Gut microbiota-host metabolic interactions in human development

Kiana A West

Thesis presented for the degree of
Doctor of Philosophy of Imperial College London
and the Diploma of Imperial College London

Supervised by
Professor Elaine Holmes
Dr Joram M Posma

Department of Metabolism, Digestion and Reproduction
Faculty of Medicine
Imperial College London

May 2020

Statement of originality

The content of this thesis, and the research to which it refers, are the work of the Author unless otherwise noted. The work and ideas of others are fully acknowledged in accordance with the standard referencing practices in the discipline.

Copyright declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abstract

The microbial communities living in the gut exist in symbiosis with the host and perform biological functions that are fundamental to human physiology. Evidence from animal studies suggests that disturbing host-microbe interactions during developmental phases has lasting consequences, but evidence in humans is lacking. Aberrant metabolism was investigated during two periods of development - pregnancy and early childhood - to define functional associations between the gut microbiota and aspects of altered development.

Maternal obesity is associated with fetal overgrowth, but the relationship between obesity, the maternal gut microbiota and neonatal outcomes has not been thoroughly investigated. This thesis details longitudinal shifts in gestational metabolism with maternal body mass index, which were linked to elevated insulin resistance during pregnancy, elevated neonatal insulin levels and high birth weight. A direct association between the maternal gut microbiota and neonatal outcomes was not observed. However, maternal insulin resistance, which is thought to contribute to fetal overgrowth, was positively correlated with bacteria-derived metabolites. Obese women with a previous weight loss surgery had reduced insulin resistance during pregnancy, which was associated with changes in gut microbial composition and host-microbe co-metabolism. The identified metabolites were linked to lower birth weight, demonstrating the metabolism of the maternal gut microbiota may influence fetal growth either directly, or through modulation of maternal insulin resistance. In addition to prospective associations, retrospective microbiota-associated signatures of altered neurodevelopment were sought in a childhood cohort, but could not be identified. However, subgroups of children with Autism Spectrum Disorder were distinguished based on patterns of host-microbe co-metabolism of dietary protein, which may be associated with gastrointestinal dysfunction.

Collectively, this thesis deciphers several associations between microbial and metabolic profiles, demonstrating the systemic capacity of gut microbes to influence human metabolism. The results presented herein provide a foundation for investigating mechanistic links between the metabolic activity of the gut microbiota during pregnancy and fetal development, but prospective birth cohorts are needed to study the relationship between gut microbial communities and altered neurodevelopment.

Acknowledgements

First and foremost, a great deal of thanks is due to my supervisors, Prof Elaine Holmes and Dr Joram Posma. I am grateful for their continuous, unfaltering support throughout this journey. I feel fortunate to have had such personable supervisors who have also helped me grow as a scientist. I would also like to thank Prof Jeremy Nicholson for his mentorship and enthusiasm in Autism research. Thanks is also due to Prof Julian Marchesi, who offered friendly support and guidance throughout my PhD. I would like to express my gratitude to the STRATiGRAD programme for its financial support and plentiful networking opportunities, and to the Kristian Gerhard Jebsen Foundation for their generous grant which enabled us to study Autism Spectrum Disorder. I also would not be where I am today without the support of Dr Anthony Williamson. Thank you for everything you have done for me, Anthony.

On a personal level, I am grateful for many people who have supported me through the past four years, and especially the past few months. Michael, I know this has been such a roller coaster for both of us, but I could not have done it without your love, support and (above all) patience. I can't thank you enough for your consistent understanding and encouragement. To the Pauls and Boltons: Thank you for welcoming me into your family with open arms. You have made being away from my own family so much easier. To my family: I know you are eagerly waiting for us to arrive on a more permanent visit. Your love and support from "across the pond" has made the years away from home bearable, but I can't wait to be back and eat tacos with you all. To my CSM (MDR) friends from the past and present: You have made SAF such a fun working environment. I feel so lucky to have met each one of you and will miss you dearly. And to my non-Imperial friends (in the US and the UK): You have all been wonderfully supportive through this and I hope you know how much I value our friendship. The Zoom/FaceTime dates have kept me sane the past couple months and I can't wait to social un-distance with you.

I want to give a special shout-out to the friends who offered invaluable advice and suggestions during the writing period, and to those who have been persistently cheering me on (especially Allison, Gonçalo, Gordana, Marine, Martha, Nan, Natasa, Olof, Simone and Solly). Finally, I want to dedicate this thesis to Grandad, who would have been so proud of this achievement.

Contributions of others

Dr Makrina Savvidou led the pregnancy study at Chelsea & Westminster Hospital (**Chapter 3** and **Chapter 4**). She and her team (especially Dr Tanya Maric and Dr Chidimma Kanu) enrolled patients, recorded metadata and collected the samples that were analysed for this thesis. Dr Kanu contributed to the ^{1}H NMR experiments on serum samples reported in **Chapter 4** and calculated OPLS-DA models for these samples.

Dr Ivan Serrano Contreras performed the 2D NMR experiments (COSY, TOCSY, HSQC) which were instrumental for identifying some of the metabolites reported in **Chapter 3** and **Chapter 4**.

Patient enrollment and coordination for the Jebsen study (**Chapter 5**) were led by Lucia Guzman and Dr Dennis Dixon at the Center for Autism and Related Disorders.

Dr Julie McDonald generated the raw 16S rRNA gene amplicon sequencing data presented in **Chapter 4** and **Chapter 5**.

^{1}H NMR experiments detailed in **Chapter 5** were carried out by the Imperial Phenome Centre (Imperial College London, UK). The work was supported by the Medical Research Council and National Institute for Health Research (grant number MC_PC_12025). Infrastructure support was provided by the National Institute for Health Research (NIHR) Imperial Biomedical Research Centre (BRC).

Figures in this thesis were created with Biorender.com

Contents

List of Figures **10**

List of Tables **12**

Abbreviations **13**

Chapter 1 **16**

1.1 Gut microbiota-host metabolic interactions **16**
1.1.1 Macronutrient metabolism 17
1.1.2 Bacterial catabolism 20
1.1.3 Bile acid metabolism 21
1.1.4 Microbial biosynthesis of host-essential compounds 21
1.1.5 Drug metabolism 22
1.1.6 Measuring host-microbe co-metabolism 23

1.2 Gut microbiota and human development **23**
1.2.1 Maternal gut microbiota during pregnancy 24
1.2.2 Maturation of the gut microbiota in offspring 25
1.2.3 Developmental deficits associated with the gut microbiota 26

1.3 Gut microbiota in health and disease **27**
1.3.1 Obesity 27
1.3.2 Autism Spectrum Disorder 28

1.4 Scope of thesis **30**

Chapter 2 **32**

2.1 Cohorts **32**
2.1.1 CWH study 33
2.1.2 Jebsen study 34

2.2 Metabolic profiling **34**
2.2.1 ^{1}H NMR spectroscopy 35
2.2.2 Sample preparation 37
2.2.3 Data acquisition 38
2.2.4 Data pre-processing 38
2.2.5 Metabolite annotation 40

2.3 Gut bacterial community profiling **42**
2.3.1 16S rRNA gene amplicon sequencing 43

2.3.2 Library preparation 44
2.3.3 Data acquisition 45
2.3.4 Data pre-processing 46
2.3.5 Normalisation and diversity estimates 47

2.4 Statistics 48
2.4.1 Linear dimensionality reduction techniques 48
2.4.1.1 Principal component analysis 49
2.4.1.2 Partial least squares regression 50
2.4.1.3 Cross-validation 52
2.4.1.4 Relevance networks 53
2.4.2 Random forests 54
2.4.3 Linear mixed-effects models 55
2.4.4 Time series analysis 55
2.4.5 Correlation 55
2.4.6 Correction for multiple hypothesis testing 56

Chapter 3 57

3.1 Introduction 57

3.2 Materials and methods 59
3.2.1 Study population, sampling and clinical data 59
3.2.2 Metabolic profiling of biofluids 60
3.2.3 Integration of metabolite data with clinical outcomes 63

3.3 Results and discussion 63
3.3.1 Characteristics of the study participants 63
3.3.2 Associations between maternal and neonatal outcomes 65
3.3.3 Sampling summary 67
3.3.4 BMI-associated serum metabolites 68
3.3.5 BMI-associated urine metabolites 74
3.3.6 Clinical associations 81

Chapter 4 85

4.1 Introduction 85

4.2 Materials and methods 87
4.2.1 Study population, sampling and clinical data 87
4.2.2 Metabolic profiling of biofluid 88
4.2.3 Gut bacterial community profiling 88
4.2.4 Integrative analysis of metabolic and gut bacterial data 89

4.3 Results and discussion **89**
4.3.1 Characteristics of the study participants 89
4.3.2 Associations between maternal and neonatal outcomes 91
4.3.3 Sampling summary 92
4.3.4 Serum metabolites associated with previous bariatric surgery 93
4.3.5 Urine metabolites associated with previous bariatric surgery 95
4.3.6 Malabsorptive metabolic profile is consistently altered throughout gestation 97
4.3.7 Malabsorptive metabolic phenotype is associated with a shift in the gut microbiota 100
4.3.8 Clinical associations 105

Chapter 5 **108**

5.1 Introduction **108**

5.2 Materials and methods **110**
5.2.1 Study population, sampling and clinical data 110
5.2.2 Metabolic profiling 111
5.2.3 Gut bacterial community profiling 113
5.2.4 Data integration 113
5.2.5 Gut bacterial associations with study groups and metabotypes 114

5.3 Results and discussion **115**
5.3.1 Characteristics of the study participants 115
5.3.2 Sampling summary 117
5.3.3 Metabolic and gut bacterial phenotypes are not overtly associated with ASD 118
5.3.4 Faecal metabolome in ASD 119
5.3.5 Urinary metabolites related to the ASD faecal metabolome 124
5.3.6 Gut bacteria related to ASD metabolome 127
5.3.7 Host-microbe co-metabolism is altered in children with ASD compared to neurotypical participants 129
5.3.8 ASD-associated metabotypes display differential microbiota 133

Chapter 6 **144**

6.1 Summary of key results **144**
6.1.1 Maternal obesity impacts pregnancy 144
6.1.2 Previous malabsorptive surgery impacts pregnancy 146
6.1.3 Gut microbial alterations in Autism Spectrum Disorder are conditional 148
6.1.4 Overall limitations 150

6.2 Future work **151**

6.3 General conclusion **153**

References **154**

Appendix A - Metabolite Identification **180**

Appendix B - Additional data **187**
Chapter 3 187
Chapter 4 193
Chapter 5 204

Appendix C - R packages and package versions **212**
Chapter 3 212
Chapter 4 213
Chapter 5 215

Appendix D - Publication **217**

List of Figures

Chapter 1

1.1 Metabolic phenotyping

1.2 Major microbiota-associated metabolites investigated in this thesis

Chapter 2

2.1 Study design and maternal sampling

2.2 ^{1}H NMR theory and data acquisition

2.3 ^{1}H NMR spectral alignment

2.4 Probabilistic quotient normalisation

2.5 NMR experiments

2.6 16S rRNA gene

2.7 16S rRNA gene amplicons with attached adapters and indices

2.8 Illumina sequencing by synthesis chemistry

2.9 Principal component analysis

2.10 Partial least squares regression

2.11 Multi-block partial least squares regression

2.12 Robust approximation of correlation based on PLS loadings

2.13 Random forests classifier

Chapter 3

3.1 Overview of analytical workflow for metabolic profiling

3.2 Associations between maternal and neonatal outcomes

3.3 Neonatal size and gestational age are associated with neonatal insulin levels at birth

3.4 Sampling summary by group

3.5 Serum median ^{1}H NMR spectrum with annotated metabolites

3.6 Serum metabolite concentrations throughout pregnancy

3.7 Metabolite concentrations in umbilical cord samples by study group

3.8 Metabolite concentrations in maternal serum compared to umbilical cord serum

3.9 Urine median ^{1}H NMR spectrum with annotated metabolites

3.10 Urine metabolite concentrations throughout pregnancy

3.11 Time series analysis of important microbiota-associated metabolites

3.12 Integrative analysis of metabolite and clinical data

Chapter 4

4.1 Neonatal insulin is associated with fat mass and weight at birth

4.2 Summary of data collection by group

4.3 Serum median ^{1}H NMR spectrum with annotated metabolites

4.4 Urine median ^{1}H NMR spectrum with annotated metabolites

4.5 Longitudinal modelling of discriminant metabolites

4.6 Gut bacterial alterations in malabsorptive patients

4.7 Differentially abundant amplicon sequence variants in malabsorptive patients

4.8 Integrated signature of previous malabsorptive surgery

4.9 Protein-related metabolites correlate with maternal insulin resistance and fetal/birth weight

Chapter 5

5.1 Overview of statistical workflow
5.2 Participants by ethnicity and state
5.3 Sampling summary
5.4 Cross-validated scores from orthogonal partial least squares discriminant analysis of metabolic and gut bacterial profiles
5.5 Sparse principal component analysis of faecal ^{1}H NMR data from children with ASD
5.6 Faecal ^{1}H NMR spectrum
5.7 Associations between faecal metabolites and ASD characteristics
5.8 Urine ^{1}H NMR spectrum
5.9 Associations between urinary metabolites and ASD characteristics
5.10 Associations between metabolites and microbiota in ASD
5.11 Comparison of metabolic phenotypes between study groups
5.12 Correlations between protein-related metabolites and faecal short-chain fatty acids
5.13 Metabotype prediction from gut bacterial profiles
5.14 Summary of metabotypes by study group
5.15 Variable importance from random forests model
5.16 Associations between bacterial genera and study groups
5.17 Associations between bacterial genera, study groups and metabotypes

Chapter 6

6.1 Observed metabolic relationship between maternal obesity and neonatal outcomes
6.2 Malabsorptive surgery-induced changes in protein metabolism
6.3 Relationship between intestinal transit time and protein putrefaction in the colon

List of Tables

Chapter 2

2.1 16S rRNA gene V1-V2 primer sequences

Chapter 3

3.1 Participant characteristics

3.2 OPLS model statistics for serum samples

3.3 Results from time series modelling of serum metabolites

3.4 OPLS model statistics for urine samples

3.5 Results from time series modelling of urine metabolites

Chapter 4

4.1 Participant characteristics

4.2 OPLS-DA model statistics for serum samples

4.3 OPLS-DA model statistics for urine samples

Chapter 5

5.1 Participant characteristics

5.2 Associations between metabolite levels and characteristics in neurotypical individuals

5.3 Random forests model statistics

Abbreviations

^{1}H NMR Proton Nuclear Magnetic Resonance Spectroscopy

1D 1-dimensional

2D 2-dimensional

ASD Autism Spectrum Disorder

BCAA Branched-Chain Amino Acid

BCFA Branched-Chain Fatty Acid

BMI Body Mass Index

bp base pairs

CARD Center for Autism and Related Disorders

CNS Central Nervous System

COSY Correlation Spectroscopy

CPMG Carr–Purcell–Meiboom–Gill

CWH Chelsea & Westminster Hospital

DNA Deoxyribonucleic Acid

ENS Enteric Nervous System

FDR False Discovery Rate

FID Free Induction Decay

FISH Fluorescence *In Situ* Hybridization

FMT Faecal Microbiota Transplant

GC-MS Gas Chromatography-Mass Spectrometry

GDM Gestational Diabetes Mellitus

GlycA *N*-acetyl Glycoproteins

GPCR G-protein-coupled receptors

HbA1c glycated haemoglobin

HDAC histone deacetylases

HOMA-IR Homeostatic Model Assessment for Insulin Resistance

HSQC Heteronuclear Single Quantum Coherence

IBD Inflammatory Bowel Disease

IBS Irritable Bowel Syndrome

IS indoxyl sulfate

LC-MS Liquid Chromatography-Mass Spectrometry

LPS lipopolysaccharide

MAL Malabsorptive

MSEP Mean Squared Error of Prediction

NGS Next-Generation Sequencing

NOESY Nuclear Overhauser Effect Spectroscopy

NW Normal Weight

OB Obese

OOB Out-Of-Bag

OPLS Orthogonal Partial Least Squares/Projections to Latent Structures

OTU Operational Taxonomic Unit

OW Overweight

PAG phenylacetylglutamine

PCA Principal Component Analysis

PCR Polymerase Chain Reaction

PCS *p*-cresyl sulfate

PHPA p-hydroxyphenylacetate

PLS Partial Least Squares/Projections to Latent Structures

ppm parts per million

qPCR quantitative (real-time) Polymerase Chain Reaction

RES Restrictive

rRNA ribosomal Ribonucleic Acid

RSPA Recursive Segment-wise Peak Alignment

SCFA Short-Chain Fatty Acid

sPCA sparse Principal Component Analysis

sPLS sparse Partial Least Squares/Projections to Latent Structures

spp. species

TMA trimethylamine

TMAO trimethylamine N-oxide

TOCSY Total Correlation Spectroscopy

TSP 3-trimethyl-silyl-[2,2,3,3-$^{2}H_{4}$]propionic acid

UK United Kingdom

Chapter 1

Introduction

1.1 Gut microbiota-host metabolic interactions

The human body contains entire ecosystems of microorganisms, referred to as microbiomes[1], and we are only beginning to appreciate their contribution to states of health and disease[2]. Different areas of the body harbour distinct microbial communities which contain populations of bacteria, viruses, fungi, archaea and other microbes. The communities are shaped by the unique microenvironments[3,4] forming ecological niches throughout and on the body. The gut microbiota, the collection of microbes residing in the gastrointestinal tract[1], is currently of much interest in clinical research because of its ability to influence human health and disease[5]. While the bacteria inhabiting the large intestine are currently the most extensively characterised members of the human gut microbiota, their activities and how these affect the host are still only loosely understood. Biochemical processes occurring within the human body consume and produce metabolites, small molecules that are start points, intermediates and endpoints of chemical reactions. Metabolites are found in all biofluids and, when considered together, depict the collective biochemistry of the host and its resident microbes as well as external factors including dietary preferences, pharmaceutical usage and environmental exposures[6] (**Figure 1.1**). Microbially synthesised metabolites make up a proportion of the metabolome - the total composition of metabolites present in a biological sample - of human tissues or biofluids and can perform important signalling functions within the host. Metabolic profiling has become a key tool for detailing both host and gut microbial metabolism, allowing us to understand the ways in which these two entities co-exist in symbiosis.

Current knowledge indicates that gut bacteria primarily assist with the digestion of nutrients which are indigestible by human metabolism[7]. Bacteria also metabolise exogenous chemicals such as pharmaceuticals[8], influencing host responses to drugs. Microbial metabolism generates compounds that are essential for host physiology including vitamins, amino acids, steroid hormones and neurotransmitters[9]. Microbe-generated compounds act locally in the gut but are

also absorbed into host circulation and can act as signalling molecules, modulating systemic function[10]. It is evident that the gut microbiota influences many aspects of human health; our understanding of these host-microbial interactions has the potential to propel clinical research and transform medical practice, offering new avenues for therapeutic intervention.

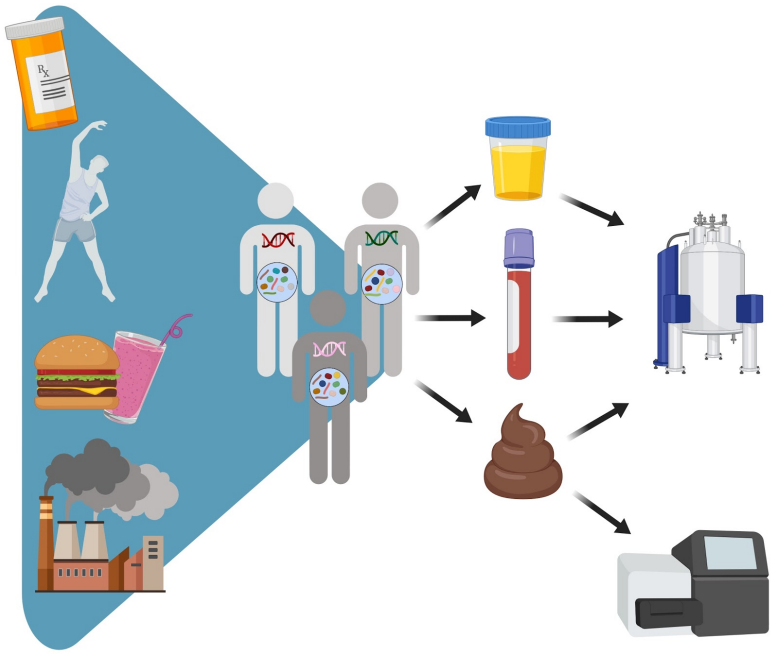


Figure 1.1: Metabolic phenotyping. Metabolic profiles can be acquired from biofluids (typically urine, blood and faeces) to give a snapshot of the biochemical reactions occurring in the human body at a given time. These profiles reflect host genetics, lifestyle choices, environmental influences and activity of the gut microbiota. Gut bacteria from faecal samples can be profiled in parallel to determine the species contributing to the host phenotype.

1.1.1 Macronutrient metabolism

The food we eat, among other factors, shapes the composition of the gut microbiota[11–13] and dictates much of the known bacterial metabolism in the colon. Humans consume three types of macronutrients which serve as fuel for the body: carbohydrates (e.g. starch, sugars, fibre), lipids (e.g. triglycerides, cholesterol) and proteins. Digestion begins in the oral cavity where food is homogenised by the action of chewing. In the mouth, salivary amylase provides the initial hydrolysis of starches, and lingual lipase begins the digestion of fats[14]. The bolus, or chewed

food, then travels to the stomach where gastric acid and enzymes initiate the digestion of proteins while the stomach contents are mixed by muscular contractions. Lipid digestion continues in the stomach. The resulting chyme is then pushed into the small intestine, the main site of chemical digestion and nutrient absorption. With the help of pancreatic enzymes and bile, nutrients are further degraded into absorbable particles. At this point, some carbohydrates have been broken down into monosaccharides (simple sugars) or oligosaccharides, lipids have been broken down to fatty acids and glycerol, and proteins have been broken down to peptides and amino acids[14]. As the digested particles move along the length of the small intestine, nutrients (including vitamins and minerals) are absorbed into the bloodstream for use by tissues and organs. Bacteria likely contribute to nutrient metabolism in the small intestine[15], though to a lesser extent than in the large intestine.

Undigested materials pass through the small intestine and into the large intestine. Dietary fibre, a type of complex carbohydrate, is not broken down by human enzymes but is a primary substrate for many bacterial genera residing in the colon[7]. The principal products of anaerobic bacterial carbohydrate fermentation in the gut are short-chain fatty acids (SCFAs) butyrate, propionate and acetate[16,17] (**Figure 1.2**). SCFAs are recognised as important modulators of human health with a diverse range of physiological effects. Butyrate is the primary source of energy for colonic epithelial cells[18] and has many beneficial roles pertaining to inflammation, gut barrier integrity and colonic homeostasis[17]. The known physiological effects of SCFAs are primarily mediated through two mechanisms - inhibition of histone deacetylases (HDACs) and activation of G-protein-coupled receptors (GPCRs)[19] - which are involved in regulating gene expression, metabolism and inflammation. SCFAs produced in the gut are readily absorbed into circulation[20,21], leaving a small fraction of the total SCFAs produced to be excreted in the faeces[22,23]. Like dietary fibre, polyphenols also pass through the small intestine in large quantities and undergo extensive biotransformation by gut bacteria[24]. Polyphenols are micronutrients found in vegetables, fruit and whole grains that have bioactive properties. Bacterial metabolism of these compounds affects their bioavailability and bioactivity, thus modulating their health benefits.

As carbohydrate availability decreases along the length of the large intestine, metabolism of proteins that escaped primary digestion increases[25,26]. Hydrolysis of proteins and anaerobic bacterial fermentation of peptides and amino acids, a process called putrefaction, also produces

SCFAs as well as branched-chain fatty acids (BCFAs), but the main products of protein putrefaction are phenols, indoles, amines, ammonia and hydrogen sulfide[27–29] (**Figure 1.2**). Unlike SCFAs and BCFAs, the main putrefaction metabolites are potentially toxic. These metabolites have been shown to exert genotoxic and cytotoxic effects, decrease the integrity of the intestinal barrier, and are uremic toxins[17,30–32]. The detrimental effects of putrefaction metabolites have not been directly demonstrated in humans[33], but colonic protein fermentation may be relevant to inflammatory bowel disease (IBD), irritable bowel syndrome (IBS) and colorectal cancer[30,31]. However, some amino acid-derived gut bacterial metabolites are involved in normal physiology such as the neuroactive compounds 4-aminobutryate (GABA), serotonin and histamine, which can be produced from bacterial metabolism of glutamate, tyrosine and histidine, respectively[30]. These molecules contribute to chemical communication within the gut-brain axis or host inflammatory responses.

Most dietary fat is metabolised before reaching the large intestine[14] but bacteria are involved in the metabolism of lipids that do arrive there[7,34]. An important group of lipid-related metabolites is the products of choline degradation[34]. Choline is an essential nutrient that needs to be obtained from the diet since the body does not produce a sufficient amount. It has been shown that the gut microbiota can modify choline bioavailability in the host by transforming it to trimethylamine (TMA)[35] which could contribute to choline deficiency. TMA is transported to the liver where it is oxidised to trimethylamine N-oxide (TMAO), a metabolite that has been shown to promote atherosclerosis[36,37]. Given the high amounts of TMAO in fish, consumption of which is associated with healthy outcomes, this metabolite may be a marker of disease rather than a mediator[38].

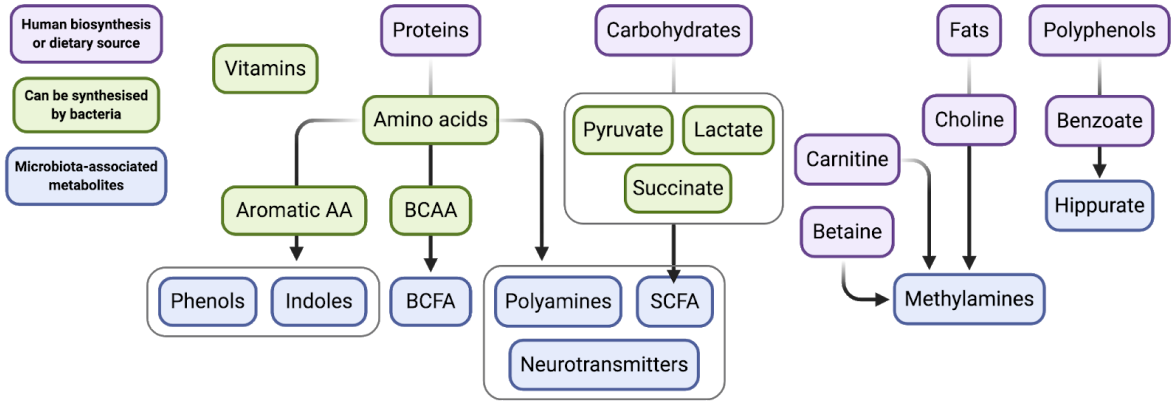


Figure 1.2: Major microbiota-associated metabolites investigated in this thesis. Figure is based on information from Holmes et al.[34] and Oliphant & Allen-Vercoe[7]. *AA amino acid; BCAA branched-chain amino acid; BCFA branched-chain fatty acid; SCFA short-chain fatty acid.*

1.1.2 Bacterial catabolism

Bacteria yield energy through the catabolism of organic compounds. Respiration, the complete oxidation of organic compounds to carbon dioxide and water, can occur in the presence or absence of oxygen and generates energy by producing adenosine triphosphate (ATP) through glucose dissimilation pathways (e.g. glycolysis), the tricarboxylic acid (TCA) cycle or glyoxylate cycle, and oxidative phosphorylation. In aerobic respiration, oxygen is the final electron acceptor while a compound other than oxygen (e.g. fumarate, sulfate, nitrate) is the final electron acceptor in anaerobic respiration. Fermentation is another anaerobic catabolic process. It differs from respiration in that the TCA cycle and oxidative phosphorylation are not utilised, therefore less ATP is generated. Pyruvate is used as a final electron acceptor, yielding products such as organic acids (e.g. lactic acid, acetic acid), ethanol and carbon dioxide.

Bacteria preferentially utilise specific carbon sources over others. As discussed in the preceding section, many species present in the human large intestine primarily metabolise dietary fibre, but protein metabolism becomes more prominent as carbohydrate availability decreases. In addition to substrate availability, chemical compositions, and chemical structures, bacterial regulatory mechanisms also dictate which substrate is utilised first. One such paradigm, catabolite repression, involves inhibition of metabolism of non-preferred substrates while the preferred, rapidly metabolisable substrate is available[39]. For example, in the presence of both glucose and lactose, *Escherichia coli* will utilise all available glucose before lactose is utilised. In

batch cultures, the presence of fermentable carbohydrate sources results in a 60-80% reduction in the production of putrefaction metabolites by gut microbiota[40,41], demonstrating the inhibition of protein fermentation while carbohydrates are available. Tryptophanase, the bacterial enzyme responsible for catalysing indole formation from tryptophan, is subject to catabolite repression by glucose, a main component of carbohydrate molecules[41].

1.1.3 Bile acid metabolism

Bile is a fluid produced by the liver, stored in the gallbladder and secreted into the duodenum (first portion of the small intestine) to aid digestion. Bile contains primary bile acids which are conjugated to glycine or taurine to form bile salts. Bile salts facilitate the emulsification and absorption of lipids. Over 95% of the secreted bile salts are reabsorbed in the ileum (final portion of the small intestine) and transported back to the liver where they will be reused in bile; this enterohepatic circulation occurs multiple times during a meal[14]. Bile salts are modified by gut bacteria to produce deconjugated and secondary bile acids[42]. These modifications contribute to bile acid recycling, permitting the reabsorption of some bile acids from the colon back into enterohepatic circulation. However, some secondary bile acids have toxic properties and have been implicated in colon and liver cancers[43]. Animal-based diets increase the abundance of bile-tolerant bacteria in the gut and the concentrations of secondary bile acids in the faeces[11], in response to increased fat intake and subsequent increase in bile secretion. Hence both the diet and the gut microbiota contribute to the composition of the circulating bile acid pool and therefore play a role in bile acid-mediated metabolic signalling, as well as the detrimental effects of modified bile acids.

1.1.4 Microbial biosynthesis of host-essential compounds

In addition to extracting nutrients from the diet, the gut microbiota has an impressive capacity to synthesise compounds that are essential for host physiology. Bacteria in the gut contribute to the amino acid pool through *de novo* synthesis, primarily using ammonia derived from dietary nitrogen sources[44]. Bacteria also synthesise essential vitamins such as biotin, cobalamin (vitamin B_{12}), folate, pantothenate (vitamin B_5), riboflavin (vitamin B_2), thiamine (vitamin B_1) and vitamin K[45,46]. From a genomic screen of 256 bacteria common to the human gut, it was reported that each of the eight B vitamins investigated was synthesised by 40–65% of the strains[46]. The extent to which humans rely on the microbial supply of vitamins is not clear, but

the large intestine is capable of absorbing bacteria-produced, water-soluble vitamins[47]. Additionally, conventional rats were shown to be protected from vitamin K deficiency when fed a vitamin K-deficient diet, while germ-free rats exhibited symptoms of vitamin K deficiency on the same diet[48]. Administration of vitamin K reversed the symptoms in germ-free mice, demonstrating an exogenous source of vitamin K was required in rats lacking a gut microbiota. The relationship between the gut microbiota and host vitamin K deficiency was similarly illustrated in a small (n=10) study of hospitalised elderly patients[49]. All patients received parenteral administration of nutrients without vitamin K. The antibiotic-treated patients developed symptoms of vitamin K deficiency while those without antibiotic treatment showed no symptoms of vitamin K deficiency, again demonstrating the host's ability to utilise bacteria-produced vitamins.

1.1.5 Drug metabolism

Like diet, the pharmaceutical compounds ingested can affect the composition and function of the gut microbiota. In addition to obvious anti-microbial actions such as those of antibiotics, some pharmaceuticals with initial host targets have been shown to inhibit the growth of bacteria common to the human gut[50]. However, some drugs can also be directly or indirectly metabolised by the gut microbiota[8]. Bacterial metabolism of pharmaceutical compounds is increasingly recognised as a substantial contributor to drug efficacy and toxicity[51]. Prodrugs, like those developed to deliver 5-amino salicylic acid to the colon for IBD treatment, exploit the enzymatic potential of the gut microbiota to hydrolyse and thus activate the drug when it reaches the colon[52]. Other pharmaceuticals, such as the cardiac drug digoxin, can be inactivated by the gut microbiota[53] through detoxification processes. Bacterial metabolism of certain drugs also mediates their toxicity to the host. For example, the bacterial enzyme β-glucuronidase is responsible for the toxicity of some anti-cancer drugs[54] (such as irinotecan) and nonsteroidal anti-inflammatory drugs (NSAIDs)[55]. Glucuronidation is a common phase II detoxification pathway which makes compounds more water-soluble and thereby marks them for excretion. β-glucuronidase cleaves the glucuronic acid from these compounds, releasing the active form of the drug in the gastrointestinal tract and causing gastrointestinal side effects. Inhibiting this enzyme alleviates the intolerable effects of irinotecan toxicity in mice, permitting higher dosage which ultimately improves the drug's efficacy[56]. Bacterial metabolism may also influence the host's ability to detoxify drug metabolites through sulfation. High urinary concentrations of the

host-microbe co-metabolite *p*-cresyl sulfate were shown to be indicative of decreased host sulfation capacity, limiting the sulfation of toxic paracetamol metabolites[57].

1.1.6 Measuring host-microbe co-metabolism

Faecal samples are a useful indicator of bacterial metabolism in the gut. However, the different rates of absorption for different compounds sometimes makes the interpretation of their clinical significance challenging[58]. Metabolites that are not utilised in the gut are either excreted in the faeces or taken up into circulation. A comparison between germ-free and conventional mice showed that many blood metabolites were only produced when the gut microbiota was present, and the concentrations of ≈10% of blood metabolites were significantly altered by the presence or absence of the gut microbiota[59]. The blood metabolome, or the collection of small molecules in a blood sample, is also a reliable predictor of alpha-diversity in the gut[60], providing further evidence of a systemic gut microbial signature. Circulating blood is continuously filtered through the kidneys, removing excess water and waste from the body which is then stored in the bladder as urine. Urine, although a waste product, reflects the biochemical processes that have occurred in the body and contains a considerable number of microbiota-associated metabolites[61–63]. Urine also provides a time-averaged metabolic profile (i.e. urine collects in the bladder) compared to the snapshot in time provided by a blood sample. Although the three biofluids described here contain complementary information about host and microbial metabolism, each will give a unique view of the system. By combining the information gleaned from these different compartments, a holistic view of gut microbiota-host co-metabolism can be achieved.

1.2 Gut microbiota and human development

Microbial influence on human health has been demonstrated at all stages of life, from birth until geriatric years[64]. The critical developmental window, starting from pregnancy and continuing into early childhood, is a period of developmental plasticity during which many factors contribute to an individual's long-term disease risk[65]. Maternal health and nutrition are important for the proper maturation of her fetus, but evidence suggests environmental cues *in utero,* including those from the vaginal and gut microbiota*,* also have a programming effect on the offspring. This paradigm is supported by epidemiological observations linking low birth weight, a manifestation

of an adverse intrauterine environment, to increased risks for childhood stunting, cardiovascular disease and all-cause mortality[66]. The fetoplacental unit derives its nutrients from the maternal supply, depending on both substrate availability and transport efficiency[67]. Given the substantial contribution of the maternal gut microbiota to maternal metabolism and the circulating metabolite pool, it is conceivable that the maternal gut microbiota influences the intrauterine environment and impacts fetal growth. From birth, the maturation of the neonatal gut microbiota parallels the maturation of other body systems[68]. The first few years of life are crucial for immune, metabolic and neurological development, which rely on external cues. Signals from the gut microbiota are thought to be key modulators of several developmental pathways during this period of plasticity; disrupting the developmental trajectory of the gut microbiota may have long-term consequences[68]. Around the age of three, the gut microbiota begins to resemble that of an adult, but there is a lack of data detailing the composition of the gut microbial community in children over three years old[69]. However, the evidence that exists suggests continued evolution of the gut microbiota past age three[69], potentially extending the window for microbiome-based interventions to optimise human development and reduce future disease risk.

1.2.1 Maternal gut microbiota during pregnancy

Pregnancy is a period of profound physiological and biochemical changes with significant remodelling of host metabolism to accommodate the growing fetus[70,71]. The extent to which the maternal gut microbiota is involved in pregnancy-related physiology is currently unknown. Some studies have shown that pregnancy induces a shift in the maternal gut microbiota[72,73] but one study with weekly sampling determined that the maternal gut microbiota was stable throughout gestation[74]. One of the clearest demonstrations of microbial impact on maternal metabolism during pregnancy is that when transplanted in germ-free mice, the third, but not first, trimester maternal gut microbiome induced inflammation, weight gain and insulin resistance[72], mimicking the adaptive changes that occur during pregnancy. Furthermore, the absence of a maternal microbiota in rodents results in offspring with deficits in immune and neural function[75], demonstrating that microbes facilitate proper development. The mechanisms of these interactions likely involve metabolic and endocrine signals from the maternal gut microbiota to the developing fetus. For example, bacteria-derived molecules were detected in the placenta and fetal circulation after pregnant mice were colonised with a strain of *Escherichia coli* only

during gestation[76]. This transient, monocolonisation stimulated postnatal development of the innate immune system in the offspring. In another animal model of pregnancy, SCFAs were shown to translocate from the maternal colon (through the bloodstream) to the embryo where they were sensed by embryonic GPCRs[77]. This maternal-embryonic signalling protected offspring from the high fat diet-induced obesity that was observed in the offspring of germ-free counterparts. Airway inflammation, a proxy for asthma, could also be prevented in the next generation by inoculating germ-free mice with a consortium of SCFA-producing bacteria[78].

A link between maternal gut microbial composition during pregnancy, fetal development and offspring outcomes has also been observed in humans, but the mechanisms have not been defined. The maternal gut microbiota has been implicated in preterm birth and low birth weight although the vaginal microbiota likely plays a larger role in the former[68]. Both probiotic and antibiotic usage during pregnancy are associated with allergic outcomes in offspring[79]. Prenatal and postnatal probiotic supplementation, in conjunction but not alone, may protect offspring from developing eczema. On the other hand, antibiotic treatment during pregnancy may increase a child's risk of developing asthma. The existing data is convincing that the maternal gut microbiota has the potential to exert transgenerational effects, potentially by priming the offspring immune system. However, large longitudinal studies of mother-child pairs with extensive clinical metadata will be required to understand which host-microbiome interactions during pregnancy are important for human development.

1.2.2 Maturation of the gut microbiota in offspring

Microbial colonisation of the placenta and fetus in healthy pregnancies remains a controversial topic[69]. From birth, an infant's gut microbiota is shaped by exposures, with its initial composition reflecting the maternal microbiota, delivery mode (vaginal or skin bacteria) and antibiotic exposure[69,80], while subsequent colonisation reflects the body site habitat[81]. Consistent patterns of bacterial assembly in the infant gut are observed across cohorts and continents in vaginally-born children without antibiotic treatment[80], illustrating ubiquitous interactions between the developing child and its resident microbes. *Bifidobacterium*, a genus recognised for its capacity to ferment human milk oligosaccharides[82], initially increases in abundance in the infant gut then starts to decrease around the age of weaning[80]. The relative abundance of *Proteobacteria* in the gut decreases over the first 24 months of life while *Clostridia* increases.

This inverse relationship marks the transition from a community composed of facultative and obligate anaerobes, to a primarily anaerobic community[80] that is more suitable for degrading dietary fibre. Although infant diet (breast- or formula-feeding) shapes the gut microbiota, the observations above remained significantly associated with age after adjusting for diet, indicating universal trends in the developmental trajectory of the gut microbiota. The community continues to evolve and increase in complexity during the first years of life, guided by the introduction of solid food and environmental exposures[68]. Host-microbe co-metabolites are increasingly produced during the first two years of life[83], showing an increase in functional potential with the changing community structure.

1.2.3 Developmental deficits associated with the gut microbiota

Collective evidence from animal models and epidemiological studies show that the early-life gut microbiome plays a key role in human development. A leading theory attributes the dramatic rise in the incidence of non-communicable diseases, particularly immune-related disorders, to interruptions in the normal progression of microbial establishment in and on the body, which likely disrupts developmental signals to the host[84]. These interruptions begin at birth with children delivered by caesarean section. Passage through the birth canal inoculates the neonate with the maternal vaginal microbiota whereas caesarean delivery inoculates the neonate with the maternal skin microbiota. Epidemiological studies have associated caesarean delivery with various detrimental outcomes in the offspring such as asthma, IBD, obesity and food allergies[84]. Similarly, disrupting the early-life gut microbiota with antibiotics may increase the risk of developing asthma, IBD, obesity and allergies[84]. Infants born by caesarean delivery show delayed vertical transmission of maternal microbiota and notably lack appropriate populations of *Bacteroides* spp. in the gut during the first few months of life[80]. The gut microbiota of caesarean-born infants can recover by age one but the delayed colonisation during this critical window may have lasting effects. Antibiotic perturbation of bifidobacteria populations also disrupts the normal colonisation pattern of the infant gut[80]. *Bifidobacterium spp.* are key members of the infant gut microbiota[68] whose abundance peaks during the first few months of life. After antibiotic treatment, populations of bifidobacteria may not recover within this window. The potential for correcting gut bacterial aberrations in early-life to improve developmental outcomes has been demonstrated in mouse models. One such study reported that

supplementation with *Bacteroides fragilis* at weaning improved immune and gastrointestinal pathophysiology, and ameliorated behavioural deficits in a mouse model of autsim[85].

1.3 Gut microbiota in health and disease

As an integral component of human physiology, the gut microbiota not only confers benefit to its host but may also be involved in disease pathogenesis. However, proving a causal link between the gut microbiome and disease phenotypes in humans is challenging. Nevertheless, understanding the compositional and functional changes observed in disease-associated communities may inform on disease pathology and open new avenues for treatments. Compositionally, the gut microbiota is highly individualised and largely depends on age, diet, pharmaceutical usage, lifestyle and health status[86–88]. The metabolic pathways encoded in the genomes of each community (the metagenome) are remarkably consistent[4], suggesting a "healthy" microbiota depends on a core set of genes or metabolic capabilities rather than a core set of species[89]. Nevertheless, variation in the gut microbiota, attributed to a case compared to a control population, has been documented in a number of gastrointestinal and extra-gastrointestinal disorders[90]. Two such conditions are investigated in this thesis in the context of human development: maternal obesity and Autism Spectrum Disorder (ASD).

1.3.1 Obesity

The prevalence of obesity and the exponential increase in its incidence are global concerns[91]. Obesity, defined as excessive fat accumulation resulting in a body mass index (BMI) of 30 kg/m^2 or greater, is a preventable condition that is associated with a multitude of medical comorbidities. Obese individuals have a high risk of cardiovascular diseases, diabetes, musculoskeletal disorders and some cancers, which ultimately increase the risk of mortality. Abnormal accumulation of fat is caused by an imbalance between caloric intake and energy expenditure. Increasing consumption of foods that are high in fat and sugar combined with the increasingly sedentary lifestyle of the Western world has fuelled the obesity epidemic. Given the metabolic capacity of the gut microbiota and its role in energy harvest[92], gut bacterial composition in relation to host BMI has been an area of focus in microbiome research. Meta-analyses have failed to identify a consistent taxonomic signature associated with obesity[90,93,94], but have found that alpha-diversity is significantly reduced in obese individuals[93]. Despite the lack of a gut microbial signature for obesity, consistent changes in

microbiota-associated metabolites have been reported[95–100]. Hippurate, a host-microbe co-metabolite of benzoic acid, was negatively associated with BMI in two studies[99,101] while TMAO, a product of the bacterial breakdown of dietary choline, was positively associated with BMI across multiple studies[100]. Higher faecal SCFA concentrations[95,96,98] but lower blood SCFA concentrations[97] have also been reported in obese individuals, but the relevance of SCFAs in obesity is not entirely understood. For example, acetate has been shown to suppress appetite[102] but also to promote metabolic syndrome[103]. Given the central role that SCFAs play in human metabolism, the relationship between these metabolites and obesity is likely complex.

Amino acid metabolism is consistently altered in obese individuals, which exhibit elevated concentrations of branched-chain and aromatic amino acids in circulation[100]. BCAAs in particular are also associated with type 2 diabetes mellitus[104–106], a common comorbidity of obesity, and have been shown to induce insulin resistance in mice on a high fat diet[104]. High blood concentrations of BCAAs in insulin-resistant individuals were positively associated with BCAA biosynthetic potential in the gut[105]. This relationship was mainly attributed to the abundance of *Prevotella copri*. Inoculating mice with *P. copri* increased blood BCAA levels and insulin resistance, demonstrating a causal relationship between the gut microbiota and obesity-associated insulin resistance. Although the relationship between the gut microbiota and obesity is equivocal, the potential of treating obesity and its comorbidities through the gut microbiota is promising. Recently, a probiotic strain was shown to correct obesity-associated metabolic deficits in mice with diet-induced obesity[107]. Transplanting the faecal virome from mice with a lean phenotype to mice with an obese phenotype was also shown to reduce high-fat diet-induced weight gain and improve glucose tolerance[108]. Such approaches are crucial to address the growing obesity epidemic, especially in cases where dietary and lifestyle interventions are ineffective.

1.3.2 Autism Spectrum Disorder

Neurodevelopmental disorders have gained attention as manifestations of altered gut-brain signalling, although the evidence is lacking in human studies[109]. The bidirectional communication system that exists between the gastrointestinal tract and the brain, termed the gut-brain axis, is mediated by the gut microbial community[10]. The gut, innervated by the enteric nervous system (ENS), is a primary host-microbe interface in the human body. In addition to

local effects, microbes residing in the gut can signal to systemic organs through a variety of neural, endocrine, immune and metabolic pathways[110]. Communication between the ENS and the central nervous system (CNS) can occur through the vagus nerve, which can be activated by gut bacteria. Gut bacteria are also involved in the production of a number of neurotransmitters including serotonin, dopamine, noradrenaline, GABA, acetylcholine and histamine[10]. Stimulation of cytokine production enables the gut microbiota to activate the hypothalamic-pituitary-adrenal axis which regulates the stress response[110]. The various routes of communication between the gut microbiota and the CNS suggest that, in addition to its involvement in developmental pathways, the gut microbiota may also play an active role in the maintenance of behaviours associated with altered neurodevelopment.

The gut-brain axis incorporates neural, immune and metabolic signalling, all of which are proposed to be affected in ASD[111]. ASD is a neurodevelopmental disorder characterised by impaired communication and sociability, with unknown aetiology and multiple comorbidities (discussed in **Chapter 5**). Two decades ago, a report of improved autistic symptoms while the children were taking an oral antibiotic[112] ignited interest in the ASD-associated gut microbiota. The prevalence of gastrointestinal dysfunction in children with ASD, and the positive correlation between the severities of gastrointestinal and behavioural symptoms[113,114], support the notion that the gut microbiota is a considerable factor in ASD pathophysiology, either causally or as a contributor to the symptomatology. Like with obesity, a definitive microbiota-associated signature in ASD has not been identified to date, although most studies have reported a perturbation in gut microbial composition or activity[115–119]. Metabolic profiling studies have found that gut bacterial metabolism is disrupted in ASD but again, there is no consensus on the directionality of the metabolic changes. For example, urinary excretion of phenols and indoles was increased[120–122] or decreased[123,124] in children with ASD compared to children with typical development. The presence of gastrointestinal symptoms, which were not factored into the analyses of the aforementioned studies, is a potential confounder[125]; these data should be collected and included as a covariate when analysing data pertaining to children with ASD. Stool consistency may be more variable in the ASD population compared to neurotypical controls due to the prevalence of gastrointestinal dysfunction. This factor has a large effect on the faecal microbiota[86] and consequently, the microbiota-associated metabolite profile[126].

The heterogeneity in ASD clinical presentation and lack of sufficiently powered studies contribute to the inconsistencies observed when studying this population[115,116]. Differential findings in the faecal bacterial composition and metabolite profile have been reported between subtypes of ASD[127], underscoring the importance of considering ASD as a spectrum and investigating variation within the population. Rather than a distinct bacterial or metabolic signature, multiple ASD phenotypes likely exist which reflect varying degrees of ASD severity, different patterns of comorbidities and gastrointestinal status. Differentiating ASD subtypes may enhance treatment efficacy in specific groups of children and may uncover targets for novel treatments. Recently, the beneficial effects of faecal microbiota transplant (FMT) on both gastrointestinal and behavioural symptoms in children with ASD were shown to be maintained two years after treatment[128]. While the results are promising, safer alternatives to FMT need to be developed due to the high risk posed by introducing a live, uncharacterised microbial community into a patient[129]. Formulating specific consortia of microorganisms or cocktails of metabolites meant to treat ASD patients with certain phenotypes would reduce the risk of transferring pathogens and unknown disease-potentiating microbes.

1.4 Scope of thesis

A primary goal of clinical research is to improve understanding of human pathophysiology and identify prophylactic measures. Current knowledge about the role of the gut microbiome in human development is limited. The purpose of the work presented in this thesis was to identify metabolic links between the human gut microbiota and altered development, generating hypotheses to test with future *in vitro* and *in vivo* experiments. First, maternal microbiota-associated metabolism during gestation was investigated with regard to fetal development. **Chapter 3** details a network of metabolic changes in obese women throughout pregnancy that were ultimately linked to elevated insulin levels and high birth weight in the neonates. However, as reported in **Chapter 4**, women with a high BMI and previous weight loss surgery were protected from obesity-induced insulin resistance during pregnancy and were less likely to have large neonates. These differential outcomes were associated with structural and functional changes in the gut microbiome, but the causal nature of these changes is yet to be determined. In addition to the prospective study, an observational study was undertaken to investigate metabolic relationships between the gut microbiome and altered neurodevelopment. **Chapter 5** details findings from a childhood cohort focused on ASD. Retrospective

microbiota-associated signatures of altered neurodevelopment were not identified, suggesting prospective birth cohorts may be necessary to catch signatures that precede ASD symptoms. Nevertheless, ASD was associated with metabolic signatures of gastrointestinal dysfunction related to the gut microbiota. The observational findings from these two projects are summarised in **Chapter 6**, along with potential mechanistic experiments to follow up the hypotheses generated from this work.

Chapter 2

Materials and Methods

Modern medicine transformed the world with hygiene, antibiotics, vaccines and health education. Despite its success eradicating deadly diseases, reducing mortality and improving longevity, the rise of chronic non-communicable diseases is a growing threat to healthcare systems worldwide[130]. Understanding disease phenotypes to work toward efficacious treatments and prophylactic measures remains a top research priority. Non-communicable diseases usually arise from interactions between genetic, environmental and lifestyle factors, resulting in complex phenotypes. Systems-level approaches are necessary to provide a comprehensive illustration of each disease and to ascertain actionable targets. The present work details high-throughput phenotypic profiling ("omics") of multiple human biospecimen sample types using two analytical platforms and statistical analysis of the resulting profiles to decipher disease- versus health-associated biology. This chapter provides an overview of the cohort studies and the methods used to interrogate the populations.

2.1 Cohorts

This thesis details observations from two human cohorts focusing on two stages of human development: 1) pregnancy and 2) childhood. The pregnancy study carried out at Chelsea & Westminster Hospital (CWH) aimed to characterise gestational metabolic disturbances associated with maternal obesity (dataset 1, **Chapter 3**) or previous bariatric surgery (dataset 2, **Chapter 4**), and the association between maternal metabolism and neonatal outcomes. A cohort of children with Autism Spectrum Disorder (ASD), their neurotypical siblings and unrelated neurotypical controls was also investigated to uncover host-microbe interactions associated with altered neurodevelopment (**Chapter 5**). From each individual studied, a combination of biospecimens was collected and profiled, often for multiple points in time, to understand the dynamic interplay between the gut microbiota, host metabolism and human health. Together, utilising data from 825 individuals and 2,404 clinical samples, the two projects

highlight the influence of bacteria residing in the gut on host metabolism, and the clinical relevance of this interaction in human development.

2.1.1 CWH study

The patient cohort is part of an ongoing prospective study[131–133]. The study was approved by the West London Research Ethics Committee (No: 14/LO/0592) and informed consent was obtained from all participants. Pregnant women were recruited between May 2015 and April 2017. Participants were enrolled during pregnancy and followed until delivery (**Figure 2.1**). At each hospital visit (**T1**: 11-14, **T2**: 20-24, **T3**: 28-30, **T4**: 30-33, **T5**: 35-37 weeks gestation and **T6**: delivery), blood serum and urine samples were collected, and biometrics were recorded. At T1, T2 and T4, faecal samples were collected from a subset of participants (dataset 2). Trans-abdominal ultrasound scans were performed at T2, T4 and T5 to estimate fetal weight and calculate percentiles for the gestation. At T3, a full oral glucose tolerance test (2 h, 75 g) was conducted as a diagnostic test for the development of gestational diabetes[134] and maternal insulin resistance was calculated using the homeostatic model assessment for insulin resistance (HOMA-IR):

$$HOMA-IR = fasting\ serum\ insulin\ (\mu U/L)\ x\ fasting\ glucose\ (mmol/L)/22.5$$ [135]

Plasma glycated haemoglobin (HbA1c) was also measured at this time point to assess the long term glucose control. HbA1c is another diagnostic marker for diabetes. Birth weight and percentiles were recorded at T6 and, where possible, umbilical cord blood serum and neonatal urine were collected. Insulin and glucose levels were measured in cord samples and neonatal insulin resistance was calculated. All samples were stored at -80°C for future analysis.

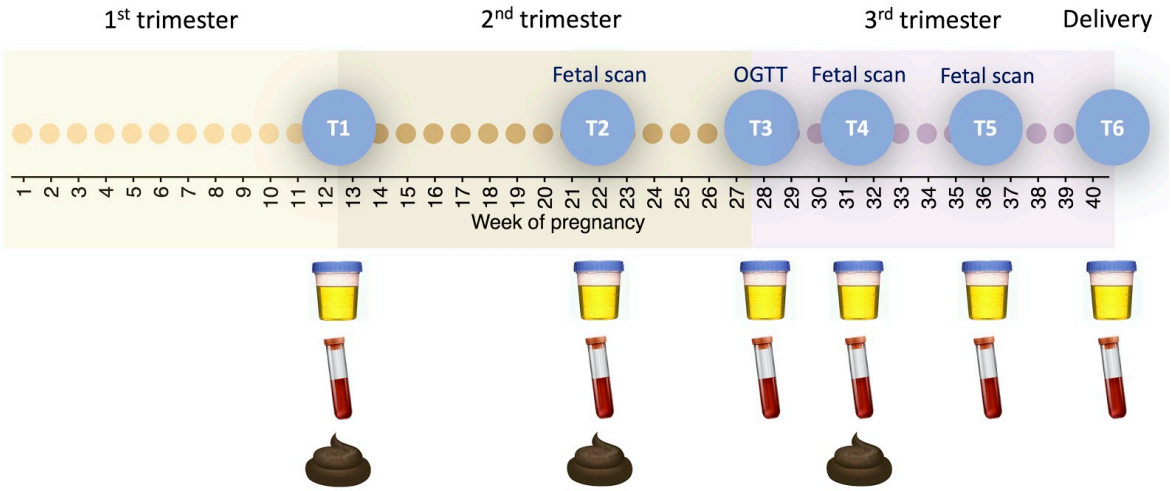


Figure 2.1: Study design and maternal sampling. Participants were seen at five time points during pregnancy (T1 - T5) and at delivery (T6). Urine and blood serum samples were collected at each visit while faecal samples were collected at T1, T2 and T4. Fetal growth scans were performed at T2, T4 and T5. An oral glucose tolerance test (OGTT) was performed at T3.

