. To assess the performance of the score in predicting prevalent IFG and incident T2D, we calculated the area under the curve (AUC) values obtained using fivefold cross-validation (caret package implemented in R [20]). Finally, logistic and Cox regressions were used to investigate the association between the IFG score (*Z*-scaled) and prevalent IFG risk and incident T2D risk, respectively.

Given the strong association between fecal metabolites and gut microbiome composition (12), we investigated to what extent the gut microbiota composition was associated with each of the replicated metabolites using random forest regressors and classifiers with compositional data and fivefold cross-validation. The performance was calculated using the average of the obtained Spearman correlations between the observed metabolite levels and the levels predicted by the model (denoted as ρ) over the fivefolds used as a test set for the regressors and the average of the obtained AUC values over the testing folds for the classifiers. For details see the Supplementary Material.

We further investigated the associations between their top 100 bacterial features and IFG by running logistic regression models adjusting for covariates and multiple testing species (FDR <0.05). Specifically, we included all of the fecal metabolites that could be predicted by the gut microbiome with an AUC >70%, and we then focused on those that had an outstanding prediction performance (AUC >90%).

Finally, we used formal mediation analysis as implemented in the R package “mediation” with 1,000 nonparametric bootstrap samples (21) to test the mediation effects of the metabolites on the total effect of the gut bacteria on IFG. The mediation model was used to quantify both the direct effect of these gut bacterial species on IFG and the indirect (mediated) effects mentioned above while controlling for age, BMI, and sex. The variance accounted for (VAF) score, which represents the ratio of indirect-to-total effect and determines the proportion of the variance explained by the mediation process, was used to determine the significance of the mediation effect.

Data and Resource Availability

The data used in this study are held by the Department of Twin Research at King’s College London. The data can be released to bona fide researchers using our normal procedures overseen by the Wellcome Trust and its guidelines as part of our core funding (https://twinsuk.ac.uk/resources-for-researchers/access-our-data/). The gut microbiome data are available on EBI (https://www.ebi.ac.uk/) under accession number PRJEB32731 (TwinsUK). The KORA FF4 datasets are available upon application through the KORA-PASST (project application self-service tool, https://www.helmholtz-munich.de/epi/research/cohorts/kora-cohort/data-use-and-access-via-korapasst/index.html).

RESULTS

We included 1,247 unrelated individuals from the TwinsUK cohort who had fecal metabolite measures along with glucose/diabetes and prediabetes information. Of these, 142 individuals had IFG (mean fasting glucose 5.9 mmol/L [SD 0.4]) and 1,105 were healthy individuals (mean fasting glucose 4.5 mmol/L [SD 0.3]). Descriptive characteristics of the discovery and replication populations are included in Table 1.

Fecal Metabolites Associated Cross-sectionally to IFG

Of the 526 known fecal metabolites analyzed in TwinsUK, the fecal abundances of 26 compounds were associated with IFG after adjusting for age, BMI, sex, and multiple testing (FDR <0.05) (Fig. 2). Identified metabolites were mainly amino acids (*n* = 7) and lipids (*n* = 7), but also included xenobiotics (*n* = 4), cofactors and vitamins (*n* = 3), nucleotides (*n* = 2), carbohydrates (*n* = 2), and one energy-related metabolite (Fig. 2). All significant metabolites, but 3-hydroxyoleate, octadecanedioate (C18-DC), azelate (C9-DC), γ-tocotrienol, and enterolactone, were positively associated with IFG (Fig. 2). Of the 26 metabolites, 18 were also measured in KORA (Supplementary Table 1), and 8 metabolites were replicated ($P < 0.1$) (Fig. 3). These were the lipid cholesterol (sterol metabolism), the carbohydrate glucuronate (aminosugar metabolism), the cofactors/vitamins nicotinate (nicotinate and nicotinamide metabolism) and protoporphyrin IX (hemoglobin and porphyrin metabolism), the xenobiotics caffeine and 1-methylxanthine (both involved in the xanthine metabolism), the amino acid serine (glycine, serine, and threonine metabolism), and the nucleotide uridine (pyrimidine metabolism). The correlation matrices for the eight fecal metabolites in TwinsUK and KORA are depicted in Supplementary Fig. 1. We combined the results from both cohorts using inverse-variance random-effect meta-analysis (Fig. 3).

IFG-Metabolite Score and Predictive Power

We then generated the IFG-metabolite score using TwinsUK individuals:

IFG-metabolite score = −8.79 + 0.07 × *glucuronate* + 0.25 × *protoporphyrin IX* + 0.09 × *1-methylxanthine* + 0.14 × *cholesterol* + 0.04 × *serine* + 0.07 × *uridine* + 0.04 × *nicotinate* + 0.17 × *caffeine* + 0.07 × *age* + 0.1 × *BMI* − 0.6 × *sex (female = 1)*

The IFG-metabolite score was associated with an increased risk of IFG in TwinsUK (odds ratio [OR] 3.9 [95% CI 43.02–5.02], $P < 0.0001$) and in KORA (OR 1.3 [95% CI 1.16–1.52], $P < 0.0001$). The association remained significant when further adjusting for clinical covariates (i.e., systolic and diastolic blood pressure, circulating levels of HDL, total cholesterol, and triglycerides, alternative health eating index [aHEI – not available in KORA], activity levels and smoking status) (Table 1) in both cohorts (TwinsUK: OR 3.4 [95% CI 2.65–4.49], $P < 0.0001$; KORA: OR 1.2 [95% CI 1.06–1.41], $P = 0.008$). Finally, the IFG-metabolite score accurately predicted prevalent IFG in TwinsUK with an AUC of 79.8% (95% CI 76.3–83.3) in fivefold stratified cross-validation and

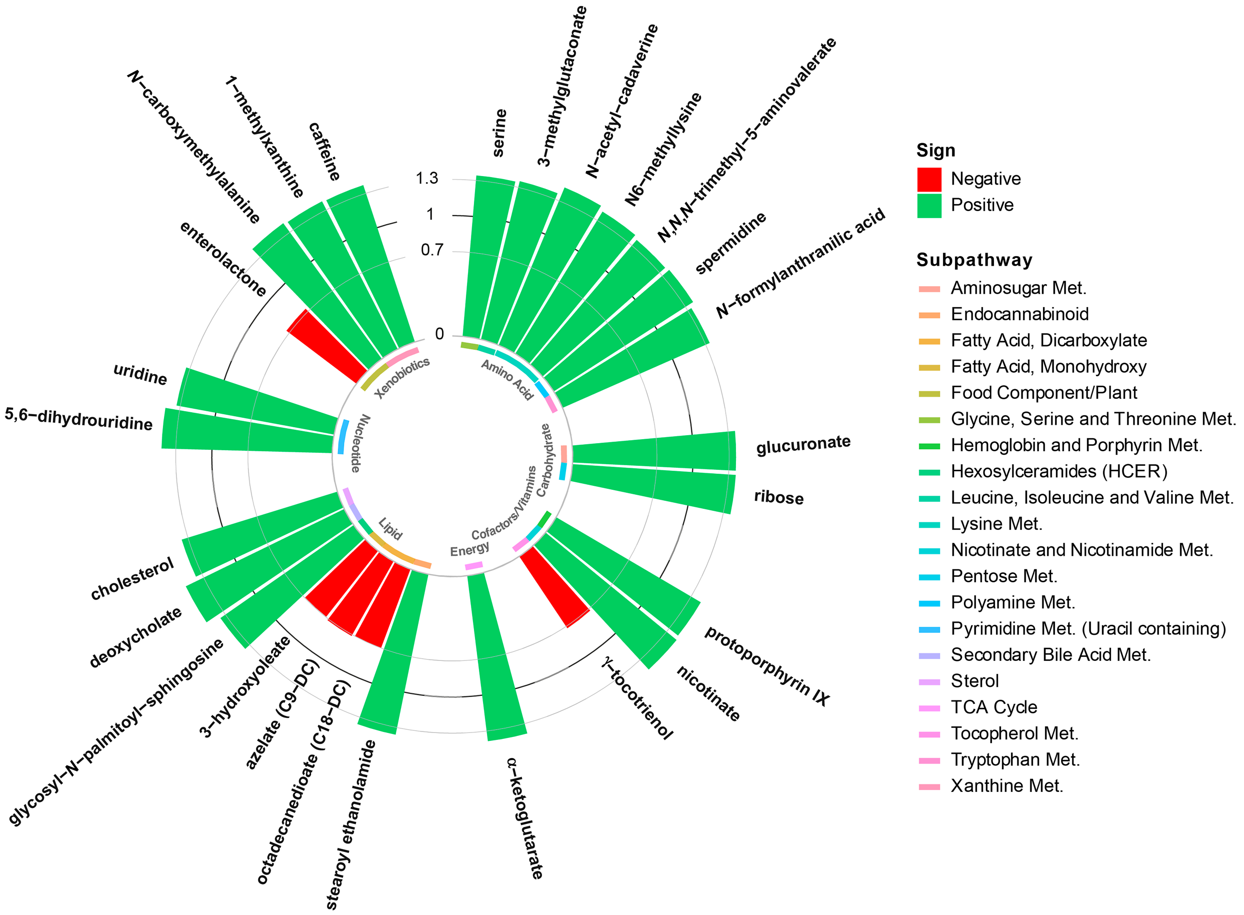


Figure 2—Fecal metabolites significantly associated with IFG in 1,247 individuals from TwinsUK after adjusting for baseline age and BMI, sex, and multiple testing (FDR <0.05). Bars represent the OR. Base labels illustrate subpathways. met., metabolism.

outperformed the model including only covariates (AUC 77.2% [95% CI = 73.6–81]) by 2.6% (Δ95% CI 2.7–2.1). In KORA, the IFG-metabolite score (top vs. lowest decile) could satisfactory predict prevalent IFG (AUC 65.4 [95% CI 57.9–73]).

Subanalysis: Incident T2D

In a small independent sample from TwinsUK (descriptive characteristics are shown in Table 1) consisting of 17 healthy individuals (different from the healthy subjects of the IFG data set) and 10 individuals with incident T2D (follow-up time between fecal metabolite measurements and incident events: mean 2.1 [SD 1.3] years), the IFG-metabolite score was also predictive of an increased risk of incident T2D (hazard ratio 4 [95% CI 1.97–8], *P* = 0.0002) in TwinsUK after further adjusting for baseline circulating glucose levels. It also accurately predicted incident T2D (AUC 83.3% [95% CI 74.4–92.2]), while a model using baseline circulating glucose levels as predictor presented a lower prediction power (AUC 72.4% [95% CI 51.8–92.9]).

Gut Microbiome–Fecal Metabolites Association

We further evaluated the extent to which the gut microbiota was associated with the fecal abundances of the eight replicated metabolites using the AUC obtained by the random forest classifiers and the Spearman correlations (denoted as ρ) between the real abundances and predicted values by the random forest regressors. We included a subset of 342 individuals from TwinsUK with concurrent gut microbiota composition assessed by shotgun metagenomics and fecal metabolites measurements. Descriptive characteristics of this subset are shown in Supplementary Table 2.

The gut microbiome composition was strongly associated with the replicated metabolites, with performance metric values ranging from an AUC of 70.7% (95% CI, 69.1–72.4) and ρ = 0.24 (95% CI, 0.23–0.25) for caffeine to an AUC of 91.4% (95% CI, 90.8–91.9) and ρ = 0.62 (95% CI, 0.62–0.62) for 1-methylxanthine (Fig. 4*A* and Supplementary Table 3). Protoporphyrin IX was the only metabolite presenting a moderate association (AUC 64.8% [95% CI 63.9–65.6]; ρ = 0.25 [95% CI 0.24–0.26]) (Fig. 4*A*).

We then investigated whether the abundances from their top 100 bacterial features based on the random

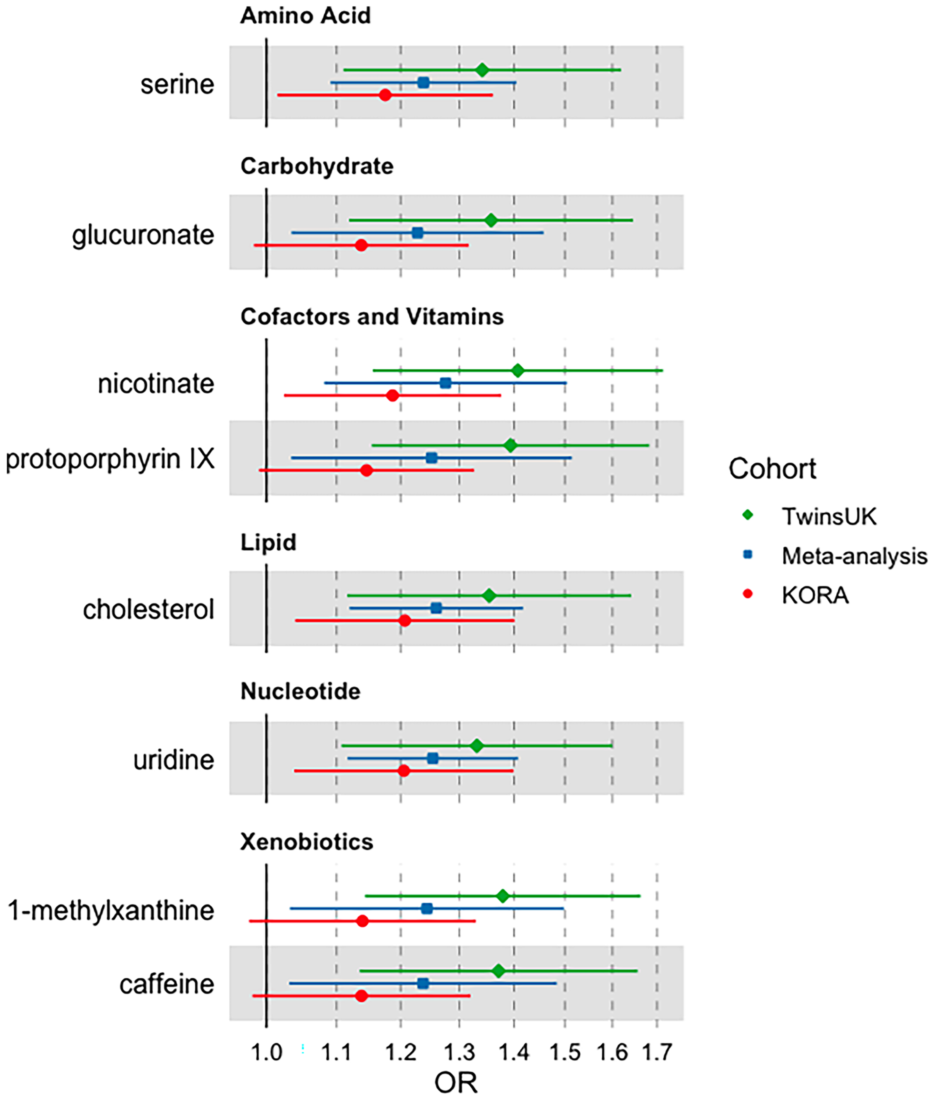


Figure 3—Fecal metabolites significantly associated with IFG after adjusting for age, BMI, and sex in TwinsUK (FDR <0.05), KORA ($P < 0.1$) and in the overall cohort (applying inverse-variance random-effect meta-analysis). The OR and 95% CI are indicated.

forest models were also significantly associated with IFG (Supplementary Table 4). We focused on the fecal metabolites that presented the strongest associations with the gut microbiome composition (AUC >90%—outstanding prediction performance; 1-methylxathine and nicotinate). We identified four characterized gut bacterial species for 1-methylxanthine and nicotinate, of which three overlapping (overlapping: *Dorea formicigenerans*, *Ruminococcus torques*, and *Faecalibacillus intestinalis*; 1-methylxanthine only: *Dorea* sp. AF24-7LB; nicotinate only: *Dorea* sp. AF36-15AT), that were positively associated with IFG after adjusting for age, BMI and sex (FDR <0.05) (Supplementary Table 4). We, therefore, performed a formal mediation analysis adjusting for age, BMI, and sex to determine whether 1-methylxanthine and/or nicotinate mediated the associations between these species and IFG. The analysis revealed that 1-methylxanthine acted as a potential mediator in the positive associations of *Dorea* sp. AF24-7LB (VAF = 10.3%, P = 0.03) and *R. torques* (VAF = 9.7%, P = 0.04) with IFG, while nicotinate acted as a potential mediator in the positive associations of *F. intestinalis* (VAF = 22.3%, P = 0.002), *D. formicigenerans* (VAF = 15.8%, P = 0.002), and *R. torques* (VAF = 14.1%, P = 0.03) with IFG (Fig. 4*B*). We further ran mediation analyses for the metabolites that could be predicted by the gut microbiome with an AUC >70%. As reported in Supplementary Fig. 2, uridine, serine, cholesterol, and caffeine were also mediators in the associations between different species (e.g., *Dorea* spp. and *Anaerobutyricum hallii*) and IFG. Models were not further adjusted for other comorbidities (e.g., systolic and diastolic blood pressure, circulating levels of HDL, total cholesterol and triglycerides, aHEI, activity

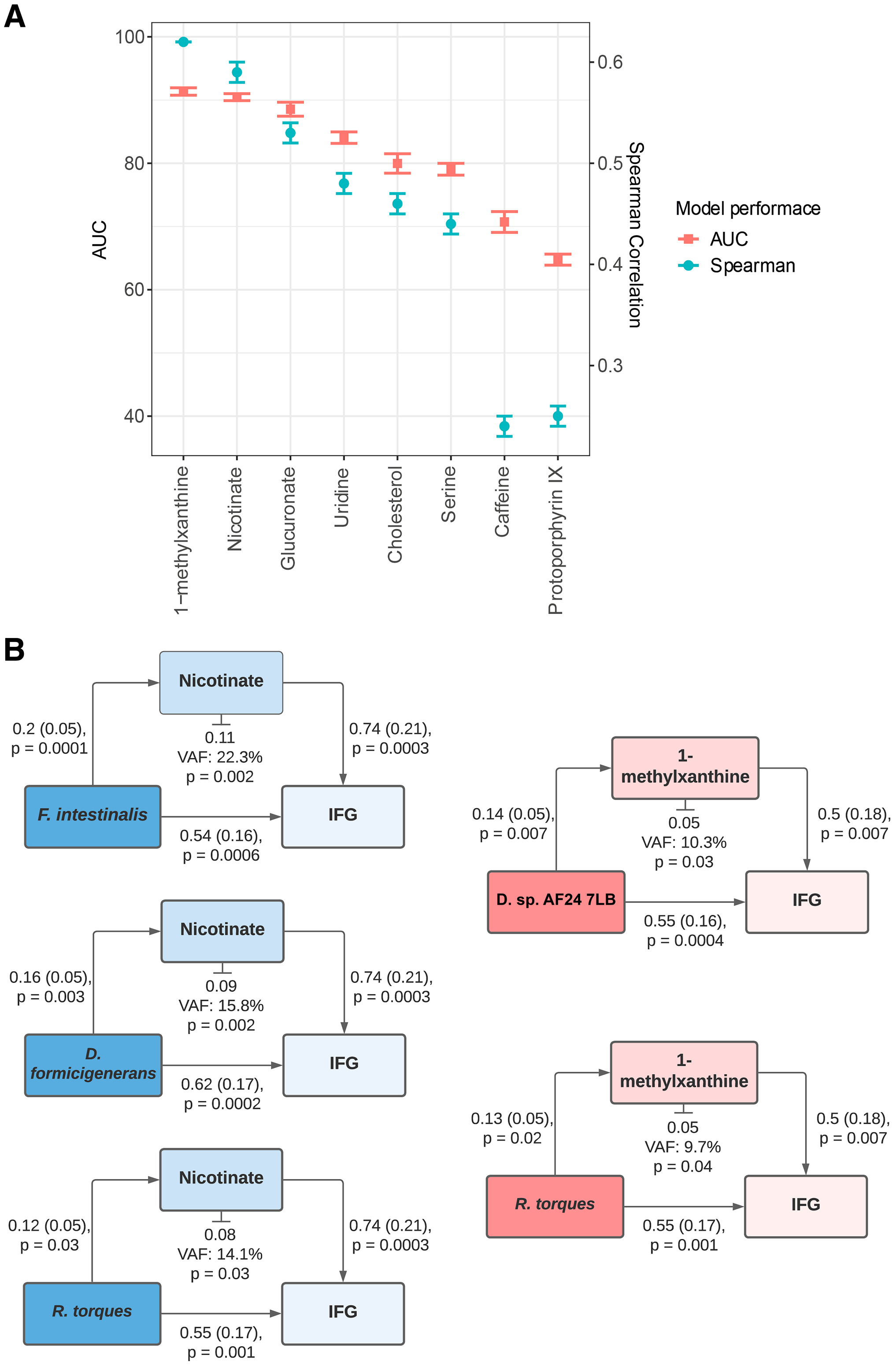


Figure 4—Associations of the gut microbiota with the eight fecal replicated metabolites and IFG in 342 TwinsUK participants. *A*: Influence of the gut microbiota composition in the fecal abundances of the eight replicated metabolites estimated by random forest regressors (Spearman correlations between the real value of each metabolite and the value predicted) and classifiers (AUC). Red and blue bars represent the mean AUC and Spearman correlations with the respective 95% CIs across fivefolds, respectively. *B*: Mediation analyses of the associations between characterized gut bacterial species and IFG. Models were adjusted for age, BMI, and sex. Path coefficients are shown beside each path, and indirect effects and VAF score are indicated below each mediator (left: nicotinate, right: 1-methylxanthine). Only metabolites with a predictive power of AUC >90% in *A* are shown.

levels, and smoking status) as these were not significantly associated with the identified bacterial species or with the metabolites making up the score (Supplementary Table 5).

DISCUSSION

Here we identify for the first time a fecal metabolite signature of IFG that is associated with prevalent IFG in two independent cohorts and is also predictive of incident T2D in a small subanalysis. The fecal metabolites making up the score are not microbial-derived metabolites but are "host metabolites" (e.g., xenobiotics, cofactors, and vitamins). However, the gut microbiome can accurately predict their fecal abundances (AUC >70%). It is well known that the gut microbiome composition can affect diseases via several mechanisms (22). Circulating microbial metabolites have been reported by us and others to be reflective of gut microbiome diversity and composition (6–8) and predictive of prevalent and incident T2D (7). Taken together, this suggests that the gut microbiome can influence T2D, not only by producing metabolites that enter the bloodstream (7) but also by regulating the absorption or excretion of host-produced compounds, thereby influencing IFG and T2D risk. This hypothesis is further supported by the results of our mediation analysis showing that metabolites making up the score act as partial mediators on the significant associations between several gut microbial species, (e.g., *F. intestinalis*, *D. formicigenerans*, *R. torques*, and *Dorea* sp. AF24-7LB) and IFG.

Studies have shown that gut microbiome composition differs between individuals with prediabetes/diabetes and healthy subjects (6,7), with compositional shifts correlated with synthesis profile changes of gut bacteria-derived metabolites, including short-chain fatty acids, indolepropionic acid, and trimethylamine (7,22). These "microbial" metabolites enter into the bloodstream and reach different tissues, where they can influence glucose homeostasis and insulin resistance by activating or inhibiting signaling pathways (22). Nevertheless, the identified signature of prediabetes in this study consists of eight metabolites of nonbacterial origin. Serine is a nonessential amino acid mainly obtained by intrinsic synthesis (23). Glucuronate is a sugar acid derived from glucose and involved in the detoxification of xenobiotic compounds (24). Protoporphyrin IX is a cofactor ubiquitously present in the human body as a heme precursor (25). Nicotinate, also known as vitamin B_3 and niacin, is a water-soluble vitamin that can be produced by the human body from tryptophan (26). Cholesterol, which is mainly produced by the liver, is an essential lipid of eukaryotic cell membranes and is also a precursor of bile acids and steroid hormones (27). Uridine is a necessary pyrimidine nucleotide for RNA synthesis produced by several reversible reactions (e.g., dephosphorylation of uridine monophosphate, deamination of a cytidine or combination of uracil and ribose 1-phosphate) (28). Caffeine and 1-methylxanthine are xenobiotics involved in the caffeine metabolism pathway (29).

Strikingly, we find that the gut microbiome is strongly associated with fecal levels of these metabolites, suggesting that the gut microbiome influences the absorption or excretion of compounds involved in various metabolic pathways (e.g., cholesterol, uridine, and glucuronate) and xenobiotics (e.g., caffeine and its derivatives), among others, and such levels of absorption or excretion are directly related to IFG. Our findings lead us to speculate that individuals with prediabetes present gut microbiome composition perturbations, which likewise influence the absorption or excretion of the identified compounds. This is further supported by the mediation analyses, which suggest that the associations between specific gut microbial species, including *F. intestinalis*, *D. formicigenerans*, *R. torques*, and *Dorea* sp. AF24-7LB, and IFG are mainly reflecting the effect of the gut microbiome in the absorption or excretion of the found compounds.

Under normal conditions, the small intestine can break down, emulsify, and absorb most nutrients, including fats, simple carbohydrates, and proteins (30). For instance, <5 g/day of fat are not absorbed and reach the colon (30). Nonetheless, the absorption capability of the gut can be limited depending on the gut microbiome composition (31). A study conducted by Basolo et al. (31) demonstrated that changes in participants' gut microbiome composition, due to diet or antibiotic use, impaired nutrient absorption. Several mechanisms might explain how gut microbiome composition might influence absorption, and thus, the disease onset (32–34). For instance, the gut microbiome can affect the gut barrier, which consists of a collection of physical and chemical structures that protect the host from pathogenic invasions and harmful stimuli (32). This can be provoked by the presence of pathogen-associated molecular patterns, such as lipopolysaccharides, in the cell walls of some gram-negative bacteria, which play an important role in intestinal absorption, blood glucose, and inflammation (33). Moreover, changes in the permeability of the gut barrier can be caused by an unbalanced increase in bacteria able to degrade mucin (the main component of mucus, which covers the epithelial surfaces of the gastrointestinal tract) (32). Indeed, in this study, we identify that individuals with prediabetes present larger abundances compared with healthy individuals of the mucin-degraders *D. formicigenerans* (35) and *R. torques* (36), which have been previously associated with lower nutrient absorption (36). Finally, some gut microbes can also reduce absorption in the jejunum by altering the expression of intestinal transporters of different types of compounds (34).

Another possible explanation for our findings could be a reduction of specific beneficial bacteria able to use these compounds, thus resulting in increased excretion (27,37). In the case of cholesterol, bacterial members of the genera *Bifidobacterium*, *Lactobacillus*, and *Peptostreptococcus* are needed to convert cholesterol into coprostanol (27). Likewise, an inefficient cholesterol-coprostanol conversion is linked to cardiometabolic diseases (27). For glucuronate, most of it is not absorbed by the small intestine; however, under normal conditions, the amounts that make it to the colon are then efficiently used by *Bifidobacterium* (37).

This work has several strengths. Our study benefits from a large, accurately phenotyped discovery cohort with metabolomic profiling and gut microbiome composition. We were also able to replicate our findings in a large independent cohort, thus strengthening our findings. Finally, a machine learning algorithm was applied to investigate the prediction of the gut microbiota to the levels of the found eight metabolites, allowing us to simultaneously integrate all the species in the models.

We also note some study limitations. First, the cross-sectional nature of the data used for our primary analysis does not allow us to determine the temporal link between IFG and the identified fecal metabolites.

Second, HbA_{1c}, postprandial glucose to derive impaired glucose tolerance, which more closely resembles the T2D state (38), and a clinician's diagnosis were not available in the discovery cohort. Thus, the division of categories in this study is derived from IFG.

Third, the sample size for the subanalysis looking at incident T2D was limited, and we were unable to seek independent replication as, to the best of our knowledge, there are no other cohorts in the world that have measured this fecal metabolome panel and incident T2D. Future studies with larger sample size are therefore needed to test the robustness of the IFG metabolite score to predict incident T2D.

Fourth, there was not a full overlap between metabolites measured in the discovery and validation data sets, which might cause the loss of metabolites of interest to study.

Fifth, the included study groups were unbalanced in age and sex. Hence, although we adjusted all analyses for them and other important clinical variants, the confidence of the results is lowered. In addition, gut microbiota composition data were only available for a subset from the discovery set, and therefore, we could not replicate the mediation analysis in KORA. Furthermore, the Spearman correlations between the predicted (from gut microbiome composition) and actual levels of the metabolites were modest. Indeed, random forest models were trained based on microbial features extracted from metagenomic data, which does not retrieve all species present in a microbiome sample for procedural and technical reasons.

Finally, this study does not include measures of permeability markers, which would contribute to a better understanding of the role of intestinal permeability in the absorption or excretion of the identified compounds.

In conclusion, we are proposing a novel mechanism of how gut microbiome composition affects prediabetes and, consequently, the onset of T2D. The gut microbiome is linked to prediabetes not only by microbial-derived metabolites but also by affecting intestinal absorption or excretion of metabolites of nonmicrobial origin, which are correlated with the risk of IFG and incident T2D. Henceforth, to better understand the onset of T2D, the effect of the gut microbiome in the excretion and/or absorption of host-produced compounds and xenobiotics also needs to be also considered.

Acknowledgments. The authors thank all the participants of TwinsUK for contributing and supporting their research and all participants of KORA for their long-term commitment to the KORA study, the staff for data collection and research data management, and the members of the KORA Study Group (https://www.helmholtz-munich.de/en/epi/cohort/kora) who are responsible for the design and conduct of the study. Finally, the authors thank Caroline Le Roy (Department of Twin Research & Genetic Epidemiology, King's College, London) for the helpful discussion. Data collection in the KORA study is done in cooperation with the University Hospital of Augsburg.

Funding. This research was funded by the Chronic Disease Research Foundation, and in part by the Wellcome Trust (grant no. 212904/Z/18/Z) and by DiabetesUK (19/0006053). TwinsUK receives funding from the Wellcome Trust, the European Commission H2020 grants SYSCID (contract #733100), the National Institute for Health Research (NIHR) Clinical Research Facility, and the Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London, the Chronic Disease Research Foundation, the UKRI Medical Research Council (MRC)/British Heart Foundation Ancestry and Biological Informative Markers for Stratification of Hypertension (AIM-HY, MR/M016560/1). This work was also supported by UKRI grant MR/W026813/1 to C.M. and A.M.V. The KORA study was initiated and financed by the Helmholtz Zentrum München–German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Stool sample collection and metabolomics analysis in KORA FF4 was supported by iMED, a research alliance within the Helmholtz Association, Germany. C.M., L.P., and A.N. are funded by the Chronic Disease Research Foundation. A.M.V. is supported by the National Institute for Health Research Nottingham Biomedical Research Centre.

Duality of Interest. This study received support from Zoe Ltd. T.D.S. is a cofounder and shareholder of Zoe Ltd. A.M.V., P.W.F., F.A., and N.S. are consultants to Zoe Ltd. K.W. and G.A.M. are employees of Metabolon Inc. No other potential conflicts of interest relevant to this article were reported.

Author Contributions. A.N., F.T., Q.D., A.V., C.C., H.G., F.A., and C.G. curated the data. A.N., F.T., Q.D., and C.M. performed the formal analyses. A.N., A.M.V., and C.M. wrote the manuscript. Q.D., P.L., A.V., C.C., T.B., J.L., H.G., N.W., F.A., K.W., A.-F.B., G.A.M., N.S., M.F., A.P., P.W.F., V.B., T.D.S., J.T.B., and C.G. contributed reagents/materials/analysis tools. T.D.S. and C.M. conceived and designed the experiments. All of the authors revised the manuscript. C.M. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

References

1. Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nat Rev Endocrinol 2018;14:88–98
2. Sun H, Saeedi P, Karuranga S, et al. IDF Diabetes Atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract 2022;183:109119
3. Kolb H, Martin S. Environmental/lifestyle factors in the pathogenesis and prevention of type 2 diabetes. BMC Med 2017;15:131
4. Knowler WC, Barrett-Connor E, Fowler SE, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med 2002;346:393–403
5. Elliott TL, Pfotenhauer KM. Classification and diagnosis of diabetes. Prim Care 2022;49:191–200
6. Aydin Ö, Nieuwdorp M, Gerdes V. The gut microbiome as a target for the treatment of type 2 diabetes. Curr Diab Rep 2018;18:55

7. Menni C, Zhu J, Le Roy CI, et al. Serum metabolites reflecting gut microbiome alpha diversity predict type 2 diabetes. Gut Microbes 2020;11: 1632–1642

8. Zhang Z, Tian T, Chen Z, Liu L, Luo T, Dai J. Characteristics of the gut microbiome in patients with prediabetes and type 2 diabetes. PeerJ 2021; 9:e10952

9. Maurice CF, Turnbaugh PJ. Quantifying the metabolic activities of human-associated microbial communities across multiple ecological scales. FEMS Microbiol Rev 2013;37:830–848

10. Krautkramer KA, Fan J, Bäckhed F. Gut microbial metabolites as multi-kingdom intermediates. Nat Rev Microbiol 2021;19:77–94

11. Visconti A, Le Roy CI, Rosa F, et al. Interplay between the human gut microbiome and host metabolism. Nat Commun 2019;10:4505

12. Zierer J, Jackson MA, Kastenmüller G, et al. The fecal metabolome as a functional readout of the gut microbiome. Nat Genet 2018;50:790–795

13. Conlon MA, Bird AR. The impact of diet and lifestyle on gut microbiota and human health. Nutrients 2014;7:17–44

14. Verdi S, Abbasian G, Bowyer RCE, et al. TwinsUK: the UK adult twin registry update. Twin Res Hum Genet 2019;22:523–529

15. ElSayed NA, Aleppo G, Aroda VR, et al.; American Diabetes Association. 2. Classification and diagnosis of diabetes: *Standards of Care in Diabetes—2023*. Diabetes Care 2023;46(Suppl. 1):S19–S40

16. Donia MS, Fischbach MA. Human microbiota. Small molecules from the human microbiota. Science 2015;349:1254766

17. Pasolli E, Asnicar F, Manara S, et al. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. Cell 2019;176:649–662.e620

18. Thissen D, Steinberg L, Kuang D. Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. J Educ Behav Stat 2002;27:77–83

19. Tugwell P, Knottnerus JA. A statistic to avoid being misled by the "winners curse". J Clin Epidemiol 2018;103:vi–viii

20. Kuhn M. Building predictive models in R using the caret package. J Stat Softw 2008;28:1–26

21. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. mediation: R package for causal mediation analysis. J Stat Softw 2014;59:1–38

22. Nogal A, Valdes AM, Menni C. The role of short-chain fatty acids in the interplay between gut microbiota and diet in cardio-metabolic health. Gut Microbes 2021;13:1–24

23. Jiang J, Li B, He W, Huang C. Dietary serine supplementation: friend or foe? Curr Opin Pharmacol 2021;61:12–20

24. Fujiwara R, Yoda E, Tukey RH. Species differences in drug glucuronidation: Humanized UDP-glucuronosyltransferase 1 mice and their application for predicting drug glucuronidation and drug-induced toxicity in humans. Drug Metab Pharmacokinet 2018;33:9–16

25. Sachar M, Anderson KE, Ma X. Protoporphyrin IX: the good, the bad, and the ugly. J Pharmacol Exp Ther 2016;356:267–275

26. Moffett JR, Namboodiri MA. Tryptophan and the immune response. Immunol Cell Biol 2003;81:247–265

27. Kriaa A, Bourgin M, Potiron A, et al. Microbial impact on cholesterol and bile acid metabolism: current status and future prospects. J Lipid Res 2019;60: 323–332

28. Urasaki Y, Pizzorno G, Le TT. Uridine affects liver protein glycosylation, insulin signaling, and heme biosynthesis. PLoS One 2014;9:e99728

29. Keijzers GB, De Galan BE, Tack CJ, Smits P. Caffeine can decrease insulin sensitivity in humans. Diabetes Care 2002;25:364–369

30. de Vos WM, Tilg H, Van Hul M, Cani PD. Gut microbiome and health: mechanistic insights. Gut 2022;71:1020–1032

31. Basolo A, Hohenadel M, Ang QY, et al. Effects of underfeeding and oral vancomycin on gut microbiome and nutrient absorption in humans. Nat Med 2020;26:589–598

32. Raimondi S, Musmeci E, Candeliere F, Amaretti A, Rossi M. Identification of mucin degraders of the human gut microbiota. Sci Rep 2021;11:11094

33. Anhê FF, Barra NG, Cavallari JF, Henriksbo BD, Schertzer JD. Metabolic endotoxemia is dictated by the type of lipopolysaccharide. Cell Rep 2021;36:109691

34. Depommier C, Van Hul M, Everard A, Delzenne NM, De Vos WM, Cani PD. Pasteurized *Akkermansia muciniphila* increases whole-body energy expenditure and fecal energy excretion in diet-induced obese mice. Gut Microbes 2020;11:1231–1245

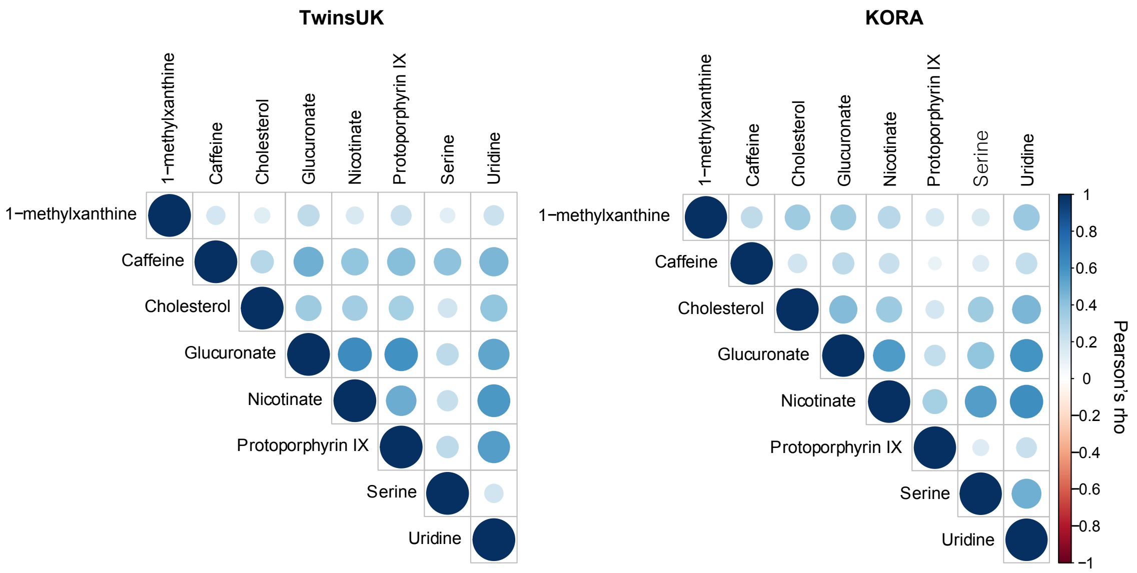
35. Vacca M, Celano G, Calabrese FM, Portincasa P, Gobbetti M, De Angelis M. The controversial role of human gut lachnospiraceae. Microorganisms 2020; 8:573

36. Kaczmarczyk M, Löber U, Adamek K, et al. The gut microbiota is associated with the small intestinal paracellular permeability and the development of the immune system in healthy children during the first two years of life. J Transl Med 2021;19:177

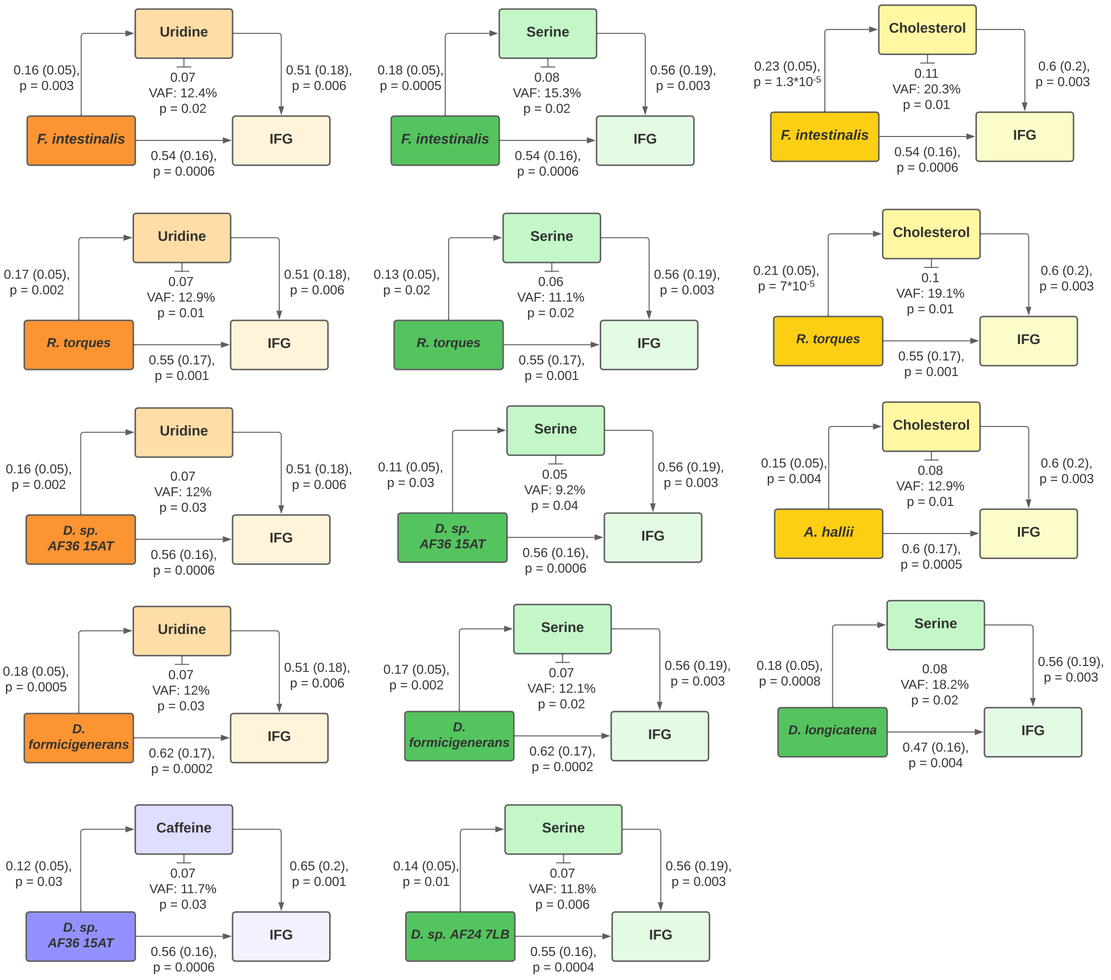
37. Asano T, Yuasa K, Kunugita K, Teraji T, Mitsuoka T. Effects of gluconic acid on human faecal bacteria. Microb Ecol Health Dis 1994;7:247–256

38. Unwin N, Shaw J, Zimmet P, Alberti KG. Impaired glucose tolerance and impaired fasting glycaemia: the current status on definition and intervention. Diabet Med 2002;19:708–723

Supplementary material



Supplementary Fig. 6.1 Pearson's correlation matrix calculated from the abundances of the 8 faecal replicated metabolites in TwinsUK (n=1247) and KORA (n=1007). The shown correlations were significant (p-value<0.05).



Supplementary Fig. 6.2 Mediation analyses of the associations between characterised gut bacterial species and IFG. Path coefficients are shown beside each path, and indirect effects and variance accounted for (VAF) score are indicated below each mediator (metabolite). Only metabolites with a predictive power of 70%>AUC<90% are shown. Glucuronate did not present a mediatory role with any species.

Supplementary Table 6.1 Complete list of the 526 included metabolites in TwinsUK measured by Metabolon Inc. with their super-pathways, sub-pathways, and KEGG and HMDB identifiers. From these, the metabolites with measurements available for KORA participants are indicated.

Large table. Access to the table is given in the attached OneDrive. A list of the entire OneDrive content is listed in ***Appendix A****.*

Supplementary Table 6.2 Descriptive characteristics of the subset of 342 individuals from TwinsUK with concurrent gut microbiota composition and faecal metabolites measurements. The p-value from a Wilcoxon test (continuous variable) or chi-squared test (categorical variable) was calculated to check whether differences between the different subject groups for the described parameters existed.