2.1.2 Jebsen study

Children were recruited for the study through the Center for Autism and Related Disorders (CARD) at various locations throughout the United States between December 2015 and June 2017. The study was approved by the CARD Institutional Review Board (IRB: 00004971) and informed consent was obtained from a legal guardian for each participating child. Individuals were asked to donate a faecal and a urine sample upon entering the study with the option of donating additional urine and faecal samples monthly for up to 8 months. Biospecimens were collected by the caregiver, stored at -20°C upon collection and kept frozen until being stored at -80°C within a CARD centre. Samples were shipped to Imperial College London on dry ice in 3 batches. For each collection time point, caregivers were asked to complete a questionnaire detailing the child's age, height, weight, current medical issues, medications and dietary information.

2.2 Metabolic profiling

Historically, small panels of known metabolites were used to detect aberrant physiology and diagnose disease. High-throughput analytical methods now allow the untargeted detection and

measurement of hundreds to thousands of metabolites from a single biospecimen, permitting a detailed biochemical profile of an individual that reflects a holistic physiological view. Metabolic profiling, also referred to as metabolomics or metabonomics, is primarily accomplished with the following analytical platforms: proton nuclear magnetic resonance (^{1}H NMR) spectroscopy and mass spectrometry-based techniques such as liquid chromatography coupled to mass spectrometry (LC-MS), gas chromatography coupled to mass spectrometry (GC-MS) and mass spectrometry imaging. ^{1}H NMR spectroscopy and LC-MS are most commonly used for global metabolic profiling of human biofluids. ^{1}H NMR is a non-destructive analytical method that is robust but lacks sensitivity, resulting in fewer detected metabolites than LC-MS. Although LC-MS is a highly sensitive analytical method, results are not as reproducible as those obtained from ^{1}H NMR[136]. Both platforms provide complementary information, and multiple assays will provide a complete picture of the metabolome. However, given the ease of sample preparation, non-destructive and quantitative nature, reproducibility, low cost and relatively fast analysis offered by ^{1}H NMR spectroscopy, it is an ideal first step in an exploratory metabolomics workflow and was the platform used for generating metabolomics data throughout this thesis.

2.2.1 ^{1}H NMR spectroscopy

NMR spectroscopy is a fundamental tool for metabolic profiling[136]. It enables the detection of specific atomic nuclei in a complex matrix such as a human biofluid. Since hydrogen is the most abundant element in biological molecules, ^{1}H NMR spectroscopy is often used for metabolic profiling although other nuclei (such as ^{13}C and ^{15}N) can yield NMR spectra. Certain atomic nuclei have a quantum mechanical property called nuclear spin (I), an intrinsic angular momentum, which gives them characteristic magnetic properties. The nucleus of a hydrogen atom contains a single positively charged proton giving it an atomic mass of approximately 1. Nuclei with an odd mass will have half-integer spins and in the case of hydrogen, I=½. The spin number of a nucleus dictates the number of possible orientations or energy levels (2I+1). In the absence of an external magnetic field, nuclei spin in random directions. When a magnetic field is applied, the magnetic moment associated with the non-zero spin will precess around the axis of the magnetic field and the orientations of the nuclei will be split over two (2(½)+1) energy levels (**Figure 2.2a**). The nuclei align with (low energy state, m=+½) or against (high energy state, m=-½) the magnetic field and the energy difference between the two states is dependent on the strength of the magnetic field. The rate of procession of the nuclei, the Larmor frequency,

is proportional to the strength of the magnetic field and the gyromagnetic ratio (ratio of nuclear magnetic moment to nuclear angular momentum)[137].

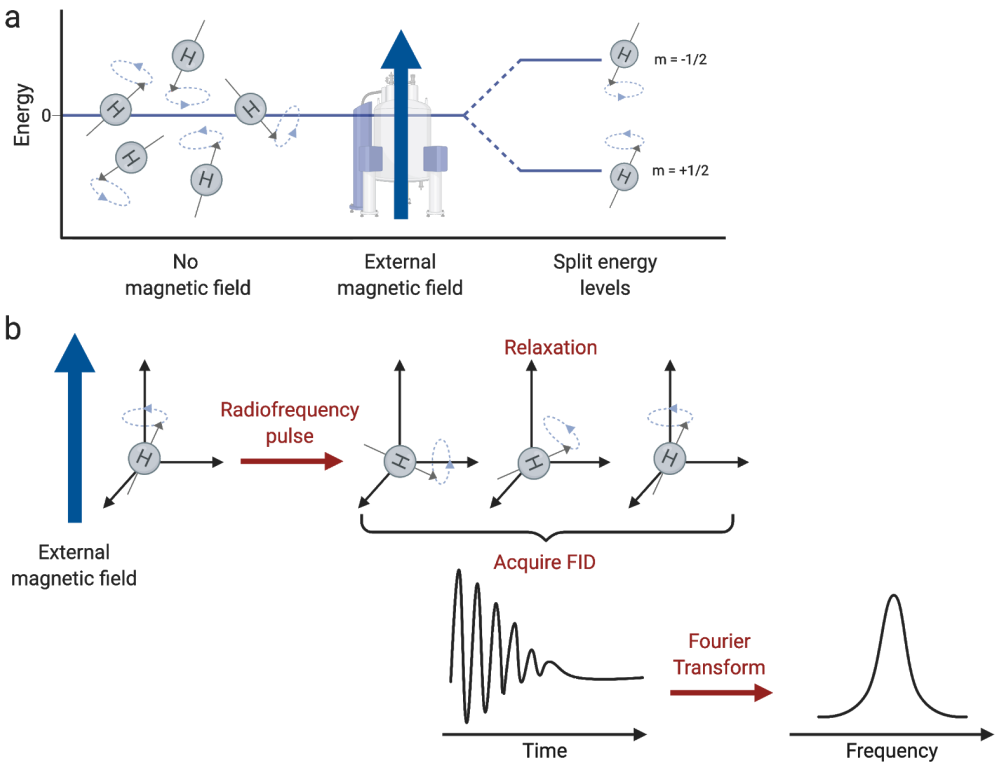


Figure 2.2: ^{1}H NMR theory and data acquisition. a) In the absence of an external magnetic field, hydrogen nuclei spin in random directions. When a magnetic field is applied, nuclei orient themselves with or against the magnetic field. **b)** A radiofrequency pulse perpendicular to the external magnetic field causes an excitation and as the nuclei relax back to equilibrium, their precessions are detected (free induction decay, FID). The signals are then transformed into frequencies.

Since a majority of proton nuclei will orient themselves with the applied magnetic field (low energy state), there is a net magnetisation in this direction. The orientations of these nuclei can be flipped from low to high energy states by supplying a specific amount of energy with a radiofrequency pulse perpendicular to the external magnetic field (**Figure 2.2b**). The frequency required to cause this excitation is equal to the Larmor frequency and to the difference in energy between the spin states. Nuclei can absorb this energy causing the magnetic moment to oppose the applied magnetic field (high energy state). After the radiofrequency pulse, nuclei relax back to the low energy state and their precessions are observed, generating a signal called the free

induction decay (FID). During this relaxation time, signals from all protons in a sample are simultaneously acquired. To separate signals from protons in different chemical environments, the FID is converted from the time-domain to the frequency-domain using Fourier transformation[137]. The transformation results in the NMR spectrum where peaks represent the intensities of resonances of protons contained within the sample. The intensity of a peak is directly proportional to the number of protons observed, hence it is a direct readout of the concentration of the relevant molecule in the original sample. Since each chemically inequivalent proton will resonate at different frequencies, every molecule in the original sample will be represented as a unique fingerprint in the NMR spectrum.

2.2.2 Sample preparation

^{1}H NMR spectra can be obtained from many different biospecimen types to provide a multifaceted view of human biochemistry and host-microbial interactions. Since urine and faecal samples can be collected conveniently and non-invasively, and blood is routinely collected at doctor visits, these are commonly used in human metabolomics studies and were the focus of this work. Upon collection, the samples were stored at -80°C until preparation and analysis. The unique characteristics of each biofluid necessitate different treatment prior to analysing the chemical composition by ^{1}H NMR spectroscopy. Urine is composed of 91–96% water[138], making the sample preparation relatively simple; urine was defrosted and centrifuged to remove any large particles then supernatant was mixed with an NMR-suitable buffer. At collection, blood needs to be partitioned to either plasma or serum[139]. Plasma is the main fluid component of blood without red blood cells, white blood cells and platelets. Serum can be obtained by allowing the blood to clot, thus clotting factors are removed from the sample. Although both plasma and serum samples are suitable for metabolic profiling and would likely lead to similar biological conclusions[140], serum is not affected by adding an anticoagulant. All data pertaining to blood in this thesis was from serum samples. Prior to analysis, serum was defrosted, centrifuged and supernatant was mixed with an NMR-suitable buffer. The preparation of faecal samples for ^{1}H NMR spectroscopy requires additional steps due to the extent of unwanted solid material. A typical faecal sample is 25% particulate waste which includes bacterial biomass and undigested dietary components[138]. To remove the solid particles, the samples were extensively homogenised in an NMR-suitable buffer and then filtered. Regardless of biospecimen type, the buffer required to analyse samples by NMR contains deuterated water (D_2O), salts with

non-exchangeable hydrogen atoms (such as phosphates) and sodium azide (to prevent microbial growth). Deuterium ions resonate at a different frequency than hydrogen atoms and the signal can therefore be used to compensate for drifts in magnetic field strength (lock system). Detailed methods for sample preparation are contained within each subsequent chapter.

2.2.3 Data acquisition

Different pulse sequences can be used depending on the desired signals. Pulse sequences are a set of instructions given to the NMR spectrometer that dictate which spectral signals will be recorded in the spectrum. A standard one-dimensional (1D) ^{1}H NMR experiment, nuclear Overhauser effect spectroscopy (NOESY) with water suppression (to eliminate signals from protons in water) was performed on all samples described in this thesis. 1D-NOESY is the most popular pulse sequence for collecting metabolic information from biological samples as it gives a general overview of all molecules present in the biofluid[141]. However, blood-based samples contain macromolecules such as lipoproteins which overshadow the signals from low molecular weight metabolites in 1D-NOESY spectra. The Carr–Purcell–Meiboom–Gill (CPMG) spin-echo sequence can be used to selectively acquire signals from the low molecular weight metabolites[141]. In addition to NOESY experiments, CPMG experiments with water suppression were also performed for all serum samples. To aid with structure elucidation (see **2.2.5**), two-dimensional (2D) experiments were also run[142]. J-resolved spectra were acquired for all samples to provide information on the multiplicity and coupling patterns of resonances. For unidentified signals, correlation spectroscopy (COSY), total correlation spectroscopy (TOCSY) and heteronuclear single quantum coherence (HSQC) experiments were run on a single sample with the highest concentration of the unidentified signals.

2.2.4 Data pre-processing

In ^{1}H NMR spectroscopy, an internal reference standard can be used to calibrate the spectra. Here, sodium 3-(trimethylsilyl) propionate-2,2,3,3-d_4 (TSP, deuterated) was used as it shows a single resonance at very low frequency in water[143]. This signal is used to calculate the chemical shifts (δ) of the resonances observed which are reported in parts per million (ppm).

$$\delta(ppm) = \frac{frequency\ of\ signal\ (Hz) - frequency\ of\ reference\ (Hz)}{frequency\ of\ spectrometer\ (MHz)} \times 10^6$$

Since different magnetic field strengths can be used, referencing the signals to a specific resonance and normalising by the spectrometer frequency results in data that is comparable across different platforms, independent of the field strength. In this case, a 600 MHz NMR spectrometer was used. Acquired spectra were automatically phased (making the peaks as narrow and symmetric as possible), baseline corrected and referenced to TSP (δ=0.00) in the Bruker TopSpin 3.1 software. Raw data were digitised in MATLAB version R2016b (The MathWorks, Inc., Natwick, MA) using in-house scripts developed by Dr Tim Ebbels[144]. Biologically irrelevant signals (e.g. TSP and water) were removed from the spectra. Chemical shifts are sensitive to pH and other sample-dependent or technical factors[142] necessitating the use of an algorithm to align peaks across all samples in a dataset. Here, spectra were aligned using the recursive segment-wise peak alignment (RSPA) algorithm[145]. The goal of the RSPA algorithm is to reposition all peaks in an NMR spectrum to match a reference spectrum (**Figure 2.3**). The reference spectrum is chosen by finding the spectrum that is most similar to all other spectra in the dataset. Alignment is done in segments and recursively; correction is performed then segments are divided into smaller segments to improve alignment and this continues until no further alignment is required.

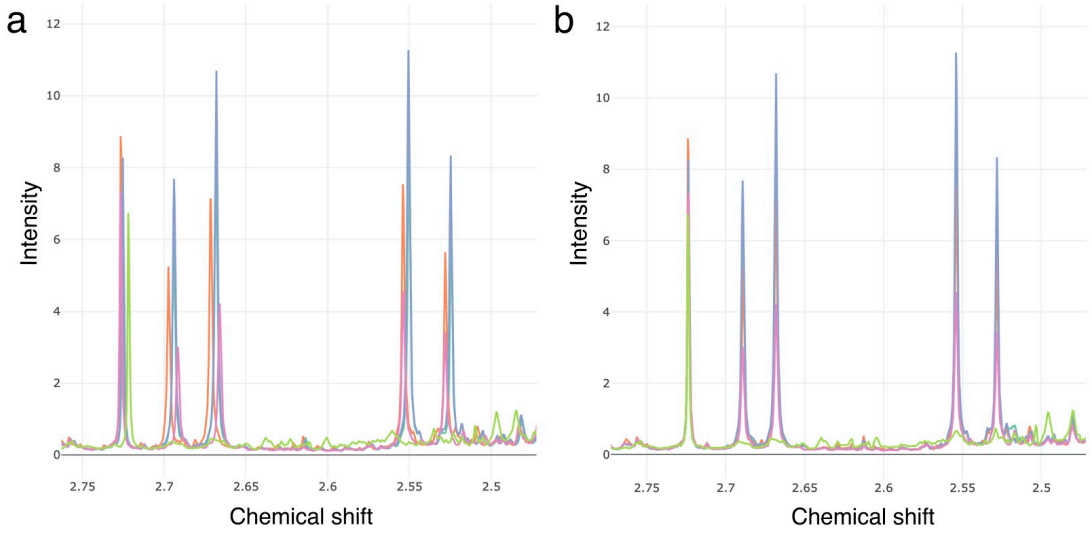


Figure 2.3: ^{1}H NMR spectral alignment. Due to slight variations in chemical shifts **(a)**, peaks need to be aligned **(b)** so that spectral variables are comparable across samples.

Another sample-dependent factor which causes unwanted variation across samples in a dataset is dilution. Biological samples have varying water content which is influenced by an individual's

diet, fluid intake and frequency of defaecation[138]. To scale samples to the same effective concentration, each dataset was normalised with the probabilistic quotient normalisation (PQN) method[146]. In the PQN method, the most probable dilution factor for each spectrum is determined by comparing it to a reference (median) spectrum. A quotient is calculated for each spectral variable by dividing its value by the value of the corresponding variable in the reference spectrum (**Figure 2.4**). The median quotient is then used as the dilution factor; every variable in the spectrum is divided by the dilution factor to calculate a normalised spectrum.

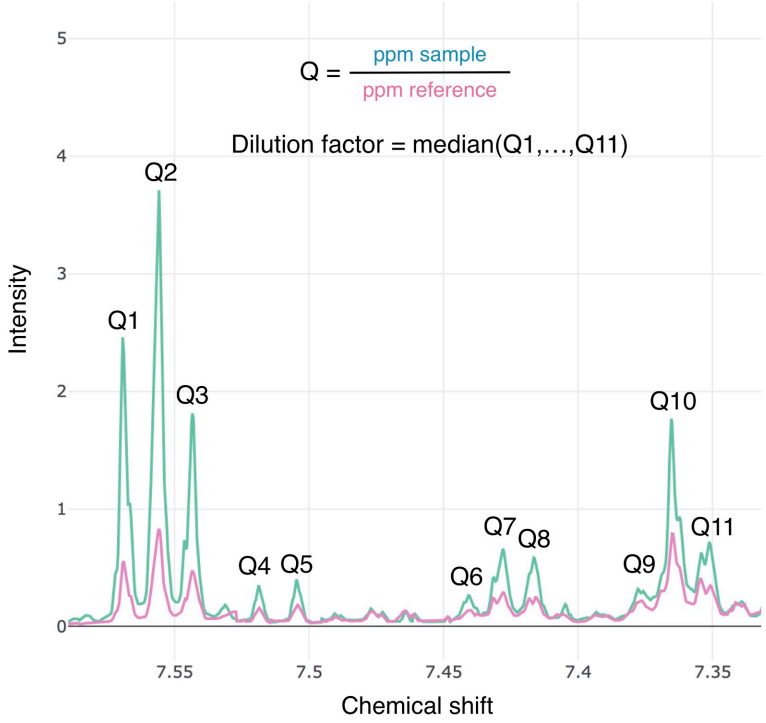


Figure 2.4: Probabilistic quotient normalisation. Each spectrum is normalised with respect to a reference (median) spectrum. A quotient (Q) is calculated for each spectral variable (simplified for illustration) and the median quotient is used as the dilution factor to normalise the spectrum.

2.2.5 Metabolite annotation

Each 1D ^{1}H NMR spectrum details tens of thousands of data points from a biological sample. The data points are organised into peaks which indicate the intensity of a signal at a specific resonance frequency. In the simplest case, a molecule can be represented by a single peak. A single ^{1}H NMR peak indicates all hydrogen atoms in the molecule are equivalent (e.g. methane,

CH_4). The position of the peak in the spectrum (chemical shift) depends on the molecular environment around the observed proton. Electrons around the nucleus shield it from the applied magnetic field. The electronegativity of the other atoms in the molecule affect the electron density around the proton. More electronegative atoms attract electrons, effectively deshielding the proton from the external magnetic field which allows it to resonate at a higher frequency. Conversely, protons close to less electronegative atoms in a molecule will be more shielded from the external magnetic field and will therefore resonate at a lower frequency.

Molecules with more complex structures will have nonequivalent hydrogens resulting in multiple signals pertaining to the same molecule. In addition, due to a phenomenon called spin-spin coupling, the resonance for a proton will be split into N + 1 peaks where N is the number of hydrogens on the atom adjacent to it. The ratio of intensities of the split peaks usually follows Pascal's triangle (**Figure 2.5a**). J-resolved spectra can be used to verify splitting patterns for each resonance (**Figure 2.5b**) and to resolve overlapping peaks. Due to the high number of peaks observed in a single ^{1}H NMR spectrum acquired from a human biospecimen, peaks representing the same molecule need to be identified using statistical methods such as Statistical Total Correlation Spectroscopy (STOCSY[147]) and SubseT Optimization by Reference Matching (STORM[148]). STOCSY calculates covariance and correlation between the intensity of the peak of interest and the intensities of all other spectral variables to visualise both structural and biological associations. This method is also implemented in STORM but is only applied to a subset of spectra containing the peak of interest. Once all visible peaks are identified for a given molecule, that molecule can be annotated by matching chemical shifts, multiplicities and J-couplings (frequencies between split peaks) to reference databases such as Human Metabolome Database (HMDB[149]) or the literature. Further structural information can be obtained for ambiguous annotations by running 2D NMR experiments (**Figure 2.5c-d**) which provide information about the interactions between nuclei. A COSY spectrum will detail correlations between protons that are 2-3 bonds apart while TOCSY will show correlations between all protons of a spin system. Using these experiments, additional proton resonances from the molecule of interest can be identified. HSQC experiments detail the chemical shift for the carbon (^{13}C) attached to each proton. Final metabolite identities are confirmed by spiking in authentic standards[139].

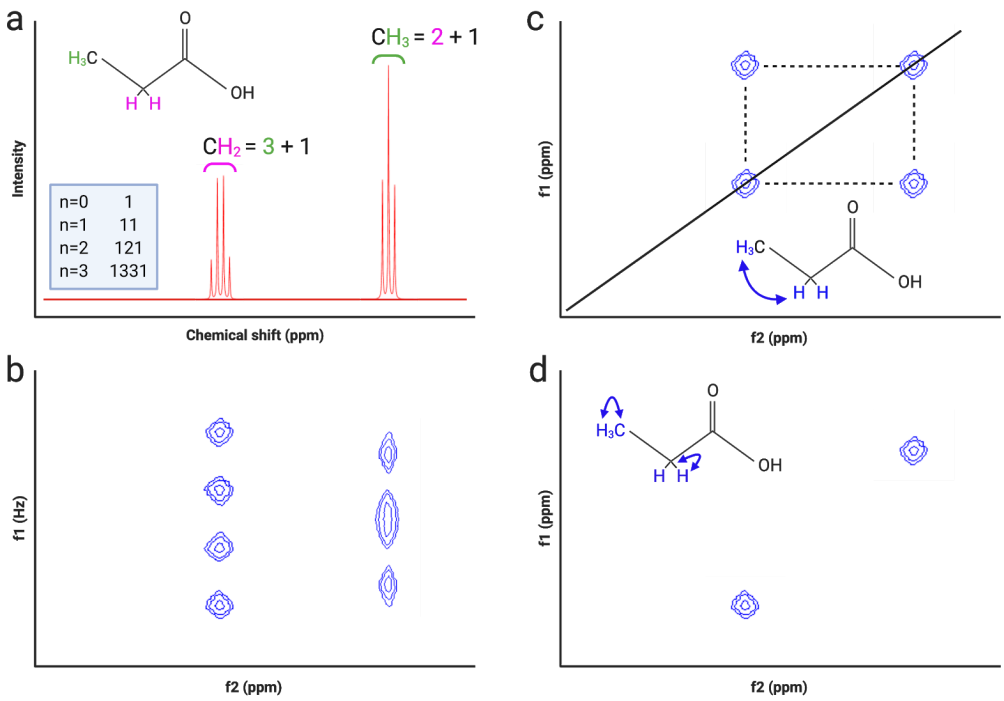


Figure 2.5: NMR experiments. a) 1D proton NMR experiments detail resonances which can be split depending on the number of neighbouring protons (n). The ratio of the peak intensities follows Pascal's triangle. **b)** The splitting pattern, or multiplicity, of each resonance can be confirmed with J-resolved spectra where proton resonances are plotted on the x-axis (f2) and coupling constants are plotted on the y-axis (f1). **c)** A COSY spectrum (pictured here) will have cross-peaks for protons that are 2-3 bonds apart while TOCSY (not pictured) will show correlations between all protons of a spin system. **d)** HSQC experiments give the chemical shift for the protons (f2) as well as the carbon (^{13}C) attached (f1).

2.3 Gut bacterial community profiling

Technological advances in DNA sequencing have substantially enhanced our ability to identify bacterial species living in complex communities. Prior to the advent of next-generation sequencing (NGS), bacteria were studied through culturing, targeted molecular biology methods (e.g. fluorescence *in situ* hybridization (FISH), quantitative polymerase chain reaction (qPCR)) and Sanger sequencing. These approaches, though still valuable in microbiome research, cannot capture the diversity of the gut microbiota[4]. The Sanger method, which only sequences a single fragment of DNA at a time, is low-throughput and has been superseded by methods which sequence millions of DNA fragments in parallel. In bacterial community profiling, these

DNA fragments give information about the taxonomic or genetic composition of the community. In the context of metabolic profiling, this added layer of information is important for teasing apart host metabolism and microbiota-associated metabolism. DNA sequencing can provide a blueprint for the metabolic capacity of the microbiota (directly or inferred from taxonomic information), and metabolic profiling gives evidence for which molecular mechanisms are invoked.

2.3.1 16S rRNA gene amplicon sequencing

To study the composition of the gut microbiota, two main approaches are employed: taxonomic profiling to determine the types of microbes present and metagenomic profiling to determine the genes contained within the community. Although metagenomic sequencing provides far more information about the community such as its functional capacity and metabolic potential, this method is considerably more expensive than taxonomic profiling. In this thesis, the taxonomic approach was used. Taxonomic profiling of microbes is accomplished using a marker gene and variations in its sequence to differentiate distinct bacterial groups. For bacteria, the most commonly used marker gene is the 16S ribosomal RNA (rRNA) gene due to its slow evolutionary rate, presence in bacterial genomes and absence in the human genome[150]. The 16S rRNA gene is 1500 base pairs (bp) and consists of 9 hypervariable regions (**Figure 2.6**). Within the hypervariable regions, genetic sequences differ substantially in disparate bacterial groups. The variable regions are flanked by conserved regions where the sequence is similar across many different types of bacteria. The conserved regions act as primer binding sites for polymerase chain reaction (PCR) and the variable regions are amplified. The amplified sequences act as barcodes for taxonomic identification and phylogenetic analysis. The portion of the 16S rRNA gene that is sequenced is limited by the platform used; each platform will have a limit for the number of base pairs in a DNA fragment that it can accurately read (read length). For this reason, one to three variable regions are typically targeted for bacterial profiling and the number of samples included in a sequencing run is limited.

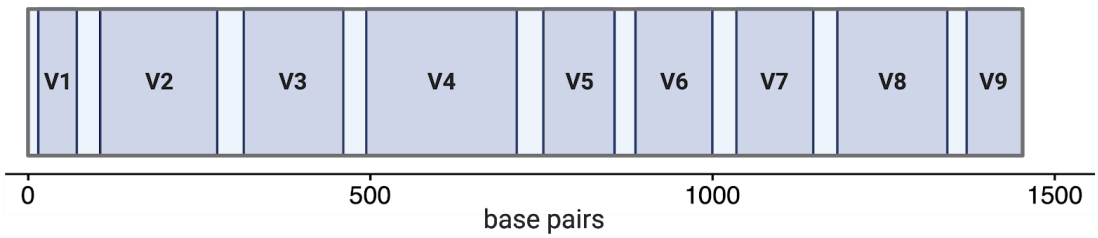


Figure 2.6: 16S rRNA gene. The 16S rRNA gene is around 1500 base pairs long and has 9 regions where sequences are hypervariable among different bacterial groups. Hypervariable regions are flanked by regions where sequences are conserved among many bacteria. Genetic sequences within the hypervariable regions can be used for taxonomic classification and phylogenetic analysis of bacteria. Figure is based on information from Yang et al.[151].

2.3.2 Library preparation

Although bacterial DNA isolated from gastrointestinal tract tissue samples can be analysed to determine the composition of the gut microbiota, faecal samples are typically preferred due to the non-invasive collection, high bacterial burden and low host DNA content. However, it must be noted that the bacterial communities associated with gut tissue are different from those passed in the stool[152]. The data presented in this thesis is based exclusively on faecal samples. DNA was extracted from 250 mg of each sample using the PowerLyzer PowerSoil DNA Isolation Kit (Mo Bio, Carlsbad, CA, USA). Samples were homogenised by bead beating in a Bullet Blender Storm (Chembio Ltd., St. Albans, UK) for 3 min at speed 8 to ensure cell lysis. The Qubit fluorometric assay (Thermo Fisher Scientific, Carlsbad, CA, USA) was used to quantify DNA. The V1-V2 region of the 16S rRNA gene was amplified by PCR using a mixture of 4 forward primers at 4:1:1:1 (28F-YM:28F-Borrellia:28FChloroflex:28F-Bifdo) and 1 reverse primer (388R), each with an overhang adapter sequence[153]. Primer sequences are detailed in **Table 2.1**. A second PCR was performed to attach multiplexing indices and Illumina sequencing adapters to overhang adapters on amplicon sequences (**Figure 2.7**). Multiplexing indices add unique barcodes to amplicons in each sample to enable massively parallel sequencing; each sequence can be traced back to the original sample. SequalPrep Normalization Plate Kits (Life Technologies, Paisley, UK) were used to clean and normalise PCR reactions which were then pooled into libraries. Libraries were quantified using the NEBNext Library Quant Kit for Illumina (New England Biolabs, Hitchin, UK) prior to being sequenced.

Table 2.1: 16S rRNA gene V1-V2 primer sequences. Bases in bold are MiSeq overhang adapter sequences.

Primer name	Primer sequence
28F-YM (forward primer)	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GAGTTTGATYMTGGCTCAG
28F-Borrellia (forward primer)	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GAGTTTGATCCTGGCTTAG
28FChloroflex (forward primer)	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GAATTTGATCTTGGTTCAG
28F-Bifdo (forward primer)	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**GGGTTCGATTCTGGCTCAG
388R (reverse primer)	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**TGCTGCCTCCCGTAGGAGT

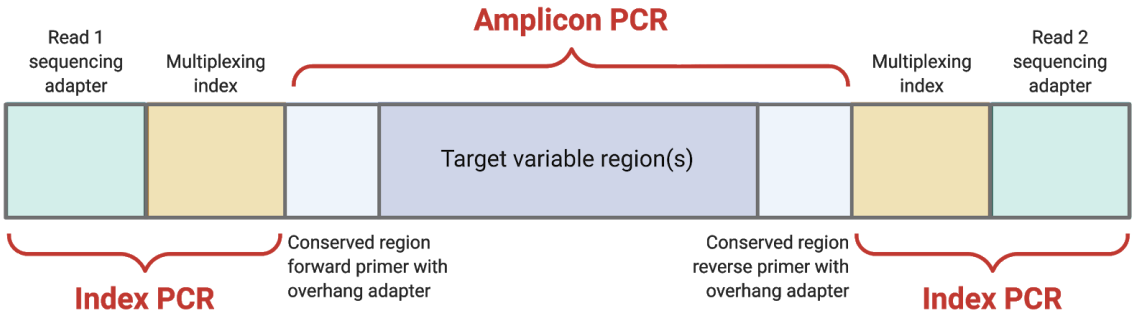


Figure 2.7: 16S rRNA gene amplicons with attached adapters and indices. Amplicon PCR amplifies the targeted variable region(s) with primers which bind to conserved regions flanking the region of interest. 16S primers contain overhang adapters which allow binding of the multiplex indices and sequencing adapters in a second PCR (index PCR).

2.3.3 Data acquisition

Amplicon libraries were sequenced on the Illumina MiSeq platform (San Diego, CA, USA) using the MiSeq Reagent Kit v3 (paired-end 300 bp). Illumina uses clonal amplification and sequencing by synthesis chemistry[154] that takes place on a glass slide called a flow cell. The flow cell contains oligonucleotides that are complementary to the sequences ligated to the ends of the amplicons (**Figure 2.7**), enabling hybridization of the amplicons to the flow cell. A cluster of clonal sequences is generated for each sequence hybridised to the flow cell. After cluster generation, sequencing of read 1 (forward) takes place. A complement to each clonal fragment in each cluster is synthesised by incorporating fluorescently labelled nucleotides. After each base is incorporated, the flow cell is imaged and the emission wavelength and intensity from each cluster are used to identify the base (**Figure 2.8**). In this case, the cycle was repeated 300 times to get a read length of 300 bp for each amplicon. For paired-end sequencing, a second 300 bp read was then generated for the reverse direction (read 2).

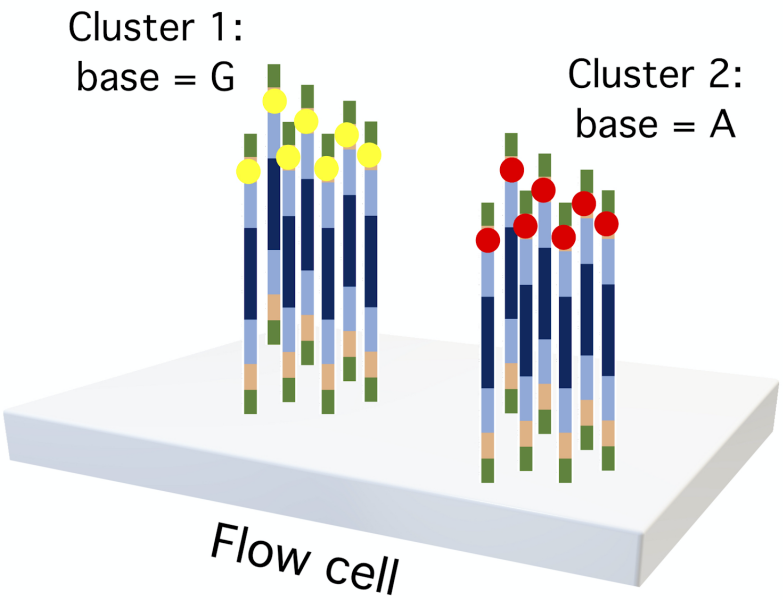


Figure 2.8: Illumina sequencing by synthesis chemistry. DNA amplicons are hybridised on the flow cell and clonal clusters are generated for each amplicon. For each clonal sequence, fluorescently labelled nucleotides are used to synthesise a complement strand. After each nucleotide is incorporated, the emission wavelength and intensity of the cluster is used to determine the base. The number of cycles (nucleotides incorporated and imaged) dictates the length of the sequenced read.

2.3.4 Data pre-processing

Fastq sequence files from pooled libraries were demultiplexed (sorted into separate files based on the multiplexing indices) and primer sequences were removed using cutadapt[155]. Raw reads were processed in the R (version 3.5.0) statistical environment[156] following a published workflow[157] which includes amplicon denoising. Prior to amplicon denoising, reads were trimmed (forward trimmed to 250 bp, reverse trimmed to 175 bp) to remove low quality base pairs at the 3' ends of the reads. The poorer quality of reverse reads is a known property of Illumina sequencing data[158]. The denoising algorithm, "DADA2"[159], utilises error profiles (based on quality scores for base calls) generated during the sequencing run to correct sequencing errors. Traditionally, sequencing errors were ignored, sequence variants were clustered by similarity (usually 97%) and the resulting clusters, called operational taxonomic units (OTUs), were assigned a taxonomic label. This approach diminishes true, biologically-relevant variation

contained within the communities since many distinct bacterial species have very similar (>98.7%) 16S rRNA gene sequences[150]. Amplicon sequence error correction permits the analysis of unique sequence variants instead of OTUs, providing fine-scale resolution of true variation within the bacterial communities[159]. After amplicon denoising, paired forward and reverse reads were merged with a minimum overlap of 50 bp. Reads without an exact matching pair were removed. Chimeric reads, those composed of two different parent reads, were also removed. For each sequence, taxonomy was assigned with reference to the RDP[160] database and assignments of statistically significant taxa were confirmed with taxonomic assignments using the SILVA[161] database. To construct a phylogenetic tree for comparing the relatedness of the sequences, the R package “phangorn”[162] was used. The sequences were aligned, Hamming distances were calculated and the tree was built using the neighbour-joining method.

2.3.5 Normalisation and diversity estimates

As previously discussed, the number of reads obtained for each sample (sequencing depth) is a function of the sequencing platform used and can vary significantly between samples, especially when multiple sequencing runs were required. The sequencing depth of a sample will affect bacterial diversity estimates as the number of distinct bacteria detected will increase with the number of reads obtained. To bypass this artefact, sequencing depth is normalised across samples but the method of normalisation depends on the downstream analysis. To estimate alpha-diversity, the numbers and proportions of distinct bacterial taxa present within a sample, rarefaction was applied. Rarefaction normalises sequencing depth by randomly subsampling a set number of reads from each sample, usually the minimum sequencing depth in the dataset. This approach is highly criticised due to the volume of data that is ignored[163]. However, in microbial ecology, it remains the standard method for normalising prior to estimating alpha-diversity.

For calculating beta-diversity, (dis)similarity measurements between samples, total sum scaling (relative abundance) or log transformations can be used[157]. Total sum scaling normalises each sample by its sequencing depth, converting counts to proportions, while log transformations make the distributions of the abundance data less skewed or more normal. Although these methods remove some unwanted technical variation from the data, they ignore the compositional properties of count data and may introduce artefacts[164]. The centred log-ratio

transformation[165] was applied to microbiome count data prior to multivariate modelling to account for compositionality. This transformation normalises each sample ($x = [x_1, x_2, \ldots x_D]$) based on the geometric mean ($G(x)$) of the counts in that sample. The geometric mean indicates the central tendency of the counts in a sample, effectively accounting for different ranges of numbers across samples.

$$G(x) = \sqrt[D]{x_1 \times x_2 \times \ldots x_D}$$

$$x_{clr} = [log(\frac{x_1}{G(x)}), log(\frac{x_2}{G(x)}), \ldots log(\frac{x_D}{G(x)})]$$

2.4 Statistics

The profiling methods employed in omics sciences generate large datasets for which both multivariate and univariate statistical analyses are used to uncover patterns and draw biological conclusions. Multivariate methods are suitable for high-dimensional data (i.e. where variables > observations) especially when the data are multicollinear (i.e. when there is strong correlation between measured variables). Using multivariate methods, relationships between many data features can be summarised. On the other hand, univariate methods summarise information about individual features in the data. The basic concepts of the statistical methods applied in this thesis are described below while specific descriptions of the methods used are detailed in the subsequent chapters. All calculations were performed in the R (version 3.5.0) statistical environment[156].

2.4.1 Linear dimensionality reduction techniques

Multivariate methods such as projection-based linear dimensionality reduction techniques are useful for visualising overall structure in a dataset, summarising relationships between variables and selecting individual variables with biological relevance to the research question. These methods can be unsupervised (models are agnostic to sample classification or group labels) or supervised (models are trained to predict specific outcomes). While unsupervised methods are used to summarise global variation in a dataset, supervised methods are employed to identify variation that is specific to a group within the dataset or related to a covariate. Prior to multivariate modelling, autoscaling was applied to mean-centre the data and transform it to unit-variance. Autoscaling standardises the values for each variable (mean=0, variance=1),

ensuring each variable is considered equally in the model regardless of differences in magnitude.

2.4.1.1 Principal component analysis

Principal component analysis[166] (PCA) is an unsupervised technique which summarises variance in a dataset using dimensionality reduction. Dimensions are reduced by summarising similar information into a minimal number of new variables calculated through matrix decomposition (**Figure 2.9**). These new latent variables are the principal components; a component will capture a subset of variation in the dataset with each successive component capturing variation that is orthogonal (uncorrelated) to the previous. The number of components calculated for a PCA model is determined for each dataset but is dependent on the percentage of variance explained by each component. The goal is to maximise the explained variance in as few latent variables as possible. Each component has two associated vectors: scores and loadings. Every observation (or individual in this case) from the original data matrix will have a score for each component. The scores are linear combinations of all original variables which depend on the loadings. Loadings are weights for each variable which indicate their influence in the subset of variation explained for a specific component. Individuals with similar scores are more similar (similar values of the original variables) to each other than individuals with less similar scores, enabling the detection of global patterns in the data. PCA is also useful for identifying outliers in a dataset; these samples usually contain abnormal values for some variable(s) making them incomparable with the rest of the samples. Commonly, such samples are contaminated with exogenous compounds (e.g. drug metabolites) and might need to be excluded from further analyses since linear multivariate methods are sensitive to outliers.

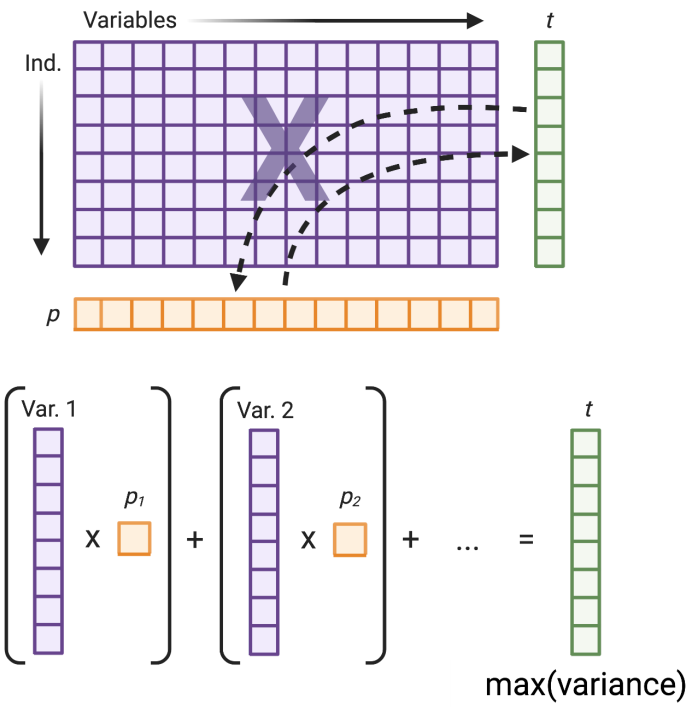


Figure 2.9: Principal component analysis. Latent variables (components) are calculated from the original data matrix (*X*). Scores (*t*) and loadings (*p*) are calculated through matrix decomposition in such a way that the first component will have maximum variance. *Ind individual; Var variable.*

An extension of the classic PCA algorithm, sparse PCA[167] (sPCA), can be used for unsupervised feature selection. As in PCA, dimensionality is reduced by linearly combining the original variables into latent variables, but sparsity is introduced such that each component is composed of a subset of variables, the important features. This is accomplished by penalising (weight=0) variables with low variation across samples thus removing non-informative features.

2.4.1.2 Partial least squares regression

The fundamental concept behind partial least squares (PLS) regression[168], also called projections to latent structures[169], is similar to PCA. The goal is to reduce dimensionality by calculating latent variables through matrix decomposition. However, PLS models the covariance structure between two data matrices, *X* (predictors) and *Y* (outcomes), and covariance will be maximised in the first component (**Figure 2.10**). Although the initial use of PLS was mainly to predict *Y* from *X*, variants of the algorithm have been developed in which *X* and *Y* play a

symmetric role in the model and the analysis is similar to canonical correlation[170]. The method is referred to as PLS discriminant analysis (PLS-DA) when the goal is to classify observations into groups therefore the outcome is categorical. As with PCA, sparse versions of PLS (sPLS and sPLS-DA) exist[171] where penalisation is introduced on the pair of loading vectors so each component is composed of a subset of variables and/or outcomes. sPLS methods are superior to non-sparse methods when the goal of the analysis is feature selection[171]. Another variant of PLS that is commonly used to model metabolomics data is orthogonal PLS (OPLS)[172]. Here, a single predictive component is calculated and subsequent components model information that is uncorrelated with (orthogonal to) *Y*. Although the predictive performance is not changed with the orthogonal signal correction, interpretability of the model may be improved since informative and uninformative variation are separated[172].

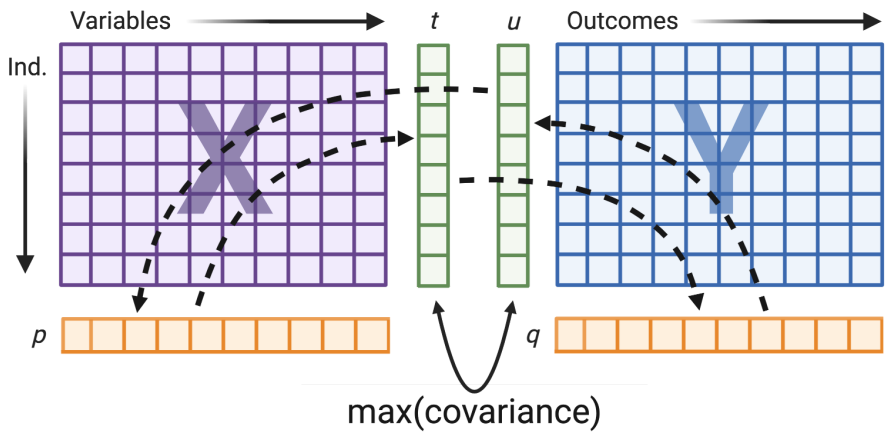


Figure 2.10: Partial least squares regression. Relationships between two data matrices, *X* and *Y*, are modelled to maximise the covariance between latent variables. This is accomplished through matrix decomposition to calculate scores (*t* and *u*) and loadings (*p* and *q*) vectors which summarise similar information from *X* and *Y*. *Ind individual.*

PLS can also be used to model relationships between more than two datasets using multi-block methods[173]. Complementary information is extracted from related datasets (i.e. multiple assays conducted on the same samples or multiple sample types collected from the same individuals) and modelled in a supervised or unsupervised manner, as a discriminant or regression analysis and with or without sparsity. In multi-block PLS methods, there are two levels of analysis (**Figure 2.11**). At the sub level, PLS components are calculated for each block (X_1, X_2 … X_n)

with respect to *Y* and scores ($t_1, t_2 \ldots t_n$) are combined in a super block (*T*). At the super level, another iteration of PLS is performed between *T* and *Y*. Covariance is maximised between components of *X* blocks and between components of *X* and *Y*.

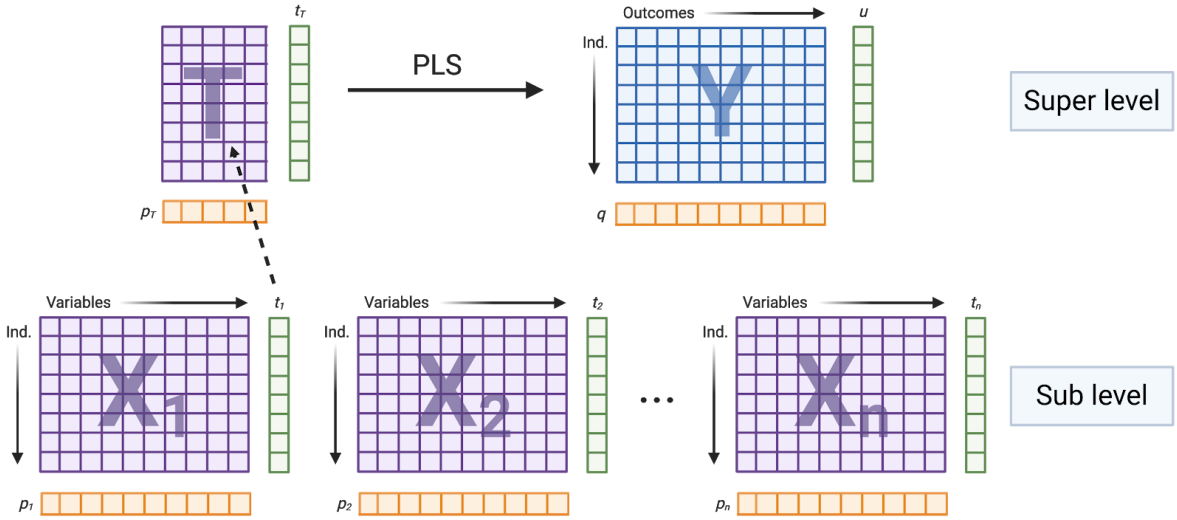


Figure 2.11: Multi-block partial least squares regression. The partial least squares algorithm is applied at the sub level for each block (X_1, $X_2 \ldots X_n$) and *Y* then scores (t_1, $t_2 \ldots t_n$) are combined into a super block (*T*). Another iteration of the algorithm is performed at the super level where covariance between *T* and *Y* is maximised.

2.4.1.3 Cross-validation

For each PLS model, performance was assessed using cross-validation. The cross-validation procedure involves splitting the data into a number of test sets and for each test set, a new model is calculated based on all data not included in the test set (training set). The performance of the new model is evaluated based on its ability to predict correct values for the test data. In other words, a model performs well if the error in prediction is low for the left out samples. Performance was measured based on the following metrics: mean squared error of prediction (MSEP)[174], R^2Y and Q^2Y[175]. MSEP is a measure of the error between the "real" values calculated from a full model (no left out data) and the predicted values (calculated on left out data). R^2Y and Q^2Y are measures of the goodness of fit and predictive ability of the model, respectively. These metrics indicate the fraction of variation in Y that can be explained (R^2Y) and predicted (Q^2Y) by the model. There are no standard acceptable values of R^2Y and Q^2Y, but it has been proposed that Q^2Y values of 0.02, 0.15 and 0.35 can be interpreted as weak,

moderate and strong evidence for predictability, respectively[176], while the R^2Y value should be close to Q^2Y. High disparity between these two values indicates overfitting[177], a case when the model can explain the training data well, but cannot be generalised to new data.

2.4.1.4 Relevance networks

In addition to describing associations between data features and outcomes, PLS models can be used to calculate associations between features to visualise the correlation structure between datasets. This is accomplished by extracting pairwise associations between features across datasets[178]. Correlation between two variables is approximated based on their projections in latent space; the angle between the two vectors indicates their degree of relatedness in the model (**Figure 2.12**). Using this approximation, a correlation matrix can be inferred from the model and a relevance network[179] can be generated; for every correlation coefficient, an edge is drawn between the two variables (nodes). Additionally, a *P* value was calculated for each correlation coefficient through permutation testing.

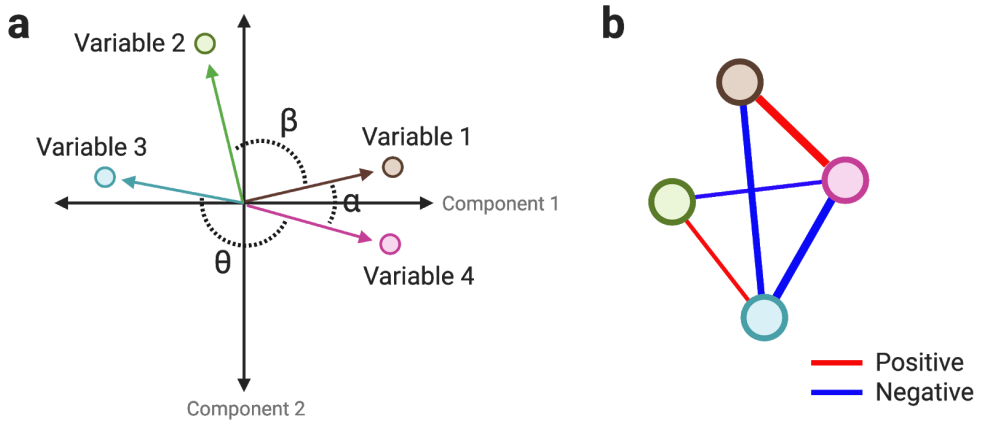


Figure 2.12: Robust approximation of correlation based on PLS loadings. a) Pairwise correlations between features (variables) in different datasets are approximated based on their projections in latent space and the angle between the two vectors. Two vectors with an acute angle ($\cos(\alpha)>0$) indicate a positive correlation while two vectors with an obtuse angle ($\cos(\theta)<0$) indicate a negative correlation. Perpendicular vectors ($\cos(\beta)\approx 0$) indicate no correlation between the two variables. **b)** A relevance network is generated by drawing an edge between two variables for each correlation. Figure (a) adapted from González et al.[178] (Creative Commons Attribution License http://creativecommons.org/licenses/by/2.0).

2.4.2 Random forests

In Chapter 5, the random forests algorithm was used to build a classifier for predicting metabolic phenotypes from microbiome data. Random forests is a simple supervised machine learning method in which a large number of decision trees operate as an ensemble[180]. The individual trees consider a random subset of variables by which split to observations (**Figure 2.13**). By default, a subset of observations (⅔) are used to train the decision trees to identify the correct classes and the model performance is evaluated on the left out (out-of-bag, OOB) samples. For classification, each OOB sample is assigned the class with the most votes, the most popular decision, among the trees. The OOB error rate, a common metric for evaluating a random forests model, is the percentage of OOB samples that are incorrectly classified. Variable importance is ranked based on the mean decrease in accuracy observed when the values of that variable are permuted in OOB samples. A large mean decrease in accuracy indicates the permuted variable is important for the classifier. Once tuned for the appropriate number of trees and the number of candidate variables to try at each split, the model can be used to predict classes of new samples.

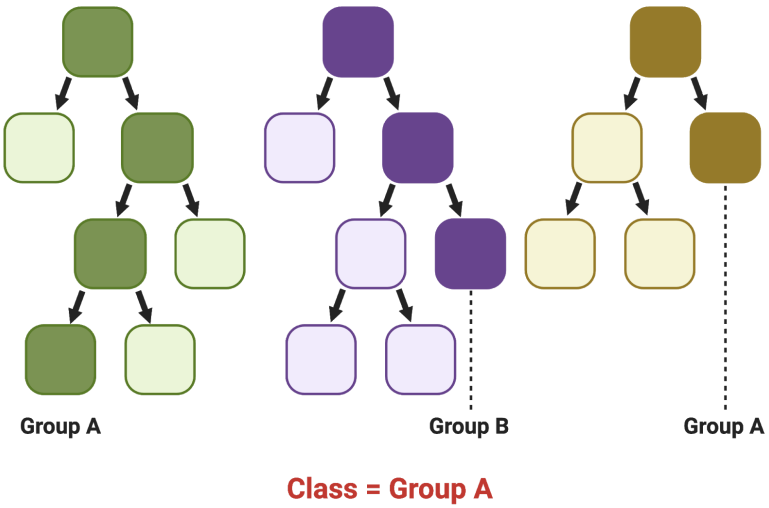


Figure 2.13: Random forests classifier. A large number of decision trees are trained to classify observations based on a subset of variables. Each variable is used to split the observations (yes/no). For a given out-of-bag sample, the class with the most votes among the trees wins.

2.4.3 Linear mixed-effects models

Linear mixed-effects models aim to explain variation in a single variable that is due to both fixed effects and random effects[181]. In the case of clinical data, fixed effects are information collected about the individuals such as gender, age or other covariates that are interesting for making biological interpretations about the data. Random effects are information associated with the random sampling of patients such as the individual (for a repeated measures design) or the location of the clinic where they were enrolled. The goal is to adjust for unwanted variation that may have been introduced due to this variable while identifying subsets of variation that can be attributed to each of the fixed effects. A regression coefficient is calculated for each fixed effect (predictor) which indicates the strength and direction of its association with the dependent variable (response). The significance of each association was evaluated using the Kenward-Roger approximation[182] for denominator degrees of freedom *F* statistics.

2.4.4 Time series analysis

Linear mixed-effects models are suitable for modelling linear relationships, but non-linear methods are necessary to model dynamic relationships such as changes in metabolic phenotypes over time. Here, a spline-fitting method[183] was used to condense the measurements for each individual into a smooth, continuous function of time. The curve then becomes the fundamental unit of analysis meaning trajectories can be compared rather than individual time points. Mean curves are then calculated (e.g. for case and control) and significance can be assessed by permutation testing. A *P* value is derived as the probability that the distance between the group mean curves is greater when group labels are shuffled than when appropriate labels are assigned.

2.4.5 Correlation

For normally distributed data, Pearson's correlation coefficients[184] were calculated while Spearman's rank correlation coefficients[185] were calculated for data with skewed distributions. Spearman's correlation is essentially Pearson's correlation calculated on the ranks of the variables, therefore it is not sensitive to outliers. In some instances, partial Pearson or Spearman correlation coefficients were calculated. A partial correlation describes the linear association between two variables after eliminating the effect of one or more covariates[186]. For

variables with repeated measures (different time points for each individual) where a linear relationship was expected, analysis of covariance was used to calculate correlations between two variables while controlling for inter-individual variability, a method named repeated measures correlation[187]. Parallel regression lines are fit for each individual with different intercepts, allowing multiple measurements from each individual to be modelled without violating the assumption that observations are independent. A correlation coefficient of 0.5-0.7 is described as moderate correlation[188].

2.4.6 Correction for multiple hypothesis testing

Performing a large number of statistical tests increases the risk of rejecting null hypotheses by chance[189]. Throughout this thesis, the rate of type I errors was reduced using the Benjamini-Hochberg procedure[190] which controls the expected proportion of false discoveries, the False Discovery Rate (FDR), by considering the number of tests performed and the ranks of the resulting *P* values. Where multiple hypothesis tests were performed, *P* values were adjusted as follows: *P* values were ranked in ascending order and multiplied by the number of tests performed then each product was divided by the rank of the *P* value. The null hypothesis was then only rejected if the resulting value (FDR) was less than the defined threshold (α=0.05).

Chapter 3

Metabolic perturbations in obese women during pregnancy

During gestation, substantial metabolic adaptations occur to support the needs of the developing fetoplacental unit. This chapter details obesity-associated maternal metabolism over the course of pregnancy as well as relationships between altered maternal metabolism and both maternal and neonatal outcomes. Clinical samples were collected from a UK cohort of women at 6 time points, and from the neonates of these women at delivery. Proton nuclear magnetic resonance (^{1}H NMR) spectroscopy was used to profile longitudinal changes in circulating (blood serum) and excreted (urine) metabolites. Obese, compared to normal weight, pregnancies were characterised by changes in amino acid metabolism including branched-chain amino acids (BCAAs), decreased concentrations of short-chain fatty acids (SCFAs), changes in energy metabolism and elevated inflammation. BMI-associated metabolites in maternal serum and urine throughout pregnancy were highly correlated with maternal insulin resistance while the circulating metabolites transferred to the fetus, measured in umbilical cord serum, were correlated with umbilical cord insulin and birth weight. The present data suggest that elevated maternal insulin resistance related to high BMI during pregnancy increases the rate of transfer of selected serum metabolites from maternal circulation to the fetus, and that these metabolites influence fetal insulin secretion and growth.

3.1 Introduction

Maternal obesity is a major risk factor for adverse obstetric outcomes and long-term cardiometabolic complications in offspring. Studies have shown that high pre-pregnancy body mass index (BMI) is associated with increased risk of miscarriage, preterm birth, development of gestational diabetes mellitus (GDM), hypertensive disorders and caesarean delivery[191–193]. Obese mothers are also more likely to give birth to large for gestational age neonates with higher fat mass percentages[193,194] and higher insulin resistance[194] than non-obese mothers.

Dietary and lifestyle interventions have not been shown to adequately reduce maternal complications and fetal overgrowth[195], suggesting additional mechanisms are at play. Obesity has been associated with changes in gut microbial community structure[95,196], but there is a lack of consensus about which, if any, bacterial taxa change with BMI[90,93,94,197] and whether any observed changes are instead associated with dietary differences[92]. Similarly, the role of gut bacteria during pregnancy is not completely understood but it is conceivable, given that the microbiome is highly metabolically active, that maternal gut microbes play an important role in fetal development and may contribute to conferred chronic health risks[75].

To date, few studies have focused on BMI-related variation in the gut microbiota during gestation and its effects on pregnancy outcomes. These reports show that abundances of some bacterial taxa vary in pregnant women with a high BMI[73,198,199], although the affected taxa are not consistent between studies. However, these studies show that BMI-associated alterations in gut bacterial composition are related to excessive gestational weight gain[73,198,200], maternal cholesterol and folic acid levels[198], and high birth weight[198]. Regardless of differences in community compositions, metabolomics studies suggest gut microbial activity is altered in individuals with a high BMI[99,101]. Concentrations of SCFAs, major fermentation products of the gut microbiota, are altered in the faeces and blood of non-pregnant obese adults[95–98] and obese children[201,202]. Maternal circulating SCFAs in the late third trimester of pregnancy have been shown to correlate with maternal leptin and adiponectin levels, gestational weight gain and neonatal body weight[203]. Increased excretion of SCFA acetate and decreased excretion of hippurate, an important host-microbial co-metabolite, have also been reported in obese pregnancies[204]. These findings highlight the importance of looking beyond gut microbial community composition and investigating functional differences in bacterial communities of obese individuals, particularly during pregnancy when fetal exposure to maternal metabolites is critical to its health[67,205].

In healthy pregnancies, considerable physiological adaptations occur affecting cardiovascular, renal, respiratory, immune and metabolic function[70,71]. Some of these adaptations actually mimic metabolic syndrome in non-pregnant individuals[206,207]. A progressive rise in insulin resistance[208] and systemic inflammation[209] occur during normal pregnancy and this is enhanced in overweight or obese women[199,210,211]. Elevated inflammatory markers during pregnancy are associated with changes in the gut microbiota[72,199] and transplantation of the third, but not first, trimester

maternal gut microbiome induces symptoms of metabolic syndrome in germ-free mice[72]. Biomarkers of inflammation in overweight or obese pregnant women have also been associated with increased BCAA levels and insulin resistance[210]. BCAAs are essential amino acids acquired through the diet and synthesised by bacteria[212,213]. These metabolites are strong, consistent biomarkers of obesity-related insulin resistance[104–106,210] that are independent of protein intake[106] but are related to variations in the gut microbiome[105]. In addition to maternal pregnancy outcomes, the maternal metabolic profile also predicts fetal and neonatal outcomes[214–217], highlighting the importance of understanding maternal pathophysiological mechanisms to improve obstetric outcomes in obese women.

Few studies have employed untargeted metabolic profiling in detailed longitudinal studies of pregnancy[218–221] and even fewer have focused on obese pregnancies[222]. Data from the current literature is limited by insufficient numbers of time points (2-3) and the use of linear methods to analyse the highly dynamic remodelling of metabolism during pregnancy. Here, both blood and urine samples collected during 6 different hospital visits provide multiple views of host and host-microbe metabolism as well as a sufficient number of sampling points to perform non-linear time series analysis. In conjunction with the extensive clinical characterisation of the present cohort and their neonates plus the neonatal biofluid samples, the current study is unique and powerful. Future follow-up of these offspring will allow unprecedented analysis of the effects of maternal obesity-associated perturbations in metabolism on fetal programming and future health risk.

3.2 Materials and methods

3.2.1 Study population, sampling and clinical data

The population, a subset of patients from the Chelsea & Westminster Hospital (CWH) study, along with clinical sampling and metadata collection is detailed in **2.1.1**. For this subpopulation (dataset 1), the aim was to compare normal weight versus obese pregnancies, therefore the following exclusion criteria were applied: previous bariatric surgery, miscarriage, still birth, Type 2 diabetes mellitus and withdrawal from the study.

Participants were classified as normal weight (NW) controls with BMI <25 kg/m^2, overweight (OW) with BMI 25-29.9 kg/m^2 or obese (OB) with BMI ≥30 kg/m^2. Hypothesis tests were performed for each case group (OW or OB) with NW as the reference group. Groupwise differences in clinical variables with numerical values were evaluated with Student's t-test or Wilcoxon rank sum (Mann-Whitney) test for normally and non-normally distributed data, respectively. One-tailed tests were performed for variables related to BMI, insulin resistance and neonatal size since these values are expected to be greater in the case groups. Chi-squared tests were used for categorical variables unless any observations had fewer than 5 values, then a Fisher's exact test was used. Associations between maternal and neonatal outcomes were also assessed with correlation analyses. For variables with multiple time point measurements, repeated measures correlation (package "rmcorr"[223]) was used. Partial correlation (Spearman, package "ppcor"[187]) was used for variables with a single measurement; coefficients were adjusted for maternal age. False discovery rates (*FDR*) were calculated with the Benjamini-Hochberg[190] procedure and an alpha of 0.05 was used throughout. All calculations were executed in the R (version 3.5.0) statistical environment[156]. A list of packages and package versions used can be found in **Appendix C**.

3.2.2 Metabolic profiling of biofluids

Sample Preparation. Serum and urine samples were prepared according to a published protocol[139]. For serum, 350 µL of sample was mixed with 350 µL serum buffer (0.075 M NaH_2PO_4, 5 mM TSP, 2 mM NaN_3, 20% D_2O, pH 7.4). For urine, 540 µL of sample was mixed with 60 µL urine buffer (1.5 M KH_2PO_4, 6 mM TSP, 2 mM NaN_3, D_2O, pH 7.4). Samples were then centrifuged for 5 min at 12000g and 4°C. Supernatant was transferred to 5 mm SampleJet tubes (Bruker Biospin Ltd. Part No. Z105684); tube caps were sealed with POM balls (Bruker Biospin Ltd. Part No. Z72497).

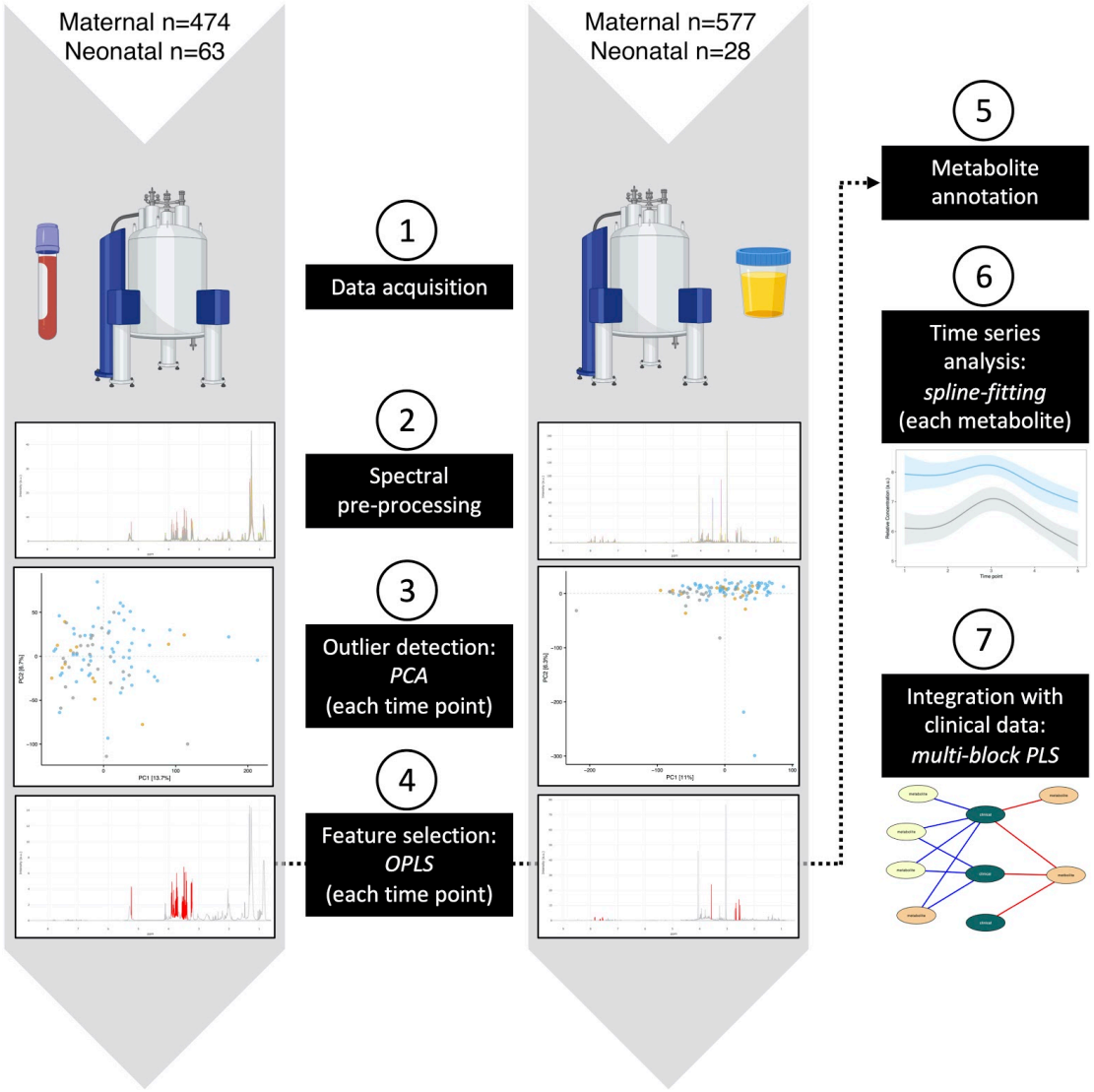


Figure 3.1: Overview of analytical workflow for metabolic profiling. Data were acquired from prepared urine and serum samples by ^{1}H NMR spectroscopy **(1)** then spectra were pre-processed **(2)**. Outlying spectra were detected at each time point by PCA **(3)** and those with contaminants or drug metabolites were removed. Features associated with BMI were selected at each time point by OPLS **(4)** and significant features were annotated **(5)**. Each semi-quantified metabolite and each quantified (B.I.QUANT) metabolite was modelled longitudinally **(6)** and those significantly increased or decreased in obese women throughout the time course were included in multi-block PLS models to integrate metabolite (serum and urine) concentrations with clinical variables **(7)**. *PCA principal component analysis; (O)PLS (orthogonal) partial least squares.*

Acquisition and pre-processing of spectral data. ^{1}H NMR spectra were acquired on a Bruker 600 MHz spectrometer (Bruker BioSpin) following a published method[139] and using Bruker's Avance® In Vitro Diagnostics research (IVDr) standard operating procedures. Experiments for serum samples were run at 310 K using a Carr-PurcellMeiboom-Gill (CPMG) spin-echo pulse sequence with water presaturation (Bruker pulse program: cpmgpr1d). Experiments for urine samples were run at 300 K using a standard one-dimensional spectroscopy pulse sequence with water presaturation (noesygppr1d). J-resolved experiments (jresgpprqf) were also run for all samples to increase capacity for structural identification of molecules. Raw spectra were pre-processed as described in **2.2.4**. Quantification modules of Bruker's Avance® IVDr platform were also run on the acquired serum (B.I.QUANT-PS™) and urine (B.I.QUANT-UR™, normalised to creatinine) data to obtain concentrations of 91 molecules for each sample. These quantified metabolites were used in longitudinal modelling and integration with clinical data only if the metabolite was detected in more than 40% of samples.

Multivariate modelling and feature selection. First, principal component analysis (PCA; package "mixOmics"[170]) was used to identify extreme outliers. Each outlying spectrum was inspected and those with high concentrations of glucose or drug metabolites were excluded prior to supervised modelling. Orthogonal partial least squares (OPLS; package "ropls") regression was performed with one predictive component for each biofluid and each time point separately to identify spectral variables associated with BMI. Models were tuned using 7-fold cross-validation to determine the optimal number of orthogonal components by comparing R^2Y and Q^2Y metrics (see **2.4.1.3**). Variables which had a correlation coefficient (absolute value) of 0.5 or greater with the predictive component and *FDR* <0.05 were annotated.

Longitudinal analysis. For each BMI-associated metabolite annotated in the cross-sectional analysis, a representative peak was selected and integrated to calculate relative concentration. For each semi-quantified metabolite and each quantified metabolite, longitudinal group differences between NW and OB women or between OB women without GDM and OB women with GDM were assessed using the spline-fitting approach implemented in the R package "santaR" (see **2.4.4**). Participants with less than four time points were excluded from time series analysis. *P* values were calculated based on 10,000 permutations of the model and distances between group mean curves. Metabolites with significantly different concentrations between the groups throughout the time series and after correction for multiple hypothesis tests were carried

forward for association to clinical outcomes. Since umbilical cord samples represented a single time point, group differences in metabolite concentrations were assessed using the Wilcoxon rank sum test (rather than time series analysis) to determine if the maternal obesity-associated metabolic profile was reflected in neonatal samples.

3.2.3 Integration of metabolite data with clinical outcomes

Multi-block partial least squares (PLS) regression (see **2.4.1.2**), as implemented in "mixOmics", was used to integrate maternal metabolite data with maternal and neonatal clinical measurements. Separate models were calculated for serum and urine metabolites, with each *X* matrix containing data from a single time point while the *Y* matrix contained outcome variables. Correlation networks were calculated as described in **2.4.1.4** and correlations with *FDR* <0.05 were plotted using Cytoscape[224] (version 3.7.0). A single-block PLS model was also calculated for umbilical cord serum metabolites to determine relationships between the cord metabolome (*X*) and neonatal outcomes (*Y*). Here, metabolites with significant group differences in the maternal serum and metabolites with significant group differences in the cord serum were included in the model.

3.3 Results and discussion

3.3.1 Characteristics of the study participants

Of the 178 women enrolled in the study, the following were excluded from the present analysis: women with a previous bariatric surgery (n=51), those who had a miscarriage or still birth (n=5), Type 2 diabetics (n=2) and withdrawals (n=1). The final study population consisted of 119 women who were classified as NW controls (n=31), OW (n=17) or OB (n=71) based on BMI at T1 (12-14 weeks gestation). Although pre-pregnancy BMI was preferable, this data was not available and given that only 2 kg are usually gained in the first trimester[225], early pregnancy BMI was deemed suitable. The OB group was composed of fewer Caucasian individuals, was younger and more likely to develop GDM than the control group (**Table 3.1**). Maternal insulin resistance was higher in both OW and OB groups than in NW, but HbA1c was higher only in the OB group. As expected, OW and OB mothers had a higher prevalence of large for gestational age (birth weight >90th percentile) neonates than controls, but this was only statistically

significant in the OB group. Nevertheless, OW and OB women gave birth to significantly heavier babies than NW women. Neonatal fat mass percentage and umbilical cord insulin levels were both higher in the OB group compared to controls.

Maternal and neonatal biophysical and biochemical characteristics in the present cohort were consistent with current medical knowledge of the effects of obesity on pregnancy. Obese women have higher insulin resistance during pregnancy and have larger babies with greater adiposity compared to women with a normal pre-pregnancy BMI[226]. One study has also shown that neonates from obese mothers are more likely to be born with insulin resistance[194]. Neonatal insulin levels in the present cohort were higher in the OB group compared to the NW group, but the difference in insulin resistance was not statistically significant (**Table 3.1**). Gestational age at delivery was similar between the study groups despite reports that overweight and obese women have a higher risk of delivering preterm compared to women with a healthy weight[227]. The present cohort also did not show statistically significant evidence of obese women having a 2-3 times higher risk of caesarean delivery compared to non-obese women as was previously reported in a large meta-analysis[228]. However, the prevalence of caesarean delivery was substantially higher in the OB group with almost half of the women delivering non-vaginally. Birth by caesarean section has long-term implications for the health of the offspring[229,230] and has been shown to impact the development of the infant's gut microbiota during the first months of life[231], although it is unclear if this effect is more related to factors associated with caesarean delivery such as antibiotic usage[232].

Table 3.1: Participant characteristics. Values reported are mean ± standard deviation. Statistical tests were performed with normal weight as the reference group.

	Normal weight (n=31)	**Overweight (n=17)**		**Obese (n=71)**	
BMI at T1 (kg/m2)	21.4 ± 1.7	27.1 ± 1.6		37.5 ± 7	
Maternal			*P* value		*P* value
Ethnicity	90.3% Caucasian	88.2% Caucasian	1	69% Caucasian	**0.040**
Age	32.5 ± 3.6	33.1 ± 4.3	0.516	29.2 ± 5.1	**1.48E-03**
Maternal fasting insulin (mU/mL)	5.9 ± 2.1	8.1 ± 2.8	**3.29E-03**	14.2 ± 6.2	**4.88E-12**
Maternal fasting glucose (mmol/L)	4.3 ± 0.4	4.4 ± 0.6	0.150	4.6 ± 0.7	**1.96E-03**
Maternal HOMA-IR	1.1 ± 0.4	1.6 ± 0.7	**4.00E-03**	3 ± 1.6	**1.93E-11**

Maternal HbA1c (mmol/mol)	30.1 ± 2.5	30.5 ± 3.3	0.230	32 ± 4.2	**4.14E-03**
Gestational diabetes mellitus	0% Yes	5.9% Yes	0.354	12.7% Yes	**0.033**
Neonatal					
Gestational age at delivery (weeks)	39.8 ± 1.6	39.6 ± 1	0.745	39.2 ± 1.6	0.093
Delivery mode	29% C-section	35.3% C-section	0.903	49.3% C-section	0.092
Sex of baby	48.4% Male	29.4% Male	0.332	56.3% Male	0.501
Birth weight (g)	3301 ± 533.2	3623.9 ± 483.5	**0.020**	3458.1 ± 612.5	0.099
Birth weight percentile	42.2 ± 30.1	65.5 ± 27.2	**4.78E-03**	59 ± 34.3	**7.92E-03**
Large for gestational age	9.7% Yes	29.4% Yes	0.091	28.2% Yes	**0.027**
Neonatal fat mass (g)	601.7 ± 215.9	689.5 ± 168.6	0.054	680.8 ± 210.5	**0.021**
Neonatal fat mass (%)	17.9 ± 4.7	18.7 ± 3	0.146	19.4 ± 4.3	**0.028**
Umbilical cord insulin (mU/mL)	5.7 ± 4.9	7.6 ± 5.9	0.086	10.7 ± 20.8	**0.035**
Umbilical cord glucose (mmol/L)	4.8 ± 1.5	4.5 ± 1.5	0.779	4.2 ± 1.1	0.963
Umbilical cord HOMA-IR	1.3 ± 1.4	1.6 ± 1.5	0.183	2.3 ± 5.3	0.116

BMI body mass index; HOMA-IR homeostatic model of insulin resistance; HbA1c glycated haemoglobin; T1 12-14 weeks gestation

3.3.2 Associations between maternal and neonatal outcomes

As expected, fetal growth during gestation exhibited a strong positive correlation with maternal weight and BMI, but less so with waist-to-hip ratio and blood pressure (**Figure 3.2a**). Waist-to-hip ratio is a poor indicator of visceral adiposity[233] and fetal growth is proposed to be directly related to maternal adiposity[234]. Although gestational age at delivery was similar between the study groups (**Table 3.1**) and not associated with maternal BMI (**Figure 3.2b**), it had an inverse correlation with maternal insulin resistance indicating that previous reports of early delivery in obese women[227] may be related to obesity-associated elevation in insulin resistance. Umbilical cord insulin levels were significantly correlated with maternal BMI and maternal insulin resistance while neonatal insulin resistance was related to maternal fasting glucose. This result is consistent with “Pedersen’s hypothesis”: maternal hyperglycaemia and increased transplacental transfer of glucose causes fetal hyperglycaemia which stimulates insulin secretion by fetal beta cells and leads to macrosomia, or higher than average birth weight[235]. Fetal insulin production also mediates fat deposition[236].

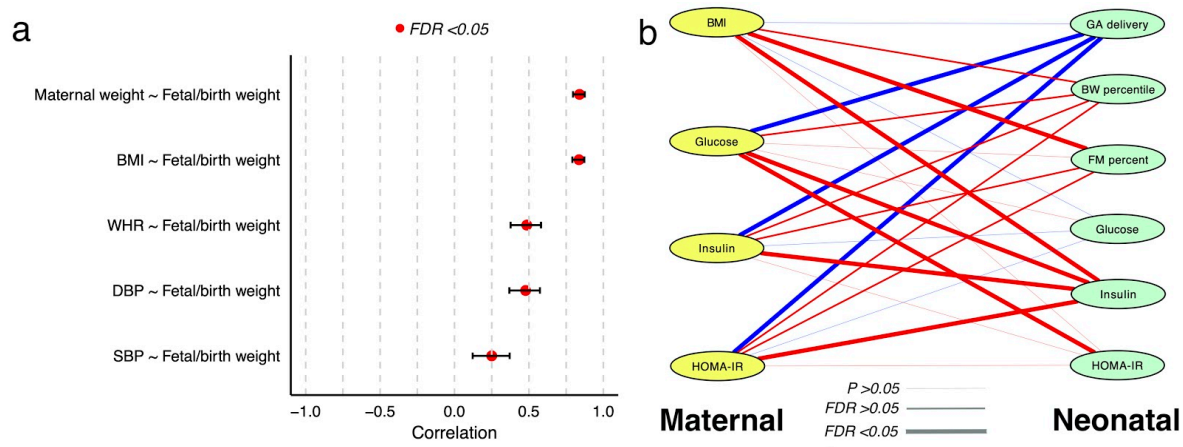


Figure 3.2: Associations between maternal and neonatal outcomes. a) Repeated measures correlation coefficients are plotted with the colour representing statistical significance and error bars denoting 95% confidence intervals. **b)** Partial Spearman's correlation coefficients (adjusted for maternal age) are plotted with red and blue lines representing positive and negative correlations, respectively. The line widths indicate statistical significance. *BMI body mass index; WHR waist-to-hip ratio; DBP diastolic blood pressure; SBP systolic blood pressure; HOMA-IR homeostatic model assessment for insulin resistance; GA gestational age; BW birth weight; FM fat mass*.

The lack of correlation between maternal measures and neonatal glucose is likely due to fetal uptake and utilisation of glucose as it is the primary energy substrate for the fetus[237]. Neonatal fat mass was mostly related to maternal BMI while birth weight percentile was positively associated with maternal BMI and maternal insulin resistance, but these associations were not significant after correcting for FDR. Although maternal measures showed weak associations with neonatal birth weight and fat mass (**Figure 3.2b**), these growth measures were highly correlated with neonatal insulin (**Figure 3.3**), again supporting Pedersen's hypothesis. The inverse association between neonatal insulin and gestational age at delivery (**Figure 3.3c**) has been reported in preterm births and persists in early childhood[238]. Together these results detail maternal obesity-induced fetal hyperinsulinaemia, the associated fetal overgrowth and potential influence on length of gestation. Birth weight in particular is thought to be a determinant of future obesity and metabolic syndrome[239,240], which underscores the importance of defining relationships between maternal insulin resistance and fetal development.

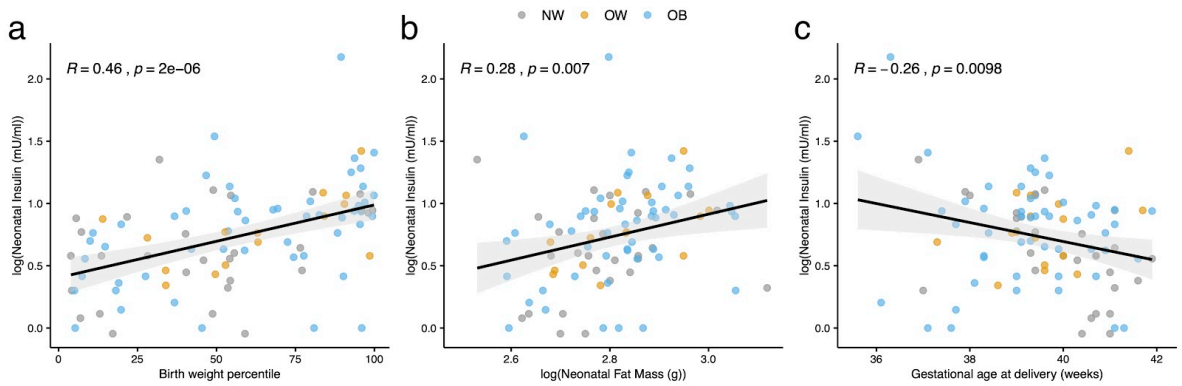


Figure 3.3: Neonatal size and gestational age are associated with neonatal insulin levels at birth. Neonatal insulin measurements compared to **a)** birth weight percentile, **b)** neonatal fat mass and **c)** gestational age at delivery. Pearson correlation coefficients are displayed as well as linear regression lines (black) and 95% confidence intervals (light grey). Points are coloured by maternal weight class. *NW normal weight; OW overweight; OB obese.*

3.3.3 Sampling summary

From the 119 women included in the present analysis, the following samples were studied: 474 maternal serum, 577 maternal urine, 63 cord serum and 28 neonatal urine (**Figure 3.4**). Due to the ongoing nature of the study, many women (particularly in the OB group) had not completed the study at the time these samples were analysed. These patients were included in the analysis; their data were used in cross-sectional feature selection but excluded from longitudinal analyses.

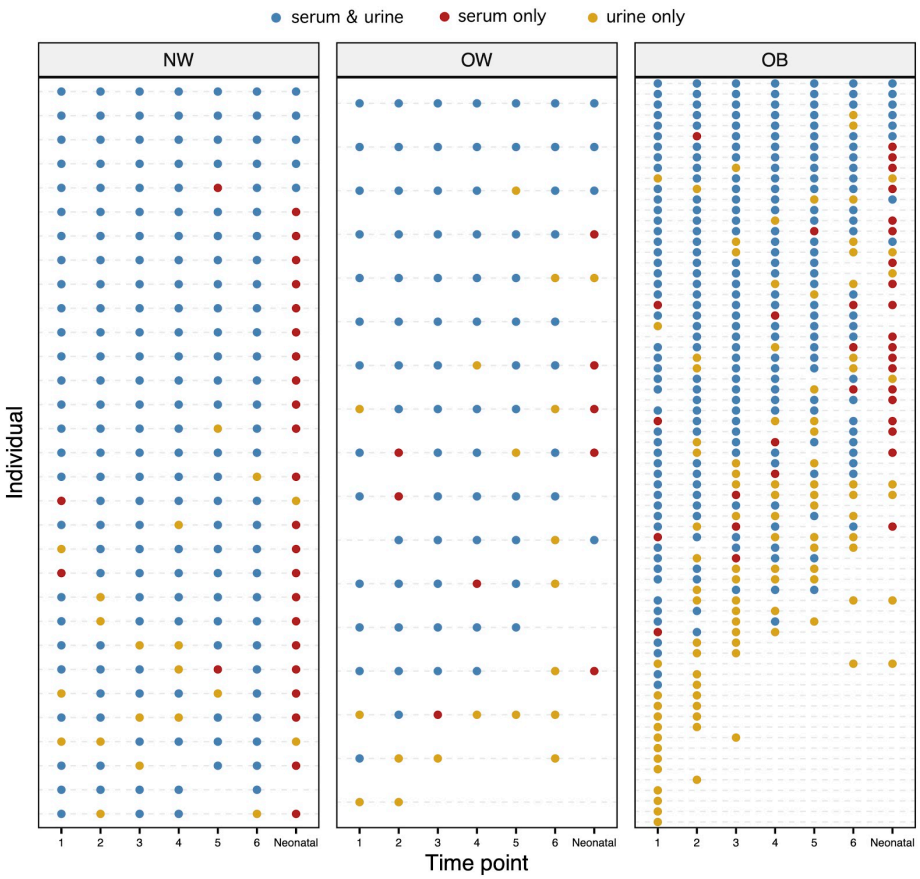


Figure 3.4: Sampling summary by group. Sampling events are represented by points. Points are coloured based on which biofluids were collected at that time point. *NW normal weight; OW overweight; OB obese.*

3.3.4 BMI-associated serum metabolites

Of all maternal serum samples analysed (see **Figure 3.S1**, Appendix B), a single outlier was detected at T6 due to high concentration of mannitol (see **Figure 1**, Appendix A), a common ingredient in pharmaceuticals. This sample was removed from further analyses. Women who developed GDM (individuals n=10; samples n=46) were excluded from feature selection to avoid selecting variables associated with the condition rather than obesity-associated variables. Cross-sectional models at each time point indicated that maternal serum metabolomes were correlated with BMI throughout gestation (**Table 3.2**). Spectral variables selected in at least one model (time point) were annotated (**Figure 3.5**) and included in time series analysis along with quantified metabolites. Loadings plots for each OPLS model are shown in **Figure 3.S3**

(Appendix B). Lipid species were excluded as ^{1}H NMR spectra do not provide enough resolution for detailed annotation of lipids.

Table 3.2: OPLS model statistics for serum samples

Time point	Samples	R^2Y	Q^2Y
T1	n=85	0.709	0.392
T2	n=75	0.711	0.308
T3	n=75	0.717	0.315
T4	n=70	0.774	0.157
T5	n=66	0.74	0.151
Umbilical Cord	n=57	0.849	-0.278

T1 11-14 weeks gestation; T2 20-24 weeks gestation; T3 28-30 weeks gestation; T4 30-33 weeks gestation; T5 35-37 weeks gestation

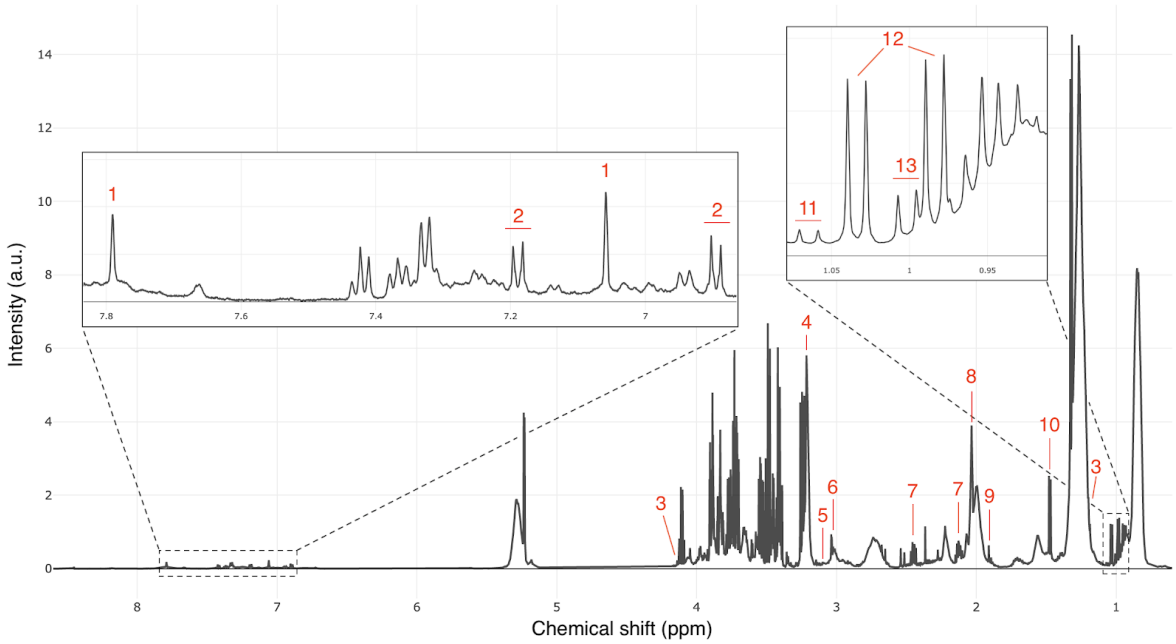


Figure 3.5: Serum median ^{1}H NMR spectrum with annotated metabolites. A median serum spectrum is plotted. Variables selected in at least one time point of the cross-sectional analysis are annotated. See **Table 1** (Appendix A) for annotation details. *1 1-methylhistidine; 2 tyrosine; 3 D-β-hydroxybutyrate; 4 choline; 5 unknown-1; 6 creatine; 7 glutamine; 8 N-acetyl glycoproteins (GlycA); 9 acetate; 10 alanine; 11 unknown-2; 12 valine; 13 isoleucine; a.u. arbitrary units*.

In total, 13 BMI-associated metabolites were extracted from the cross-sectional analysis and 13 quantified metabolites (creatinine, glutamate, glycine, histidine, leucine, phenylalanine, threonine, citrate, lactate, acetoacetate, acetone, pyruvate and glucose) were included. Concentrations of 9 metabolites were significantly changed throughout pregnancy in OB compared to NW women (**Table 3.3**). Six of these metabolites remained significant after adjusting for multiple hypothesis tests. Proteinogenic amino acids alanine (**Figure 3.6a**), glutamine (**Figure 3.6c**) and tyrosine (**Figure 3.6e**) were decreased in the OB group along with acetate (**Figure 3.6b**) and 1-methylhistidine (**Figure 3.6f**). GlycA, a composite signal from *N*-acetyl glycoproteins and a marker of systemic inflammation[241,242], increased throughout pregnancy in both groups but was markedly higher in the OB group (**Figure 3.6d**). Four BMI-associated metabolite concentrations were altered in OB women who developed GDM (**Table 3.3**), although these differences were not significant after adjusting for multiple hypothesis tests. This was due to large confidence intervals in the GDM group since only 10 patients in this dataset developed GDM.

Table 3.3: Results from time series modelling of serum metabolites

Model	*P*	*FDR*	Association with Case	Metabolite
NW vs OB	0.0493	0.1424	↓	Choline
	0.0029	**0.0151**	↓	Glutamine
	0.0006	**0.0061**	↑	GlycA
	0.0023	**0.0149**	↓	Acetate
	0.0001	**0.0026**	↓	Alanine
	0.0088	**0.0381**	↓	Tyrosine
	0.0007	**0.0061**	↓	1-methylhistidine
	0.0360	0.1170	↓	Unknown-1
	0.0237	0.0880	↑	Glutamate
OB vs GDM	0.0229	0.1638	↑	Alanine
	0.0228	0.1638	↑	Valine
	0.0252	0.1638	↑	Acetoacetate
	0.0184	0.1638	↑	Glucose

NW normal weight; OB obese; GDM gestational diabetes mellitus; GlycA N-acetyl glycoproteins; NS not significant

The decrease in some maternal proteinogenic amino acids throughout pregnancy, namely glutamine, tyrosine and alanine (in early gestation), is consistent with increased protein synthesis[243] and placental uptake to accommodate fetal protein synthesis[244]. In accordance, amino acid concentrations are reported to be higher in fetal plasma than in maternal plasma[205]. Lower amino acid concentrations in OB women during gestation could indicate increased transfer of these metabolites to support the growth of the larger fetus. Serum glutamine was previously shown to inversely correlate with BMI and risk of diabetes[245]. The same study showed the administration of glutamine improved glucose tolerance and blood pressure in mice. This suggests glutamine supplementation is a potential therapeutic for mitigating the risk of gestational diabetes and hypertensive disorders in obese pregnancies[246]. On the contrary, tyrosine and alanine are elevated in obese compared to lean non-pregnant individuals[104] and positively associated with insulin resistance[106], suggesting lower serum concentrations of these metabolites in OB women in this cohort were pregnancy-related. Since alanine and glutamine are two main glucogenic amino acids, depletion of these substrates in OB women during pregnancy could also be related to upregulation of gluconeogenesis[207,247].

The observation of reduced circulating concentrations of acetate, a major SCFA produced by gut bacteria, in pregnant obese women has not been reported but is an interesting finding given the recent mouse study linking low maternal SCFAs to high fat diet-induced obesity in offspring[77]. Decreased serum acetate in humans has also been associated with an increased risk of pre-eclampsia[248], a hypertensive disorder that is common in obese pregnancies[191]. Furthermore, acetate supplementation has been shown to prevent hypertension in mice[249]. Serum acetate was increased in obese pregnant women with improved diet and physical activity during a randomised control trial investigating the effects of a lifestyle intervention on gestational metabolic profiles[222]. The same study found that GlycA increased during obese pregnancies without the intervention at a rate higher than that of healthy pregnancies[250]. Inflammation is a hallmark metabolic adaptation during pregnancy[209], but heightened maternal inflammation during gestation is strongly associated with adverse maternal and neonatal outcomes[251] and has been linked to altered neurodevelopment[252], adiposity[253] and chronic immune-mediated diseases[254] in offspring. The metabolite 1-methylhistidine, enriched in serum of the NW group, may be associated with dietary preferences as it is derived from the metabolism of the dipeptide anserine found in animal meat, especially poultry[255].

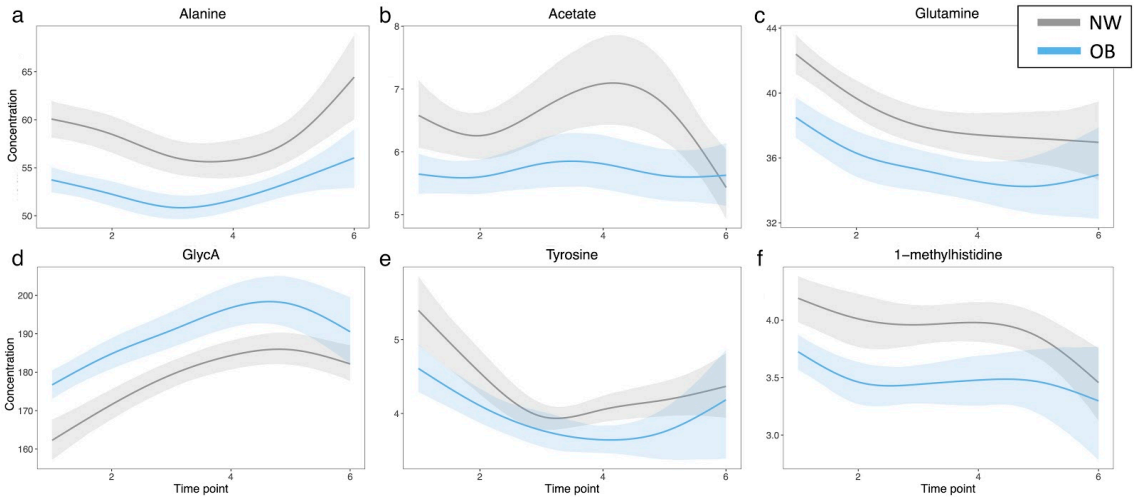


Figure 3.6: Serum metabolite concentrations throughout pregnancy. Time series analysis of BMI-associated metabolites comparing normal weight (NW) women to obese (OB) women. Curves represent group means and shaded areas denote 95% confidence bands. Metabolites plotted have significantly (*FDR* <0.05) altered serum concentrations throughout the time course. Concentrations are in arbitrary units.

Overall, the metabolic profiles of umbilical cord serum samples were not associated with maternal BMI (**Table 3.2**). Univariate analysis revealed subtle differences in individual metabolite concentrations between the study groups (**Figure 3.7a**) although no differences were significant after correction for multiple hypothesis tests. BCAAs isoleucine and valine were increased in cord serum from neonates with OW and OB mothers with the highest average concentrations found in those whose mothers developed GDM. BCAAs are consistently linked with obesity and insulin resistance[104–106,210] but are also important for fetal growth[256]. The fact that maternal and fetal concentrations of these amino acids are tightly correlated[256,257] indicates they could be involved in programming the fetus for future metabolic function[258]. Maternal products of BCAA metabolism are also positively correlated with fetal growth[219]. Cord serum creatine, histidine, lactate and pyruvate were negatively correlated with maternal BMI (**Figure 3.7b**) which likely reflects increased utilisation of these metabolites in larger fetuses.

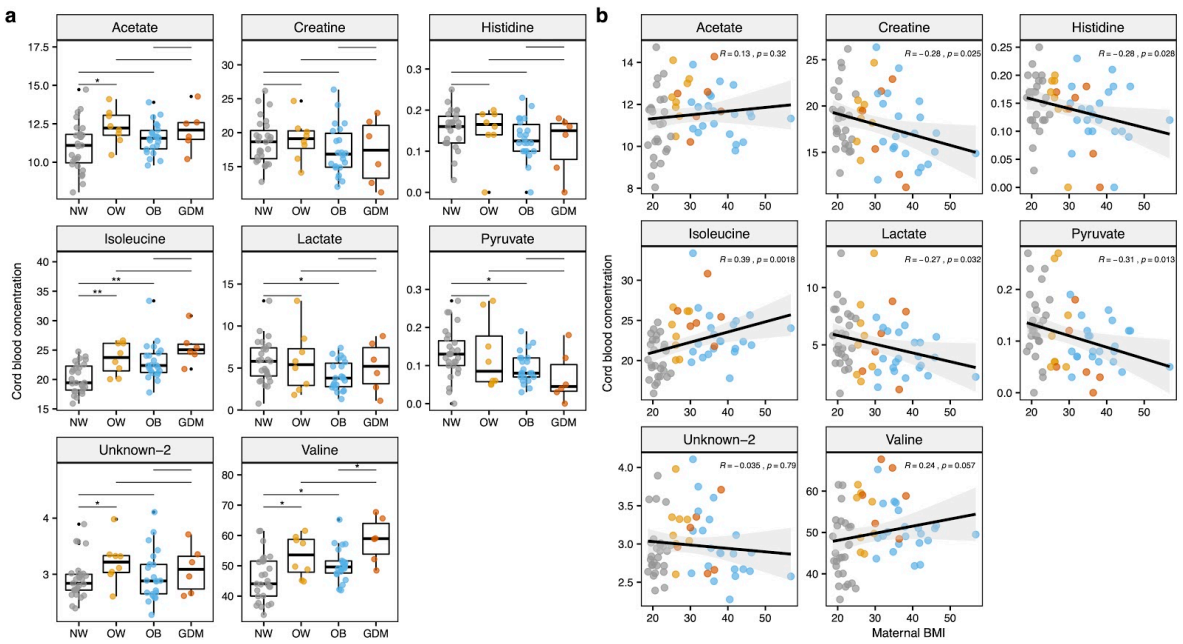


Figure 3.7: Metabolite concentrations in umbilical cord samples by study group. a) Metabolite concentrations are plotted by study group. A separate boxplot is shown for neonates whose mothers developed gestational diabetes (GDM). **b)** Metabolite concentrations are plotted according to maternal BMI with a linear regression line and confidence intervals. Spearman's correlation coefficients are displayed for each plot. Concentrations are reported in arbitrary units or, for quantified metabolites, in mmol/L. *NW normal weight; OW overweight; OB obese; *P <0.05; **P <0.01*.

Paired analysis of maternal and cord serum metabolite concentrations indicated significant fetal uptake of amino acids, lactate, creatine and acetate (**Figure 3.8**). Maternal amino acids are considered to be key determinants for fetal growth[259], so high concentrations of these metabolites in the umbilical cord are expected[205,260]. Lactate is another major energy substrate for the fetus, secondary to glucose and amino acids[261], therefore high umbilical cord concentrations of lactate are also expected. In addition to the maternal supply, lactate is produced from glucose in the placenta[262], which may explain the seemingly less consistent transfer of this metabolite from mother to fetus. Less is known about fetal creatine metabolism but in mice, the fetus relies on maternal creatine until late pregnancy[263]. Creatine supplementation during pregnancy increases placental concentration and protects the fetus from hypoxic insult[264]. Another study in mice showed that maternal acetate also crosses the placenta and influences acetate concentration in the fetus[265]. Maternal serum acetate during pregnancy is important for proper immune development in offspring[248] and suppresses allergic airway disease[265]. GlycA was the only obesity-associated metabolite that remained higher in

maternal samples reflecting the selective transport of nutrients to the fetus[67]. The paired observations of histidine and pyruvate between maternal and umbilical cord samples were not correlated, which may indicate less reliance on the maternal supply of these metabolites. In fact, it has been shown that the placenta produces pyruvate[266].

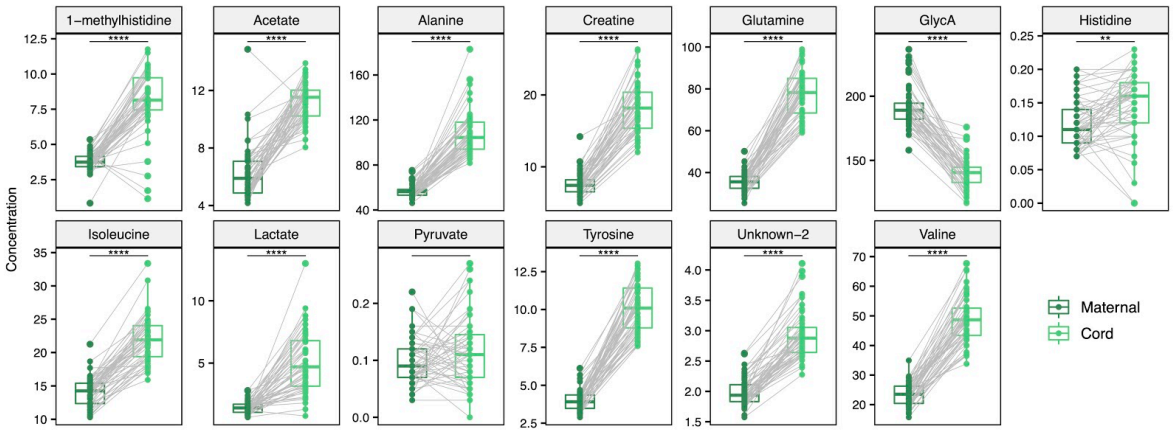


Figure 3.8: Metabolite concentrations in maternal serum compared to umbilical cord serum. Metabolite concentrations are plotted in arbitrary units or, for quantified metabolites, in mmol/L. For maternal samples, T5 (the last time point before delivery) was used. *P* values are from a Wilcoxon signed-rank test. ***P <0.01; ****P <0.0001*.

3.3.5 BMI-associated urine metabolites

Many outlying urine spectra, especially at T6, were detected by PCA (see **Figure 3.S2**, Appendix B). Outliers contained high concentrations of paracetamol metabolites (n=63; see **Figure 2**, Appendix A) or high concentrations of glucose (n=1; see **Figure 3**, Appendix A). These samples, along with samples collected from women who developed GDM (individuals n=10; samples n=48), were excluded from cross-sectional feature selection. Due to the high number of maternal samples excluded at T6, this time point was not included in the cross-sectional analysis. As observed in maternal serum, maternal urine metabolomes were correlated with BMI at all time points during gestation (**Table 3.4**). Spectral variables selected in at least one model were annotated (**Figure 3.9**) and included in time series analysis along with quantified metabolites. See **Figure 3.S4** (Appendix B) for individual loadings plots from each OPLS model. Multivariate analysis of neonatal urine was not performed due to contamination with compounds from cotton wool[267] used to collect the sample.

Table 3.4: OPLS model statistics for urine samples

Time point	Samples	R^2Y	Q^2Y
T1	n=97	0.701	0.319
T2	n=91	0.707	0.316
T3	n=84	0.743	0.358
T4	n=78	0.803	0.379
T5	n=76	0.785	0.283

T1 11-14 weeks gestation; T2 20-24 weeks gestation; T3 28-30 weeks gestation; T4 30-33 weeks gestation; T5 35-37 weeks gestation

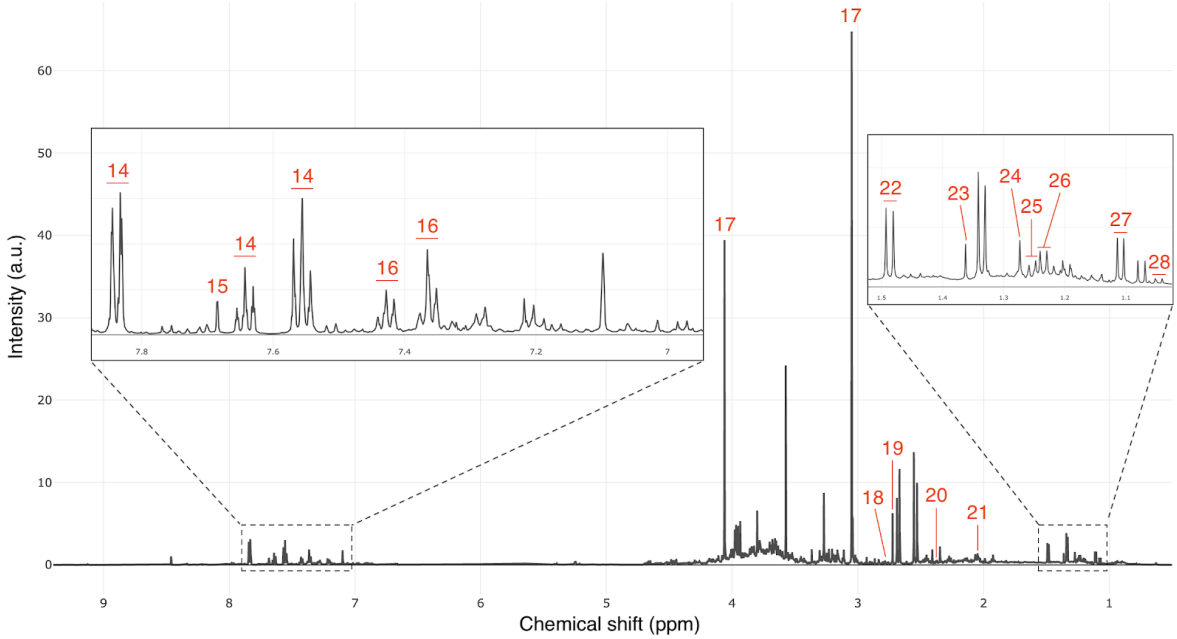


Figure 3.9: Urine median ^{1}H NMR spectrum with annotated metabolites. A median urine spectrum is plotted. Variables selected in at least one time point of the cross-sectional analysis are annotated. See **Table 1** (Appendix A) for annotation details. *14 hippurate; 15 unknown-3; 16 phenylacetylglutamine (PAG); 17 creatinine; 18 unknown-4; 19 dimethylamine; 20 oxaloacetate; 21 N-acetyl glycoproteins (GlycA); 22 alanine; 23 2-hydroxyisobutyrate; 24 3-hydroxyisovalerate; 25 methylmalonate; 26 4-deoxythreonic acid (4-DTA); 27 4-deoxyerythronic acid (4-DEA); 28 valine; a.u. arbitrary units*.

Of the 15 metabolites extracted from the cross-sectional analysis and the 17 quantified metabolites (betaine, creatine, glycine, guanidinoacetate, N,N-dimethylglycine, acetate, citrate, formate, lactate, proline betaine, succinate, tartrate, acetoacetate, acetone, pyruvate and

D-glucose), 19 compounds were significantly increased or decreased in OB women compared to NW women throughout pregnancy (**Table 3.5**). Eighteen metabolites remained significant after correcting for multiple hypothesis tests. As in the serum, no urine metabolites associated with GDM remained significant after correcting for multiple hypothesis tests (**Table 3.5**).

Table 3.5: Results from time series modelling of urine metabolites

Model	*P*	*FDR*	Association with Case	Metabolite
NW vs OB	0.0001	**0.0004**	↑	Creatinine
	0.0001	**0.0004**	↑	Dimethylamine
	0.0001	**0.0004**	↑	Unknown-3
	0.0003	**0.0009**	↑	Valine
	0.0001	**0.0004**	↑	2-hydroxyisobutyrate
	0.0136	**0.0281**	↑	3-hydroxyisovalerate
	0.0159	**0.0308**	↑	Methylmalonate
	0.0247	**0.0425**	↑	4-deoxythreonic acid
	0.0181	**0.0330**	↓	PAG
	0.0003	**0.0009**	↓	Alanine
	0.0073	**0.0174**	↓	Betaine
	0.0055	**0.0155**	↓	Guanidinoacetate
	0.0407	0.0664	↓	Hippurate
	0.0001	**0.0004**	↓	Acetate
	0.0062	**0.0160**	↓	Citrate
	0.0001	**0.0004**	↓	Formate
	0.0001	**0.0004**	↓	Succinate
	0.0002	**0.0008**	↓	Acetone
	0.0132	**0.0281**	↓	Oxaloacetate
OB vs GDM	0.0068	0.1054	↑	GlycA
	0.0291	0.1804	↑	Betaine
	0.0046	0.1054	↑	Creatine
	0.0109	0.1126	↑	Acetone
	0.0231	0.1790	↑	D-Glucose

NW normal weight; OB obese; GDM gestational diabetes mellitus; PAG phenylacetylglutamine; GlycA N-acetyl glycoproteins; NS not significant

Consistent with serum findings, lower concentrations of alanine (**Figure 3.10b**) and acetate (**Figure 3.10f**) were excreted during pregnancy in OB compared to NW women. In contrast to longitudinal observations in serum, concentrations of alanine tended to increase over the course of gestation. Increased excretion of specific amino acids including alanine was previously reported in a cross-sectional study where urine metabolic profiles of 16 non-pregnant women and 55 pregnant women at different stages of gestation were studied[268]. Aminoaciduria during pregnancy is not completely understood[269], but is thought to be associated with pregnancy-related changes in glomerular filtration rate and tubular reabsorption[270]. Ketonuria is also common during pregnancy[271]; low levels of acetone (**Figure 3.10a**) were detected in urine throughout pregnancy with lower concentrations in the OB group. Ketone bodies, produced by the metabolism of fatty acids when carbohydrate stores are depleted, are typically associated with diabetes or fasting, but metabolism of lipids is upregulated during pregnancy to conserve glucose and amino acids for the fetus[207]. Since ketone bodies can also be utilised by the fetus as fuel[272], differing urinary acetone levels between NW and OB mothers could reflect different rates of transport into the placenta.