	Total	Healthy individuals	IFG individuals	Differences between groups (p-value)
N	342	297	45	-
Females, %	83.9	86.2	68.9	0.006
Age, yrs	56 (16.6)	54.3 (16.8)	67.4 (9.3)	<0.0001
BMI, kg/m2	25.6 (5)	25 (4.6)	29.4 (6.1)	<0.0001
Fasting glucose, mmol/L	4.7 (0.5)	4.5 (0.2)	5.8 (0.3)	<0.0001
SBP, mmHg	126.7 (13.7)	126 (12.6)	134 (17.7)	0.002
DBP, mmHg	74.2 (7.5)	73.8 (6.8)	76.9 (10.8)	0.07
Circulating total cholesterol, mmol/L	4.1 (0.5)	4.1 (0.5)	4 (0.7)	0.25
Circulating HDL, mmol/L	1.87 (1.3)	1.91 (1.3)	1.6 (0.8)	0.15
Circulating triglycerides, mmol/L	0.94 (1.1)	0.91 (1.1)	1.11 (1)	0.03
aHEI	70.6 (5.7)	70.6 (5.6)	70.8 (6.3)	0.88
Current Smoker	No: 331 Yes: 11	No: 286 Yes: 11	No: 45	0.39
Activity level	Low: 25 Moderate: 259 High: 58	Low: 22 Moderate: 225 High: 50	Low: 3 Moderate: 34 High: 8	0.98

Supplementary Table 6.3 Influence of the gut microbiota composition in the faecal abundances of the 8 replicated metabolites in 342 participants from TwinsUK. Influence estimated by regression (using Spearman's correlations) and classification (using AUC values) Random Forest models.

Metabolite name	AUC (%) [95% CI]	Spearman's rho [95% CI]
Protoporphyrin IX	64.8 [63.9,65.6]	0.25 [0.24,0.26]
Caffeine	70.7 [69.1,72.4]	0.24 [0.23,0.25]
Serine	79.1 [78.1,80]	0.44 [0.43,0.45]
Cholesterol	80 [78.4,81.5]	0.46 [0.45,0.47]
Uridine	84.1 [83.1,85]	0.48 [0.47,0.49]
Glucuronate	88.6 [87.4,89.7]	0.53 [0.52,0.54]
Nicotinate	90.5 [89.9,91]	0.59 [0.58,0.6]
1-methylxanthine	91.4 [90.8,91.9]	0.62 [0.62,0.62]

Supplementary Table 6.4 Associations between the gut microbiota composition and impaired fasting glucose (IFG). Specifically, the top 100 features from the Random Forest models predicting the faecal metabolite abundances from the gut microbiome composition with an AUC>70% are shown. The linear regression models were adjusted for age, BMI, sex and multiple testing (false discovery rate – FDR). The prevalence of each gut bacteria is also indicated.

Large table. Access to the table is given in the attached OneDrive. A list of the entire OneDrive content is listed in ***Appendix A****.*

Supplementary Table 6.5 Associations of comorbidities with the 8 metabolites making up the score and the bacterial species involved in the mediation analyses. Pearson's correlations run for the continuous comorbidities (systolic and diastolic blood pressure, circulating HDL, total cholesterol and triglycerides levels, and aHEI) whereas a two-proportion z-test was used for the categorical comorbidities (activity level and smoking status).

Large table. Access to the table is given in the attached OneDrive. A list of the entire OneDrive content is listed in ***Appendix A****.*

Supplementary Table 6.6 List of gut species represented using species-level genome bins (SGBs) that were profiled in 342 participants from TwinsUK. Prevalence and if the composition of a species presents variance zero and/or near zero are indicated.

Large table. Access to the table is given in the attached OneDrive. A list of the entire OneDrive content is listed in ***Appendix A****.*

Supplementary Text. Methodology details.

Metabolomics profiling

Metabolite concentrations were measured from faecal samples by Metabolon Inc. (Durham, USA) using an untargeted LC-MS platform. All samples were maintained at -80°C until processing. As a means of quality control, several recovery standards were added prior to the first step in the extraction process. Briefly, to remove protein, dissociate small molecules bound to proteins or trapped within the precipitated protein matrix, and to recover chemically diverse metabolites, proteins were precipitated in methanol and vigorously shaken for 2 minutes (Glen Mills GenoGrinder 2000), then centrifuged. The resulting extract was divided into five fractions; both aliquots (i) and (ii) were analysed using acidic positive ion conditions and chromatographically optimised for hydrophilic and hydrophobic compounds respectively, aliquot (iii) was analysed using basic negative ion optimised conditions using a dedicated separate dedicated C18 column, aliquot (iv) was analysed using negative ionisation following elution from a hydrophilic interaction liquid chromatography column, while aliquot (v) was reserved as a back-up.

Several controls were analysed in concert with experimental samples. (i) a pooled sample generated from a small volume of each experimental sample of interest served as a technical replicate throughout the platform run; (ii) extracted water samples served as process blanks; (iii) and a cocktail of standards, known not to interfere with measurements, spiked into every analysed sample facilitated instrument performance monitoring and aided chromatographic alignment. Instrument variability was determined by calculating the median relative standard deviation (RSD) for the standards that were added to each sample prior to injection into the mass spectrometers. Overall process variability was determined by calculating the median RSD for all endogenous metabolites (i.e., non-instrument standards) present in 100% or more of the pooled technical replicate samples. Experimental samples and controls were randomised across the platform run.

Compound identification

Metabolites were identified by comparison of the ion features in the experimental samples to a reference library of chemical standard entries that included retention time/index, molecular weight (m/z), and MS spectra. Identification of known chemical entities is based on comparison across all 3 features to metabolomic library entries of purified standards. More than 3300 commercially available purified standard compounds have been acquired and registered into the library, while additional mass spectral entries have been created for structurally unnamed biochemicals, which have been identified by virtue of their recurrent nature (both chromatographic and mass spectral). These compounds have the potential to be identified by future acquisition of a matching purified standard or by classical structural analysis.

Metabolite quantification and normalisation

Peaks were quantified using area-under-the-curve. Raw area counts for each metabolite in each sample were normalised to correct for variation resulting from instrument inter-day tuning differences by the median value for each run-day, therefore, setting the medians to 1.0 for each run. This preserved variation between samples but allowed metabolites of widely different raw peak areas to be compared on a similar graphical scale.

Metagenomic assessment in TwinsUK

Faecal sample collection

Participants collected stool samples at home in pre-labelled kits (containing 2 x 25ml tube or 1 x 25ml tube and 1 x 10ml Zymo buffer), which were posted to them before their clinic visit date and brought with them to the visit. In the laboratory, samples were homogenised, aliquoted into 4 bijou tubes, and stored at 80°C, within 2 hours of receipt.

DNA extraction, library preparation, and sequencing

To isolate genomic DNA from faecal material, bijou tubes were removed from the freezer and grounded with glass beads and 5-6ml distilled water (Spex Grinder, 10 seconds, 800 strokes per minute). The supernatant was centrifuged and further grounded (5 minutes, 1000 strokes per minute) before 200-300µl of the sample was mixed with 10µl PK solution and 720µl of Lysis/Bind Master Mix). Proteins were degraded by the binding solution and subsequently extracted by KingFisher Flex robot. DNA was washed in 2 steps using washing solutions and eluted in MagMax Core Elution Buffer in 100µl. Library preparation and sequencing were performed by GenomeScan.

Metagenome quality control and preprocessing

Sequenced metagenomes were processed using the YAMP pipeline (v. 0.9.5.3). Briefly, identical reads were removed. Reads were filtered to remove adapters, known artefacts, phix174, and then quality trimmed (PhRED quality score<10). Reads that became too short after trimming (N<60bp) were discarded. We retained singleton reads (i.e., reads whose mate has been discarded) to retain as much information as possible. Contaminant reads belonging to the host genome were removed (build: GRCh37), and low-quality samples (i.e., samples with <10M reads after QC) were discarded.

Microbiome taxonomic profiling

The metagenomic analysis was conducted following the general guidelines and based on the bioBakery computational environment. High-resolution taxonomic profiling of the metagenomes was performed using MetaPhlAn 4.beta.2 with the January 2021 database and default parameters.

Statistical analysis

We run random forest regression (1000 trees and a third of features number as the number of variables randomly sampled as candidates at each split) and classification models (1000 trees and the square root of features number as the number of variables randomly sampled as candidates at each split) with compositional data using 5-folds cross-validation. Before running the models, gut microbiota variables with variance zero or near to zero were excluded using the nearZeroVar function implemented in R in the caret package (the included/excluded SGBs are shown in **Supplementary Table 6.6**. For the classifiers, the continuous response was converted into two classes based on the top and bottom quartiles. The features were ranked based on node purity.

Chapter 7

Genetic and gut microbiome determinants of SCFA levels, and their links with inflammatory responses

Modulating SCFA levels could be a potential target to treat CMD as discussed in the introduction. It is thus crucial to understand whether circulating (fasting and postprandial) and faecal SCFA levels are potentially modifiable and the gut microbiome contribution. SCFAs might positively influence inflammatory responses, and in turn CMD. However, their role in acute inflammatory responses is still unknown.

In this chapter, I comprehensively assess the host genetics and gut microbiota contribution to a panel of eight SCFAs measured in serum and stool, examined their postprandial changes, and explored their links with chronic and acute inflammatory responses.

The findings illustrate that SCFA levels might be modifiable, the gut microbiome is mainly predictive of their faecal level, and circulating levels change postprandially. Finally, the obtained results indicate that SCFAs might play a key role in chronic and acute inflammatory responses.

Andrei-Florin Balean processed the stool samples for the gut microbiome sequencing. Dr Alessia Visconti and Dr Francesco Asnicar generated and performed the QC of the shotgun metagenome data in TwinsUK and ZOE PREDICT-1, respectively. I performed the QC of the SCFA data, ran the statistical analyses, with the assistance of Dr Massimo Mangino for estimating heritability, and wrote the original draft of the manuscript.

This chapter has been published in *Gut microbes* (Nogal et al., 2023).

GUT MICROBES
2023, VOL. 15, NO. 1, 2240050
https://doi.org/10.1080/19490976.2023.2240050

Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

OPEN ACCESS Check for updates

Genetic and gut microbiome determinants of SCFA circulating and fecal levels, postprandial responses and links to chronic and acute inflammation

Ana Nogal[a], Francesco Asnicar[b], Amrita Vijay[c], Afroditi Kouraki[c], Alessia Visconti[a], Panayiotis Louca[a], Kari Wong[d], Andrei-Florin Baleanu[a], Francesca Giordano[e], Jonathan Wolf[e], George Hadjigeorgiou[e], Richard Davies[e], Gregory A. Michelotti[d], Paul W. Franks[f,g], Sarah E. Berry[h], Mario Falchi[a], Adeel Ikram[c], Benjamin J. Ollivere[c], Amy Zheng[c], Jessica Nightingale[c], Massimo Mangino[a,i], Nicola Segata[b], William J. Bulsiewicz[e], Tim D. Spector[a], Ana M. Valdes[c*], and Cristina Menni[a*]

[a]Department of Twin Research, King's College London, London, UK; [b]Department of Cellular, Computational and Integrative Biology, University of Trento, Trento, Italy; [c]Nottingham NIHR Biomedical Research Centre at the School of Medicine, University of Nottingham, Nottingham, UK; [d]Metabolon, Metabolon, Inc. Research Triangle Park, Morrisville, NC, USA; [e]Zoe Limited, London, UK; [f]Lund University Diabetes Center, Lund University, Malmö, Sweden; [g]Department of Clinical Sciences, Lund University, Malmö, Sweden; [h]Department of Nutritional Sciences, King's College London, London, UK; [i]NIHR Biomedical Research Centre at Guy's and St Thomas' Foundation Trust, London, UK

ABSTRACT

Short-chain fatty acids (SCFA) are involved in immune system and inflammatory responses. We comprehensively assessed the host genetic and gut microbial contribution to a panel of eight serum and stool SCFAs in two cohorts (TwinsUK, $n = 2507$; ZOE PREDICT-1, $n = 328$), examined their postprandial changes and explored their links with chronic and acute inflammatory responses in healthy individuals and trauma patients. We report low concordance between circulating and fecal SCFAs, significant postprandial changes in most circulating SCFAs, and a heritable genetic component (average h^2: serum = 14%(SD = 14%); stool = 12%(SD = 6%)). Furthermore, we find that gut microbiome can accurately predict their fecal levels (AUC>0.71) while presenting weaker associations with serum. Finally, we report different correlation patterns with inflammatory markers depending on the type of inflammatory response (chronic or acute trauma). Our results illustrate the breadth of the physiological relevance of SCFAs on human inflammatory and metabolic responses highlighting the need for a deeper understanding of this important class of molecules.

ARTICLE HISTORY
Received 16 December 2022
Revised 20 April 2023
Accepted 19 July 2023

KEYWORDS
Short-chain fatty acids; postprandial; heritability; gut microbiota; inflammatory response

Introduction

Short-chain fatty acids (SCFA) are carboxylic acids formed by an aliphatic chain of 1–5 carbons[1] mainly produced by colonic bacteria through the saccharolytic fermentation of resistant carbohydrates such as inulin, resistant starch and fructo-oligosaccharides, which escape digestion and absorption.[2] The major formed SCFAs are acetate, propionate and butyrate, which account for approximately 80% of all SCFAs.[3]

Once produced, SCFAs can either be absorbed by the enterocytes or go into the bloodstream and reach different systemic tissues, exerting regulatory functions in gut barrier integrity, lipid and glucose metabolism, blood pressure, and immune function.[3] Several studies have shown that SCFAs can also influence postprandial responses including postprandial glucose and insulin.[4,5] For instance, a host-genetic-driven increase in gut production of butyrate was associated with improved postprandial insulin response.[4]

SCFAs can also exert anti-inflammatory effects, influencing chronic inflammation, by reducing the recruitment and migration of macrophages, dendritic cells, and neutrophils, and by altering T and B cell differentiation.[6] Previously, levels of fecal SCFAs have been reported to be associated with mortality in critically ill patients with sepsis.[7,8]

CONTACT Ana M Valdes ana.valdes@nottingham.ac.uk Nottingham NIHR Biomedical Research Centre at the School of Medicine, University of Nottingham, Nottingham NG5 1PB, UK; Cristina Menni cristina.menni@kcl.ac.uk Department of Twin Research, King's College London, St. Thomas' Hospital Campus, London SE1 7EH, UK

*Equal contribution.

Supplemental data for this article can be accessed online at https://doi.org/10.1080/19490976.2023.2240050

However, the role of SCFAs in blood in acute inflammatory responses, such as those seen in acute trauma cases, has not been explored to date.

In human studies, SCFA levels have been associated both with gut bacterial species, including *Coprococcus*, *Bifidobacterium* and *Roseburia*,[3,9] and with host genetics.[4] Nonetheless, these studies do not simultaneously integrate SCFAs measured in both serum and stool, along with concurrent gut microbiome composition and genetic information. They also include only a subset of SCFAs available and focus on circulating fasting levels. Indeed, though humans spend most of their days in a postprandial state, postprandial SCFAs responses have only been investigated in animal models.[10,11]

By integrating data from a large population-based cohort and a postprandial interventional study, using healthy individuals we aimed to investigate (i) changes in SCFA levels after a meal challenge and (ii) the contribution of the host genetics and gut microbiome composition to their levels in serum and stool. We further aimed to understand the role of circulating SCFAs in chronic and acute inflammation by assessing their correlations with a set of pro- and anti-inflammatory markers in healthy individuals and in an acute fracture case-control study, as well as changes in their levels in response to acute inflammatory responses to trauma.

Results

A flowchart of the study design is presented in Figure 1.

We included 2507 individuals (age mean = 57.9 ± 15.4 years, 2110 females) from the TwinsUK cohort and 328 females (age mean = 53.8 ± 7.1 years) from the ZOE PREDICT-1 cohort that had eight SCFAs, including acetate, propionate and butyrate, measured in stool and in fasting serum. Additionally, postprandial serum SCFAs were measured in ZOE PREDICT-1 participants. Both cohorts included twins. The demographic characteristics of the study populations and the mean baseline levels of the SCFAs in serum and stool are presented in Table 1.

The correlations between circulating and fecal SCFAs calculated in TwinsUK and ZOE PREDICT-1 separately are presented in Figure 2a. We found non-significant or low correlations ($\rho < \pm 0.15$) between the circulating SCFAs with their respective fecal SCFAs in both studies. We also detected a low correlation between serum and fecal SCFAs and age and body mass index (BMI) ($\rho < \pm 0.2$) (Supplementary Figure S1).

Postprandial changes in SCFA levels

As individuals spend most of the day in a postprandial state (i.e., not fasting), we investigated whether circulating SCFA levels change after a standardized meal and their inter-individual variability. For that, we took advantage of the ZOE PREDICT-1 cohort, in which SCFA levels were measured in a tightly controlled clinic setting at fasting, 30 min, 2 h, and 4 h after a meal challenge. Each participant consumed a standardized muffin, as described in the Methods section, that

Table 1. Demographic characteristics and SCFA levels of the study populations TwinsUK and ZOE PREDICT-1.

Study	TwinsUK		ZOE PREDICT-1	
n	2507		328	
Females, (%)	84%		100%	
Age, yrs	57.9 (15.4)		53.8 (7.1)	
BMI, kg/m^2	25.87 (5)		26.24 (5.6)	
MZ:DZ:Singlet ons	1376:776:355		164:50:114	
SCFA	Fasting serum (ng/ml)	Stool (μg/g)	Fasting serum (ng/ml)	Stool (μg/g)
Acetate	3030 (1600)	4060 (1700)	3500 (1620)	4020 (1710)
Propionate	83.3 (29)	1320 (624)	84.9 (32.3)	1340 (676)
Butyrate	40.4 (19.5)	1420 (893)	41.4 (20.0)	1380 (924)
Isobutyrate	72.2 (24.3)	202 (87.7)	90.0 (27.1)	220 (88.8)
Methylbutyrate	117 (42.6)	157 (77)	134 (48.9)	170 (72.6)
Valerate	8.57 (4.47)	254 (119)	7.34 (4.76)	287 (129)
Isovalerate	78.7 (30)	193 (91.9)	93.2 (29.8)	212 (92.7)
Hexanoate	47.2 (17.2)	105 (114)	45.3 (20.1)	130 (121)

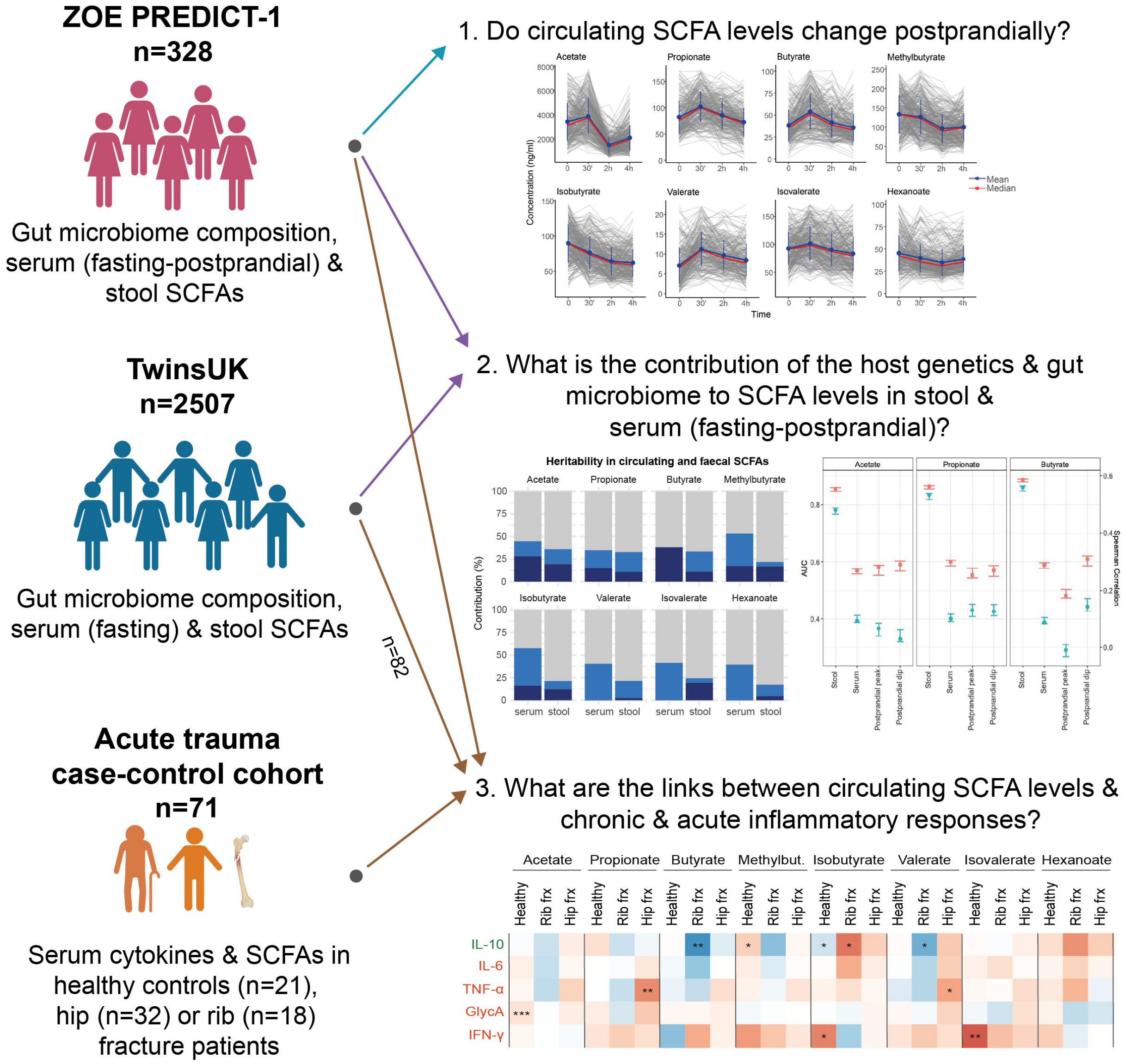

ZOE PREDICT-1
n=328
Gut microbiome composition, serum (fasting-postprandial) & stool SCFAs
TwinsUK
n=2507
Gut microbiome composition, serum (fasting) & stool SCFAs
n=82
Acute trauma case-control cohort
n=71
Serum cytokines & SCFAs in healthy controls (n=21), hip (n=32) or rib (n=18) fracture patients
1. Do circulating SCFA levels change postprandially?
Acetate
Propionate
Butyrate
Methylbutyrate
Isobutyrate
Valerate
Isovalerate
Hexanoate
Concentration (ng/ml)
Time
0
30'
2h
4h
Mean
Median
2. What is the contribution of the host genetics & gut microbiome to SCFA levels in stool & serum (fasting-postprandial)?
Heritability in circulating and faecal SCFAs
Contribution (%)
serum stool
AUC
Spearman Correlation
Stool
Serum
Postprandial peak
Postprandial dip
3. What are the links between circulating SCFA levels & chronic & acute inflammatory responses?
Methylbut.
Healthy
Rib frx
Hip frx
IL-10
IL-6
TNF-α
GlycA
IFN-γ

Figure 1. Flowchart of the study design. The three cohorts are independent. From each cohort, we included the subset of individuals who had short-chain fatty acids (SCFA) measured. TwinsUK and ZOE PREDICT-1 cohorts consist of healthy individuals, while the acute trauma case-control includes fracture patients and healthy individuals.

included on average 2.25 g of dietary fiber (50 g fat and 85 g carbohydrate). The changes in the SCFA levels in response to it are depicted in Figure 2b. We report a significant postprandial change in all SCFAs from fasting (Wilcoxon test, $p < 0.05$) except for peak hexanoate and butyrate, and dip valerate (Supplementary Table S1). Compared to the fasting measure, SCFA levels changed on average by at least 1.03-fold for methylbutyrate and by as much as 2.6 folds for acetate. The coefficient of variation (CV) indicated a moderate postprandial inter-individual variability in their highest (CV range = 26.5–39.8%) and lowest concentrations (CV range = 32–46.7%) (Supplementary Table S1). We found only weak non-statistically significant associations between postprandial SCFAs and postprandial lipemic and glycemic parameters (2-h glucose iAUC, rise in triglyceride at 6 h

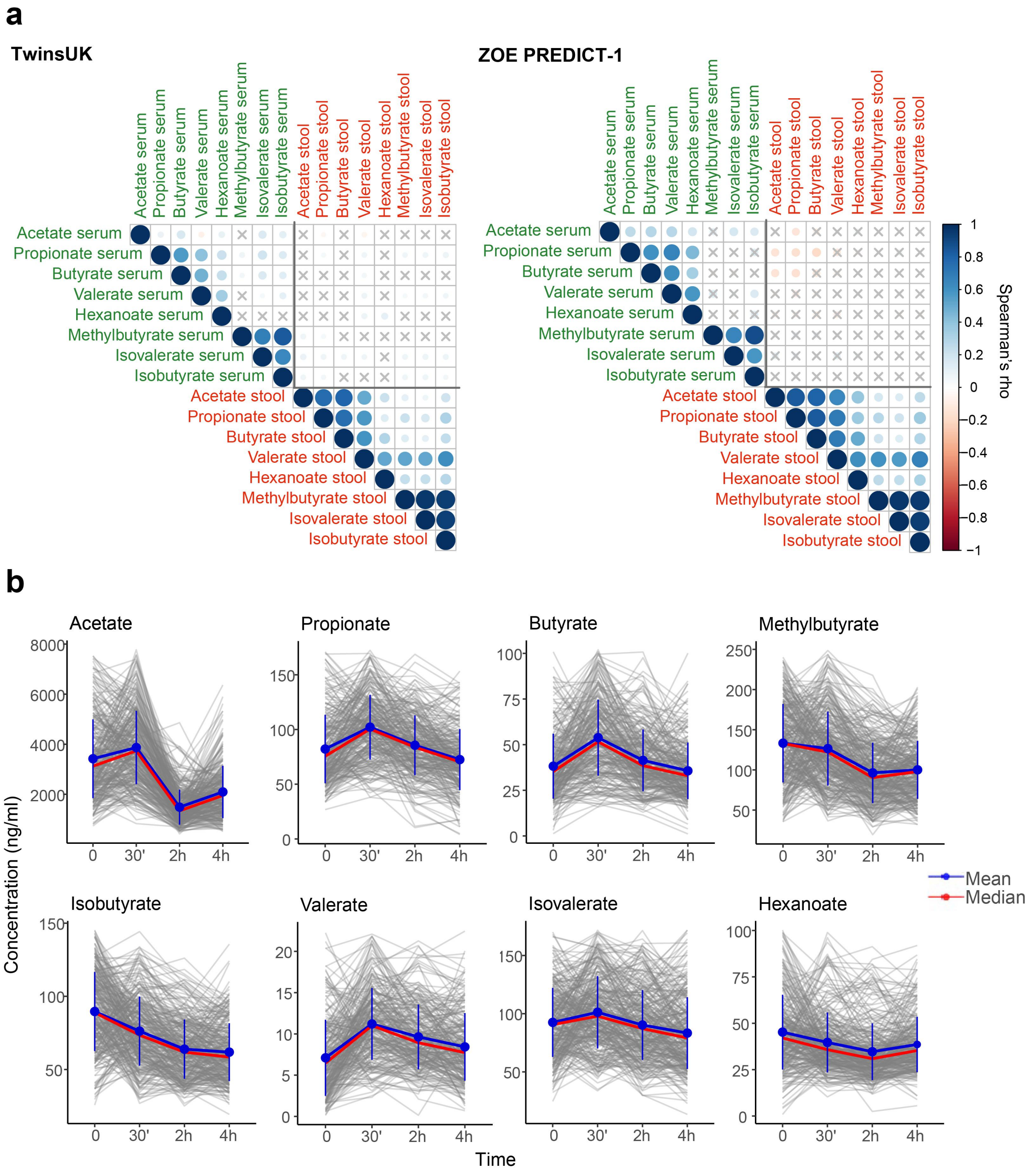


Figure 2. (a) Fasting circulating and faecal SCFAs correlation in TwinsUK (*n*=2229) and ZOE PREDICT-1 (*n*=328). Spearman's correlations are presented. Non-significant correlations (FDR≥0.05) are indicated with a 'X'. (b) Postprandial changes in circulating SCFA levels for 328 ZOE PREDICT-1 participants in response to a meal challenge under a controlled clinic setting at baseline and after 30 min, 2h and 4h. The bars indicate the standard deviations for each time point.

postprandially, rise in insulin at 2 h postprandially and rise in C-peptide at 2 h postprandially), with the exception of the significant associations with the rise of triglycerides at 6 h postprandially with dip acetate, and 2-h glucose iAUC with dip valerate and acetate (Supplementary Table S2).

Host genetics contribution to SCFA levels

We estimated the contribution of host genetics to the SCFA levels in serum and stool by calculating heritability estimates using structural equation modeling adjusted for covariates and pooling together TwinsUK and ZOE PREDICT-1 participants (serum: $n = 2835$, 1540 monozygotic (MZ) pairs, 826 dizygotic (DZ) pairs, and 469 singletons; stool: $n = 2557$, 1258 MZ pairs, 734 DZ pairs, and 565 singletons). The estimated additive genetic component (h^2) for circulating SCFA levels were on average 14% (SD = 14%), ranging from 0% for valerate (95%CI = 0%,0%), isovalerate (95%CI = 0%,19%), and hexanoate (95%CI = 0%,0%), to 38% (95%CI = 32%,44%) for butyrate. In stool, the estimated additive genetic component was on average 12% (SD = 6%) ranging from 3% (95%CI = 0%,27%) for valerate to 19% for acetate (95%CI = 0%,42%) and isovalerate (95%CI = 0%,44%) (Figure 3a). Acetate, propionate and butyrate presented larger heritability estimates for serum (average h^2 = 27%(SD = 12%)) than for stool levels (average h^2 = 14%(SD = 5%)). In a sub-analysis, we also explored the host genetics contribution to the postprandial SCFA levels in ZOE PREDICT-1 participants ($n = 328$, 164 MZ pairs, 50 DZ pairs and 114 singletons). The ACE model obtained for the peak and dip calculated for each SCFA revealed that postprandial levels of propionate, isobutyrate and isovalerate are environmentally driven, whereas hexanoate and acetate have a large genetic component with heritability estimates of 54% (95% CI = 37%,72%) and 10% (95%CI = 0%,82%) for the peak, and of 28% (95%CI = 4%,52%) and 58% (95% CI = 43%,73%) for the dip, respectively (Figure 3a).

Gut microbiota contribution to SCFA levels

As SCFAs are gut microbial-derived metabolites, we investigated the gut microbiome contribution to SCFA levels in serum and stool using RF classifiers and regressors trained on relative abundance values of gut microbiome species. The performance was evaluated with the median AUC values for the classifiers and the median Spearman's correlation values (defined as "ρ") for the regressors over 100 bootstrap folds (see Methods). We found that the gut microbiota was able to accurately predict the fecal SCFA levels (AUC>0.71 and ρ>0.29), with the strongest associations observed for acetate: AUC[95%CI] = 0.85 [0.85,0.86], ρ[95%CI] = 0.48 [0.47,0.49]; propionate: AUC[95%CI] = 0.86 [0.85,0.87], ρ[95%CI] = 0.53 [0.52,0.54] and butyrate: AUC[95%CI] = 0.89 [0.88, 0.89], ρ[95%CI] = 0.56 [0.55,0.56] (Figure 3b). We also identified *Akkermansia muciniphila*, *Faecalibacterium prausnitzii* and *Roseburia* spp., among others, as important predictors (Supplementary Figure S2). On the other hand, a moderate association was found between the gut microbiota and circulating SCFAs with an AUC and ρ average of 0.63 (95%CI = 0.61,0.63) and 0.15 (95% CI = 0.14,0.16), respectively (Figure 3b). These findings were consistent with the results obtained for TwinsUK and ZOE PREDICT-1 separately (Supplementary Table S3). In a sub-analysis, we also report that postprandial SCFA levels are poorly linked with gut microbiome composition (average AUC = 0.53 (95%CI = 0.51,0.54) and ρ =0 .07(95% CI = 0.05,0.08) for the peak, and average AUC = 0.56 (95%CI = 0.54,0.58) and ρ = 0.07(95%CI = 0.05,0.09) for the dip (Figure 3b).

Circulating SCFA levels in chronic and acute inflammation

SCFAs are known to modulate immune responses by regulating the production of immune cells and cytokines [6]. We thus investigated the role of circulating SCFAs in chronic and acute inflammatory responses. A deep understanding of how SCFAs can influence inflammation is crucial for developing strategies for the management and recovery of patients with inflammatory disorders.

We first investigated the relationship between circulating SCFA levels and systemic inflammation. For that, we performed Pearson's correlations between serum levels of SCFAs and cytokines in the 328 individuals from ZOE PREDICT-1, a subset of 82 women from TwinsUK, and in 21 healthy individuals from the acute trauma case-control study (see Supplementary Table S4 for descriptive characteristics). We then combined the results from the different studies using an inverse variance random effect meta-analysis (Figure 4a). We found that healthy

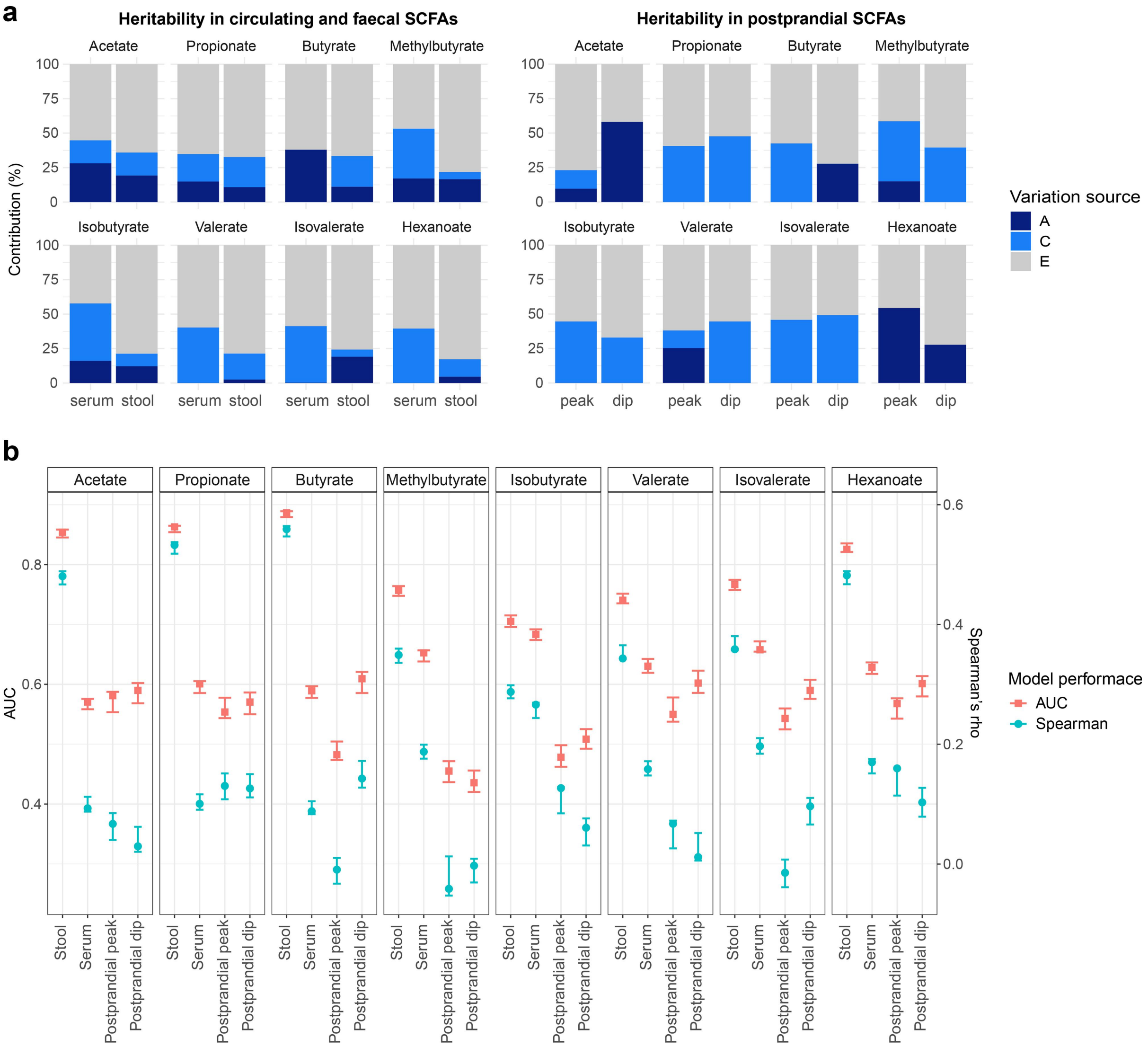


Figure 3. Contribution of host genetics and gut microbiome composition to SCFA levels in serum and stool. Analyses using serum at fasting and stool measurements were performed using the TwinsUK and ZOE PREDICT-1 participants together, while analyses using postprandial measurements were run using ZOE PREDICT-1 participants. Postprandial measures were defined as peak (the maximum SCFA concentration in the 4 hours following the test meal challenge minus the fasting level) and dip (the fasting level minus the minimum SCFA concentration in the 4 hours following the test meal challenge) (a) Heritability estimates of (left) fasting circulating and fecal SCFAs, and (right) postprandial circulating SCFAs. A, C and E labeling indicates the amount of variance attributed to the additive genetic factors or heritability, common/shared environmental factors, and unique environmental factors/error, respectively. (b) Influence of the gut microbiota composition in fecal and circulating (fasting and postprandial) SCFA levels estimated by Random Forest regression (using Spearman's correlations) and classification (using AUC) models. Blue bars indicate the median and the 95% confidence intervals of the correlation between the real value of each component and the value predicted by regression models across 100 training/testing folds. Red bars represent the median AUC and the 95% confidence intervals across 100 folds for a corresponding binary classifier between the highest and lowest quartile.

individuals tended to have negative correlations with pro-inflammatory markers including interferon-gamma (IFN-γ) (isovalerate: ρ=-0.61, p-value = 0.004; isobutyrate: ρ=-0.5, p-value = 0.03) and GlycA (acetate: ρ=-0.14, p-value = 0.0005).

We further explored whether there were any links between fasting and postprandial changes in SCFA and postprandial interleukin-6 (IL-6) and GlycA levels [12] in the ZOE PREDICT-1 participants (Figure 4b). We found postprandial GlycA (measured

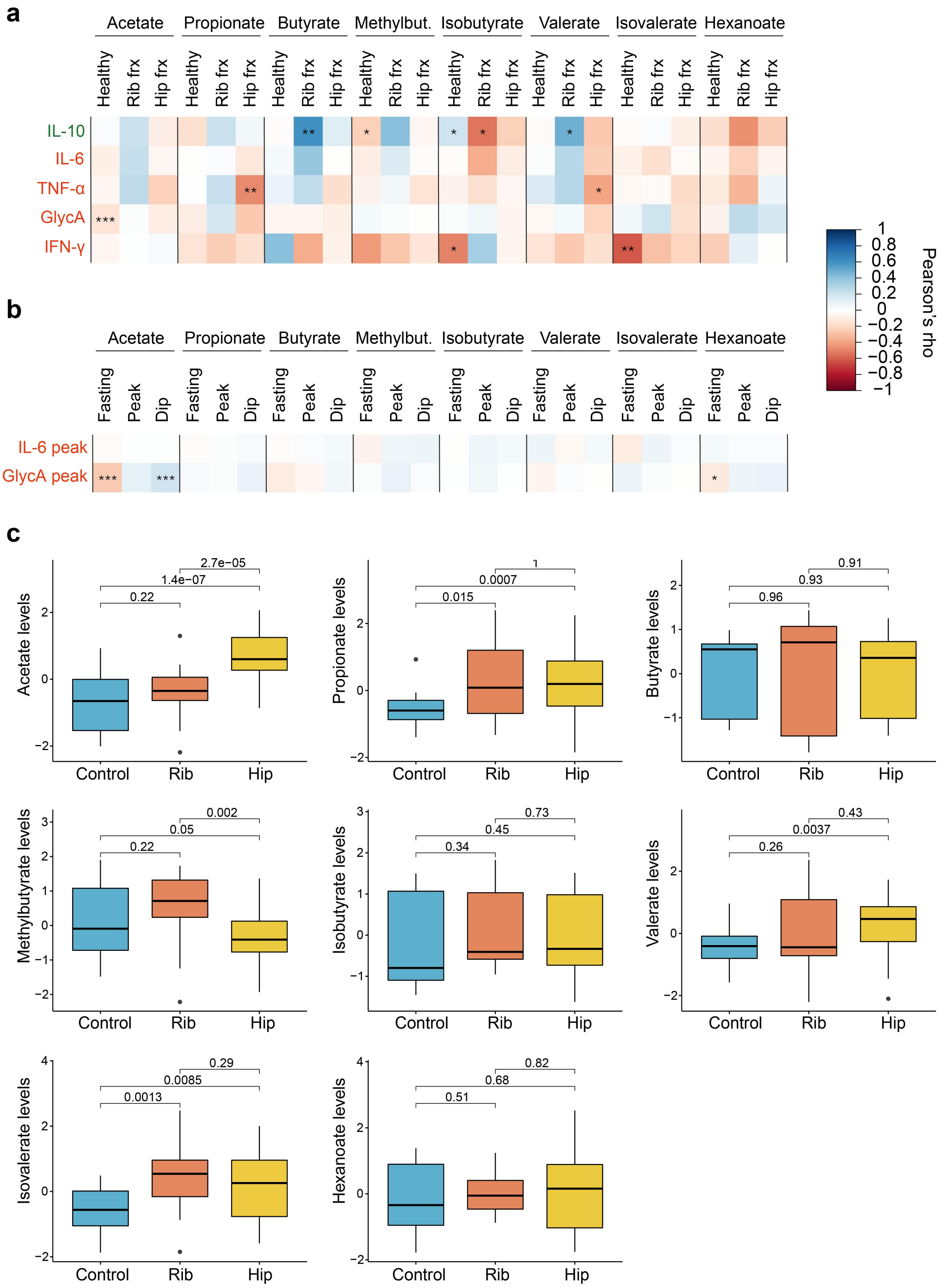


Figure 4. Role of circulating SCFAs in chronic and acute inflammation. (a) Pearson's correlations between SCFA levels and anti- (IL-10) and pro-inflammatory (IL-6, TNF-α, GlycA,IFN-γ) markers stratified by healthy individuals and acute trauma cases. For the healthy group, correlation results obtained in the healthy individuals from the acute trauma case-control cohort, in the subset from TwinsUK and ZOE PREDICT-1 were combined by applying inverse variance random effect meta-analysis. Cases are from the acute trauma case-control cohort. The controls illustrate the links between SCFAs and chronic inflammation, whereas the cases show the links between SCFAs and acute inflammatory responses. (b) Pearson's correlations between fasting and postprandial SCFA levels and the postprandial pro-inflammatory markers available in ZOE PREDICT-1 (IL-6 and GlycA). (c) Differences in the SCFA levels between

as the highest concentration within 6 h postprandially) to be strongly correlated with fasting (ρ=-0.26, *p*-value = 9.2×10^{-6}) and postprandial (ρ = 0.2, p-value = 0.0008) acetate, while no significant correlations were found with postprandial IL-6.

We then investigated whether there were links between serum SCFAs and acute inflammatory responses measuring SCFA levels and their correlations to inflammatory markers in serum samples taken immediately preoperatively from individuals who had undergone either fragility hip fractures (*n* = 32) or multiple rib fractures (*n* = 18) requiring surgery (see Supplementary Table S4 for descriptive characteristics). The fragility fractures are measured in individuals with frailty (i.e., with high systemic inflammation)[13] whereas the rib fractures cases are individuals who were otherwise healthy before the trauma.