Lower concentrations of betaine (**Figure 3.10d**) were also excreted in the OB compared to NW group, which is consistent with lower circulating levels of its precursor, choline, detected in OB serum (**Table 3.3**). While these two metabolites are strongly linked to dietary intake, they are essential for human physiology and important during fetal development[273]. Maternal plasma choline and betaine concentrations in the second trimester have been shown to positively correlate with infant cognitive scores at 18 months[274], conveying the importance of maternal nutrition during pregnancy. Dimethylamine, a microbiota-associated metabolite of choline[275], was excreted in higher amounts in the OB compared to NW group (**Figure 3.10g**). In mice, conversion of choline to methylamines by gut bacteria reduces choline bioavailability and predisposes the host to impaired glucose homeostasis and non-alcoholic fatty liver disease[276]. High excretion of dimethylamine in the present cohort may be another indicator of obesity-associated perturbation in the gut microbiota and subsequent choline deficiency during pregnancy. Formate, another product of choline metabolism[277] and a major gut bacteria-derived SCFA[278], was also reduced in OB compared to NW urine throughout pregnancy (**Figure 3.10h**). Formate is a central molecule in one carbon metabolism and evidence from animal studies suggests that formate is important for embryonic development[279]. Low formate during pregnancy has also been associated with preterm birth[216]. Reduced concentrations of formate (excreted)

and acetate (circulating and excreted) are somewhat inconsistent with previous reports of higher faecal concentrations of SCFA in non-pregnant obese individuals[95,96,201] and one report of increased excretion of acetate in obese pregnant women at a single first trimester time point[204]. However, negative correlations of urinary formate with BMI[99], serum acetate with BMI[97] and serum acetate with visceral adipose tissue[280] have been observed outside of pregnancy. Furthermore, circulating SCFA concentrations have been shown to be independent of faecal concentrations[58,97] which are poor measures of SCFA production due to rapid mucosal absorption and metabolism by epithelial cells[58]. These data suggest the present results should not be compared to studies investigating faecal SCFA compositions.

Formate and acetate are also important precursors in the citric acid cycle. Consistent with reduced excretion of these precursors, decreased concentrations of citric acid cycle intermediates citrate (**Figure 3.10e**), succinate (**Figure 3.10m**) and oxaloacetate (**Figure 3.10k**) were observed in urine of OB women. Negative correlations between urinary citrate/succinate and BMI were also reported in a large epidemiological study of non-pregnant individuals[99]. In addition to reduced availability of energy substrates, decreased excretion of citric acid cycle intermediates could be explained by high-fat diet-induced suppression of the citric acid cycle[281,282] or changes in renal function[99]. High excretion of creatinine, a metabolite of creatine, is well-established in obese subjects[283] and creatinine excretion decreases over the course of healthy pregnancies[268], two observations that were confirmed in the current study (**Figure 3.10c**). Guanidinoacetate, a creatine precursor, was excreted in lower amounts in OB compared to NW women (**Figure 3.10i**), reflecting the increased need for creatine metabolism in individuals with a high BMI[284].

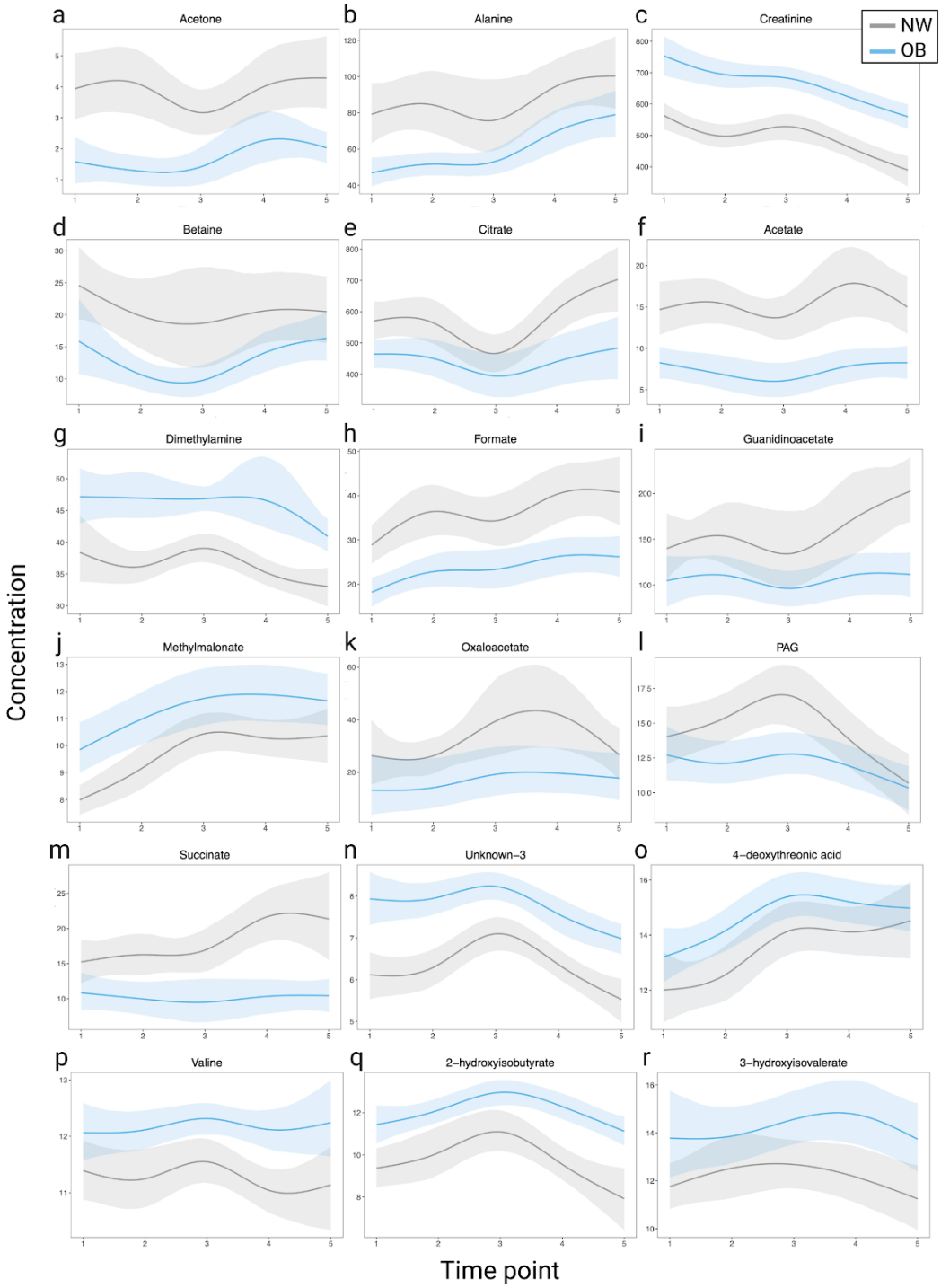


Figure 3.10: Urine metabolite concentrations throughout pregnancy. Longitudinal modelling of BMI-associated metabolites comparing normal weight (NW) women to obese (OB) women. Curves represent group means and shaded areas denote 95% confidence bands. Metabolites plotted have significantly (*FDR* <0.05) altered urine concentrations throughout the time course. Concentrations are in arbitrary units or mmol/L for quantified metabolites.

Another metabolite that was elevated in the urine of OB women during pregnancy was 4-deoxythreonic acid (**Figure 3.10o**). Excretion of this metabolite increased with gestation in both groups. Not much is known about the biological significance of this metabolite, but it is a product of threonine metabolism present in normal human urine[285] and has been shown to increase during pregnancy[219,268]. Significant increases in excretion of valine (**Figure 3.10p**) and BCAA metabolites 3-hydroxyisovalerate (**Figure 3.10r**), 2-hydroxyisobutyrate (**Figure 3.10q**) and methylmalonate (**Figure 3.10j**) were also detected in the OB compared to NW group during gestation. In addition to the previous associations between these metabolites, obesity and insulin resistance[101,104–106,210,286,287], 3-hydroxyisovalerate and 2-hydroxyisobutyrate have been linked to GDM[217]. As previously discussed, maternal BCAAs are important for fetal development[256] and are positively correlated with fetal weight[219]. However, excess amounts of BCAAs and their metabolites may contribute to insulin resistance during pregnancy and may influence insulin resistance in offspring[258]. The role of the gut microbiota in BCAA metabolism is not completely understood but gut bacteria can both synthesise and utilise BCAAs[44,213,288]. Additionally, urinary 2-hydroxyisobutyrate has been shown to positively correlate with faecal *Faecalibacterium prausnitzii*[289]. The unknown metabolite exhibited a remarkably similar excretion pattern to 2-hydroxyisobutyrate (**Figure 3.10n**) indicating it may be part of the same BCAA degradation pathway although structurally, these two metabolites are very different (**Table 1**, Appendix A).

Microbiota-associated metabolite phenylacetylglutamine (PAG) was found in lower urinary concentrations in the OB compared to the NW group (**Figure 3.10l**). PAG is produced through gut bacterial putrefaction of protein in the large intestine[41] and is a major urinary marker of gut microbial metabolism; it is one of the main excreted host-microbe co-metabolites along with hippurate and *p*-cresyl sulfate[290]. Longitudinal analysis of hippurate and PCS are shown in **Figure 3.11**. Hippurate was excreted in significantly lower amounts in the OB compared to the control group, but due to the high interindividual variability in hippurate excretion in controls, this association was not significant after adjusting for multiple hypothesis tests (**Table 3.5**). Outside of pregnancy, reduced hippurate excretion is a consistent finding in obesity[99,101]. Excretion of PAG/PCS and hippurate are influenced by intake of protein[291] and polyphenol-rich foods[292], respectively, but their production relies extensively on the metabolism of gut bacteria[59,293]. The longitudinal excretion of PAG and PCS in the present cohort are interesting given that the trend was very different in NW compared to OB women with NW PAG/PCS concentrations rising until

the beginning of the third trimester then falling to the level of OB women by the end of the third trimester. OB PAG/PCS concentrations remained relatively stable throughout gestation. Succinate also exhibited different trends in the two study groups with NW succinate excretion increasing in the third trimester while OB succinate was excreted at a constant rate during pregnancy. In addition to being a prominent host metabolite, succinate is also produced by gut bacteria and is a primary cross-feeding metabolite between members of the gut microbiota[294]. These results could indicate differential adaptations of the gut microbiota to pregnancy in NW women that are absent in OB women. The associations between BMI, choline (negative), betaine (negative) and dimethylamine (positive) prompted the longitudinal analysis of trimethylamine N-oxide (TMAO), a host-microbe co-metabolite of choline and betaine[35]. However, TMAO was not associated with obesity during pregnancy in the present cohort.

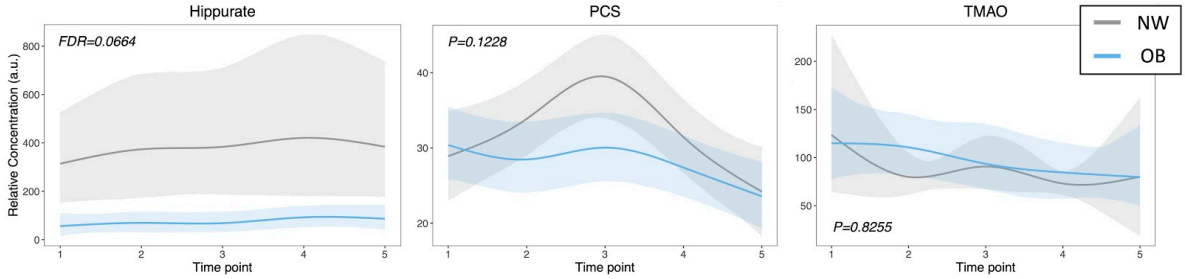


Figure 3.11: Time series analysis of important microbiota-associated metabolites. Longitudinal modelling of host-microbe co-metabolites hippurate, *p*-cresyl sulfate (PCS) and trimethylamine N-oxide (TMAO) comparing normal weight (NW) women to obese (OB) women. Curves represent group means and shaded areas denote 95% confidence bands. Concentration is in arbitrary units (a.u.).

3.3.6 Clinical associations

Relationships between maternal metabolic profile and maternal or neonatal outcomes were assessed using multi-block PLS regression (**Figure 3.12a**). Correlations based on variable loadings in the models revealed that maternal fasting insulin and insulin resistance levels mid-pregnancy were highly related to serum and urine metabolite concentrations throughout pregnancy. As indicated in a previous animal study[276], urinary dimethylamine was significantly positively correlated with maternal insulin resistance at all time points during gestation. Urine 2-hydroxyisobutyrate also positively correlated with maternal insulin resistance at all time points, which was expected given its reported association with diabetes[217,295]. These two associations point to a strong link between gut bacterial metabolism and increased maternal insulin resistance during pregnancy. Maternal insulin resistance, which is associated with

inflammation[296], was also strongly correlated with serum inflammatory marker[241,242] GlycA. Amino acids, amino acid derivatives, citric acid cycle intermediates and a ketone body found in lower concentrations in OB compared to NW pregnant women (alanine, betaine, glutamine, 1-methylhistidine, tyrosine, citrate, succinate and acetone) exhibited a negative relationship with maternal insulin resistance that was stronger with fasting insulin than fasting glucose levels. This conflicts with previous reports in non-pregnant individuals where alanine and tyrosine were positively associated with insulin resistance[106,297,298], as well as one report in non-obese pregnant women where alanine measured at one time point positively correlated with insulin resistance[299]. Conversely, glutamine has been associated with the insulin sensitive phenotype[245] and with a reduced risk of developing diabetes[287], which are consistent with observations in the present cohort. Since insulin resistance during pregnancy is designed to shunt adequate amounts of glucose to the growing fetus[300], negative correlations between metabolites that are also transferred to the fetus (**Figure 3.8**) and maternal insulin resistance seem plausible. This suggests the degree of maternal insulin resistance and the degree of nutrient transfer to the fetus are related[235,301].

Urinary alanine, betaine, citrate and acetate exhibited the most consistent correlations (negative) with maternal fasting glucose (**Figure 3.12a**). Acetate was the only metabolite to exhibit correlations (negative) with HbA1c. Previous studies have shown beneficial effects of acetate on glucose homeostasis in humans although results are not always consistent[302]. The effects are thought to be mediated through improvements in adipose tissue functioning, lipid metabolism, glucose-stimulated insulin secretion and insulin sensitivity[302]. Less is known about the role of formate in glucose homeostasis, but like acetate, this SCFA can activate some G protein-coupled receptors which are directly involved in the development of obesity-induced insulin resistance[303]. Reduced levels of formate and betaine in conjunction with increased levels of dimethylamine may also be indicative of disrupted choline metabolism which can predispose the host to insulin resistance[276]. This idea is supported by the present data where formate and betaine were inversely correlated with maternal insulin resistance while dimethylamine exhibited a positive correlation with the phenotype.

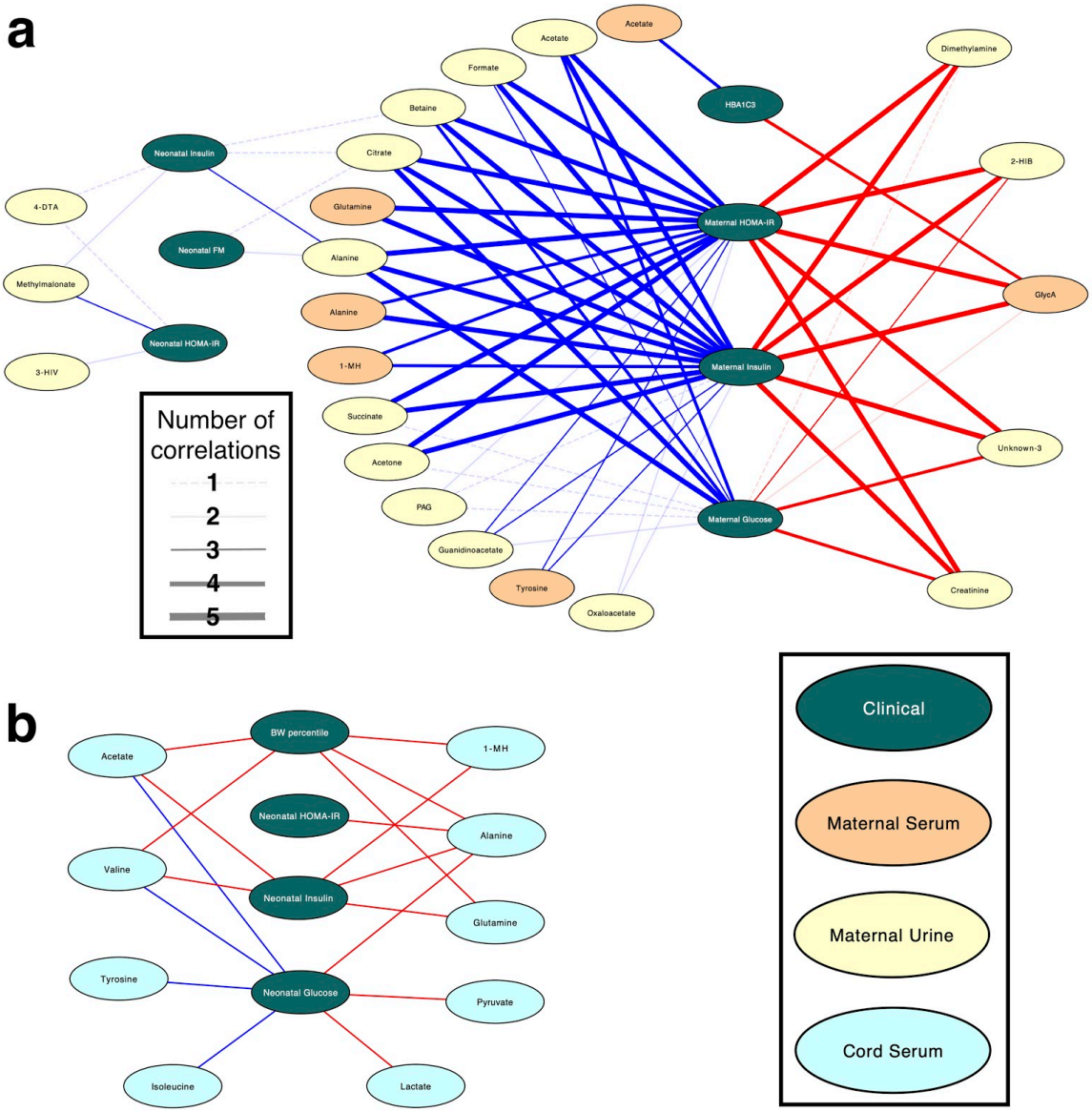


Figure 3.12: Integrative analysis of metabolite and clinical data. a) Significant (*FDR* <0.05) correlations between serum or urine metabolites and maternal or neonatal outcomes. Widths of edges depict the number of time points in which the correlation was significant. **b)** Significant (*FDR* <0.05) correlations between umbilical cord metabolites and neonatal outcomes. Positive and negative correlations are shown in red and blue, respectively. See **Tables 3.S1** and **3.S2** (Appendix B) for tabulated values. *4-DTA 4-deoxythreonic acid; 3-HIV 3-hydroxyisovalerate; 1-MH 1-methylhistidine; PAG phenylacetylglutamine; HbA1c glycated haemoglobin; HOMA-IR homeostatic model assessment for insulin resistance; 2-HIB 2-hydroxyisobutyrate; GlycA N-acetyl glycoproteins; BW birth weight.*

While maternal metabolites were only weakly related to neonatal insulin and insulin resistance levels (**Figure 3.12a**), metabolite concentrations in umbilical cord serum were highly related to neonatal outcomes (**Figure 3.12b**). Amino acids and energy metabolites were positively correlated with neonatal insulin, insulin resistance and birth weight percentile but were not related to neonatal fat mass. However, neonatal fat mass may be more related to lipid species than small molecules[304]. Although umbilical cord acetate and valine were associated with higher neonatal insulin levels, these metabolites, along with tyrosine and isoleucine, were associated with lower neonatal glucose levels. This could be explained by the direct influence of acetate[103] and branched-chain/aromatic amino acids[305] on insulin secretion and subsequent glucose uptake. In addition, increasing placental amino acid levels has been shown to correct fetal beta cell deficits, proving their role in effective insulin secretion[306]. Alanine was the only metabolite in umbilical cord samples to correlate (positive) with neonatal insulin resistance.

Together with current medical knowledge of metabolic adaptations during pregnancy, these findings indicate that the maternal metabolic phenotype during pregnancy, although influenced by obesity, may also be highly dependent on maternal-fetal metabolite transfer which is likely influenced by maternal insulin resistance. Elevated maternal insulin resistance causes maternal hyperglycaemia and glucose transfer to the fetus is subsequently increased. The present data suggest that other maternal circulating metabolites, particularly amino acids, may also be transferred to the fetus at a higher rate in pregnancies with high maternal insulin resistance, as indicated by the negative correlations between these metabolites and maternal HOMA-IR as well as the lower concentrations of these metabolites in OB compared to NW women throughout pregnancy. The resulting fetal metabolic phenotype is dependent on the selective uptake of maternal metabolites and may play a role in fetal hyperinsulinaemia followed by fetal overgrowth. The present data suggest that manipulation of the gut microbiota to regulate maternal insulin resistance during pregnancy should be investigated in mouse models to determine its potential for mediating obesity-related fetal overgrowth.

Chapter 4

Gestational metabolism in women with previous bariatric surgery

Bariatric surgery, particularly the malabsorptive type, induces a drastic and persistent shift in gut microbial community structure. The altered activity of the gut microbiota in bariatric patients is believed to be involved in improvements in obesity-related comorbidities, but its effect on pregnancy outcomes has not been investigated. Longitudinal metabolic profiling of 111 women was performed to identify changes in host-microbe co-metabolism related to bariatric surgery that persist throughout pregnancy, and that may be related to neonatal outcomes. Blood (serum) and urine samples were collected at six time points and analysed by ^{1}H NMR spectroscopy. While no changes in metabolism were identified in women with previous restrictive surgery, those with a previous malabsorptive surgery showed significant changes in branched-chain and aromatic amino acid metabolism that were linked to a shift in the gut microbiota toward facultative anaerobes and species typically found in the upper gastrointestinal tract. The gestational metabolic signature was correlated with maternal insulin resistance but did not have a strong association with neonatal outcomes. However, phenylacetylglutamine, a host-microbial co-metabolite that was elevated in the urine of malabsorptive mothers, was also elevated in the urine of neonates born from these mothers. This observation suggests the altered maternal metabolic phenotype has the potential to impact the child, which may have long-term health implications. The data presented in this chapter has been published[132] (see **Appendix D**).

4.1 Introduction

Due to the associated comorbidities and increased health risks imposed by obesity, weight loss treatments are important for addressing this global epidemic. When dietary and lifestyle interventions fail, bariatric surgery is an alternative option. Bariatric procedures fall into two categories: restrictive treatments, such as gastric banding, aim to limit food intake and malabsorptive treatments, such as gastric bypass, aim to decrease nutrient absorption[307].

Compared to restrictive surgeries, the malabsorptive types result in a higher percentage of weight loss and greater resolution of obesity-related comorbidities. Post-gastric bypass, patients lose an average of 14 kg/m^2 in the first year, around 80% of patients recover from dyslipidaemia and hypertension, and more than 90% of patients achieve diabetes remission[308]. Improvement in insulin resistance can occur within two weeks of surgery[309], before a significant amount of weight or body fat is lost, suggesting weight-independent mechanisms of diabetes resolution. In gastric bypass, the stomach is partitioned to form a small pouch which is attached to the distal part of the small intestine, bypassing a portion of the proximal small intestine[307]. The anatomical rearrangement results in decreased production of stomach acid and allows more oxygen to enter the large intestine[310,311]. Decreased pH and increased oxygen content in the gastrointestinal tract, in conjunction with altered food intake and impaired digestion, contribute to the changes in microbial community structure observed in gastric bypass patients[311–316]. These conditions create a favourable environment for the translocation of oral-associated microbiota, and the proliferation of facultative anaerobes and acid-tolerant bacteria. A consistent and sustained gut bacterial signature has been observed in patients post-malabsorptive surgery; an increase in relative abundance of *Gammaproteobacteria* is the strongest biomarker associated with gastric bypass in humans[317] and in animal models[318,319]. Knowledge of the downstream effects of this altered microbial community on the host is still lacking, although evidence suggests that changes in the gut microbiota may play a role in insulin sensitivity following gastric bypass surgery[105,320]. Patients who have elected for a restrictive procedure lose less weight, have lower diabetes remission rates and worse maintenance of weight loss than those with malabsorptive procedures[308]. However, restrictive procedures are associated with less complications[308]. Minimal changes in the gut microbiota are observed following restrictive procedures, most likely due to the minor changes in gastrointestinal physiology[313].

Bariatric surgery has proven to be the most effective solution to morbid obesity, but the long-term and transgenerational effects of these procedures are poorly understood. Women who have had bariatric surgery, especially a malabsorptive surgery, are at a high risk of delivering small for gestational age (birth weight <10th percentile) neonates[321,322] and may be more likely to deliver preterm compared to obese women without bariatric surgery[321]. In spite of the poor outcomes in neonates, women with previous malabsorptive surgery maintain significantly lower insulin resistance levels during gestation compared to women with a similar BMI[131], and are therefore less likely to develop gestational diabetes[322]. Associations between the

gut microbiota and host glucose homeostasis have been identified[105,320], but a potential link between improved insulin sensitivity and bariatric-surgery induced changes in the gut microbiota has not been elucidated. Similarly, the effects of the altered community on fetal development and the long-term health of the offspring have not been investigated to date.

To begin to unravel the complex changes in host-microbe metabolic relationships in bariatric patients and the effects of these changes on pregnancy outcomes, longitudinal data from 41 pregnant women with previous bariatric surgery were analysed compared to 70 pregnant women with similar early pregnancy BMI, but with no history of weight loss surgery. Maternal and neonatal outcomes were evaluated in association with observed changes in metabolic profiles.

4.2 Materials and methods

4.2.1 Study population, sampling and clinical data

The patient cohort, biospecimen sampling and clinical data collection for the Chelsea & Westminster Hospital (CWH) study are detailed in section **2.1.1**. The results in this chapter were derived from dataset 2, a subset of the cohort defined to compare patients with a previous bariatric surgery to patients without bariatric surgery. Exclusion criteria for dataset 2 included: BMI <25 kg/m^2 or BMI >50 kg/m^2 (to match the range of bariatric patients), miscarriage, still birth, Type 2 diabetes mellitus, gestational diabetes mellitus (due to the influence of bariatric surgery on insulin sensitivity) and withdrawal from the study.

Participants were classified into 3 groups depending on the type or lack of previous bariatric surgery: restrictive (RES), malabsorptive (MAL) or no bariatric surgery (NBS). Hypothesis tests were performed for each case group (RES or MAL) with NBS as reference. Student's t-test or Wilcoxon rank sum (Mann-Whitney) test were used to assess groupwise differences in numerical variables for normally and non-normally distributed data, respectively. For categorical variables, the chi-squared test was used unless the variable had fewer than 5 values, a Fisher's exact test was used. The Benjamini-Hochberg[190] procedure was used to control the false discovery rate (FDR) and an alpha of 0.05 was used throughout. All calculations (except where stated otherwise) were executed in the R (version 3.5.0) statistical environment[156]. A list of packages and package versions used can be found in **Appendix C**.

4.2.2 Metabolic profiling of biofluid

Samples were prepared and ^{1}H NMR spectra were acquired as detailed in **3.2.2**. Raw spectral data were pre-processed as described in **2.2.4**, and multivariate modelling on autoscaled data was carried out in SIMCA V.14.1 (Sartorius Stedim Biotech). Outlying spectra identified through Principal Component Analysis (PCA) were inspected for contamination, drug metabolites or high concentrations of glucose. Spectra containing these signals were excluded from supervised analysis. Spectral variables associated with a previous bariatric surgery were identified using orthogonal partial least squares discriminant analysis (OPLS-DA). Data were modelled at each time point and for each pairwise comparison (NBS vs RES, RES vs MAL or NBS vs MAL). A single orthogonal component was fit in each model. Models were evaluated by 7-fold cross validation using R^2Y and Q^2Y metrics. Variables which had a correlation coefficient (absolute value) of 0.5 or greater with the predictive component were annotated. A representative peak from each annotated metabolite was integrated to calculate relative concentration, and time series analysis of each metabolite was conducted as detailed in **3.2.2**.

Relationships between metabolite relative concentrations at each time point and clinical variables were assessed using Spearman's correlation (package "Hmisc"). Partial correlations were then calculated ("ppcor"[187]) to observe the metabolite-clinical relationships without the effects of potential confounders (maternal age, BMI and insulin resistance). Biospecimens with missing data were removed from the analysis. The correlations were visualised as a network using Cytoscape[224] (version 3.7.0).

4.2.3 Gut bacterial community profiling

Detailed descriptions of bacterial DNA extraction from faecal samples, amplification of the 16S rRNA gene and library sequencing are provided in **2.3.2** and **2.3.3**. For each faecal sample analysed, alpha-diversity was estimated (Shannon diversity index[323], package "vegan") after rarefying to the minimum sequencing depth. Beta-diversity was calculated using Bray-Curtis dissimilarity[324] on log-transformed (pseudocount of 1 added to each value to avoid log(0)) data. Differences in bacterial community structure between NBS and MAL patients were tested using permutational multivariate analysis of variance (PERMANOVA)[325]. Changes in relative abundance of specific bacterial taxa at different taxonomic levels were assessed using the

Mann-Whitney U test. At the amplicon sequence variant level, differential abundance was tested using "DESeq2"[326] on raw counts. The method implemented in "DESeq2" includes dispersion estimation and correction which allows a more robust calculation of fold change for sparse count data. Here, data were pooled for each individual instead of calculating differential abundance at distinct time points.

4.2.4 Integrative analysis of metabolic and gut bacterial data

Multi-block sparse PLS-DA[173] (see **2.4.1.2**) was employed to uncover a systems-level signature of serum metabolites, urine metabolites and gut bacteria that discriminated case from control patients. The model contained 3 blocks: serum metabolites, urine metabolites and bacterial genera (log-transformed). Metabolites were chosen based on their behaviour over the time course; those significantly (*FDR <0.05*) altered in the case group throughout gestation were included in the model. Sparsity was introduced only for the block containing taxonomic data with the goal of selecting bacterial genera that were highly associated with the metabolites of interest. Since faecal samples were only collected at T1, T2 and T4, additional time points were not included in the model.

4.3 Results and discussion

4.3.1 Characteristics of the study participants

Of the 178 women recruited for this study, 67 were excluded from the present analysis due to BMI (n=36), diagnosis of diabetes (n=19), miscarriage/still birth (n=7), withdrawal from the study (n=3) or no clinical samples provided (n=2). The final study population included 111 women (NBS n=70; RES n=16; MAL n=25). Within the RES group, 8 women had gastric banding and 8 women had a sleeve gastrectomy while all MAL patients had gastric bypass. Age distribution differed significantly between the groups (**Table 4.1**), but BMI was similar among all groups. In accordance with previous literature[321,322], bariatric patients had lower insulin resistance during pregnancy than NBS and this was more pronounced in the MAL group[131]. As detailed in **Chapter 3**, high maternal BMI was associated with elevated maternal insulin resistance during pregnancy which resulted in fetal overgrowth, increased neonatal adiposity and elevated neonatal insulin levels at birth. However, women with a high BMI and a previous malabsorptive

surgery were able to maintain insulin resistance during pregnancy at the level of women with a healthy BMI (see **Figure 4.S1**, Appendix B). Increased insulin resistance during pregnancy is a normal physiological adaptation to salvage glucose for the growing fetus[300], but abnormal elevation in insulin resistance, as observed in obese women, is associated with fetal overgrowth[235]. Understanding the mechanisms by which a previous malabsorptive surgery contributes to maintained insulin sensitivity during pregnancy in women with a high BMI could have translational potential.

Although the improvement in insulin resistance associated with bariatric surgery was not reflected in neonatal insulin levels at birth (**Table 4.1**), neonatal birth weight was lower in both case groups regardless of gestational age at delivery[133]. The prevalence of neonates born large for gestational age (birth weight >90th percentile) was reduced in women with a previous bariatric surgery compared to NBS, although the prevalence of neonates born small for gestational age (particularly in the MAL group) was slightly increased. In spite of reports suggesting bariatric women are at high risk of delivering small for gestational age neonates, the bariatric women in the present cohort delivered babies with comparable birth weight percentile to normal weight women (**Figure 4.S2**, Appendix B). In addition, the trajectories of fetal growth in bariatric women were similar to that of normal weight women (**Figure 4.S3**, Appendix B). This discrepancy with the literature may be an artefact of the small sample size.

Table 4.1: Participant characteristics. Values reported are mean ± standard deviation. Statistical tests were performed with no bariatric surgery as the reference group.

	No Bariatric Surgery (n=70)	**Restrictive Surgery (n=16)**	***P* value**	**Malabsorptive Surgery (n=25)**	***P* value**
Maternal					
Ethnicity	77.1% Caucasian	75% Caucasian	1	76% Caucasian	1
Age	29.6 ± 5.3	32.2 ± 3.8	**0.029**	34 ± 4.8	**3.85E-04**
BMI at T1 (kg/m^2)	34.1 ± 5.7	32.8 ± 6.2	0.184	32.7 ± 3.8	0.117
Maternal fasting insulin (mU/mL)	12.3 ± 5.5	10.1 ± 6.8	**0.027**	6.9 ± 4	**1.50E-06**
Maternal fasting glucose (mmol/L)	4.5 ± 0.6	4.5 ± 0.5	0.405	4.1 ± 0.4	**1.82E-03**
Maternal HOMA-IR	2.5 ± 1.3	2.1 ± 1.5	**0.037**	1.3 ± 0.7	**5.57E-07**
Maternal HbA1c (mmol/mol)	31.3 ± 4.1	31.6 ± 3.2	0.565	31.7 ± 2.6	0.747
Neonatal					

Gestational age at delivery (weeks)	39.4 ± 1.4	39 ± 2.3	0.768	38.5 ± 2.3	**0.042**
Delivery mode	45.7% C-section	62.5% C-section	0.350	40% C-section	0.795
Sex of baby	51.4% Male	56.2% Male	0.943	56% Male	0.873
Birth weight (g)	3520.9 ± 603.9	3096.8 ± 519.1	**4.23E-03**	3040.6 ± 638.1	**1.08E-03**
Birth weight percentile	59.5 ± 33.3	39.3 ± 26.3	**6.94E-03**	38.8 ± 28.6	**2.33E-03**
Small for gestational age	11.4% Yes	12.5% Yes	1	16% Yes	0.726
Large for gestational age	28.6% Yes	6.2% Yes	0.052	8% Yes	**0.029**
Neonatal fat mass (g)	687.2 ± 199.2	572.1 ± 189	**0.037**	553.8 ± 240.9	**6.83E-03**
Neonatal fat mass (%)	19.3 ± 4.2	17.8 ± 4.6	0.158	17.2 ± 5.1	0.093
Umbilical cord insulin (mU/mL)	7.6 ± 5.8	6 ± 4.5	0.198	8.4 ± 6.7	0.592
Umbilical cord glucose (mmol/L)	4.2 ± 1.2	4.2 ± 1	0.621	4.7 ± 1.7	0.923
Umbilical cord HOMA-IR	1.5 ± 1.3	1.2 ± 1	0.205	2 ± 2.5	0.681

BMI body mass index; HOMA-IR homeostatic model of insulin resistance; HbA1c glycated haemoglobin; T1 12-14 weeks gestation

4.3.2 Associations between maternal and neonatal outcomes

Unlike the observations made with regard to obese pregnancies (see **3.3.2**), maternal insulin resistance in bariatric patients was not related to neonatal insulin levels at birth (**Figure 4.1a**) despite the large disparity in maternal insulin resistance between MAL and NBS patients[131]. However, in agreement with the previous findings, insulin concentration in the umbilical cord at birth was positively correlated with both birth weight percentile (**Figure 4.1c**) and neonatal fat mass (**Figure 4.1d**). The disconnect between maternal and neonatal insulin levels suggests additional factors may contribute to reduced fetal growth in bariatric women. Maternal nutrient deficiency, particularly in malabsorptive patients, likely plays a role[327]. Although the MAL group had shorter gestation compared to NBS (**Table 4.1**), the gestational age at delivery was not associated with neonatal umbilical cord insulin at birth (**Figure 4.1b**). Again, the discrepancy between findings in obese (see **Figure 3.3**) and MAL patients indicate divergent mechanisms which influence neonatal outcomes in either population. Understanding changes in maternal metabolism that result from a previous bariatric surgery are key to revealing the effect of these surgeries on fetal health.

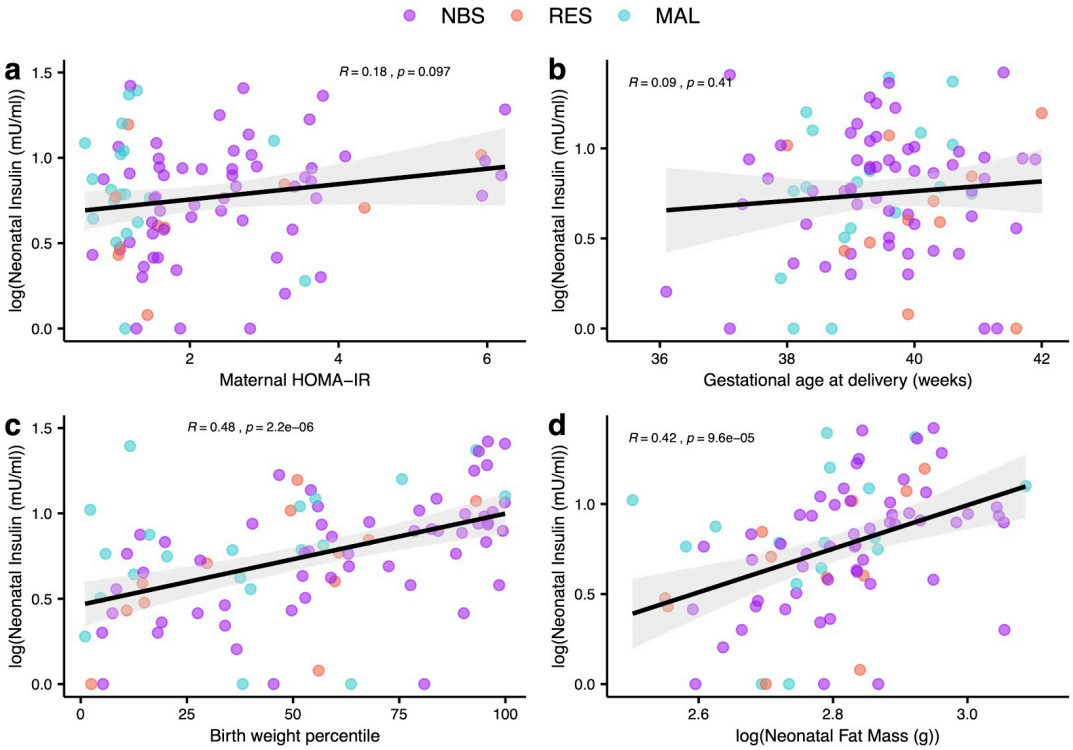


Figure 4.1: Neonatal insulin is associated with fat mass and weight at birth. The concentration of insulin in the umbilical cord was not related to **a)** maternal insulin resistance or **b)** gestational age at delivery. Umbilical cord insulin was positively correlated with **c)** weight and **d)** fat mass at birth. Correlation coefficients (Pearson) are displayed as well as linear regression lines (black) and 95% confidence intervals (light grey). Colours of data points indicate maternal surgery status. *NBS no bariatric surgery; RES restrictive; MAL malabsorptive.*

4.3.3 Sampling summary

The dynamics of gestational metabolism in bariatric patients were studied using 491 maternal urine, 394 maternal blood, 68 maternal faeces, 54 umbilical cord blood and 28 neonatal urine samples (**Figure 4.2**) collected from the 111 women included here and 67 of their offspring. Two women (RES n=1, MAL n=1) contributed only neonatal samples. Since the study is ongoing, many women had incomplete sample sets at the time of analysis. However, a cross-sectional approach was taken to interrogate the data initially, and individuals with more than 2 missing time points were excluded from longitudinal analysis. Faecal samples were only collected from a subset of NBS and MAL patients at T1, T2 and T4.

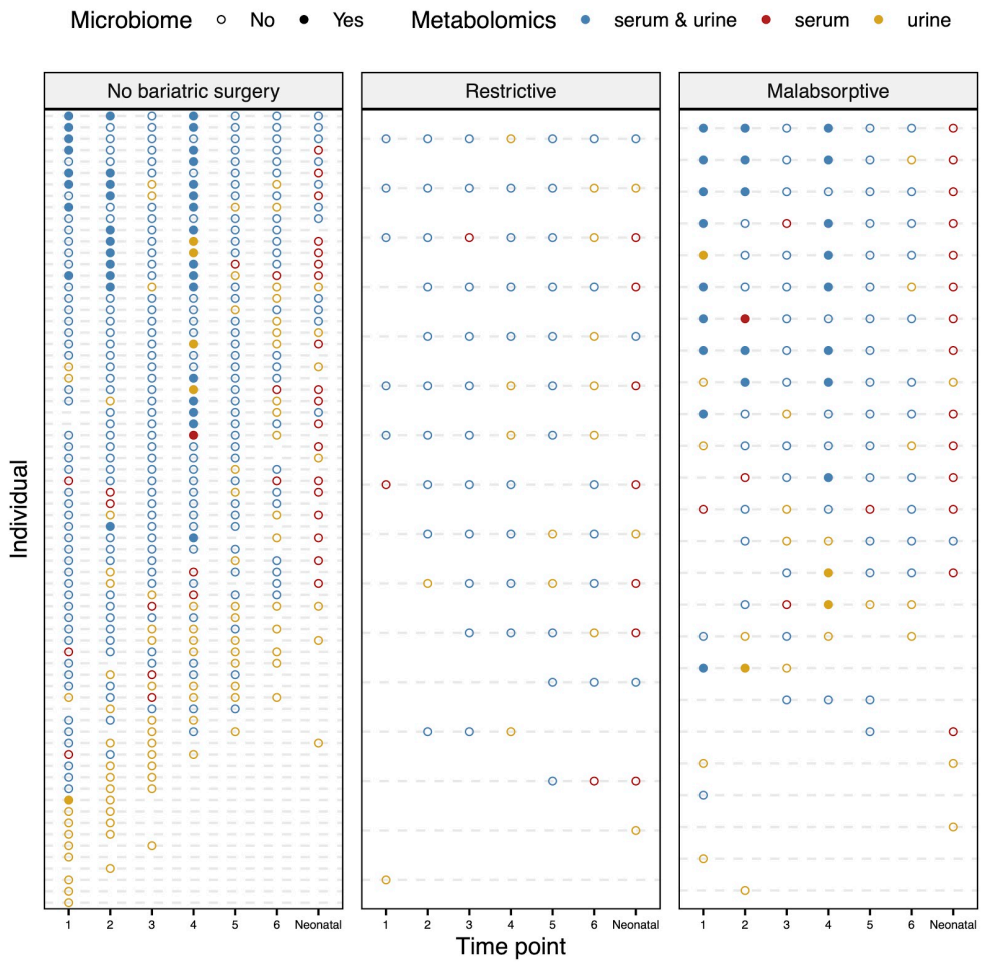


Figure 4.2: Summary of data collection by group. Each point represents a sampling event for a single individual. Colours denote which biofluids were collected for ^{1}H NMR spectroscopy (both serum and urine, only serum or only urine). Solid points indicate faecal samples were also collected for 16S rRNA gene amplicon sequencing.

4.3.4 Serum metabolites associated with previous bariatric surgery

All maternal serum samples analysed (see **Figure 4.S4**, Appendix B) were included in cross-sectional OPLS-DA models to uncover metabolic signatures associated with restrictive or malabsorptive bariatric surgery. Metabolic profiles of RES patients could not be discriminated from either NBS or MAL patients at any time point investigated (see **Table 4.S1**, Appendix B). Since sleeve gastrectomy induces malabsorptive effects while gastric banding is purely

restrictive, the RES group likely represented a metabolic profile intermediate to NBS and MAL patients. However, maternal metabolism was altered in MAL compared to NBS patients at T4 and T5 (**Table 4.2**), corresponding to the third trimester of pregnancy, although the predictive ability of these two models were moderate (Q^2Y between 0.15 and 0.35)[176]. Spectral variables selected from OPLS-DA models at T4 and T5 (see **Figure 4.S6**, Appendix B) were annotated (**Figure 4.3**) and significance of associations between metabolites and the MAL phenotype were tested by time series analysis. Lipid species were not investigated as ^{1}H NMR spectra do not provide enough resolution for detailed annotation of lipids. The metabolic profiles acquired from umbilical cord blood were not overtly affected by maternal surgery status (**Table 4.2**).

Table 4.2: OPLS-DA model statistics for serum samples

Time point	Samples	R^2Y	Q^2Y
T1	NBS n = 54 MAL n = 12	0.749	-0.156
T2	NBS n = 47 MAL n = 15	0.825	0.114
T3	NBS n = 47 MAL n = 15	0.744	0.091
T4	NBS n = 43 MAL n = 15	0.775	0.178
T5	NBS n = 38 MAL n = 17	0.847	0.245
T6	NBS n = 27 MAL n = 11	0.789	0.028
Umbilical Cord	NBS n = 30 MAL n = 14	0.885	0.031

NBS no bariatric surgery; MAL malabsorptive; T1 11-14 weeks gestation; T2 20-24 weeks gestation; T3 28-30 weeks gestation; T4 30-33 weeks gestation; T5 35-37 weeks gestation; T6 delivery

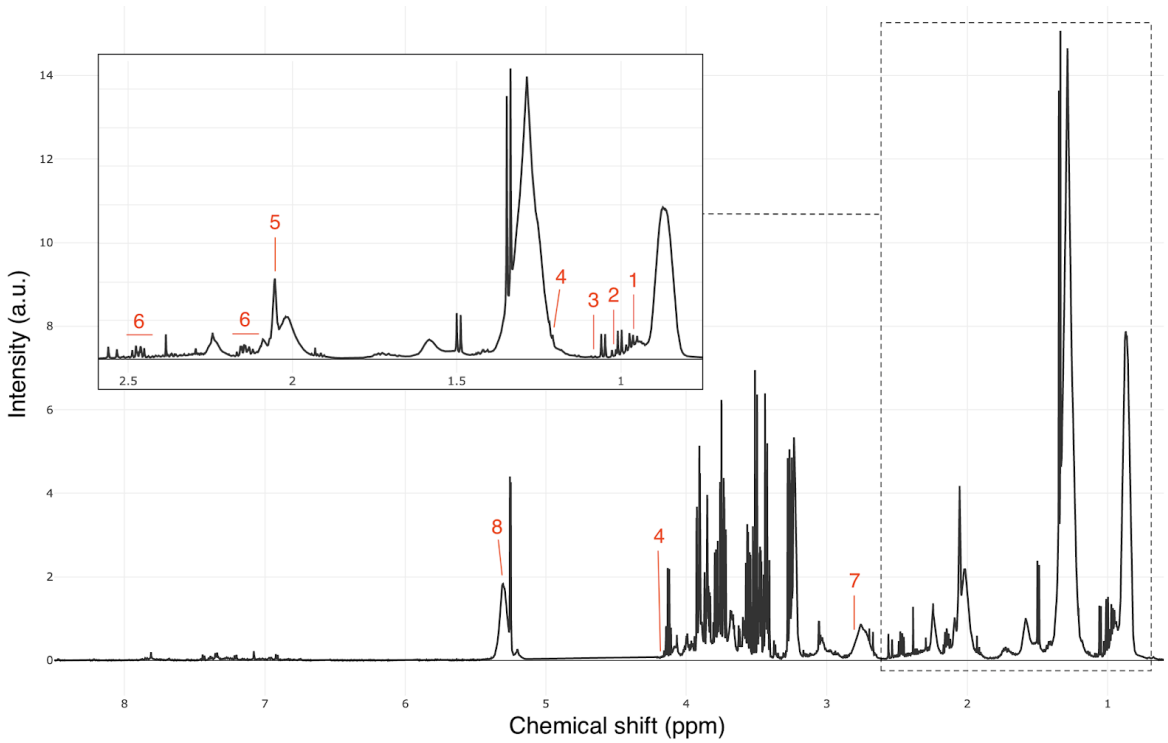


Figure 4.3: Serum median ^{1}H NMR spectrum with annotated metabolites. Spectral peaks which contributed to the discrimination between women with malabsorptive surgery and women without bariatric surgery are annotated. See **Table 2** (Appendix A) for annotation details. *1 leucine; 2 isoleucine; 3 isobutyrate; 4 D-β-hydroxybutyrate; 5 N-acetyl glycoproteins (GlycA); 6 glutamine; 7 lipids (C=CCH2C=C); 8 unsaturated lipids (CH=CH).*

4.3.5 Urine metabolites associated with previous bariatric surgery

Exclusion of some maternal urine samples (n=24) was necessary due to the presence of drug metabolites (mainly paracetamol, see **Figure 2**, Appendix A) or high concentrations of glucose (see **Figure 3**, Appendix A). Like in serum samples, the RES metabolic profile in urine was not appreciably altered compared to either NBS or MAL patients (see **Table 4.S1**, Appendix B). However, the MAL urinary metabolic profile could be discriminated from that of NBS at all time points during gestation (**Table 4.3**). The overall performance of the models was higher using urine compared to serum samples based on the Q^2Y metric. From each individual time point model (except T6: see **Figure 4.S5**, Appendix B), variables were selected based on their contribution to the discrimination between NBS and MAL groups (see **Figure 4.S7**, Appendix B), then annotated (**Figure 4.4**). Significance of each individual metabolite was tested by time

series analysis. Due to contamination with compounds found in cotton wool[267] that was used to collect neonatal urine, multivariate modelling of these samples was not conducted. A single discriminant urinary metabolite could not be identified (unknown; resonances at 7.34 (d) and 7.48 (t); see **Table 2**, Appendix A).

Table 4.3: OPLS-DA model statistics for urine samples

Time point	Samples	R^2Y	Q^2Y
T1	NBS n = 63 MAL n = 16	0.83	0.159
T2	NBS n = 61 MAL n = 16	0.843	0.324
T3	NBS n = 58 MAL n = 17	0.843	0.407
T4	NBS n = 52 MAL n = 18	0.812	0.344
T5	NBS n = 51 MAL n = 18	0.852	0.483
T6	NBS n = 41 MAL n = 16	0.767	0.0666

NBS no bariatric surgery; MAL malabsorptive; T1 11-14 weeks gestation; T2 20-24 weeks gestation; T3 28-30 weeks gestation; T4 30-33 weeks gestation; T5 35-37 weeks gestation; T6 delivery

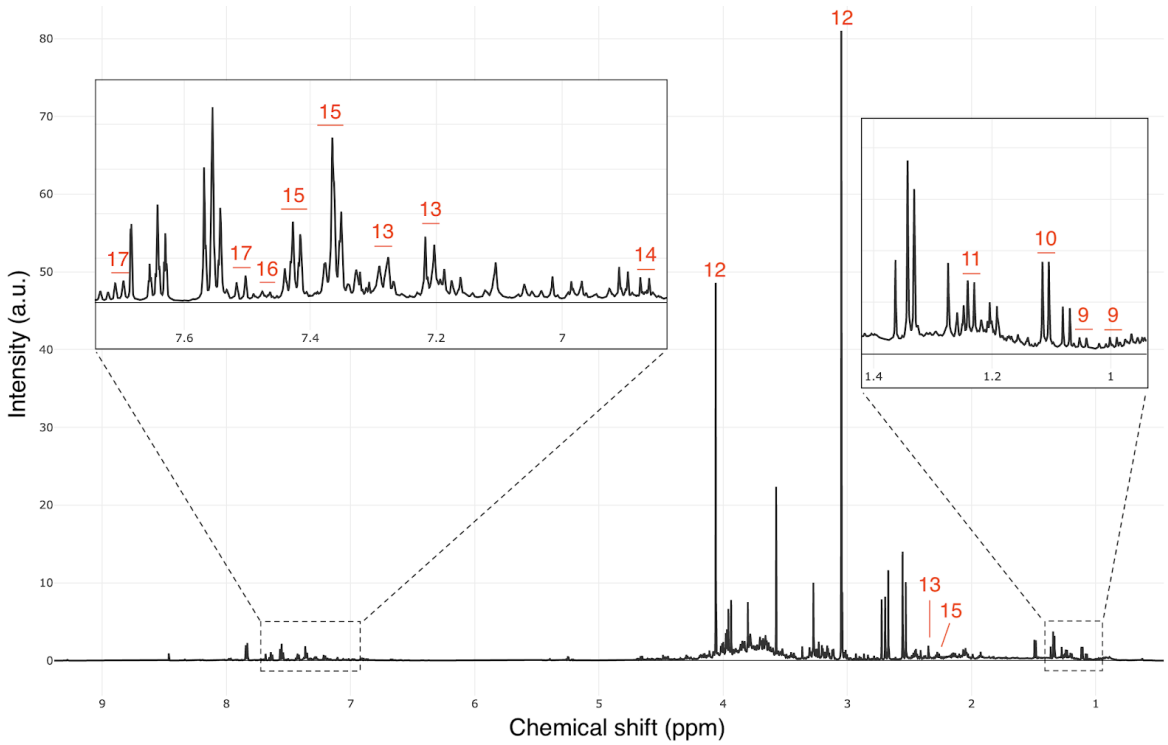


Figure 4.4: Urine median ^{1}H NMR spectrum with annotated metabolites. Spectral peaks which contributed to the discrimination between women with malabsorptive surgery and women without bariatric surgery are annotated. See **Table 2** (Appendix A) for annotation details. *9 valine; 10 α-ketoisovalerate; 11 methylmalonate; 12 creatinine; 13 p-cresyl sulfate; 14 p-hydroxyphenylacetate; 15 phenylacetylglutamine; 16 unknown; 17 indoxyl sulfate.*

4.3.6 Malabsorptive metabolic profile is consistently altered throughout gestation

To avoid spurious findings, each metabolite identified in the cross-sectional analysis of the maternal samples was subjected to time series analysis with the goal of establishing a metabolic signature of malabsorptive surgery that persisted throughout pregnancy. Patients in the MAL group had consistently reduced serum concentrations of branched-chain amino acids (BCAAs) leucine and isoleucine as well as branched-chain fatty acid (BCFA) isobutyrate (**Figure 4.5a**). As discussed in the previous chapter, BCAAs are associated with obesity-related insulin resistance[104–106,210]. Diminished circulation of BCAAs in women with a previous malabsorptive surgery is in line with the reduced insulin resistance observed in this population compared to BMI-matched women without bariatric surgery. The composite signal from *N*-acetyl glycoproteins (GlycA), a marker of systemic inflammation[241,242], was also decreased in the MAL

group, though only in the third trimester of pregnancy. In the previous chapter, an abnormal increase in inflammation in obese women over the course of gestation was demonstrated. The present data showed a previous malabsorptive surgery attenuated the rise in pregnancy-related inflammation. Controlling systemic inflammation during pregnancy may protect against both maternal and neonatal adverse outcomes[251], as well as chronic health complications in offspring. As previously discussed, elevated maternal inflammation during pregnancy is associated with altered neurodevelopment[252], adiposity[253] and chronic immune-mediated diseases[254] in offspring.

The most striking metabolic difference between MAL and NBS women came from host-microbe co-metabolites of protein putrefaction. Phenylacetylglutamine (PAG), *p*-cresyl sulfate (PCS), indoxyl sulfate (IS) and *p*-hydroxyphenylacetate (PHPA) were excreted in higher concentrations in women with a previous malabsorptive procedure compared to those with similar BMI but no history of bariatric surgery (**Figure 4.5a**). This urinary signature reflects increased bacterial fermentation of aromatic amino acids in the gut[41,328]. Malabsorptive surgeries result in decreased cleavage and absorption of dietary protein[329], hypothetically allowing more undigested protein to reach the large intestine where it is metabolised by resident bacteria, releasing free amino acids which are transformed to phenols and indoles by bacterial enzymes (see **Figure 1.2**). Hindered digestion and malabsorption of protein could also explain the reduction in circulating leucine and isoleucine observed in MAL patients since serum concentrations are directly related to the absorption of BCAAs in the small intestine[14]. In addition, reduced excretion of α-ketoisovalerate, a metabolite of valine, in MAL patients is consistent with impaired absorption of BCAAs. Increased protein putrefaction in the gut is generally considered detrimental to human health due to the potential toxicity of the phenolic and indolic compounds formed by this process[30,330,331]. One MAL-associated host-microbe co-metabolite, PAG, was significantly elevated in urine of neonates from the MAL group compared to neonatal urine from either NBS or RES groups (**Figure 4.5b**), suggesting the MAL metabolic phenotype has the potential to impact offspring. Due to the small number of infant urine samples collected in the present cohort, this observation should be treated with caution. However, investigation of the effects of increased fetal exposure to phenolic and indolic compounds is warranted.