When we assessed SCFA-cytokines correlations in the acute trauma cases (Figure 4a), we identified different patterns depending on whether the individual had a rib or a hip fracture. Specifically, the hip fracture patients presented significant negative correlations between the pro-inflammatory marker tumor necrosis alpha (TNF-α) and two SCFAs, namely propionate (ρ=-0.47, p-value = 0.007) and valerate (ρ=-0.39, p-value = 0.03). On the other hand, patients with multiple rib fractures presented significant associations with the interleukin-10 (IL-10), either positive or negative. Butyrate and valerate were positively associated with IL-10 levels (butyrate: ρ = 0.6, p-value = 0.007; valerate: ρ = 0.48, p-value = 0.04), whereas isobutyrate was presenting the opposing direction (ρ=-0.54, p-value = 0.02). No significant associations were found with IL-6 levels. Likewise, when comparing the circulating SCFA levels between healthy controls, rib or hip fracture patients using pairwise t.tests, we observed that acetate levels were significantly different in the three groups, whereas propionate and isovalerate levels in trauma cases were significantly higher than in the controls, and valerate levels were higher in the hip fracture patients in comparison with the controls. On the other hand, patients with a hip fracture presented significantly lower levels than patients with multiple rib fractures (Figure 4c). Results were consistent when running linear models and when further adjusting for age and sex (Supplementary Table S5).

Discussion

To our knowledge, this is the first study to date investigating simultaneously the contribution of host genetics and gut microbiome to the fasting and postprandial levels of eight SCFAs in serum and stool in two independent cohorts of healthy individuals. Specifically, we have shown that (i) there is a very low concordance between fecal and circulating SCFA levels, which might be due to the fact that most absorbed SCFAs act as an energy source for the enterocytes and are not systemically transported[3], (ii) SCFA levels change postprandially and there are substantial inter-individual differences in these responses, (iii) stool and serum SCFA levels are heritable, with the exception of circulating valerate, isovalerate and hexanoate that are environmentally determined, (iv) most postprandial SCFA levels appear to be environmentally driven, (v) the gut microbiome composition is an important contributor of fecal levels, but presents weaker associations with circulating levels. Importantly, using an independent acute trauma case-control cohort, we report for the first time that circulating SCFA levels vary between trauma patients and controls and that there is a different relationship between pro- and anti-inflammatory cytokines and SCFAs depending on the type of inflammatory response (chronic or acute).

We found that a large proportion of the SCFA levels in serum and stool are explained by environmental factors, which is in line with the proposed

controls and cases in the acute trauma case-control cohort. The p-values obtained from t-tests between groups are indicated. For these analyses, only individuals with both serum SCFAs and cytokines are included (i.e., TwinsUK, n=82; ZOE PREDICT-1, n= 328; acute trauma case-control cohort: controls, n= 21, rib fracture, n= 18, hip fracture, n=32). Levels were log-transformed and Z-scaled. P-value: *0.05; **0.01; ***0.001. Pro- and anti-inflammatory cytokines are colour-coded in red and green, respectively. Abbreviations: Frx: Fracture; IFN, interferon; IL, interleukin; Methylbut., methylbutyrate; TNF, tumour necrosis factor.

importance of non-genetic factors in the SCFAs formation, including the different environmental factors modulating the gut microbiota.[14] Of note, we found that the three most widely studied SCFAs (acetate, butyrate and propionate) had moderate heritability estimates, with a larger genetic contribution to serum (average $h^2 = 27\%(\text{SD} = 12\%)$) than to stool levels (average $h^2 = 14\%(\text{SD} = 5\%)$). Our results are supportive of Sanna and coworkers previous report suggesting that host genetics influence the microbial expression of propionate and butyrate.[4] The lower heritability for stool level we found is not surprising given that fecal levels are more likely to reflect SCFA bacterial generation, whereas serum levels will reflect absorption from the gut but also synthesis by the host (e.g., acetate is a metabolite involved in the tricarboxylic acid (TCA) cycle).[15] On the other hand, we were unable to detect a genetic contribution for the postprandial levels of most SCFAs, similar to what we observed in this same cohort for postprandial c-peptide and insulin responses (see[16]). The only exceptions were postprandial acetate and hexanoate both of which presented large heritability estimates, which can be due to the fact that, as previously mentioned, acetate is involved in the TCA cycle,[15] and hexanoate can be also generated by hepatic peroxisomal beta-oxidation of long-chain fatty acids.[17]

When investigating the contribution of the gut microbiome, we found that it can accurately predict SCFA levels in stool (AUC>0.71), however, the predictive power is reduced for serum levels, both fasting and postprandial. This is consistent with our findings indicating that SCFA levels in serum and stool are not correlated with each other. These observations highlight the fact that fecal levels are not representative of the actual absorption and suggest that caution should be taken when inferring microbiome-disease associations,[18] from either serum or fecal SCFA levels. Both types of measurements are needed to fully understand the role of SCFAs in health. Moreover, we were able to identify the key gut microbial species modulating SCFAs fecal abundances. These include the widely known SCFA producers *F. prausnitzii*,[19] *Roseburia* spp,[20] or *C. comes*,[9] among others, positively correlated with acetate, propionate and butyrate, which also confirms the robustness of our methodology. On the other hand, we identify *Alistipes* spp. showing negative associations. *Alistipes* spp. has been associated with gut dysbiosis and chronic inflammation diseases.[21] Of note, some of the identified species were showing an opposing direction between acetate, propionate and butyrate, and isobutyrate, methylbutyrate and isovalerate (e.g., *R. lactatiformans* is negative for the first three and positive for the last three). This can be explained by the distinct substrate used to produce SCFAs.[22] Indeed, acetate, butyrate and propionate are mainly produced by the fermentation of resistant carbohydrates,[23] while isobutyrate, methylbutyrate and isovalerate are mainly produced from the amino acid fermentation.[24]

We also explored the links between serum SCFAs and chronic and acute inflammation. When examining the correlations between cytokines and SCFA levels in healthy individuals, we observed that SCFAs are linked to lower systemic inflammation consistent with these compounds being involved in downregulating pro-inflammatory markers (e.g., isovalerate and isobutyrate vs IFN-γ) and their postprandial responses (e.g., fasting and postprandial acetate vs postprandial GlycA). This is in line with the already reported benefits of SCFAs in chronic inflammation.[25] In acute trauma patients we report that hip fragility fractures and multiple rib impact fractures led to differences in some circulating SCFA levels with respect to healthy individuals and with each other. According to animal studies, SCFAs are mostly metabolized by the muscles and kidneys.[11] After a fragility fracture, like hip fractures, acute kidney injury is a frequent complication, whereas it is significantly less frequent in impact fractures, like rib fractures,[26] which might explain some of the observed differences in SCFA levels between distinct trauma patients and healthy individuals. When we analyzed the observed differences in the acute trauma patients in relation to their correlations with inflammatory markers, we noted that SCFA levels were positively correlated with the anti-inflammatory cytokine IL-10 (butyrate and isovalerate) but only among rib fracture, whereas SCFA levels were negatively associated with pro-inflammatory cytokines in hip fracture cases (e.g., TNF-α vs propionate and valerate). Therefore, the associations between SCFAs and cytokines are different in rib fractures (young and healthy before trauma) and hip fractures (frail elderly individuals). Importantly, mortality rates within 30 days of fractures are < 1% for rib fractures for individuals under the age of 65[27] and > 10% for fragility hip fractures[28] which is also one of the global top 10 causes of disability in adults.[29] The differences in inflammatory responses in these two trauma scenarios suggest that SCFAs and their links to pro- and anti-inflammatory pathways might be related also to the recovery process. Taken together, SCFAs might help to dampen the inflammatory response in acute inflammation, while they might contribute to the maintenance of a low-grade inflammatory state in systemic inflammation by influencing fasting and postprandial inflammation.

We acknowledge the following study limitations. First, in ZOE PREDICT-1, SCFAs were measured only in women, and therefore, postprandial results might not be generalized to men. Unfortunately, we could not compare the postprandial findings in TwinsUK, which include men and women, as postprandial measurements are not available for this cohort. Likewise, it was not possible to assess the postprandial responses in the acute trauma case-control cohort as samples were collected in an acute hospital setting. The heritability analyses that exclusively include the ZOE PREDICT-1 participants lack statistical power, and the results might differ if more participants are included. Lack of power also prevented us from examining the genetic factors that influence postprandial levels of SCFAs. Although we meta-analyzed the correlations obtained from the acute trauma case-control, a subset of TwinsUK and ZOE PREDICT-1 studies, not all of them presented the same cytokines measures. Besides, postprandial levels were only available for GlycA and IL-6 in the ZOE PREDICT-1 study, and the data used in this study does not allow us to infer causality between SCFA levels and the studied inflammatory markers. Finally, we were unable to evaluate variations in SCFA levels over time as longitudinal SCFA data was unavailable for the included cohorts.

Despite the above limitations, we benefit from an independent and well-characterized large population-based study, a detailed postprandial interventional study and an acute fracture case-control study that allowed us to investigate the link between circulating SCFAs and acute inflammation.

In conclusion, in the most comprehensive study to date examining the contribution of host genetics and gut microbiome composition to fecal and circulating levels in two independent population-based cohorts, our findings indicate that SCFA levels are mostly modifiable and change postprandially, and fecal SCFAs reflect the gut microbiome composition. We also show for the first time that the SCFA profile and their correlations with inflammatory markers change depending on the type of inflammatory response (chronic or acute trauma). Taken together, our results illustrate the breadth of the physiological relevance of SCFAs on human inflammatory and metabolic responses highlighting the need for a deeper understanding of this important class of molecules.

Patients and methods

Study populations

This study consists of three completely independent cohorts. TwinsUK and ZOE PREDICT-1 consist of healthy individuals, whereas the acute trauma case-control cohort includes three subsets of individuals (healthy individuals, patients with rib fractures and patients with hip fractures). The acute trauma case-control cohort was included to exemplify the role of circulating SCFA levels in an acute inflammatory situation, as the rest of the work focused on SCFA levels in healthy individuals.

TwinsUK

TwinsUK registry is a national register of adult twins recruited as volunteers without selecting for any particular disease or trait.[30] We included 2507 and 2229 individuals with serum and fecal SCFA measurements, respectively. For those, 2197 had measurements in both stool and serum. Along with the SCFA measurements, shotgun metagenomes from the gut microbiome were also available. A subset of 82 individuals also had measurements of circulating cytokines. The study was approved by NRES Committee London – Westminster, and all twins provided informed written consent.

ZOE PREDICT-1

We included a subset of 328 individuals from the UK-based ZOE PREDICT-1 study with SCFAs measured in serum and stool, and gut microbiome composition assessed with shotgun metagenomes. The ZOE PREDICT-1 study[16] was a single-arm nutritional intervention conducted between June 2018 and May 2019. Study participants were healthy individuals (thus eliminating potential confounders brought about by the presence of infections or other comorbidities) aged between 18–65 years recruited from the TwinsUK registry,[30] and the general population using online advertising. Although the ZOE PREDICT- 1 participants were recruited from the TwinsUK registry, in this study the two cohorts, ZOE PREDICT-1 and TwinsUK, are completely independent and there is no overlap in participants. Participants attended a full-day clinical visit consisting of test meal challenges followed by a 13-day home-based phase, as previously described.[16]

Test meal challenge. Within a tightly controlled clinical setting, participants consumed meal 1: breakfast muffins and a milkshake (890 kcal, 85.5 g carbohydrate (38.4%), 52.7 g fat (53.3%), 16.1 g protein (7.2%), and 2.3 g fiber at the 0-hour timepoint, following baseline blood draw). Venous blood samples were collected at 15, 30, 60, 120, 180, 240, 300, 360 minutes post-meal 1.

Acute trauma case-control cohort

Patients were all recruited at Queens Medical Hospital part of the Nottingham University Hospital's (NUH) NHS Trust.

Rib fracture cohort (OPERA). Inclusion criteria were: adult patients (16 years and above) presenting multiple (3+) rib fractures suitable for surgical repair and having, as per British Orthopaedic Association Audit Standards For Trauma (BOAST-15) Standard 8, indications for fixation as: clinical flail chest; respiratory difficulty requiring respiratory support or uncontrollable pain using standard modalities; was a surgical candidate.

Patients were excluded if: they had a head or thoracic injury requiring emergency intervention; could not be operated on within 72 hours as unfit for surgery; presented with significant thoracic injury requiring surgery where conservative management would be inappropriate. Blood samples were taken at the time of the patient going into anesthesia ahead of entering the operating theater.

Hip fracture cohort (FEMUR). Inclusion criteria: age over 65 (no upper age limit), Rockwood frailty score ≥ 4, fractured hip sustained following a fall that required surgery. Good understanding of spoken and written English language, ability to give informed consent or to provide assent and availability of a legally acceptable surrogate to provide consent. Exclusion criteria: those who fell and sustained the hip fracture more than 12 hours prior to hospitalization. Patients who had fallen and sustained a hip fracture whilst in-patient. Surgery that had to be delayed to 96 hours or more after the fall.

Control cohort with SCFAs and cytokines. Healthy students from the School of Medicine at the University of Nottingham or healthcare workers.

Ethics

TwinsUK: This study was carried out under TwinsUK BioBank ethics, approved by North West – Liverpool Central Research Ethics Committee (REC reference 19/NW/0187), IRAS ID 258,513. This approval supersedes earlier approvals granted to TwinsUK by the St Thomas' Hospital Research Ethics Committee, later London – Westminster Research Ethics Committee (REC reference EC04/015), which have now been subsumed within the TwinsUK BioBank.
ZOE PREDICT-1: The study was approved by the London – Hampstead Research Ethics Committee (REC reference 18/LO/0663) and the trial was registered on ClinicalTrials.gov (registration number: NCT03479866).

The rib fractures cohort was collected as part of The Operative Rib Fixation (ORiF) Study (REC Reference: 18/SC/066, IRAS 248,460, IRSCTN 10,777,575). The hip fracture cohort was collected under Functioning of Elder Muscle; Understanding Recovery (FEMUR) study (REC approval: 20/LO/0841 clinicaltrials.gov registration NCT04764617). The control individuals collected alongside were collected under REC ref FMHS 302–0621 by the internal review board by the University of Nottingham School of Medicine.

All participants provided written informed consent.

SCFA measurements

Metabolomic profiling was performed on 2906 serum samples (2507, 328 and 71 participants from TwinsUK, ZOE PREDICT-1 and the acute trauma case-control cohort, respectively) and 2557 stool samples (2229 and 328 participants from TwinsUK and ZOE PREDICT-1, respectively) by Metabolon Inc. using liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS), as previously described.[31] For TwinsUK and ZOE PREDICT-1 cohorts, stool and fasting serum samples were available, whereas only fasting serum samples were available for the acute trauma case-control cohort, as samples were collected in a hospital setting. For the ZOE PREDICT-1 participants, postprandial (30 min, 2 h and 4 h) serum samples were also collected after consuming a standardized meal (see Methods:Test meal challenge). Full details and quality control are included in Supplementary Text 1. In all the samples, the SCFA acetate, propionate, butyrate, methylbutyrate, isobutyrate, valerate and isovalerate, and the medium-chain fatty acid hexanoate were measured. For the sake of ease of reading, hexanoate is included in the definition of SCFA.

Postprandial metrics

For each SCFA, we defined as (i) peak the maximum SCFA concentration in the 4 hours following the test meal challenge minus the fasting level, and (ii) dip the fasting level minus the minimum SCFA concentration in the 4 hours following the test meal challenge. Postprandial lipemic and glycemic parameters (the 2-h glucose iAUC, rise in triglyceride at 6 h postprandially, rise in insulin at 2 h postprandially and rise in C-peptide at 2 h postprandially - see Berry et al., 2020[16] for more details), and cytokines (the highest concentration of GlycA and IL-6 within 6 h postprandially) were also available.

Microbiome sequencing and profiling

Deep shotgun metagenomic sequencing in stool samples from TwinsUK and ZOE PREDICT-1 was performed as previously described,[22,32], and as detailed below.

Faecal sample collection

TwinsUK and ZOE PREDICT-1 participants collected stool samples at home in pre-labeled kits (containing 2 × 25 ml tube or 1 × 25 ml tube and 1 × 10 ml Zymo buffer) posted to them prior to their clinic visit date and brought with them to the visit. Alternatively, samples can be posted to the clinic using blue Royal Mail safe boxes. In the laboratory, samples were homogenized, aliquoted into 4 bijou tubes, and stored at – 80°C, within 2 hours of receipt.

DNA extraction, library preparation and DNA sequencing

To isolate genomic DNA from fecal material in TwinsUK, bijou tubes are removed from the freezer and ground with glass beads and 5-6 ml distilled water (Spex Grinder, 10 seconds, 800 strokes per minute). The supernatant is centrifuged and ground further (5 minutes, 1000 strokes per minute) before 200–300 µl of the sample is mixed with 10 µl PK solution and 720 µl of Lysis/Bind Master Mix). Proteins are degraded by the binding solution and subsequently extracted by KingFisher Flex robot. DNA is washed in 2 steps by washing solutions and eluted in MagMax Core Elution Buffer in 100 µl. In ZOE PREDICT-1, DNA was isolated by QIAGEN Genomic Services using DNeasy 96 PowerSoil Pro from the microbiome samples. Library preparation and sequencing were performed by GenomeScan for TwinsUK. For ZOE PREDICT-1, the quality and yield after sample preparation were measured with the Fragment Analyzer system following the manufacturer's guidelines. The size of the resulting product was consistent with the expected size of approximately 500–700 bp. Libraries were sequenced for 300 bp paired-end reads using the Illumina NovaSeq6000 platform according to the manufacturer's protocols. 1.1 nM library was used for flow cell loading. NovaSeq control software NCS v1.5 was used. Image analysis, base calling, and the quality check were performed with the Illumina data analysis pipeline RTA3.3.5 and Bcl2fastq v2.20.

Metagenome quality control and preprocessing

TwinsUK sequenced metagenomes were processed using the YAMP pipeline (v. 0.9.5.3).[33] Briefly, identical reads were removed. Reads were filtered to remove adapters, known artifacts, phi X 174, and then quality trimmed (PhRED quality score < 10). Reads that became too short after trimming ($N < 60$ bp) were discarded. We retained singleton reads (i.e., reads whose mate has been discarded) to retain as much information as possible. Contaminant reads belonging to the host genome were removed (build: GRCh37). Low-quality samples, i.e., samples with <10 M reads after QC were discarded ($n = 4$). Sequenced metagenomes in ZOE PREDICT-1 were QCed using the pipeline implemented in https://github.com/SegataLab/preprocessing.

Microbiome taxonomic profiling

The metagenomic analysis was conducted following the general guidelines[34] and based on the bioBakery computational environment.[35,36] High-resolution taxonomic profiling of the TwinsUK and ZOE PREDICT-1 metagenomes was performed using MetaPhlAn 4.beta.2 with the Jan21 database that comprises 26,970 species-level genome bins, with default parameters.[37]

Inflammatory markers measurements

Pro-inflammatory markers TNF-α, IFN-γ, GlycA and IL-6, and the anti-inflammatory marker IL-10 were measured by ELISA by Affinity biomarkers in the acute trauma case-control cohort. In TwinsUK, IL-10, TNF-α, and IL-6 were measured using the bead-based high-sensitivity human cytokine kit (HSCYTO-60SK, Linco-Millipore) according to the manufacturer's instructions. In ZOE PREDICT-1, IL-6 was measured by Affinity Biomarkers Lab using a Sandwich Immunoassay by Meso Scale Diagnostics. In TwinsUK and ZOE PREDICT-1, GlycA was measured using the high-throughput NMR metabolomic (Nightingale) 2016 panel.

Data availability statement

The data used in this study are held by the Department of Twin Research at King's College London. The data can be released to bona fide researchers using our normal procedures overseen by the Wellcome Trust and its guidelines as part of our core funding (https://twinsuk.ac.uk/resources-for-researchers/access-our-data/). The gut microbiome data is available on EBI (https://www.ebi.ac.uk/) under accession numbers PRJEB39223 (ZOE- PREDICT-1) and PRJEB32731 (TwinsUK).

Statistical analyses

Statistical analyses were performed using RStudio version 1.3.1093, and python. All the models were corrected for multiple testing using false discovery rate (FDR – Benjamini and Hochberg method).[38] If not indicated otherwise, the level of statistical significance was set at FDR < 0.05 in all the analyses. Before running the analyses, outliers of SCFA measures defined as values 4 standard deviations from the mean were excluded, and values were Z-scaled. Analyses with postprandial SCFA levels were performed only in ZOE PREDICT-1. To achieve the second aim of this study, TwinsUK cohort and ZOE PREDICT-1 were processed together, and findings were checked for consistency with results obtained for each individual cohort. To achieve the last aim, data from ZOE PREDICT-1, a subset of TwinsUK, and the acute trauma case-control cohort was included (see Figure 1).

Correlations between circulating and fecal SCFA levels

To investigate the correlations between circulating and fecal SCFA levels, we used non-parametric Spearman's correlations as the SCFA measurements in TwinsUK and ZOE PREDICT-1 did not follow a normal distribution.

Changes in postprandial SCFA levels and associations with postprandial lipemic and glycemic parameters

We examined postprandial changes from fasting in each SCFA using the Wilcoxon test, and the inter-individual variability in the highest and lowest postprandial concentration of each SCFA using the coefficient of variation (CV – calculated as SD/mean, %). We assessed the associations with postprandial SCFA levels and Z-scaled log-transformed postprandial lipemic and glycemic parameters running linear mixed models adjusting for age, BMI and family relatedness as random effects.

Host genetics contribution to SCFA levels: heritability estimates

To estimate the heritability of the SCFA levels in serum (fasting and postprandial) and stool, we utilized the classical twin model and compared the degree of similarity among monozygotic (MZ) twins, who share 100% of their genetic make-up, and dizygotic (DZ) twins, who share on average 50% of their segregating genes. Under the equal environment assumption (EEA), the variance of the trait/phenotype (P) is explained by three latent parameters: additive genetic variance (A), shared (familial) environmental variance (C) and individual-specific environmental variance/error (E).[39] To estimate the heritability, we utilized the structural equation models (SEM), which uses the observed covariances from both MZ and DZ pairs to establish a causal relationship between the covariances and the latent parameters. We performed the heritability analysis using the twinlm function (R METs package).[40] Heritability of the SCFA levels in fasting serum and stool was calculated in TwinsUK and ZOE PREDICT- 1 participants together to increase the sample size ensuring accurate estimates. Heritability models of the fecal and fasting circulating SCFA levels were adjusted for age, sex and BMI, whereas models of the postprandial levels were adjusted for age and BMI (sex was not included as all the participants were women).

Gut microbiota contribution to SCFA levels: random forest models. The machine learning framework employed is based on the scikit-learn Python package.[41] The ML algorithms used for the prediction of SCFAs in serum (fasting and postprandial) and stool from the species-level relative abundances (as estimated by MetaPhlAn 4.beta.2 and normalized using the arcsin-sqrt transformation for compositional data) are based on Random Forest (RF) classification and regression. We selected RF-based methods a priori as it has been repeatedly shown to be particularly suitable and robust to the statistical challenges inherent to microbiome abundance data.[42] A cross-validation approach was implemented, based on 100 bootstrap iterations and an 80/20 random split into training and testing folds. To specifically avoid overfitting due to the twin nature of our data and their shared factors, we removed any twin from the training fold if their twin was present in the test fold.

For the classifiers, we divided the continuous features into two classes: the top and bottom quartiles. From the scikit-learn package, we used the RandomForestClassifier function with n_estimators = 1000, max_features='sqrt' parameters. For the regressors, we trained an RF regressor to learn the feature to predict and simple linear regression to calibrate the output for the test folds on the range of values in the training folds. From the scikit-learn package, we used the RandomForestRegressor function with n_estimators = 1000, criterion='mse', max_features='sqrt' parameters and LinearRegres sion with default parameters.

As an additional control, we verified that when randomly swapping the target labels or values (classification and regression, respectively), the performances were reflecting a random prediction, hence an area under the ROC curve (AUC) very close to 0.5 and a non-significant correlation between the real and predicted values approaching 0.

Links between circulating SCFA levels and chronic and acute inflammatory responses, and differences in SCFA levels between controls and acute fracture patients. Circulating SCFA levels and cytokines were log-transformed to obtain a normal distribution and then Z-scaled. We first assessed the associations between SCFA levels and cytokines in healthy individuals from the acute trauma case-control study, a subset of TwinsUK with measurements of circulating cytokines and SCFAs, and ZOE PREDICT-1 by running Pearson's correlations. We then combined the results from the different studies using an inverse variance random effect meta-analysis. Moreover, Pearson's correlations between each marker (IL-10, TNF-α, IFN-γ GlycA, IL-6) and SCFA stratifying by the type of acute trauma (hip fracture or rib fracture) were also run to investigate the potential link between SCFA levels and acute inflammatory responses. Pairwise t.test and linear models were employed to test differences in circulating SCFAs between trauma patients (hip/rib fracture) and controls. Linear models were further adjusted for age (set as a two levels factor defined by age ≥ 50 and age < 50) and sex.

Disclosure statement

TDS is co-founder and shareholder of ZOE Ltd ("Zoe").SEB is a consultant to ZOE and has options in ZOE. AMV, WJB, PWF, FA, NS are consultants to Zoe. JW, GH, RD, FG are employees of Zoe. KW and GAM are employees of Metabolon Inc. Other authors have no conflict of interest to declare.

Funding

This research was funded by the Chronic Disease Research Foundation and in part by the Wellcome Trust [Grant number: 212904/Z/18/Z] and by Diabetes UK [19/0006053]. For the purpose of open access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. TwinsUK receives funding from the Wellcome Trust, the European Commission H2020 grants SYSCID (contract #733100); the National Institute for Health Research (NIHR) Clinical Research Facility and the Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London, the Chronic Disease Research foundation, the UKRI Medical Research Council (MRC)/British Heart Foundation Ancestry and Biological Informative Markers for Stratification of Hypertension (AIM-HY; MR/M016560/1), and Zoe Limited. This work was also supported by UKRI grant MR/W026813/1 to CM and AMV. MM is supported by the National Institute for Health Research (NIHR) Clinical Research Facility and the Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. AN, PL and CM are funded by the Chronic Disease Research Foundation. CM is also supported by the UKRI Medical Research Council (MRC)/British Heart Foundation Ancestry and Biological Informative Markers for Stratification of Hypertension (AIM-HY; MR/M016560/1). AMV is supported by the National Institute for Health Research Nottingham Biomedical Research Centre. The collection of samples from trauma patients was supported by the NIHR Research Nottingham Biomedical Research Centre and by NIHR grants NIHR132240 (OPERA) and 16/61/10 (ORiF) to BJO. The ZOE PREDICT-1 cohort is supported by Zoe Limited. We thank all the participants of TwinsUK, ZOE PREDICT-1 and the acute trauma case-control cohorts for contributing and supporting our research.

ORCID

Cristina Menni http://orcid.org/0000-0001-9790-0571

References

1. Cummings JH, Pomare EW, Branch WJ, Naylor CP, Macfarlane GT. Short chain fatty acids in human large intestine, portal, hepatic and venous blood. Gut. 1987;28(10):1221–1227. doi:10.1136/gut.28.10.1221.
2. Topping DL, Clifton PM. Short-chain fatty acids and human colonic function: roles of resistant starch and nonstarch polysaccharides. Physiol Rev. 2001;81 (3):1031–1064. doi:10.1152/physrev.2001.81.3.1031.
3. Nogal A, Valdes AM, Menni C. The role of short-chain fatty acids in the interplay between gut microbiota and diet in cardio-metabolic health. Gut Microbes. 2021;13 (1):1–24. doi:10.1080/19490976.2021.1897212.
4. Sanna S, van Zuydam NR, Mahajan A, Kurilshikov A, Vich Vila A, Võsa U, Mujagic Z, Masclee AAM, Jonkers DMAE, Oosting M, et al. Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. Nat Genet. 2019;51(4):600–605. doi:10.1038/s41588-019-0350-x.
5. Vitale M, Giacco R, Laiola M, Della Pepa G, Luongo D, Mangione A, Salamone D, Vitaglione P, Ercolini D, Rivellese AA, et al. Acute and chronic improvement in postprandial glucose metabolism by a diet resembling the traditional Mediterranean dietary pattern: Can SCFAs play a role? Clin Nutr. 2021;40 (2):428–437. doi:10.1016/j.clnu.2020.05.025.
6. Yao Y, Cai X, Fei W, Ye Y, Zhao M, Zheng C. The role of short-chain fatty acids in immunity, inflammation and metabolism. Crit Rev Food Sci Nutr. 2022;62 (1):1–12. doi:10.1080/10408398.2020.1854675.
7. Nakahori Y, Shimizu K, Ogura H, Asahara T, Osuka A, Yamano S, Tasaki O, Kuwagata Y, Shimazu T. Impact of fecal short-chain fatty acids on prognosis in critically ill patients. Acute Med Surg. 2020;7(1):e558. doi:10.1002/ams2.558.
8. Valdés-Duque BE, Giraldo-Giraldo NA, Jaillier-Ramírez AM, Giraldo-Villa A, Acevedo- Castaño I, Yepes-Molina MA, Barbosa-Barbosa J, Barrera-Causil CJ, Agudelo-Ochoa GM. Stool short-chain fatty acids in critically Ill patients with sepsis. J Am Coll Nutr. 2020;39(8):706–712. doi:10.1080/07315724.2020.1727379.
9. Nogal A, Louca P, Zhang X, Wells PM, Steves CJ, Spector TD, Falchi M, Valdes AM, Menni C. Circulating levels of the short-chain fatty acid acetate mediate the effect of the gut microbiome on visceral fat. Front Microbiol. 2021;12:711359. doi:10.3389/fmicb.2021.711359.
10. Meyer RK, Lane AI, Weninger SN, Martinez TM, Kangath A, Laubitz D, Duca FA. Oligofructose restores postprandial short-chain fatty acid levels during

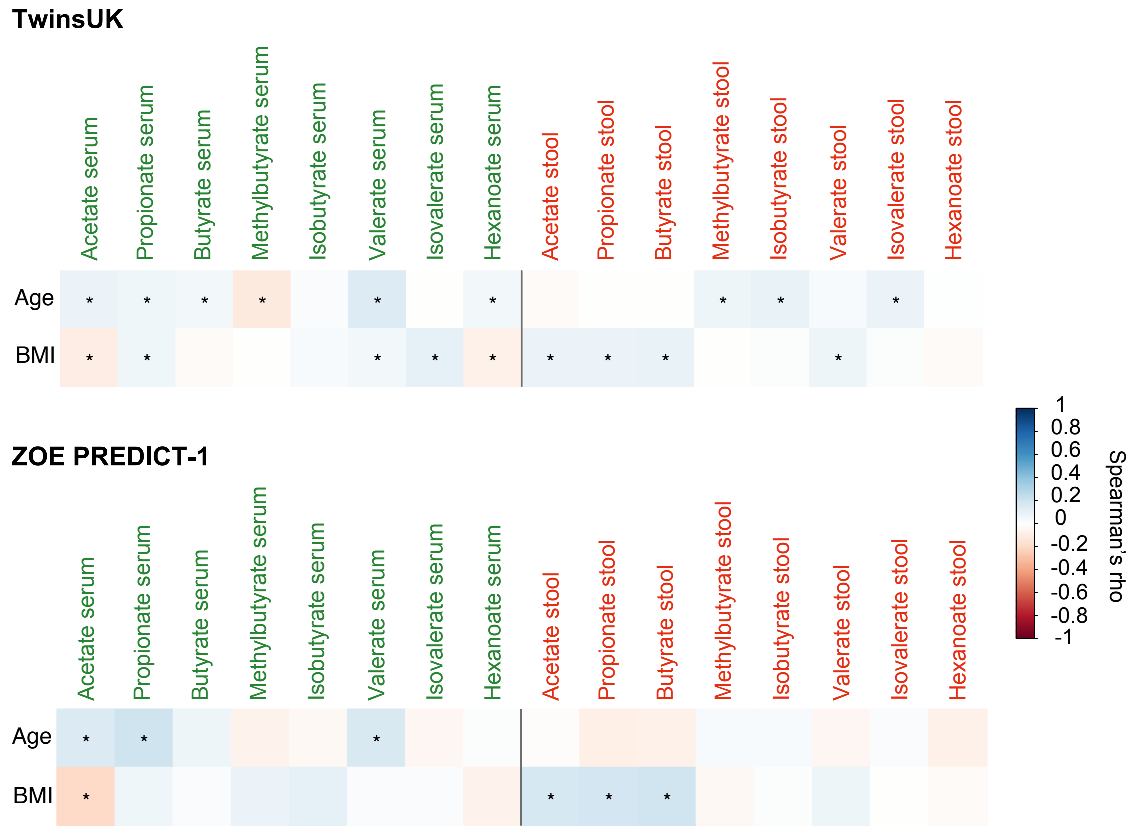
high-fat feeding. Obesity. 2022;30(7):1442–1452. doi:10.1002/oby.23456.

11. Kirschner SK, Ten Have GA, Engelen MP, Deutz NE. Transorgan short-chain fatty acid fluxes in the fasted and postprandial state in the pig. American J Physiol-Endocrinol Metab. 2021;321(5):E665–E673. doi:10.1152/ajpendo.00121.2021.
12. Wang ZC, Jiang W, Chen X, Yang L, Wang H, Liu YH. Systemic immune-inflammation index independently predicts poor survival of older adults with hip fracture: a prospective cohort study. BMC Geriatr. 2021;21(1):155. doi:10.1186/s12877-021-02102-3.
13. Mazidi M, Valdes AM, Ordovas JM, Hall WL, Pujol JC, Wolf J, Hadjigeorgiou G, Segata N, Sattar N, Koivula R, et al. Meal-induced inflammation: postprandial insights from the Personalised REsponses to DIetary Composition Trial (PREDICT) study in 1000 participants. Am J Clin Nutr. 2021;114(3):1028–1038. doi:10.1093/ajcn/nqab132.
14. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, Kurilshikov A, Bonder MJ, Valles-Colomer M, Vandeputte D, et al. Population-level analysis of gut microbiome variation. Science. 2016;352(6285):560–564. doi:10.1126/science.aad3503.
15. Akram M. Citric acid cycle and role of its intermediates in metabolism. Cell Biochem Biophys. 2014;68(3):475–478. doi:10.1007/s12013-013-9750-1.
16. Berry SE, Valdes AM, Drew DA, Asnicar F, Mazidi M, Wolf J, Capdevila J, Hadjigeorgiou G, Davies R, Al Khatib H, et al. Human postprandial responses to food and potential for precision nutrition. Nat Med. 2020;26(6):964–973. doi:10.1038/s41591-020-0934-0.
17. Saresella M, Marventano I, Barone M, La Rosa F, Piancone F, Mendozzi L, d'Arma A, Rossi V, Pugnetti L, Roda G, et al. Alterations in circulating fatty acid are associated with gut microbiota dysbiosis and inflammation in multiple sclerosis. Front Immunol. 2020;11:1390. doi:10.3389/fimmu.2020.01390.
18. Deng K, Xu J-J, Shen L, Zhao H, Gou W, Xu F, Fu Y, Jiang Z, Shuai M, Li B-Y, et al. Comparison of fecal and blood metabolome reveals inconsistent associations of the gut microbiota with cardiometabolic diseases. Nat Commun. 2023;14(1):571. doi:10.1038/s41467-023-36256-y.
19. Louis P, Young P, Holtrop G, Flint HJ. Diversity of human colonic butyrate-producing bacteria revealed by analysis of the butyryl-CoA: acetate CoA-transferase gene. Environ Microbiol. 2010;12(2):304–314. doi:10.1111/j.1462-2920.2009.02066.x.
20. Duncan SH, Hold GL, Barcenilla A, Stewart CS, Flint HJ. Roseburia intestinalis sp. nov., a novel saccharolytic, butyrate-producing bacterium from human faeces. Int J Syst Evol Microbiol. 2002;52(5):1615–1620. doi:10.1099/00207713-52-5-1615.
21. Parker BJ, Wearsch PA, Veloo ACM, Rodriguez-Palacios A. The genus: Gut bacteria with emerging implications to inflammation, cancer, and mental health. Front Immunol. 2020;11:906. doi:10.3389/fimmu.2020.00906.
22. Asnicar F, Berry SE, Valdes AM, Nguyen LH, Piccinno G, Drew DA, Leeming E, Gibson R, Le Roy C, Khatib HA, et al. Microbiome connections with host metabolism and habitual diet from 1,098 deeply phenotyped individuals. Nat Med. 2021;27(2):321–332. doi:10.1038/s41591-020-01183-8.
23. Bergman EN. Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. Physiol Rev. 1990;70(2):567–590. doi:10.1152/physrev.1990.70.2.567.
24. Rasmussen HS, Holtug K, Mortensen PB. Degradation of amino acids to short-chain fatty acids in humans: An in vitro study. Scand J Gastroenterol. 1988;23(2):178–182. doi:10.3109/00365528809103964.
25. Campos-Perez W, Martinez-Lopez E. Effects of short chain fatty acids on metabolic and inflammatory processes in human health. Biochim Biophys Acta Mol Cell Biol Lipids. 2021;1866(5):158900. doi:10.1016/j.bbalip.2021.158900.
26. Porter CJ, Moppett IK, Juurlink I, Nightingale J, Moran CG, Devonald MA. Acute and chronic kidney disease in elderly patients with hip fracture: prevalence, risk factors and outcome with development and validation of a risk prediction model for acute kidney injury. BMC Nephrol. 2017;18(1):1–11. doi:10.1186/s12882-017-0437-5.
27. Bankhead-Kendall B, Radpour S, Luftman K, Guerra E, Ali S, Getto C, Brown CVR. Rib fractures and mortality: Breaking the causal relationship. Am Surg. 2019;85(11):1224–1227. doi:10.1177/000313481908501127.
28. Foss NB, Kehlet H. Mortality analysis in hip fracture patients: implications for design of future outcome trials. Br J Anaesth. 2005;94(1):24–29. doi:10.1093/bja/aei010.
29. Bhandari M, Swiontkowski M, Solomon CG. Management of acute hip fracture. N Engl J Med. 2017;377(21):2053–2062. doi:10.1056/NEJMcp1611090.
30. Moayyeri A, Hammond CJ, Valdes AM, Spector TD. Cohort profile: TwinsUK and healthy ageing twin study. Int J Epidemiol. 2013;42(1):76–85. doi:10.1093/ije/dyr207.
31. Evans AM, DeHaven CD, Barrett T, Mitchell M, Milgram E. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. Anal Chem. 2009;81(16):6656–6667. doi:10.1021/ac901536h.
32. Visconti A, Le Roy CI, Rosa F, Rossi N, Martin TC, Mohney RP, Li W, de Rinaldis E, Bell JT, Venter JC, et al. Interplay between the human gut microbiome and

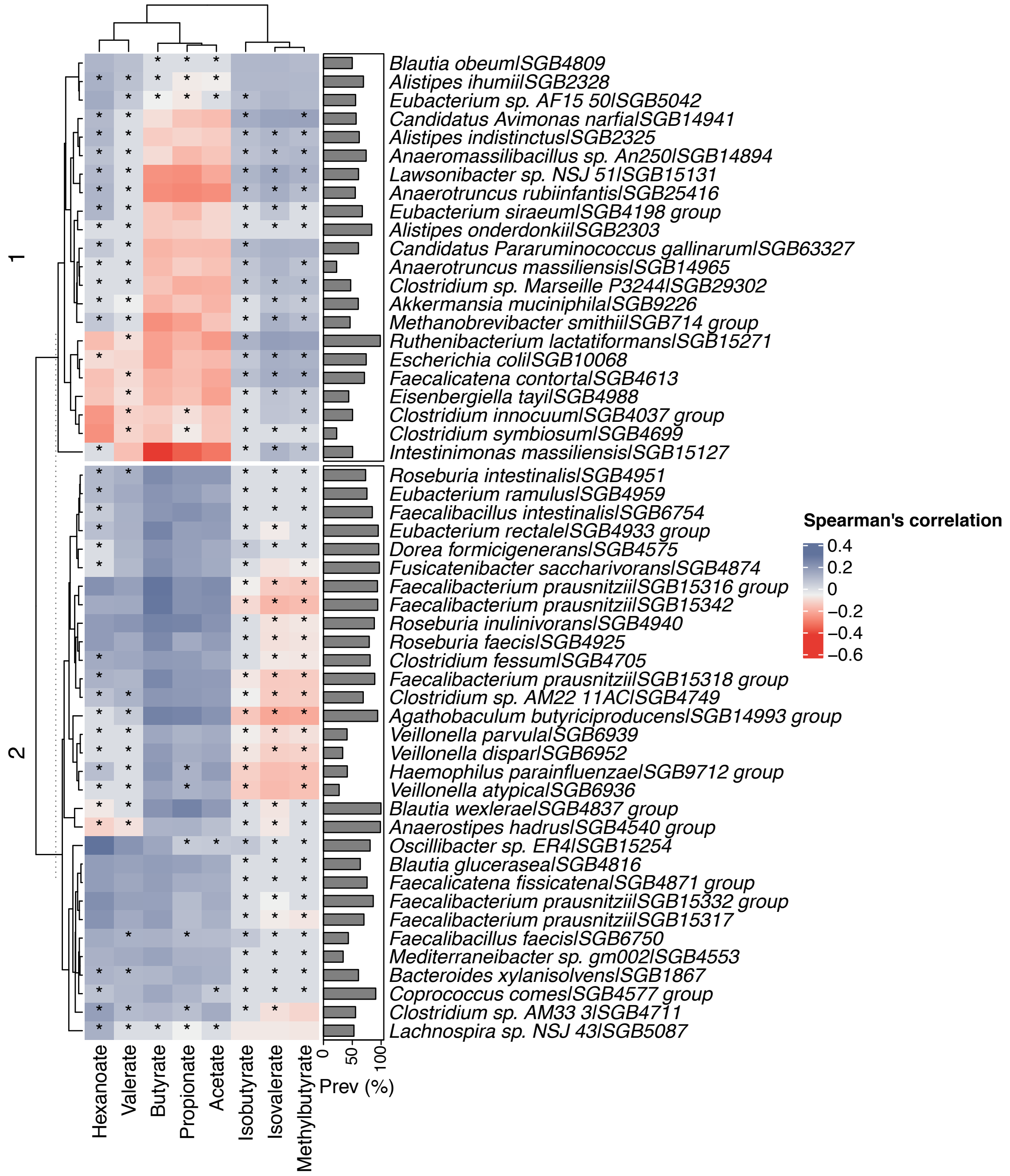
host metabolism. Nat Commun. 2019;10(1):4505. doi:10.1038/s41467-019-12476-z.

33. Visconti A, Martin TC, Falchi M. YAMP: a containerized workflow enabling reproducibility in metagenomics research. Gigascience. 2018;7(7). doi:10.1093/gigascience/giy072.
34. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. Nat Biotechnol. 2017;35(9):833–844. doi:10.1038/nbt.3935.
35. McIver LJ, Abu-Ali G, Franzosa EA, Schwager R, Morgan XC, Waldron L, Segata N, Huttenhower C. bioBakery: a meta'omic analysis environment. Bioinformatics. 2018;34(7):1235–1237. doi:10.1093/bioinformatics/btx754.
36. Beghini F, McIver LJ, Blanco-Míguez A, Dubois L, Asnicar F, Maharjan S, Mailyan A, Manghi P, Scholz M, Thomas AM, et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. eLife. 2021;10. doi:10.7554/eLife.65088.
37. Blanco-Miguez A, Beghini F, Cumbo F, McIver LJ, Thompson KN, Zolfo M, Manghi P, Dubois L, Huang KD, Thomas, AM, et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species with MetaPhlAn 4. Nat Biotechnol. 2023. doi:10.1038/s41587-023-01688-w.
38. Thissen D, Steinberg L, Kuang D. Quick and easy implementation of the Benjamini- Hochberg procedure for controlling the false positive rate in multiple comparisons. J Educ Behav Stat. 2002;27(1):77–83. doi:10.3102/10769986027001077.
39. Neale M, Cardon LR. Methodology for genetic studies of twins and families. Dordrecht: Kluwer Academic Publishers; 1992.
40. Scheike TH, Holst KK, Hjelmborg JB. Estimating heritability for cause specific mortality based on twin studies. Lifetime Data Anal. 2014;20(2):210–233. doi:10.1007/s10985-013-9244-x.
41. Nelli F. Machine learning with scikit-learn. Python Data Anal. Berkeley, CA: Apress; 2015. p. 237–264.
42. Pasolli E, Truong DT, Malik F, Waldron L, Segata N, Eisen JA. Machine learning meta-analysis of large metagenomic datasets: Tools and biological insights. PLoS Comput Biol. 2016;12(7):e1004977. doi:10.1371/journal.pcbi.1004977.

Supplementary material



Supplementary Fig. 7.1 Spearman's correlations between age and BMI, and SCFAs in serum and stool in participants from the TwinsUK and ZOE PREDICT-1 cohorts. Significant correlations (FDR <0.05) are indicated with an asterix.



Supplementary Fig. 7.2 Partial Spearman's correlations between abundances of single gut microbial species and faecal SCFAs levels for 1178 individuals from TwinsUK and ZOE PREDICT-1. Correlations were adjusted for age, BMI and sex. Characterised species with a prevalence>20%, presenting significant correlations in the 3 datasets - TwinsUK together with ZOE PREDICT-1 (FDR<0.2), TwinsUK, and ZOE PREDICT-1 (nominal p-value=0.05) -, and with at least 3 different SCFAs were presented. Correlations that were not replicated in TwinsUK and/or ZOE PREDICT-1 (nominal p-value$\leq$0.05) are indicated with an asterisk. The species are presented using their Species-level Genome Bins (SGBs) identifiers. The species and SCFAs were hierarchically clustered (complete linkage, Euclidean distance). The horizontal bars indicate the prevalence values (%) for each species.

Supplementary Table 7.1 Postprandial changes from fasting and inter-individual variability for each SCFA assessed using Wilcoxon tests and coefficient of variation (CV-calculated as SD/mean, %).

SCFA	Postprandial measure	CV	Wilcoxon p-value
Acetate	Peak	36.8	9.6×10^{-5}
	Dip	41.6	4.1×10^{-52}
Propionate	Peak	28.2	3.6×10^{-23}
	Dip	37.9	1.3×10^{-11}
Butyrate	Peak	33.6	4.0×10^{-28}
	Dip	41.3	6.1×10^{-11}
Methylbutyrate	Peak	29.8	0.37
	Dip	36.5	1.5×10^{-35}
Isobutyrate	Peak	26.5	9.6×10^{-5}
	Dip	32	6.9×10^{-40}
Valerate	Peak	32.7	5.9×10^{-32}
	Dip	46.7	0.08
Isovalerate	Peak	25.4	1.4×10^{-17}
	Dip	33.8	1.2×10^{-18}
Hexanoate	Peak	39.8	0.37
	Dip	39.2	2.1×10^{-27}

Supplementary Table 7.2 Associations between postprandial SCFA levels and postprandial lipaemic and glycaemic parameters in ZOE PREDICT-1 participants. The beta estimates and p-values (FDR) (shown in parenthesis) obtained from linear mixed models adjusted for age, BMI and family relatedness are reported.

SCFA	Postprandial measure	Triglycerides 6h-rise	Glucose iAUC0-2h	C-peptide 2h-rise	Insulin 2h-rise
Acetate	Peak	0.15 (0.11)	0.04 (0.91)	0.02 (0.95)	0.08 (0.68)
	Dip	0.22 (0.001)	-0.2 (0.005)	-0.01 (0.97)	0.06 (0.64)
Propionate	Peak	0.01 (0.95)	0.15 (0.11)	0.08 (0.7)	0.04 (0.91)
	Dip	0.06 (0.63)	0.15 (0.07)	0.12 (0.13)	0.07 (0.54)
Butyrate	Peak	-0.06 (0.91)	0.14 (0.11)	0.06 (0.91)	0.04 (0.91)
	Dip	0.03 (0.93)	0.14 (0.08)	0.12 (0.11)	0.07 (0.55)
Methylbutyrate	Peak	0.06 (0.91)	0.04 (0.91)	0.05 (0.91)	0.04 (0.91)
	Dip	0.02 (0.94)	-0.08 (0.5)	-0.03 (0.93)	-0.03 (0.93)
Isobutyrate	Peak	0.03 (0.91)	-0.01 (0.97)	0 (0.97)	0 (0.97)
	Dip	0.01 (0.98)	-0.07 (0.54)	0 (0.98)	-0.02 (0.94)
Valerate	Peak	-0.03 (0.91)	0.14 (0.14)	0.02 (0.95)	0.05 (0.91)
	Dip	-0.03 (0.93)	0.14 (0.08)	0.03 (0.93)	0.02 (0.94)
Isovalerate	Peak	-0.02 (0.95)	-0.03 (0.91)	0.02 (0.95)	0 (0.97)
	Dip	0 (0.98)	-0.16 (0.04)	-0.02 (0.94)	-0.01 (0.94)
Hexanoate	Peak	-0.03 (0.95)	0.07 (0.83)	0.01 (0.95)	-0.01 (0.97)
	Dip	-0.03 (0.93)	0.05 (0.69)	0 (0.98)	-0.04 (0.81)

Supplementary Table 7.3 Influence of the gut microbiota composition in faecal and circulating SCFA levels estimated by Random Forest regression (using Spearman's correlations) and classification (using AUC) models. The median AUC and the 95% confidence intervals across 100 folds for a corresponding binary classifier between the highest and lowest quartile, and the median values and the 95% confidence intervals of the Spearman's correlation between the real value of each component and the value predicted by regression models across 100 training/testing folds are shown.

Sample	SCFA	Cohort	AUC	AUC 95% CI	Spearman's rho	Spearman's rho 95% CI
Serum	Acetate	TwinsUK	0.51	0.51,0.53	0.07	0.06,0.09
		ZOE PREDICT-1	0.57	0.56,0.59	0.05	0.02,0.07
	Propionate	TwinsUK	0.51	0.51,0.53	0.03	0,0.03
		ZOE PREDICT-1	0.67	0.65,0.68	0.17	0.15,0.19
	Butyrate	TwinsUK	0.6	0.59,0.61	0.06	0.04,0.07
		ZOE PREDICT-1	0.61	0.6,0.63	0.12	0.09,0.14
	Methylbutyrate	TwinsUK	0.6	0.59,0.61	0.1	0.08,0.11
		ZOE PREDICT-1	0.52	0.5,0.54	0	-0.01,0.03
	Isobutyrate	TwinsUK	0.56	0.56,0.58	0.08	0.07,0.09
		ZOE PREDICT-1	0.52	0.5,0.53	0.01	0,0.04
	Valerate	TwinsUK	0.61	0.61,0.63	0.14	0.13,0.16
		ZOE PREDICT-1	0.59	0.57,0.6	0.08	0.05,0.1
	Isovalerate	TwinsUK	0.5	0.48,0.51	-0.02	-0.04,-0.01
		ZOE PREDICT-1	0.57	0.54,0.58	0.05	0.04,0.08
	Hexanoate	TwinsUK	0.63	0.62,0.64	0.19	0.16,0.19
		ZOE PREDICT-1	0.56	0.53,0.57	0.02	-0.01,0.04
Stool	Acetate	TwinsUK	0.82	0.81,0.82	0.43	0.41,0.43
		ZOE PREDICT-1	0.91	0.9,0.92	0.59	0.56,0.59
	Propionate	TwinsUK	0.82	0.81,0.82	0.47	0.45,0.48
		ZOE PREDICT-1	0.93	0.91,0.93	0.62	0.6,0.63
	Butyrate	TwinsUK	0.86	0.85,0.86	0.55	0.54,0.55
		ZOE PREDICT-1	0.91	0.89,091	0.61	0.59,0.62
	Methylbutyrate	TwinsUK	0.78	0.77,0.79	0.41	0.39,0.42
		ZOE PREDICT-1	0.64	0.62,0.66	0.19	0.15,0.2
	Isobutyrate	TwinsUK	0.75	0.74,0.76	0.33	0.31,0.34
		ZOE PREDICT-1	0.6	0.58,0.62	0.1	0.08,0.13
	Valerate	TwinsUK	0.75	0.73,0.75	0.35	0.34,0.36
		ZOE PREDICT-1	0.78	0.75,0.78	0.33	0.31,0.35
	Isovalerate	TwinsUK	0.78	0.78,0.79	0.42	0.4,0.43
		ZOE PREDICT-1	0.66	0.64,0.67	0.23	0.2,0.25
	Hexanoate	TwinsUK	0.83	0.83,0.84	0.46	0.45,0.47
		ZOE PREDICT-1	0.82	0.8,0.83	0.39	0.37,0.42

Supplementary Table 7.4 Demographic characteristics of the participants from the subset of TwinsUK with measurements of circulating SCFAs and cytokines, and the acute trauma case-control cohort.

Cohort	Type	n	Females, (%)	Age, yrs
Acute trauma case-control	Healthy (controls)	21	55%	38.7 (14.97)
	Rib fracture	18	38%	59.6 (16.18)
	Hip fracture	32	80%	88.7 (5.03)
TwinsUK	Healthy	82	100%	67.6 (10.9)

Supplementary Table 7.5 Associations between circulating SCFA levels and fracture in individuals from the acute trauma case-control cohort. Results are presented without adjusting and after adjusting for age and sex.

		Without adjusting			Adjusting for age and sex		
SCFA	**Compared groups**	**Beta**	**SE**	**P-value**	**Beta**	**SE**	**P-value**
Acetate	Hip-Control	1.51	0.23	0	1.82	0.36	0
	Rib-Control	0.39	0.26	0.14	0.61	0.32	0.06
	Rib-Hip	-1.11	0.23	0	-1.21	0.26	0
Propionate	Hip-Control	0.75	0.28	0.01	1.11	0.43	0.01
	Rib-Control	0.79	0.32	0.02	1.13	0.39	0
	Rib-Hip	0.04	0.28	0.88	0.03	0.32	0.93
Butyrate	Hip-Control	0.07	0.31	0.83	0.34	0.46	0.47
	Rib-Control	0.13	0.34	0.71	0.44	0.41	0.29
	Rib-Hip	0.06	0.31	0.84	0.1	0.34	0.76
Methylbutyrate	Hip-Control	-0.52	0.28	0.07	-0.51	0.43	0.24
	Rib-Control	0.48	0.31	0.13	0.46	0.38	0.24
	Rib-Hip	1	0.28	0	0.97	0.32	0
Isobutyrate	Hip-Control	0.12	0.3	0.69	0.19	0.47	0.69
	Rib-Control	0.2	0.34	0.55	0.15	0.42	0.71
	Rib-Hip	0.08	0.3	0.79	-0.03	0.35	0.92
Valerate	Hip-Control	0.63	0.3	0.04	1.14	0.44	0.01
	Rib-Control	0.36	0.33	0.29	0.82	0.4	0.04
	Rib-Hip	-0.27	0.3	0.36	-0.32	0.33	0.33
Isovalerate	Hip-Control	0.66	0.28	0.02	0.84	0.44	0.06
	Rib-Control	0.95	0.32	0	1	0.39	0.01
	Rib-Hip	0.28	0.28	0.32	0.16	0.32	0.63
Hexanoate	Hip-Control	0.02	0.3	0.96	0.19	0.47	0.69
	Rib-Control	0.08	0.34	0.82	0.14	0.42	0.75
	Rib-Hip	0.06	0.3	0.84	-0.05	0.35	0.88

Supplementary Text 7.1 Full details and quality control of the SCFA measurements.

Human serum and stool samples were spiked with stable labelled internal standards, homogenized and subjected to protein precipitation with an organic solvent. After centrifugation, an aliquot of the supernatant is derivatized. The reaction mixture was injected onto an Agilent 1290/AB Sciex QTrap 5500 LC MS/MS system equipped with a C18 reversed-phase UHPLC column. The mass spectrometer is operated in negative mode using electrospray ionization (ESI). The peak area of the individual analyte product ions was measured against the peak area of the product ions of the corresponding internal standards. Quantitation was performed using a weighted linear least squares regression analysis generated from fortified calibration standards prepared immediately prior to each run. LC-MS/MS raw data were collected and processed using AB SCIEX software Analyst 1.6.3 and processed using SCIEX OS-MQ software v1.7

Sample analyses were carried out in a 96-well plate format containing two calibration curves. Accuracy was evaluated using the corresponding QC replicates in the sample runs. QCs met acceptance criteria at all levels for all analytes (QC acceptance criteria: At least 50% of QC samples at each concentration level per analyte should be within ±20.0% of the corresponding historical mean, and at least 2/3 of all QC samples per analyte should fall within ±20.0% of the corresponding historical mean).

Chapter 8

Mediatory effect of acetate between the gut microbiome and visceral fat

As discussed in the introduction (**Chapter 1**), acetate is one of the major SCFAs, and which has been associated with different cardiometabolic traits. However, integrating different types of data is necessary to gain further insights into the host-microbial cross-talk involving its circulating levels and its implications in CMD.

In this chapter, I assess the associations between circulating acetate levels, gut microbiome composition and diversity, and visceral fat. Furthermore, I explore the phylogenetic diversity and metabolic complexity of the identified acetate-associated gut genera by performing genomic analyses.

The obtained results show the beneficial effects of circulating acetate on visceral fat, and its mediatory role in the influence of the gut microbiome with visceral fat. Moreover, the findings highlight the role of different gut microbiome species in CMD.

Collaborator Dr Philippa M. Wells cleaned and generated the amplicon sequence variants for the gut microbiome data. I performed the statistical analyses and wrote the original draft of the manuscript.

This chapter has been published in *Frontiers in Microbiology* (Nogal *et al.*, 2021).

frontiers in Microbiology

ORIGINAL RESEARCH
published: 15 July 2021
doi: 10.3389/fmicb.2021.711359

Circulating Levels of the Short-Chain Fatty Acid Acetate Mediate the Effect of the Gut Microbiome on Visceral Fat

Ana Nogal[1†], Panayiotis Louca[1†], Xinyuan Zhang[1], Philippa M. Wells[1], Claire J. Steves[1], Tim D. Spector[1†], Mario Falchi[1], Ana M. Valdes[1,2†] and Cristina Menni[1†]*

[1] Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom, [2] Nottingham NIHR Biomedical Research Centre at the School of Medicine, Nottingham City Hospital, University of Nottingham, Nottingham, United Kingdom

OPEN ACCESS

Edited by:
Hui Wu,
East China University of Science and Technology, China

Reviewed by:
Congying Chen,
Jiangxi Agricultural University, China
Katie Lynn Summers,
United States Department of Agriculture (USDA), United States

****Correspondence:***
Cristina Menni
cristina.menni@kcl.ac.uk

†ORCID:
Ana Nogal
orcid.org/0000-0002-0973-4313
Panayiotis Louca
orcid.org/0000-0001-5956-1433
Tim D. Spector
orcid.org/0000-0002-9795-0365
Ana M. Valdes
orcid.org/0000-0003-1141-4471
Cristina Menni
orcid.org/0000-0001-9790-0571

Specialty section:
This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology

Received: *18 May 2021*
Accepted: *21 June 2021*
Published: *15 July 2021*

Citation:
Nogal A, Louca P, Zhang X, Wells PM, Steves CJ, Spector TD, Falchi M, Valdes AM and Menni C (2021) Circulating Levels of the Short-Chain Fatty Acid Acetate Mediate the Effect of the Gut Microbiome on Visceral Fat. Front. Microbiol. 12:711359. doi: 10.3389/fmicb.2021.711359

Background: Acetate is a short-chain fatty acid (SCFA) produced by gut bacteria, which has been implicated in cardio-metabolic health. Here we examine the relationships of circulating acetate levels with gut microbiome composition and diversity and with visceral fat in a large population-based cohort.

Results: Microbiome alpha-diversity was positively correlated with circulating acetate levels (Shannon, Beta [95%CI] = 0.12 [0.06, 0.18], $P = 0.002$) after adjustment for covariates. Six serum acetate-associated bacterial genera were also identified, including positive correlations with *Coprococcus*, *Barnesiella*, *Ruminococcus*, and *Ruminococcaceae NK4A21* and negative correlations were observed with *Lachnoclostridium* and *Bacteroides.* We also identified a correlation between visceral fat and serum acetate levels (Beta [95%CI] = −0.07 [−0.11, −0.04], $P = 2.8 \times 10^{-4}$) and between visceral fat and *Lachnoclostridium* (Beta [95%CI] = 0.076 [0.042, 0.11], $P = 1.44 \times 10^{-5}$). Formal mediation analysis revealed that acetate mediates ∼10% of the total effect of *Lachnoclostridium* on visceral fat. The taxonomic diversity showed that *Lachnoclostridium* and *Coprococcus* comprise at least 18 and 9 species, respectively, including novel bacterial species. By predicting the functional capabilities, we found that *Coprococcus* spp. present pathways involved in acetate production and metabolism of vitamins B, whereas we identified pathways related to the biosynthesis of trimethylamine (TMA) and CDP-diacylglycerol in *Lachnoclostridium* spp.

Conclusions: Our data indicates that gut microbiota composition and diversity may influence circulating acetate levels and that acetate might exert benefits on certain cardio-metabolic disease risk by decreasing visceral fat. *Coprococcus* may play an important role in host health by its production of vitamins B and SCFAs, whereas *Lachnoclostridium* might have an opposing effect by influencing negatively the circulating levels of acetate and being involved in the biosynthesis of detrimental lipid compounds.

Keywords: acetate, *Lachnoclostridium*, *Coprococcus*, human gut microbiota, visceral fat

INTRODUCTION

Acetate is a short-chain fatty acid (SCFA) produced by colonic bacteria through the saccharolytic fermentation of fibres (e.g., resistant starch, polysaccharides and simple sugars), which escape digestion and absorption (Topping and Clifton, 2001). The molar ratio of acetate in the colon is three times larger than that of the two other major SCFAs, butyrate and propionate (Cummings et al., 1987). Enteric bacteria, including *Ruminococcus* spp., *Prevotella* spp., *Bifidobacterium* spp., and *Akkermansia muciniphila* are suggested to be the main acetate-producing bacteria (Rey et al., 2010).

Recently SCFAs have received increasing attention as they have been shown to play an important role in cardio-metabolic diseases (CMD), including obesity, type-2 diabetes (T2D), arterial stiffness and atherosclerosis (Den Besten et al., 2013). Once these bacteria-derived metabolites are synthetised, they have the capacity to reach different systematic tissues, improving the gut barrier integrity, glucose, cholesterol and lipid metabolism, and regulating the immune system and anti-inflammatory response, energy intake, and blood pressure (Martin-Gallausiaux et al., 2020). For instance, acetate was shown to decrease appetite by impacting directly on the hypothalamus (Frost et al., 2014), inhibit endogenous lipolysis (Hron et al., 1978), enhance hepatic uptake of blood cholesterol (Zhao Y. et al., 2017) and reduce hyperglycaemia (Sakakibara et al., 2006). However, to gain further insight into the host-microbial cross-talk involving circulating acetate levels and its implications in cardio-metabolic health (CMH), it is important to integrate different types of data.

In this study, we analyzed the associations between circulating acetate levels, gut microbiome composition and diversity and visceral fat in a cohort of 948 women from TwinsUK. Furthermore, by performing genomic analyses, we have explored the phylogenetic diversity and metabolic complexity of the acetate-associated gut genera.

MATERIALS AND METHODS

Study Subjects

Study subjects were female twins enrolled in the TwinsUK registry, a national register of adult twins recruited as volunteers without selecting for any particular disease or trait (Moayyeri et al., 2013). In this study, we analyzed data from 948 female twins with concurrent measures of 16S gut microbiome composition, serum acetate levels and visceral fat. The study was approved by NRES Committee London–Westminster, and all twins provided informed written consent. A flowchart of the study design is presented in **Figure 1A**.

Measurements

Microbiome Analysis

Fecal samples were collected and the composition of the gut microbiome was determined by 16S rRNA gene sequencing carried out as previously described (Goodrich et al., 2016). Briefly, the V4 region of the 16S rRNA gene was amplified and sequenced on Illumina MiSeq. 16S sequences were demultiplexed in QIIME 1 (Caporaso et al., 2010). The following analyses were conducted in RStudio version 1.3.1093. Amplicon sequence variants (ASV) were then generated using the "DADA2" R package following the pipeline described by Wells and colleagues (Wells et al., 2020). The ASVs were grouped into genera and the samples with less than 10,000 reads were discarded. The indices of microbiome alpha-diversity, quantified as Shannon, inverse Simpson, Gini Simpson diversity, CHAO1 and number of observed ASVs were calculated using the "microbiome" package (Lahti and Shetty, 2018).

Acetate Measure

Circulating levels of acetate were measured from serum by Nightingale Health Ltd. (Helsinki, Finland; previously known as Brainshake Ltd.) using a targeted NMR spectroscopy platform that has been extensively applied for biomarker profiling in epidemiological studies (Würtz et al., 2015) as previously described (Barrios et al., 2018).

Visceral Fat Measure

Measurements of whole body composition were performed for 948 female twins aged 48 to 87 years using the DXA fan-beam technology (Hologic QDR; Hologic, Inc., Waltham, MA, United States) as was indicated by Menni and colleagues (Menni et al., 2016). This DXA-based measurement has been validated against VF measured by CT scan (Kaul et al., 2012) and shown to be reliable and reproducible.

Briefly, subjects were positioned in a supine position wearing only a gown. The DXA machine was calibrated following the manufacturer's suggestions. The scans were analyzed using the QDR System Software v12.6. Regions of interest were defined manually by the same operator following the SOP (derived from the manufacturer's guidelines). The lower and upper horizontal margins were placed just above the iliac crest and at the half of the distance between the acromions and the iliac crest, respectively. The vertical margins were adjusted at the external body borders so that all the soft tissue was included.

Fibre Intake

A validated 131-item semi-quantitative Food Frequency Questionnaire (FFQ) established for the EPIC (European Prospective Investigations into Cancer and Nutrition)-Norfolk study (Bingham et al., 2001) was used to assess dietary intake. Estimated intakes of fiber (in grams per day) were derived from the UK Nutrient Database (McCance and Widdowson, 2014) and were adjusted for energy intake using the residual method prior to analysis (Willett and Stampfer, 1986).

Abbreviations: ANI, average nucleotide identity; ASV, amplicon sequence variants; CMD, cardio-metabolic diseases; CMH, cardio-metabolic health; CVD, cardiovascular diseases; KEGG, Kyoto Encyclopedia of Genes and Genomes; MAG, metagenome-assembled genomes; MetaCyc, Metabolic Pathway Database; NCBI, National Center for Biotechnology Information; SCFA, short-chain fatty acids; T2D, type-2 diabetes; TMA, trimethylamine; TMAO, trimethylamine-N-oxide; UHGG, Unified Human Gastrointestinal Genome.

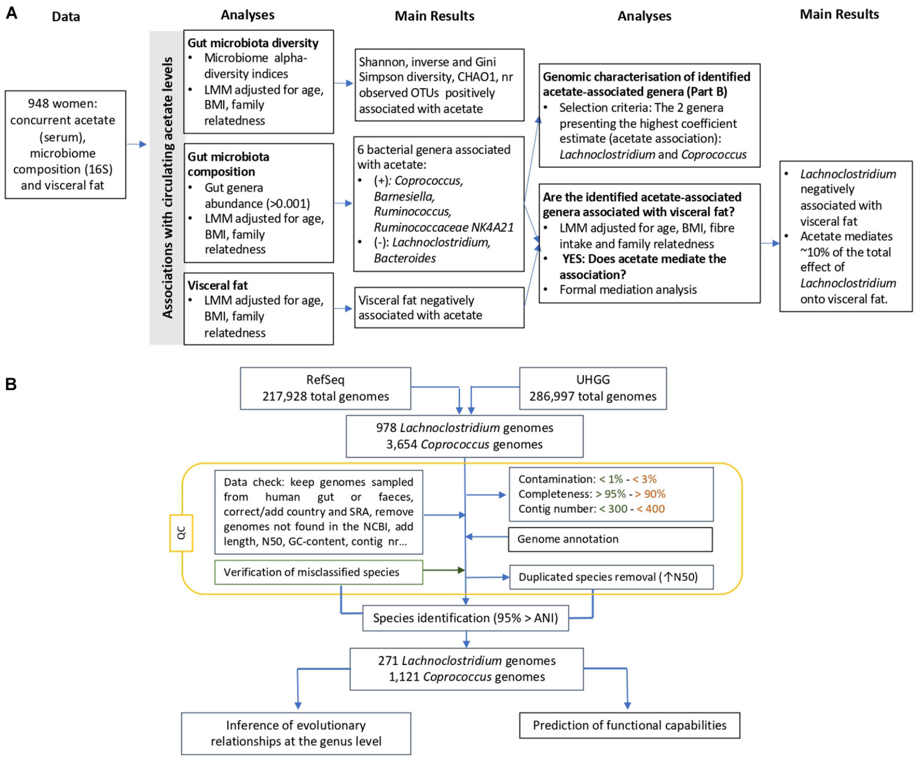

A
Data
Analyses
Main Results
Analyses
Main Results
948 women: concurrent acetate (serum), microbiome composition (16S) and visceral fat
Associations with circulating acetate levels
Gut microbiota diversity
• Microbiome alpha-diversity indices
• LMM adjusted for age, BMI, family relatedness
Gut microbiota composition
• Gut genera abundance (>0.001)
• LMM adjusted for age, BMI, family relatedness
Visceral fat
• LMM adjusted for age, BMI, family relatedness
Shannon, inverse and Gini Simpson diversity, CHAO1, nr observed OTUs positively associated with acetate
6 bacterial genera associated with acetate:
• (+): Coprococcus, Barnesiella, Ruminococcus, Ruminococcaceae NK4A21
• (-): Lachnoclostridium, Bacteroides
Visceral fat negatively associated with acetate
Genomic characterisation of identified acetate-associated genera (Part B)
• Selection criteria: The 2 genera presenting the highest coefficient estimate (acetate association): Lachnoclostridium and Coprococcus
Are the identified acetate-associated genera associated with visceral fat?
• LMM adjusted for age, BMI, fibre intake and family relatedness
• YES: Does acetate mediate the association?
• Formal mediation analysis
• Lachnoclostridium negatively associated with visceral fat
• Acetate mediates ~10% of the total effect of Lachnoclostridium onto visceral fat.
B
RefSeq 217,928 total genomes
UHGG 286,997 total genomes
978 Lachnoclostridium genomes 3,654 Coprococcus genomes
QC
Data check: keep genomes sampled from human gut or faeces, correct/add country and SRA, remove genomes not found in the NCBI, add length, N50, GC-content, contig nr...
Verification of misclassified species
Contamination: < 1% - < 3%
Completeness: > 95% - > 90%
Contig number: < 300 - < 400
Genome annotation
Duplicated species removal (↑N50)
Species identification (95% > ANI)
271 Lachnoclostridium genomes 1,121 Coprococcus genomes
Inference of evolutionary relationships at the genus level
Prediction of functional capabilities

FIGURE 1 | Overview of the flowchart **(A)** integrating gut microbiota composition and diversity, visceral fat and circulating acetate levels, and **(B)** showing the applied filters and conducted steps to genomically characterize the *Lachnoclostridium* spp. and *Coprococcus* spp. Steps exclusively applied for *Lachnoclostridium* and *Coprococcus* are indicated in orange and green, respectively, whereas the rest of steps were conducted in the genomes from both species. ANI, average nucleotide identity; LMM, linear mixed model; UHGG, Unified Human Gastrointestinal Genome; QC, quality control.

Statistical Analyses

Statistical analyses were conducted in RStudio version 1.3.1093. We assessed the association between circulating acetate and (i) indices of alpha-diversity (Shannon, inverse Simpson, Gini Simpson, CHAO1, and number of observed OTUs), (ii) gut bacterial genera abundance (genera with abundance >0.001), (iii) visceral fat using linear mixed model adjusting for age, BMI, family relatedness and multiple testing using false discovery rate [Benjamini and Hochberg (Thissen et al., 2002)]. Indices of alpha diversity were also adjusted for sequencing depth. Then, linear mixed models were further employed to investigate the association between visceral fat and any acetate-associated genera. All variables included in the models were Z-score normalized.

Finally, we employed mediation analysis as implemented in the R package "mediation" (Tingley et al., 2014) with 1,000 Monte Carlo draws for a quasi-Bayesian approximation, to test the mediation effects of acetate (indirect effect) on the total effect of *Lachnoclostridium* on visceral fat adjusting for BMI, age and fiber intake. We constructed a mediation model to quantify both the direct effect *Lachnoclostridium* on visceral fat and the indirect (mediated) effects mentioned above. The variance accounted for (VAF) score, which represents the ratio of indirect-to-total effect and determines the proportion of the variance explained by the mediation process, was further used to determine the significance of mediation effect.

Genomic Characterization of the Identified Acetate-Associated Gut Genera

A flowchart of the steps conducted for the genomic characterisation is presented in **Figure 1B**.

Selection of Genome Sequences and Preliminary Filtering

Genomes belonging to the acetate-associated gut genera (*Lachnoclostridium* and *Coprococcus)* and their corresponding metadata were obtained from the UHGG catalog and RefSeq dataset (January, 2021), respectively (Almeida et al., 2020). We removed the RefSeq genomes derived from metagenomes and not sampled from human faeces, stool or the gastrointestinal tract, Inconsistencies related to the variable country were corrected and the missing sample accessions were added. Genomes from sample identifiers not found in the National Center for Biotechnology Information (NCBI) (Sayers et al., 2019) were discarded. The two datasets were merge and we then filtered by completeness, contamination and number of contigs (>90%, <3%, and <400 for *Lachnoclostridium* and >95%, <1%, and <300 for *Coprococcus*). The thresholds in *Lachnoclostridium* were less strict due to the scarcity of genomes presenting higher standards. Duplicated genomes were discarded, keeping the one with the highest N50 value. In total, we downloaded 271 *Lachnoclostridium* and 1,121 *Coprococcus* high-quality genomes (**Supplementary Table 1**). Finally, genomes from uncharacterized species or misclassified species were renamed based on the cluster given by fastANI classification (see section "Materials and Methods").

Quality Assessment of Genome Assemblies and Genome Annotation

Completeness and contamination were estimated with CheckM version 1.1.3 (Parks et al., 2015) using the "lineage_wf" workflow. QUAST version 5.0.2 (Gurevich et al., 2013) was run to retrieve the total length, GC-content, contig number and N50. Genome annotation was performed using Prokka version 1.12 (Seemann, 2014) using the default parameters.

Average Nucleotide Identity-Based Taxonomic Classification

FastANI version 1.32 (Jain et al., 2018) was separately run on *Lachnoclostridium* and *Coprococcus* genomes to calculate the average nucleotide identity (ANI) between all pairs of sequences (**Supplementary Tables 2, 3, 4**). Results were filtered by the alignment fraction (>0.4), and symmetric pairwise ANI dissimilarities (100-95, ANI = 95%) were calculated from the ANI values to construct a dendrogram for each genus using the single linkage hierarchical clustering method ["hclust" R function, stats package (R Core Team and DC, 2019)]. Two networks analyses based on the information given by the dendrograms were conducted using the "layoutwithdrl" layout implemented in the "igraph" R package (Csardi and Nepusz, 2006) with an expansion and simmer attraction of 0, and an innit, liquid and crunch temperature of 100, 50, and 50, respectively.

Verification of Misclassified *Coprococcus* Species

The inconsistencies in the taxonomic classification were verified using BLASTn. For that, barrnap v0.9 was run to predict the 16S rRNA sequences of genomes from *C. eutactus, C. sp. BIOML-A2, C. sp. BIOML-A1, C. sp. NSJ-10*, *C. sp900066115*, and *C. sp000154*245. These were used as query and subject to perform a BLASTn search. The matches were filtered by 99% of identity and a query cover of 50%.

Phylogeny Inference at the Genus Level

Evolutionary relationships among the *Coprococcus* and *Lachnoclostridium* species were inferred using ezTree version 0.1 (Wu, 2018). For each species, up to three genomes (depending on the number of available genomes) sequenced from isolates were used as input. If genomes sequenced from isolates were not available, then the metagenome-assembled genomes (MAGs) with the highest completeness percentage were selected.

Prediction of the Functional Capabilities of *Coprococcus* spp. and *Lachnoclostridium* spp.

Metabolic Pathway Database (Metacyc) (Caspi et al., 2018) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2014) information for each genome was retrieved using the enzyme commission numbers from the gff files generated by Prokka and MinPath (Minimal set of Pathways) (Ye and Doak, 2009; **Supplementary Table 5**). For *Coprococcus* spp, only the KEGG and MetaCyc pathways related to metabolism, and fermentation, biosynthesis and degradation, respectively, were kept. *C. sp6* was not included in the analyses due to its scarcity of genomes (n = 1). The retrieved information was utilized to construct heatmaps ["Heatmap" R function implemented in the "ComplexHeatmap" package (Gu et al., 2016)] showing the genome percentage of each species with a given pathway. For KEGG data, only the highly different pathways between species were selected (for a given pathway, at least one species has a percentage <5% and another species has a percentage >80%). Moreover, a principal component analysis (PCA) was performed using the presence/absence matrix with the MetaCyc biosynthesis/degradation pathways using the "prcomp" R function within the "stats" package. For the three major species of *Lachnoclostridium* (species with >15 genomes), only the MetaCyc pathways related to the lipid metabolism were selected and utilized to construct a heatmap as previously indicated.

RESULTS

Associations Between Circulating Acetate Levels, Gut Microbiota Composition and Diversity and Visceral Fat

The descriptive characteristics of the study participants are depicted in **Table 1**. Overall, 948 women were included, aged between 48 and 87 years, with an average BMI of 26.2 km/m^2 (SD = 4.9) and concurrent measures of serum acetate levels, 16S microbiome data and visceral fat.

As shown in **Figure 2**, circulating acetate levels were positively correlated with several measures of microbiome

TABLE 1 | Descriptive characteristics of the study population.

Phenotype	*N*	%
N	948	
Females	948	100
	Mean	**SD**
Age, years	65	7.84
BMI, km/m^2	26.25	4.90
Acetate, mmol/l (log)	−0.745	0.594
Fiber intake, gr	20.3	5.70
Visceral fat, gr	613	294
Indices of microbiome alpha-diversity		
Shannon diversity	3.8	0.505
CHAO1	230	67.3
Number of observed OTUs	224	64.5
Inverse Simpson diversity	23.1	12.2
Gini Simpson diversity	0.938	0.05

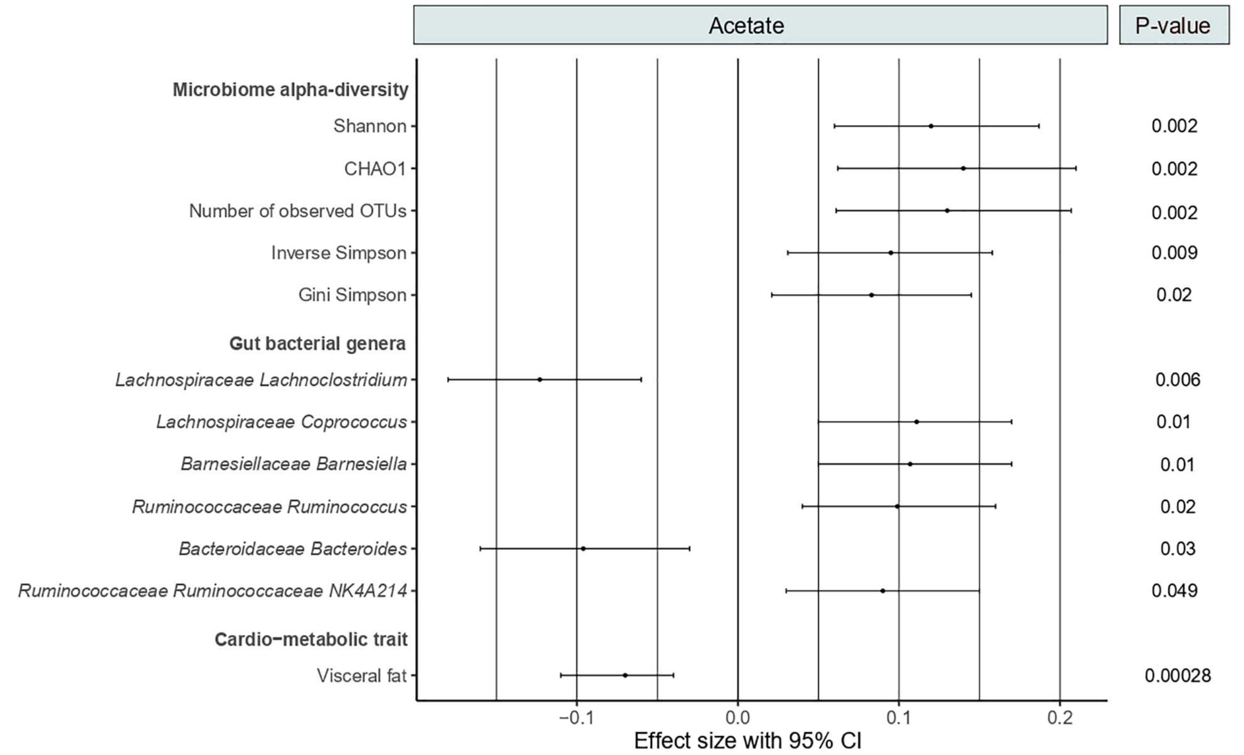


FIGURE 2 | Forest plot showing the significant associations of acetate with microbiome alpha-diversity, gut bacterial genera and visceral fat. *P*-values are FDR-adjusted.

alpha-diversity, including Shannon (Beta [95%CI] = 0.12 [0.06, 0.18], P = 0.002), CHAO1 (Beta [95%CI] = 0.14 [0.06, 0.21], P = 0.002), number of observed OTUs (Beta [95%CI] = 0.13 [0.06, 0.21], P = 0.002), inverse Simpson (Beta [95%CI] = 0.095 [0.03, 0.16], P = 0.009) and Gini Simpson (Beta [95%CI] = 0.083 [0.021, 0.15], P = 0.02). We then examined the association between acetate and bacterial genera abundances (genera with abundance >0.001). We identified six genera significantly associated with acetate levels after adjusting for age, BMI, family relatedness and multiple testing using FDR correction (FDR < 0.05) (**Figure 2**). These include *Coprococcus, Barnesiella*, *Ruminococcus*, and *Ruminococcaceae NK4A214* positively associated with acetate levels and two genera negatively associated, namely, *Lachnoclostridium* and *Bacteroides.* Among them, *Lachnoclostridium* presented the most robust association (P = 0.006).

As SCFAs exert benefits on CMH, we tested the correlation between serum levels of acetate and the cardio-metabolic trait visceral fat. We found a strong negative association between both variables (Beta [95%CI] = −0.07 [−0.11, −0.04], $P = 2.8 \times 10^{-4}$) (**Figure 2**).

We then assessed the correlation between the acetate-associated gut genera and visceral fat. We found a strong positive correlation between *Lachnoclostridium* abundances and visceral fat (Beta [95%CI] = 0.076 [0.042, 0.11], $P = 1.44 \times 10^{-5}$). No significant associations were identified for the remaining five genera. We therefore conducted a formal mediation analysis to determine the indirect effects of acetate on the total effect of *Lachnoclostridium* on visceral fat. The analysis revealed that acetate acted as a potential partial mediator in the positive association between *Lachnoclostridium* and visceral fat (VAF = 10.3%, $P = 2 \times 10^{-16}$). These associations remained significant even after adjusting for dietary fiber intake.

Among the bacterial genera identified, we then genomically characterized *Lachnoclostridium* and *Coprococcus* because they presented the largest coefficient estimates in the association with acetate (**Figure 2**).

Genomic-Based Taxonomic Classification and Phylogenetic Relationships of *Coprococcus* and *Lachnoclostridium* Species

The dendrograms created from the symmetric pairwise ANI values revealed the grouping of the 271 *Lachnoclostridium* and 1,121 *Coprococcus* genomes in 18 and 9 different species, respectively. Among them, most *Lachnoclostridium* species has been characterized (14 species), whereas *Coprococcus* presented four novel bacterial species and one has not been formally characterized so far.

In addition, we found that genomes identified as *C. sp900066115, C. sp00015424, C. sp. BIOML-A2, C. sp.*

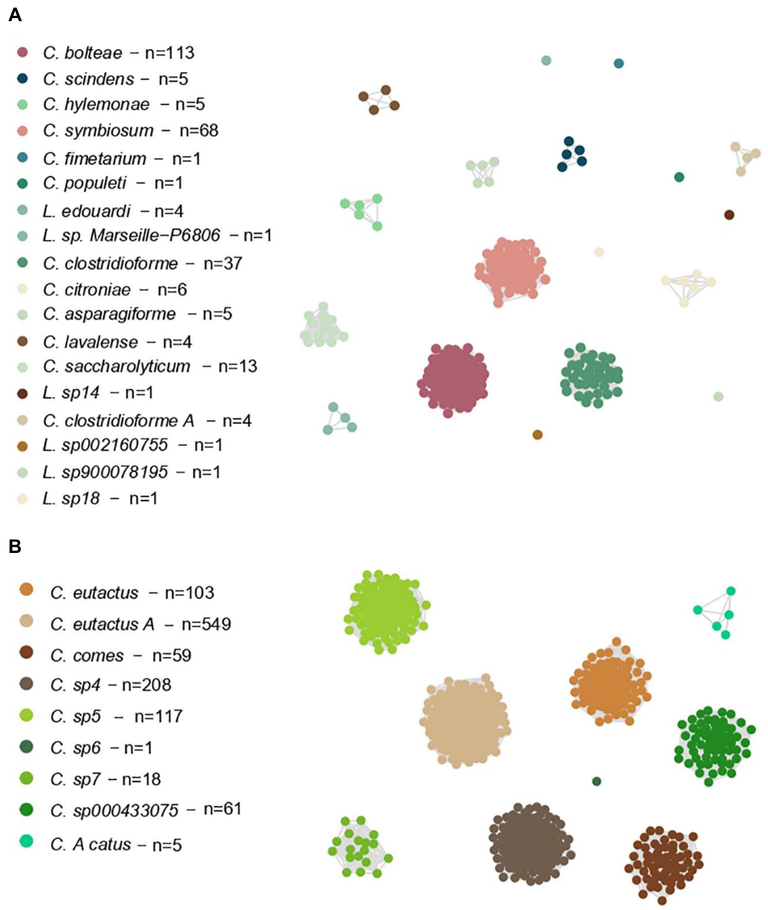


FIGURE 3 | Weighted undirected graph based on symmetric pairwise ANI values of Lachnoclostridium **(A)** and Coprococcus **(B)**. Each node represents a single genome, and each edge represents a connection between two genomes that share a mean ANI value higher than 95%. Nodes are colored by the belonging Coprococcus and Lachnoclostridium species. The species that have not been formally characterized yet, have been named using "sp" followed by the cluster number given by the dendrogram. "n" indicates the genome number of each species.

BIOML-A1, and *C. sp. NSJ-10* were assigned to the clusters of *C. eutactus, C. sp4* and *C. sp5* by the dendrogram. These misclassifications were further verified using their 16S rRNA sequences in a BLASTn search.

We also constructed two networks based on these ANI values, allowing us to study the genomes as members of a connected system (**Figure 3**). Here, the largest clusters of *Lachnoclostridium* were shown by *C. bolteae* (113 genomes, 41% of the total), *C. symbiosum* (68 genomes, 25% of the total) and *C. clostridioforme* (37 genomes, 13% of the total), whereas *C. eutactus A* (549 genomes, 49% of the total), *C. sp4* (208 genomes, 19% of the total), *C. sp5* (117 genomes, 10% of the total) and *C. eutactus* (103 genomes, 9% of the total) presented the largest clusters of *Coprococcus.*

Additionally, the computed maximum-likelihood phylogenetic trees show the existing diversity of the gut bacteria *Lachnoclostridium* and *Coprococcus* (**Figure 4**). We observed that all the identified *Lachnoclostridium* and *Coprococcus* species represented well-defined independent lineages.

Of note, the shown *Clostridium* species in Part A of **Figures 3**, **4** belong to *Lachnoclostridium* (Yutin and Galperin, 2013).

Prediction of the Functional Capabilities of *Coprococcus* spp. and *Lachnoclostridium* spp.