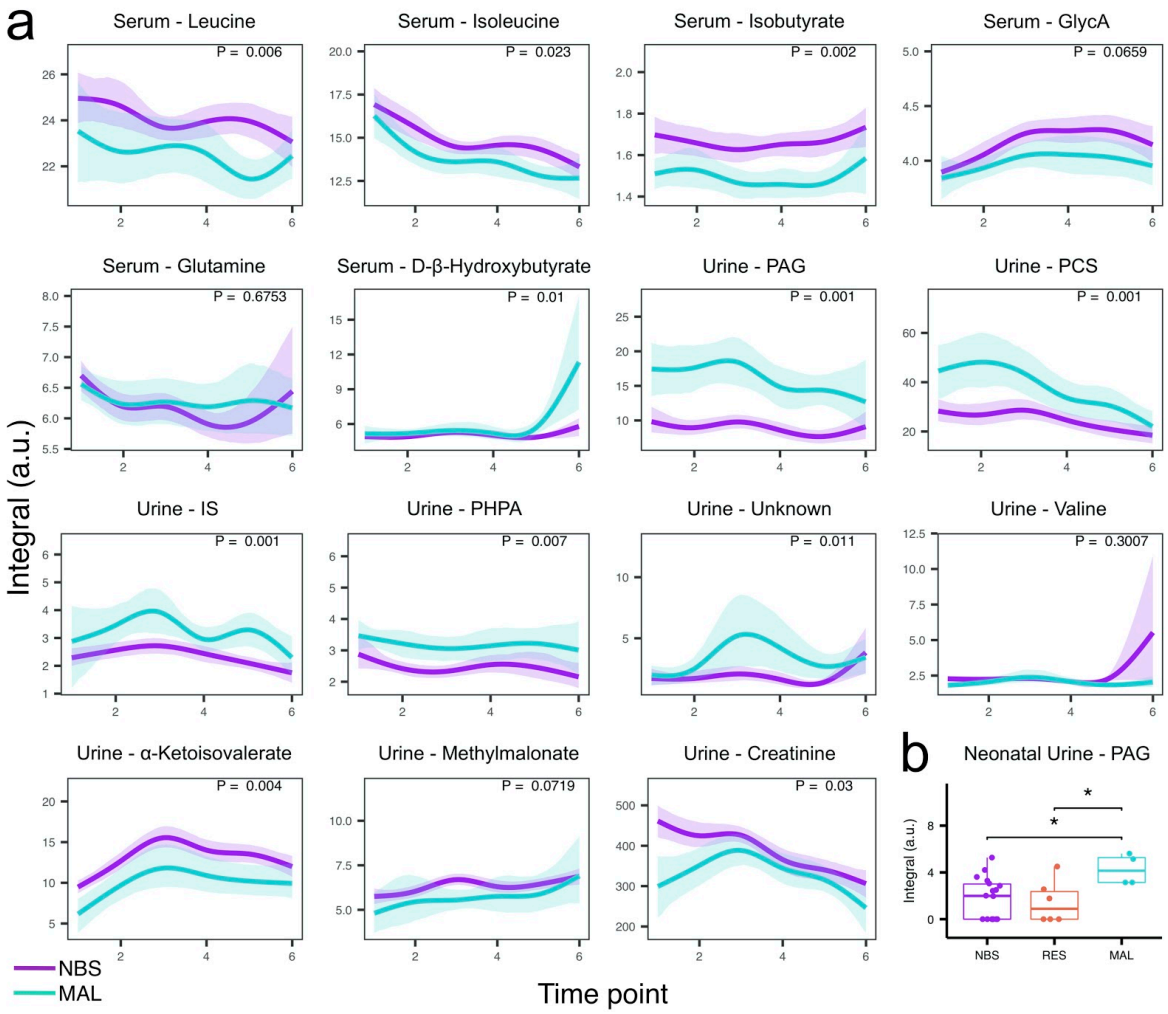


Figure 4.5: Longitudinal modelling of discriminant metabolites. a) For each metabolite identified in the cross-sectional analysis, significance was assessed by time series analysis ("santaR"). **b)** Relative concentrations of phenylacetylglutamine (PAG) in neonatal urine depend on maternal surgery status. *a.u. arbitrary units; *P<0.05 (Mann-Whitney U test); GlycA N-acetyl glycoproteins; IS indoxyl sulfate; PCS p-cresyl sulfate; PHPA p-hydroxyphenylacetate.*

The effect of previous malabsorptive surgery on maternal protein metabolism is seemingly biased toward aromatic amino acids (tyrosine, tryptophan and phenylalanine), which may reflect surgery-induced changes in the gut microbiota of MAL women. Isobutyrate, a BCFA which originates from bacterial metabolism of BCAA valine[332], was reduced in concentration in the MAL compared to the NBS group (**Figure 4.5a**). Little is known about the functional significance of BCFAs in human health, but unlike metabolites of aromatic amino acids, BCFAs are not generally considered to be toxic. Concentration differences in valine between MAL and NBS

patients were not detected during pregnancy, but urinary excretion of valine at the time of delivery increased in the NBS group; this trend was not observed in the MAL group. The opposite pattern was observed in the excretion of D-β-hydroxybutyrate, a ketone body. At the time of delivery, MAL patients excreted high concentrations of this metabolite but NBS patients did not. The two study groups clearly exhibit a differential response to labour. In particular, women with a previous malabsorptive surgery may be more prone to entering ketosis during labour. Despite the differential findings at T6, the metabolisms of the two study groups converge at this time point which was also evident in the cross-sectional OPLS-DA model (**Table 4.3**). The labour process itself induces metabolic changes[333] which likely overshadowed the malabsorptive signature. Urinary excretion of creatinine, which was much higher in NBS than MAL patients during the first two trimesters but not the third, shows that some pregnancy-related metabolic changes can also overshadow the malabsorptive signature during gestation. The trajectories of glutamine and methylmalonate exhibit the utility of collecting longitudinal metabolic data and analysing it as a time series. In addition to modelling the dynamics of metabolism, particularly during a period of drastic metabolic adaptations, the rate of identifying false positives is reduced.

4.3.7 Malabsorptive metabolic phenotype is associated with a shift in the gut microbiota

Gut bacterial profiles between NBS and MAL groups were compared at three time points during pregnancy which represented the end of the first (T1), middle of the second (T2) and middle of the third (T4) trimester. A single faecal sample from the NBS group was excluded due to low sequencing depth (2,181 reads). After the exclusion, the mean sequencing depth was 22,405 (±5,535 standard deviation) and the minimum sequencing depth was 12,967. MAL patients tended to have higher alpha-diversity than NBS patients (**Figure 4.6a**) although this was only statistically significant at T4. Gut microbial community structure was stable throughout pregnancy, but was substantially altered in the MAL compared to the NBS group (**Figure 4.6b**). The alteration was driven by increased relative abundance of three genera: *Enterococcus, Escherichia/Shigella* and *Streptococcus* (**Figure 4.6c**). The differential abundance of these genera manifested at all taxonomic ranks. *Rothia* and its family *Micrococcaceae* were also found in higher abundance in MAL patients while *Anaerostipes* was the only genus found in

lower abundance in these patients. Taxa which only displayed differential abundance at a single time point are not plotted, but are reported in **Table 4.S2** (Appendix B).

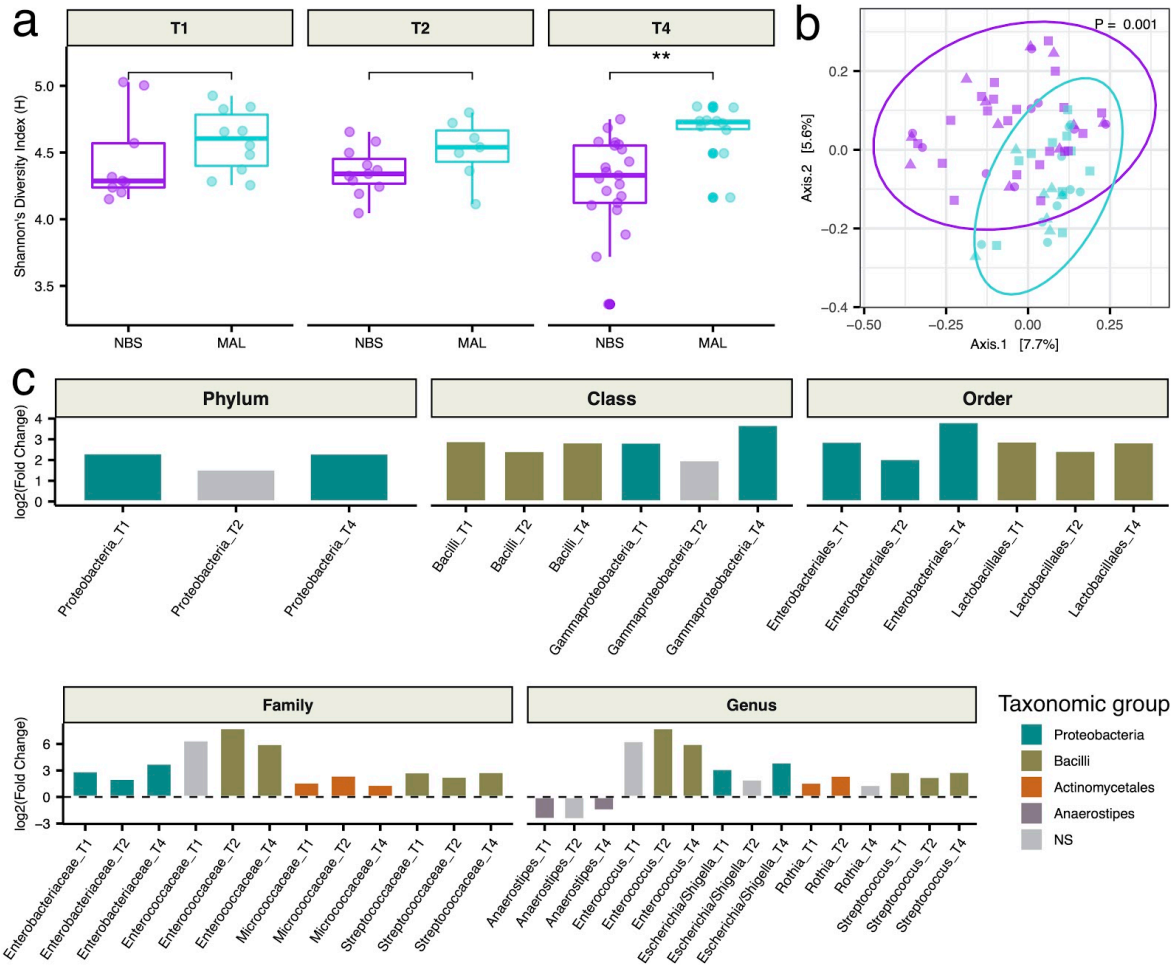


Figure 4.6: Gut bacterial alterations in malabsorptive patients. a) Alpha-diversity (Shannon) estimates for no bariatric surgery (NBS) and malabsorptive (MAL) groups at 11–14 (T1), 20–24, (T2) and 30–33 (T4) weeks gestation. **b)** Beta-diversity (Bray-Curtis dissimilarity) with principal coordinates analysis ordination. MAL (blue) and NBS (purple) patients at T1 (circle), T2 (triangle), and T4 (square) are plotted. *P* value obtained from permutational multivariate analysis of variance. **c)** Differentially abundant (Mann-Whitney U test; *FDR* <0.05) taxa in the MAL compared to the NBS group. ***P* <0.01 (Mann-Whitney U test); NS not significant.

The findings pertaining to the gut microbiota of pregnant MAL patients are consistent with previous reports of increased abundance of facultative anaerobes, especially *Gammaproteobacteria*, in faecal samples from non-pregnant patients post-gastric bypass surgery[311–316]. As discussed previously, the anatomical rearrangement of a gastric bypass results in decreased pH and increased oxygen in the gut as well as impaired digestion[310,311]. These factors contribute to the changes observed in the gut microbiota post-surgery. The current study is the first to demonstrate that the shift in the gut bacterial community induced by gastric bypass is persistent throughout pregnancy. Further studies to investigate the effects of the gastric bypass-induced gut microbiota on fetal health and offspring outcomes are warranted given the potential pathogenicity of the implicated bacteria. Increased relative abundance of facultative anaerobes, particularly enterobacteria, in the gut is associated with inflammatory bowel disease[334] and colorectal cancer[335]. *Enterobacteriaceae* are typically present in low densities in the healthy human gut, but blooms of *Enterobacteriaceae* are often observed in the inflamed gut which appears to be a favourable environment for this group of bacteria[336]. Here, the relative increase in *Enterobacteriaceae* was attributed to increased abundance of *Escherichia/Shigella* with some amplicon sequence variants identified as *E. coli* (**Figure 4.7**). Although some strains of *E. coli* have virulence properties, many are normal residents of the human gastrointestinal tract. However, some commensal *E. coli* strains can promote intestinal inflammation by triggering the release of proinflammatory cytokines[337].

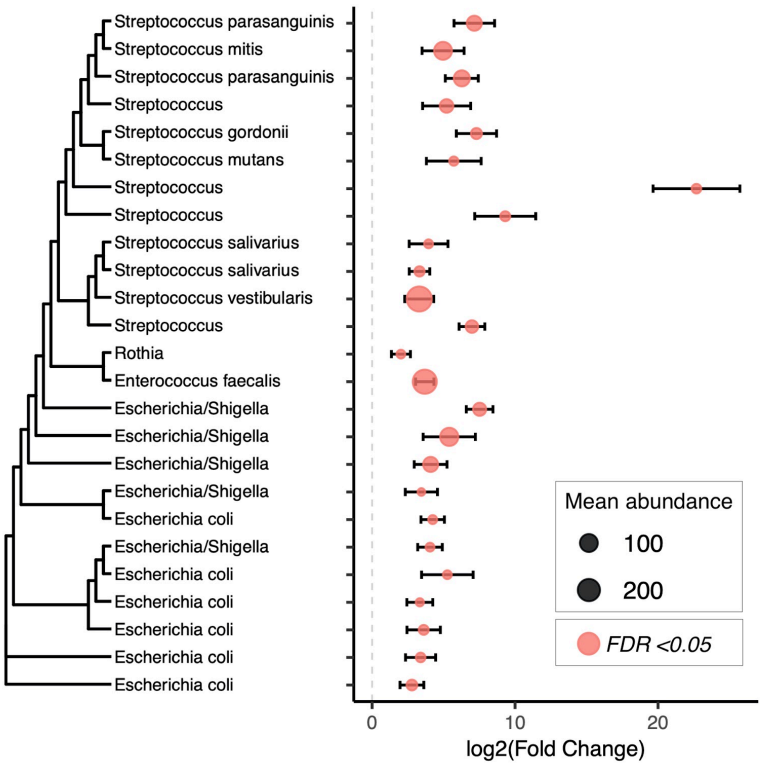


Figure 4.7: Differentially abundant amplicon sequence variants in malabsorptive patients. Amplicon sequence variants significantly (*FDR* <0.05) associated with the MAL group are plotted with their taxonomic assignments and phylogenetic relatedness. Points represent effect sizes with reference to the NBS group as well mean abundances (normalised). Black bars represent standard error.

Many different *Streptococcus* spp. were also more abundant in faecal samples from MAL compared to NBS patients (**Figure 4.7**). *Streptococcus* spp. are typically more abundant in the upper gastrointestinal tract than in the colon; they are dominant species in the oral cavity[338], stomach[339] and small intestine[340]. The implications for the translocation of these bacteria to the colon are unknown but like *E. coli*, some *Streptococcus* spp. are well-known pathogens. Of the 6 species identified in the present data, all have been implicated in various infections[341]. *Enterococcus faecalis*[342], a species closely related to *Streptococcus,* and *Rothia*[343] were both found in higher abundance in MAL compared to NBS patients and are also opportunistic pathogens. The pathogenic potential of the MAL-associated gut microbiota could contribute to the increased risk of developing colon cancer following bariatric surgery[344].

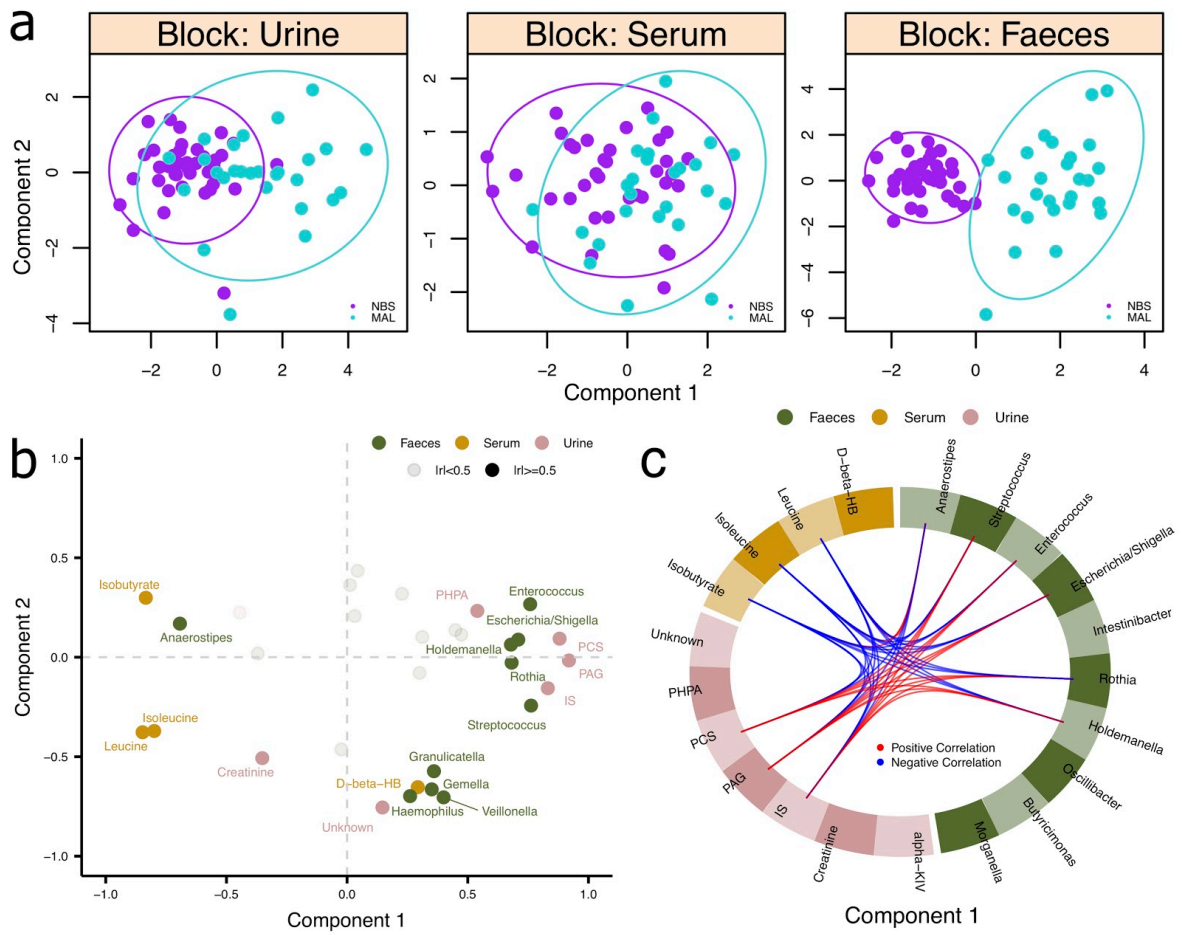


Figure 4.8: Integrated signature of previous malabsorptive surgery. a) Score plots from multi-block PLS-DA show samples from each biofluid type projected in latent space. Discrimination between no bariatric surgery (NBS) and malabsorptive (MAL) patients is achieved on component 1. **b)** Each variable (coloured by biofluid) from the model is plotted based on its correlation (Pearson) with each component. Variables with a high correlation on either component are highlighted. **c)** Pairwise correlations (see **2.4.1.4**) between variables along component 1 are plotted. *alpha-KIV, α-ketoisovalerate; D-beta-HB, D-β-hydroxybutyrate; IS, indoxyl sulfate; PAG, phenylacetylglutamine; PCS, p-cresol sulfate; PHPA, p-hydroxyphenylacetate.*

Data from the three biofluids were integrated using multi-block PLS-DA. Faecal samples provided the most discrimination between MAL and NBS patients, while serum samples provided the least discrimination (**Figure 4.8a**). A group of 5 bacterial genera (*Enterococcus, Escherichia/Shigella, Holdemanella, Rothia and Streptococcus*) were tightly clustered with the urinary protein fermentation metabolites (PAG, PCS, IS and PHPA), and this cluster was highly associated with the MAL group (**Figure 4.8b**). Genera and metabolites in the aforementioned cluster were negatively correlated with leucine, isoleucine and isobutyrate (**Figure 4.8c**), which

were highly associated with the NBS group. The multi-biofluid MAL signature demonstrates that the malabsorptive condition affects maternal protein metabolism, which, in addition to the anatomical and physiological changes to the gastrointestinal tract, likely favours the proliferation of specific bacteria in the colon. The species that bloom in the colon following malabsorptive surgery are presumably proteolytic and preferentially ferment aromatic amino acids tryptophan, tyrosine and phenylalanine, which give rise to indole, *p*-cresol and phenylacetate, respectively. PHPA can be produced from both tyrosine and phenylacetate[41]. *E. coli* and *E. faecalis* produce both phenols and indoles from tyrosine and tryptophan, respectively[345,346]. Following bacterial transformation of the aromatic amino acids, indole and *p*-cresol are sulfated to IS and PCS, respectively, while phenylacetate is conjugated to glutamine forming PAG. PAG and PCS were previously shown to correlate with the phylum *Proteobacteria* in general[347]. *Streptococcus* spp. have not been shown to produce the above metabolites, but these species have proteolytic activity[348] and are involved in the putrefactive process[349], which likely contributes to the MAL signature.

The data presented here is the first report of increased host-microbe co-metabolism of aromatic amino acids in humans post-malabsorptive surgery, a metabolic change that has been reported in gastric bypass-operated rodents[318,350]. The upregulation of bacterial protein fermentation following gastric bypass has been observed[312,313,315] through the sampling of faecal samples, but the present study has gone a step further, demonstrating an interaction between the altered microbiota and host biology. In addition, the integration of the three datasets unravelled the relationships between MAL-associated metabolites and the gut microbiota; it is evident that some metabolites are directly linked to changes in the gut microbiota following gastric bypass while others (e.g. creatinine, unknown and D-β-hydroxybutyrate) may be associated with other post-surgery factors (**Figure 4.8b**).

4.3.8 Clinical associations

Although phenolic and indolic compounds have known toxic properties[30,330,331], the present data indicate that the production of PAG, PCS and IS (to a lesser extent) may be associated with the maintenance of insulin sensitivity in MAL patients during pregnancy (**Figure 4.9**). A causal link between the gut microbiome and insulin sensitivity has been demonstrated through the use of faecal microbiota transplant (FMT) from lean donors to obese patients[105,320]. Both trials pointed

to SCFA-producing bacteria as the main drivers of improved insulin sensitivity following FMT. Conversely, a causal link between *Prevotella copri* and insulin resistance was demonstrated in mice, which the authors proposed was mediated through BCAA production[105]. The potential of the gut microbiota to influence glucose homeostasis is clear, though multiple mechanisms probably exist. The relationship between protein fermentation metabolites and insulin sensitivity in MAL patients likely reflects impaired protein digestion and the concomitant reduction in circulating BCAAs. However, the data indicates that the process of colonic protein putrefaction itself is more related to maternal insulin resistance than the presence of BCAA in pregnant MAL patients. PAG and PCS were negatively correlated with maternal insulin resistance at five of six time points while leucine and isoleucine were positively correlated with maternal insulin resistance at only two time points (**Figure 4.9**).

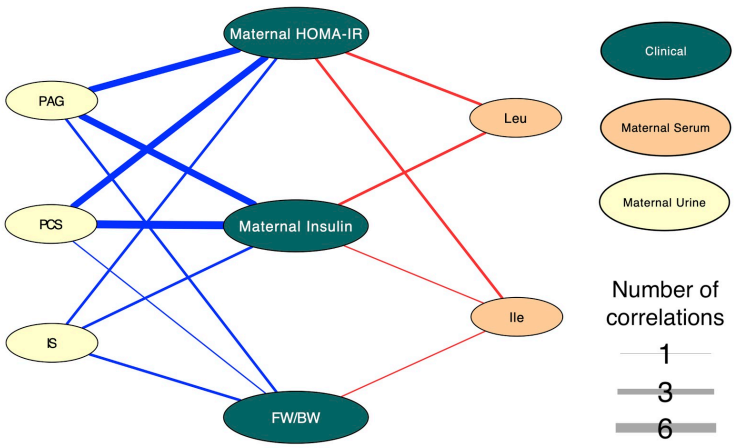


Figure 4.9: Protein-related metabolites correlate with maternal insulin resistance and fetal/birth weight. Network representation of Spearman correlation between maternal metabolite relative concentrations and clinical measurements. Edges represent significant (*FDR* <0.05) correlations at multiple (1-6) time points. Positive and negative correlations are denoted by red and blue lines, respectively. Values are reported in **Table 4.S3** (Appendix B). *PAG phenylacetylglutamine; PCS p-cresyl sulfate; IS indoxyl sulfate; Leu leucine; Ile isoleucine; HOMA-IR homeostatic model of insulin resistance; FW/BW fetal weight/birth weight.*

In addition to the association between maternal insulin resistance and maternal metabolites, estimated fetal weight and birth weight were correlated with PAG, PCS, IS and isoleucine (**Figure 4.9**). Glucose-stimulated insulin secretion is a key regulator of fetal growth[351]. Lower

maternal insulin resistance in MAL patients results in reduced maternal blood glucose levels, which may contribute to slower fetal growth[133,352]. However, the correlations between fetal/birth weight, PAG, PCS and IS remained significant after adjusting for maternal insulin resistance (**Table 4.S3**, Appendix B), indicating an independent association with the maternal gut microbiota, albeit weak. The nutritional deficiencies induced by malabsorptive procedures[353] may also affect fetal growth in these patients, but the reports of protein malnutrition and protein intolerance post-malabsorptive surgery[353,354] are important indicators that diet was not likely a significant contributor to the increase in protein fermentation metabolites observed in MAL compared to NBS patients. The urinary concentrations of these metabolites in the MAL group were also not related to the percentage of weight lost or the number of years since the surgery was performed (**Figure 4.S8**, Appendix B), providing evidence that the signature is persistent.

In summary, patients with a previous malabsorptive, but not restrictive, bariatric surgery exhibit a shift in the gut microbiota in favour of facultative anaerobes and bacteria typically found in the upper gastrointestinal tract. The malabsorption of protein and altered gut bacterial community in malabsorptive patients are associated with decreased serum concentrations of BCAAs and BCFA isobutyrate, as well as increased urinary excretion of protein putrefaction metabolites that arise from the fermentation of aromatic amino acids. The malabsorptive metabolic signature is highly correlated with maternal insulin resistance and weakly correlated with fetal growth. The data presented has considerable translational potential for the maintenance of an adequate level of insulin sensitivity during pregnancy by manipulating the gut microbiota. Since the bacteria implicated are potentially pathogenic, and the metabolic endpoints implicated are potentially toxic, a mechanism of action needs to be identified in order to develop a safe, targeted therapeutic.

Chapter 5

Metabolic and gut bacterial perturbations associated with Autism Spectrum Disorder

Maternal-fetal health is a key factor in obstetric outcomes, but also programs offspring for chronic conditions. Genetic predisposition combined with maternal health status, insults *in utero* and early-life environmental influences likely contribute to the development of Autism Spectrum Disorder (ASD) and related conditions. The high prevalence of gastrointestinal dysfunction in this population indicates a potential role of the gut microbiota in the pathophysiology of ASD, a relationship between the gut microbiota and ASD symptomatology, or both. This chapter details metabolic (urine and faecal) and gut bacterial (faecal) characteristics of a large multi-centre cohort of children with ASD (n=298), their neurotypical siblings (SIB; n=132) and unrelated neurotypical controls (NTC; n=122). The present cohort represented a heterogeneous group of children with ASD who displayed a spectrum of phenotypes potentially related to the degree of protein fermentation in the gut. The phenotype representing reduced gut protein fermentation was more prevalent in the ASD group compared to unrelated neurotypical controls and this observation was mirrored in the neurotypical siblings, reflecting shared environmental factors. While metabolic profiles were variable over time, gut bacterial profiles related to faecal amino acid composition were remarkably stable within an individual. Differentially abundant genera were observed between ASD and control children, and between subgroups of ASD children with different metabolic phenotypes. These findings support the gut microbiota as a potential target for ASD treatment but highlight the need for further research, and for adopting a precision medicine approach in this population.

5.1 Introduction

Since the first comprehensive description of Autism was published by Leo Kanner in 1943, the clinical definition of this complex disorder has continued to evolve. Early observations of autistic

children describe abnormal behavioural manifestations present from the first years of life: preference for “aloneness”, deficiency in social communication, obsessive or restricted interests, and a tendency toward ritualistic behaviours[355]. Other highly related neurodevelopmental conditions have since been described, including Asperger’s Disorder, Autistic Disorder, Child Disintegrative Disorder, and Pervasive Developmental Disorder – Not Otherwise Specified. The core behavioural symptoms of these disorders are generally similar but vary in severity. However, each patient will exhibit a unique set of clinical features, creating a vastly heterogeneous population representing a spectrum of cases. In practice, distinctions between the pervasive developmental disorders are difficult to ascertain, leading to inconsistent diagnoses among different clinics[356]. From 2013, the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) has incorporated all of these conditions under the single umbrella diagnosis of ASD[357], removing the ambiguity of ASD subtypes. Still, there is a clinical need to move beyond behavioural characterisation to stratify the patient population and provide targeted treatments that are formulated for individual cases or groups of children that exhibit similar phenotypes. A multifaceted systems biology approach must be undertaken to understand the biochemical mechanisms that contribute to the phenotypic diversity observed in ASD.

In the United States, the most recent estimates (from 2016) are that ASD affects 1 in every 54 children and is 4.3 times more prevalent in males than in females[358]. Worldwide, the prevalence rate is approximately 1%[359]. The cause of ASD is unknown, although it is likely multifactorial with contributions from both genetic and environmental factors. Genetically predisposed individuals might accumulate insults from prenatal and postnatal events such as maternal immune activation, maternal pharmaceutical drug usage, and exposure to pollutants or toxicants[360]. Comparison of monozygotic and dizygotic twin pairs has proven that the shared environment has a substantial effect on the concordance rates of ASD[361]. This complex interplay between genes, environment, and the immune system could also explain the heterogeneous manifestations and occurrence of comorbid medical conditions observed in the ASD population[362]. For patients with ASD, quality of life is significantly decreased by multiple comorbid afflictions. Autism has traditionally been viewed as a neurodevelopmental disorder, yet it is becoming clear that the entire body system is affected. These children commonly suffer from gastrointestinal issues and immune dysfunction as well as additional neurological conditions. Comorbidities can range from mood disorders, sleep disorders, anxiety, attention

deficit disorders, and epilepsy to autoimmunity, allergies, mitochondrial disorders and metabolic disorders[111]. The underlying systemic pathology of ASD and its associated conditions is complex, but at the intersection of neural, immune, and gastrointestinal physiology, sits an entity that acts as a significant mediator of several communication axes in the human body[363]. The gut microbiota is now considered a fundamental constituent of human biology, therefore it is probable that this entity plays a role in a variety of health conditions. It has also been shown that gastrointestinal dysfunction correlates with ASD symptom severity[113,114], providing evidence that the gut microbiota might be involved in the persistence of some autistic behaviours. It is possible that dietary patterns or eating behaviours common to these children[364] contribute to this phenomenon, however the evidence to date is inconclusive and must be investigated further.

Sufficient literature exists to support that there is some extent of gut microbial dysfunction in children diagnosed with ASD. However, a definitive ASD-associated signature has not been identified[115–119]. Due to the heterogeneity in this population, large sample sizes are necessary and within-group variation must be investigated. To explore host-microbe metabolic interactions, which vary across the ASD population and with respect to neurotypical children, a large multi-site cohort was studied. Faecal and urine samples were collected from children at multiple time points; metabolic (faecal and urine) and faecal bacterial profiles were acquired using proton nuclear magnetic resonance (^{1}H NMR) spectroscopy and 16S ribosomal RNA (rRNA) gene amplicon sequencing, respectively. The breadth of the data collected surpasses that of any other published cohort to date, allowing unprecedented phenotypic characterisation of ASD.

5.2 Materials and methods

5.2.1 Study population, sampling and clinical data

The population and design of the study are detailed in **2.1.2**. For all metadata collected, hypothesis tests were performed in a pairwise manner (NTC vs ASD, SIB vs ASD or NTC vs SIB). Differences in numerical variables were evaluated with Student's t-test or Wilcoxon rank sum (Mann-Whitney) test for normally and non-normally distributed data, respectively. Chi-squared or Fisher's exact (variables with fewer than 5 observations) tests were used for categorical variables. *P* values were adjusted according to the false discovery rate (*FDR*) and an alpha of 0.05 was used throughout. All calculations were executed in the R (version 3.5.0)

statistical environment[156]. A list of packages and package versions used can be found in **Appendix C**.

5.2.2 Metabolic profiling

Analysis of biospecimens. Metabolites were extracted from faecal samples using acetonitrile buffer (ACN):H_2O (1:3) containing 2% w/v sodium azide and bead-beating. Tubes containing 625 µl buffer, 250 mg faeces and 1.0 mm Zirconia beads (Stratech) were shaken for 45 sec in a Biospec bead-beater and centrifuged for 20 min at 17,000 x g and 4°C. Supernatant was filtered through 0.45 µm Spin-X® centrifuge tube filters (Corning) and stored at -80°C. Global metabolic profiles of faecal extracts and urine samples were acquired by ^{1}H NMR spectroscopy at the Imperial Phenome Centre (Imperial College London, UK). A standard one-dimensional pulse sequence with water presaturation was used[139]. J-resolved experiments were also run on each sample to assist with metabolite identification[142].

Modelling of metabolic profiles. Raw spectral data were pre-processed using an in-house pipeline[365]. Data were normalised using the probabilistic quotient normalisation method[146] and multivariate modelling (see **2.4.1** for detailed methods) was conducted on autoscaled data. Principal component analysis (PCA) was used to identify extreme outliers. Spectra with contaminants or drug metabolites were excluded. For all initial analyses, one time point from each participant was used. Orthogonal partial least squares discriminant analysis (OPLS-DA) was employed to identify group-wise differences in metabolic profiles. Models were evaluated using 7-fold cross validation. Sparse PCA (sPCA) was applied to ASD faecal metabolic profiles (**Figure 5.1**) to select microbiota-associated metabolites that varied across the ASD population. The sparsity introduced ensures that non-informative variables are not selected. Spectral variables with a correlation coefficient of |0.6| or higher on component 1 were annotated, and a representative peak for each metabolite was integrated to calculate relative concentration. To determine which urinary metabolites in ASD were associated with the selected faecal metabolites, a partial least squares (PLS) regression model of faecal metabolites (*X*) and urine ^{1}H NMR spectra (*Y*) was generated. For this model, “canonical” mode was used as the goal was not to predict the urine metabolites but to calculate correlations between the two datasets. Urinary spectral variables that correlated (≥|0.6|) with the faecal metabolites were annotated and metabolite relative concentrations were calculated by integrating a representative peak.

Multivariate models were calculated using the R package “mixOmics”[170], except for OPLS-DA which was calculated using “MetaboMate”.

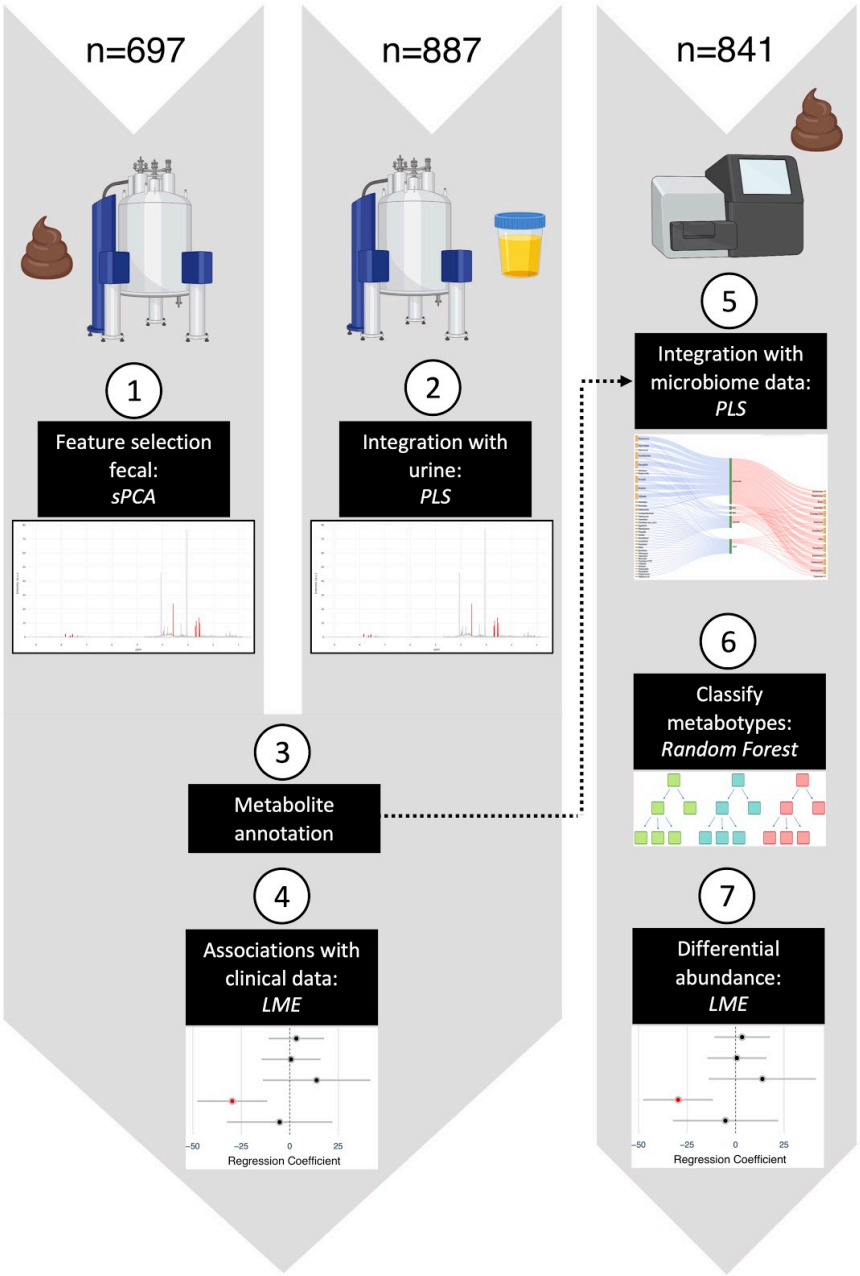


Figure 5.1: Overview of statistical workflow. Following spectral pre-processing, feature selection was performed using sparse principal component analysis (sPCA) on faecal data from the ASD group to identify metabolites that varied across the population **(1)**. Urine metabolites in ASD that were related to the selected faecal metabolites were identified using partial least squares (PLS) regression to integrate the two datasets **(2)**. Selected spectral features from

either dataset were annotated **(3)** and clinical associations with these metabolites and ASD comorbidities were assessed using linear mixed-effects (LME) models **(4)**. The metabolites were integrated with faecal microbiome data generated using 16S rRNA gene amplicon sequencing **(5)** to determine microbiota-metabolite relationships in the ASD population. A random forests classifier was trained on microbiome data to classify metabotypes (identified in faecal metabolite data) for each individual in the study **(6)**. LME models were used to determine bacterial genera associated with each study group and within each metabotype **(7)**.

Statistical analysis. Linear mixed-effects models (R package "lme4") were used to test the association between metabolites and potential confounding variables while controlling for the random effect of different recruitment centres. Gender, age, the presence of comorbidities (psychiatric, attention deficit, allergies, asthma, lower gastrointestinal), diet (consumption of meat or dairy) and dietary supplements (probiotics, melatonin, B12, fish oil, herbal supplements) were included as fixed effects in the model. *P* values were calculated with the Kenward-Roger approximation[182] implemented in "pbkrtest"[366] (see **2.4.3**). Faecal samples with ^{1}H NMR data were clustered into "metabotypes" (groups of metabolically similar individuals) using 2-means clustering on the relative concentrations of selected metabolites.

5.2.3 Gut bacterial community profiling

Details pertaining to the processing of faecal samples and subsequent sequencing of the 16S rRNA gene are provided in **2.3.2** and **2.3.3**. To investigate group-wide taxonomic differences in the gut microbiota, the first time point from each participant was modelled using OPLS-DA. Abundance data was filtered to remove taxa with less than 5% prevalence and a centred log-ratio transformation (with a pseudocount of 1 added to each value to avoid log(0)) was applied to account for compositionality[164] (see **2.3.5**). The multivariate model was calculated using the R package "MetaboMate" and was evaluated using 7-fold cross validation.

5.2.4 Data integration

Associations between bacterial genera and selected (faecal and urine) metabolites were assessed using PLS regression as implemented in "mixOmics". Abundance data was aggregated at genus level. Genera present in less than 10% of individuals were removed and a centred log-ratio transformation was applied. The multivariate model was calculated with microbiome data as the predictors (*X* matrix) and metabolite data as the outcomes (*Y* matrix). A relevance network was generated using a pairwise similarity matrix inferred from the model by

calculating the correlations between the variables projected in latent space. The resulting values are robust approximations of Pearson correlation[178] (see **2.4.1.4**). A Sankey diagram (package "networkD3") was used to visualise the relevance network.

5.2.5 Gut bacterial associations with study groups and metabotypes

Prediction of metabotypes from microbiome data. To further investigate the relationship between the gut microbiota and faecal metabolome, as well as the persistence of the observed phenotypes over time, a random forests (see **2.4.2**) model (package "randomForest") was trained on 16S rRNA sequencing data to classify faecal samples into two metabotypes (two clusters as defined in **5.2.2**). Abundance data were aggregated at genus level and filtered to include only genera present in at least 10% of samples then the centred log-ratio transformation was applied. A random forests model was trained on a dataset containing one random faecal sample from each individual. The model was first assessed based on the training data using permutation (n=1000) testing and leave-one-out cross-validation, then the model was tested on the remaining samples to answer two questions: 1) can the composition of the gut microbiota predict the metabolic phenotype in the children under study and 2) is the metabolic phenotype stable in these children over time? Statistics for model performance based on a confusion matrix of the observed and predicted classes were calculated with the R package "caret". *P* values were calculated by testing whether the overall accuracy rate was greater than the rate of the largest class, and by McNemar's test[367] which is based on the incorrect predictions.

Differential abundance according to study group and metabotype. The overall metabotype for each individual was determined based on metabotype classifications at all sampling time points for that individual. If classifications were not consistent across multiple sampling points, the metabotype was labelled "ambiguous". Linear mixed-effects models were calculated for each of the top 30 (based on mean decrease in accuracy) most important genera in the random forests classifier. First, the models were calculated with study group and age as fixed effects, but without including metabotype. Another set of models was generated, after excluding individuals with an ambiguous metabotype, to investigate the associations between genera abundances, study groups, age and metabotype as well as the interaction between study group

and metabotype. All models were adjusted for repeated sampling and different recruitment centres.

5.3 Results and discussion

5.3.1 Characteristics of the study participants

Children diagnosed with ASD (n=298), their neurotypical siblings (SIB; n=132) and unrelated neurotypical controls (NTC; n=122) were recruited from 72 offices located across 13 states (**Figure 5.2**). As expected, males were over-represented in the ASD group (**Table 5.1**) since males are 4.3 times more likely to develop ASD than females[368]. Children ranged from ages 1 to 14 and were significantly older in the ASD group than the neurotypical groups, likely because ASD is typically diagnosed around age 4.5 with less severe cases diagnosed even later[369]. Comorbidities and supplement consumption were more prevalent in the ASD compared to both SIB and NTC groups. The prevalence of attention deficit disorders and fish oil consumption were also higher in siblings compared to controls. Seizure disorders, attention deficit disorders and psychiatric disorders (e.g. anxiety, depression) are common comorbidities of ASD, and many children with ASD suffer from multiple concurrent comorbidities[370,371]. In addition to central nervous system deficits, ASD is also associated with immune[372] and gastrointestinal[111] dysfunction. In the current cohort, asthma was equally prevalent among ASD and NTC, but seasonal allergies were reported in a higher percentage of ASD children than both siblings and controls. Lower gastrointestinal afflictions (e.g. diarrhoea, constipation) were significantly higher in the ASD group than SIB and NTC groups which is consistent with current literature[111] and points to a potential link between ASD and the gut microbiota. In accordance, high probiotic usage was observed in the ASD group due to the common belief that some ASD symptoms are related to deficits in the gut microbiota[364]. High usage of melatonin was also observed in the ASD group, which fits with evidence that sleep disorders, which may be caused by deficits in melatonin synthesis[373], are common in these children[111]. Vitamins and other dietary supplements are often administered to children with ASD[374] as complementary and alternative medicine practices[375,376]. Fish oil, which contains omega-3 fatty acids that are important for neuronal development[377], was also commonly used in the ASD group. Evidence shows that supplementation of omega-3 fatty acids may improve some ASD symptoms including stereotypy and hyperactivity, although the effect size is small[378]. The characteristics of the present cohort

are indicative of the known heterogeneity of the ASD population. Each case is unique in its clinical presentation, core symptom severity[357] and pattern of comorbid medical conditions[362].

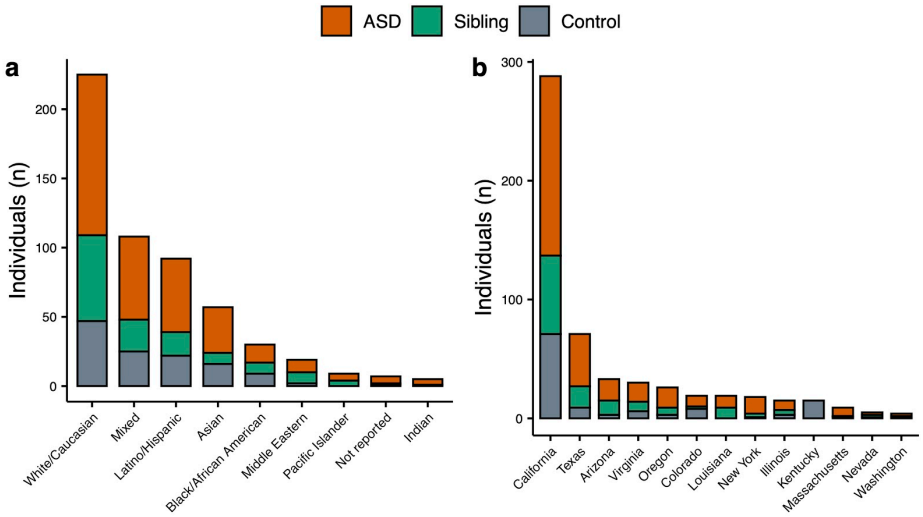


Figure 5.2: Participants by ethnicity and state. The number of participants is plotted for **a)** each reported ethnicity and **b)** each state where recruitment took place.

Table 5.1: Participant characteristics. Values reported are percentages except age which is presented as mean ± standard deviation. Statistical tests were performed as pairwise comparisons between study groups.

	Control (n=122)	**Sibling (n=132)**	***P value* (Sibling ~ Control)**	**ASD (n=298)**	***P value* (ASD ~ Control)**	***P value* (ASD ~ Sibling)**
Gender	54.1% Male	44.7% Male	0.170	84.6% Male	**8.89E-11**	**4.24E-17**
Age at first collection	4.8 ± 2.7	5.3 ± 2.8	0.156	5.9 ± 2.2	**6.65E-06**	**6.62E-03**
Ethnicity	38.5% Caucasian	47% Caucasian	0.218	38.9% Caucasian	0.975	0.166
Comorbidities and Medications						
Psychiatric comorbidities	0% Yes	3% Yes	0.054	10.7% Yes	**1.70E-04**	**7.41E-03**
Medication	0% Yes	2.3% Yes	0.248	10.1% Yes	**3.17E-05**	**5.06E-03**
Attention deficit disorder	0% Yes	3.8% Yes	**0.031**	16.1% Yes	**2.66E-06**	**3.04E-04**
Medication	0% Yes	1.5% Yes	0.499	10.1% Yes	**3.17E-05**	**1.09E-03**
Epilepsy	0% Yes	0% Yes	-	3.4% Yes	**0.041**	**0.034**
Medication	0% Yes	0% Yes	-	2.3% Yes	0.201	0.106
Asthma	7.4% Yes	2.3% Yes	0.056	7.7% Yes	0.906	**0.029**

Medication	7.4% Yes	2.3% Yes	0.075	7.7% Yes	1	**0.029**
Allergies	10.7% Yes	11.4% Yes	0.875	19.1% Yes	**0.035**	**0.043**
Medication	10.7% Yes	10.6% Yes	1	19.1% Yes	**0.049**	**0.040**
Upper GI conditions	0.8% Yes	0.8% Yes	0.960	4.4% Yes	0.067	0.053
Medication	0.8% Yes	0.8% Yes	1	4% Yes	0.120	0.074
Lower GI conditions	2.5% Yes	1.5% Yes	0.592	12.1% Yes	**1.69E-03**	**3.27E-04**
Medication	0% Yes	0% Yes	0.530	8.1% Yes	**2.92E-04**	**1.48E-04**
Antibiotics	0% Yes	0.8% Yes	1	2.3% Yes	0.201	0.444
Supplements						
Multivitamin	34.4% Yes	38.6% Yes	0.572	42.6% Yes	0.149	0.505
Probiotics	2.5% Yes	6.1% Yes	0.220	19.5% Yes	**1.04E-06**	**6.47E-04**
Antimicrobials	0% Yes	0% Yes	0.530	2.7% Yes	0.112	0.113
Melatonin	2.5% Yes	6.8% Yes	0.140	13.8% Yes	**3.22E-04**	0.056
Digestive enzymes	0.8% Yes	0% Yes	0.480	4.4% Yes	0.076	**0.012**
Folate	0.8% Yes	0.8% Yes	1	2.3% Yes	0.447	0.444
Vitamin B12	0% Yes	0.8% Yes	1	4.4% Yes	**0.013**	0.074
Fish oil	0.8% Yes	7.6% Yes	**0.011**	11.1% Yes	**1.14E-04**	0.347
Herbal supplements	0% Yes	0% Yes	0.530	4.7% Yes	**0.013**	**7.07E-03**
Iron	1.6% Yes	2.3% Yes	1	3.7% Yes	0.363	0.565
>1 Faecal samples donated	40.2% Yes	40.9% Yes		43.6% Yes		
>1 Urine samples donated	41% Yes	49.2% Yes		44.3% Yes		

ASD autism spectrum disorder; GI gastrointestinal

5.3.2 Sampling summary

From the 552 children recruited for the present study, 848 faecal and 887 urine samples were collected and analysed. Between 40-50% of the participants in each group donated samples on at least 2 separate occasions (**Table 5.1**). However, longitudinal sampling (more than 3 time points) was only obtained from ASD participants and one neurotypical sibling (**Figure 5.3**). 15 urine spectra and 52 faecal spectra were excluded for contamination or high concentrations of drug metabolites (see examples in **Figures 4-6**, Appendix A). From the sequencing data, two

samples were excluded for low sequencing depth (<5000 reads). Initial analyses were based only on one time point from each individual (ASD n=206, SIB n=97, NTC n=99). The earliest time point with all assays completed (faecal ^{1}H NMR, urine ^{1}H NMR and faecal 16S rRNA gene amplicon sequencing) was selected. Children without all assays completed for a single time point were excluded (n=150) from initial analyses, but were included in the metabotype analysis.

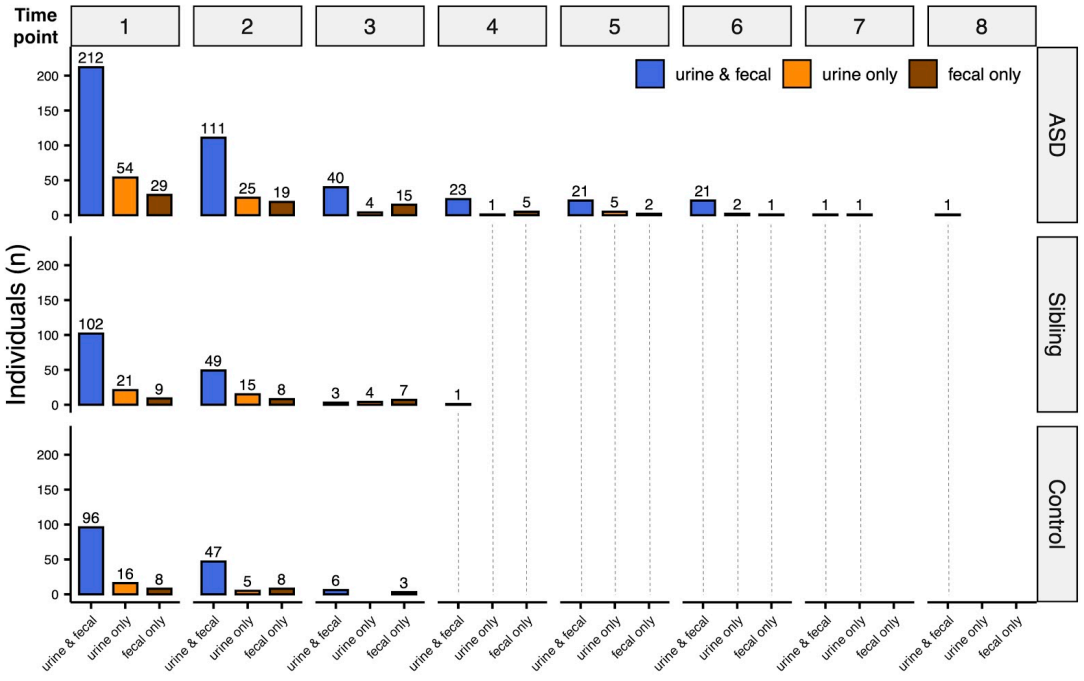


Figure 5.3: Sampling summary. The number of individuals who donated samples at each time point for each study group are detailed. Numbers of individuals who donated matched urine and faecal samples are reported separately from those who donated only one biospecimen type.

5.3.3 Metabolic and gut bacterial phenotypes are not overtly associated with ASD

Faecal and urinary metabolic profiles and gut bacterial profiles were compared between ASD and NTC groups by PLS-DA. Study groups could not be discriminated (**Figure 5.4**), as indicated by low Q^2Y values and large discrepancies between R^2Y and Q^2Y values in all models. Although previous studies have reported phenotypic differences in children with ASD compared to those with typical development using metabolomics (urine[120–125,379] and faeces[127,380]) and next generation sequencing (faeces[127,380–383]), consistent ASD-associated signatures have not been identified[115–119]. To date, published studies have investigated small cohorts ranging from 21 to

134 participants. In addition, many of the aforementioned studies have focused on more severe cases of Autism[121,125,127,382,384] and children with a variety of gastrointestinal issues. The current population is more than double the size of any published study and represents children across the spectrum. This presents a significant challenge for case-control classification given the heterogeneity of the disorder.

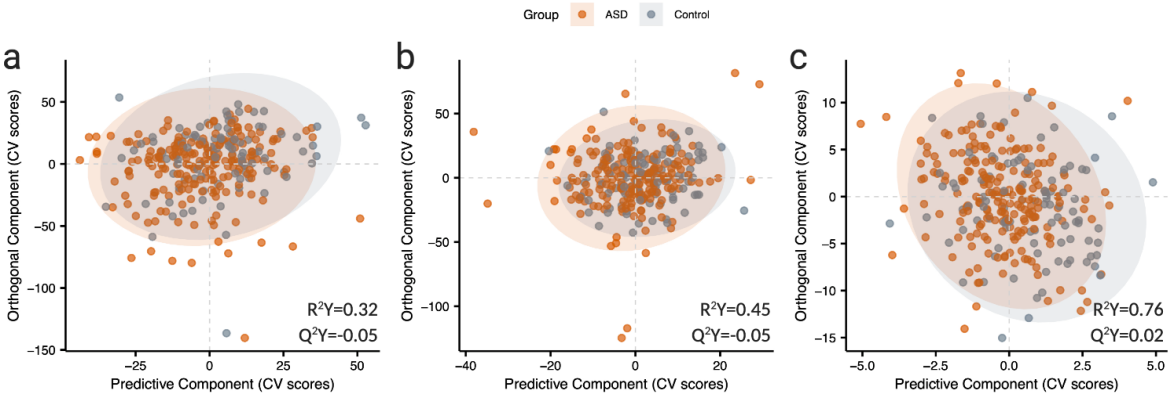


Figure 5.4: Cross-validated scores from orthogonal partial least squares discriminant analysis of metabolic and gut bacterial profiles. a) Faecal and **b)** urine ^{1}H NMR data as well as **c)** faecal 16S data were modelled to compare phenotypes between children with ASD and unrelated neurotypical controls. Discrimination between study groups was not achieved using any of these assays. R^2Y and Q^2Y metrics were calculated using 7-fold cross-validation (CV).

5.3.4 Faecal metabolome in ASD

Since ASD is a heterogeneous disorder potentially associated with disruption of the gut microbiota, variation in faecal metabolite profiles within this group of children was assessed by sPCA (**Figure 5.5**). The largest source of variation in ASD metabolic profiles, or the variables associated with component 1 (**Figure 5.6**), were signals from aromatic amino acids (tryptophan, phenylalanine and tyrosine), branched-chain amino acids (leucine, isoleucine and valine), glycine, lysine, aspartate, methionine, glutamate, alanine and uracil (see **Table 3** in Appendix A for annotation details). Due to overlapping signals in the ^{1}H NMR spectra, relative concentrations of lysine were not calculated. Uracil, the only non-amino acid metabolite selected, is a pyrimidine that has many functions in the human body. Studies in mice have shown that faecal concentrations of uracil decrease significantly after antibiotic treatment[385,386], suggesting this metabolite is also associated with the gut microbiota. Amino acids in the large intestine arise from bacterial biosynthesis as well as bacterial catabolism of protein[44,288]. Gut bacteria preferentially ferment carbohydrates over proteins[40,387], but as carbohydrate availability

decreases along the length of the large intestine, proteins become important substrates[25,26]. These proteins, endogenous or those that have escaped host primary digestion, are subjected to bacterial proteases[348,388] and peptidases[389] then free amino acids can be further fermented to short-chain fatty acids (SCFAs), phenols, indoles, branched-chain fatty acids (BCFAs) and amines[30,31,328], which are absorbed into circulation or utilised by colonic epithelial cells[390].

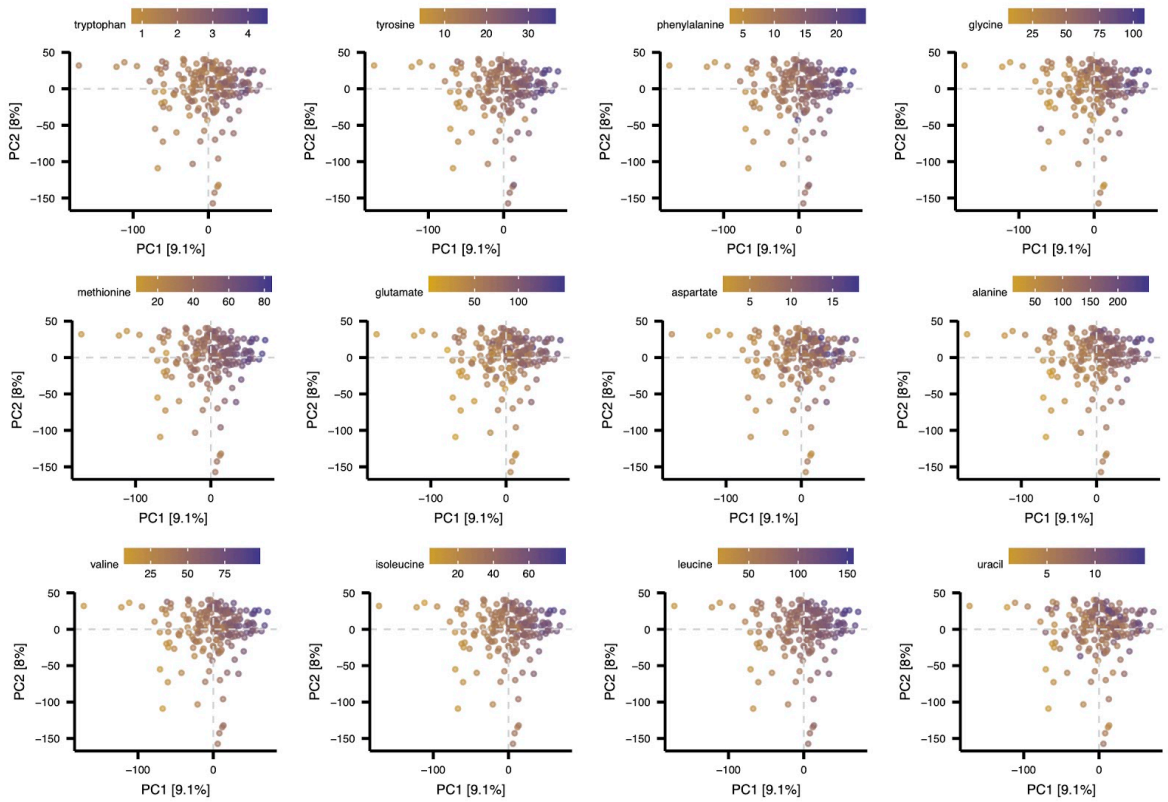


Figure 5.5: Sparse principal component analysis of faecal ^{1}H NMR data from children with ASD. Score plots are shown with points coloured by metabolite relative concentrations in arbitrary units. A gradient of amino acid concentrations can be observed within the ASD group. *PC principal component.*

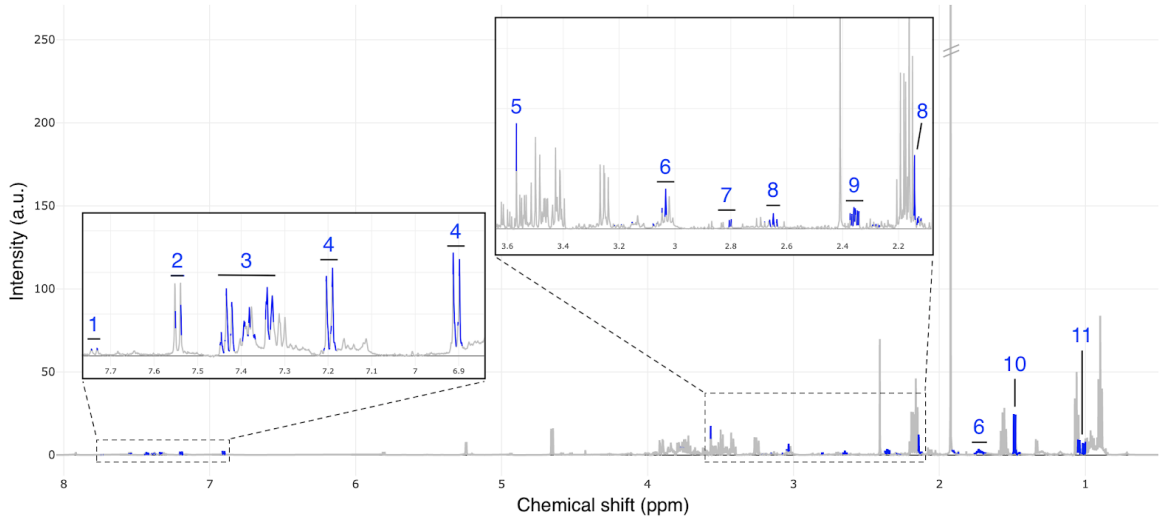


Figure 5.6: Faecal ^{1}H NMR spectrum. A representative faecal ^{1}H NMR spectrum from the ASD group with features that correlate with the first component of the ASD sPCA model highlighted (blue). See **Table 3** in Appendix A for annotation details. *1 tryptophan; 2 uracil; 3 phenylalanine; 4 tyrosine; 5 glycine; 6 lysine; 7 aspartate; 8 methionine; 9 glutamate; 10 alanine; 11 branched-chain amino acids (leucine, isoleucine, valine); a.u. arbitrary units.*

Of the faecal amino acids identified in the ASD group, aspartate, glycine and glutamate can act as neurotransmitters. In the human gut, these three amino acids can be produced by bacteria, are readily fermented by bacteria and are incorporated into bacterial cells[44], affecting their bioavailability. Glutamate is also a precursor to the primary inhibitory neurotransmitter 4-aminobutryate (GABA) which can be produced via host or bacterial decarboxylation of glutamate[391]. Supplementation with a GABA-producing probiotic strain of *Bifidobacterium dentium* has been shown to attenuate visceral hypersensitivity in rats[392], a phenotype that is associated with functional bowel disorders. Another study showed a *Lactobacillus brevis* strain producing large amounts of GABA rescued depressive behaviours in mice[393]. Furthermore, administration of a *Lactobacillus rhamnosus* strain to healthy mice increased glutamate, glutamine and GABA concentrations in the brain[394]. The same strain induced neurochemical and behavioural changes in mice that were dependent on the vagus nerve[395], a major gut-brain communication route[110]. Other studies have shown that germ-free mice have altered concentrations of BCAAs, aromatic amino acids, glutamine, aspartate, arginine, alanine, cysteine and serine in the brain compared to colonised mice[396,397]. The aromatic amino acid tryptophan is another neurotransmitter precursor, and serotonin production from tryptophan is regulated by the gut microbiota[398]. Collectively, this evidence strongly suggests that bacterial

modulation of amino acids and amino acid-derived molecules plays an important role in the gut-brain axis.

The faecal metabolic profiles in the present cohort unfortunately could not be attributed to symptom severity due to the absence of this data, but other factors were investigated using a linear mixed-effects model to account for variation due to different recruitment centres. Low faecal amino acid levels in children with ASD were significantly associated with the presence of psychiatric comorbidities (e.g. anxiety and depression) as well as fish oil consumption (**Figure 5.7**). Of the 206 ASD children included in the model, 20 had reported psychiatric comorbidities and 26 reported consumption of fish oil. Only one child was reported to have both psychiatric comorbidities and fish oil consumption. It is important to note that 18 of the 20 children with reported psychiatric comorbidities were taking prescription medications which may have influenced the results. Fish oil consumption is also a potential confounder. Over 50% of the modelled variation in faecal metabolic profiles of ASD children could be attributed to this variable. Previous reports show that fish oil supplementation mainly affects lipid metabolism[399–401]. To date, there are no reports detailing changes in faecal amino acids due to fish oil consumption.

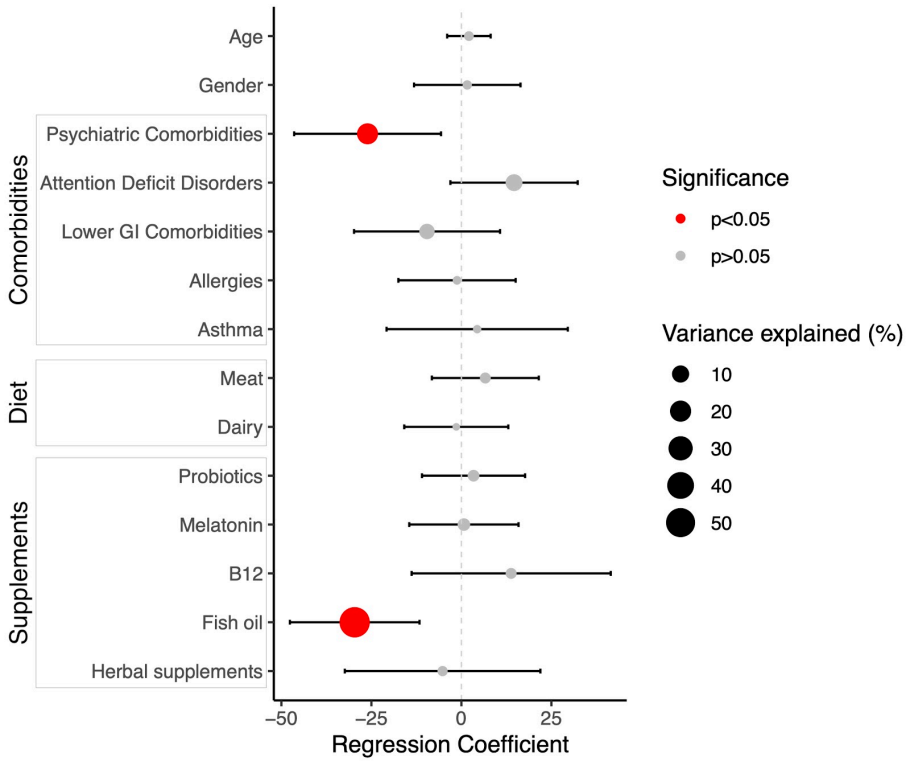


Figure 5.7: Associations between faecal metabolites and ASD characteristics. Regression coefficients from a linear mixed-effects model are plotted with 95% confidence intervals. Recruitment centre was included as a random effect. Coefficients coloured in red are significant (*P <0.05*). Size of points represents the percentage of variance explained by each fixed effect.

Faecal metabolic profiles are markedly influenced by the microorganisms residing in the large intestine. A study of 786 adult twins (predominantly female) showed that almost 70% of the variance in the faecal metabolome could be explained by the composition of the faecal microbiota and this was particularly true for amino acids[402]. Variance in concentrations of some amino acids could be 100% explained by the faecal bacterial community. Elevated concentrations of amino acids could indicate increased bacterial biosynthesis or decreased utilisation of amino acids while reduced concentrations could indicate decreased bacterial biosynthesis or increased utilisation, although more complex metabolic interactions are likely occurring within the community. Currently, there is no evidence that amino acids are reabsorbed in large quantities in the human large intestine[403], so this factor presumably did not contribute to the observed ASD faecal metabolic phenotypes.

Changes in faecal amino acid concentrations have been reported in paediatric inflammatory bowel disease (IBD) patients. Treatment-naïve children with IBD had higher faecal concentrations of histidine, tryptophan, phenylalanine, leucine, valine, and serine compared to children without IBD[404]. In another study of paediatric Crohn's Disease, one of the two main IBDs, nearly all faecal amino acids detected were higher compared to controls; these metabolites positively correlated with intestinal inflammation and were not related to protein intake[405]. Higher concentrations of faecal amino acids[406,407] and BCFAs[406] have also been reported in colorectal cancer patients. In the current cohort, faecal amino acid levels were not associated with gastrointestinal dysfunction (**Figure 5.7**) but these data were reported indirectly through the primary caregivers of the children, so it is possible that many cases were missed, especially given the communication deficits present in this population[408]. Given that the estimated prevalence of gastrointestinal problems in children with ASD is between 40%[114] and 60%[409], it can be assumed that these are under-reported (12.1%) in the present cohort. ASD children with high faecal amino acids also had high concentrations of faecal uracil. Pathogenic bacteria such as *Vibrio fluvialis, Klebsiella pneumoniae, Shigella sonnei, Pseudomonas aeruginosa*, and *Serratia marcescens* produce uracil[410]. In drosophila, pathogens but not commensals, produce uracil which activates the immune system and causes inflammation[410].

5.3.5 Urinary metabolites related to the ASD faecal metabolome

To determine if shifts in faecal amino acid levels in the ASD group resulted in changes in host-microbe co-metabolism, as reflected in urine samples, PLS regression was employed. Only two urinary metabolites, phenylacetylglutamine (PAG) and *p*-cresyl sulfate (PCS), were associated (positive) with the selected faecal amino acids and uracil in ASD children (**Figure 5.8**). PAG and PCS are microbiota-associated metabolites of phenylalanine and tyrosine, respectively, that are produced via protein fermentation in the colon[41]. Bacterial enzymes convert phenylalanine to phenylacetate[345] which is conjugated to glutamine by host metabolism before excretion. Tyrosine is converted to *p*-cresol by bacterial enzymes[345] and *p*-cresol is sulfated to PCS prior to excretion. Urinary concentrations of these phenolic compounds are strong indicators of gut bacterial activity[32,411].

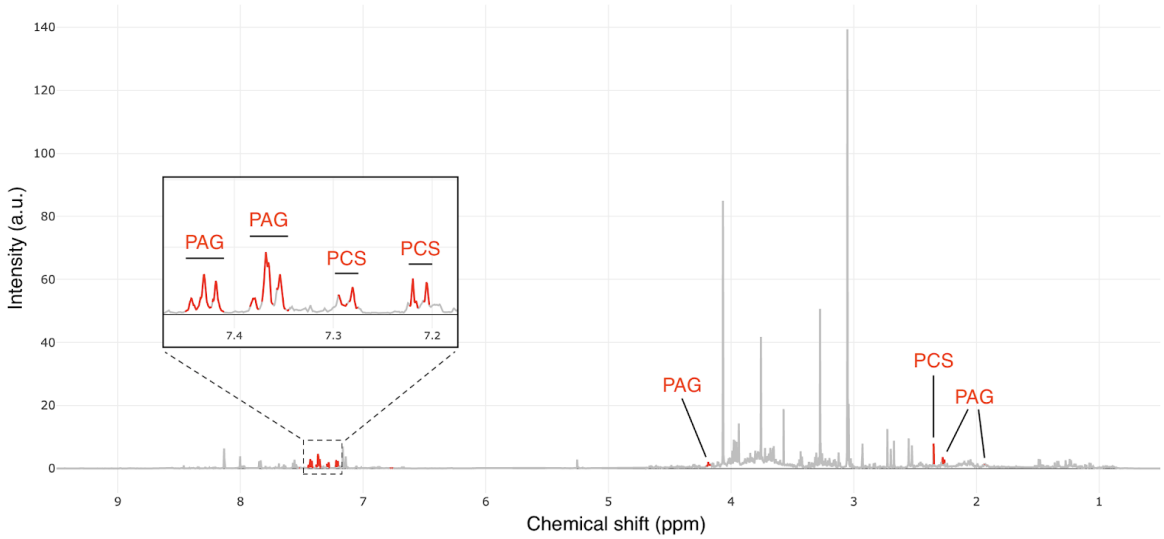


Figure 5.8: Urine ^{1}H NMR spectrum. A representative urine ^{1}H NMR spectrum from the ASD group with selected features highlighted (red). Urinary phenylacetylglutamine (PAG) and *p*-cresyl sulfate (PCS) correlate positively with faecal metabolites in the ASD group. *a.u. arbitrary units.*

Linear mixed-effects models were used to identify factors associated with urinary PAG and PCS excretion in children with ASD. Both metabolites were present in higher concentration in children with seasonal allergies (**Figure 5.9**). Of the 206 children included in the models, 33 children were reportedly affected by allergies and 16 of these children were taking an allergy medication during the study. PCS, but not PAG, was also higher in children (n=31) whose caregiver reported melatonin usage. For 5 children, both allergy medication and melatonin supplementation were reported. Although the cause of high excretion of PAG and PCS in some ASD children cannot be determined in the present study, urinary concentrations of these metabolites have been shown to increase with protein intake[27,29,291] and decrease with fibre intake[29]. A high protein, low carbohydrate diet has also been shown to increase faecal phenolic compounds including phenylacetate[330]. However, intestinal transit time can affect urinary concentrations of PAG and PCS independent of diet[412]. Delayed colonic transit results in depletion of carbohydrates and a switch to protein fermentation, but is also associated with a change in the faecal microbiota[413]. It has also been proposed that the formation of PAG and other microbiota-associated molecules is an important regulation of neurotransmitter levels in the human body[414]. Amino acids such as glycine, glutamate and glutamine are conjugated to aromatic acids prior to excretion thus decreasing their availability to the host. Altered urinary concentrations of PAG and PCS have previously been detected in children with ASD[120,121,126,415], in obese adults[99], in centennials[347] and in patients with chronic kidney disease[416]. Serum

concentrations of these metabolites have been associated with cardiovascular disease in renal patients[291,417]. PAG and PCS, although affected by diet, are reflective of gut microbial activity, are clearly associated with human health and are indicative of certain phenotypes.

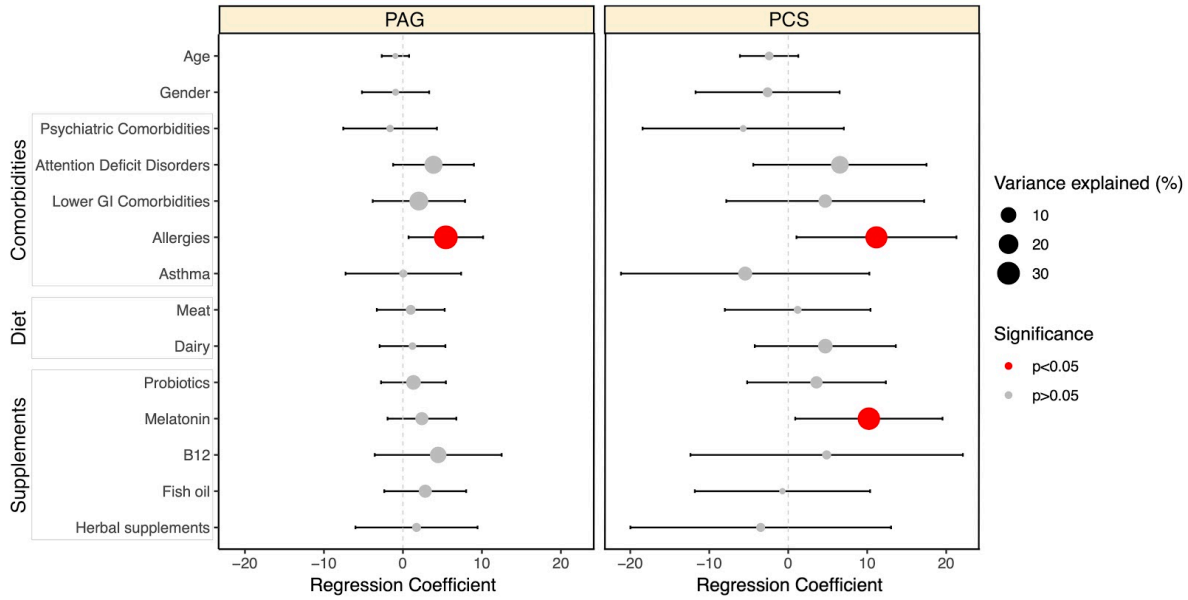


Figure 5.9: Associations between urinary metabolites and ASD characteristics. Regression coefficients from linear mixed-effects models are plotted with 95% confidence intervals. Recruitment centre was included as a random effect. Coefficients coloured in red are significant (*P <0.05*). Size of points represents the percentage of variance explained by each fixed effect. *PAG phenylacetylglutamine; PCS p-cresyl sulfate.*

In the current cohort, children with ASD displayed a spectrum of metabolic phenotypes. On one end of the spectrum, children had low faecal amino acid and low urinary PAG/PCS concentrations. Children on this end of the metabolic spectrum also tended to have psychiatric comorbidities and to consume fish oil as a supplement. On the other end of the metabolic spectrum, children had high faecal amino acid and high urinary PAG/PCS concentrations. These children tended to suffer from seasonal allergies and potentially sleep disturbances, as inferred by the prevalence of melatonin supplementation. One likely explanation for the metabolic findings is variation in intestinal transit time. Many children with ASD suffer from chronic constipation[114], a hallmark of delayed colonic transit which results in higher bacterial protein fermentation[412]. This theory is supported by a report of high PCS excretion that was related to increased intestinal transit time in children with ASD[126]. High faecal concentration of *p*-cresol was also detected in a group of children with ASD and high gastrointestinal severity

symptoms compared to neurotypical children with low gastrointestinal severity symptoms[381]. Conversely, decreased intestinal transit time and consequent reduction in bacterial protein fermentation in the gut, could also explain the lower concentrations of faecal amino acids, urinary PAG and urinary PCS observed in some children. The consistency of a faecal sample, an indicator of intestinal transit time, is also one of the main explanatory variables for the faecal microbiome[86]. However, given the association between metabolic phenotype and different metadata covariates, the metabolic heterogeneity within the ASD group likely reflects many factors. In addition to the inherent symptomatic diversity and the presence of gastrointestinal dysfunction in a subset of the population, children with ASD often have difficult eating behaviours[418,419]. Diet is a major determinant of gut microbial community structure[11,420] and therefore faecal metabolome[402,420], as well as urinary metabotype[420,421]. Although metabolic phenotypes in ASD may be confounded by the consumption of supplements and medications[422], multi-biofluid systems-level analysis provides a depiction of the population encompassing endogenous physiology and differential behaviours, as well as environmental and microbial influences, capturing the heterogeneity of ASD.

5.3.6 Gut bacteria related to ASD metabolome

Given the proposed link between metabolic variation within the ASD group and microbiota-associated metabolism, relationships between faecal/urinary metabolites and bacterial genera detected in faecal samples were assessed using PLS regression. Of the 191 genera detected, 77 were included in the model after prevalence filtering. Major saccharolytic bacteria in the human gut[7,423] such as *Roseburia* spp., *Faecalibacterium* spp., and *Prevotella* spp. were negatively correlated with both faecal amino acids and urinary phenols while amino acid-fermenting bacteria such as *Clostridium* spp.[40,424] were positively correlated with these metabolites (**Figure 5.10**). This finding supports the idea that the observed metabolic spectrum in children with ASD is related to the trade-off between carbohydrate and protein fermentation, which may reflect intestinal transit time. High faecal amino acids and urinary phenols in this population were also associated with increased relative abundances of genera *Dorea* and *Ruminococcus.* Low *Faecalibacterium* with high *Clostridium* and *Ruminococcus* is a faecal signature of irritable bowel syndrome (IBS)[425]. All genera positively associated with faecal amino acids and urinary phenols belong to just three families (*Ruminococcaceae, Lachnospiraceae*

and Clostridiales Incertae Sedis XIII) in one order (*Clostridiales*). Faecal abundances of *Clostridiales* taxa have been associated with increased intestinal transit time[86,426].

All faecal amino acids exhibited consistent relationships with bacterial genera except glutamate. A large number of taxa were negatively correlated with glutamate but were not correlated with other amino acids (**Figure 5.10**). The present data suggest that reduced faecal glutamate in some ASD children is associated with increased abundance of many genera in the gut which could be due to a reduction in lactic acid bacteria which are known to be glutamate-producers[427]. Downregulation of gut bacterial glutamate metabolism and reduced faecal amino acids have previously been reported in children with ASD compared to typically developing children and this was unrelated to diet[380]. Glutamate is an important substrate for many bacterial species and is one of the main amino acids incorporated into bacterial cells[44], so reduction in faecal glutamate could also indicate increased bacterial utilisation. As previously discussed, glutamate is an important neurotransmitter and GABA precursor. Increasing this metabolite through modulation of the gut microbiota[392–395] could benefit a subgroup of children with ASD. In support of this theory, lower faecal concentrations of GABA were reported in children with Pervasive Developmental Disorder - Not Otherwise Specified, a subtype of ASD, compared to neurotypical children and those with classical autism[127]. Uracil shared correlation patterns with glutamate, but had little overlap with the other amino acids. Uracil degradation is a common pathway in bacteria[428,429], but the relevance of the pathway to host health is unknown.

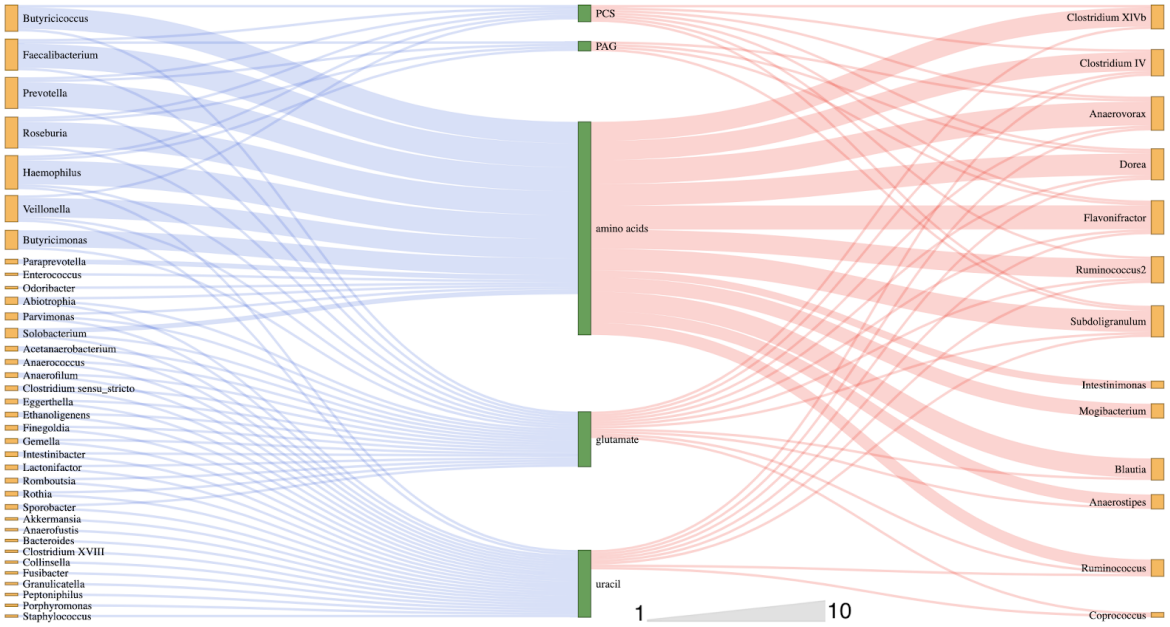


Figure 5.10: Associations between metabolites and microbiota in ASD. Significant (*FDR* <0.05) associations between bacterial genera (orange) and metabolites (green) are plotted with negative and positive associations represented by blue and red ribbons, respectively. Data for amino acids except glutamate are summarised; the ribbon width represents the number of amino acids (1-10) with significant associations. See **Table 5.S1** (Appendix B) for tabulated values. *PAG phenylacetylglutamine; PCS p-cresyl sulfate.*

5.3.7 Host-microbe co-metabolism is altered in children with ASD compared to neurotypical participants

Metabolite (tryptophan, phenylalanine, tyrosine, leucine, isoleucine, valine, glycine, lysine, aspartate, methionine, glutamate, alanine and uracil) concentrations in each faecal sample were summarised by PCA to compare signatures between study groups. The first principal component explained 77.4% of the variance in these metabolites. Linear mixed-effects models were calculated to explore relationships between faecal metabolite concentrations, urinary PAG, urinary PCS and covariates in the SIB and NTC groups while controlling for the random effect of different recruitment centres. The only significant association observed was between faecal metabolite concentrations and age (negative) in the NTC group (**Table 5.2**). This association was not present in the SIB or the ASD group (see **Figures 5.7** and **5.9**).

Table 5.2: Associations between metabolite levels and characteristics in neurotypical individuals. Statistical output from linear mixed-effects models in neurotypical siblings or unrelated neurotypical controls. Metadata variables with less than 5 observations in a group were omitted from the models.

Group	Response	Fixed Effect	Regression Coefficient	Standard Error	*P*	Group	Response	Fixed Effect	Regression Coefficient	Standard Error	*P*
Sibling	Faecal metabolites	Age	0.037	0.318	0.910	Control	Faecal metabolites	Age	-1.056	0.324	**0.002**
		Gender	-0.095	0.707	0.897			Gender	-0.140	0.634	0.828
		Allergies	0.419	0.927	0.659			Allergies	-0.424	1.075	0.700
		Meat	1.113	1.047	0.302			Meat	1.660	1.213	0.181
		Dairy	-0.693	1.003	0.498			Dairy	0.340	1.397	0.814
		Probiotics	1.961	1.302	0.147			Asthma	-1.993	1.282	0.129
		Melatonin	-0.386	1.165	0.746						
		Fish oil	0.121	1.419	0.934						
Sibling	PAG	Age	0.006	0.020	0.746	Control	PAG	Age	0.011	0.023	0.633
		Gender	0.033	0.043	0.458			Gender	-0.015	0.045	0.743
		Allergies	-0.013	0.057	0.828			Allergies	-0.058	0.072	0.433
		Meat	0.054	0.063	0.406			Meat	0.156	0.088	0.086
		Dairy	-0.038	0.062	0.542			Dairy	-0.068	0.090	0.459
		Probiotics	0.072	0.079	0.380			Asthma	0.055	0.091	0.556
		Melatonin	0.051	0.070	0.478						
		Fish oil	-0.008	0.086	0.932						
Sibling	PCS	Age	0.017	0.020	0.425	Control	PCS	Age	-0.008	0.024	0.760
		Gender	0.053	0.045	0.252			Gender	0.057	0.047	0.247
		Allergies	-0.035	0.059	0.561			Allergies	0.030	0.076	0.697
		Meat	0.099	0.066	0.148			Meat	0.088	0.092	0.351
		Dairy	0.021	0.064	0.746			Dairy	0.058	0.094	0.545
		Probiotics	0.061	0.082	0.473			Asthma	0.107	0.095	0.274
		Melatonin	0.062	0.073	0.405						
		Fish oil	-0.038	0.090	0.682						

PAG phenylacetylglutamine; PCS p-cresyl sulfate

Children with ASD and their neurotypical siblings had significantly lower faecal concentrations of amino acids and uracil compared to the NTC group (**Figure 5.11a**). In spite of the association between faecal amino acids/uracil and age in the NTC group, age did not have a significant

effect on group differences in these metabolites; group designation accounted for over 80% of the variation explained in the model (**Figure 5.11b**). Paired analysis revealed that ASD and SIB groups had similar faecal metabolite profiles (**Figure 5.11c**). Gut bacterial compositions of related individuals tend to be more similar than those of unrelated individuals[87] and this is especially true for individuals living together[430]. Therefore, it is not unexpected to see highly similar faecal metabolomes in siblings who likely eat similar meals. In addition, global urinary metabolomes of children with ASD and their neurotypical siblings have previously been shown to be highly similar although single metabolite concentrations varied between the two groups[124]. In spite of the similarity in faecal metabolite profiles observed between ASD and SIB groups in the present cohort, the correlations between faecal metabolites and urinary PAG (**Figure 5.11d**) or PCS (**Figure 5.11e**) that were observed in the ASD group were not present in the SIB group. These relationships were also not found in the NTC group. This result indicates that the degree of protein fermentation in the gut varies in children with ASD, potentially due to gastrointestinal status, and this variation is not present in the neurotypical population of the present cohort. Bacterial fermentation of protein in the gut produces both phenols (such as PAG and PCS) and indoles[41]. Previous studies focused on discriminating ASD versus neurotypical metabotypes have found higher excretion of phenolic and indolic compounds in both urine[120–122,126,415] and faecal[127,381] samples from ASD children. However, at least two studies have reported lower excretion of phenolic and indolic compounds in urine from children with ASD compared to neurotypical children[123,124]. At least two studies have reported no statistically significant differences in phenolic and indolic compounds when comparing urine samples from children with and without ASD[122,379,431]. To avoid this inconsistency, gastrointestinal problems should be considered when investigating ASD metabolic phenotypes.

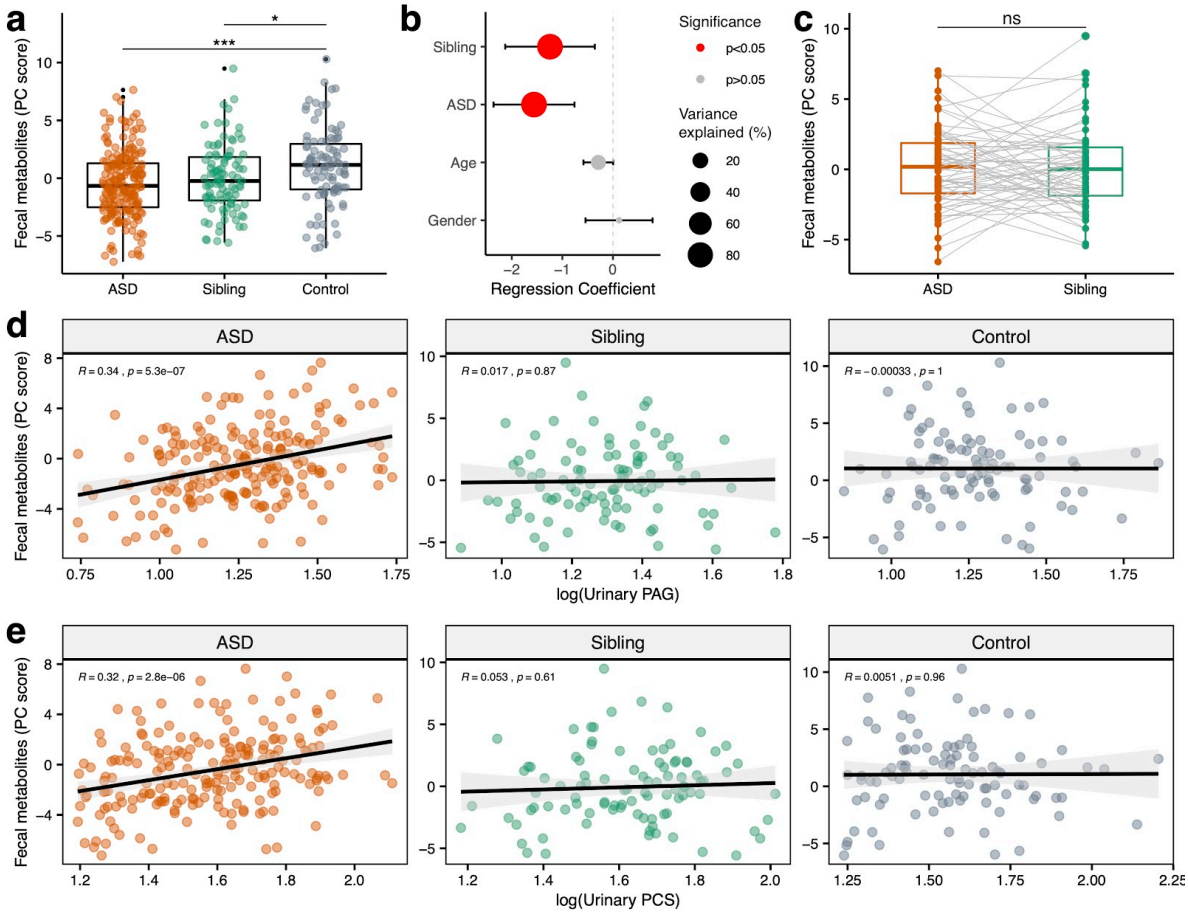


Figure 5.11: Comparison of metabolic phenotypes between study groups. a) Faecal metabolite levels in each study group represented by a latent variable from principal component analysis. *P* values were calculated by t-test. **b)** Regression coefficients from a linear mixed-effects model regressing faecal metabolite concentrations against study group, age and gender while adjusting for the random effect of different recruitment centres. Fixed effects with significant influence in the model are plotted in red. **c)** Paired t-test of faecal metabolite concentrations between ASD and their neurotypical siblings. **d)** Faecal metabolite levels in ASD correlate (Pearson) with urinary phenols phenylacetylglutamine (PAG) and d) *p*-cresyl sulfate (PCS) but this association was not observed in the sibling or control groups. Linear regression lines are shown (black) with 95% confidence intervals (light grey). *PC principal component; ns not significant; *P<0.05; ***P<0.001*.

As previously discussed, there is a trade-off between carbohydrate and protein fermentation in the gut; proteins become increasingly important substrates as carbohydrates are depleted[25,26]. To support the hypothesis that the observed metabolic phenotypes in children with ASD were related to the degree of protein fermentation in the gut, faecal concentrations of the main SCFAs (acetate, butyrate and propionate) produced by carbohydrate fermentation[23] were correlated with faecal metabolite profiles and urinary phenols. Acetate and butyrate were negatively

correlated with urinary markers of protein fermentation, PAG and PCS, and negatively correlated with faecal amino acid/uracil concentrations (**Figure 5.12**), indicating that carbohydrate fermentation was reduced in children with higher rates of protein fermentation and elevated in children with lower rates of protein fermentation. Significant correlations with propionate were not observed. One study showed that faecal concentrations of acetate and butyrate were affected by a low-carbohydrate diet while the relative proportion of faecal propionate was unaffected by carbohydrate intake[423], indicating that carbohydrate availability may not be the main regulator of propionate production in the gut.

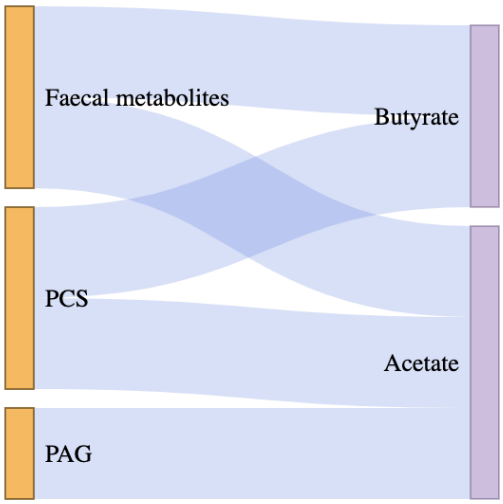


Figure 5.12: Correlations between protein-related metabolites and faecal short-chain fatty acids. Significant (*FDR* <0.05) Spearman correlations between protein-associated metabolites (orange) and carbohydrate-associated metabolites (purple) are plotted with negative correlations represented by blue bands. *PAG phenylacetylglutamine; PCS p-cresyl sulfate.*

5.3.8 ASD-associated metabotypes display differential microbiota

All faecal samples analysed by ^{1}H NMR spectroscopy (n=697) were clustered (2-means) into groups based on relative concentrations of tryptophan, phenylalanine, tyrosine, leucine, isoleucine, valine, glycine, lysine, aspartate, methionine, glutamate, alanine and uracil. These were the observed metabotypes. A random forests model (ntrees=351, mtry=8) was trained on

the matching microbiome data (i.e. from the same faecal samples) to classify the samples into the two groups, labelled low amino acid (lowAA) and high amino acid (highAA) metabotypes. The training data was composed of a single, random time point from each individual (n=419) and the model was tested on all samples left out of the training set (n=271). Samples with faecal ^{1}H NMR data but without matching microbiome data were excluded from this analysis (n=7). The random forests model was significant (*P* <0.001, permutation test) although it had a high out-of-bag (OOB) error rate of 36.3%. Cross-validation revealed that 65.2% of samples could be classified correctly. When tested with new samples, the model performed with similar accuracy which, although low, was significant (**Table 5.3**). McNemar's test revealed that there was a significant difference in the performance of the model for the two classes. Samples with the lowAA phenotype were predicted correctly (sensitivity) in 73.1% of samples while this was only true in 53.3% of samples classified as highAA, which could indicate that the lowAA phenotype had a greater association with the composition of the gut microbiota.

Table 5.3: Random forests model statistics

Metric	Value
Accuracy	0.631
P value (accuracy)	2.12E-05
P value (McNemar's test)	0.007
Sensitivity (lowAA)	0.731
Sensitivity (high AA)	0.533
Prevalence (lowAA/highAA)	0.494/0.506

AA amino acid

For most children included in the present study, those with the highest or lowest faecal metabolite concentrations exhibited consistent observed metabotypes across multiple sampling time points (**Figure 5.13**). However, those with intermediate levels of the faecal metabolites had inconsistent observed metabotype classifications. Metabotype predictions based on gut bacterial profiles were remarkably consistent for individuals across sampling time points in spite of the disagreement between observed and predicted metabotypes in some children (**Figure**

5.13), demonstrating a degree of stability in the gut bacterial profiles of most children in the study. A small number of children (ASD n=11; SIB n=1) had inconsistent predicted metabotype classes across sampling time points. Five of these children reported illnesses (2 prior to joining the study and 3 during the study). It is possible that instability in the gut microbiota of these children is a reflection of illness or use of medications[50,422], although none of these children reported antibiotic usage.

From the remainder of the 16S rRNA gene amplicon sequencing data (samples without matching ^{1}H NMR data), metabotypes were predicted using the random forests model. For each individual, the overall metabotype was assigned as the majority class observed and predicted (**Figure 5.14**). In cases where the class label was inconsistent between sampling points or between observations and predictions, no class label was assigned for that individual. The distribution of metabotypes was significantly different between the study groups (P=0.00119, chi-squared test) with the lowAA phenotype over-represented in children with ASD and under-represented in the unrelated controls while the SIB group had almost equal numbers of children with the two phenotypes. This result is consistent with the faecal metabolite pattern observed in the single time point analysis (see **Figure 5.11**).

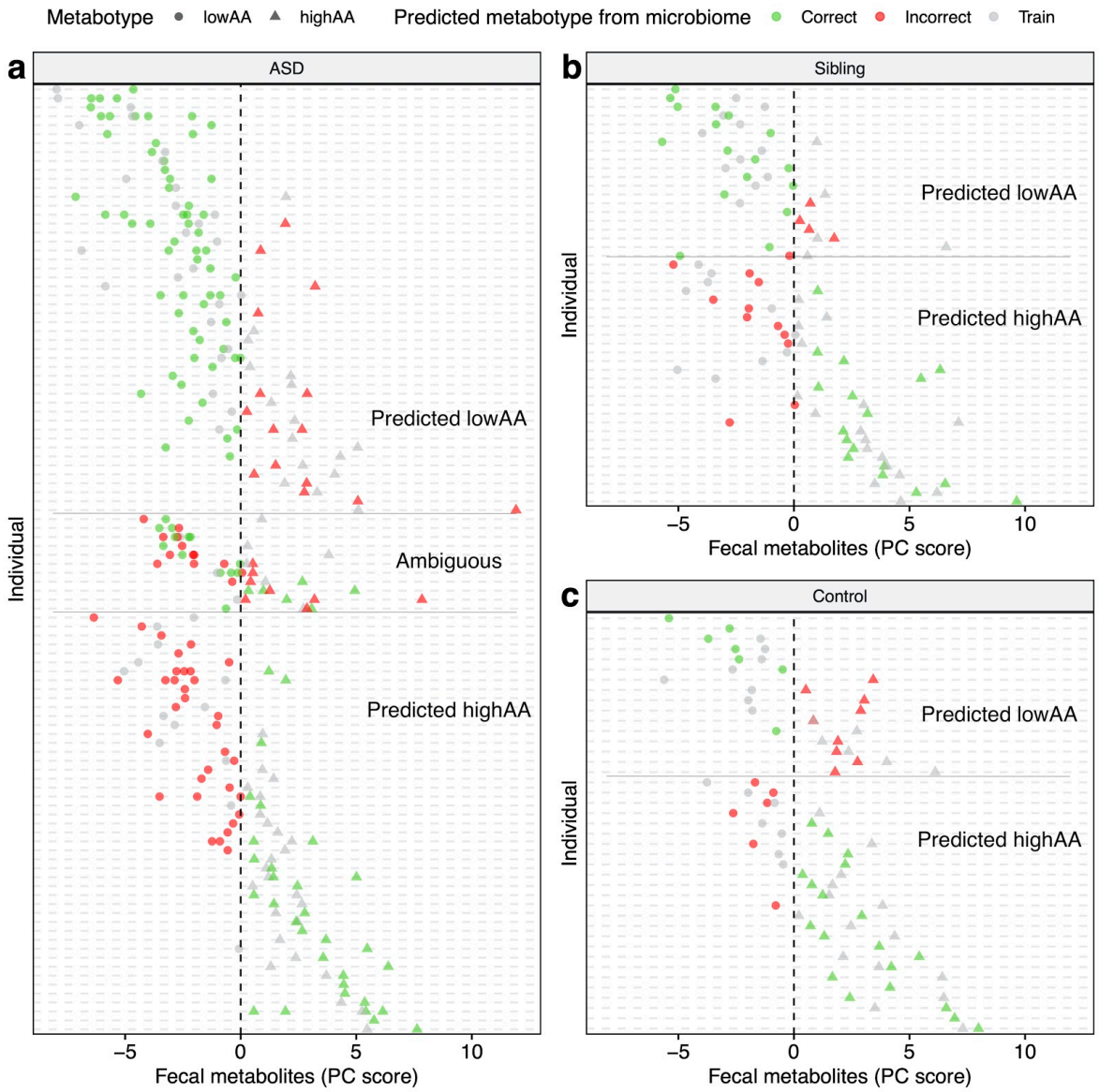


Figure 5.13: Metabotype prediction from gut bacterial profiles. For each individual, faecal metabolite profiles of repeated samples are represented by a latent variable (PCA component score) which is a linear combination of the selected metabolites. An observed metabotype, low amino acid (lowAA; circle) or high amino acid (highAA; triangle), for each of the samples was determined using 2-means clustering of the relative concentrations of the selected metabolites. One time point from each individual was used to train a random forests model (grey) to classify the samples by metabotype based on microbiome data and the model was used to predict metabotypes of the remaining samples. Correct classifications (prediction agreed with observed metabotype) are coloured in green and incorrect classifications (prediction did not agree with observed metabotype) are coloured in red. Children with **a)** ASD, **b)** neurotypical siblings and **c)** unrelated neurotypical controls are plotted separately. *PC principal component; AA amino acid*.

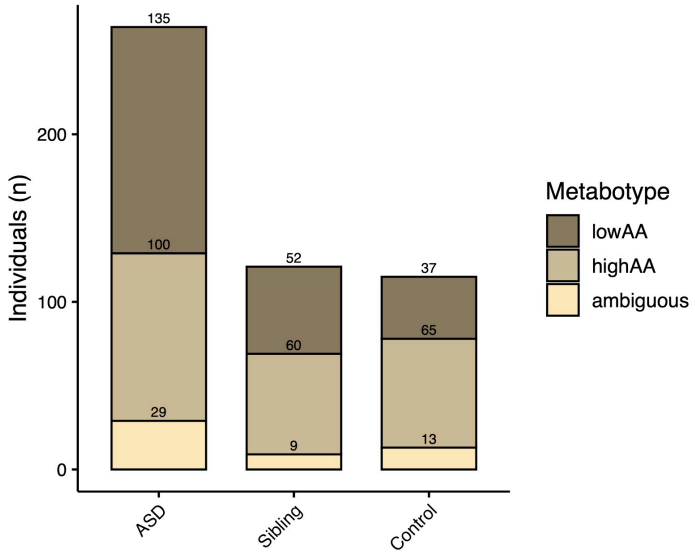


Figure 5.14: Summary of metabotypes by study group. Low amino acid (lowAA) and high amino acid (highAA) metabotypes were assigned for each individual except in cases where the metabotype was inconsistent (ambiguous) among repeated samples or between observations and predictions for that individual.

The top 30 most important genera for classifying metabotypes using the random forests model are displayed in **Figure 5.15**. Only two of these genera, *Lactococcus* (lower in ASD) and *Anaerococcus* (higher in ASD), exhibited differential abundance in children with ASD compared to NTC when differences in metabotypes were not considered (**Figure 5.16**). Here, the variation in abundances was mostly explained by the ages of the children. However, when the models were adjusted for differences in metabotypes, previously undetected associations between bacterial abundances and ASD were revealed (**Figure 5.17**). *Anaerococcus* remained significantly higher in the ASD group while the reduced abundance of *Lactococcus* was instead attributed to the prevalence of the lowAA phenotype in this group. *Anaerostipes*, the most influential genus in the random forests model, are abundant SCFA-producing bacteria in the human gut microbiota of healthy individuals[432,433]. These bacteria are also capable of synthesising many amino acids and at least one species is an indole-producer[434]. In the present cohort, *Anaerostipes* abundance was negatively associated with age and elevated in children with the highAA phenotype (**Figure 5.17**) which agrees with the positive correlation previously observed between *Anaerostipes* and faecal amino acids in the ASD group (see **Figure 5.10**). However, this association was not unique to any study group.

Haemophilus, the second most influential genus in the random forests model, was also negatively associated with age but was specifically associated with the lowAA phenotype in children with ASD and their siblings (**Figure 5.17**), mirroring the negative correlation observed between *Haemophilus* and protein fermentation-associated metabolites in the ASD group (**Figure 5.10**). A previous study investigating the ASD microbiome found that *Haemophilus parainfluenzae* was significantly reduced in faecal samples from children with ASD compared to neurotypical controls[381]. *H. parainfluenzae* was also found in higher relative abundance in paediatric IBS patients[435] and in children with diarrhoea[436]. These findings together with the results from the current population suggest that *Haemophilus* could be associated with gastrointestinal dysfunction in a subtype of ASD. The study by Kang et al.[381] included a small number of ASD children (n=23) who had higher constipation than diarrhoea severity scores. Since constipation is presumably associated with the highAA phenotype (as discussed in **5.3.5**), this could explain the discordance with the present study. Like *Haemophilus, Butyricicoccus* was specifically enriched in ASD and SIB children with the lowAA phenotype and was negatively correlated with protein fermentation-associated metabolites in ASD (see **Figure 5.10**). This genus includes butyrate-producing saccharolytic bacteria[434] which is in line with the hypothesis that the lowAA metabotype is associated with increased carbohydrate fermentation compared to the highAA metabotype.

Flavonifractor, the third highest ranked genus in the random forests model, had reduced abundance in the ASD compared to NTC group (**Figure 5.17**). However, within the ASD group, *Flavonifractor* abundance was significantly lower in the lowAA subgroup compared to the highAA subgroup, again demonstrating phenotypic differences specific to a subpopulation of ASD. As before, consistency with the PLS model in ASD is observed; *Flavonifractor* was positively correlated with faecal amino acids (see **Figure 5.10**). *Flavonifractor* spp. are flavonoid-metabolising bacteria[437] which produce bioactive metabolites including phenolic compounds[438]. Flavonoid metabolites have been shown to exert a number of beneficial effects on the host such as limiting oxidative damage, inhibiting inflammation and anti-cancer actions[438]. A closely related genus, *Pseudoflavonifractor*, was also found in lower abundance in lowAA ASD although it was generally enriched in lowAA metabotypes in the neurotypical population (**Figure 5.17**). It is evident that some children with ASD may benefit from modulation of the gut

microbiota to increase biotransformation of flavonoids.