The percentage of genomes in which a fermentative pathway was predicted in each *Coprococcus* species is depicted in **Figure 5A**. Genomes from all the *Coprococcus* species presented fermentative pathways involved in the acetate formation from pyruvate (range = 90–100% genomes), acetoin biosynthesis (100% genomes), butanediol biosynthesis (100% genomes) and butyrate formation from acetyl-CoA (range = 90–100% genomes). The pyruvate fermentation to acetone and propionate (acrylate pathway) was exclusively present in genomes from *C. catus* (100% genomes), whereas the production of ethanol from pyruvate was mainly found in genomes from *C. comes* (80%

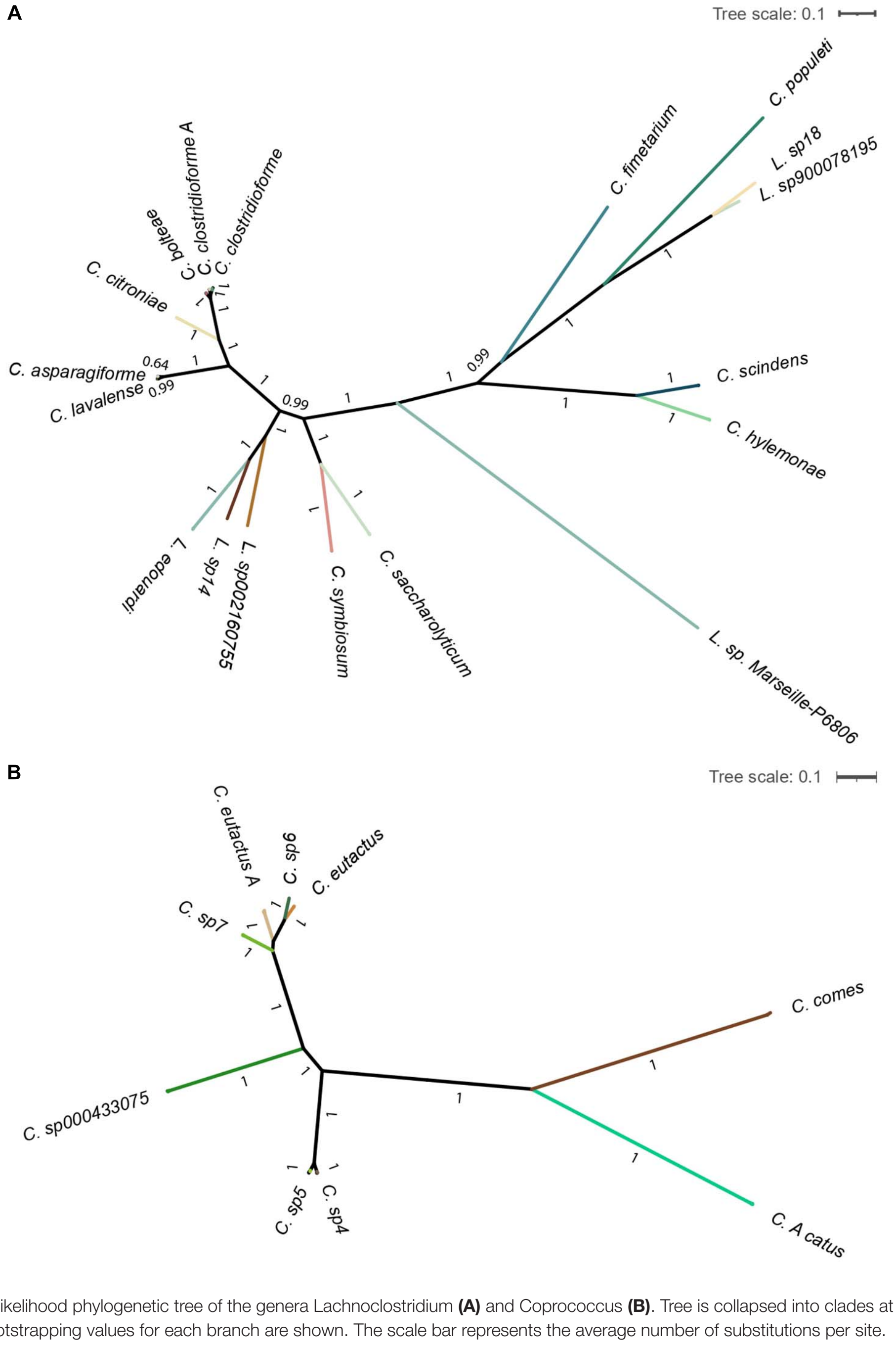


FIGURE 4 | Maximum-likelihood phylogenetic tree of the genera Lachnoclostridium **(A)** and Coprococcus **(B)**. Tree is collapsed into clades at the species level and colored by species. Bootstrapping values for each branch are shown. The scale bar represents the average number of substitutions per site.

genomes). Production of lactate from pyruvate was predicted in the latter two species (100% genomes).

The PCA performed using the presence/absence matrix with the biosynthesis and degradation pathways show each species formed a well-defined cluster, being *C. sp4* and *C. sp5,* and *C. eutactus A, C. eutactus*, and *C. sp7* closely grouped (**Figure 5B**).

As shown in the heatmap of the KEGG metabolic pathways (**Figure 5C**), differences in the functional capabilities exit between *Coprococcus* species. For instance, beta-alanine metabolism was present in all the species (∼100% genomes), except in *C. catus*, whereas chloroalkene degradation was found only in *C. catus* and *C. comes* (100% genomes). On the other hand, some metabolic pathways were present in all the *Coprococcu* species in a percentage higher than 90% (**Supplementary Table 5**). The majority of them were related to the metabolism of amino acids such as alanine, aspartate, glutamate, arginine,

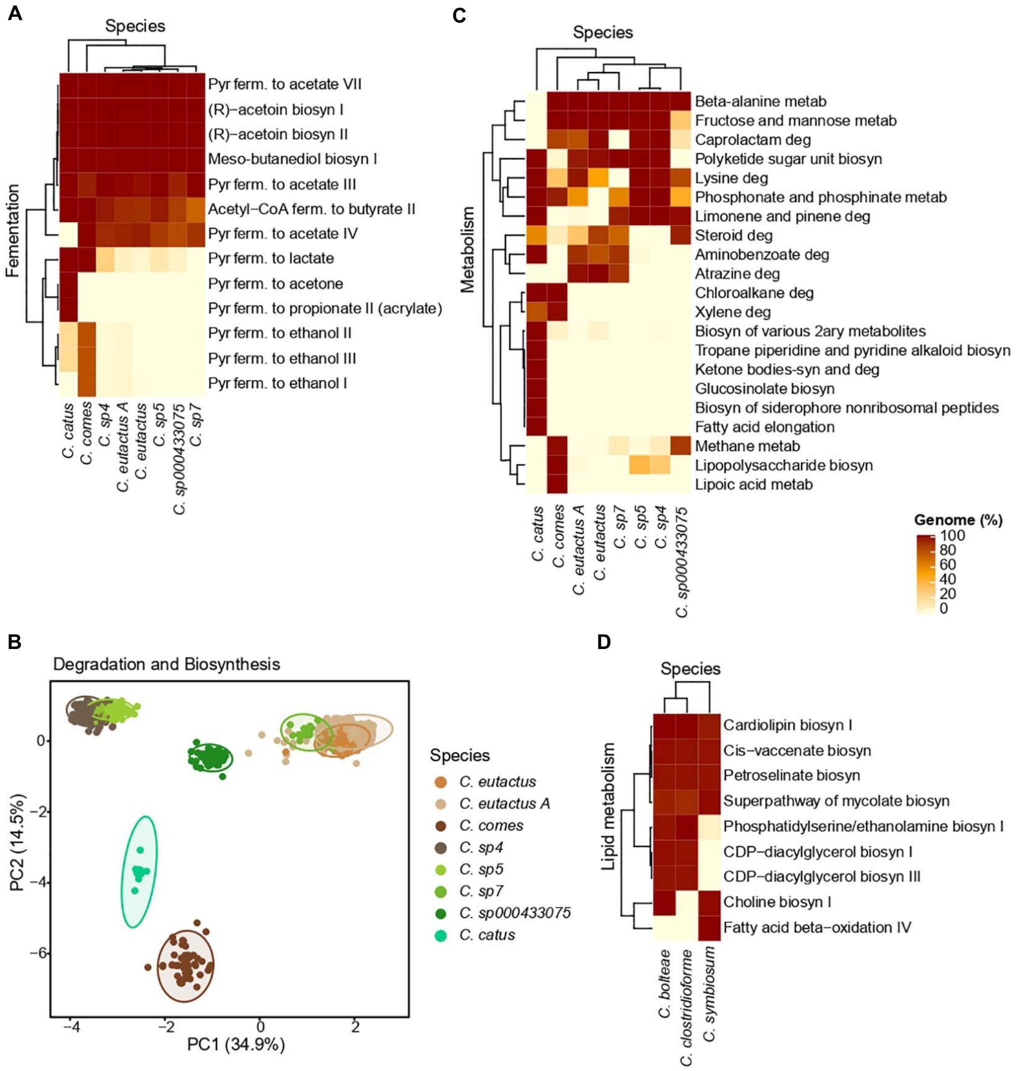


FIGURE 5 | MetaCyc and KEGG results obtained for Coprococcus spp **(A–C)** and Lachnoclostridium spp **(D)**. **(A)** Heatmap of the MetaCyc fermentative pathways. **(B)** PCA clustering based on the MetaCyc data related to the biosynthesis and degradation processes of each species. **(C)** Heatmap of the KEGG metabolic pathways. Only the pathways highly differential (for a given pathway, at least one species has an abundance value <5% and another species has a value >80%) between species are shown. **(D)** Heatmap of the MetaCyc pathways related to lipid metabolism in the three major species of Lachnoclostridium (species with >15 genomes). Of note, the shown Clostridium species belong to Lachnoclostridium (Yutin and Galperin, 2013). For all the heatmaps, the pathways and species are hierarchically clustered, and color intensities represent the percentage of genomes of a given species with a specific metabolic pathway. Pyr, pyruvate; ferm., fermentation; metab, metabolism; deg, degradation; biosyn, biosynthesis.

proline, cysteine, glycine, serine and tyrosine, as well as essential amino acids such histidine, lysine, methionine, threonine, phenylalanine and tryptophan. Additionally, all the species presented the metabolism of several vitamins B, including vitamin B1 (thiamine), B2 (riboflavin), B3 (nicotinate), B6 (pyridoxine), B7 (biotin), and B9 (folate), and pathways involved in the carbohydrate metabolism, such as the pentose phosphate pathway and the starch and sucrose metabolism.

As we found a positive association between visceral fat and *Lachnoclostridium*, we focused on the metabolic pathways related to lipid metabolism. As depicted in **Figure 5D**, a high homogeneity in the functional capabilities related to lipid metabolism is presented in the three major species of *Lachnoclostridium*, above all between *C. bolteae* and *C. clostridioforme.* Furthermore, all the predicted pathways belong to the higher category of lipid biosynthesis, such as biosynthesis of choline I, CDP-diacylglycerol I and III, with the exception of the fatty acid beta-oxidation IV pathway, which belongs to the lipid degradation category. Moreover, genomes from the three major species presented pathways involved in the production of trimethylamine (TMA), including the biosynthesis of choline (∼100% genomes of *C. bolteae* and *C. symbiosum*) and phosphatidylethanolamine (∼100% genomes of *C. bolteae* and *C. clostridioforme*).

DISCUSSION

In what is to our knowledge the largest study to date investigating the associations of circulating acetate levels with gut microbiome composition and diversity and visceral fat, we report that circulating acetate levels are positively associated with microbiome alpha-diversity, while different gut bacterial genera are associated with either higher or lower acetate levels, and higher serum levels of acetate are correlated with lower visceral fat. We have also shown for the first time that the identified acetate-associated genus *Lachnoclostridium* has a strong positive correlation with visceral fat, and such association is partially mediated by acetate. Moreover, this is the first study genomically characterizing the acetate-associated gut genera *Lachnoclostridium* and *Coprococcus*, specifically, presenting their diversity and evolution at the genus level and annotating the functional capabilities of their species.

The identified positive associations between acetate and *Barnesiella* and *Ruminococcus* are consistent with the fact they contain genes involved in acetate production (Rey et al., 2010; Lustgarten, 2019). We also found a negative correlation between acetate levels and *Bacteroides*. Strikingly, *Bacteroides* spp. are acetate producers (Miller, 1978; Robert et al., 2007). We speculate that a plausible reason why *Bacteroides* present a negative correlation is the co-presence of other bacteria that might utilize the acetate produced by *Bacteroides* to generate other metabolites.

Among the bacterial genera identified, we genomically characterized *Lachnoclostridium* and *Coprococcus* as they presented the largest coefficient estimates in the association with acetate, as well as the positive association between visceral fat and *Lachnoclostridium*. In addition, these two genera presented opposing effects on acetate levels, even though they are within the same family, and thus, their genomic characterisation can provide a more holistic perspective of the influence of the gut bacteria on human health.

The dendrogram of *Coprococcus* revealed that half of the identified species remained uncharacterized, indicating that *Coprococcus* is still a poorly known genus, whereas most *Lachnoclostridium* species presented less than six genomes suggesting that its members are very rare (low prevalence) or that may be present in the human gut at such extremely low abundances that are difficult to detect.

In addition, the dendrogram allowed us to identify misclassified genomes, emphasizing the importance of performing quality controls and taxonomic classification. We could further confirm that the groups of species obtained using an ANI threshold of 95% were correct, since all the identified species formed completely independent lineages. It is important to note that the bacterial species delineation was not affected by the high proportion of MAGs used (92% of *Coprococcus* genomes and 71% of *Lachnoclostridium* genomes from the total number). Therefore, the genomic methods proposed here can be generalizable to genomes from other bacterial species, independently of the genome type (reconstructed from metagenomes or sequenced from isolates).

Furthermore, to the best of our knowledge, the phylogenetic results represent the most complete overview of the phylogenetic relationships of species from the genera *Coprococcus* and *Lachnoclostridium* so far, as it includes non-characterized species.

The annotation of the fermentative pathways confirmed that the identified *Coprococcus* species present genes involved in the formation of acetate, explaining the found positive association between this genus and acetate. Moreover, *Coprococcus* species are known as butyrate producers (Pryde et al., 2002), supporting with our results, which show that the formation of butyrate was predicted in all species. Our results are also in line with the fact that *C. catus* can produce propionate via the acrylate pathway (Reichardt et al., 2014). *C. catus* and *C. comes* present fermentative pathways (e.g., ethanol and acetone production) which are not found in other species. Interestingly, these species clustered in a different clade in the phylogenetic tree at the genus level. Additionally, both might produce lactate. It is known that *C. comes* can also produce lactate and *C. catus* can produce propionate from this compound (Reichardt et al., 2014), however, *C. catus* is not recognized as a lactate producer. We hypothesize that the produced lactate in *C. catus* might be used to generate propionate or that this fermentative pathway is not active, as this genomic approach facilitates the prediction of the functional capabilities of this genus, but unable to infer active pathways.

When we analyzed the diversity of the functional capabilities related to the biosynthesis and degradation of compounds using a PCA, we observed a considerable functional diversity among species. Of note, *C. sp5* and *C. sp6,* and *C. eutactus A, C. eutactus* and *C. sp7* were closely clustered, again, these are closely related according to the phylogenetic tree, and thus, the lack of differences might be due to their evolutionary closeness. These results suggest that different species might be distinguished by their metabolic functional capabilities.

We also noted differences in several KEGG metabolic pathways between species. Some of these pathways have been associated with CMH. For instance, a higher aminobenzoate degradation has been associated with a body weight decrease (Pataky et al., 2016). Our results show that genomes from *C. catus*, *C. eutactus A, C. eutactus* and *C. sp7* might degrade aminobenzoate, and thus, positively influencing body weight.

Regarding the shared KEGG metabolic pathways, all the genomes presented starch and sucrose metabolism and pentose phosphate pathway, which are necessary to produce SCFAs (Topping and Clifton, 2001; Basen and Kurrer, 2020), and metabolism of essential amino acids, which can be absorbed meeting the amino acids requirements (Fuller and Tomé, 2005). Furthermore, all the species might be able to metabolize several vitamins/nutrients; including vitamins B, which has been associated with protective pathways involved in CMH; folate levels, which have been correlated with a lower metabolic syndrome score, plasma fasting glucose and a higher plasma HDL cholesterol (Navarrete-Muñoz et al., 2020); biotin, which has been shown to be involved in the glucose and lipid homeostasis

(Fernandez-Mejia, 2005); thiamine, which may attenuate hypertension (Alaei-Shahmiri et al., 2015); and pyridoxine, might decrease triglyceride levels (Mottaghian et al., 2020).

Finally, we examined the lipid metabolism of the three major species of *Lachnoclostridium* as we found a positive association with visceral fat, as well as several studies have reported it to be related to diet-induced obesity (Zhao L. et al., 2017; Li et al., 2019; Sun et al., 2020), total cholesterol and LDL-C (Wang et al., 2020). Additionally, the mechanisms by through *Lachnoclostridium* impacts obesity remain unknown. Our results suggest that *Lachnoclostridium* spp. might negatively impact obesity and T2D. For instance, *Lachnoclostridium* spp might biosynthesize choline and phosphatidylethanolamine. Phosphatidylethanolamine can be methylated producing choline (Li and Vance, 2008), which can be subsequently used to produce TMA, and then trimethylamine-N-oxide (TMAO) in the liver (Zhu et al., 2018). This is in line with the fact that *Lachnoclostridium* has been suggested to be a TMA-producing bacteria (Jameson et al., 2016). Likewise, TMAO pathway has been associated with CMD in humans such as obesity and T2D (Dambrova et al., 2016; Schugar et al., 2017). Moreover, we identified in *C. bolteae* and *C. clostridioforme* two pathways involved in the biosynthesis of CDP-diacylglycerol, which might be a potential mediator of insulin resistance (Petersen and Shulman, 2018).

We are aware of some limitations in this study. The study sample includes only woman, and thus, our results might not be generalisable to men or different ranges of age. Only measures of acetate were available in this study, and therefore, we could not assess the associations between other relevant SCFAs, such as butyrate and propionate, and gut microbiota and visceral fat. These measures were performed using NMR, which provides different levels as compared to the gold standard LC-MS methodology. Furthermore, the association study was performed using 16S rRNA gene sequencing data. Our findings would have benefited from metagenomic sequencing analyses and an independent dataset to replicate our results or in vitro demonstrations.

Notwithstanding the above limitations, we have shown for the first time that higher abundances of *Lachnoclostridium* lead to lower circulating levels of acetate, resulting in increasing visceral fat. In addition, *Coprococcus* may play an important role in host health by its production of vitamins B and SCFAs, whereas *Lachnoclostridium* might have a negative impact on CMH by influencing negatively the circulating levels of acetate and being involved in the biosynthesis of harmful lipid compounds, such as TMA and CDP-diacylglycerol. We have also presented a dataset that compiles 271 and 1,121 high-quality genomes of *Lachnoclostridium* and *Coprococcus,* respectively, which can be very useful for scientists working in this area.

DATA AVAILABILITY STATEMENT

16S sequencing data used for this study is deposited in the European Nucleotide Archive (ERP015317). All other TwinsUK data are available upon request on the department website (http://www.twinsuk.ac.uk/dataaccess/accessmanagement/). All the metagenome data generated during the current study are included in the **Supplementary Material**.

ETHICS STATEMENT

Twins provided informed written consent and the study was approved by St. Thomas' Hospital Research Ethics Committee (REC Ref: EC04/015).

AUTHOR CONTRIBUTIONS

CM and AMV conceived and designed the experiments. AN analyzed the data. PL, XZ, PMW, CJS, TDS, and MF contributed reagents/materials/analysis tools. AN drafted the first version of the manuscript. AMV and CM edited and revised the manuscript. All authors have read the final manuscript and approved it for publication.

FUNDING

The Department of Twin Research receives support from grants from the Wellcome Trust (212904/Z/18/Z), the Medical Research Council (MRC)/British Heart Foundation (BHF) Ancestry and Biological Informative Markers for Stratification of Hypertension (AIM-HY; MR/M016560/1), the European Union, the Chronic Disease Research Foundation (CDRF), Zoe Global Ltd., and the NIHR Clinical Research Facility and Biomedical Research Centre (based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London). CM is funded by the Chronic Disease Research Foundation and by the MRC AIM-HY project grant. PL and AN are funded by the Chronic Disease Research Foundation. AMV is supported by the National Institute for Health Research Nottingham Biomedical Research Centre.

ACKNOWLEDGMENTS

We thank all the participants of TwinsUK for contributing and supporting our research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.711359/full#supplementary-material

Supplementary Table 1 | The created dataset containing 1,121 ultra high-quality genomes belonging to *Coprococcus* and 271 high-quality genomes belonging to *Lachnoclostridium* and their respective metadata.

Supplementary Tables 2, 3, 4 | Average nucleotide identity (ANI) values obtained for each pair of genomes belonging to *Lachnoclostridium* and *Coprococcus* using FastANI. Previously, genomes were filtered by alignment fraction (>0.4).

Supplementary Table 5 | KEGG and MetaCyc pathways obtained for *Coprococcus* and *Lachnoclostridium* genomes along with the percentage of genomes of each species presenting a given pathway.

REFERENCES

Alaei-Shahmiri, F., Soares, M., Zhao, Y., and Sherriff, J. (2015). The impact of thiamine supplementation on blood pressure, serum lipids and C-reactive protein in individuals with hyperglycemia: a randomised, double-blind cross-over trial. *Diabetes Metab. Syndrome: Clin. Res. Rev.* 9, 213–217. doi: 10.1016/j.dsx.2015.04.014

Almeida, A., Nayfach, S., Boland, M., Strozzi, F., Beracochea, M., Shi, Z. J., et al. (2020). A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat. Biotechnol.* 39, 105–114. doi: 10.1038/s41587-020-0603-3

Barrios, C., Zierer, J., Würtz, P., Haller, T., Metspalu, A., Gieger, C., et al. (2018). Circulating metabolic biomarkers of renal function in diabetic and non-diabetic populations. *Sci. Rep.* 8;15249.

Basen, M., and Kurrer, S. E. (2020). A close look at pentose metabolism of gut bacteria. *FEBS J.* 288, 1804–1808. doi: 10.1111/febs.15575

Bingham, S. A., Welch, A. A., McTaggart, A., Mulligan, A. A., Runswick, S. A., Luben, R., et al. (2001). Nutritional methods in the European prospective investigation of cancer in Norfolk. *Public Health Nutr.* 4, 847–858. doi: 10.1079/phn2000102

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods.* 7, 335–336.

Caspi, R., Billington, R., Fulcher, C. A., Keseler, I. M., Kothari, A., Krummenacker, M., et al. (2018). The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Res.* 46, D633–D639.

Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. *InterJ. Complex Syst.* 1695, 1–9.

Cummings, J., Pomare, E., Branch, W., Naylor, C., and Macfarlane, G. (1987). Short chain fatty acids in human large intestine, portal, hepatic and venous blood. *Gut* 28, 1221–1227. doi: 10.1136/gut.28.10.1221

Dambrova, M., Latkovskis, G., Kuka, J., Strele, I., Konrade, I., Grinberga, S., et al. (2016). Diabetes is associated with higher trimethylamine N-oxide plasma levels. *Exp. Clin. Endocrinol. Diabetes* 124, 251–256. doi: 10.1055/s-0035-1569330

Den Besten, G., Van Eunen, K., Groen, A. K., Venema, K., Reijngoud, D.-J., and Bakker, B. M. (2013). The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *J. Lipid Res.* 54, 2325–2340. doi: 10.1194/jlr.r036012

Fernandez-Mejia, C. (2005). Pharmacological effects of biotin. *J. Nutr. Biochem.* 16, 424–427. doi: 10.1016/j.jnutbio.2005.03.018

Frost, G., Sleeth, M. L., Sahuri-Arisoylu, M., Lizarbe, B., Cerdan, S., Brody, L., et al. (2014). The short-chain fatty acid acetate reduces appetite via a central homeostatic mechanism. *Nat. Commun.* 5, 1–11.

Fuller, M. F., and Tomé, D. (2005). In vivo determination of amino acid bioavailability in humans and model animals. *J. AOAC Int.* 88, 923–934. doi: 10.1093/jaoac/88.3.923

Goodrich, J. K., Davenport, E. R., Beaumont, M., Jackson, M. A., Knight, R., Ober, C., et al. (2016). Genetic determinants of the gut microbiome in UK twins. *Cell Host Microbe* 19, 731–743. doi: 10.1016/j.chom.2016.04.017

Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847–2849. doi: 10.1093/bioinformatics/btw313

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

Hron, W., Menahan, L., and Lech, J. (1978). Inhibition of hormonal stimulation of lipolysis in perfused rat heart by ketone bodies. *J. Mol. Cell. Cardiol.* 10, 161–174. doi: 10.1016/0022-2828(78)90040-8

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9, 1–8.

Jameson, E., Doxey, A. C., Airs, R., Purdy, K. J., Murrell, J. C., and Chen, Y. (2016). Metagenomic data-mining reveals contrasting microbial populations responsible for trimethylamine formation in human gut and marine ecosystems. *Microb. Genomics* 2:e000080.

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 42, D199–D205.

Kaul, S., Rothney, M. P., Peters, D. M., Wacker, W. K., Davis, C. E., Shapiro, M. D., et al. (2012). Dual-energy X-ray absorptiometry for quantification of visceral fat. *Obesity* 20, 1313–1318. doi: 10.1038/oby.2011.393

Lahti, L., and Shetty, S. (2018). *Introduction to the Microbiome R package.*

Li, S., Li, J., Mao, G., Yan, L., Hu, Y., Ye, X., et al. (2019). Effect of the sulfation pattern of sea cucumber-derived fucoidan oligosaccharides on modulating metabolic syndromes and gut microbiota dysbiosis caused by HFD in mice. *J. Funct. Foods* 55, 193–210. doi: 10.1016/j.jff.2019.02.001

Li, Z., and Vance, D. E. (2008). Thematic review series: glycerolipids. Phosphatidylcholine and choline homeostasis. *J. Lipid Res.* 49, 1187–1194. doi: 10.1194/jlr.r700019-jlr200

Lustgarten, M. S. (2019). The role of the gut microbiome on skeletal muscle mass and physical function: 2019 update. *Front. Physiol.* 10:1435. doi: 10.3389/fphys.2019.01435

Martin-Gallausiaux, C., Marinelli, L., Blottière, H. M., Larraufie, P., and Lapaque, N. (2020). SCFA: mechanisms and functional importance in the gut. *Proc. Nutr. Soc.* 80, 37–49. doi: 10.1017/s0029665120006916

McCance, R. A., and Widdowson, E. M. (2014). *McCance and Widdowson's the Composition of Foods.* Royal Society of Chemistry.

Menni, C., Migaud, M., Glastonbury, C. A., Beaumont, M., Nikolaou, A., Small, K. S., et al. (2016). Metabolomic profiling to dissect the role of visceral fat in cardiometabolic health. *Obesity (Silver Spring)* 24, 1380–1388. doi: 10.1002/oby.21488

Miller, T. L. (1978). The pathway of formation of acetate and succinate from pyruvate by *Bacteroides* succinogenes. *Arch. Microbiol.* 117, 145–152. doi: 10.1007/bf00402302

Moayyeri, A., Hammond, C. J., Valdes, A. M., and Spector, T. D. (2013). Cohort profile: TwinsUK and healthy ageing twin study. *Int. J. Epidemiol.* 42, 76–85. doi: 10.1093/ije/dyr207

Mottaghian, M., Salehi, P., Teymoori, F., Mirmiran, P., Hosseini-Esfahani, F., and Azizi, F. (2020). Nutrient patterns and cardiometabolic risk factors among Iranian adults: tehran lipid and glucose study. *BMC Public Health* 20:653. doi: 10.1186/s12889-020-08767-6

Navarrete-Muñoz, E.-M., Vioque, J., Toledo, E., Oncina-Canovas, A., Martínez-González, M. A., and Salas-Salvado, J. (2020). Dietary folate intake and metabolic syndrome in participants of PREDIMED-Plus study: a cross-sectional study. *Eur. J. Nutr.* 60, 1125–1136.

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114

Pataky, Z., Genton, L., Spahr, L., Lazarevic, V., Terraz, S., Gaïa, N., et al. (2016). Impact of hypocaloric hyperproteic diet on gut microbiota in overweight or obese patients with nonalcoholic fatty liver disease: a pilot study. *Digest. Dis. Sci.* 61, 2721–2731. doi: 10.1007/s10620-016-4179-1

Petersen, M. C., and Shulman, G. I. (2018). Mechanisms of insulin action and insulin resistance. *Physiol. Rev.* 98, 2133–2223. doi: 10.1152/physrev.00063.2017

Pryde, S. E., Duncan, S. H., Hold, G. L., Stewart, C. S., and Flint, H. J. (2002). The microbiology of butyrate formation in the human colon. *FEMS Microbiol. Lett.* 217, 133–139. doi: 10.1111/j.1574-6968.2002.tb11467.x

R Core Team and DC, R. (2019). *A language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing.

Reichardt, N., Duncan, S. H., Young, P., Belenguer, A., Leitch, C. M., Scott, K. P., et al. (2014). Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. *ISME J.* 8, 1323–1335. doi: 10.1038/ismej.2014.14

Rey, F. E., Faith, J. J., Bain, J., Muehlbauer, M. J., Stevens, R. D., Newgard, C. B., et al. (2010). Dissecting the in vivo metabolic potential of two human gut acetogens. *J. Biol. Chem.* 285, 22082–22090. doi: 10.1074/jbc.m110.117713

Robert, C., Chassard, C., Lawson, P. A., and Bernalier-Donadille, A. (2007). *Bacteroides* cellulosilyticus sp. nov., a cellulolytic bacterium from the human gut microbial community. *Int. J. Syst. Evol. Microbiol.* 57, 1516–1520. doi: 10.1099/ijs.0.64998-0

Sakakibara, S., Yamauchi, T., Oshima, Y., Tsukamoto, Y., and Kadowaki, T. (2006). Acetic acid activates hepatic AMPK and reduces hyperglycemia in diabetic KK-A (y) mice. *Biochem. Biophys. Res. Commun.* 344, 597–604. doi: 10.1016/j.bbrc.2006.03.176

Sayers, E. W., Agarwala, R., Bolton, E. E., Brister, J. R., Canese, K., Clark, K., et al. (2019). Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 47:D23.

Schugar, R. C., Shih, D. M., Warrier, M., Helsley, R. N., Burrows, A., Ferguson, D., et al. (2017). The TMAO-producing enzyme flavin-containing monooxygenase 3 regulates obesity and the beiging of white adipose tissue. *Cell Rep.* 19, 2451–2461. doi: 10.1016/j.celrep.2017.05.077

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Sun, X., Zhao, H., Liu, Z., Sun, X., Zhang, D., Wang, S., et al. (2020). Modulation of gut microbiota by fucoxanthin during alleviation of obesity in high-fat diet-fed mice. *J. Agric. Food Chem.* 68, 5118–5128. doi: 10.1021/acs.jafc.0c01467

Thissen, D., Steinberg, L., and Kuang, D. (2002). Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *J. Educ. Behav. Stat.* 27, 77–83. doi: 10.3102/10769986027001077

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., and Imai, K. (2014). *Mediation: R package for Causal Mediation Analysis*.

Topping, D. L., and Clifton, P. M. (2001). Short-chain fatty acids and human colonic function: roles of resistant starch and nonstarch polysaccharides. *Physiol. Rev.* 81, 1031–1064. doi: 10.1152/physrev.2001.81.3.1031

Wang, Y., Ye, X., Ding, D., and Lu, Y. (2020). Characteristics of the intestinal flora in patients with peripheral neuropathy associated with type 2 diabetes. *J. Int. Med. Res.* 48:0300060520936806.

Wells, P. M., Adebayo, A. S., Bowyer, R. C. E., Freidin, M. B., Finckh, A., Strowig, T., et al. (2020). Associations between gut microbiota and genetic risk for rheumatoid arthritis in the absence of disease: a cross-sectional study. *Lancet Rheumatol.* 2, e418–e427.

Willett, W., and Stampfer, M. J. (1986). Total energy intake: implications for epidemiologic analyses. *Am. J. Epidemiol.* 124, 17–27. doi: 10.1093/oxfordjournals.aje.a114366

Wu, Y.-W. (2018). ezTree: an automated pipeline for identifying phylogenetic marker genes and inferring evolutionary relationships among uncultivated prokaryotic draft genomes. *BMC Genomics* 19:921. doi: 10.1186/s12864-017-4327-9

Würtz, P., Havulinna, A. S., Soininen, P., Tynkkynen, T., Prieto-Merino, D., Tillin, T., et al. (2015). Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation* 131, 774–785. doi: 10.1161/circulationaha.114.013116

Ye, Y., and Doak, T. G. (2009). A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. *PLoS Comput. Biol.* 5:e1000465. doi: 10.1371/journal.pcbi.1000465

Yutin, N., and Galperin, M. Y. (2013). A genomic update on clostridial phylogeny: g ram-negative spore formers and other misplaced clostridia. *Environ. Microbiol.* 15, 2631–2641.

Zhao, L., Zhang, Q., Ma, W., Tian, F., Shen, H., and Zhou, M. (2017). A combination of quercetin and resveratrol reduces obesity in high-fat diet-fed rats by modulation of gut microbiota. *Food Funct.* 8, 4644–4656. doi: 10.1039/c7fo01383c

Zhao, Y., Liu, J., Hao, W., Zhu, H., Liang, N., He, Z., et al. (2017). Structure-specific effects of short-chain fatty acids on plasma cholesterol concentration in male syrian hamsters. *J. Agric. Food Chem.* 65, 10984–10992. doi: 10.1021/acs.jafc.7b04666

Zhu, W., Buffa, J., Wang, Z., Warrier, M., Schugar, R., Shih, D., et al. (2018). Flavin monooxygenase 3, the host hepatic enzyme in the metaorganismal trimethylamine N-oxide-generating pathway, modulates platelet responsiveness and thrombosis risk. *J. Thrombosis Haemostasis* 16, 1857–1872. doi: 10.1111/jth.14234

Conflict of Interest: TDS is co-founder of Zoe Global Ltd. AMV is a consultant for Zoe Global Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supplementary material

Supplementary Table 8.1 The created dataset containing 1,121 ultra high-quality genomes belonging to *Coprococcus* and 271 high-quality genomes belonging to *Lachnoclostridium* and their respective metadata.

Large table. Access to the table is given in the attached OneDrive. A list of the entire OneDrive content is listed in ***Appendix A****.*

Supplementary Tables 8.2, 8.3, 8.4 Average nucleotide identity (ANI) values obtained for each pair of genomes belonging to *Lachnoclostridium* and *Coprococcus* using FastANI. Previously, genomes were filtered by alignment fraction (>0.4).

*Large tables. Access to the tables are given in the attached OneDrive. A list of the entire OneDrive content is listed in **Appendix A**.*

Supplementary Table 8.5 KEGG and MetaCyc pathways obtained for *Coprococcus* and *Lachnoclostridium* genomes along with the percentage of genomes of each species presenting a given pathway.

Large table. Access to the table is given in the attached OneDrive. A list of the entire OneDrive content is listed in ***Appendix A****.*

Chapter 9

Discussion and conclusions

In this concluding chapter, I discuss the findings from the previous chapters and I frame them in the context of the thesis' aims and hypotheses. Moreover, I identify the limitations and strengths of this work. Finally, I provide some suggestions on future research lines based on the observed results, and I highlight the scientific, social, and economic implications of my research.

By leveraging data from multiple population-based cohorts and applying a variety of statistical and computational approaches, this thesis tested two interconnected hypotheses: (i) that specific metabolites contribute to the individual metabolic risk and are useful biomarkers of prevalent and incident CMD; and (ii) that circulating and faecal gut microbial-derived metabolites, such as SCFAs, are important determinants of CMD and represent specific pathways to be targeted by gut microbiome interventions.

To test the first hypothesis, I searched for biomarkers of prevalent and incident CMD. I assessed the association between metabolites measured in serum and stool and different cardiometabolic traits, including incident cardiovascular mortality, incident MI, and prevalent prediabetes. Moreover, I explored the underlying molecular pathways and estimated the gut microbiota contribution to the faecal abundances of the identified metabolites.

To test the second hypothesis, I investigated the role of the gut bacteria-derived metabolites SCFAs in the interplay between gut microbiota and CMD. I comprehensively assessed the host genetics and gut microbiota contribution to a panel of eight SCFAs in serum and stool, examined their changes after a meal challenge, and explored the links with inflammatory responses. I then focused on acetate, one of the major SCFAs, and explored the associations between its circulating levels, the gut microbiota and visceral fat. Finally, I genomically characterised the identified acetate-associated gut genera.

In this final chapter, I first provide a summary and a comprehensive discussion of these results. I then summarise the limitations and strengths of the presented work. Lastly, I propose potential future research lines based on the present findings.

9.1 Summary and discussion of findings

This thesis identified specific circulating and faecal metabolites, including the gut bacteria-derived metabolites SCFAs, as key contributors to the onset and progression of CMD. My results confirmed previous associations (e.g., the positive effect of SCFAs in chronic inflammation), thus highlighting the robustness of my approach and they also identified novel biomarkers and metabolic signatures (e.g., the faecal metabolic signature associated with a higher risk of prediabetes). In addition, the results of this thesis showed the metabolic pathways in which the identified metabolites are involved, and the gut microbiota contribution to their levels. Taken together, these findings open promising avenues for future research and potential treatments for CMD (discussed in detail in **Section 9.4**).

In **Chapters 4, 5** and **6**, I searched for circulating and faecal biomarkers of prevalent and incident CMD. In **Chapter 4**, I identified a panel of 21 circulating metabolites cross-sectionally associated with ASCVD at two timepoints explaining 9.3% of the variance not already explained by environmental and traditional risk factors. This panel added an incremental predictive value of incident cardiac disease and CVD mortality over the aforementioned factors. Then in **Chapter 5**, I found 56 circulating biomarkers (10 novel) of incident MI in the largest MWAS of MI to date, consisting of 7897 individuals from 6 intercontinental COMETS cohorts. Finally, in **Chapter 6**, I identified a faecal metabolite signature of prediabetes, and found that the gut microbiome affects this metabolic condition by influencing the absorption and/or excretion of host-produced metabolites.

In **Chapters 7** and **8**, I then focused on the role of the gut bacteria-derived metabolites SCFAs as biomarkers of CMD and their interplay with the gut microbiota. In **Chapter 7**, I reported that SCFA levels are mostly modifiable and change postprandially, and that faecal SCFAs reflect gut microbiome composition. I also showed for the first time that the SCFA levels and their correlations with inflammatory markers change depending on the type of inflammatory response (chronic or acute trauma). Finally, in **Chapter 8**, I reported that gut microbiota diversity and specific bacteria are associated with circulating acetate levels and identified the mediatory role of acetate in the association between gut microbiota and visceral fat.

Below, I discuss the main findings, moving from the broad identification of metabolites as biomarkers of CMD to the investigation of the role of the gut bacteria-derived metabolites SCFAs in the interplay between gut microbiota and CMD.

Biomarkers of prevalent and incident CMD: In **Chapters 4** and **5**, I identified circulating biomarkers of incident cardiovascular traits. Both studies presented similarities with respect to the obtained results. First, the serum metabolome of individuals with cardiovascular morbidities or mortality starts to be dysregulated before the disease is well-established, underscoring the potential of metabolomics as a proactive tool in the screening and prevention of CVD. Second, the dysregulated metabolites are primarily lipids (47% in ASCVD and 37% in MI) and amino acids (23% in ASCVD and 30% in MI), highlighting the importance of these types of molecules in the onset of CVD, and their potential use as targets for CVD prevention. However, these dysregulated metabolites consist of a wide range of sub-classes, particularly in the case of lipids (e.g., lysophospholipids, steroids and sterols), pointing out the complexity of the aetiology of these diseases. Third, enrichment pathway analyses revealed that the identified metabolites are involved in pathways that had already been previously reported to be associated with CVD (e.g., sphingomyelin metabolism, choline biosynthesis, and glycine, serine and threonine metabolism). Therefore, these pathways should be further studied and validated using experimental models to better understand the mechanisms underlying these diseases. Of note, most studies investigating circulating metabolites associated with cardiovascular traits to date have used univariate and linear approaches, and/or a limited number of participants [250, 291]. Unlike these, the metabolite panel associated with ASCVD was identified using a machine learning algorithm that goes beyond linear and univariate associations.Likewise, the COMETS study benefits from a large number of individuals (n=7897) from diverse races and backgrounds with metabolomics profiling and prospective MI information, allowing us to identify novel universal biomarkers.

As highlighted in **Chapter 1**, faecal metabolomics can provide mechanistic insights into microbiome-linked host phenotypes [236, 292]. In **Chapter 6**, I identified a faecal metabolite signature associated with a higher risk of prediabetes and potentially predictive of incident T2D. Though the 8 metabolites making up the signature are not produced by the gut microbiome but by the host (e.g., cofactors and vitamins), gut microbiome composition can accurately predict their faecal abundances (AUC>70%). Prediabetes had been linked to the gut microbiome composition, however, the underlying mechanisms remained unclear [293]. My results suggest that the gut microbiome might affect prediabetes risk by regulating the absorption or excretion of host-produced compounds, possibly via changes in gut barrier permeability or shifts in beneficial bacteria populations. For instance, this might occur due to alterations in gut barrier permeability caused by mucin-degrading bacteria or a decrease in beneficial bacteria utilising these metabolites. This is further supported by the mediation analysis results, showing the metabolites in the signature act as partial mediators on the significant associations between several gut microbial species (e.g., *Faecalibacillus intestinalis*, *Dorea spp.* and *Ruminococcus torques*) and prediabetes.

The role of SCFAs in the interplay between gut microbiota and CMD: In **Chapters 7** and **8**, I then explored the gut bacteria-derived metabolites SCFAs as biomarkers of CMD. Specifically, I focused on their role in inflammatory responses and visceral fat, which are key elements for the understanding of CMD (see **Section 1.1**). In **Chapter 7**, I explored the links between circulating SCFAs and chronic and acute inflammatory responses. When examining the links with chronic inflammation, I observed that SCFAs are linked to lower systemic inflammation, which is in line with previous reports [294]. Moreover, for the first time, I analysed their role in acute inflammatory responses, such as those seen in acute trauma cases. I found that certain types of fractures (rib or hip fractures) led to changes in circulating SCFA levels with respect to healthy individuals and with each other. The differences in inflammatory responses in different trauma scenarios suggest that SCFAs may play a role in the recovery process, potentially by dampening the inflammatory response in acute inflammation and contributing to the maintenance of a low-grade inflammatory state in systemic inflammation. On the other hand, in **Chapter 8**, I assessed the association between circulating levels of acetate and visceral fat, and I found that higher levels were associated with less visceral fat. This suggests that the potential role of circulating SCFAs in exerting benefits in CMD is by regulating inflammatory responses and decreasing visceral fat.

I also explored the role of host genetics in regulating SCFA levels (**Chapter 7**), as this would enable the understanding of how we can modulate their levels to enhance cardiometabolic health. I found that a large proportion of the SCFA levels in serum and stool are explained by environmental factors (average h^2: serum=14%(SD=5%); stool=12%(SD=6%)). These findings reinforce the importance of non-genetic factors in SCFA formation. Differences in individual dietary and lifestyle patterns, including alcohol intake, smoking intensity, exercise, and sleep patterns, might be driving the observed variations, as some studies have suggested for other metabolites in serum and stool [236, 295]. These factors have been reported to modulate the gut microbiota composition and function [83], which might result in changes in SCFA levels [2].

Consequently, I assessed the contribution of the gut microbiota to SCFA levels. When integrating the gut microbiota compositional data using a machine learning approach to predict SCFA levels in serum and stool (**Chapter 7**), I identified that the gut microbiome can accurately predict their faecal levels (AUC>0.71) with *Akkermansia muciniphila*, *Faecalibacterium prausnitzii* and *Roseburia spp.* as important predictors, while presenting weaker associations with serum. This is consistent with the obtained low correlation between circulating and faecal levels of SCFAs. These observations highlight the fact that faecal levels are not representative of the actual absorption and suggest that caution should be taken when inferring microbiome-disease associations [67] from either serum or faecal SCFA levels.