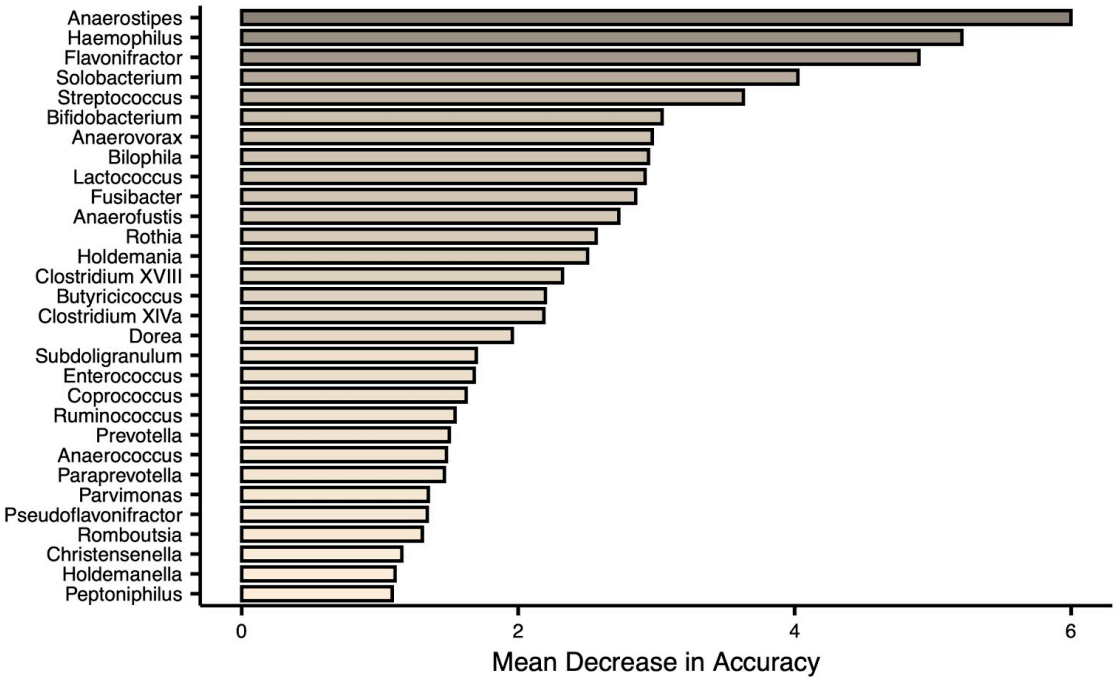


Figure 5.15: Variable importance from random forests model. Bacterial genera are ranked based on mean decrease in accuracy. A high decrease in accuracy for a specific variable indicates the model is more accurate when that variable is included.

Two additional genera, *Anaerovorax* and *Ruminococcus,* had lower abundance in ASD compared to NTC and this was especially true for ASD children with the lowAA metabotype (**Figure 5.17**). The abundances of these genera were also reduced in lowAA SIB, but the relationship with metabotype was weak in the general population. In the ASD group, *Anaerovorax* and *Ruminococcus* were positively correlated with faecal amino acids (see **Figure 5.10**) which is consistent with *A. odorimutans,* the only known species of *Anaerovorax,* being a putrescine-fermenter[439] (putrescine is an amino acid breakdown product) and the mucin-degrading capacity of *Ruminococcus* species[440]. Mucin degradation results in the production of both amino acids and SCFAs[441]. The relationship between abundance of *Anaerovorax* in the human gut and host health is not clear although this genus is negatively associated with obesity[90], albeit higher prevalence in colorectal cancer patients[442]. There is far more evidence for *Ruminococcus* spp. as key members of the human gut microbiota[89,443]. Relative abundances of *Ruminococcus* and its family, *Ruminococcaceae,* are often reported to

be altered in children with ASD compared to neurotypical controls, but the effect sizes are small and have inconsistent directions[115–119].

In the present cohort, *Holdemanella* was increased in the ASD compared to the NTC group and was further enriched in children with the lowAA phenotype. The positive association between *Holdemanella* and ASD has been reported[444]. Consistent with the current finding, *Holdemanella* was not only found in higher abundance in faecal samples from children with ASD, but was also enriched in ASD children without constipation compared to children with ASD and constipation. Although *Holdemanella* is prevalent in the gut microbiota of healthy individuals, it has been associated with an unhealthy lipid profile[445] and the progression of chronic kidney disease[446]. *Fusibacter* and *Anaerococcus* abundances were also found to be higher in ASD regardless of metabotype although the effect size was small. Many of the genera deemed important for classification of metabotypes were associated with age although the variance explained by age was smaller than that explained by group or metabotype for ASD-associated genera, with the exception of *Haemophilus* and *Holdemanella* (**Figure 5.17**). The gut microbiota develops rapidly over the first three years of life but continues to stabilise into early childhood[69]. In particular, *Bifidobacterium* is highly abundant in children and the abundance of this genus seems to decrease into adulthood[69], a trend that is observed in the present cohort.

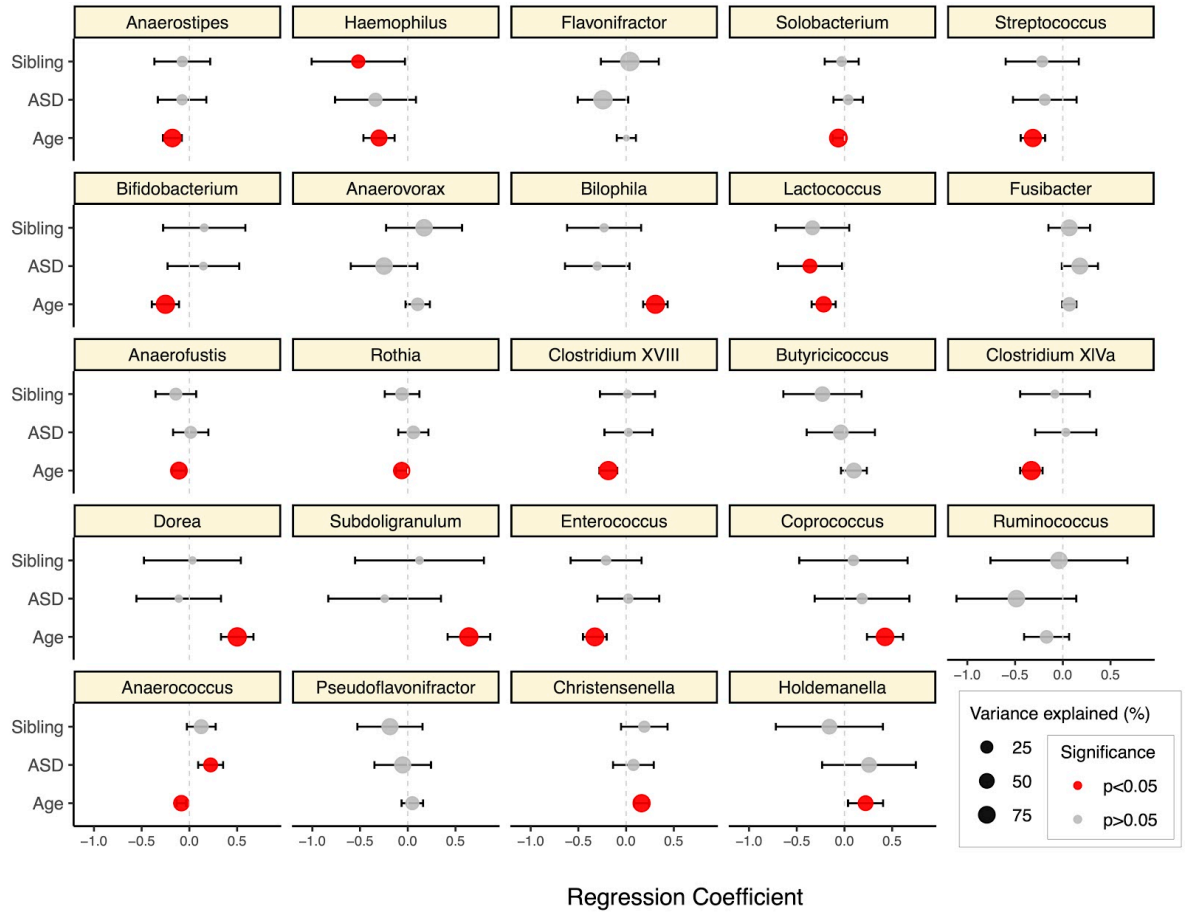


Figure 5.16: Associations between bacterial genera and study groups. Regression coefficients from linear mixed-effects models are plotted with 95% confidence intervals for each genus. Genera that were not significantly associated with any of the fixed effects investigated are not plotted. Individual and recruitment centre were included as random effects. Percentage of variance explained by each fixed effect is represented by the size of the points while statistical significance is represented by the colour of the points. Regression coefficients for ASD and Sibling are with respect to Control children.

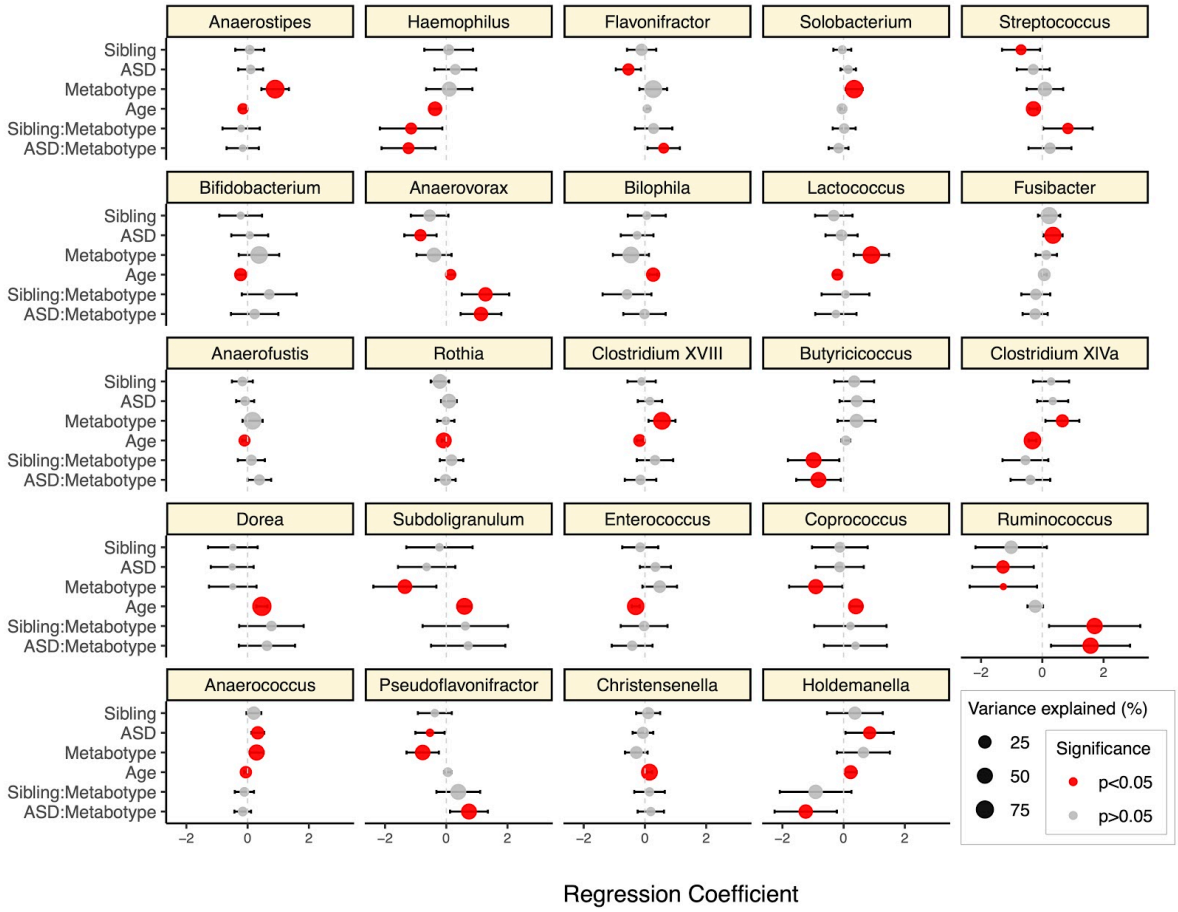


Figure 5.17: Associations between bacterial genera, study groups and metabotypes. For each genus, regression coefficients (linear mixed-effects models) are plotted with 95% confidence intervals. Models were adjusted for repeated measures and different recruitment centres. For each fixed effect, the percentage of variation explained and significance are denoted by the size and colour of the points, respectively. For metabotype, the positive direction indicates association with the high amino acid group while the negative direction indicates association with the low amino acid group. Regression coefficients for ASD and Sibling are with respect to Control children.

In summary, the findings presented here provide evidence that there is significant heterogeneity in both the metabolome and microbiome within the ASD population. Some of the variation can potentially be attributed to differences in gastrointestinal status with the strongest evidence pointing toward phenotypic differences due to increased intestinal transit time. Although the limited metadata available for the present study was not able to support this hypothesis, the phenotypes associated with the highAA group (high concentrations of faecal amino acids, high urinary excretion of phenolic compounds, negative correlations between these metabolites and

faecal SCFAs, and positive correlations between these metabolites and protein/amino acid-fermenting bacteria) are consistent with increased gut bacterial fermentation of protein. Constipation, a hallmark of delayed intestinal transit, results in depletion of carbohydrates in the colon and a switch to protein fermentation[412]. The differential findings within subpopulations of ASD recapitulate the importance of sample size and addressing within-group variation when studying heterogeneous conditions such as ASD and IBS[413,447]. Notwithstanding, these results support the idea that modulation of the gut microbiota may be an effective intervention in ASD though a precision medicine approach would be necessary.

Chapter 6

General discussion

6.1 Summary of key results

The work presented in this thesis details the intricate metabolic relationships that exist between the gut microbiota and its host at different stages of human development with respect to two underlying health conditions that affect early life development: maternal obesity and Autism Spectrum Disorder (ASD). Results were obtained from two observational cohorts of 825 individuals who donated 2,404 clinical samples. Throughout this thesis, the utility of longitudinal multi-omic profiling of multiple biofluids was demonstrated. A consistent theme observed in both studies was the relevance of amino acid metabolism by the gut microbiota to many aspects of human health. In particular, anaerobic fermentation of aromatic amino acids in the gut may be involved in the maintenance of maternal insulin sensitivity during pregnancy but may also be an indication of increased intestinal transit time (and therefore gastrointestinal dysfunction) in children with Autism Spectrum Disorder (ASD). The digestion of protein in the large intestine by resident bacteria - i.e. protein putrefaction - is an important physiological process that is often overlooked[31] but is of interest to human health and disease.

6.1.1 Maternal obesity impacts pregnancy

During gestation, development of the fetus is regulated by maternal physiology. Maternal obesity impacts fetal development, as evidenced by the prevalence of macrosomia among neonates born from mothers with a high pre-pregnancy BMI[193,194], but the relationship between the maternal gut microbiota and fetal development in obese pregnancies is not clear. The present study (CWH dataset 1; **Chapter 3**) demonstrated that maternal insulin resistance was elevated in obese women during pregnancy and was positively correlated with urinary metabolites produced by gut bacterial metabolism. Maternal insulin resistance was negatively correlated with amino acids in maternal serum and urine, and these amino acids were found in significantly reduced concentrations in samples from obese women. Since amino acids are

transported from maternal circulation to the fetus[67] to promote fetal protein synthesis[244], it is tempting to speculate that this transport occurs at a higher rate in obese compared to normal weight pregnancies. The rate of transfer may be inversely proportional to maternal insulin resistance. Increased transfer of these amino acids to the fetus would leave the maternal supply reduced and contribute to enhanced fetal growth (**Figure 6.1**). This supposition fits with Pedersen's hypothesis[235] that elevated maternal insulin resistance leads to increased glucose transfer to the fetus which stimulates fetal insulin secretion, thereby resulting in fetal overgrowth. Positive correlations between amino acid concentrations and insulin levels in umbilical cord serum, as well as positive correlations between umbilical cord amino acids and birth weight identified from the current analyses further support this theory.

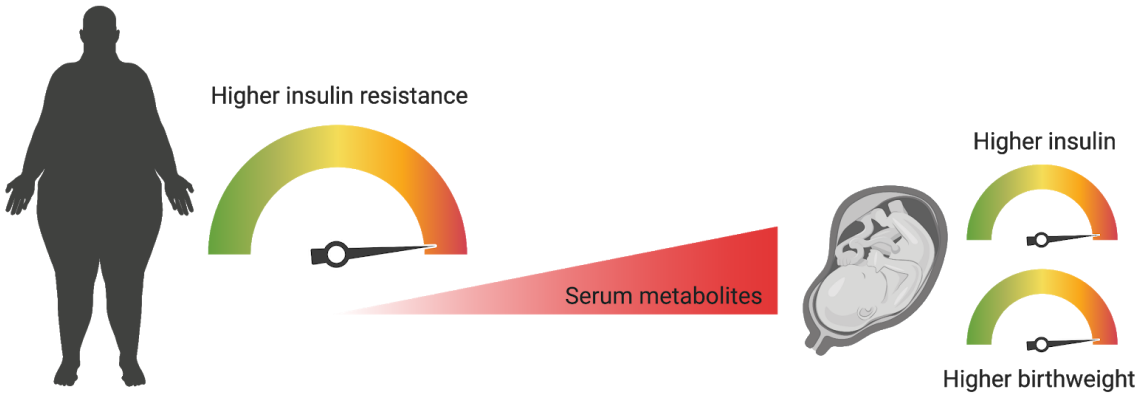


Figure 6.1: Observed metabolic relationship between maternal obesity and neonatal outcomes. Maternal obesity is associated with elevated insulin resistance during pregnancy. The data presented in Chapter 3 indicate higher levels of maternal insulin resistance are associated with increased transfer of some metabolites, namely amino acids, from maternal circulation to the fetus. This results in lower maternal concentrations of these metabolites compared to women with normal weight and lower insulin resistance. High umbilical cord concentrations of these metabolites are associated with high insulin concentration in the umbilical cord and increased weight of the neonate at birth.

Although umbilical cord samples provide an approximation of fetal/neonatal metabolism, they may not reflect the exact metabolic profiles of the neonates, which is a limitation when interpreting the data presented in **Chapter 3**. In addition, umbilical cord amino acid concentrations have been shown to be dependent upon delivery mode[448]. Similarly, insulin concentration in the umbilical cord may not reflect the true insulin level in the neonate and may be confounded by mode of delivery, fetal distress and maternal glucose intake[131]. The confounders detailed above could also explain the lack of association between maternal BMI and umbilical cord amino acid concentrations. The relationship between delivery mode (elective

caesarean section, emergency caesarean section, standard vaginal birth and instrumental vaginal birth) and biochemical measurements in the umbilical cord will be investigated as the cohort grows. The current cohort is not likely to have sufficient statistical power to analyse differences between delivery modes due to the small number of umbilical cord samples collected.

Despite the limitations described above, this dataset provided unparalleled information about the dynamic maternal metabolic response to pregnancy in obese women compared to those with normal weight. The findings show that metabolic trajectories of many metabolites are not linear during this period (see **Figure 3.10** for example), hence the sampling throughout pregnancy was necessary to model the metabolic changes. Overall, the maternal metabolic profile was driven by physiological adaptations to pregnancy and was strongly associated with maternal insulin resistance. The maternal gut microbiota may be associated with elevated insulin resistance in obese women during pregnancy. This increase in insulin resistance may result in higher transfer of some maternal metabolites to the fetus and fetal overgrowth.

6.1.2 Previous malabsorptive surgery impacts pregnancy

Bariatric surgery has become a popular solution to morbid obesity due to its success as a treatment for weight loss and obesity-related comorbidities[308]. Although bariatric surgery induces dramatic weight loss, individuals tend to remain in the obese category[308]. Some pregnancy complications, such as elevated insulin resistance[308], are improved in obese women with a previous bariatric surgery, but these women are at a high risk of having small for gestational age neonates[321,322]. The present study (CWH dataset 2; **Chapter 4**) investigated changes in maternal metabolism associated with a previous restrictive or malabsorptive bariatric surgery compared to women with similar early-pregnancy BMI, and whether these changes influenced fetal development. Restrictive surgeries did not induce significant changes in maternal metabolism during pregnancy, although the number of patients with these surgeries was small (n=16), therefore this could indicate the statistical test does not have sufficient statistical power. In addition, sleeve gastrectomy has been shown to have malabsorptive effects whereas gastric banding is purely restrictive. The two subgroups were not investigated separately due to the insufficient number of patients, but will be analysed separately in the future. Malabsorptive patients, all of which had a previous gastric bypass surgery, had distinct gut microbial

communities (compared to no surgery controls) that were associated with increased excretion of protein putrefaction metabolites throughout pregnancy. These metabolites were negatively correlated with maternal insulin resistance. The proposed mechanism is that protein digestion is inhibited in malabsorptive patients[329], which allows higher amounts of intact protein to reach the colon where it is fermented by gut bacteria[348] (**Figure 6.2**). The change in nutrient balance may also contribute to the observed alterations in the gut microbiota. This theory is supported by the reduction in circulating amino acids, which result from peptide and amino acid absorption in the small intestine[14]. The link between protein putrefaction and insulin sensitivity is not clear. Rather than being a direct connection, it is possible that insulin sensitivity is related to the degree of protein malabsorption which is reflected in the excretion of putrefaction metabolites. The negative correlation between fetal growth and putrefaction metabolites may similarly reflect inhibition of fetal growth due to maternal malabsorption-induced malnutrition[353]. However, the detection of high concentrations of putrefaction metabolite, phenylacetylglutamine, in urine of neonates born from malabsorptive mothers suggests that the maternal gut microbiota of malabsorptive patients has the potential to directly influence fetal health.

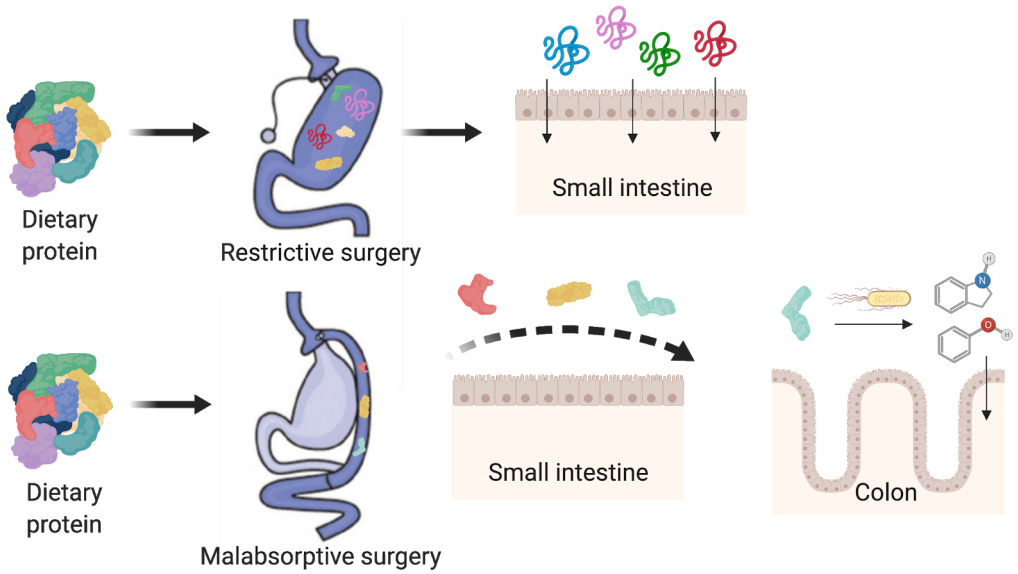


Figure 6.2: Malabsorptive surgery-induced changes in protein metabolism. Restrictive surgeries, which aim to reduce food intake, do not alter digestion. Dietary protein is metabolised in the stomach and small intestine as normal; peptides and amino acids are absorbed into circulation through the small intestine. On the other hand, metabolism is altered after malabsorptive surgeries. In addition to inhibited digestion in the stomach, part of the small intestine is bypassed. Hypothetically, this altered digestion results in high amounts of intact protein reaching the large intestine, where it is broken down and fermented by the gut microbiota.

The standard oral glucose tolerance test was used to measure insulin resistance in the present cohort. However, its use for gastric bypass patients has been questioned due to dumping syndrome[449] (the rapid emptying of gastric contents). Capillary blood glucose monitoring has been suggested as an alternative and the team at Chelsea & Westminster Hospital have since revised their methods. The small number of neonatal urine samples also causes concern with respect to drawing biological conclusions and therefore, the results pertaining to these samples should be treated with caution.

6.1.3 Gut microbial alterations in Autism Spectrum Disorder are conditional

In **Chapter 5** (Jebsen study), a population of children with ASD were characterised and compared to their neurotypical siblings and unrelated neurotypical controls, with the goal of uncovering microbiota-associated signatures of altered neurodevelopment. In line with recent reviews and meta-analyses[115–119], a distinct bacterial signature of ASD was not detected in the present study, potentially due to heterogeneity within the ASD group. Children with ASD displayed variation in metabolic and gut bacterial profiles that was consistent with variations in intestinal transit time. On the extremes, ASD children had either low or high concentrations of faecal amino acids and urinary putrefaction metabolites. High concentrations of these metabolites are consistent with increased intestinal transit time, which results in depletion of carbohydrates in the colon and a switch to proteolytic fermentation[412] (**Figure 6.3**). In accordance, the metabolites were negatively correlated with carbohydrate-fermenting bacteria and faecal short-chain fatty acids, and positively correlated with protein-fermenting bacteria. The two metabolic phenotypes identified in children with ASD likely differentiate children who tend to have constipation from children who tend to have diarrhoea. However, gastrointestinal symptoms were considerably under-reported (12.1%) in the present study compared to the expected prevalence (40%[114] to 60%[409]), so a definite association between the observed metabotypes and gastrointestinal status cannot be made. Low and high amino acid metabotypes were not associated with variations in protein putrefaction metabolites in the sibling or control children, which potentially reflects a lack of gastrointestinal problems. Of the 254 sibling and control children included in this study, only 5 were reported to have gastrointestinal symptoms.

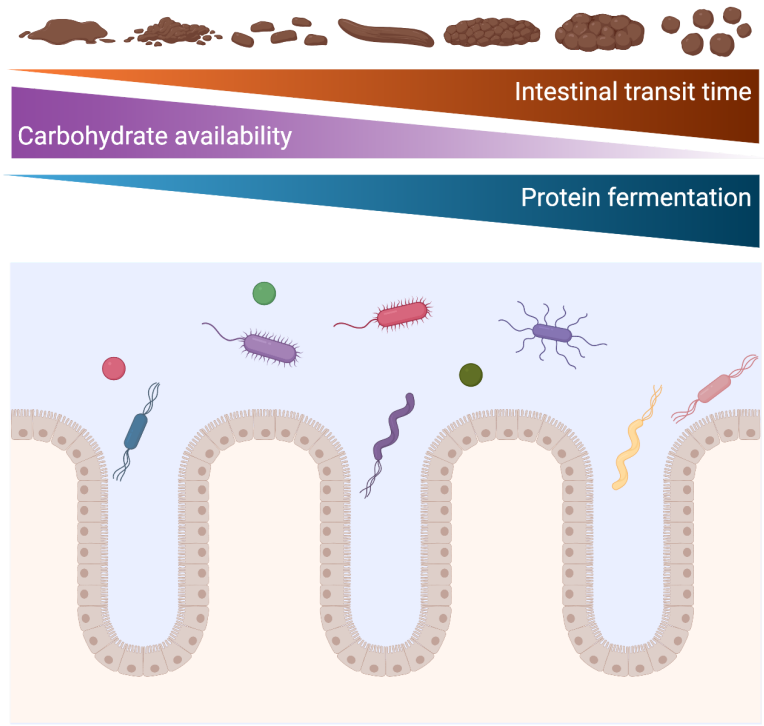


Figure 6.3: Relationship between intestinal transit time and protein putrefaction in the colon. Constipation, a proxy for increased intestinal transit time, results in the depletion of carbohydrates in the large intestine. Resident bacteria switch to protein fermentation which produces phenolic and indolic compounds.

Despite the distribution of the two metabotypes, children in the present cohort with ASD had significantly lower faecal amino acid concentrations compared to unrelated neurotypical controls. Low faecal amino acids in the ASD group were associated with psychiatric comorbidities and the use of fish oil as a supplement. Although fish oil consumption may have been a confounder, it would have been interesting to relate fish oil consumption and the presence of psychiatric comorbidities with ASD symptom severity, had the data been collected. It is possible that children with problematic behaviours are more likely to be given this supplement that is proposed to reduce stereotypy and hyperactivity in children with ASD[378]. Therefore, the low faecal amino acid metabotype could potentially be associated with symptom severity in addition to intestinal transit time. Many faecal amino acids are precursors to important neurotransmitters produced in the gut, which adds further merit to this theory. Neurotypical siblings also had slightly lower faecal amino acid concentrations compared to the

controls, indicating this metabotype may also reflect lifestyle factors, though faecal amino acid concentrations were not associated with fish oil consumption in neurotypical siblings.

The present study highlights the necessity of large sample sizes, repeated sampling, and adequate clinical metadata for microbiome and metabolomics studies investigating ASD. To date, studies investigating the gut microbiome and metabolic profile of ASD have been small (up to 134 participants) and have only focused on case-control contrasts. The inherent heterogeneity in the population, presence of challenging eating behaviours[364] and prevalence of gastrointestinal dysfunction will significantly confound ASD signatures in omics data. However, capturing within-group variation and factoring this into the analyses proved useful. This approach uncovered many associations between bacterial taxa and ASD that were undetected in the standard case-control contrasts. Collectively, the results presented in **Chapter 5** suggest that simple case-control comparisons are not suitable in ASD studies since the population likely consists of many subtypes, some of which are characterised by opposing symptoms. A validated gastrointestinal symptom scoring system should be used for every study participant, in addition to scoring each faecal sample according to the Bristol stool scale as an indicator of intestinal transit time[450]. Stool consistency is highly associated with the faecal microbiome composition[86] and is highly variable in ASD, again consistent with the notion of multiple ASD subtypes. Finally, another limitation of this study was the collection of the patient metadata. Self-reporting, even if done by a caregiver, can be inaccurate for children who have impaired communication or are non-verbal.

6.1.4 Overall limitations

The data presented in this thesis was based on purely observational studies and therefore, any claims made herein were associations by nature and not indicative of causal relationships. Power calculations were also not performed as part of the study designs, so it is unclear whether a sufficient number of individuals were included in each study. Additionally, the platforms used for profiling have limitations which need to be considered when interpreting the data. Metabolic profiling by ^{1}H NMR spectroscopy is a reliable and robust method but suffers from a lack of sensitivity. The results presented are biased toward molecules that are present in sufficient quantities to detect using this analytical method. However, ^{1}H NMR is able to assay a wide variety of compounds, which is not the case for other metabolic profiling platforms. This

method also suffers from poor spectral resolution of some metabolites where peak overlap occurs, resulting in discarding potentially important metabolites from the analysis. The problem can be solved by using spectra acquired with other pulse sequences (e.g. J-Resolved) for profiling instead of the standard 1D pulse sequence. Taxonomic profiling by 16S rRNA gene amplicon sequencing is also a low-resolution method. Amplicon sequence variants are very short (usually less than 300 bp) segments of the full length gene, and taxonomic discrimination between bacterial species, and even some genera, is poor[150]. No standard genetic sequence similarity threshold exists for assigning bacterial taxonomies. Furthermore, discrimination between taxa in the *Enterobacteriaceae* family and *Streptococcus* genus, two taxonomic groups associated with malabsorptive surgery in **Chapter 4**, are sometimes difficult to differentiate by the 16S rRNA gene sequence[150]. Differentiation between bacterial species and even strains is crucial for understanding their metabolic capabilities. Another limitation of taxonomic profiling using the 16S rRNA gene is that this gene is not a single copy gene in all bacterial genomes, which may artificially inflate the relative abundance of some strains compared to others. Metagenomic sequencing is a good alternative to amplicon-based sequencing, although the computational challenges are far greater.

6.2 Future work

As the CWH study is ongoing, the findings presented in this thesis pertaining to pregnancy will be validated in a larger cohort, although validation in a separate independent cohort will be crucial for generalising the findings. The offspring will also be followed longer-term (up to age 5) to assess the transgenerational metabolic and neurodevelopmental effects of maternal obesity and maternal malabsorption. Additional experiments will build on the results presented in **Chapter 3** and **Chapter 4**. Spectral signals representing lipid molecules in maternal serum samples were significantly associated with both obesity and malabsorptive surgery, but these metabolites cannot be identified by ^{1}H NMR chemical shifts. Mass spectrometry-based metabolic profiling is more suitable for identifying lipid species and will be applied to maternal serum and umbilical cord serum samples going forward. Vaginal swabs will also be analysed to determine obesity- and bariatric surgery-associated changes in the vaginal microbiome during pregnancy and the clinical significance of any microbial alterations. In **Chapter 3**, associations between maternal gut microbiota-associated metabolites and maternal insulin resistance were identified, but due to financial and time constraints, microbiome data were not available for this

subset of the cohort (dataset 1). Taxonomic analysis of faecal samples for dataset 1 will fill an important gap in the current story, enabling a link between the obesity-associated metabolic profile and the gut microbiota. In addition, metabolic profiling and metagenomic analysis of all faecal samples (**Chapter 3** and **Chapter 4**) would provide another layer of information about the functionality of the gut microbiota in obese and malabsorptive patients during pregnancy. Elevated concentrations of amino acids and increased abundances of the relevant genes in faecal samples from malabsorptive patients would support the speculation that protein putrefaction is upregulated in these patients. Higher activities of bacterial proteases and peptidases in faecal samples would also provide further support for this hypothesis.

To move past associations and investigate mechanisms, *in vitro* and *in vivo* experiments will be necessary. The main theme that emerged from the CWH study was that the maternal gut microbiota influences maternal insulin sensitivity/resistance during pregnancy. This concept has been demonstrated in a mouse model but, to my knowledge, has not been investigated in relation to maternal obesity or previous bariatric surgery. In a similar experiment to Koren et al.[72], the gut microbiota of obese, malabsorptive and normal weight pregnant women from the first and third trimesters could be transplanted into germ-free mice. To support the current hypothesis, mice inoculated with the third trimester obese microbiota should exhibit the highest levels of insulin resistance while those inoculated with malabsorptive or normal weight microbiota should exhibit lower levels of insulin resistance. It would also be interesting to investigate the effect of protein malabsorption on insulin resistance, both during and outside of pregnancy. Batch cultures or artificial gut models could be used to first measure the putrefactive capacity of faecal samples from malabsorptive patients compared to no surgery controls. These microbial communities could then be transplanted into germ-free diabetic mice on a diet of resistant protein to assess whether an upregulation in gut microbial protein fermentation improves insulin resistance. It would be interesting to observe the outcomes of pups born from dams inoculated with faecal samples from malabsorptive patients compared to no surgery controls.

The data generated from the Jebsen study (**Chapter 5**) undoubtedly has much more information to be mined. There is a lack of data describing the gut microbiota of school-aged children[69], and the Jebsen study characterised the gut microbiota in over 500 school-aged children from across the United States. Analysis of these data with respect to age will be pursued in the future. Given

the importance of amino acids in discriminating subtypes of ASD as well as ASD from neurotypical controls, running targeted assays on the samples collected would be worthwhile to measure differences in amino acids (especially D- versus L-amino acids to assess bacterial origin) and their derivatives, which are markers of gut bacterial metabolism (e.g. tryptophan pathway assay[451]). It would also be interesting to study the causal role of the gut microbiota in gastrointestinal symptoms associated with ASD. Again, using faecal transplant in germ-free mouse models would be one approach. Batch cultures or artificial gut models could also be used to determine the fermentative capacity of different communities with respect to their metabolic phenotype. In other words, do communities associated with the high amino acid metabotype in ASD children have a higher propensity to ferment protein, or is the metabolic activity purely based on the available substrates? With regard to future studies, it will be critical to study ASD in the context of birth cohorts to understand the role of the gut microbiota and its metabolites in the development of ASD.

6.3 General conclusion

Microbiome research is emerging as a prominent field in healthcare and nutrition. In this thesis, I detailed gut microbiota-associated metabolic pathways linked to human disease phenotypes. These observations suggest a close relationship between human health and our resident microbes, and highlight the prospect of manipulating human health by harnessing the metabolic potential of the gut microbiota. The translational potential of microbiome research is promising, but characterisation of the metabolic activity of the gut microbiota is lacking. There is a growing need to isolate and culture bacteria relevant to human health to map metabolic capacities of specific strains and the metabolic interactions between strains. As a community, we should focus more effort on following up observations with mechanistic experiments to prove causality between the gut microbiota and disease phenotypes, as well as identify protective mechanisms where the gut microbiota is associated with health. The data presented in this thesis lays a framework for investigating the mechanisms of interactions between the gut microbiota and its host in relation to human development. Future work building on the results detailed herein will help solidify our understanding of the microbial influence on our developmental outcomes.

References

1. Marchesi, J. R. & Ravel, J. The vocabulary of microbiome research: a proposal. *Microbiome* **3**, 31 (2015).
2. Gilbert, J. A. *et al.* Current understanding of the human microbiome. *Nat. Med.* **24**, 392–400 (2018).
3. Costello, E. K. *et al.* Bacterial community variation in human body habitats across space and time. *Science* **326**, 1694–1697 (2009).
4. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
5. Lynch, S. V. & Pedersen, O. The Human Intestinal Microbiome in Health and Disease. *N. Engl. J. Med.* **375**, 2369–2379 (2016).
6. German, J. B., Hammock, B. D. & Watkins, S. M. Metabolomics: building on a century of biochemistry to guide human health. *Metabolomics* **1**, 3–9 (2005).
7. Oliphant, K. & Allen-Vercoe, E. Macronutrient metabolism by the human gut microbiome: major fermentation by-products and their impact on host health. *Microbiome* **7**, 91 (2019).
8. Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R. & Goodman, A. L. Mapping human microbiome drug metabolism by gut bacteria and their genes. *Nature* **570**, 462–467 (2019).
9. Sharon, G. *et al.* Specialized metabolites from the microbiome in health and disease. *Cell Metab.* **20**, 719–730 (2014).
10. Clarke, G. *et al.* Minireview: Gut microbiota: the neglected endocrine organ. *Mol. Endocrinol.* **28**, 1221–1238 (2014).
11. David, L. A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014).
12. Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105–108 (2011).
13. Walker, A. W. *et al.* Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *ISME J.* **5**, 220–230 (2011).
14. *The Gastrointestinal System: Gastrointestinal, Nutritional and Hepatobiliary Physiology*. (Springer, Dordrecht, 2014).
15. Martinez-Guryn, K. *et al.* Small Intestine Microbiota Regulate Host Digestive and Absorptive Adaptive Responses to Dietary Lipids. *Cell Host Microbe* **23**, 458–469.e5 (2018).
16. Rowland, I. *et al.* Gut microbiota functions: metabolism of nutrients and other food components. *Eur. J. Nutr.* **57**, 1–24 (2018).
17. Macfarlane, G. T. & Macfarlane, S. Bacteria, colonic fermentation, and gastrointestinal health. *J. AOAC Int.* **95**, 50–60 (2012).
18. Roediger, W. E. Role of anaerobic bacteria in the metabolic welfare of the colonic mucosa in man. *Gut* **21**, 793–798 (1980).
19. Tan, J. *et al.* The role of short-chain fatty acids in health and disease. *Adv. Immunol.* **121**,

91–119 (2014).
20. Ruppin, H., Bar-Meir, S., Soergel, K. H., Wood, C. M. & Schmitt, M. G., Jr. Absorption of short-chain fatty acids by the colon. *Gastroenterology* **78**, 1500–1507 (1980).
21. McNeil, N. I., Cummings, J. H. & James, W. P. Short chain fatty acid absorption by the human large intestine. *Gut* **19**, 819–822 (1978).
22. Von Engelhardt, W., Rönnau, K., Rechkemmer, G. & Sakata, T. Absorption of short-chain fatty acids and their role in the hindgut of monogastric animals. *Anim. Feed Sci. Technol.* **23**, 43–53 (1989).
23. 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 (1987).
24. Manach, C., Scalbert, A., Morand, C., Rémésy, C. & Jiménez, L. Polyphenols: food sources and bioavailability. *Am. J. Clin. Nutr.* **79**, 727–747 (2004).
25. Macfarlane, G. T., Gibson, G. R. & Cummings, J. H. Comparison of fermentation reactions in different regions of the human colon. *J. Appl. Bacteriol.* **72**, 57–64 (1992).
26. Macfarlane, G. T., Gibson, G. R., Beatty, E. & Cummings, J. H. Estimation of short-chain fatty acid production from protein by human intestinal bacteria based on branched-chain fatty acid measurements. *FEMS Microbiol. Lett.* **101**, 81–88 (1992).
27. Geypens, B. *et al.* Influence of dietary protein supplements on the formation of bacterial metabolites in the colon. *Gut* **41**, 70–76 (1997).
28. Macfarlane, G. T. & Allison, C. Utilisation of protein by human gut bacteria. *FEMS Microbiol. Lett.* **38**, 19–24 (1986).
29. Cummings, J. H., Hill, M. J., Bone, E. S., Branch, W. J. & Jenkins, D. J. The effect of meat protein and dietary fiber on colonic function and metabolism. II. Bacterial metabolites in feces and urine. *Am. J. Clin. Nutr.* **32**, 2094–2101 (1979).
30. Portune, K. J. *et al.* Gut microbiota role in dietary protein metabolism and health-related outcomes: The two sides of the coin. *Trends Food Sci. Technol.* **57**, 213–232 (2016).
31. Yao, C. K., Muir, J. G. & Gibson, P. R. Review article: insights into colonic protein fermentation, its modulation and potential health implications. *Aliment. Pharmacol. Ther.* **43**, 181–196 (2016).
32. Aronov, P. A. *et al.* Colonic contribution to uremic solutes. *J. Am. Soc. Nephrol.* **22**, 1769–1776 (2011).
33. Windey, K. *et al.* Modulation of protein fermentation does not affect fecal water toxicity: a randomized cross-over study in healthy subjects. *PLoS One* **7**, e52387 (2012).
34. Holmes, E., Li, J. V., Athanasiou, T., Ashrafian, H. & Nicholson, J. K. Understanding the role of gut microbiome-host metabolic signal disruption in health and disease. *Trends Microbiol.* **19**, 349–359 (2011).
35. Romano, K. A., Vivas, E. I., Amador-Noguez, D. & Rey, F. E. Intestinal microbiota composition modulates choline bioavailability from diet and accumulation of the proatherogenic metabolite trimethylamine-N-oxide. *MBio* **6**, e02481 (2015).
36. Wang, Z. *et al.* Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57–63 (2011).
37. Koeth, R. A. *et al.* Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat,

promotes atherosclerosis. *Nat. Med.* **19**, 576–585 (2013).
38. Velasquez, M. T., Ramezani, A., Manal, A. & Raj, D. S. Trimethylamine N-Oxide: The Good, the Bad and the Unknown. *Toxins* **8**, (2016).
39. Deutscher, J. The mechanisms of carbon catabolite repression in bacteria. *Curr. Opin. Microbiol.* **11**, 87–93 (2008).
40. Smith, E. A. & Macfarlane, G. T. Enumeration of amino acid fermenting bacteria in the human large intestine: effects of pH and starch on peptide metabolism and dissimilation of amino acids. *FEMS Microbiol. Ecol.* **25**, 355–368 (1998).
41. Smith, E. A. & Macfarlane, G. T. Formation of Phenolic and Indolic Compounds by Anaerobic Bacteria in the Human Large Intestine. *Microb. Ecol.* **33**, 180–188 (1997).
42. Ridlon, J. M., Kang, D.-J. & Hylemon, P. B. Bile salt biotransformations by human intestinal bacteria. *J. Lipid Res.* **47**, 241–259 (2006).
43. Ridlon, J. M., Harris, S. C., Bhowmik, S., Kang, D.-J. & Hylemon, P. B. Consequences of bile salt biotransformations by intestinal bacteria. *Gut Microbes* **7**, 22–39 (2016).
44. Dai, Z.-L., Wu, G. & Zhu, W.-Y. Amino acid metabolism in intestinal bacteria: links between gut ecology and host health. *Front. Biosci.* **16**, 1768–1786 (2011).
45. Hill, M. J. Intestinal flora and endogenous vitamin synthesis. *Eur. J. Cancer Prev.* **6 Suppl 1**, S43–5 (1997).
46. Magnúsdóttir, S., Ravcheev, D., de Crécy-Lagard, V. & Thiele, I. Systematic genome assessment of B-vitamin biosynthesis suggests co-operation among gut microbes. *Front. Genet.* **6**, 148 (2015).
47. Said, H. M. & Mohammed, Z. M. Intestinal absorption of water-soluble vitamins: an update. *Curr. Opin. Gastroenterol.* **22**, 140–146 (2006).
48. Gustafsson, B. E. *et al.* Effects of vitamin K-active compounds and intestinal micro-organisms in vitamin K-deficient germfree rats. *J. Nutr.* **78**, 461–468 (1962).
49. Frick, P. G., Riedler, G. & Brögli, H. Dose response and minimal daily requirement for vitamin K in man. *J. Appl. Physiol.* **23**, 387–389 (1967).
50. Maier, L. *et al.* Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature* **555**, 623–628 (2018).
51. Wilson, I. D. & Nicholson, J. K. Gut microbiome interactions with drug metabolism, efficacy, and toxicity. *Transl. Res.* **179**, 204–222 (2017).
52. Jain, A., Gupta, Y. & Jain, S. K. Azo chemistry and its potential for colonic delivery. *Crit. Rev. Ther. Drug Carrier Syst.* **23**, 349–400 (2006).
53. Haiser, H. J. *et al.* Predicting and manipulating cardiac drug inactivation by the human gut bacterium Eggerthella lenta. *Science* **341**, 295–298 (2013).
54. Wallace, B. D. *et al.* Alleviating cancer drug toxicity by inhibiting a bacterial enzyme. *Science* **330**, 831–835 (2010).
55. LoGuidice, A., Wallace, B. D., Bendel, L., Redinbo, M. R. & Boelsterli, U. A. Pharmacologic targeting of bacterial β-glucuronidase alleviates nonsteroidal anti-inflammatory drug-induced enteropathy in mice. *J. Pharmacol. Exp. Ther.* **341**, 447–454 (2012).
56. Bhatt, A. P. *et al.* Targeted inhibition of gut bacterial β-glucuronidase activity enhances anticancer drug efficacy. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 7374–7381 (2020).

57. Clayton, T. A., Baker, D., Lindon, J. C., Everett, J. R. & Nicholson, J. K. Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 14728–14733 (2009).
58. Sakata, T. Pitfalls in short-chain fatty acid research: A methodological review. *Anim. Sci. J.* **90**, 3–13 (2019).
59. Wikoff, W. R. *et al.* Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 3698–3703 (2009).
60. Wilmanski, T. *et al.* Blood metabolome predicts gut microbiome α-diversity in humans. *Nat. Biotechnol.* **37**, 1217–1228 (2019).
61. Marcobal, A. *et al.* A metabolomic view of how the human gut microbiota impacts the host metabolome using humanized and gnotobiotic mice. *ISME J.* **7**, 1933–1943 (2013).
62. Claus, S. P. *et al.* Systemic multicompartmental effects of the gut microbiome on mouse metabolic phenotypes. *Mol. Syst. Biol.* **4**, 219 (2008).
63. Swann, J. R. *et al.* Variation in antibiotic-induced microbial recolonization impacts on the host metabolic phenotypes of rats. *J. Proteome Res.* **10**, 3590–3603 (2011).
64. O'Toole, P. W. & Claesson, M. J. Gut microbiota: Changes throughout the lifespan from infancy to elderly. *Int. Dairy J.* **20**, 281–291 (2010).
65. Gluckman, P. D., Cutfield, W., Hofman, P. & Hanson, M. A. The fetal, neonatal, and infant environments-the long-term consequences for disease risk. *Early Hum. Dev.* **81**, 51–59 (2005).
66. Belbasis, L., Savvidou, M. D., Kanu, C., Evangelou, E. & Tzoulaki, I. Birth weight in relation to health and disease in later life: an umbrella review of systematic reviews and meta-analyses. *BMC Med.* **14**, 147 (2016).
67. Hay, W. W., Jr. Placental transport of nutrients to the fetus. *Horm. Res.* **42**, 215–222 (1994).
68. Robertson, R. C., Manges, A. R., Brett Finlay, B. & Prendergast, A. J. The Human Microbiome and Child Growth – First 1000 Days and Beyond. *Trends in Microbiology* vol. 27 131–147 (2019).
69. Derrien, M., Alvarez, A.-S. & de Vos, W. M. The Gut Microbiota in the First Decade of Life. *Trends Microbiol.* **27**, 997–1010 (2019).
70. Soma-Pillay, P., Nelson-Piercy, C., Tolppanen, H. & Mebazaa, A. Physiological changes in pregnancy. *Cardiovasc. J. Afr.* **27**, 89–94 (2016).
71. Lockitch, G. Clinical biochemistry of pregnancy. *Crit. Rev. Clin. Lab. Sci.* **34**, 67–139 (1997).
72. Koren, O. *et al.* Host remodeling of the gut microbiome and metabolic changes during pregnancy. *Cell* **150**, 470–480 (2012).
73. Collado, M. C., Isolauri, E., Laitinen, K. & Salminen, S. Distinct composition of gut microbiota during pregnancy in overweight and normal-weight women. *Am. J. Clin. Nutr.* **88**, 894–899 (2008).
74. DiGiulio, D. B. *et al.* Temporal and spatial variation of the human microbiota during pregnancy. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 11060–11065 (2015).
75. Jašarević, E. & Bale, T. L. Prenatal and postnatal contributions of the maternal microbiome

on offspring programming. *Front. Neuroendocrinol.* **55**, 100797 (2019).
76. Gomez de Agüero, M. *et al.* The maternal microbiota drives early postnatal innate immune development. *Science* **351**, 1296–1302 (2016).
77. Kimura, I. *et al.* Maternal gut microbiota in pregnancy influences offspring metabolic phenotype in mice. *Science* **367**, (2020).
78. Arrieta, M.-C. *et al.* Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Sci. Transl. Med.* **7**, 307ra152 (2015).
79. Abrahamsson, T. R., Wu, R. Y. & Jenmalm, M. C. Gut microbiota and allergy: the importance of the pregnancy period. *Pediatr. Res.* **77**, 214–219 (2015).
80. Korpela, K. & de Vos, W. M. Early life colonization of the human gut: microbes matter everywhere. *Curr. Opin. Microbiol.* **44**, 70–78 (2018).
81. Chu, D. M. *et al.* Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. *Nat. Med.* **23**, 314–326 (2017).
82. Marcobal, A. & Sonnenburg, J. L. Human milk oligosaccharide consumption by intestinal microbiota. *Clin. Microbiol. Infect.* **18 Suppl 4**, 12–15 (2012).
83. Giallourou, N. *et al.* Metabolic maturation in the first 2 years of life in resource-constrained settings and its association with postnatal growths. *Sci Adv* **6**, eaay5969 (2020).
84. Tamburini, S., Shen, N., Wu, H. C. & Clemente, J. C. The microbiome in early life: implications for health outcomes. *Nat. Med.* **22**, 713–722 (2016).
85. Hsiao, E. Y. *et al.* Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. *Cell* **155**, 1451–1463 (2013).
86. Falony, G. *et al.* Population-level analysis of gut microbiome variation. *Science* **352**, 560–564 (2016).
87. Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227 (2012).
88. Power, S. E., O'Toole, P. W., Stanton, C., Ross, R. P. & Fitzgerald, G. F. Intestinal microbiota, diet and health. *Br. J. Nutr.* **111**, 387–402 (2014).
89. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
90. Duvallet, C., Gibbons, S. M., Gurry, T., Irizarry, R. A. & Alm, E. J. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat. Commun.* **8**, 1784 (2017).
91. Information sheet on obesity and overweight. *World Health Organization* https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight.
92. Jumpertz, R. *et al.* Energy-balance studies reveal associations between gut microbes, caloric load, and nutrient absorption in humans. *Am. J. Clin. Nutr.* **94**, 58–65 (2011).
93. Sze, M. A. & Schloss, P. D. Looking for a Signal in the Noise: Revisiting Obesity and the Microbiome. *MBio* **7**, (2016).
94. Walters, W. A., Xu, Z. & Knight, R. Meta-analyses of human gut microbes associated with obesity and IBD. *FEBS Lett.* **588**, 4223–4233 (2014).
95. Schwiertz, A. *et al.* Microbiota and SCFA in lean and overweight healthy subjects. *Obesity* **18**, 190–195 (2010).