Furthermore, in **Chapter 8**, I identified alpha-diversity metrics to be positively associated with circulating acetate levels, and the abundances of 6 gut bacterial genera to be either positively (e.g., *Lachnoclostridium*) or negatively (e.g., *Coprococcus*) associated with its levels. Importantly, I found *Lachnoclostridium* to be also positively associated with visceral fat, and acetate was partially mediating such an association (VAF=10%). Additionally, in **Chapter 8**, these results were combined with the identification of the species belonging to *Lachnoclostridium* and *Coprococcus* followed by their functional characterisation to further confirm the observed associations and to better understand the mechanisms through which they might impact human health. I identified genes involved in acetate formation, aligning with the obtained results. Moreover, while *Lachnoclostridium spp.* might negatively impact obesity and T2D by potentially biosynthesising choline, phosphatidylethanolamine [296] and CDP-diacylglycerol [297], *Coprococcus spp.* might be able to metabolise several vitamins B, which have been associated with protective pathways involved in cardiometabolic health [298–300].

Finally, in **Chapter 7**, I examined for the first time the changes in SCFA levels after a meal challenge. Although humans spend most of their days in a postprandial state, postprandial SCFA responses had only been investigated in animal models [301, 302]. I found that there are significant individual differences in these responses. Moreover, the heritability analysis revealed that for most SCFAs these were largely environmentally driven. However, the gut microbiome presented a weak association with these levels, as happened with the circulating levels at fasting. Therefore, other potential factors underlying these changes need to be investigated.

9.2 Limitations

This thesis presents several limitations that should be carefully considered.

- **Phenotype collection in TwinsUK**: Although TwinsUK provides an extensive, well-phenotyped, population-based cohort [218], concurrent data is not always available. Given that the gut microbiome and metabolome are inherently dynamic and susceptible to extrinsic environmental perturbations [303], these variations could potentially introduce different sources of bias into the results, including inter-individual variability, temporal bias and confounding bias. Furthermore, the availability of data varies across different subsets of the cohort. When integrating different subsets, the sample size containing all available data tends to decrease, affecting the statistical power of the conducted analyses. Finally, TwinsUK mainly consists of middle-aged Caucasian women [218]. As such, the generalisability of my findings may vary when extrapolating to populations with distinct demographic characteristics.

- **Cross-sectional design**: Most analyses conducted in this thesis have a cross-sectional design rather than a prospective design. Longitudinal data for some of the outcomes of interest was not available (e.g., prediabetes) or was limited to a small number of participants (e.g., T2D). This restricts the ability to evaluate temporal changes in the metabolome and determine causation. For example, in **Chapter 8**, the effect of circulating acetate levels on visceral fat is cross-sectionally explored, thus, its impact on visceral fat over time cannot be inferred. Furthermore, while the included cohorts provide a wealth of observational data, experimental analyses are needed for a more comprehensive understanding of the underlying mechanisms. However, the obtained findings can serve as a foundation for other researchers to test some of the hypotheses that arise from my results.

- **Relative measures for Metabolon metabolomics data**: Metabolites profiled by Metabolon Inc. are relative rather than absolute quantifications. Consequently, the effect sizes obtained from the analyses incorporating such data (e.g., **Chapters 4, 5** and **6**) have no direct biological significance and might be influenced by the sample size. Nonetheless, I was still able to identify the deregulated metabolites and their directions in the association with CMD. Also, I was able to meta-analyse Metabolon data with metabolites measured by other metabolomics providers (i.e., Broad Institute and Nightingale) in **Chapter 5**.

- **Data collection bias**: In addition to the biases previously discussed, other biases might have also arisen during the process of data collection. For metabolomics, 16S rRNA data and metagenomics, procedural and technical sources might introduce variability in the metabolite and gut bacterial abundances. For instance, there could be species not properly extracted from wet lab procedures as well as species not represented as not enough genomic data sequenced or either below the limit of detection of taxonomic profiling tools. Despite this, I was able to replicate in different cohorts the identified associations between gut microbiota-metabolites-cardiometabolic traits (e.g., in **Chapter 6**, the associations between the identified faecal metabolite signature and prediabetes were replicated in TwinsUK and KORA, while in **Chapter 7**, the power of the gut microbiome composition to predict SCFA levels was consistent in TwinsUK and ZOE PREDICT-1). Moreover, diverse kits and wet-lab assays were employed to measure the cytokines used in **Chapter 7** across studies, including TwinsUK, ZOE PREDICT-1 and the acute trauma case-control cohort. Variations in sample collection protocols, handling, and storage could have led to systematic variability, potentially affecting the results. To mitigate potential measurement bias, inverse variance random effect meta-analysis was applied to combine the estimates derived from each cohort. Finally, MI and comorbidities were recorded from self-reported questionnaires, which might suffer from misreporting bias, instead of being directly extracted from medical records. However, when running sensitivity analysis in **Chapter 5** excluding cohorts where MI was assessed by self-reported questionnaires, results remained consistent.

9.3 Strengths

Despite the above-mentioned limitations, the presented work has several strengths.

- **Data variety in TwinsUK**: TwinsUK is one of the most genotyped and phenotyped population-based cohorts in the world, enriched by the availability of several omics data, including serum and stool metabolomics, and shotgun metagenomes. This has allowed me to explore the interplay between the gut microbiome and metabolome and their impact on several cardiometabolic traits. Moreover, this cohort consists of twins enabling the study of genetic/hereditary factors (e.g., in **Chapter 7**, the heritability of circulating and faecal SCFA levels was estimated).

- **Replication of findings**: I had access to independent cohorts to replicate my findings, highlighting the robustness of my results. This was the case for **Chapters 4, 5, 6** and **7**. In particular, for **Chapter 5**, I had access to data from 6 intercontinental COMETS cohorts, which provided me with a high number of participants, increasing the power of my statistical analyses, and allowed me to study the influence of demographic diversity in the identified MI-metabolite associations. Furthermore, despite the extensive range of tests conducted in these chapters, analyses have been extensively adjusted for covariates and multiple testing, mitigating the risk of false positives and enhancing the reliability of the findings.

- **Comprehensive assessment of SCFA levels in two independent cohorts**: SCFA levels were measured in serum and stool for two cohorts, namely TwinsUK and ZOE PREDICT-1, providing a more holistic understanding of their interrelation and how the host genetics and the gut microbiome influence both levels. In the ZOE PREDICT-1 cohort, postprandial measurements were also available, which provides a dynamic picture of their physiological responses.

- **Variety of statistical and computational analyses**: Throughout this thesis, I have applied a wide range of statistical methods, ranging from univariate and traditional linear models to machine learning algorithms, which allow the study of the contribution of many features to a given response, and explore beyond linear relations. For instance, in **Chapters 6** and **7**, I applied Random Forest models, allowing me to integrate the compositional profiles of all the detected gut microbiome species and assess their associations with the levels of different metabolites, including SCFAs. Likewise, different computational approaches were employed to genomically characterise acetate-associated gut bacterial genera, shedding light on their functional capabilities at a deeper taxonomic level (i.e., species level) and their potential impact on human health.

9.4 Future directions

In this thesis, I identified novel biomarkers and metabolic signatures of CMD risk, including sphingolipid molecules in MI and SCFAs in acute inflammation. Moreover, I provided mechanistic insights into the pathways regulating CMD and the interplay of the gut microbiota and metabolites in CMD.

Future studies should validate these biomarkers and develop strategies that modulate the levels of these biomarkers to prevent the onset and development of different CMD.

Biomarker validation: To validate these biomarkers as reliable indicators of CMD risk, causality, directionality, and the underlying molecular mechanisms need to be further investigated. It is still unclear whether metabolome changes contribute to the onset of CMD or are a result of it [304]. Therefore, moving from associations to causation is a crucial step for biomarker validation. This implies establishing a cause-and-effect relationship between the biomarker and the CMD; and ensuring that deregulation of the biomarker levels appears before the disease onset. Data from large population-based studies with a prospective study design and experimental models are pivotal for that purpose.

Population-based cohorts should include participants with different demographic characteristics (e.g., different races and ranges of age, and a balanced representation of genders), extensive metadata and repeated measures over time. This would enable the identification of more generalised and reproducible biomarkers, while accounting for potential confounders.

Human microbiota-associated rodent studies, gnotobiotics, *in vitro* models, human organotypic cultures, synthetic cultures, microbiome-depleted, germ-free, and Wildling mouse models could be applied in the context of experimental models for the microbial-derived metabolites [292, 305]. For instance, human microbiota-associated rodent studies provide insights into the host-microbiome interplay and its implications for CMD by replicating human microbial communities in a rodent host. Human organotypic cultures are another model that enables the investigation of human cellular and tissue responses while mimicking the *in vivo* conditions. Specifically, the tissue-specific effects of the gut microbiota and their metabolic products can be studied [305], providing valuable insights into the localised responses and molecular processes that may contribute to CMD progression. On the other hand, synthetic cultures can be used to dissect the metabolic interactions between defined microbial communities and the host. Furthermore, microbiome-depleted models, including germ-free animals, are invaluable in assessing the role of microbiota—or the lack thereof—in the development of CMD. These models can identify causal relationships between microbial presence, or specific microbial constituents, and host phenotypes [292, 305]. Lastly, Wildling mouse models, which are captured from wild environments and thus harbour a naturally acquired microbiota, provide a more realistic depiction of how environmental exposures to various microbial communities affect the host and potentially modulate CMD risk.

To understand the molecular mechanisms underlying the association between the biomarker and CMD, the biochemical pathways and interactions that lead to the presence of these biomarkers need to be further understood. Genomic, proteomic, and metabolomic analyses, along with animal models and clinical trials could be employed to elucidate these mechanisms. Likewise, my results could be complemented with flux balance analyses (FBA). While metabolomics provides a snapshot of the metabolic profile under particular conditions, FBA study biological networks in a quantitative manner [306], shedding light on the potential metabolic capacity and metabolic fluxes in a network. Such a comprehensive understanding of biomarkers would not only enhance our knowledge of the disease process, but also pave the way for reproducible and generalised targeted therapeutic strategies.

Translational strategies: Based on the validated biomarkers, different translational strategies with potential for future clinical implementation could be then investigated and developed. As a large proportion of the metabolites are influenced by the gut microbiome [1], here I will focus on discussing potential strategies that modulate the levels of specific biomarkers by targeting the gut microbiome (e.g., dietary interventions, pathobiont depletion, pre-, pro- and postbiotic usage, and whole community transfer). Future research could focus on designing personalised dietary approaches that target the gut microbiome, modulating specific biomarkers, for effective CMD prevention and treatment. However, there are still scientific challenges that also need to be tackled to achieve this. These include understanding (i) the inter-individual heterogeneity in metabolic responses to dietary interventions due to the temporal and inter-individual variability of the gut microbiome, and (ii) the impact of single foods and dietary compounds in the gut microbiome and metabolome. Leveraging data from large cohorts with extensive omics data and applying different sophisticated machine learning algorithms, such as neural networks and kernel-based methods, could provide insights into these scientific gaps. Non-dietary strategies, such as the administration of pre-, pro- and postbiotics, are other potential promising alternatives to modulate the biomarkers of interest. Notably, the Food and Drug Administration (FDA) has approved certain live biotherapeutic products to treat various conditions [307]. However, some still present issues in relation to the dosage (dose-specific effects of target metabolite) and variability of response between individuals. Trials supplementing these compounds and determining their effects on the identified biomarkers are needed to fully understand their clinical applicability. In contrast, whole community transfer, including therapies like faecal microbiota transplantation (FMT), which involves the transfer of complete microbiota from a healthy donor to a recipient to restore a balanced microbial ecosystem, has been shown to be effective in treating *C. difficile* infections [308] and is being explored for other conditions [309]. Additionally, when compared with other FDA-approved live biotherapeutic products, which are limited to one or a few bacteria strains, these types of therapies might provide a longer-lasting effect in the patient (recipient) and be more cost effective. Nevertheless, for FMT to be widely adopted as a therapeutic strategy, it must undergo rigorous standardisation. Protocols for donor selection, stool processing, and delivery methods need to be established [310]. In parallel, validation studies are imperative to ensure that these treatments are not only safe and effective but also reproducible. Inconsistencies in the effectiveness of FMT for treating CMD, such as metabolic syndrome, might also be addressed through such standardisation, as observed in inconsistent effects from studies citing progressive loss of donor microbes [311, 312]. In this context, the work of Karen Madsen has shown the potential of orally-administered FMT combined with fibre supplementation to improve insulin sensitivity in severe obesity and metabolic syndrome patients [313]. FMT would be able to modify the recipient's microbial ecology, thereby improving insulin sensitivity, while the fibre supplementation would enhance or maintain these effects [313].

Further work: Finally, my findings could be further expanded in future work. For instance, a significant proportion of the profiled metabolites are unknown compounds and they do not have any match in public databases [314]. Nonetheless, many of them are highly likely to play an important role in the interplay between gut microbiota and human health [315]. In this context, future research could apply recent approaches known as guilt-by-association, in which the unknown compounds can be inferred based on their associations with other known compounds [292]. The findings can shed light on the biological processes in which these unknown compounds participate, increasing our knowledge of the metabolome in cardiometabolic health. Moreover, the gut microbiome profile assessment in the presented studies of this thesis was performed from stool samples, which tends to reflect the luminal microbiome content rather than the microbiome residing in the intestinal wall [316]. Importantly, the mucosa-associated microbiome has been suggested to play key roles in the host's immunity and metabolism [317]. Therefore, future studies should also integrate mucosa-associated microbiome analyses, enabling a more holistic picture of the whole gut microbiome community. Furthermore, to better understand the gut microbiome-metabolites-CMD interactions is necessary to integrate transkingdom analyses with other omics. The shown findings in this thesis primarily focus on gut bacteria members, however, the gut microbiome community consists of other members, including viruses, archeae and fungi, which, as discussed in **Chapter 1**, also play an important role in human health. The compositional data of these members can now be extracted using novel or updated computational approaches. For instance, Soverini and colleagues have recently developed a tool called HumanMycobiomeScan that allows the characterisation of the fungal fraction from metagenomic samples [318]. Additionally, the virome can now be profiled using the novel pipeline ViroProfiler [319] or by running the gut microbiome compositional profiling tool MetaPhlAn, which in its latest version (MetaPhlAn 4), integrates a novel viral catalogue [248, 320].

Scientific, social, and economic implications: My results along with the findings from the discussed research will have profound scientific, social, and economic implications. From the scientific perspective, they will unlock fundamental mechanisms underlying gut microbiome-metabolites-cardiometabolic health interactions, which can be transferred to other research lines and applied to other diseases, such as cancer, autoimmune and neurodegenerative diseases. From a social and economic perspective, they will directly improve patients' lives by providing low-risk and non-expensive strategies, which are aligned with the sustainable development goals (SDGs) established by the United Nations (e.g., numbers 3 and 10), for the treatment and prevention of CMD.

9.5 Conclusions

The findings of this thesis illustrate the breadth of the physiological relevance of metabolites, particularly SCFAs, on CMD, and highlight the importance of the gut microbiota in the pathogenesis of CMD not only by producing metabolic products but also by affecting intestinal absorption/excretion of host-produced metabolites. Future studies should determine causality and explore translational strategies that could modulate the identified metabolites by for example targeting the gut microbiota.

References from main text

1. Visconti, A. *et al.* Interplay between the human gut microbiome and host metabolism. *Nature communications* **10,** 1–10. ISSN: 2041-1723 (2019).
2. Nogal, A., Valdes, A. M. & Menni, C. The role of short-chain fatty acids in the interplay between gut microbiota and diet in cardio-metabolic health. *Gut microbes* **13,** 1897212 (2021).
3. Ounpuu, S., Anand, S & Yusuf, S. The global burden of cardiovascular disease. *Medscape Cardiology* **4** (2002).
4. Rudd, K. E. *et al.* Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. *The Lancet* **395,** 200–211. ISSN: 0140-6736 (2020).
5. Jagannathan, R., Patel, S. A., Ali, M. K. & Narayan, K. V. Global updates on cardiovascular disease mortality trends and attribution of traditional risk factors. *Current diabetes reports* **19,** 1–12 (2019).
6. Goodrich, J. K. *et al.* Human genetics shape the gut microbiome. *Cell* **159,** 789–799. ISSN: 0092-8674 (2014).
7. Meyer, K. A. & Bennett, B. J. Diet and gut microbial function in metabolic and cardiovascular disease risk. *Current diabetes reports* **16,** 93. ISSN: 1534-4827 (2016).
8. Fan, Y. & Pedersen, O. Gut microbiota in human metabolic health and disease. *Nature Reviews Microbiology,* 1–17. ISSN: 1740-1534 (2020).
9. Shah, S. H. *et al.* Association of a peripheral blood metabolic profile with coronary artery disease and risk of subsequent cardiovascular events. *Circulation: Cardiovascular Genetics* **3,** 207–214 (2010).
10. Cheng, S. *et al.* Potential impact and study considerations of metabolomics in cardiovascular health and disease: a scientific statement from the American Heart Association. *Circulation: Cardiovascular Genetics* **10,** e000032 (2017).
11. Brown, J. M. & Hazen, S. L. The gut microbial endocrine organ: bacterially derived signals driving cardiometabolic diseases. *Annual review of medicine* **66,** 343–359. ISSN: 0066-4219 (2015).
12. Larraufie, P. *et al.* SCFAs strongly stimulate PYY production in human enteroendocrine cells. *Scientific reports* **8,** 1–9. ISSN: 2045-2322 (2018).
13. Jin, Y., Liang, J., Hong, C., Liang, R. & Luo, Y. Cardiometabolic multimorbidity, lifestyle behaviours, and cognitive function: a multicohort study. *The Lancet Healthy Longevity* (2023).
14. Danaei, G. *et al.* Cardiovascular disease, chronic kidney disease, and diabetes mortality burden of cardiometabolic risk factors from 1980 to 2010: a comparative risk assessment. *Lancet Diabetes & Endocrinology* (2014).

15. Karagkiozaki, V., Logothetidis, S. & Pappa, A.-M. Nanomedicine for atherosclerosis: molecular imaging and treatment. *Journal of biomedical nanotechnology* **11,** 191–210 (2015).
16. Goff Jr, D. C. *et al.* 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation* **129,** S49–S73 (2014).
17. Moran, A. E. *et al.* Variations in ischemic heart disease burden by age, country, and income: the Global Burden of Diseases, Injuries, and Risk Factors 2010 study. *Global heart* **9,** 91–99 (2014).
18. Virani, S. S. *et al.* Heart disease and stroke statistics—2020 update: a report from the American Heart Association. *Circulation* **141,** e139–e596 (2020).
19. Knowler, W. C. Diabetes Prevention Program Research Group: Reduction in the incidence of type 2 diabetes with life-style intervention or metformin. *N. Engl. J. Med.* **346,** 393–403 (2002).
20. Elliott, T. L. & Pfotenhauer, K. M. Classification and diagnosis of diabetes. *Primary Care: Clinics in Office Practice* **49,** 191–200 (2022).
21. Kopelman, P. G. Obesity as a medical problem. *Nature* **404,** 635–643 (2000).
22. Nuttall, F. Q. Body mass index: obesity, BMI, and health: a critical review. *Nutrition today* **50,** 117 (2015).
23. Galic, S., Oakhill, J. S. & Steinberg, G. R. Adipose tissue as an endocrine organ. *Molecular and cellular endocrinology* **316,** 129–139 (2010).
24. Kissebah, A. H. *et al.* Relation of body fat distribution to metabolic complications of obesity. *The Journal of Clinical Endocrinology & Metabolism* **54,** 254–260 (1982).
25. Palomer, X., Salvadó, L., Barroso, E. & Vázquez-Carrera, M. An overview of the crosstalk between inflammatory processes and metabolic dysregulation during diabetic cardiomyopathy. *International journal of cardiology* **168,** 3160–3172 (2013).
26. Tona, F. *et al.* Systemic inflammation is related to coronary microvascular dysfunction in obese patients without obstructive coronary disease. *Nutrition, Metabolism and Cardiovascular Diseases* **24,** 447–453 (2014).
27. Mathers, C. D. & Loncar, D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS medicine* **3,** e442. ISSN: 1549-1676 (2006).
28. Basak, T., Varshney, S., Akhtar, S. & Sengupta, S. Understanding different facets of cardiovascular diseases based on model systems to human studies: A proteomic and metabolomic perspective. *Journal of proteomics* **127,** 50–60 (2015).
29. O'Rahilly, S. Human genetics illuminates the paths to metabolic disease. *Nature* **462,** 307–314 (2009).
30. Newgard, C. B. Metabolomics and metabolic diseases: where do we stand? *Cell metabolism* **25,** 43–56 (2017).
31. Brunius, C., Shi, L. & Landberg, R. Metabolomics for improved understanding and prediction of cardiometabolic diseases—Recent findings from human studies. *Current Nutrition Reports* **4,** 348–364 (2015).
32. Nicholson, J. K. & Lindon, J. C. Metabonomics. *Nature* **455,** 1054–1056 (2008).
33. Wishart, D. S. *et al.* HMDB 5.0: the human metabolome database for 2022. *Nucleic acids research* **50,** D622–D631 (2022).
34. Barrios, C., Spector, T. D. & Menni, C. Blood, urine and faecal metabolite profiles in the study of adult renal disease. *Archives of biochemistry and biophysics* **589,** 81–92 (2016).
35. Sparkman, O. D. & Price, P. *Mass spectrometry desk reference* (2006).

36. Feng, X., Liu, X., Luo, Q. & Liu, B.-F. Mass spectrometry in systems biology: an overview. *Mass spectrometry reviews* **27,** 635–660 (2008).
37. Markley, J. L. *et al.* The future of NMR-based metabolomics. *Current opinion in biotechnology* **43,** 34–40 (2017).
38. Menni, C. *et al.* Omega-3 fatty acids correlate with gut microbiome diversity and production of N-carbamylglutamate in middle-aged and elderly women. *Scientific Reports* **7,** 1–11. ISSN: 2045-2322 (2017).
39. Bar, N. *et al.* A reference map of potential determinants for the human serum metabolome. *Nature* **588,** 135–140 (2020).
40. Ussher, J. R., Elmariah, S., Gerszten, R. E. & Dyck, J. R. The emerging role of metabolomics in the diagnosis and prognosis of cardiovascular disease. *Journal of the American College of Cardiology* **68,** 2850–2870. ISSN: 0735-1097 (2016).
41. Shah, S. H., Kraus, W. E. & Newgard, C. B. Metabolomic profiling for the identification of novel biomarkers and mechanisms related to common cardiovascular diseases: form and function. *Circulation* **126,** 1110–1120 (2012).
42. McGarrah, R. W., Crown, S. B., Zhang, G.-F., Shah, S. H. & Newgard, C. B. Cardiovascular metabolomics. *Circulation research* **122,** 1238–1258 (2018).
43. Lin, Y.-T. *et al.* Global plasma metabolomics to identify potential biomarkers of blood pressure progression. *Arteriosclerosis, thrombosis, and vascular biology* **40,** e227–e237 (2020).
44. Wishart, D. S. Emerging applications of metabolomics in drug discovery and precision medicine. *Nature reviews Drug discovery* **15,** 473–484 (2016).
45. Johnson, C. H., Ivanisevic, J. & Siuzdak, G. Metabolomics: beyond biomarkers and towards mechanisms. *Nature reviews Molecular cell biology* **17,** 451–459 (2016).
46. Newgard, C. B. *et al.* A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell metabolism* **9,** 311–326 (2009).
47. Du Clos, T. W. Function of C-reactive protein. *Annals of medicine* **32,** 274–278 (2000).
48. Shen, J. & Ordovas, J. M. Impact of genetic and environmental factors on hsCRP concentrations and response to therapeutic agents. *Clinical chemistry* **55,** 256–264 (2009).
49. Haffner, S. M. The metabolic syndrome: inflammation, diabetes mellitus, and cardiovascular disease. *The American journal of cardiology* **97,** 3–11 (2006).
50. Idicula, T. T., Brogger, J., Naess, H., Waje-Andreassen, U. & Thomassen, L. Admission C-reactive protein after acute ischemic stroke is associated with stroke severity and mortality: The'Bergen stroke study'. *BMC neurology* **9,** 1–9 (2009).
51. Ridker, P. M., Glynn, R. J. & Hennekens, C. H. C-reactive protein adds to the predictive value of total and HDL cholesterol in determining risk of first myocardial infarction. *Circulation* **97,** 2007–2011 (1998).
52. Elliott, P. *et al.* Genetic Loci associated with C-reactive protein levels and risk of coronary heart disease. *Jama* **302,** 37–48 (2009).
53. Kuppa, A., Tripathi, H., Al-Darraji, A., Tarhuni, W. M. & Abdel-Latif, A. C-Reactive Protein Levels and Risk of Cardiovascular Diseases: A Two-Sample Bidirectional Mendelian Randomization Study. *International Journal of Molecular Sciences* **24,** 9129 (2023).
54. Zhang, S., Zeng, X., Ren, M., Mao, X. & Qiao, S. Novel metabolic and physiological functions of branched-chain amino acids: a review. *Journal of animal science and biotechnology* **8,** 1–12 (2017).

55. Zappe, D. H. *et al.* Metabolic and antihypertensive effects of combined angiotensin receptor blocker and diuretic therapy in prediabetic hypertensive patients with the cardiometabolic syndrome. *The Journal of Clinical Hypertension* **10,** 894–903 (2008).
56. Ding, Y. *et al.* Plasma glycine and risk of acute myocardial infarction in patients with suspected stable angina pectoris. *Journal of the American Heart Association* **5,** e002621 (2015).
57. Wang-Sattler, R. *et al.* Novel biomarkers for pre-diabetes identified by metabolomics. *Molecular systems biology* **8,** 615 (2012).
58. Li, X. *et al.* Association of serum glycine levels with metabolic syndrome in an elderly Chinese population. *Nutrition & metabolism* **15,** 1–9 (2018).
59. Zaric, B. L. *et al.* Atherosclerosis linked to aberrant amino acid metabolism and immunosuppressive amino acid catabolizing enzymes. *Frontiers in Immunology* **11,** 551758 (2020).
60. Palmnäs, M. S. *et al.* Serum metabolomics of activity energy expenditure and its relation to metabolic syndrome and obesity. *Scientific reports* **8,** 3308 (2018).
61. Calder, P. C. Functional roles of fatty acids and their effects on human health. *Journal of parenteral and enteral nutrition* **39,** 18S–32S (2015).
62. Würtz, P. *et al.* Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation* **131,** 774–785 (2015).
63. Juturu, V. Omega-3 fatty acids and the cardiometabolic syndrome. *Journal of the cardiometabolic syndrome* **3,** 244–253 (2008).
64. Menni, C. *et al.* Serum metabolites reflecting gut microbiome alpha diversity predict type 2 diabetes. *Gut Microbes* **11,** 1632–1642 (2020).
65. Cavus, E. *et al.* Association of circulating metabolites with risk of coronary heart disease in a European population: results from the Biomarkers for Cardiovascular Risk Assessment in Europe (BiomarCaRE) consortium. *JAMA cardiology* **4,** 1270–1279. ISSN: 2380-6583 (2019).
66. Randrianarisoa, E. *et al.* Relationship of serum trimethylamine N-oxide (TMAO) levels with early atherosclerosis in humans. *Scientific reports* **6,** 1–9. ISSN: 2045-2322 (2016).
67. Deng, K. *et al.* Comparison of fecal and blood metabolome reveals inconsistent associations of the gut microbiota with cardiometabolic diseases. *Nature Communications* **14,** 571 (2023).
68. De la Cuesta-Zuluaga, J. *et al.* Higher fecal short-chain fatty acid levels are associated with gut microbiome dysbiosis, obesity, hypertension and cardiometabolic disease risk factors. *Nutrients* **11,** 51 (2018).
69. Solar, I. *et al.* Short-chain fatty acids are associated with adiposity, energy and glucose homeostasis among different metabolic phenotypes in the Nutritionists' Health Study. *Endocrine,* 1–12 (2023).
70. Zhao, J. V., Fan, B. & Burgess, S. Using genetics to examine the overall and sex-specific associations of branch-chain amino acids and the valine metabolite, 3-hydroxyisobutyrate, with ischemic heart disease and diabetes: A two-sample Mendelian randomization study. *Atherosclerosis* **381,** 117246 (2023).
71. Jiang, W. *et al.* Mendelian Randomization Analysis Provides Insights into the Pathogenesis of Serum Levels of Branched-Chain Amino Acids in Cardiovascular Disease. *Metabolites* **13,** 403 (2023).
72. Liu, L. *et al.* Association of plasma branched-chain amino acids with overweight: A Mendelian randomization analysis. *Obesity* **29,** 1708–1718 (2021).

73. Georgiou, A. N. *et al.* Appraising the causal role of risk factors in coronary artery disease and stroke: A systematic review of Mendelian Randomization studies. *Journal of the American Heart Association* **12,** e029040 (2023).
74. Park, S. *et al.* Causal effects of serum levels of n-3 or n-6 polyunsaturated fatty acids on coronary artery disease: Mendelian randomization study. *Nutrients* **13,** 1490 (2021).
75. Sanna, S. *et al.* Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. *Nat Genet* **51,** 600–605. ISSN: 1061-4036 (Print) 1061-4036 (2019).
76. Dodd, D. *et al.* A gut bacterial pathway metabolizes aromatic amino acids into nine circulating metabolites. *Nature* **551,** 648–652. ISSN: 1476-4687 (2017).
77. Savage, D. C. Microbial ecology of the gastrointestinal tract. *Annual review of microbiology* **31,** 107–133. ISSN: 0066-4227 (1977).
78. Thursby, E. & Juge, N. Introduction to the human gut microbiota. *Biochemical journal* **474,** 1823–1836 (2017).
79. Matijašić, M. *et al.* Gut microbiota beyond bacteria—mycobiome, virome, archaeome, and eukaryotic parasites in IBD. *International journal of molecular sciences* **21,** 2668 (2020).
80. Hooper, L. V. & Gordon, J. I. Commensal host-bacterial relationships in the gut. *Science* **292,** 1115–1118. ISSN: 0036-8075 (2001).
81. Eckburg, P. B. *et al.* Diversity of the human intestinal microbial flora. *science* **308,** 1635–1638. ISSN: 0036-8075 (2005).
82. Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S. & Medzhitov, R. Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell* **118,** 229–241. ISSN: 0092-8674 (2004).
83. Falony, G. *et al.* Population-level analysis of gut microbiome variation. *Science* **352,** 560–564. ISSN: 0036-8075 (2016).
84. Nagata, N. *et al.* Population-level metagenomics uncovers distinct effects of multiple medications on the human gut microbiome. *Gastroenterology* **163,** 1038–1052 (2022).
85. Malla, M. A. *et al.* Exploring the human microbiome: The potential future role of next-generation sequencing in disease diagnosis and treatment. *Frontiers in Immunology* **9,** 2868. ISSN: 1664-3224 (2019).
86. Montalto, M, D'onofrio, F, Gallo, A, Cazzato, A & Gasbarrini, G. Intestinal microbiota and its functions. *Digestive and Liver Disease Supplements* **3,** 30–34. ISSN: 1594-5804 (2009).
87. Korecka, A. & Arulampalam, V. The gut microbiome: scourge, sentinel or spectator? *Journal of oral microbiology* **4,** 9367. ISSN: 2000-2297 (2012).
88. Iebba, V. *et al.* Eubiosis and dysbiosis: the two sides of the microbiota. *New Microbiol* **39,** 1–12 (2016).
89. Schippa, S. & Conte, M. P. Dysbiotic events in gut microbiota: impact on human health. *Nutrients* **6,** 5786–5805 (2014).
90. Katsimichas, T. *et al.* The intestinal microbiota and cardiovascular disease. *Cardiovascular research* **115,** 1471–1486. ISSN: 0008-6363 (2019).
91. Tang, W. W. & Hazen, S. L. Microbiome, trimethylamine N-oxide, and cardiometabolic disease. *Translational Research* **179,** 108–115. ISSN: 1931-5244 (2017).
92. Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Human gut microbes associated with obesity. *nature* **444,** 1022–1023. ISSN: 1476-4687 (2006).

93. Tanti, J.-F., Ceppo, F., Jager, J. & Berthou, F. Implication of inflammatory signaling pathways in obesity-induced insulin resistance. *Frontiers in endocrinology* **3,** 181. ISSN: 1664-2392 (2013).
94. Philpott, D. J., Sorbara, M. T., Robertson, S. J., Croitoru, K. & Girardin, S. E. NOD proteins: regulators of inflammation in health and disease. *Nature Reviews Immunology* **14,** 9–23. ISSN: 1474-1741 (2014).
95. Curtiss, L. K. & Tobias, P. S. Emerging role of Toll-like receptors in atherosclerosis. *Journal of lipid research* **50,** S340–S345. ISSN: 0022-2275 (2009).
96. Ma, Y., You, X., Mai, G., Tokuyasu, T. & Liu, C. A human gut phage catalog correlates the gut phageome with type 2 diabetes. *Microbiome* **6,** 1–12 (2018).
97. Mar Rodríguez, M *et al.* Obesity changes the human gut mycobiome. *Scientific reports* **5,** 14600 (2015).
98. Martin-Gallausiaux, C., Marinelli, L., Blottière, H. M., Larraufie, P. & Lapaque, N. SCFA: mechanisms and functional importance in the gut. *Proceedings of the Nutrition Society,* 1–13. ISSN: 0029-6651 (2020).
99. Zhu, W. *et al.* Gut microbial metabolite TMAO enhances platelet hyperreactivity and thrombosis risk. *Cell* **165,** 111–124. ISSN: 0092-8674 (2016).
100. Wang, Z. *et al.* Non-lethal inhibition of gut microbial trimethylamine production for the treatment of atherosclerosis. *Cell* **163,** 1585–1595. ISSN: 0092-8674 (2015).
101. Dambrova, M *et al.* Diabetes is associated with higher trimethylamine N-oxide plasma levels. *Experimental and clinical endocrinology & diabetes* **124,** 251–256. ISSN: 0947-7349 (2016).
102. Schugar, R. C. *et al.* The TMAO-producing enzyme flavin-containing monooxygenase 3 regulates obesity and the beiging of white adipose tissue. *Cell reports* **19,** 2451–2461. ISSN: 2211-1247 (2017).
103. Wahlström, A., Sayin, S. I., Marschall, H.-U. & Bäckhed, F. Intestinal crosstalk between bile acids and microbiota and its impact on host metabolism. *Cell metabolism* **24,** 41–50. ISSN: 1550-4131 (2016).
104. Hylemon, P. B. *et al.* Bile acids as regulatory molecules. *Journal of lipid research* **50,** 1509–1520. ISSN: 0022-2275 (2009).
105. Joyce, S. A. & Gahan, C. G. Disease-associated changes in bile acid profiles and links to altered gut microbiota. *Digestive Diseases* **35,** 169–177. ISSN: 0257-2753 (2017).
106. Vítek, L. Bile acids in the treatment of cardiometabolic diseases. *Annals of hepatology* **16,** 43–52 (2018).
107. Brown, J. M. & Hazen, S. L. Microbial modulation of cardiovascular disease. *Nature Reviews Microbiology* **16,** 171–181. ISSN: 1740-1534 (2018).
108. Masella, R *et al.* Protocatechuic acid and human disease prevention: biological activities and molecular mechanisms. *Current medicinal chemistry* **19,** 2901–2917. ISSN: 0929-8673 (2012).
109. Menni, C. *et al.* Circulating levels of the anti-oxidant indoleproprionic acid are associated with higher gut microbiome diversity. *Gut microbes* **10,** 688–695. ISSN: 1949-0976 (2019).
110. Al-Waiz, M, Mikov, M, Mitchell, S. & Smith, R. The exogenous origin of trimethylamine in the mouse. *Metabolism* **41,** 135–136. ISSN: 0026-0495 (1992).
111. Zhu, W *et al.* Flavin monooxygenase 3, the host hepatic enzyme in the metaorganismal trimethylamine N-oxide-generating pathway, modulates platelet responsiveness and thrombosis risk. *Journal of Thrombosis and Haemostasis* **16,** 1857–1872. ISSN: 1538-7933 (2018).

112. Tang, W. W. *et al.* Gut microbiota-dependent trimethylamine N-oxide (TMAO) pathway contributes to both development of renal insufficiency and mortality risk in chronic kidney disease. *Circulation research* **116,** 448–455. ISSN: 0009-7330 (2015).
113. Heianza, Y., Ma, W., Manson, J. E., Rexrode, K. M. & Qi, L. Gut microbiota metabolites and risk of major adverse cardiovascular disease events and death: a systematic review and meta-analysis of prospective studies. *Journal of the American Heart Association* **6,** e004947. ISSN: 2047-9980 (2017).
114. Yang, S. *et al.* Gut microbiota-dependent marker TMAO in promoting cardiovascular disease: inflammation mechanism, clinical prognostic, and potential as a therapeutic target. *Frontiers in Pharmacology* **10** (2019).
115. Keitel, V., Kubitz, R. & Häussinger, D. Endocrine and paracrine role of bile acids. *World journal of gastroenterology: WJG* **14,** 5620 (2008).
116. Fiorucci, S. & Distrutti, E. Bile acid-activated receptors, intestinal microbiota, and the treatment of metabolic disorders. *Trends in molecular medicine* **21,** 702–714. ISSN: 1471-4914 (2015).
117. Khurana, S., Raufman, J. & Pallone, T. L. Bile acids regulate cardiovascular function. *Clinical and translational science* **4,** 210–218. ISSN: 1752-8054 (2011).
118. Aura, A.-M. *et al.* In vitro metabolism of anthocyanins by human gut microflora. *European journal of nutrition* **44,** 133–142. ISSN: 1436-6207 (2005).
119. De Mello, V. D. *et al.* Indolepropionic acid and novel lipid metabolites are associated with a lower risk of type 2 diabetes in the Finnish Diabetes Prevention Study. *Scientific reports* **7,** 46337. ISSN: 2045-2322 (2017).
120. Moss, G., Smith, P. & Tavernier, D. Glossary of class names of organic compounds and reactivity intermediates based on structure (IUPAC Recommendations 1995). *Pure and applied chemistry* **67,** 1307–1375. ISSN: 1365-3075 (1995).
121. Layden, B. T., Angueira, A. R., Brodsky, M., Durai, V. & Lowe Jr, W. L. Short chain fatty acids and their receptors: new metabolic targets. *Translational Research* **161,** 131–140. ISSN: 1931-5244 (2013).
122. Richards, L. B., Li, M., van Esch, B. C., Garssen, J. & Folkerts, G. The effects of short-chain fatty acids on the cardiovascular system. *PharmaNutrition* **4,** 68–111. ISSN: 2213-4344 (2016).
123. Henningsson, A. M., Bjorck, I. M. & Nyman, E. M. G. Combinations of indigestible carbohydrates affect short-chain fatty acid formation in the hindgut of rats. *The Journal of nutrition* **132,** 3098–3104. ISSN: 0022-3166 (2002).
124. Bergman, E. Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiological reviews* **70,** 567–590. ISSN: 0031-9333 (1990).
125. Macfarlane, J. Proteolysis and amino acid fermentation. *Human colonic bacteria,* 75–100 (1995).
126. Darzi, J., Frost, G. S. & Robertson, M. D. Do SCFA have a role in appetite regulation? *Proceedings of the Nutrition Society* **70,** 119–128. ISSN: 1475-2719 (2011).
127. McNeil, N. I., Cummings, J. & James, W. Short chain fatty acid absorption by the human large intestine. *Gut* **19,** 819–822. ISSN: 0017-5749 (1978).
128. Holtug, K, Clausen, M., Hove, H, Christiansen, J & Mortensen, P. The colon in carbohydrate malabsorption: short-chain fatty acids, pH, and osmotic diarrhoea. *Scandinavian journal of gastroenterology* **27,** 545–552. ISSN: 0036-5521 (1992).
129. Cummings, J., Pomare, E., Branch, W., Naylor, C. & Macfarlane, G. Short chain fatty acids in human large intestine, portal, hepatic and venous blood. *Gut* **28,** 1221–1227. ISSN: 0017-5749 (1987).

130. Macfarlane, S. & Macfarlane, G. T. Regulation of short-chain fatty acid production. *Proceedings of the Nutrition Society* **62,** 67–72. ISSN: 1475-2719 (2003).
131. Macfarlane, G., Gibson, G. & Cummings, J. Comparison of fermentation reactions in different regions of the human colon. *Journal of Applied Bacteriology* **72,** 57–64. ISSN: 0021-8847 (1992).
132. Flint, H. J., Duncan, S. H., Scott, K. P. & Louis, P. Links between diet, gut microbiota composition and gut metabolism. *Proceedings of the Nutrition Society* **74,** 13–22. ISSN: 0029-6651 (2015).
133. Morrison, D. J. & Preston, T. Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. *Gut microbes* **7,** 189–200. ISSN: 1949-0976 (2016).
134. Miller, T. L. & Wolin, M. J. Pathways of acetate, propionate, and butyrate formation by the human fecal microbial flora. *Applied and environmental microbiology* **62,** 1589–1592. ISSN: 0099-2240 (1996).
135. Rey, F. E. *et al.* Dissecting the in vivo metabolic potential of two human gut acetogens. *Journal of biological chemistry* **285,** 22082–22090. ISSN: 0021-9258 (2010).
136. Ragsdale, S. W. & Pierce, E. Acetogenesis and the Wood–Ljungdahl pathway of CO2 fixation. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **1784,** 1873–1898. ISSN: 1570-9639 (2008).
137. Reichardt, N. *et al.* Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. *The ISME journal* **8,** 1323–1335. ISSN: 1751-7370 (2014).
138. Macy, J. M., Ljungdahl, L. G. & Gottschalk, G. Pathway of succinate and propionate formation in Bacteroides fragilis. *Journal of bacteriology* **134,** 84–91. ISSN: 0021-9193 (1978).
139. Macy, J. M. & Probst, I. The biology of gastrointestinal bacteroides. *Annual Reviews in Microbiology* **33,** 561–594. ISSN: 0066-4227 (1979).
140. Marchandin, H. *et al.* Negativicoccus succinicivorans gen. nov., sp. nov., isolated from human clinical samples, emended description of the family Veillonellaceae and description of Negativicutes classis nov., Selenomonadales ord. nov. and Acidaminococcaceae fam. nov. in the bacterial phylum Firmicutes. *International journal of systematic and evolutionary microbiology* **60,** 1271–1279. ISSN: 1466-5026 (2010).
141. Saxena, R., Anand, P., Saran, S., Isar, J. & Agarwal, L. Microbial production and applications of 1, 2-propanediol. *Indian journal of microbiology* **50,** 2–11. ISSN: 0046-8991 (2010).
142. Bobik, T. A., Havemann, G. D., Busch, R. J., Williams, D. S. & Aldrich, H. C. The propanediol utilization (pdu) operon of Salmonella enterica serovar Typhimurium LT2 includes genes necessary for formation of Polyhedral organelles involved in coenzyme B12-dependent 1, 2-propanediol degradation. *Journal of bacteriology* **181,** 5967–5975. ISSN: 0021-9193 (1999).
143. Scott, K. P., Martin, J. C., Campbell, G., Mayer, C.-D. & Flint, H. J. Whole-genome transcription profiling reveals genes up-regulated by growth on fucose in the human gut bacterium “Roseburia inulinivorans”. *Journal of bacteriology* **188,** 4340–4349. ISSN: 0021-9193 (2006).
144. Belzer, C. *et al.* Microbial metabolic networks at the mucus layer lead to diet-independent butyrate and vitamin B12 production by intestinal symbionts. *MBio* **8.** ISSN: 2150-7511 (2017).

145. Ze, X., Le Mougen, F., Duncan, S. H., Louis, P. & Flint, H. J. Some are more equal than others: the role of "keystone" species in the degradation of recalcitrant substrates. *Gut Microbes* **4,** 236–240. ISSN: 1949-0976 (2013).

146. Duncan, S. H., Louis, P. & Flint, H. J. Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product. *Applied and environmental microbiology* **70,** 5810–5817. ISSN: 0099-2240 (2004).

147. Louis, P. & Flint, H. J. Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. *FEMS microbiology letters* **294,** 1–8. ISSN: 0378-1097 (2009).

148. Duncan, S. H., Barcenilla, A., Stewart, C. S., Pryde, S. E. & Flint, H. J. Acetate utilization and butyryl coenzyme A (CoA): acetate-CoA transferase in butyrate-producing bacteria from the human large intestine. *Applied and environmental microbiology* **68,** 5186–5190. ISSN: 0099-2240 (2002).

149. Louis, P., Young, P., Holtrop, G. & Flint, H. J. Diversity of human colonic butyrate-producing bacteria revealed by analysis of the butyryl-CoA: acetate CoA-transferase gene. *Environmental microbiology* **12,** 304–314. ISSN: 1462-2912 (2010).

150. Pluznick, J. L. Gut microbiota in renal physiology: focus on short-chain fatty acids and their receptors. *Kidney international* **90,** 1191–1198. ISSN: 0085-2538 (2016).

151. Pluznick, J. L. *et al.* Olfactory receptor responding to gut microbiota-derived signals plays a role in renin secretion and blood pressure regulation. *Proceedings of the National Academy of Sciences* **110,** 4410–4415. ISSN: 0027-8424 (2013).

152. Dalile, B., Van Oudenhove, L., Vervliet, B. & Verbeke, K. The role of short-chain fatty acids in microbiota–gut–brain communication. *Nature Reviews Gastroenterology & Hepatology,* 1. ISSN: 1759-5053 (2019).

153. Milligan, G., Stoddart, L. A. & Smith, N. J. Agonism and allosterism: the pharmacology of the free fatty acid receptors FFA2 and FFA3. *British journal of pharmacology* **158,** 146–153. ISSN: 0007-1188 (2009).

154. Thangaraju, M. *et al.* GPR109A is a G-protein–coupled receptor for the bacterial fermentation product butyrate and functions as a tumor suppressor in colon. *Cancer research* **69,** 2826–2832. ISSN: 0008-5472 (2009).

155. Donohoe, D. R. *et al.* The Warburg effect dictates the mechanism of butyrate-mediated histone acetylation and cell proliferation. *Molecular cell* **48,** 612–626. ISSN: 1097-2765 (2012).

156. Bose, P., Dai, Y. & Grant, S. Histone deacetylase inhibitor (HDACI) mechanisms of action: emerging insights. *Pharmacology & therapeutics* **143,** 323–336. ISSN: 0163-7258 (2014).

157. Alex, S. *et al.* Short-chain fatty acids stimulate angiopoietin-like 4 synthesis in human colon adenocarcinoma cells by activating peroxisome proliferator-activated receptor γ. *Molecular and cellular biology* **33,** 1303–1316. ISSN: 0270-7306 (2013).

158. Marinelli, L. *et al.* Identification of the novel role of butyrate as AhR ligand in human intestinal epithelial cells. *Scientific reports* **9,** 1–14. ISSN: 2045-2322 (2019).

159. Puddu, A., Sanguineti, R., Montecucco, F. & Viviani, G. L. Evidence for the gut microbiota short-chain fatty acids as key pathophysiological molecules improving diabetes. *Mediators of Inflammation* **2014.** ISSN: 0962-9351 (2014).

160. Theodorakis, M. J. *et al.* Human duodenal enteroendocrine cells: source of both incretin peptides, GLP-1 and GIP. *American Journal of Physiology-Endocrinology and Metabolism* **290,** E550–E559. ISSN: 0193-1849 (2006).

161. De Silva, A. & Bloom, S. R. Gut hormones and appetite control: a focus on PYY and GLP-1 as therapeutic targets in obesity. *Gut and liver* **6,** 10 (2012).

162. Savage, A., Adrian, T., Carolan, G, Chatterjee, V. & Bloom, S. Effects of peptide YY (PYY) on mouth to caecum intestinal transit time and on the rate of gastric emptying in healthy volunteers. *Gut* **28,** 166–170. ISSN: 0017-5749 (1987).
163. Naslund, E. *et al.* GLP-1 slows solid gastric emptying and inhibits insulin, glucagon, and PYY release in humans. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **277,** R910–R916. ISSN: 1522-1490 (1999).
164. Frost, G. *et al.* The short-chain fatty acid acetate reduces appetite via a central homeostatic mechanism. *Nature communications* **5,** 1–11. ISSN: 2041-1723 (2014).
165. Soliman, M. *et al.* Inverse regulation of leptin mRNA expression by short-and long-chain fatty acids in cultured bovine adipocytes. *Domestic animal endocrinology* **33,** 400–409. ISSN: 0739-7240 (2007).
166. Lee, S. & Hossner, K. Coordinate regulation of ovine adipose tissue gene expression by propionate. *Journal of animal science* **80,** 2840–2849. ISSN: 0021-8812 (2002).
167. Al-Lahham, S. H. *et al.* Regulation of adipokine production in human adipose tissue by propionic acid. *Eur J Clin Invest* **40,** 401–7. ISSN: 0014-2972 (2010).
168. Canfora, E. E., Jocken, J. W. & Blaak, E. E. Short-chain fatty acids in control of body weight and insulin sensitivity. *Nature Reviews Endocrinology* **11,** 577. ISSN: 1759-5037 (2015).
169. Zadeh-Tahmasebi, M. *et al.* Activation of short and long chain fatty acid sensing machinery in the ileum lowers glucose production in vivo. *Journal of Biological Chemistry* **291,** 8816–8824. ISSN: 0021-9258 (2016).
170. Pingitore, A. *et al.* The diet-derived short-chain fatty acid propionate improves beta-cell function in humans and stimulates insulin secretion from human islets in vitro. *Diabetes, Obesity and Metabolism* **19,** 257–265. ISSN: 1462-8902 (2017).
171. Li, Q., Chen, H., Zhang, M., Wu, T. & Liu, R. Altered short chain fatty acid profiles induced by dietary fiber intervention regulate AMPK levels and intestinal homeostasis. *Food & function* **10,** 7174–7187 (2019).
172. Yoshida, H., Ishii, M. & Akagawa, M. Propionate suppresses hepatic gluconeogenesis via GPR43/AMPK signaling pathway. *Archives of biochemistry and biophysics* **672,** 108057. ISSN: 0003-9861 (2019).
173. Chambers, E. S. *et al.* Effects of targeted delivery of propionate to the human colon on appetite regulation, body weight maintenance and adiposity in overweight adults. *Gut* **64,** 1744–1754. ISSN: 0017-5749 (2015).
174. Yang, L., Lin, H., Lin, W. & Xu, X. Exercise Ameliorates Insulin Resistance of Type 2 Diabetes through Motivating Short-Chain Fatty Acid-Mediated Skeletal Muscle Cell Autophagy. *Biology* **9,** 203 (2020).
175. Velazquez, O. C., Lederer, H. M. & Rombeau, J. L. in *Dietary fiber in health and disease* 123–134 (Springer, 1997).
176. Lupton, J. R. Microbial degradation products influence colon cancer risk: the butyrate controversy. *The Journal of nutrition* **134,** 479–482. ISSN: 0022-3166 (2004).
177. Wang, H. *et al.* Dietary non-digestible polysaccharides ameliorate intestinal epithelial barrier dysfunction in IL-10 knockout mice. *Journal of Crohn's and Colitis* **10,** 1076–1086. ISSN: 1876-4479 (2016).
178. Hung, T. V. & Suzuki, T. Dietary fermentable fibers attenuate chronic kidney disease in mice by protecting the intestinal barrier. *The Journal of Nutrition* **148,** 552–561. ISSN: 0022-3166 (2018).
179. Kelly, C. J. *et al.* Crosstalk between microbiota-derived short-chain fatty acids and intestinal epithelial HIF augments tissue barrier function. *Cell host & microbe* **17,** 662–671. ISSN: 1931-3128 (2015).

180. Munford, R. S. Endotoxemia—menace, marker, or mistake? *Journal of leukocyte biology* **100,** 687–698. ISSN: 0741-5400 (2016).

181. Filardo, S., Di Pietro, M., Farcomeni, A., Schiavoni, G. & Sessa, R. Chlamydia pneumoniae-mediated inflammation in atherosclerosis: a meta-analysis. *Mediators of inflammation* **2015.** ISSN: 0962-9351 (2015).

182. Koren, O. *et al.* Human oral, gut, and plaque microbiota in patients with atherosclerosis. *Proceedings of the National Academy of Sciences* **108,** 4592–4598. ISSN: 0027-8424 (2011).

183. Hamer, H. M. *et al.* The role of butyrate on colonic function. *Alimentary pharmacology & therapeutics* **27,** 104–119. ISSN: 0269-2813 (2008).

184. Gaudier, E *et al.* Butyrate specifically modulates MUC gene expression in intestinal epithelial goblet cells deprived of glucose. *American Journal of Physiology-Gastrointestinal and Liver Physiology* **287,** G1168–G1174. ISSN: 0193-1857 (2004).

185. Zhao, Y. *et al.* GPR43 mediates microbiota metabolite SCFA regulation of antimicrobial peptide expression in intestinal epithelial cells via activation of mTOR and STAT3. *Mucosal immunology* **11,** 752–762. ISSN: 1935-3456 (2018).

186. Lührs, H. *et al.* Butyrate inhibits NF-$\varkappa$B activation in lamina propria macrophages of patients with ulcerative colitis. *Scandinavian journal of gastroenterology* **37,** 458–466. ISSN: 0036-5521 (2002).

187. Maeda, T., Towatari, M., Kosugi, H. & Saito, H. Up-regulation of costimulatory/adhesion molecules by histone deacetylase inhibitors in acute myeloid leukemia cells. *Blood, The Journal of the American Society of Hematology* **96,** 3847–3856. ISSN: 1528-0020 (2000).

188. Glauben, R. & Siegmund, B. Inhibition of histone deacetylases in inflammatory bowel diseases. *Molecular medicine* **17,** 426–433. ISSN: 1076-1551 (2011).

189. Zaki, M. H. *et al.* The NLRP3 inflammasome protects against loss of epithelial integrity and mortality during experimental colitis. *Immunity* **32,** 379–391. ISSN: 1074-7613 (2010).

190. Macia, L. *et al.* Metabolite-sensing receptors GPR43 and GPR109A facilitate dietary fibre-induced gut homeostasis through regulation of the inflammasome. *Nature communications* **6,** 6734. ISSN: 2041-1723 (2015).

191. Huang, N., Katz, J. P., Martin, D. R. & Wu, G. D. Inhibition of IL-8 gene expression in Caco-2 cells by compounds which induce histone hyperacetylation. *Cytokine* **9,** 27–36. ISSN: 1043-4666 (1997).

192. Atarashi, K. *et al.* T reg induction by a rationally selected mixture of Clostridia strains from the human microbiota. *Nature* **500,** 232–236. ISSN: 1476-4687 (2013).

193. Nastasi, C. *et al.* The effect of short-chain fatty acids on human monocyte-derived dendritic cells. *Scientific reports* **5,** 1–10. ISSN: 2045-2322 (2015).

194. Slavin, J. Fiber and prebiotics: mechanisms and health benefits. *Nutrients* **5,** 1417–1435 (2013).

195. Cobo, E. R., Kissoon-Singh, V., Moreau, F., Holani, R. & Chadee, K. MUC2 mucin and butyrate contribute to the synthesis of the antimicrobial peptide cathelicidin in response to Entamoeba histolytica-and dextran sodium sulfate-induced colitis. *Infection and immunity* **85.** ISSN: 0019-9567 (2017).

196. Schauber, J., Dorschner, R. A., Yamasaki, K., Brouha, B. & Gallo, R. L. Control of the innate epithelial antimicrobial response is cell-type specific and dependent on relevant microenvironmental stimuli. *Immunology* **118,** 509–519. ISSN: 0019-2805 (2006).

197. Furusawa, Y. *et al.* Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* **504,** 446–450. ISSN: 1476-4687 (2013).
198. Al-Lahham, S. *et al.* Propionic acid affects immune status and metabolism in adipose tissue from overweight subjects. *European journal of clinical investigation* **42,** 357–364. ISSN: 0014-2972 (2012).
199. Li, G., Yao, W. & Jiang, H. Short-chain fatty acids enhance adipocyte differentiation in the stromal vascular fraction of porcine adipose tissue. *The Journal of nutrition* **144,** 1887–1895. ISSN: 1541-6100 (2014).
200. Hong, Y.-H. *et al.* Acetate and propionate short-chain fatty acids stimulate adipogenesis via GPCR43. *Endocrinology* **146,** 5092–5099. ISSN: 0013-7227 (2005).
201. Zhao, Y. *et al.* Structure-specific effects of short-chain fatty acids on plasma cholesterol concentration in male syrian hamsters. *Journal of agricultural and food chemistry* **65,** 10984–10992. ISSN: 0021-8561 (2017).
202. Demigné, C. *et al.* Effect of propionate on fatty acid and cholesterol synthesis and on acetate metabolism in isolated rat hepatocytes. *British journal of nutrition* **74,** 209–219. ISSN: 1475-2662 (1995).
203. Kang, C. *et al.* Gut microbiota mediates the protective effects of dietary capsaicin against chronic low-grade inflammation and associated obesity induced by high-fat diet. *MBio* **8.** ISSN: 2150-7511 (2017).
204. Cani, P. D. *et al.* Selective increases of bifidobacteria in gut microflora improve high-fat-diet-induced diabetes in mice through a mechanism associated with endotoxaemia. *Diabetologia* **50,** 2374–2383. ISSN: 0012-186X (2007).
205. Dao, M. C. *et al.* Akkermansia muciniphila and improved metabolic health during a dietary intervention in obesity: relationship with gut microbiome richness and ecology. *Gut* **65,** 426–436. ISSN: 0017-5749 (2016).
206. Kimura, I. *et al.* The gut microbiota suppresses insulin-mediated fat accumulation via the short-chain fatty acid receptor GPR43. *Nature communications* **4,** 1–12. ISSN: 2041-1723 (2013).
207. Lin, H. V. *et al.* Butyrate and propionate protect against diet-induced obesity and regulate gut hormones via free fatty acid receptor 3-independent mechanisms. *PloS one* **7,** e35240. ISSN: 1932-6203 (2012).
208. Gao, Z. *et al.* Butyrate improves insulin sensitivity and increases energy expenditure in mice. *Diabetes* **58,** 1509–1517. ISSN: 0012-1797 (2009).
209. Sakakibara, S., Yamauchi, T., Oshima, Y., Tsukamoto, Y. & Kadowaki, T. Acetic acid activates hepatic AMPK and reduces hyperglycemia in diabetic KK-A (y) mice. *Biochemical and biophysical research communications* **344,** 597–604. ISSN: 0006-291X (2006).
210. Roshanravan, N. *et al.* Effect of butyrate and inulin supplementation on glycemic status, lipid profile and glucagon-like peptide 1 level in patients with type 2 diabetes: a randomized double-blind, placebo-controlled trial. *Hormone and metabolic research* **49,** 886–891. ISSN: 0018-5043 (2017).
211. Bartolomaeus, H. *et al.* Short-chain fatty acid propionate protects from hypertensive cardiovascular damage. *Circulation* **139,** 1407–1421. ISSN: 0009-7322 (2019).
212. Kim, K. N., Yao, Y. & Ju, S. Y. Short chain fatty acids and fecal microbiota abundance in humans with obesity: A systematic review and meta-analysis. *Nutrients* **11,** 2512 (2019).
213. Tang, C. *et al.* Loss of FFA2 and FFA3 increases insulin secretion and improves glucose tolerance in type 2 diabetes. *Nature medicine* **21,** 173–177 (2015).

214. Müller, M. *et al.* Circulating but not faecal short-chain fatty acids are related to insulin sensitivity, lipolysis and GLP-1 concentrations in humans. *Scientific Reports* **9,** 1–9 (2019).

215. Vinolo, M. A. *et al.* Short-chain fatty acids stimulate the migration of neutrophils to inflammatory sites. *Clinical science* **117,** 331–338 (2009).

216. Vinolo, M. A., Rodrigues, H. G., Nachbar, R. T. & Curi, R. Regulation of inflammation by short chain fatty acids. *Nutrients* **3,** 858–876 (2011).

217. Halili, M. A. *et al.* Differential effects of selective HDAC inhibitors on macrophage inflammatory responses to the Toll-like receptor 4 agonist LPS. *Journal of leukocyte biology* **87,** 1103–1114 (2010).

218. Verdi, S. *et al.* TwinsUK: the UK adult twin registry update. *Twin Research and Human Genetics* **22,** 523–529. ISSN: 1832-4274 (2019).

219. Moayyeri, A., Hammond, C. J., Valdes, A. M. & Spector, T. D. Cohort Profile: TwinsUK and healthy ageing twin study. *Int J Epidemiol* **42,** 76–85. ISSN: 0300-5771 (Print) 0300-5771 (2013).

220. Office for National Statistics. *Population estimates for the UK, England and Wales, Scotland and Northern Ireland* Government Document. 2021.

221. Verdi, S. *et al.* TwinsUK: The UK Adult Twin Registry Update. *Twin Res Hum Genet* **22,** 523–529. ISSN: 1832-4274 (Print) 1832-4274 (Linking). https://www.ncbi.nlm.nih.gov/pubmed/31526404 (2019).

222. Jarrar, Z. A. *et al.* Definitive Zygosity Scores in the Peas in the Pod Questionnaire is a Sensitive and Accurate Assessment of the Zygosity of Adult Twins. *Twin Res Hum Genet* **21,** 146–154. ISSN: 1832-4274 (Print) 1832-4274 (2018).

223. Stergiou, G. S. *et al.* 2021 European Society of Hypertension practice guidelines for office and out-of-office blood pressure measurement. *Journal of Hypertension* **39,** 1293–1302. ISSN: 0263-6352. https://journals.lww.com/jhypertension/Fulltext/2021/07000/2021_European_Society_of_Hypertension_practice.5.aspx (2021).

224. Association, A. D. 2. Classification and diagnosis of diabetes: standards of medical care in diabetes—2019. *Diabetes care* **42,** S13–S28 (2019).

225. Kaul, S. *et al.* Dual-energy X-ray absorptiometry for quantification of visceral fat. *Obesity* **20,** 1313–1318 (2012).

226. Bingham, S. A. *et al.* Nutritional methods in the European Prospective Investigation of Cancer in Norfolk. *Public Health Nutr* **4,** 847–58 (2001).

227. Bingham, S. A. *et al.* Validation of dietary assessment methods in the UK arm of EPIC using weighed records, and 24-hour urinary nitrogen and potassium and serum vitamin C and carotenoids as biomarkers. *International journal of epidemiology* **26,** S137. ISSN: 1464-3685 (1997).

228. Mulligan, A. A. *et al.* A new tool for converting food frequency questionnaire data into nutrient and food group values: FETA research methods and availability. *BMJ Open* **4,** e004503. ISSN: 2044-6055 (Print) 2044-6055 (2014).

229. Holland, B. *et al. McCance and Widdowson's The Composition of Foods* 5th, xiii + 462 pp. ISBN: 0851863914 (Royal Society of Chemistry, Cambridge, 1991).

230. Mompeo, O. *et al.* Genetic and environmental influences of dietary indices in a UK female twin cohort. *Twin Research and Human Genetics* **23,** 330–337. ISSN: 1832-4274. https://www.cambridge.org/core/article/genetic-and-environmental-influences-of-dietary-indices-in-a-uk-female-twin-cohort/48632224BDF4C60E69FDE67F8E21A1C7 (2020).

231. Kennedy, E. T., Ohls, J., Carlson, S. & Fleming, K. The Healthy Eating Index: Design and Applications. *Journal of the American Dietetic Association* **95,**

1103–1108. ISSN: 0002-8223. https://doi.org/10.1016/S0002-8223(95)00300-2 (1995).

232. Schwingshackl, L., Bogensberger, B. & Hoffmann, G. Diet Quality as Assessed by the Healthy Eating Index, Alternate Healthy Eating Index, Dietary Approaches to Stop Hypertension Score, and Health Outcomes: An Updated Systematic Review and Meta-Analysis of Cohort Studies. *J Acad Nutr Diet* **118,** 74–100.e11. ISSN: 2212-2672 (Print) 2212-2672 (2018).
233. T Kennedy, E., Ohls, J., Carlson, S. & Fleming, K. The healthy eating index: design and applications. *Journal of the American dietetic association* **95,** 1103–1108 (1995).
234. Chiuve, S. E. *et al.* Alternative dietary indices both strongly predict risk of chronic disease. *The Journal of nutrition* **142,** 1009–1018 (2012).
235. Menni, C. *et al.* Metabolomic markers reveal novel pathways of ageing and early development in human populations. *International journal of epidemiology* **42,** 1111–1119. ISSN: 1464-3685 (2013).
236. Zierer, J. *et al.* The fecal metabolome as a functional readout of the gut microbiome. *Nature genetics* **50,** 790–795. ISSN: 1546-1718 (2018).
237. Vehtari, A. *et al.* A novel Bayesian approach to quantify clinical variables and to determine their spectroscopic counterparts in 1H NMR metabonomic data. *Bmc Bioinformatics* **8,** 1–9 (2007).
238. Goodrich, J. K. *et al.* Genetic determinants of the gut microbiome in UK twins. *Cell host & microbe* **19,** 731–743 (2016).
239. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature methods* **7,** 335–336 (2010).
240. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nature methods* **13,** 581–583 (2016).
241. Wells, P. M. *et al.* Associations between gut microbiota and genetic risk for rheumatoid arthritis in the absence of disease: a cross-sectional study. *The Lancet Rheumatology* **2,** e418–e427 (2020).
242. Lahti, L., Shetty, S., *et al.* Introduction to the microbiome R package. *Bioconductor 2018Available online: https://www. bioconductor. org/packages/release/bioc/html/microbiome. html (accessed on 15 October 2022)* (2018).
243. Visconti, A., Martin, T. C. & Falchi, M. YAMP: a containerized workflow enabling reproducibility in metagenomics research. *Gigascience* **7,** giy072 (2018).
244. Jones, M. B. *et al.* Library preparation methodology can influence genomic and functional predictions in human microbiome research. *Proceedings of the National Academy of Sciences* **112,** 14024–14029 (2015).
245. Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J. & Segata, N. Shotgun metagenomics, from sampling to analysis. *Nature biotechnology* **35,** 833–844 (2017).
246. McIver, L. J. *et al.* bioBakery: a meta'omic analysis environment. *Bioinformatics* **34,** 1235–1237 (2018).
247. Beghini, F. *et al.* Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *elife* **10,** e65088 (2021).
248. Blanco-Míguez, A. *et al.* Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nature Biotechnology,* 1–12 (2023).

249. Berry, S. *et al.* Personalised REsponses to DIetary Composition Trial (PREDICT): an intervention study to determine inter-individual differences in postprandial response to foods (2020).
250. Yu, B. *et al.* The Consortium of Metabolomics Studies (COMETS): metabolomics in 47 prospective cohort studies. *American journal of epidemiology* **188,** 991–1012 (2019).
251. Wright, J. D. *et al.* The ARIC (atherosclerosis risk in communities) study: JACC focus seminar 3/8. *Journal of the American College of Cardiology* **77,** 2939–2959 (2021).
252. Price, J. F. *et al.* The Edinburgh type 2 diabetes study: study protocol. *BMC endocrine disorders* **8,** 1–10 (2008).
253. Sierra, A. *et al.* The GenoDiabMar Registry: A Collaborative Research Platform of Type 2 Diabetes Patients. *Journal of Clinical Medicine* **11,** 1431 (2022).
254. Santanasto, A. J. *et al.* Body composition remodeling and mortality: the health aging and body composition study. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences* **72,** 513–519 (2017).
255. Suhre, K. *et al.* Human metabolic individuality in biomedical and pharmaceutical research. *Nature* **477,** 54–60. ISSN: 1476-4687 (2011).
256. Paynter, N. P. *et al.* Metabolic predictors of incident coronary heart disease in women. *Circulation* **137,** 841–853 (2018).
257. Study, T. W. H. I. *et al.* Design of the Women's Health Initiative clinical trial and observational study. *Controlled clinical trials* **19,** 61–109 (1998).
258. Team, R. C. *R: A language and environment for statistical computing.* Computer Program. 2020. https://www.R-project.org/.
259. Thissen, D., Steinberg, L. & Kuang, D. Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *J. Educ. Behav. Stat.* **27,** 77–83. ISSN: 1076-9986. http://journals.sagepub.com/doi/10.3102/10769986027001077http://dx.doi.org/10.3102/10769986027001077 (2002).
260. Team, R. D. C. A language and environment for statistical computing. *http://www.R-project. org* (2009).
261. Therneau, T. & Lumley, T. R survival package. *R Core Team* (2013).
262. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. lmerTest package: tests in linear mixed effects models. *Journal of statistical software* **82,** 1–26 (2017).
263. Borenstein, M., Hedges, L. V., Higgins, J. P. & Rothstein, H. R. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research synthesis methods* **1,** 97–111 (2010).
264. Schwarzer, G. *et al.* meta: An R package for meta-analysis. *R news* **7,** 40–45 (2007).
265. Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *The American Journal of Human Genetics* **88,** 586–598 (2011).
266. Krämer, A., Green, J., Pollard Jack, J. & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30,** 523–530. ISSN: 1367-4803. https://doi.org/10.1093/bioinformatics/btt703 (2013).
267. Pang, Z. *et al.* MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights, journal = Nucleic acids research. **49,** W388–W396. ISSN: 0305-1048 (2021).
268. Couronné, R., Probst, P. & Boulesteix, A.-L. Random forest versus logistic regression: a large-scale benchmark experiment. *BMC bioinformatics* **19,** 1–14 (2018).

269. Qi, Y. Random forest for bioinformatics. *Ensemble machine learning: Methods and applications,* 307–323 (2012).
270. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. *Advances in neural information processing systems* **30** (2017).
271. Asnicar, F. *et al.* Microbiome connections with host metabolism and habitual diet from 1,098 deeply phenotyped individuals. *Nature Medicine* **27,** 321–332 (2021).
272. Kuhn, M. Building predictive models in R using the caret package. *Journal of statistical software* **28,** 1–26 (2008).
273. Liaw, A., Wiener, M., *et al.* Classification and regression by randomForest. *R news* **2,** 18–22 (2002).
274. Berry, S. *et al.* Influence of gut microbial communities on fasting and postprandial lipids and circulating metabolites: the PREDICT 1 study. *Current Developments in Nutrition* **4,** 4141547 (2020).
275. Tingley, D., Yamamoto, T., Hirose, K., Keele, L. & Imai, K. Mediation: R package for causal mediation analysis. ISSN: 1548-7660 (2014).
276. Neale, M. & Cardon, L. R. *Methodology for Genetic Studies of Twins and Families* 496. ISBN: 9789401580182. https://play.google.com/store/books/details?id=EKYyBwAAQBAJ (Springer Science & Business Media, 2013).
277. Scheike, T. H., Holst, K. K. & Hjelmborg, J. B. Estimating heritability for cause specific mortality based on twin studies. *Lifetime data analysis* **20,** 210–233. ISSN: 1380-7870 (2014).
278. Almeida, A. *et al.* A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nature biotechnology* **39,** 105–114 (2021).
279. Sayers, E. W. *et al.* Database resources of the national center for biotechnology information. *Nucleic acids research* **49,** D10 (2021).
280. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research* **25,** 1043–1055 (2015).
281. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29,** 1072–1075 (2013).
282. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30,** 2068–2069 (2014).
283. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature communications* **9,** 5114 (2018).
284. Csardi, G., Nepusz, T., *et al.* The igraph software package for complex network research. *InterJournal, complex systems* **1695,** 1–9 (2006).
285. Wu, Y.-W. ezTree: an automated pipeline for identifying phylogenetic marker genes and inferring evolutionary relationships among uncultivated prokaryotic draft genomes. *BMC genomics* **19,** 7–16 (2018).
286. Caspi, R. *et al.* The MetaCyc database of metabolic pathways and enzymes. *Nucleic acids research* **46,** D633–D639 (2018).
287. Kanehisa, M. *et al.* Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic acids research* **42,** D199–D205 (2014).
288. Ye, Y. & Doak, T. G. A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. *PLoS computational biology* **5,** e1000465 (2009).
289. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32,** 2847–2849 (2016).

290. Nogal, A. *et al.* Circulating levels of the short-chain fatty acid acetate mediate the effect of the gut microbiome on visceral fat. *Frontiers in microbiology* **12,** 711359 (2021).
291. Wang, Z. *et al.* Metabolomic pattern predicts incident coronary heart disease: findings from the Atherosclerosis Risk in Communities Study. *Arteriosclerosis, thrombosis, and vascular biology* **39,** 1475–1482. ISSN: 1079-5642 (2019).
292. Bhosle, A., Wang, Y., Franzosa, E. A. & Huttenhower, C. Progress and opportunities in microbial community metabolomics. *Current Opinion in Microbiology* **70,** 102195 (2022).
293. Zhang, Z. *et al.* Characteristics of the gut microbiome in patients with prediabetes and type 2 diabetes. *PeerJ* **9,** e10952 (2021).
294. Campos-Perez, W. & Martinez-Lopez, E. Effects of short chain fatty acids on metabolic and inflammatory processes in human health. *Biochim Biophys Acta Mol Cell Biol Lipids* **1866,** 158900. ISSN: 1388-1981 (2021).
295. Menni, C. *et al.* Targeted metabolomics profiles are strongly correlated with nutritional patterns in women. *Metabolomics* **9,** 506–514 (2013).
296. Li, Z. & Vance, D. E. Thematic review series: glycerolipids. Phosphatidylcholine and choline homeostasis. *Journal of lipid research* **49,** 1187–1194. ISSN: 0022-2275 (2008).
297. Petersen, M. C. & Shulman, G. I. Mechanisms of insulin action and insulin resistance. *Physiological reviews* **98,** 2133–2223. ISSN: 0031-9333 (2018).
298. Navarrete-Muñoz, E.-M. *et al.* Dietary folate intake and metabolic syndrome in participants of PREDIMED-Plus study: a cross-sectional study. *European journal of nutrition,* 1–12. ISSN: 1436-6215 (2020).
299. Fernandez-Mejia, C. Pharmacological effects of biotin. *The Journal of nutritional biochemistry* **16,** 424–427. ISSN: 0955-2863 (2005).
300. Alaei-Shahmiri, F, Soares, M., Zhao, Y & Sherriff, J. The impact of thiamine supplementation on blood pressure, serum lipids and C-reactive protein in individuals with hyperglycemia: a randomised, double-blind cross-over trial. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* **9,** 213–217. ISSN: 1871-4021 (2015).
301. Meyer, R. K. *et al.* Oligofructose restores postprandial short-chain fatty acid levels during high-fat feeding. *Obesity* **30,** 1442–1452. ISSN: 1930-7381 (2022).
302. Kirschner, S. K., Ten Have, G. A., Engelen, M. P. & Deutz, N. E. Transorgan short-chain fatty acid fluxes in the fasted and postprandial state in the pig. *American Journal of Physiology-Endocrinology and Metabolism* **321,** E665–E673. ISSN: 0193-1849 (2021).
303. Gibbons, S. M., Kearney, S. M., Smillie, C. S. & Alm, E. J. Two dynamic regimes in the human gut microbiome. *PLoS computational biology* **13,** e1005364 (2017).
304. DeJong, E. N., Surette, M. G. & Bowdish, D. M. The gut microbiota and unhealthy aging: disentangling cause from consequence. *Cell Host & Microbe* **28,** 180–189 (2020).
305. VanEvery, H., Franzosa, E. A., Nguyen, L. H. & Huttenhower, C. Microbiome epidemiology and association studies in human health. *Nature Reviews Genetics* **24,** 109–124 (2023).
306. Lee, J. M., Gianchandani, E. P. & Papin, J. A. Flux balance analysis in the era of metabolomics. *Briefings in bioinformatics* **7,** 140–150 (2006).
307. Cordaillat-Simmons, M., Rouanet, A. & Pot, B. Live biotherapeutic products: the importance of a defined regulatory framework. *Experimental & molecular medicine* **52,** 1397–1406 (2020).

308. Yadegar, A. *et al.* Beneficial effects of fecal microbiota transplantation in recurrent Clostridioides difficile infection. *Cell Host & Microbe* **31,** 695–711 (2023).
309. Hediyal, T. A. *et al.* Protective effects of fecal microbiota transplantation against ischemic stroke and other neurological disorders: an update. *Frontiers in Immunology* **15,** 1324018 (2024).
310. Bénard, M. V. *et al.* Challenges and costs of donor screening for fecal microbiota transplantations. *PLoS One* **17,** e0276323 (2022).
311. Kootte, R. S. *et al.* Improvement of insulin sensitivity after lean donor feces in metabolic syndrome is driven by baseline intestinal microbiota composition. *Cell metabolism* **26,** 611–619 (2017).
312. Vrieze, A. *et al.* Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome. *Gastroenterology* **143,** 913–916 (2012).
313. Mocanu, V. *et al.* Fecal microbial transplantation and fiber supplementation in patients with severe obesity and metabolic syndrome: a randomized double-blind, placebo-controlled phase 2 trial. *Nature Medicine* **27,** 1272–1279 (2021).
314. Roume, H. *et al.* A biomolecular isolation framework for eco-systems biology. *The ISME journal* **7,** 110–121 (2013).
315. Dorrestein, P. C., Mazmanian, S. K. & Knight, R. Finding the missing links among metabolites, microbes, and the host. *Immunity* **40,** 824–832 (2014).
316. Sun, S. *et al.* On the robustness of inference of association with the gut microbiota in stool, rectal swab and mucosal tissue samples. *Scientific Reports* **11,** 14828 (2021).
317. Juge, N. Relationship between mucosa-associated gut microbiota and human diseases. *Biochemical Society Transactions* **50,** 1225–1236 (2022).
318. Soverini, M. *et al.* HumanMycobiomeScan: a new bioinformatics tool for the characterization of the fungal fraction in metagenomic samples. *BMC genomics* **20,** 1–7 (2019).
319. Ru, J., Khan Mirzaei, M., Xue, J., Peng, X. & Deng, L. ViroProfiler: a containerized bioinformatics pipeline for viral metagenomic data analysis. *Gut Microbes* **15,** 2192522 (2023).
320. Zolfo, M. *et al.* Discovering and exploring the hidden diversity of human gut viruses using highly enriched virome samples. *bioRxiv,* 2024–02 (2024).
321. Jannasch, F. *et al.* Associations between exploratory dietary patterns and incident type 2 diabetes: a federated meta-analysis of individual participant data from 25 cohort studies. *European Journal of Nutrition* **61,** 3649–3667 (2022).
322. Ndanuko, R. N., Tapsell, L. C., Charlton, K. E., Neale, E. P. & Batterham, M. J. Dietary patterns and blood pressure in adults: a systematic review and meta-analysis of randomized controlled trials. *Advances in Nutrition* **7,** 76–89 (2016).
323. Kirkpatrick, C. F. *et al.* Nutrition interventions for adults with dyslipidemia: a clinical perspective from the National Lipid Association. *Journal of Clinical Lipidology* (2023).
324. Wang, L. *et al.* Methods to determine intestinal permeability and bacterial translocation during liver disease. *Journal of immunological methods* **421,** 44–53 (2015).

References from Chapter 4

1. Berry, S. E. *et al.* Human postprandial responses to food and potential for precision nutrition. *Nature medicine* **26,** 964–973. ISSN: 1546-170X (2020).
2. Cavus, E. *et al.* Association of circulating metabolites with risk of coronary heart disease in a European population: results from the Biomarkers for Cardiovascular Risk Assessment in Europe (BiomarCaRE) consortium. *JAMA cardiology* **4,** 1270–1279. ISSN: 2380-6583 (2019).
3. Murthy, V. L. *et al.* Comprehensive metabolic phenotyping refines cardiovascular risk in young adults. *Circulation* **142,** 2110–2127. ISSN: 0009-7322 (2020).
4. Ussher, J. R., Elmariah, S., Gerszten, R. E. & Dyck, J. R. The emerging role of metabolomics in the diagnosis and prognosis of cardiovascular disease. *Journal of the American College of Cardiology* **68,** 2850–2870. ISSN: 0735-1097 (2016).
5. Wang, Z. *et al.* Metabolomic pattern predicts incident coronary heart disease: findings from the Atherosclerosis Risk in Communities Study. *Arteriosclerosis, thrombosis, and vascular biology* **39,** 1475–1482. ISSN: 1079-5642 (2019).

References from Chapter 5

1. Virani, S. S. *et al.* Heart disease and stroke statistics—2021 update: a report from the American Heart Association. *Circulation* **143,** e254–e743. ISSN: 0009-7322 (2021).
2. Yusuf, S. *et al.* Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *The lancet* **364,** 937–952. ISSN: 0140-6736 (2004).
3. Koeth, R. A. *et al.* Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nature medicine* **19,** 576–585. ISSN: 1546-170X (2013).
4. Tang, W. W. *et al.* Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *New England Journal of Medicine* **368,** 1575–1584. ISSN: 0028-4793 (2013).
5. Ding, Y. *et al.* Plasma glycine and risk of acute myocardial infarction in patients with suspected stable angina pectoris. *Journal of the American Heart Association* **5,** e002621. ISSN: 2047-9980 (2015).
6. McKirnan, M. D. *et al.* Metabolomic analysis of serum and myocardium in compensated heart failure after myocardial infarction. *Life sciences* **221,** 212–223. ISSN: 0024-3205 (2019).
7. Hunter, W. G., Kelly, J. P., McGarrah, R. W., Kraus, W. E. & Shah, S. H. Metabolic dysfunction in heart failure: diagnostic, prognostic, and pathophysiologic insights from metabolomic profiling. *Current heart failure reports* **13,** 119–131. ISSN: 1546-9549 (2016).
8. Shah, S. H. & Hunter, W. G. *Realizing the potential of metabolomics in heart failure: signposts on the path to clinical utility* Generic. 2017.
9. Shah, S. H. *et al.* Association of a peripheral blood metabolic profile with coronary artery disease and risk of subsequent cardiovascular events. *Circulation: Cardiovascular Genetics* **3,** 207–214. ISSN: 1942-325X (2010).
10. Cheng, S. *et al.* Potential impact and study considerations of metabolomics in cardiovascular health and disease: a scientific statement from the American Heart Association. *Circulation: Cardiovascular Genetics* **10,** e000032. ISSN: 1942-325X (2017).
11. Paynter, N. P. *et al.* Metabolic predictors of incident coronary heart disease in women. *Circulation* **137,** 841–853. ISSN: 0009-7322 (2018).
12. Nogal, A. *et al.* Incremental Value of a Panel of Serum Metabolites for Predicting Risk of Atherosclerotic Cardiovascular Disease. *Journal of the American Heart Association* **11,** e024590. ISSN: 2047-9980 (2022).
13. Schmidt, D. R. *et al.* Metabolomics in cancer research and emerging applications in clinical oncology. *CA: a cancer journal for clinicians* **71,** 333–358. ISSN: 0007-9235 (2021).

14. Cicalini, I. *et al.* Integrated lipidomics and metabolomics analysis of tears in multiple sclerosis: an insight into diagnostic potential of lacrimal fluid. *International journal of molecular sciences* **20,** 1265 (2019).
15. Yu, B. *et al.* The consortium of metabolomics studies (COMETS): Metabolomics in 47 prospective cohort studies. *American journal of epidemiology* **188,** 991–1012. ISSN: 0002-9262 (2019).
16. Paynter, N. P. *et al.* Metabolic Predictors of Incident Coronary Heart Disease in Women. *Circulation* **137,** 841–853. ISSN: 0009-7322 (Print) 0009-7322 (2018).
17. Wright, J. D. *et al.* The ARIC (atherosclerosis risk in communities) study: JACC focus seminar 3/8. *Journal of the American College of Cardiology* **77,** 2939–2959. ISSN: 1558-3597 (2021).
18. Price, J. F. *et al.* The Edinburgh type 2 diabetes study: study protocol. *BMC endocrine disorders* **8,** 1–10. ISSN: 1472-6823 (2008).
19. Sierra, A. *et al.* The GenoDiabMar Registry: A Collaborative Research Platform of Type 2 Diabetes Patients. *Journal of clinical medicine* **11,** 1431. ISSN: 2077-0383 (2022).
20. Santanasto, A. J. *et al.* Body composition remodeling and mortality: the health aging and body composition study. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences* **72,** 513–519. ISSN: 1079-5006 (2017).
21. Han, S. *et al.* TIGER: technical variation elimination for metabolomics data using ensemble learning architecture. *Briefings in Bioinformatics* (2022).
22. Verdi, S. *et al.* TwinsUK: the UK adult twin registry update. *Twin Research and Human Genetics* **22,** 523–529. ISSN: 1832-4274 (2019).
23. Group, W. H. I. S. Design of the women's health initiative clinical trial and observation study. *Control Clin Trials* **19,** 61–109 (1998).
24. Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *The American Journal of Human Genetics* **88,** 586–598. ISSN: 0002-9297 (2011).
25. Thissen, D., Steinberg, L. & Kuang, D. Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of educational and behavioral statistics* **27,** 77–83. ISSN: 1076-9986 (2002).
26. Pang, Z. *et al.* MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic acids research* **49,** W388–W396. ISSN: 0305-1048 (2021).
27. Higgins, J. P. *et al. Cochrane handbook for systematic reviews of interventions* ISBN: 1119536618 (John Wiley & Sons, 2019).
28. Menni, C. *et al.* Metabolomic identification of a novel pathway of blood pressure regulation involving hexadecanedioate. *Hypertension* **66,** 422–429. ISSN: 0194-911X (2015).
29. Menni, C. *et al.* Biomarkers for type 2 diabetes and impaired fasting glucose using a nontargeted metabolomics approach. *Diabetes* **62,** 4270–4276. ISSN: 0012-1797 (2013).
30. Menni, C. *et al.* Metabolomic profiling to dissect the role of visceral fat in cardiometabolic health. *Obesity* **24,** 1380–1388. ISSN: 1930-7381 (2016).
31. Chen, J. *et al.* Essential role of nonessential amino acid glutamine in atherosclerotic cardiovascular disease. *DNA and Cell Biology* **39,** 8–15. ISSN: 1044-5498 (2020).
32. Ruiz-Canela, M. *et al.* Plasma branched-chain amino acids and incident cardiovascular disease in the PREDIMED trial. *Clinical chemistry* **62,** 582–592. ISSN: 0009-9147 (2016).

33. Li, Y.-F. *et al.* Lysophospholipids and their G protein-coupled receptors in atherosclerosis. *Frontiers in bioscience (Landmark edition)* **21,** 70 (2016).
34. Khurana, S., Raufman, J. & Pallone, T. L. Bile acids regulate cardiovascular function. *Clinical and translational science* **4,** 210–218. ISSN: 1752-8054 (2011).
35. Liu, L., Su, J., Li, R. & Luo, F. Changes in Intestinal Flora Structure and Metabolites Are Associated With Myocardial Fibrosis in Patients With Persistent Atrial Fibrillation. *Frontiers in Nutrition* **8** (2021).
36. Cruz, D. E. *et al.* Metabolomic Analysis of Coronary Heart Disease in an African American Cohort From the Jackson Heart Study. *JAMA cardiology* **7,** 184–194. ISSN: 2380-6583 (2022).
37. Ridlon, J. M., Kang, D.-J. & Hylemon, P. B. Bile salt biotransformations by human intestinal bacteria. *Journal of lipid research* **47,** 241–259. ISSN: 0022-2275 (2006).
38. Higuchi, H. *et al.* The bile acid glycochenodeoxycholate induces trail-receptor 2/DR5 expression and apoptosis. *Journal of biological chemistry* **276,** 38610–38618. ISSN: 0021-9258 (2001).
39. Xanthopoulos, A., Starling, R. C., Kitai, T. & Triposkiadis, F. Heart failure and liver disease: cardiohepatic interactions. *JACC: Heart Failure* **7,** 87–97. ISSN: 2213-1779 (2019).
40. Kand'ár, R. & Žáková, P. Allantoin as a marker of oxidative stress in human erythrocytes. *Clinical chemistry and laboratory medicine* **46,** 1270–1274. ISSN: 1437-4331 (2008).
41. Bos, M. J., Koudstaal, P. J., Hofman, A., Witteman, J. C. & Breteler, M. M. Uric acid is a risk factor for myocardial infarction and stroke: the Rotterdam study. *stroke* **37,** 1503–1507. ISSN: 0039-2499 (2006).
42. Chlopicki, S *et al.* 1-Methylnicotinamide (MNA), a primary metabolite of nicotinamide, exerts anti-thrombotic activity mediated by a cyclooxygenase-2/prostacyclin pathway. *British journal of pharmacology* **152,** 230–239. ISSN: 0007-1188 (2007).
43. Surendran, A., Aliani, M. & Ravandi, A. Metabolomic characterization of myocardial ischemia-reperfusion injury in ST-segment elevation myocardial infarction patients undergoing percutaneous coronary intervention. *Scientific reports* **9,** 1–13. ISSN: 2045-2322 (2019).
44. Rutkowski, B. *et al.* N-methyl-2-pyridone-5-carboxamide: a novel uremic toxin? *Kidney International* **63,** S19–S21. ISSN: 0085-2538 (2003).
45. Falconi, C. A. *et al.* Uremic toxins: an alarming danger concerning the cardiovascular system. *Frontiers in Physiology* **12** (2021).
46. Zou, Y. *et al.* The regulatory roles of aminoacyl-tRNA synthetase in cardiovascular disease. *Molecular Therapy-Nucleic Acids* **25,** 372–387. ISSN: 2162-2531 (2021).
47. Amin, A. M. The metabolic signatures of cardiometabolic diseases: Does the shared metabotype offer new therapeutic targets? *Lifestyle Medicine* **2,** e25. ISSN: 2688-3740 (2021).
48. Zaric, B. L. *et al.* Atherosclerosis linked to aberrant amino acid metabolism and immunosuppressive amino acid catabolizing enzymes. *Frontiers in Immunology,* 2341. ISSN: 1664-3224 (2020).
49. Jarmusch, A. K. *et al.* Enhanced characterization of drug metabolism and the influence of the intestinal microbiome: a pharmacokinetic, microbiome, and untargeted metabolomics study. *Clinical and translational science* **13,** 972–984. ISSN: 1752-8054 (2020).
50. Han, X. *et al.* Statin in the treatment of patients with myocardial infarction: a meta-analysis. *Medicine* **97** (2018).

51. De Vera, M. A., Bhole, V., Burns, L. C. & Lacaille, D. Impact of statin adherence on cardiovascular disease and mortality outcomes: a systematic review. *British journal of clinical pharmacology* **78,** 684–698. ISSN: 0306-5251 (2014).
52. Schwartz, G. G. *et al.* Effects of atorvastatin on early recurrent ischemic events in acute coronary syndromes: the MIRACL study: a randomized controlled trial. *Jama* **285,** 1711–1718. ISSN: 0098-7484 (2001).
53. Wang, C.-Y., Liu, P.-Y. & Liao, J. K. Pleiotropic effects of statin therapy: molecular mechanisms and clinical results. *Trends in molecular medicine* **14,** 37–44. ISSN: 1471-4914 (2008).
54. Walker, R., Stewart, L. & Simmonds, M. Estimating interactions in individual participant data meta-analysis: a comparison of methods in practice. *Systematic Reviews* **11,** 211. ISSN: 2046-4053 (2022).
55. Stewart, L. A. & Tierney, J. F. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Evaluation & the health professions* **25,** 76–97. ISSN: 0163-2787 (2002).
56. Suhre, K. *et al.* Human metabolic individuality in biomedical and pharmaceutical research. *Nature* **477,** 54–60. ISSN: 1476-4687 (2011).

References from Chapter 6

1. Zheng, Y., Ley, S. H. & Hu, F. B. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nature reviews endocrinology* **14,** 88–98. ISSN: 1759-5037 (2018).
2. Sun, H. *et al.* IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes research and clinical practice* **183,** 109119. ISSN: 0168-8227 (2022).
3. Kolb, H. & Martin, S. Environmental/lifestyle factors in the pathogenesis and prevention of type 2 diabetes. *BMC medicine* **15,** 1–11. ISSN: 1741-7015 (2017).
4. Knowler, W. C. *et al.* Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin (2002).
5. Elliott, T. L. & Pfotenhauer, K. M. Classification and Diagnosis of Diabetes. *Primary Care: Clinics in Office Practice* **49,** 191–200. ISSN: 0095-4543 (2022).
6. Aydin, Nieuwdorp, M. & Gerdes, V. The gut microbiome as a target for the treatment of type 2 diabetes. *Current diabetes reports* **18,** 1–11. ISSN: 1539-0829 (2018).
7. Menni, C. *et al.* Serum metabolites reflecting gut microbiome alpha diversity predict type 2 diabetes. *Gut Microbes* **11,** 1632–1642. ISSN: 1949-0976 (2020).
8. Zhang, Z. *et al.* Characteristics of the gut microbiome in patients with prediabetes and type 2 diabetes. *PeerJ* **9,** e10952. ISSN: 2167-8359 (Print) 2167-8359 (2021).
9. Maurice, C. F. & Turnbaugh, P. J. Quantifying the metabolic activities of human-associated microbial communities across multiple ecological scales. *FEMS microbiology reviews* **37,** 830–848. ISSN: 1574-6976 (2013).
10. Krautkramer, K. A., Fan, J. & Bäckhed, F. Gut microbial metabolites as multi-kingdom intermediates. *Nat Rev Microbiol* **19,** 77–94. ISSN: 1740-1526 (2021).
11. Visconti, A. *et al.* Interplay between the human gut microbiome and host metabolism. *Nature communications* **10,** 1–10. ISSN: 2041-1723 (2019).
12. Zierer, J. *et al.* The fecal metabolome as a functional readout of the gut microbiome. *Nature genetics* **50,** 790–795. ISSN: 1546-1718 (2018).
13. Conlon, M. A. & Bird, A. R. The impact of diet and lifestyle on gut microbiota and human health. *Nutrients* **7,** 17–44. ISSN: 2072-6643 (2014).
14. Verdi, S. *et al.* TwinsUK: the UK adult twin registry update. *Twin Research and Human Genetics* **22,** 523–529. ISSN: 1832-4274 (2019).
15. ElSayed, N. A. *et al.* 2. Classification and diagnosis of diabetes: standards of care in diabetes—2023. *Diabetes care* **46,** S19–S40. ISSN: 0149-5992 (2023).
16. Donia, M. S. & Fischbach, M. A. Small molecules from the human microbiota. *Science* **349,** 1254766. ISSN: 0036-8075 (2015).
17. Pasolli, E. *et al.* Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell* **176,** 649–662. e20. ISSN: 0092-8674 (2019).

18. Thissen, D., Steinberg, L. & Kuang, D. Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *J. Educ. Behav. Stat.* **27,** 77–83. ISSN: 1076-9986 (2002).
19. Tugwell, P. & Knottnerus, J. A. A statistic to avoid being misled by the "winners curse". *Journal of Clinical Epidemiology* **103,** vi–viii. ISSN: 0895-4356 (2018).
20. Kuhn, M. Building predictive models in R using the caret package. *Journal of statistical software* **28,** 1–26. ISSN: 1548-7660 (2008).
21. Tingley, D., Yamamoto, T., Hirose, K., Keele, L. & Imai, K. Mediation: R package for causal mediation analysis. ISSN: 1548-7660 (2014).
22. Nogal, A., Valdes, A. M. & Menni, C. The role of short-chain fatty acids in the interplay between gut microbiota and diet in cardio-metabolic health. *Gut Microbes* **13,** 1–24. ISSN: 1949-0976 (2021).
23. Jiang, J., Li, B., He, W. & Huang, C. Dietary serine supplementation: Friend or foe? *Current Opinion in Pharmacology* **61,** 12–20. ISSN: 1471-4892 (2021).
24. Fujiwara, R., Yoda, E. & Tukey, R. H. Species differences in drug glucuronidation: Humanized UDP-glucuronosyltransferase 1 mice and their application for predicting drug glucuronidation and drug-induced toxicity in humans. *Drug metabolism and pharmacokinetics* **33,** 9–16. ISSN: 1347-4367 (2018).
25. Sachar, M., Anderson, K. E. & Ma, X. Protoporphyrin IX: the good, the bad, and the ugly. *Journal of Pharmacology and Experimental Therapeutics* **356,** 267–275. ISSN: 0022-3565 (2016).
26. Moffett, J. R. & Namboodiri, M. A. Tryptophan and the immune response. *Immunology and cell biology* **81,** 247–265. ISSN: 0818-9641 (2003).
27. Kriaa, A. *et al.* Microbial impact on cholesterol and bile acid metabolism: current status and future prospects. *Journal of lipid research* **60,** 323–332. ISSN: 0022-2275 (2019).
28. Urasaki, Y., Pizzorno, G. & Le, T. T. Uridine affects liver protein glycosylation, insulin signaling, and heme biosynthesis. *PloS one* **9,** e99728. ISSN: 1932-6203 (2014).
29. Keijzers, G. B., De Galan, B. E., Tack, C. J. & Smits, P. Caffeine can decrease insulin sensitivity in humans. *Diabetes care* **25,** 364–369. ISSN: 0149-5992 (2002).
30. De Vos, W. M., Tilg, H., Van Hul, M. & Cani, P. D. Gut microbiome and health: mechanistic insights. *Gut* **71,** 1020–1032. ISSN: 0017-5749 (2022).
31. Basolo, A. *et al.* Effects of underfeeding and oral vancomycin on gut microbiome and nutrient absorption in humans. *Nature Medicine* **26,** 589–598. ISSN: 1078-8956 (2020).
32. Raimondi, S, Musmeci, E, Candeliere, F, Amaretti, A & Rossi, M. *Identification of mucin degraders of the human gut microbiota. Sci Rep 11: 11094* Generic. 2021.
33. Anhê, F. F., Barra, N. G., Cavallari, J. F., Henriksbo, B. D. & Schertzer, J. D. Metabolic endotoxemia is dictated by the type of lipopolysaccharide. *Cell Reports* **36,** 109691. ISSN: 2211-1247 (2021).
34. Depommier, C. *et al.* Pasteurized Akkermansia muciniphila increases whole-body energy expenditure and fecal energy excretion in diet-induced obese mice. *Gut Microbes* **11,** 1231–1245. ISSN: 1949-0976 (2020).
35. Vacca, M. *et al.* The controversial role of human gut lachnospiraceae. *Microorganisms* **8,** 573. ISSN: 2076-2607 (2020).
36. Kaczmarczyk, M. *et al.* The gut microbiota is associated with the small intestinal paracellular permeability and the development of the immune system in healthy children during the first two years of life. *Journal of Translational Medicine* **19,** 1–26. ISSN: 1479-5876 (2021).

37. Asano, T, Yuasa, K, Kunugita, K, Teraji, T & Mitsuoka, T. Effects of gluconic acid on human faecal bacteria. *Microbial ecology in health and disease* **7,** 247–256. ISSN: 1651-2235 (1994).

38. Unwin, N., Shaw, J., Zimmet, P. & Alberti, K. Impaired glucose tolerance and impaired fasting glycaemia: the current status on definition and intervention. *Diabetic medicine: a journal of the British Diabetic Association* **19,** 708–723. ISSN: 0742-3071 (2002).

References from Chapter 7

1. Cummings, J. H., Pomare, E. W., Branch, W. J., Naylor, C. P. & Macfarlane, G. T. Short chain fatty acids in human large intestine, portal, hepatic and venous blood. *Gut* **28,** 1221–1227. ISSN: 0017-5749 (1987).
2. Topping, D. L. & Clifton, P. M. Short-chain fatty acids and human colonic function: roles of resistant starch and nonstarch polysaccharides. *Physiol. Rev.* **81,** 1031–1064. ISSN: 0031-9333 (2001).
3. Nogal, A., Valdes, A. M. & Menni, C. The role of short-chain fatty acids in the interplay between gut microbiota and diet in cardio-metabolic health. *Gut Microbes* **13,** 1–24. ISSN: 1949-0976 (2021).
4. Sanna, S. *et al.* Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. *Nature genetics* **51,** 600–605. ISSN: 1061-4036 (2019).
5. Vitale, M. *et al.* Acute and chronic improvement in postprandial glucose metabolism by a diet resembling the traditional Mediterranean dietary pattern: Can SCFAs play a role? *Clinical Nutrition* **40,** 428–437. ISSN: 0261-5614 (2021).
6. Yao, Y. *et al.* The role of short-chain fatty acids in immunity, inflammation and metabolism. *Crit. Rev. Food Sci. Nutr.* **62,** 1–12. ISSN: 1040-8398 (2022).
7. Nakahori, Y. *et al.* Impact of fecal short-chain fatty acids on prognosis in critically ill patients. *Acute Med Surg* **7,** e558. ISSN: 2052-8817 (2020).
8. Valdés-Duque, B. E. *et al.* Stool Short-Chain Fatty Acids in Critically Ill Patients with Sepsis. *J. Am. Coll. Nutr.* **39,** 706–712. ISSN: 0731-5724 (2020).
9. Nogal, A. *et al.* Circulating Levels of the Short-Chain Fatty Acid Acetate Mediate the Effect of the Gut Microbiome on Visceral Fat. *Front. Microbiol.* **12,** 711359. ISSN: 1664-302X (2021).
10. Meyer, R. K. *et al.* Oligofructose restores postprandial short-chain fatty acid levels during high-fat feeding. *Obesity* **30,** 1442–1452. ISSN: 1930-7381 (2022).
11. Kirschner, S. K., Ten Have, G. A., Engelen, M. P. & Deutz, N. E. Transorgan short-chain fatty acid fluxes in the fasted and postprandial state in the pig. *American Journal of Physiology-Endocrinology and Metabolism* **321,** E665–E673. ISSN: 0193-1849 (2021).
12. Mazidi, M. *et al.* Meal-induced inflammation: postprandial insights from the Personalised REsponses to DIetary Composition Trial (PREDICT) study in 1000 participants. *Am J Clin Nutr* **114,** 1028–1038. ISSN: 0002-9165 (Print) 0002-9165 (2021).
13. Wang, Z. C. *et al.* Systemic immune-inflammation index independently predicts poor survival of older adults with hip fracture: a prospective cohort study. *BMC Geriatr* **21,** 155. ISSN: 1471-2318 (2021).
14. Falony, G. *et al.* Population-level analysis of gut microbiome variation. *Science* **352,** 560–564. ISSN: 0036-8075 (2016).

15. Akram, M. Citric Acid Cycle and Role of its Intermediates in Metabolism. *Cell Biochemistry and Biophysics* **68,** 475–478 (2014).
16. Berry, S. E. *et al.* Human postprandial responses to food and potential for precision nutrition. *Nat. Med.* **26,** 964–973. ISSN: 1078-8956 (2020).
17. Saresella, M. *et al.* Alterations in Circulating Fatty Acid Are Associated With Gut Microbiota Dysbiosis and Inflammation in Multiple Sclerosis. *Front. Immunol.* **11,** 1390. ISSN: 1664-3224 (2020).
18. Deng, K. *et al.* Comparison of fecal and blood metabolome reveals inconsistent associations of the gut microbiota with cardiometabolic diseases. *Nature Communications* **14,** 571. ISSN: 2041-1723 (2023).
19. Louis, P., Young, P., Holtrop, G. & Flint, H. J. Diversity of human colonic butyrate-producing bacteria revealed by analysis of the butyryl-CoA:acetate CoA-transferase gene. *Environ. Microbiol.* **12,** 304–314. ISSN: 1462-2912 (2010).
20. Duncan, S. H., Hold, G. L., Barcenilla, A., Stewart, C. S. & Flint, H. J. Roseburia intestinalis sp. nov., a novel saccharolytic, butyrate-producing bacterium from human faeces. *Int. J. Syst. Evol. Microbiol.* **52,** 1615–1620. ISSN: 1466-5026 (2002).
21. Parker, B. J., Wearsch, P. A., Veloo, A. C. M. & Rodriguez-Palacios, A. The Genus : Gut Bacteria With Emerging Implications to Inflammation, Cancer, and Mental Health. *Front. Immunol.* **11,** 906. ISSN: 1664-3224 (2020).
22. Asnicar, F. *et al.* Microbiome connections with host metabolism and habitual diet from 1,098 deeply phenotyped individuals. *Nat. Med.* **27,** 321–332. ISSN: 1078-8956 (2021).
23. Bergman, E. N. Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiol. Rev.* **70,** 567–590. ISSN: 0031-9333 (1990).
24. Rasmussen, H. S., Holtug, K. & Mortensen, P. B. Degradation of Amino Acids to Short-Chain Fatty Acids in Humans: An in Vitro Study. *Scandinavian Journal of Gastroenterology* **23,** 178–182 (1988).
25. Campos-Perez, W. & Martinez-Lopez, E. Effects of short chain fatty acids on metabolic and inflammatory processes in human health. *Biochim Biophys Acta Mol Cell Biol Lipids* **1866,** 158900. ISSN: 1388-1981 (2021).
26. Porter, C. J. *et al.* Acute and chronic kidney disease in elderly patients with hip fracture: prevalence, risk factors and outcome with development and validation of a risk prediction model for acute kidney injury. *BMC nephrology* **18,** 1–11 (2017).
27. Bankhead-Kendall, B. *et al.* Rib Fractures and Mortality: Breaking the Causal Relationship. *Am. Surg.* **85,** 1224–1227. ISSN: 0003-1348 (2019).
28. Foss, N. B. & Kehlet, H. Mortality analysis in hip fracture patients: implications for design of future outcome trials. *Br. J. Anaesth.* **94,** 24–29. ISSN: 0007-0912 (2005).
29. Bhandari, M. & Swiontkowski, M. Management of Acute Hip Fracture. *N. Engl. J. Med.* **377,** 2053–2062. ISSN: 0028-4793 (2017).
30. Moayyeri, A., Hammond, C. J., Valdes, A. M. & Spector, T. D. Cohort Profile: TwinsUK and Healthy Ageing Twin Study. *International Journal of Epidemiology* **42,** 76–85 (2013).
31. Berry, S. *et al.* Personalised REsponses to DIetary Composition Trial (PREDICT): an intervention study to determine inter-individual differences in postprandial response to foods.
32. Evans, A. M., DeHaven, C. D., Barrett, T., Mitchell, M. & Milgram, E. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Anal. Chem.* **81,** 6656–6667. ISSN: 0003-2700 (2009).

33. Visconti, A. *et al.* Interplay between the human gut microbiome and host metabolism. *Nat. Commun.* **10,** 4505. ISSN: 2041-1723 (2019).
34. Visconti, A., Martin, T. C. & Falchi, M. YAMP: a containerized workflow enabling reproducibility in metagenomics research. *Gigascience* **7.** ISSN: 2047-217X (2018).
35. Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J. & Segata, N. Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology* **35,** 833–844 (2017).
36. McIver, L. J. *et al.* bioBakery: a meta'omic analysis environment. *Bioinformatics* **34,** 1235–1237 (2018).
37. Beghini, F. *et al.* Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3 (2021).
38. Blanco-Miguez, A. *et al.* Extending and improving metagenomic taxonomic profiling with uncharacterized species with MetaPhlAn 4. *bioRxiv* (2022).
39. Thissen, D., Steinberg, L. & Kuang, D. Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *J. Educ. Behav. Stat.* **27,** 77–83. ISSN: 1076-9986 (2002).
40. Neale, M. & Cardon, L. R. *Methodology for Genetic Studies of Twins and Families* 496. ISBN: 9789401580182 (Springer Science & Business Media, 2013).
41. Scheike, T. H., Holst, K. K. & Hjelmborg, J. B. Estimating heritability for cause specific mortality based on twin studies. *Lifetime data analysis* **20,** 210–233. ISSN: 1380-7870 (2014).
42. Nelli, F. Machine Learning with scikit-learn. *Python Data Analytics,* 237–264 (2015).
43. Pasolli, E., Truong, D. T., Malik, F., Waldron, L. & Segata, N. Machine Learning Meta-analysis of Large Metagenomic Datasets: Tools and Biological Insights. *PLoS Comput. Biol.* **12,** e1004977. ISSN: 1553-734X (2016).

References from Chapter 8

1. Alaei-Shahmiri, F, Soares, M., Zhao, Y & Sherriff, J. The impact of thiamine supplementation on blood pressure, serum lipids and C-reactive protein in individuals with hyperglycemia: a randomised, double-blind cross-over trial. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* **9,** 213–217. ISSN: 1871-4021 (2015).
2. Almeida, A. *et al.* A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nature Biotechnology,* 1–10. ISSN: 1546-1696 (2020).
3. Barrios, C. *et al.* Circulating metabolic biomarkers of renal function in diabetic and non-diabetic populations. *Sci Rep* **8,** 15249. ISSN: 2045-2322 (2018).
4. Basen, M. & Kurrer, S. E. A close look at pentose metabolism of gut bacteria. *The FEBS Journal.* ISSN: 1742-464X (2020).
5. Bingham, S. A. *et al.* Nutritional methods in the European prospective investigation of cancer in Norfolk. *Public health nutrition* **4,** 847–858. ISSN: 1475-2727 (2001).
6. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature methods* **7,** 335–336. ISSN: 1548-7105 (2010).
7. Caspi, R. *et al.* The MetaCyc database of metabolic pathways and enzymes. *Nucleic acids research* **46,** D633–D639. ISSN: 0305-1048 (2018).
8. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal, complex systems* **1695,** 1–9 (2006).
9. Cummings, J., Pomare, E., Branch, W., Naylor, C. & Macfarlane, G. Short chain fatty acids in human large intestine, portal, hepatic and venous blood. *Gut* **28,** 1221–1227. ISSN: 0017-5749 (1987).
10. Dambrova, M *et al.* Diabetes is associated with higher trimethylamine N-oxide plasma levels. *Experimental and clinical endocrinology & diabetes* **124,** 251–256. ISSN: 0947-7349 (2016).
11. Den Besten, G. *et al.* The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *Journal of lipid research* **54,** 2325–2340. ISSN: 0022-2275 (2013).
12. Fernandez-Mejia, C. Pharmacological effects of biotin. *The Journal of nutritional biochemistry* **16,** 424–427. ISSN: 0955-2863 (2005).
13. Frost, G. *et al.* The short-chain fatty acid acetate reduces appetite via a central homeostatic mechanism. *Nature communications* **5,** 1–11. ISSN: 2041-1723 (2014).
14. Fuller, M. F. & Tomé, D. In vivo determination of amino acid bioavailability in humans and model animals. *Journal of AOAC International* **88,** 923–934. ISSN: 1060-3271 (2005).
15. Goodrich, J. K. *et al.* Genetic determinants of the gut microbiome in UK twins. *Cell host & microbe* **19,** 731–743. ISSN: 1931-3128 (2016).

16. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32,** 2847–2849. ISSN: 1460-2059 (2016).
17. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29,** 1072–1075. ISSN: 1460-2059 (2013).
18. Hron, W., Menahan, L. & Lech, J. Inhibition of hormonal stimulation of lipolysis in perfused rat heart by ketone bodies. *Journal of molecular and cellular cardiology* **10,** 161–174. ISSN: 0022-2828 (1978).
19. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature communications* **9,** 1–8. ISSN: 2041-1723 (2018).
20. Jameson, E. *et al.* Metagenomic data-mining reveals contrasting microbial populations responsible for trimethylamine formation in human gut and marine ecosystems. *Microbial genomics* **2** (2016).
21. Kanehisa, M. *et al.* Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic acids research* **42,** D199–D205. ISSN: 1362-4962 (2014).
22. Kaul, S. *et al.* Dual-energy X-ray absorptiometry for quantification of visceral fat. *Obesity* **20,** 1313–1318. ISSN: 1930-7381 (2012).
23. Lahti, L. & Sudarshan, S. *microbiome R package* Computer Program. 2017. http://microbiome.github.io/microbiome.
24. Li, S. *et al.* Effect of the sulfation pattern of sea cucumber-derived fucoidan oligosaccharides on modulating metabolic syndromes and gut microbiota dysbiosis caused by HFD in mice. *Journal of Functional Foods* **55,** 193–210. ISSN: 1756-4646 (2019).
25. Li, Z. & Vance, D. E. Thematic review series: glycerolipids. Phosphatidylcholine and choline homeostasis. *Journal of lipid research* **49,** 1187–1194. ISSN: 0022-2275 (2008).
26. Lustgarten, M. S. The role of the gut microbiome on skeletal muscle mass and physical function: 2019 update. *Frontiers in physiology* **10,** 1435. ISSN: 1664-042X (2019).
27. Martin-Gallausiaux, C., Marinelli, L., Blottière, H. M., Larraufie, P. & Lapaque, N. SCFA: mechanisms and functional importance in the gut. *Proceedings of the Nutrition Society,* 1–13. ISSN: 0029-6651 (2020).
28. McCance, R. A. & Widdowson, E. M. *McCance and Widdowson's the Composition of Foods* ISBN: 1849736367 (Royal Society of Chemistry, 2014).
29. Menni, C. *et al.* Metabolomic profiling to dissect the role of visceral fat in cardiometabolic health. *Obesity (Silver Spring)* **24,** 1380–8. ISSN: 1930-7381 (Print) 1930-7381 (2016).
30. Miller, T. L. The pathway of formation of acetate and succinate from pyruvate by Bacteroides succinogenes. *Archives of microbiology* **117,** 145–152. ISSN: 1432-072X (1978).
31. Moayyeri, A., Hammond, C. J., Valdes, A. M. & Spector, T. D. Cohort Profile: TwinsUK and healthy ageing twin study. *International journal of epidemiology* **42,** 76–85. ISSN: 1464-3685 (2013).
32. Mottaghian, M. *et al.* Nutrient patterns and cardiometabolic risk factors among Iranian adults: Tehran lipid and glucose study. *BMC public health* **20,** 1–12 (2020).
33. Navarrete-Muñoz, E.-M. *et al.* Dietary folate intake and metabolic syndrome in participants of PREDIMED-Plus study: a cross-sectional study. *European journal of nutrition,* 1–12. ISSN: 1436-6215 (2020).

34. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research* **25,** 1043–1055. ISSN: 1088-9051 (2015).
35. Pataky, Z. *et al.* Impact of hypocaloric hyperproteic diet on gut microbiota in overweight or obese patients with nonalcoholic fatty liver disease: a pilot study. *Digestive diseases and sciences* **61,** 2721–2731. ISSN: 1573-2568 (2016).
36. Petersen, M. C. & Shulman, G. I. Mechanisms of insulin action and insulin resistance. *Physiological reviews* **98,** 2133–2223. ISSN: 0031-9333 (2018).
37. Pryde, S. E., Duncan, S. H., Hold, G. L., Stewart, C. S. & Flint, H. J. The microbiology of butyrate formation in the human colon. *FEMS microbiology letters* **217,** 133–139. ISSN: 1574-6968 (2002).
38. Reichardt, N. *et al.* Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. *The ISME journal* **8,** 1323–1335. ISSN: 1751-7370 (2014).
39. Rey, F. E. *et al.* Dissecting the in vivo metabolic potential of two human gut acetogens. *Journal of biological chemistry* **285,** 22082–22090. ISSN: 0021-9258 (2010).
40. Robert, C., Chassard, C., Lawson, P. A. & Bernalier-Donadille, A. Bacteroides cellulosilyticus sp. nov., a cellulolytic bacterium from the human gut microbial community. *International journal of systematic and evolutionary microbiology* **57,** 1516–1520. ISSN: 1466-5026 (2007).
41. Sakakibara, S., Yamauchi, T., Oshima, Y., Tsukamoto, Y. & Kadowaki, T. Acetic acid activates hepatic AMPK and reduces hyperglycemia in diabetic KK-A (y) mice. *Biochemical and biophysical research communications* **344,** 597–604. ISSN: 0006-291X (2006).
42. Sayers, E. W. *et al.* Database resources of the national center for biotechnology information. *Nucleic acids research* **47,** D23 (2019).
43. Schugar, R. C. *et al.* The TMAO-producing enzyme flavin-containing monooxygenase 3 regulates obesity and the beiging of white adipose tissue. *Cell reports* **19,** 2451–2461. ISSN: 2211-1247 (2017).
44. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30,** 2068–2069. ISSN: 1460-2059 (2014).
45. Sun, X. *et al.* Modulation of gut microbiota by fucoxanthin during alleviation of obesity in high-fat diet-fed mice. *Journal of agricultural and food chemistry* **68,** 5118–5128. ISSN: 0021-8561 (2020).
46. Team, R. C. & DC, R. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2012. *URL https://www. R-project. org* (2019).
47. Thissen, D., Steinberg, L. & Kuang, D. Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of educational and behavioral statistics* **27,** 77–83. ISSN: 1076-9986 (2002).
48. Tingley, D., Yamamoto, T., Hirose, K., Keele, L. & Imai, K. Mediation: R package for causal mediation analysis. ISSN: 1548-7660 (2014).
49. Topping, D. L. & Clifton, P. M. Short-chain fatty acids and human colonic function: roles of resistant starch and nonstarch polysaccharides. *Physiological reviews* (2001).
50. Wang, Y., Ye, X., Ding, D. & Lu, Y. Characteristics of the intestinal flora in patients with peripheral neuropathy associated with type 2 diabetes. *Journal of International Medical Research* **48,** 0300060520936806. ISSN: 0300-0605 (2020).

51. Wells, P. M. *et al.* Associations between gut microbiota and genetic risk for rheumatoid arthritis in the absence of disease: a cross-sectional study. *The Lancet Rheumatology* **2,** e418–27 (2020).
52. Willett, W. & Stampfer, M. J. Total energy intake: implications for epidemiologic analyses. *American journal of epidemiology* **124,** 17–27. ISSN: 0002-9262 (1986).
53. Wu, Y.-W. ezTree: an automated pipeline for identifying phylogenetic marker genes and inferring evolutionary relationships among uncultivated prokaryotic draft genomes. *BMC genomics* **19,** 921. ISSN: 1471-2164 (2018).
54. Würtz, P. *et al.* Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation* **131,** 774–785. ISSN: 0009-7322 (2015).
55. Ye, Y. & Doak, T. G. A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. *PLoS Comput Biol* **5,** e1000465. ISSN: 1553-7358 (2009).
56. Yutin, N. & Galperin, M. Y. A genomic update on clostridial phylogeny: G ram-negative spore formers and other misplaced clostridia. *Environmental microbiology* **15,** 2631–2641. ISSN: 1462-2912 (2013).
57. Zhao, L. *et al.* A combination of quercetin and resveratrol reduces obesity in high-fat diet-fed rats by modulation of gut microbiota. *Food & function* **8,** 4644–4656 (2017).
58. Zhao, Y. *et al.* Structure-specific effects of short-chain fatty acids on plasma cholesterol concentration in male syrian hamsters. *Journal of agricultural and food chemistry* **65,** 10984–10992. ISSN: 0021-8561 (2017).
59. Zhu, W *et al.* Flavin monooxygenase 3, the host hepatic enzyme in the metaorganismal trimethylamine N-oxide-generating pathway, modulates platelet responsiveness and thrombosis risk. *Journal of Thrombosis and Haemostasis* **16,** 1857–1872. ISSN: 1538-7933 (2018).

Appendix A

OneDrive files

Chapter 5

- **Supplementary Table 5.1 Metabolites significantly associated (meta-analysis FDR<0.05) with incident MI.** TE and SE refer to estimated overall treatment effect and standard error, respectively.
- **Supplementary Table 5.2 Literature references for the metabolites previously associated with any cardiac diseases, and the super- and sub-pathways for metabolites associated with incident MI.** For the metabolites that did not remain significant after further adjusting the meta-analyses for prevalent hypertension, dyslipidaemia and type-2 diabetes, references showing their associations with any of these 3 conditions are indicated.
- **Supplementary Table 5.4 Meta-analysis results from the 56 metabolites significantly associated with incident MI when the analyses were run excluding the cohorts in which MI was assessed by self-reported questionnaires (TwinsUK and ET2DS**). TE and SE refer to estimated overall treatment effect and standard error, respectively.
- **Supplementary Table 5.5 Meta-analysis results from the 56 metabolites significantly associated with incident MI when the models were further**

adjusted for prevalent hypertension, prevalent type-2 diabetes and prevalent dyslipidemia. Significant associations are marked in red. TE and SE refer to estimated overall treatment effect and standard error, respectively.

- **Supplementary Table 5.6 Meta-analysis results from the 56 metabolites significantly associated with incident MI when the models were stratified by race (White individuals and Black individuals).** Significant associations are marked in red. TE and SE refer to estimated overall treatment effect and standard error, respectively.
- **Supplementary Table 5.7 Metabolites associated (meta-analysis nominal p-value<0.05) with prevalent MI, and that are also significantly associated with incident MI (meta-analysis FDR<0.05)**. TE and SE refer to estimated overall treatment effect and standard error, respectively.

Chapter 6

- **Supplementary Table 6.1 Complete list of the 526 included metabolites in TwinsUK measured by Metabolon Inc. with their super-pathways, sub-pathways, and KEGG and HMDB identifiers.** From these, the metabolites with measurements available for KORA participants are indicated.
- **Supplementary Table 6.4 Associations between the gut microbiota composition and impaired fasting glucose (IFG).** Specifically, the top 100 features from the Random Forest models predicting the faecal metabolite abundances from the gut microbiome composition with an AUC>70% are shown. The linear regression models were adjusted for age, BMI, sex and multiple testing (false discovery rate – FDR). The prevalence of each gut bacteria is also indicated.
- **Supplementary Table 6.5 Associations of comorbidities with the 8 metabolites making up the score and the bacterial species involved in the mediation analyses.** Pearson's correlations run for the continuous comorbidities (systolic and diastolic blood pressure, circulating HDL, total cholesterol and triglycerides levels, and aHEI)

whereas a two-proportion z-test was used for the categorical comorbidities (activity level and smoking status).

- **Supplementary Table 6.6 List of gut species represented using species-level genome bins (SGBs) that were profiled in 342 participants from TwinsUK.** Prevalence and if the composition of a species presents variance zero and/or near zero are indicated.

Chapter 8

- **Supplementary Table 8.1 The created dataset containing 1,121 ultra-high-quality genomes belonging to *Coprococcus* and 271 high-quality genomes belonging to *Lachnoclostridium* and their respective metadata.**
- **Supplementary Tables 8.2, 8.3, 8.4 Average nucleotide identity (ANI) values obtained for each pair of genomes belonging to *Lachnoclostridium* and *Coprococcus* using FastANI.** Previously, genomes were filtered by alignment fraction (>0.4).
- **Supplementary Table 8.5 KEGG and MetaCyc pathways obtained for *Coprococcus* and *Lachnoclostridium* genomes along with the percentage of genomes of each species presenting a given pathway.**

Appendix B

Extended discussion

In the following sections, some important points/limitations from **Chapters 5** ("Circulating biomarkers of incident myocardial infarction") and **6** ("A faecal metabolite signature of prediabetes)", which are not included in their respective published manuscripts, are discussed in greater detail.

Chapter 5

Unbalanced case-control ratio

In the Discussion section, the imbalanced ratio of cases to controls (the control number is 5.7-fold larger than the incident MI case number), is acknowledged as a limitation. An unbalanced distribution of the response might increase the variance, resulting in wider confidence intervals and reducing the statistical power. Nevertheless, the conducted meta-analysis identified 56 metabolites with significantly altered levels between incident MI cases and controls, underscoring notable findings despite the challenges presented by the nature of our data.

Potential confounding variables

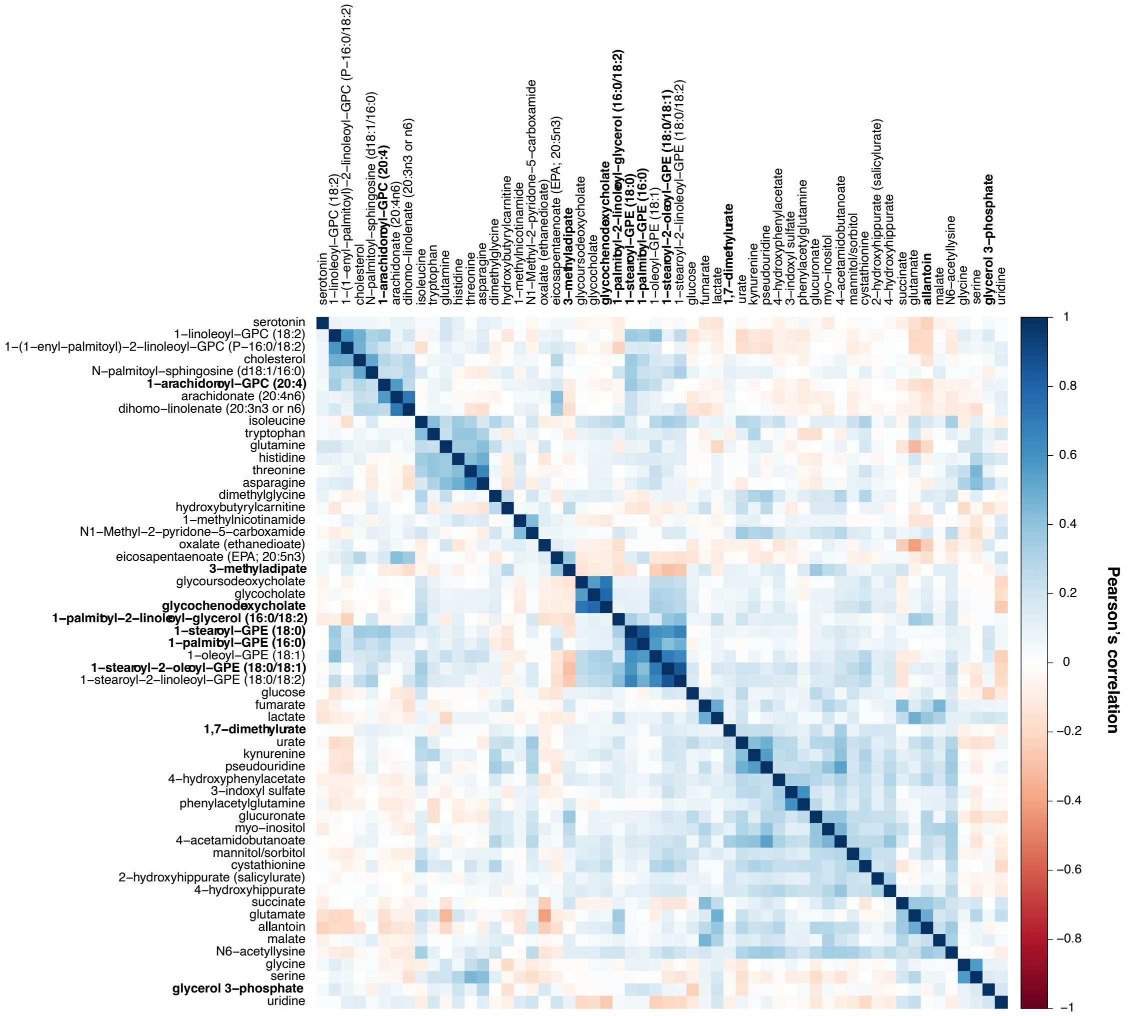
Diet and smoking intensity might be important confounders in the reported metabolite-incident MI associations. Unfortunately, this data could not be retrieved for the included cohorts. However, the models were adjusted for multiple potential confounders including age, sex, race, BMI, education level, physical activity levels, alcohol consumption and smoking status. Likewise, sensitivity analyses were also performed to confirm the reported associations after further adjusting the models for prevalent T2D, hypertension and dyslipidaemia – conditions closely related to diet and potentially reflective of the dietary patterns [321–323].

Imputation approach

Missing physical activity was replaced by the medium category of physical activity level (0=low, 1=medium, 2=high). The variable design used across cohorts was decided following the COnsortium of METabolomics Studies (COMETS)' advice. This imputation approach was selected since physical activity was not the main outcome, but one of the eight potential confounders used in the models. Moreover, only a very small proportion of individuals had this variable imputed (<10%). However, I acknowledge that this imputation approach might have introduced a certain degree of bias to the results.

Independence of the obtained results

To get insights on the independence of the findings, especially for the novel metabolites identified, a pairwise correlation analysis for the identified incident MI-associated metabolites was performed in TwinsUK (**Appendix B - Figure 1**). The results suggest that the findings are independent as most metabolites present low correlations with each other. Specifically, 89% of the pairwise correlations presented *rho* values between 0.25 and -0.25, and 88% of the pairwise correlations with the 10 novel identified metabolites presented *rho* values between 0.25 and -0.25. Ideally, these results would benefit from conditional analyses. Conditional analyses would enable the assessment of each metabolite's effect after adjusting for the others, thereby clarifying the unique contribution of each metabolite to the risk of incident MI. This is particularly crucial for the novel metabolites, as it would provide stronger evidence for their potential role as independent biomarkers or causal factors in the development of MI. Unfortunately, the execution of conditional analyses would require the independent implementation of these analyses by each contributing cohort, which was not possible primarily due to the logistical challenges involved.



Appendix B - Fig.1 Pearson's correlation matrix calculated from the abundances in TwinsUK (n=911) of the 56 incident MI-associated metabolites in 6 cohorts from the COnsortiun of METabolomics Studies (COMETS). The novel identified metabolites are indicated in bold.

Chapter 6

Methodological approach for building the IFG-metabolite score

To construct the IFG-metabolite score, univariate analyses in TwinsUK (discovery set) and KORA (replication set) were conducted separately. Faecal metabolites that were significant and showing the same directional association in both cohorts were selected and linearly combined. Although other methods such as elastic net regression and lasso could have been applied to derive the score, their use was considered but not implemented. These approaches might identify a set of metabolites able to accurately distinguish IFG cases from healthy individuals in the discovery set. However, such a metabolite set may not generalise well across different datasets or populations. To develop an IFG-metabolite score potentially more representative of diverse populations with varying demographic characteristics, such as those in KORA, the score was based on metabolites replicated in the KORA cohort.

Potential selection bias in the subset with gut microbiome profiling

As the gut microbiome was available only from a subset of individuals from the original dataset, there might have been any potential selection bias. To investigate this, the baseline characteristics of the individuals with the microbiome profiled were compared with those of the individuals who did not have the microbiome profiled. As it is observed in **Appendix B – Table 1**, there were no significant differences in the demographic characteristics between these two groups of subjects, thus mitigating the concern of potential selection bias.

Appendix B - Table 1 Descriptive characteristics of the individuals from TwinsUK with and without concurrent gut microbiota composition and faecal metabolites measurements. The p-value from a Wilcoxon test (continuous variable) or chi-squared test (categorical variable) was calculated to check whether differences between the different subject groups for the described parameters existed.

	Individuals with gut microbiome	Individuals without gut microbiome	Differences between groups (p-value)
N	342	905	-
Females, %	83.9	89.2	0.02
Age, yrs	56 (16.6)	58.5 (14)	0.23
BMI, kg/m2	25.6 (5)	25.5 (4.7)	0.82
Circulating total cholesterol, mmol/L	4.1 (0.5)	4.1 (0.5)	0.78
Fasting glucose, mmol/L	4.7 (0.5)	4.7 (0.5)	0.85
Alternate health eating index	70.6 (5.7)	70.3 (6.7)	0.64
Current smoker	No: 331 Yes: 11	No: 868 Yes: 37	0.58
Activity level	Low: 25 Moderate: 259 High: 58	Low: 88 Moderate: 645 High: 172	0.24

Potential permeability markers for finding validation

In **Chapter 6**, the potential utility of measuring permeability markers to elucidate the role of intestinal permeability in the absorption or excretion of the eight identified metabolites is discussed. A variety of permeability markers could be used in conjunction with the existing data presented in this study. For instance, an ELISA assay could be used to detect and quantify biomarkers related to intestinal permeability, including faecal or circulating zonulin (a protein that modulates the permeability of tight junctions between epithelial cells), faecal alpha-1 antitrypsin (a protease inhibitor that reflects the protein loss into the intestinal lumen) and circulating LPS (which under normal conditions are prevented from entering the bloodstream by the gut barrier) [324]. Elevated levels of these biomarkers would suggest a potential disruption of the gut barrier integrity [324].