96. Fernandes, J., Su, W., Rahat-Rozenbloom, S., Wolever, T. M. S. & Comelli, E. M. Adiposity, gut microbiota and faecal short chain fatty acids are linked in adult humans. *Nutr. Diabetes* **4**, e121 (2014).
97. Müller, M. *et al.* Circulating but not faecal short-chain fatty acids are related to insulin sensitivity, lipolysis and GLP-1 concentrations in humans. *Sci. Rep.* **9**, 12515 (2019).
98. Rahat-Rozenbloom, S., Fernandes, J., Gloor, G. B. & Wolever, T. M. S. Evidence for greater production of colonic short-chain fatty acids in overweight than lean humans. *Int. J. Obes.* **38**, 1525–1531 (2014).
99. Elliott, P. *et al.* Urinary metabolic signatures of human adiposity. *Sci. Transl. Med.* **7**, 285ra62 (2015).
100. Rangel-Huerta, O. D., Pastor-Villaescusa, B. & Gil, A. Are we close to defining a metabolomic signature of human obesity? A systematic review of metabolomics studies. *Metabolomics* **15**, 93 (2019).
101. Calvani, R. *et al.* Gut microbiome-derived metabolites characterize a peculiar obese urinary metabotype. *Int. J. Obes.* **34**, 1095–1098 (2010).
102. Frost, G. *et al.* The short-chain fatty acid acetate reduces appetite via a central homeostatic mechanism. *Nat. Commun.* **5**, 3611 (2014).
103. Perry, R. J. *et al.* Acetate mediates a microbiome-brain-β-cell axis to promote metabolic syndrome. *Nature* **534**, 213–217 (2016).
104. 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 Metab.* **9**, 311–326 (2009).
105. Pedersen, H. K. *et al.* Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature* **535**, 376–381 (2016).
106. Tai, E. S. *et al.* Insulin resistance is associated with a metabolic profile of altered protein metabolism in Chinese and Asian-Indian men. *Diabetologia* **53**, 757–767 (2010).
107. Walsh, C. J., Healy, S., O'Toole, P. W., Murphy, E. F. & Cotter, P. D. The probiotic L. casei LC-XCAL™ improves metabolic health in a diet-induced obesity mouse model without altering the microbiome. *Gut Microbes* 1–17 (2020).
108. Rasmussen, T. S. *et al.* Faecal virome transplantation decreases symptoms of type 2 diabetes and obesity in a murine model. *Gut* (2020) doi:10.1136/gutjnl-2019-320005.
109. Warner, B. B. The contribution of the gut microbiome to neurodevelopment and neuropsychiatric disorders. *Pediatr. Res.* **85**, 216–224 (2019).
110. Dinan, T. G. & Cryan, J. F. The Microbiome-Gut-Brain Axis in Health and Disease. *Gastroenterol. Clin. North Am.* **46**, 77–89 (2017).
111. Bauman, M. L. Medical comorbidities in autism: challenges to diagnosis and treatment. *Neurotherapeutics* **7**, 320–327 (2010).
112. Sandler, R. H. *et al.* Short-term benefit from oral vancomycin treatment of regressive-onset autism. *J. Child Neurol.* **15**, 429–435 (2000).
113. Adams, J. B., Johansen, L. J., Powell, L. D., Quig, D. & Rubin, R. A. Gastrointestinal flora and gastrointestinal status in children with autism--comparisons to typical children and correlation with autism severity. *BMC Gastroenterol.* **11**, 22 (2011).
114. Wang, L. W., Tancredi, D. J. & Thomas, D. W. The prevalence of gastrointestinal problems

in children across the United States with autism spectrum disorders from families with multiple affected members. *J. Dev. Behav. Pediatr.* **32**, 351–360 (2011).
115. Bezawada, N., Phang, T. H., Hold, G. L. & Hansen, R. Autism Spectrum Disorder and the Gut Microbiota in Children: A Systematic Review. *Ann. Nutr. Metab.* 1–14 (2020).
116. Ho, L. K. H. *et al.* Gut microbiota changes in children with autism spectrum disorder: a systematic review. *Gut Pathog.* **12**, 6 (2020).
117. Iglesias-Vázquez, L., Van Ginkel Riba, G., Arija, V. & Canals, J. Composition of Gut Microbiota in Children with Autism Spectrum Disorder: A Systematic Review and Meta-Analysis. *Nutrients* **12**, (2020).
118. Xu, M., Xu, X., Li, J. & Li, F. Association Between Gut Microbiota and Autism Spectrum Disorder: A Systematic Review and Meta-Analysis. *Front. Psychiatry* **10**, 473 (2019).
119. Lacorte, E. *et al.* A Systematic Review of the Microbiome in Children With Neurodevelopmental Disorders. *Front. Neurol.* **10**, 727 (2019).
120. Diémé, B. *et al.* Metabolomics Study of Urine in Autism Spectrum Disorders Using a Multiplatform Analytical Methodology. *J. Proteome Res.* **14**, 5273–5282 (2015).
121. Gevi, F., Zolla, L., Gabriele, S. & Persico, A. M. Urinary metabolomics of young Italian autistic children supports abnormal tryptophan and purine metabolism. *Mol. Autism* **7**, 47 (2016).
122. Noto, A. *et al.* The urinary metabolomics profile of an Italian autistic children population and their unaffected siblings. *J. Matern. Fetal. Neonatal Med.* **27 Suppl 2**, 46–52 (2014).
123. Emond, P. *et al.* GC-MS-based urine metabolic profiling of autism spectrum disorders. *Anal. Bioanal. Chem.* **405**, 5291–5300 (2013).
124. Yap, I. K. S. *et al.* Urinary metabolic phenotyping differentiates children with autism from their unaffected siblings and age-matched controls. *J. Proteome Res.* **9**, 2996–3004 (2010).
125. Ming, X., Stein, T. P., Barnes, V., Rhodes, N. & Guo, L. Metabolic perturbance in autism spectrum disorders: a metabolomics study. *J. Proteome Res.* **11**, 5856–5862 (2012).
126. Gabriele, S. *et al.* Slow intestinal transit contributes to elevate urinary p-cresol level in Italian autistic children. *Autism Res.* **9**, 752–759 (2016).
127. De Angelis, M. *et al.* Fecal microbiota and metabolome of children with autism and pervasive developmental disorder not otherwise specified. *PLoS One* **8**, e76993 (2013).
128. Kang, D.-W. *et al.* Long-term benefit of Microbiota Transfer Therapy on autism symptoms and gut microbiota. *Sci. Rep.* **9**, 5821 (2019).
129. Giles, E. M., D'Adamo, G. L. & Forster, S. C. The future of faecal transplants. *Nat. Rev. Microbiol.* **17**, 719 (2019).
130. Organization, W. H. & Others. *Global status report on noncommunicable diseases 2014*. (World Health Organization, 2014).
131. Maric, T., Kanu, C., Johnson, M. R. & Savvidou, M. D. Maternal, neonatal insulin resistance and neonatal anthropometrics in pregnancies following bariatric surgery. *Metabolism* **97**, 25–31 (2019).
132. West, K. A. *et al.* Longitudinal metabolic and gut bacterial profiling of pregnant women with previous bariatric surgery. *Gut* (2020) doi:10.1136/gutjnl-2019-319620.
133. Maric, T. *et al.* Fetal growth and fetoplacental circulation in pregnancies following bariatric

surgery: a prospective study. *BJOG* (2020) doi:10.1111/1471-0528.16105.
134. Diabetes in pregnancy: management from preconception to the postnatal period. *National Institute for Health and Care Excellence* https://www.nice.org.uk/guidance/ng3 (2015).
135. Matthews, D. R. *et al.* Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* **28**, 412–419 (1985).
136. Emwas, A.-H. *et al.* NMR Spectroscopy for Metabolomics Research. *Metabolites* **9**, (2019).
137. Keeler, J. *Understanding NMR Spectroscopy*. (John Wiley & Sons, 2011).
138. Rose, C., Parker, A., Jefferson, B. & Cartmell, E. The Characterization of Feces and Urine: A Review of the Literature to Inform Advanced Treatment Technology. *Crit. Rev. Environ. Sci. Technol.* **45**, 1827–1879 (2015).
139. Dona, A. C. *et al.* Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping. *Anal. Chem.* **86**, 9887–9894 (2014).
140. Yu, Z. *et al.* Differences between human plasma and serum metabolite profiles. *PLoS One* **6**, e21230 (2011).
141. Beckonert, O. *et al.* Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat. Protoc.* **2**, 2692–2703 (2007).
142. Dona, A. C. *et al.* A guide to the identification of metabolites in NMR-based metabonomics/metabolomics experiments. *Comput. Struct. Biotechnol. J.* **14**, 135–153 (2016).
143. Pohl, L. & Eckle, M. Sodium 3-Trimethylsilyltetradeuteriopropionate, a New Water-Soluble Standard for1H-NMR. *Angew. Chem. Int. Ed Engl.* **8**, 381–381 (1969).
144. Haggart, G. A. *csmsoftware/IMPaCTS: Version 1.1.1*. (2019). doi:10.5281/zenodo.3077413.
145. Veselkov, K. A. *et al.* Recursive segment-wise peak alignment of biological 1H NMR spectra for improved metabolic biomarker recovery. *Anal. Chem.* **81**, 56–66 (2009).
146. Dieterle, F., Ross, A., Schlotterbeck, G. & Senn, H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Anal. Chem.* **78**, 4281–4290 (2006).
147. Cloarec, O. *et al.* Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic 1H NMR data sets. *Anal. Chem.* **77**, 1282–1289 (2005).
148. Posma, J. M. *et al.* Subset optimization by reference matching (STORM): an optimized statistical approach for recovery of metabolic biomarker structural information from 1H NMR spectra of biofluids. *Anal. Chem.* **84**, 10694–10701 (2012).
149. Wishart, D. S. *et al.* HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.* **46**, D608–D617 (2018).
150. Janda, J. M. & Abbott, S. L. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J. Clin. Microbiol.* **45**, 2761–2764 (2007).
151. Yang, B., Wang, Y. & Qian, P.-Y. Sensitivity and correlation of hypervariable regions in 16S

rRNA genes in phylogenetic analysis. *BMC Bioinformatics* **17**, 135 (2016).
152. Gevers, D. *et al.* The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* **15**, 382–392 (2014).
153. Mullish, B. H. *et al.* Functional microbiomics: Evaluation of gut microbiota-bile acid metabolism interactions in health and disease. *Methods* **149**, 49–58 (2018).
154. An introduction to Next-Generation Sequencing Technology. *Illumina* https://emea.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf.
155. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
156. R Core Team. *R: A language and environment for statistical computing*. (Vienna, Austria, 2013).
157. Callahan, B. J., Sankaran, K., Fukuyama, J. A., McMurdie, P. J. & Holmes, S. P. Bioconductor Workflow for Microbiome Data Analysis: from raw reads to community analyses. *F1000Res.* **5**, 1492 (2016).
158. Kwon, S., Park, S., Lee, B. & Yoon, S. In-depth analysis of interrelation between quality scores and real errors in Illumina reads. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2013**, 635–638 (2013).
159. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
160. Cole, J. R. *et al.* Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* **42**, D633–42 (2014).
161. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–6 (2013).
162. Schliep, K. P. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2011).
163. McMurdie, P. J. & Holmes, S. Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput. Biol.* **10**, e1003531 (2014).
164. Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V. & Egozcue, J. J. Microbiome Datasets Are Compositional: And This Is Not Optional. *Front. Microbiol.* **8**, 2224 (2017).
165. Aitchison, J. The Statistical Analysis of Compositional Data. *J. R. Stat. Soc. Series B Stat. Methodol.* **44**, 139–160 (1982).
166. Pearson, K. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **2**, 559–572 (1901).
167. Shen, H. & Huang, J. Z. Sparse principal component analysis via regularized low rank matrix approximation. *J. Multivar. Anal.* **99**, 1015–1034 (2008).
168. Wold, H. Estimation of principal components and related models by iterative least squares. *J. Multivar. Anal.* 391–420 (1966).
169. Wold, S., Sjöström, M. & Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemometrics Intellig. Lab. Syst.* **58**, 109–130 (2001).
170. Rohart, F., Gautier, B., Singh, A. & Lê Cao, K.-A. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* **13**, e1005752 (2017).
171. Lê Cao, K.-A., Rossouw, D., Robert-Granié, C. & Besse, P. A sparse PLS for variable selection when integrating omics data. *Stat. Appl. Genet. Mol. Biol.* **7**, Article 35 (2008).

172. Trygg, J. & Wold, S. Orthogonal projections to latent structures (O-PLS). *J. Chemom.* **16**, 119–128 (2002).
173. Singh, A. *et al.* DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinformatics* **35**, 3055–3062 (2019).
174. Mevik, B.-H. & Cederkvist, H. R. Mean squared error of prediction (MSEP) estimates for principal component regression (PCR) and partial least squares regression (PLSR). *J. Chemom.* **18**, 422–429 (2004).
175. Chavent, M. & Patouille, B. Calcul des coefficients de régression et du PRESS en régression PLS1. *Modulad n* **30**, 1–11 (2003).
176. Hair, J. F., Ringle, C. M. & Sarstedt, M. Editorial - Partial Least Squares Structural Equation Modeling: Rigorous Applications, Better Results and Higher Acceptance. *Long range planning* (2013).
177. Worley, B. & Powers, R. Multivariate Analysis in Metabolomics. *Curr Metabolomics* **1**, 92–107 (2013).
178. González, I., Cao, K.-A. L., Davis, M. J. & Déjean, S. Visualising associations between paired 'omics' data sets. *BioData Min.* **5**, 19 (2012).
179. Butte, A. J., Tamayo, P., Slonim, D., Golub, T. R. & Kohane, I. S. Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 12182–12186 (2000).
180. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. (Springer Science & Business Media, 2009).
181. Baltagi, B. H. *Econometric Analysis of Panel Data*. (Wiley, 2005).
182. Kenward, M. G. & Roger, J. H. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* **53**, 983–997 (1997).
183. Wolfer, A. Time-dependent metabolic phenotyping of inflammatory dysregulation. (Imperial College London, 2016).
184. Pearson, K. Notes on Regression and Inheritance in the Case of Two Parents. in *Proceedings of the Royal Society of London* vol. 58 240–242 (1895).
185. Spearman, C. The proof and measurement of association between two things. *Am. J. Psychol.* **15**, 72–101 (1904).
186. Whittaker, J. *Graphical Models in Applied Multivariate Statistics*. (Wiley Publishing, 2009).
187. Kim, S. ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients. *Commun Stat Appl Methods* **22**, 665–674 (2015).
188. Hinkle, D. E., Wiersma, W. & Jurs, S. G. *Applied statistics for the behavioral sciences*. vol. 663 (Houghton Mifflin College Division, 2003).
189. McDonald, J. H. *Handbook of Biological Statistics*. vol. 2 (Sparky House Publishing, 2009).
190. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).
191. Catalano, P. M. & Ehrenberg, H. M. The short-and long-term implications of maternal obesity on the mother and her offspring. *BJOG* **113**, 1126–1133 (2006).
192. Leddy, M. A., Power, M. L. & Schulkin, J. The impact of maternal obesity on maternal and fetal health. *Rev. Obstet. Gynecol.* **1**, 170–178 (2008).

193. Santangeli, L., Sattar, N. & Huda, S. S. Impact of maternal obesity on perinatal and childhood outcomes. *Best Pract. Res. Clin. Obstet. Gynaecol.* **29**, 438–448 (2015).
194. Catalano, P. M., Presley, L., Minium, J. & Hauguel-de Mouzon, S. Fetuses of obese mothers develop insulin resistance in utero. *Diabetes Care* **32**, 1076–1080 (2009).
195. Catalano, P. M. & Shankar, K. Obesity and pregnancy: mechanisms of short term and long term adverse consequences for mother and child. *BMJ* **356**, j1 (2017).
196. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457**, 480–484 (2009).
197. Finucane, M. M., Sharpton, T. J., Laurent, T. J. & Pollard, K. S. A taxonomic signature of obesity in the microbiome? Getting to the guts of the matter. *PLoS One* **9**, e84689 (2014).
198. Santacruz, A. *et al.* Gut microbiota composition is associated with body weight, weight gain and biochemical parameters in pregnant women. *Br. J. Nutr.* **104**, 83–92 (2010).
199. Zacarías, M. F. *et al.* Pregestational overweight and obesity are associated with differences in gut microbiota composition and systemic inflammation in the third trimester. *PLoS One* **13**, e0200305 (2018).
200. Stanislawski, M. A. *et al.* Pre-pregnancy weight, gestational weight gain, and the gut microbiota of mothers and their infants. *Microbiome* **5**, 113 (2017).
201. Payne, A. N. *et al.* The metabolic activity of gut microbiota in obese children is increased compared with normal-weight children and exhibits more exhaustive substrate utilization. *Nutr. Diabetes* **1**, e12–e12 (2011).
202. Murugesan, S. *et al.* Study of the diversity and short-chain fatty acids production by the bacterial community in overweight and obese Mexican children. *Eur. J. Clin. Microbiol. Infect. Dis.* **34**, 1337–1346 (2015).
203. Priyadarshini, M. *et al.* Maternal short-chain fatty acids are associated with metabolic parameters in mothers and newborns. *Transl. Res.* **164**, 153–157 (2014).
204. Desert, R., Canlet, C., Costet, N., Cordier, S. & Bonvallot, N. Impact of maternal obesity on the metabolic profiles of pregnant women and their offspring at birth. *Metabolomics* **11**, 1896–1907 (2015).
205. Young, M. & Prenton, M. A. Maternal and fetal plasma amino acid concentrations during gestation and in retarded fetal growth. *J. Obstet. Gynaecol. Br. Commonw.* **76**, 333–334 (1969).
206. Lain, K. Y. & Catalano, P. M. Metabolic changes in pregnancy. *Clin. Obstet. Gynecol.* **50**, 938–948 (2007).
207. Di Cianni, G., Miccoli, R., Volpe, L., Lencioni, C. & Del Prato, S. Intermediate metabolism in normal pregnancy and in gestational diabetes. *Diabetes. Metab. Res. Rev.* **19**, 259–270 (2003).
208. Sonagra, A. D., Biradar, S. M., K, D. & Murthy D S, J. Normal pregnancy- a state of insulin resistance. *J. Clin. Diagn. Res.* **8**, CC01–3 (2014).
209. Belo, L. *et al.* Fluctuations in C-reactive protein concentration and neutrophil activation during normal human pregnancy. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **123**, 46–51 (2005).
210. Houttu, N., Mokkala, K. & Laitinen, K. Overweight and obesity status in pregnant women are related to intestinal microbiota and serum metabolic and inflammatory profiles. *Clin.*

Nutr. **37**, 1955–1966 (2018).
211. Friis, C. M. *et al.* Adiposity-related inflammation: effects of pregnancy. *Obesity* **21**, E124–30 (2013).
212. Park, J. H. & Lee, S. Y. Fermentative production of branched chain amino acids: a focus on metabolic engineering. *Appl. Microbiol. Biotechnol.* **85**, 491–506 (2010).
213. Amorim Franco, T. M. & Blanchard, J. S. Bacterial Branched-Chain Amino Acid Biosynthesis: Structures, Mechanisms, and Drugability. *Biochemistry* **56**, 5849–5865 (2017).
214. Diaz, S. O. *et al.* Second trimester maternal urine for the diagnosis of trisomy 21 and prediction of poor pregnancy outcomes. *J. Proteome Res.* **12**, 2946–2957 (2013).
215. Kadakia, R. *et al.* Maternal metabolites during pregnancy are associated with newborn outcomes and hyperinsulinaemia across ancestries. *Diabetologia* **62**, 473–484 (2019).
216. Maitre, L. *et al.* Urinary metabolic profiles in early pregnancy are associated with preterm birth and fetal growth restriction in the Rhea mother-child cohort study. *BMC Med.* **12**, 110 (2014).
217. Diaz, S. O. *et al.* Metabolic biomarkers of prenatal disorders: an exploratory NMR metabonomics study of second trimester maternal urine and blood plasma. *J. Proteome Res.* **10**, 3732–3742 (2011).
218. Wang, M. *et al.* Normal pregnancy induced glucose metabolic stress in a longitudinal cohort of healthy women: Novel insights generated from a urine metabolomics study. *Medicine* **97**, e12417 (2018).
219. Maitre, L. *et al.* Maternal urinary metabolic signatures of fetal growth and associated clinical and environmental factors in the INMA study. *BMC Med.* **14**, 177 (2016).
220. Sachse, D. *et al.* Metabolic changes in urine during and after pregnancy in a large, multiethnic population-based cohort study of gestational diabetes. *PLoS One* **7**, e52399 (2012).
221. Liu, X. *et al.* Urinary metabolic variation analysis during pregnancy and application in Gestational Diabetes Mellitus and spontaneous abortion biomarker discovery. *Sci. Rep.* **9**, 2605 (2019).
222. Mills, H. L. *et al.* The effect of a lifestyle intervention in obese pregnant women on gestational metabolic profiles: findings from the UK Pregnancies Better Eating and Activity Trial (UPBEAT) randomised controlled trial. *BMC Med.* **17**, 15 (2019).
223. Bakdash, J. Z. & Marusich, L. R. Repeated Measures Correlation. *Front. Psychol.* **8**, 456 (2017).
224. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
225. Rasmussen, K. M. *et al.* Recommendations for weight gain during pregnancy in the context of the obesity epidemic. *Obstet. Gynecol.* **116**, 1191–1195 (2010).
226. Nelson, S. M., Matthews, P. & Poston, L. Maternal metabolism and obesity: modifiable determinants of pregnancy outcome. *Hum. Reprod. Update* **16**, 255–275 (2010).
227. Cnattingius, S. *et al.* Maternal obesity and risk of preterm delivery. *JAMA* **309**, 2362–2370 (2013).
228. Chu, S. Y. *et al.* Maternal obesity and risk of cesarean delivery: a meta-analysis. *Obes.*

Rev. **8**, 385–394 (2007).
229. Li, H.-T., Zhou, Y.-B. & Liu, J.-M. The impact of cesarean section on offspring overweight and obesity: a systematic review and meta-analysis. *Int. J. Obes.* **37**, 893–899 (2013).
230. Sevelsted, A., Stokholm, J., Bønnelykke, K. & Bisgaard, H. Cesarean section and chronic immune disorders. *Pediatrics* **135**, e92–8 (2015).
231. Rutayisire, E., Huang, K., Liu, Y. & Tao, F. The mode of delivery affects the diversity and colonization pattern of the gut microbiota during the first year of infants' life: a systematic review. *BMC Gastroenterol.* **16**, 86 (2016).
232. Stinson, L. F., Payne, M. S. & Keelan, J. A. A Critical Review of the Bacterial Baptism Hypothesis and the Impact of Cesarean Delivery on the Infant Microbiome. *Front. Med.* **5**, 135 (2018).
233. Clasey, J. L. *et al.* The use of anthropometric and dual-energy X-ray absorptiometry (DXA) measures to estimate total abdominal and abdominal visceral fat in men and women. *Obes. Res.* **7**, 256–264 (1999).
234. Lawlor, D. A., Relton, C., Sattar, N. & Nelson, S. M. Maternal adiposity—a determinant of perinatal and offspring outcomes? *Nat. Rev. Endocrinol.* **8**, 679–688 (2012).
235. Pedersen, J. Diabetes and pregnancy; blood sugar of newborn infants during fasting and glucose administration. *Ugeskr. Laeger* **114**, 685 (1952).
236. HAPO Study Cooperative Research Group. Hyperglycemia and Adverse Pregnancy Outcome (HAPO) Study: associations with neonatal anthropometrics. *Diabetes* **58**, 453–459 (2009).
237. Hay, W. W., Jr. Recent observations on the regulation of fetal metabolism by glucose. *J. Physiol.* **572**, 17–24 (2006).
238. Wang, G. *et al.* Preterm birth and random plasma insulin levels at birth and in early childhood. *JAMA* **311**, 587–596 (2014).
239. Westermeier, F., Sáez, P. J., Villalobos-Labra, R., Sobrevia, L. & Farías-Jofré, M. Programming of fetal insulin resistance in pregnancies with maternal obesity by ER stress and inflammation. *Biomed Res. Int.* **2014**, 917672 (2014).
240. Oken, E. & Gillman, M. W. Fetal origins of obesity. *Obes. Res.* **11**, 496–506 (2003).
241. Connelly, M. A., Otvos, J. D., Shalaurova, I., Playford, M. P. & Mehta, N. N. GlycA, a novel biomarker of systemic inflammation and cardiovascular disease risk. *J. Transl. Med.* **15**, 219 (2017).
242. Ritchie, S. C. *et al.* The Biomarker GlycA Is Associated with Chronic Inflammation and Predicts Long-Term Risk of Severe Infection. *Cell Syst* **1**, 293–301 (2015).
243. Kalhan, S. C. Protein metabolism in pregnancy. *Am. J. Clin. Nutr.* **71**, 1249S–55S (2000).
244. Duggleby, S. L. & Jackson, A. A. Higher weight at birth is related to decreased maternal amino acid oxidation during pregnancy. *Am. J. Clin. Nutr.* **76**, 852–857 (2002).
245. Cheng, S. *et al.* Metabolite profiling identifies pathways associated with metabolic risk in humans. *Circulation* **125**, 2222–2231 (2012).
246. Abboud, K. Y. *et al.* Oral Glutamine Supplementation Reduces Obesity, Pro-Inflammatory Markers, and Improves Insulin Sensitivity in DIO Wistar Rats and Reduces Waist Circumference in Overweight and Obese Humans. *Nutrients* **11**, (2019).
247. Kalhan, S., Rossi, K., Gruca, L., Burkett, E. & O'Brien, A. Glucose turnover and

gluconeogenesis in human pregnancy. *J. Clin. Invest.* **100**, 1775–1781 (1997).
248. Hu, M. *et al.* Decreased maternal serum acetate and impaired fetal thymic and regulatory T cell development in preeclampsia. *Nat. Commun.* **10**, 3031 (2019).
249. Marques, F. Z. *et al.* High-Fiber Diet and Acetate Supplementation Change the Gut Microbiota and Prevent the Development of Hypertension and Heart Failure in Hypertensive Mice. *Circulation* **135**, 964–977 (2017).
250. Wang, Q. *et al.* Metabolic profiling of pregnancy: cross-sectional and longitudinal evidence. *BMC Med.* **14**, 205 (2016).
251. Vecchié, A. *et al.* C-Reactive Protein Levels at the Midpregnancy Can Predict Gestational Complications. *Biomed Res. Int.* **2018**, 1070151 (2018).
252. Rudolph, M. D. *et al.* Maternal IL-6 during pregnancy can be estimated from newborn brain connectivity and predicts future working memory in offspring. *Nat. Neurosci.* **21**, 765–772 (2018).
253. Gaillard, R., Rifas-Shiman, S. L., Perng, W., Oken, E. & Gillman, M. W. Maternal inflammation during pregnancy and childhood adiposity. *Obesity* **24**, 1320–1327 (2016).
254. Stokholm, J., Sevelsted, A., Anderson, U. D. & Bisgaard, H. Preeclampsia Associates with Asthma, Allergy, and Eczema in Childhood. *Am. J. Respir. Crit. Care Med.* **195**, 614–621 (2017).
255. Block, W. D., Hubbard, R. W. & Steele, B. F. Excretion of Histidine and Histidine Derivatives by Human Subjects Ingesting Protein from Different Sources. *J. Nutr.* **85**, 419–425 (1965).
256. Cetin, I. *et al.* Umbilical amino acid concentrations in appropriate and small for gestational age infants: a biochemical difference present in utero. *Am. J. Obstet. Gynecol.* **158**, 120–126 (1988).
257. Józwik, M. *et al.* Effects of branched-chain amino acids on placental amino acid transfer and insulin and glucagon release in the ovine fetus. *Am. J. Obstet. Gynecol.* **185**, 487–495 (2001).
258. Gürke, J. *et al.* Maternal Diabetes Leads to Adaptation in Embryonic Amino Acid Metabolism during Early Pregnancy. *PLoS One* **10**, e0127465 (2015).
259. Liu, L. X. & Arany, Z. Maternal cardiac metabolism in pregnancy. *Cardiovasc. Res.* **101**, 545–553 (2014).
260. Holm, M. B. *et al.* Uptake and release of amino acids in the fetal-placental unit in human pregnancies. *PLoS One* **12**, e0185760 (2017).
261. Battaglia, F. C. & Meschia, G. Principal substrates of fetal metabolism. *Physiol. Rev.* **58**, 499–527 (1978).
262. Hauguel, S., Challier, J. C., Cedard, L. & Olive, G. Metabolism of the human placenta perfused in vitro: glucose transfer and utilization, O2 consumption, lactate and ammonia production. *Pediatr. Res.* **17**, 729–732 (1983).
263. Ireland, Z., Russell, A. P., Wallimann, T., Walker, D. W. & Snow, R. Developmental changes in the expression of creatine synthesizing enzymes and creatine transporter in a precocial rodent, the spiny mouse. *BMC Dev. Biol.* **9**, 39 (2009).
264. Dickinson, H. *et al.* Creatine supplementation during pregnancy: summary of experimental studies suggesting a treatment to improve fetal and neonatal morbidity and reduce

mortality in high-risk human pregnancy. *BMC Pregnancy Childbirth* **14**, 150 (2014).
265. Thorburn, A. N. *et al.* Evidence that asthma is a developmental origin disease influenced by maternal diet and bacterial metabolites. *Nat. Commun.* **6**, 7320 (2015).
266. Burd, L. I. *et al.* Placental production and foetal utilisation of lactate and pyruvate. *Nature* **254**, 710–711 (1975).
267. Jackson, F. *et al.* Development of a Pipeline for Exploratory Metabolic Profiling of Infant Urine. *J. Proteome Res.* **15**, 3432–3440 (2016).
268. Diaz, S. O. *et al.* Following healthy pregnancy by nuclear magnetic resonance (NMR) metabolic profiling of human urine. *J. Proteome Res.* **12**, 969–979 (2013).
269. Hytten, F. E. The renal excretion of nutrients in pregnancy. *Postgrad. Med. J.* **49**, 625–629 (1973).
270. Odutayo, A. & Hladunewich, M. Obstetric nephrology: renal hemodynamic and metabolic physiology in normal pregnancy. *Clin. J. Am. Soc. Nephrol.* **7**, 2073–2080 (2012).
271. Robinson, H. L., Barrett, H. L., Foxcroft, K., Callaway, L. K. & Dekker Nitert, M. Prevalence of maternal urinary ketones in pregnancy in overweight and obese women. *Obstet Med* **11**, 79–82 (2018).
272. Herrera, E. Lipid metabolism in pregnancy and its consequences in the fetus and newborn. *Endocrine* **19**, 43–55 (2002).
273. Lever, M. & Slow, S. The clinical significance of betaine, an osmolyte with a key role in methyl group metabolism. *Clin. Biochem.* **43**, 732–744 (2010).
274. Wu, B. T. F., Dyer, R. A., King, D. J., Richardson, K. J. & Innis, S. M. Early second trimester maternal plasma choline and betaine are related to measures of early cognitive development in term infants. *PLoS One* **7**, e43448 (2012).
275. al-Waiz, M., Mikov, M., Mitchell, S. C. & Smith, R. L. The exogenous origin of trimethylamine in the mouse. *Metabolism* **41**, 135–136 (1992).
276. Dumas, M.-E. *et al.* Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 12511–12516 (2006).
277. Morrow, G. P. *et al.* In vivo kinetics of formate metabolism in folate-deficient and folate-replete rats. *J. Biol. Chem.* **290**, 2244–2250 (2015).
278. 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 (2016).
279. Pietzke, M., Meiser, J. & Vazquez, A. Formate metabolism in health and disease. *Mol Metab* **33**, 23–37 (2020).
280. Layden, B. T., Yalamanchi, S. K., Wolever, T. M., Dunaif, A. & Lowe, W. L., Jr. Negative association of acetate with visceral adipose tissue and insulin levels. *Diabetes Metab. Syndr. Obes.* **5**, 49–55 (2012).
281. Chen, M. *et al.* Metabolomics insights into the modulatory effects of long-term compound polysaccharide intake in high-fat diet-induced obese rats. *Nutr. Metab.* **15**, 8 (2018).
282. An, Y. *et al.* High-fat diet induces dynamic metabolic alterations in multiple biological matrices of rats. *J. Proteome Res.* **12**, 3755–3768 (2013).
283. Gerchman, F. *et al.* Body mass index is associated with increased creatinine clearance by a mechanism independent of body fat distribution. *J. Clin. Endocrinol. Metab.* **94**,

3781–3788 (2009).
284. Heymsfield, S. B., Arteaga, C., McManus, C., Smith, J. & Moffitt, S. Measurement of muscle mass in humans: validity of the 24-hour urinary creatinine method. *Am. J. Clin. Nutr.* **37**, 478–494 (1983).
285. Appiah-Amponsah, E. *et al.* Identification of 4-deoxythreonic acid present in human urine using HPLC and NMR techniques. *J. Pharm. Biomed. Anal.* **50**, 878–885 (2009).
286. Tremblay, F. *et al.* Overactivation of S6 kinase 1 as a cause of human insulin resistance during increased amino acid availability. *Diabetes* **54**, 2674–2684 (2005).
287. Wang, T. J. *et al.* Metabolite profiles and the risk of developing diabetes. *Nat. Med.* **17**, 448–453 (2011).
288. Neis, E. P. J. G., Dejong, C. H. C. & Rensen, S. S. The role of microbial amino acid metabolism in host metabolism. *Nutrients* **7**, 2930–2946 (2015).
289. Li, M. *et al.* Symbiotic gut microbes modulate human metabolic phenotypes. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 2117–2122 (2008).
290. Wijeyesekera, A. *et al.* Quantitative UPLC-MS/MS analysis of the gut microbial co-metabolites phenylacetylglutamine, 4-cresyl sulphate and hippurate in human urine: INTERMAP Study. *Anal. Methods* **4**, 65–72 (2012).
291. Poesen, R. *et al.* Microbiota-Derived Phenylacetylglutamine Associates with Overall Mortality and Cardiovascular Disease in Patients with CKD. *J. Am. Soc. Nephrol.* **27**, 3479–3487 (2016).
292. Lees, H. J., Swann, J. R., Wilson, I. D., Nicholson, J. K. & Holmes, E. Hippurate: the natural history of a mammalian-microbial cometabolite. *J. Proteome Res.* **12**, 1527–1546 (2013).
293. Olthof, M. R., Hollman, P. C. H., Buijsman, M. N. C. P., van Amelsvoort, J. M. M. & Katan, M. B. Chlorogenic acid, quercetin-3-rutinoside and black tea phenols are extensively metabolized in humans. *J. Nutr.* **133**, 1806–1814 (2003).
294. Fernández-Veledo, S. & Vendrell, J. Gut microbiota-derived succinate: Friend or foe in human metabolic diseases? *Rev. Endocr. Metab. Disord.* **20**, 439–447 (2019).
295. Li, X. *et al.* Comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry for metabonomics: Biomarker discovery for diabetes mellitus. *Anal. Chim. Acta* **633**, 257–262 (2009).
296. de Luca, C. & Olefsky, J. M. Inflammation and insulin resistance. *FEBS Lett.* **582**, 97–105 (2008).
297. Palmer, N. D. *et al.* Metabolomic profile associated with insulin resistance and conversion to diabetes in the Insulin Resistance Atherosclerosis Study. *J. Clin. Endocrinol. Metab.* **100**, E463–8 (2015).
298. Cirulli, E. T. *et al.* Profound Perturbation of the Metabolome in Obesity Is Associated with Health Risk. *Cell Metab.* **29**, 488–500.e2 (2019).
299. Jacob, S. *et al.* Targeted Metabolomics Demonstrates Distinct and Overlapping Maternal Metabolites Associated With BMI, Glucose, and Insulin Sensitivity During Pregnancy Across Four Ancestry Groups. *Diabetes Care* **40**, 911–919 (2017).
300. Kampmann, U., Knorr, S., Fuglsang, J. & Ovesen, P. Determinants of Maternal Insulin Resistance during Pregnancy: An Updated Overview. *J Diabetes Res* **2019**, 5320156

(2019).
301. Butte, N. F. Carbohydrate and lipid metabolism in pregnancy: normal compared with gestational diabetes mellitus. *Am. J. Clin. Nutr.* **71**, 1256S–61S (2000).
302. Hernández, M. A. G., Canfora, E. E., Jocken, J. W. E. & Blaak, E. E. The Short-Chain Fatty Acid Acetate in Body Weight Control and Insulin Sensitivity. *Nutrients* **11**, (2019).
303. Riddy, D. M., Delerive, P., Summers, R. J., Sexton, P. M. & Langmead, C. J. G Protein–Coupled Receptors Targeting Insulin Resistance, Obesity, and Type 2 Diabetes Mellitus. *Pharmacol. Rev.* **70**, 39–67 (2018).
304. Hellmuth, C. *et al.* Cord Blood Metabolome Is Highly Associated with Birth Weight, but Less Predictive for Later Weight Development. *Obes. Facts* **10**, 85–100 (2017).
305. Floyd, J. C., Jr, Fajans, S. S., Conn, J. W., Knopf, R. F. & Rull, J. Stimulation of insulin secretion by amino acids. *J. Clin. Invest.* **45**, 1487–1502 (1966).
306. Brown, L. D. *et al.* Chronically Increased Amino Acids Improve Insulin Secretion, Pancreatic Vascularity, and Islet Size in Growth-Restricted Fetal Sheep. *Endocrinology* **157**, 3788–3799 (2016).
307. Elder, K. A. & Wolfe, B. M. Bariatric surgery: a review of procedures and outcomes. *Gastroenterology* **132**, 2253–2271 (2007).
308. Chang, S.-H. *et al.* The effectiveness and risks of bariatric surgery: an updated systematic review and meta-analysis, 2003-2012. *JAMA Surg.* **149**, 275–287 (2014).
309. Rao, R. S., Yanagisawa, R. & Kini, S. Insulin resistance and bariatric surgery. *Obes. Rev.* **13**, 316–328 (2012).
310. Aron-Wisnewsky, J., Doré, J. & Clement, K. The importance of the gut microbiota after bariatric surgery. *Nat. Rev. Gastroenterol. Hepatol.* **9**, 590–598 (2012).
311. Zhang, H. *et al.* Human gut microbiota in obesity and after gastric bypass. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 2365–2370 (2009).
312. Tremaroli, V. *et al.* Roux-en-Y Gastric Bypass and Vertical Banded Gastroplasty Induce Long-Term Changes on the Human Gut Microbiome Contributing to Fat Mass Regulation. *Cell Metab.* **22**, 228–238 (2015).
313. Ilhan, Z. E. *et al.* Distinctive microbiomes and metabolites linked with weight loss after gastric bypass, but not gastric banding. *ISME J.* **11**, 2047–2058 (2017).
314. Graessler, J. *et al.* Metagenomic sequencing of the human gut microbiome before and after bariatric surgery in obese patients with type 2 diabetes: correlation with inflammatory and metabolic parameters. *The Pharmacogenomics Journal* vol. 13 514–522 (2013).
315. Palleja, A. *et al.* Roux-en-Y gastric bypass surgery of morbidly obese patients induces swift and persistent changes of the individual gut microbiota. *Genome Medicine* vol. 8 (2016).
316. Kong, L.-C. *et al.* Gut microbiota after gastric bypass in human obesity: increased richness and associations of bacterial genera with adipose tissue genes. *Am. J. Clin. Nutr.* **98**, 16–24 (2013).
317. Guo, Y. *et al.* Modulation of the gut microbiome: a systematic review of the effect of bariatric surgery. *Eur. J. Endocrinol.* **178**, 43–56 (2018).
318. Li, J. V., Ashrafian, H., Bueter, M., Kinross, J. & Sands, C. Metabolic surgery profoundly influences gut microbial–host metabolic cross-talk. *Gut* (2011).

319. Liou, A. P. *et al.* Conserved shifts in the gut microbiota due to gastric bypass reduce host weight and adiposity. *Sci. Transl. Med.* **5**, 178ra41 (2013).
320. Kootte, R. S. *et al.* Improvement of Insulin Sensitivity after Lean Donor Feces in Metabolic Syndrome Is Driven by Baseline Intestinal Microbiota Composition. *Cell Metab.* **26**, 611–619.e6 (2017).
321. Kwong, W., Tomlinson, G. & Feig, D. S. Maternal and neonatal outcomes after bariatric surgery; a systematic review and meta-analysis: do the benefits outweigh the risks? *Am. J. Obstet. Gynecol.* **218**, 573–580 (2018).
322. Johansson, K. *et al.* Outcomes of pregnancy after bariatric surgery. *N. Engl. J. Med.* **372**, 814–824 (2015).
323. Shannon, C. E. A Mathematical Theory of Communication. *Bell System Technical Journal* vol. 27 623–656 (1948).
324. Bray, J. R. & Curtis, J. T. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecol. Monogr.* **27**, 325–349 (1957).
325. Anderson, M. J. A new method for non-parametric multivariate analysis of variance: NON-PARAMETRIC MANOVA FOR ECOLOGY. *Austral Ecol.* **26**, 32–46 (2001).
326. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
327. Jans, G. *et al.* Maternal micronutrient deficiencies and related adverse neonatal outcomes after bariatric surgery: a systematic review. *Adv. Nutr.* **6**, 420–429 (2015).
328. Smith, E. A. & Macfarlane, G. T. Dissimilatory amino Acid metabolism in human colonic bacteria. *Anaerobe* **3**, 327–337 (1997).
329. Ponsky, T. A., Brody, F. & Pucci, E. Alterations in gastrointestinal physiology after Roux-en-Y gastric bypass. *J. Am. Coll. Surg.* **201**, 125–131 (2005).
330. Russell, W. R. *et al.* High-protein, reduced-carbohydrate weight-loss diets promote metabolite profiles likely to be detrimental to colonic health. *Am. J. Clin. Nutr.* **93**, 1062–1072 (2011).
331. Windey, K., De Preter, V. & Verbeke, K. Relevance of protein fermentation to gut health. *Mol. Nutr. Food Res.* **56**, 184–196 (2012).
332. Barker, H. A. Fermentations of nitrogenous organic compounds. in *Metabolism* 151–207 (Elsevier, 1961).
333. Caboni, P. *et al.* Urinary metabolomics of pregnant women at term: a combined GC/MS and NMR approach. *J. Matern. Fetal. Neonatal Med.* **27 Suppl 2**, 4–12 (2014).
334. Nagalingam, N. A. & Lynch, S. V. Role of the microbiota in inflammatory bowel diseases. *Inflamm. Bowel Dis.* **18**, 968–984 (2012).
335. Liu, Z., Cao, A. T. & Cong, Y. Microbiota regulation of inflammatory bowel disease and colorectal cancer. *Seminars in Cancer Biology* vol. 23 543–552 (2013).
336. Zeng, M. Y., Inohara, N. & Nuñez, G. Mechanisms of inflammation-driven bacterial dysbiosis in the gut. *Mucosal Immunol.* **10**, 18–26 (2017).
337. Kittana, H. *et al.* Commensal Escherichia coli Strains Can Promote Intestinal Inflammation via Differential Interleukin-6 Production. *Front. Immunol.* **9**, 2318 (2018).
338. Keijser, B. J. F. *et al.* Pyrosequencing analysis of the oral microflora of healthy adults. *J. Dent. Res.* **87**, 1016–1020 (2008).

339. Bik, E. M. *et al.* Molecular analysis of the bacterial microbiota in the human stomach. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 732–737 (2006).
340. van den Bogert, B. *et al.* Diversity of human small intestinal Streptococcus and Veillonella populations. *FEMS Microbiol. Ecol.* **85**, 376–388 (2013).
341. Krzyściak, W., Pluskwa, K. K., Jurczak, A. & Kościelniak, D. The pathogenicity of the Streptococcus genus. *Eur. J. Clin. Microbiol. Infect. Dis.* **32**, 1361–1376 (2013).
342. Murray, B. E. The life and times of the Enterococcus. *Clin. Microbiol. Rev.* (1990).
343. Ramanan, P., Barreto, J. N., Osmon, D. R. & Tosh, P. K. Rothia bacteremia: a 10-year experience at Mayo Clinic, Rochester, Minnesota. *J. Clin. Microbiol.* **52**, 3184–3189 (2014).
344. Tao, W. *et al.* Colon and rectal cancer risk after bariatric surgery in a multicountry Nordic cohort study. *Int. J. Cancer* (2019) doi:10.1002/ijc.32770.
345. Gryp, T., Vanholder, R., Vaneechoutte, M. & Glorieux, G. p-Cresyl Sulfate. *Toxins* **9**, (2017).
346. Lee, J.-H. & Lee, J. Indole as an intercellular signal in microbial communities. *FEMS Microbiol. Rev.* **34**, 426–444 (2010).
347. Collino, S. *et al.* Metabolic signatures of extreme longevity in northern Italian centenarians reveal a complex remodeling of lipids, amino acids, and gut microbiota metabolism. *PLoS One* **8**, e56564 (2013).
348. Macfarlane, G. T., Cummings, J. H. & Allison, C. Protein degradation by human intestinal bacteria. *J. Gen. Microbiol.* **132**, 1647–1656 (1986).
349. Kaur, H., Das, C. & Mande, S. S. In Silico Analysis of Putrefaction Pathways in Bacteria and Its Implication in Colorectal Cancer. *Front. Microbiol.* **8**, 2166 (2017).
350. Seyfried, F. *et al.* Urinary phenotyping indicates weight loss-independent metabolic effects of Roux-en-Y gastric bypass in mice. *J. Proteome Res.* **12**, 1245–1253 (2013).
351. Hill, D. J. & Milner, R. D. Insulin as a growth factor. *Pediatr. Res.* **19**, 879–886 (1985).
352. Dalfrà, M. G. *et al.* Elevated insulin sensitivity and β-cell function during pregnancy in mothers of growth-restricted newborns. *Am. J. Physiol. Endocrinol. Metab.* **301**, E25–30 (2011).
353. Lupoli, R. *et al.* Bariatric surgery and long-term nutritional issues. *World J. Diabetes* **8**, 464–474 (2017).
354. Steenackers, N., Gesquiere, I. & Matthys, C. The relevance of dietary protein after bariatric surgery: what do we know? *Curr. Opin. Clin. Nutr. Metab. Care* **21**, 58–63 (2018).
355. Kanner, L. Autistic disturbances of affective contact. *Nerv. Child* **2**, 217–250 (1943).
356. Lord, C. *et al.* A multisite study of the clinical diagnosis of different autism spectrum disorders. *Arch. Gen. Psychiatry* **69**, 306–313 (2012).
357. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (5th ed.)*. (2013).
358. Maenner, M. J. *et al.* Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2016. *MMWR Surveill. Summ.* **69**, 1–12 (2020).
359. Elsabbagh, M. *et al.* Global prevalence of autism and other pervasive developmental disorders. *Autism Res.* **5**, 160–179 (2012).

360. Matelski, L. & Van de Water, J. Risk factors in autism: Thinking outside the brain. *J. Autoimmun.* **67**, 1–7 (2016).
361. Hallmayer, J. *et al.* Genetic heritability and shared environmental factors among twin pairs with autism. *Arch. Gen. Psychiatry* **68**, 1095–1102 (2011).
362. Mannion, A. & Leader, G. Comorbidity in autism spectrum disorder: A literature review. *Res. Autism Spectr. Disord.* **7**, 1595–1616 (2013).
363. Holmes, E., Li, J. V., Marchesi, J. R. & Nicholson, J. K. Gut microbiota composition and activity in relation to host metabolic phenotype and disease risk. *Cell Metab.* **16**, 559–564 (2012).
364. Mulle, J. G., Sharp, W. G. & Cubells, J. F. The gut microbiome: a new frontier in autism research. *Curr. Psychiatry Rep.* **15**, 337 (2013).
365. Sands, C. J. *et al.* The nPYc-Toolbox, a Python module for the pre-processing, quality-control and analysis of metabolic profiling datasets. *Bioinformatics* **35**, 5359–5360 (2019).
366. Halekoh, U. & Højsgaard, S. A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models. *Journal of Statistical Software, Articles* **59**, 1–32 (2014).
367. McNemar, Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* **12**, 153–157 (1947).
368. Developmental Disabilities Monitoring Network Surveillance Year 2010 Principal Investigators & Centers for Disease Control and Prevention (CDC). Prevalence of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *MMWR Surveill. Summ.* **63**, 1–21 (2014).
369. Brett, D., Warnell, F., McConachie, H. & Parr, J. R. Factors Affecting Age at ASD Diagnosis in UK: No Evidence that Diagnosis Age has Decreased Between 2004 and 2014. *J. Autism Dev. Disord.* **46**, 1974–1984 (2016).
370. Simonoff, E. *et al.* Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. *J. Am. Acad. Child Adolesc. Psychiatry* **47**, 921–929 (2008).
371. Kohane, I. S. *et al.* The co-morbidity burden of children and young adults with autism spectrum disorders. *PLoS One* **7**, e33224 (2012).
372. Chen, M.-H. *et al.* Comorbidity of allergic and autoimmune diseases in patients with autism spectrum disorder: A nationwide population-based study. *Res. Autism Spectr. Disord.* **7**, 205–212 (2013).
373. Melke, J. *et al.* Abnormal melatonin synthesis in autism spectrum disorders. *Mol. Psychiatry* **13**, 90–98 (2008).
374. Stewart, P. A. *et al.* Dietary Supplementation in Children with Autism Spectrum Disorders: Common, Insufficient, and Excessive. *J. Acad. Nutr. Diet.* **115**, 1237–1248 (2015).
375. Levy, S. E. & Hyman, S. L. Complementary and alternative medicine treatments for children with autism spectrum disorders. *Child Adolesc. Psychiatr. Clin. N. Am.* **17**, 803–20, ix (2008).
376. Whitehouse, A. J. O. Complementary and alternative medicine for autism spectrum

disorders: rationale, safety and efficacy. *J. Paediatr. Child Health* **49**, E438–42:quiz E442 (2013).
377. Salem, N., Jr, Litman, B., Kim, H. Y. & Gawrisch, K. Mechanisms of action of docosahexaenoic acid in the nervous system. *Lipids* **36**, 945–959 (2001).
378. Cheng, Y.-S. *et al.* Supplementation of omega 3 fatty acids may improve hyperactivity, lethargy, and stereotypy in children with autism spectrum disorders: a meta-analysis of randomized controlled trials. *Neuropsychiatr. Dis. Treat.* **13**, 2531–2543 (2017).
379. Nadal-Desbarats, L. *et al.* Combined 1H-NMR and 1H-13C HSQC-NMR to improve urinary screening in autism spectrum disorders. *Analyst* **139**, 3460–3468 (2014).
380. Wang, M. *et al.* Alterations in Gut Glutamate Metabolism Associated with Changes in Gut Microbiota Composition in Children with Autism Spectrum Disorder. *mSystems* **4**, (2019).
381. Kang, D.-W. *et al.* Differences in fecal microbial metabolites and microbiota of children with autism spectrum disorders. *Anaerobe* **49**, 121–131 (2018).
382. Strati, F. *et al.* New evidences on the altered gut microbiota in autism spectrum disorders. *Microbiome* **5**, 24 (2017).
383. Kang, D.-W. *et al.* Reduced incidence of Prevotella and other fermenters in intestinal microflora of autistic children. *PLoS One* **8**, e68322 (2013).
384. Son, J. S. *et al.* Comparison of Fecal Microbiota in Children with Autism Spectrum Disorders and Neurotypical Siblings in the Simons Simplex Collection. *PLoS One* **10**, e0137725 (2015).
385. Zhao, Y. *et al.* Gut microbiota composition modifies fecal metabolic profiles in mice. *J. Proteome Res.* **12**, 2987–2999 (2013).
386. Yap, I. K. S. *et al.* Metabonomic and microbiological analysis of the dynamic effect of vancomycin-induced gut microbiota modification in the mouse. *J. Proteome Res.* **7**, 3718–3728 (2008).
387. Geboes, K. P. *et al.* The influence of inulin on the absorption of nitrogen and the production of metabolites of protein fermentation in the colon. *Br. J. Nutr.* **96**, 1078–1086 (2006).
388. Gibson, S. A., McFarlan, C., Hay, S. & MacFarlane, G. T. Significance of microflora in proteolysis in the colon. *Appl. Environ. Microbiol.* **55**, 679–683 (1989).
389. Wallace, R. J. & McKain, N. Peptidase activity of human colonic bacteria. *Anaerobe* **3**, 251–257 (1997).
390. Davila, A.-M. *et al.* Intestinal luminal nitrogen metabolism: role of the gut microbiota and consequences for the host. *Pharmacol. Res.* **68**, 95–107 (2013).
391. Barrett, E., Ross, R. P., O'Toole, P. W., Fitzgerald, G. F. & Stanton, C. γ-Aminobutyric acid production by culturable bacteria from the human intestine. *J. Appl. Microbiol.* **113**, 411–417 (2012).
392. Pokusaeva, K. *et al.* GABA-producing Bifidobacterium dentium modulates visceral sensitivity in the intestine. *Neurogastroenterol. Motil.* **29**, (2017).
393. Ko, C. Y., Lin, H.-T. V. & Tsai, G. J. Gamma-aminobutyric acid production in black soybean milk by Lactobacillus brevis FPA 3709 and the antidepressant effect of the fermented product on a forced swimming rat model. *Process Biochem.* **48**, 559–568 (2013).
394. Janik, R. *et al.* Magnetic resonance spectroscopy reveals oral Lactobacillus promotion of

increases in brain GABA, N-acetyl aspartate and glutamate. *Neuroimage* **125**, 988–995 (2016).
395. Bravo, J. A. *et al.* Ingestion of Lactobacillus strain regulates emotional behavior and central GABA receptor expression in a mouse via the vagus nerve. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 16050–16055 (2011).
396. Kawase, T. *et al.* Gut microbiota of mice putatively modifies amino acid metabolism in the host brain. *Br. J. Nutr.* **117**, 775–783 (2017).
397. Matsumoto, M. *et al.* Cerebral low-molecular metabolites influenced by intestinal microbiota: a pilot study. *Front. Syst. Neurosci.* **7**, 9 (2013).
398. Yano, J. M. *et al.* Indigenous bacteria from the gut microbiota regulate host serotonin biosynthesis. *Cell* **161**, 264–276 (2015).
399. Hall, J. A., Brockman, J. A. & Jewell, D. E. Dietary fish oil alters the lysophospholipid metabolomic profile and decreases urinary 11-dehydro thromboxane B_2 concentration in healthy Beagles. *Vet. Immunol. Immunopathol.* **144**, 355–365 (2011).
400. Zheng, J.-S. *et al.* Serum metabolomics profiles in response to n-3 fatty acids in Chinese patients with type 2 diabetes: a double-blind randomised controlled trial. *Sci. Rep.* **6**, 29522 (2016).
401. Kawasaki, M., Miura, Y., Funabiki, R. & Yagasaki, K. Effects of simultaneous dietary fish oil ingestion and sulfur amino acid supplementation on the lipid metabolism in hepatoma-bearing rats with hyperlipidemia. *J. Nutr. Sci. Vitaminol.* **56**, 247–254 (2010).
402. Zierer, J. *et al.* The fecal metabolome as a functional readout of the gut microbiome. *Nat. Genet.* **50**, 790–795 (2018).
403. van der Wielen, N., Moughan, P. J. & Mensink, M. Amino Acid Absorption in the Large Intestine of Humans and Porcine Models. *J. Nutr.* **147**, 1493–1498 (2017).
404. Bosch, S. *et al.* Fecal Amino Acid Analysis Can Discriminate De Novo Treatment-Naïve Pediatric Inflammatory Bowel Disease From Controls. *J. Pediatr. Gastroenterol. Nutr.* **66**, 773–778 (2018).
405. Ni, J. *et al.* A role for bacterial urease in gut dysbiosis and Crohn's disease. *Sci. Transl. Med.* **9**, (2017).
406. Weir, T. L. *et al.* Stool microbiome and metabolome differences between colorectal cancer patients and healthy adults. *PLoS One* **8**, e70803 (2013).
407. Monleón, D. *et al.* Metabolite profiling of fecal water extracts from human colorectal cancer. *NMR Biomed.* **22**, 342–348 (2009).
408. Mayer, E. A., Padua, D. & Tillisch, K. Altered brain-gut axis in autism: comorbidity or causative mechanisms? *Bioessays* **36**, 933–939 (2014).
409. Mazefsky, C. A., Schreiber, D. R., Olino, T. M. & Minshew, N. J. The association between emotional and behavioral problems and gastrointestinal symptoms among children with high-functioning autism. *Autism* **18**, 493–501 (2014).
410. Lee, K.-A. *et al.* Bacterial-derived uracil as a modulator of mucosal immunity and gut-microbe homeostasis in Drosophila. *Cell* **153**, 797–811 (2013).
411. Seakins, J. W. The determination of urinary phenylacetylglutamine as phenylacetic acid. Studies on its origin in normal subjects and children with cystic fibrosis. *Clin. Chim. Acta* **35**, 121–131 (1971).

412. Roager, H. M. *et al.* Colonic transit time is related to bacterial metabolism and mucosal turnover in the gut. *Nat Microbiol* **1**, 16093 (2016).
413. Jeffery, I. B. *et al.* An irritable bowel syndrome subtype defined by species-specific alterations in faecal microbiota. *Gut* **61**, 997–1006 (2012).
414. Beyoğlu, D., Smith, R. L. & Idle, J. R. Dog bites man or man bites dog? The enigma of the amino acid conjugations. *Biochem. Pharmacol.* **83**, 1331–1339 (2012).
415. Gabriele, S. *et al.* Urinary p-cresol is elevated in young French children with autism spectrum disorder: a replication study. *Biomarkers* **19**, 463–470 (2014).
416. Posada-Ayala, M. *et al.* Identification of a urine metabolomic signature in patients with advanced-stage chronic kidney disease. *Kidney Int.* **85**, 103–111 (2014).
417. Meijers, B. K. I. *et al.* p-Cresol and cardiovascular risk in mild-to-moderate kidney disease. *Clin. J. Am. Soc. Nephrol.* **5**, 1182–1189 (2010).
418. Bandini, L. G. *et al.* Food selectivity in children with autism spectrum disorders and typically developing children. *J. Pediatr.* **157**, 259–264 (2010).
419. Kral, T. V. E., Eriksen, W. T., Souders, M. C. & Pinto-Martin, J. A. Eating behaviors, diet quality, and gastrointestinal symptoms in children with autism spectrum disorders: a brief review. *J. Pediatr. Nurs.* **28**, 548–556 (2013).
420. Filippis, F. D. *et al.* High-level adherence to a Mediterranean diet beneficially impacts the gut microbiota and associated metabolome. *Gut* vol. 65 1812–1821 (2016).
421. Garcia-Perez, I. *et al.* Objective assessment of dietary patterns by use of metabolic phenotyping: a randomised, controlled, crossover trial. *Lancet Diabetes Endocrinol* **5**, 184–195 (2017).
422. Vich Vila, A. *et al.* Impact of commonly used drugs on the composition and metabolic function of the gut microbiota. *Nat. Commun.* **11**, 362 (2020).
423. Duncan, S. H. *et al.* Reduced dietary intake of carbohydrates by obese subjects results in decreased concentrations of butyrate and butyrate-producing bacteria in feces. *Appl. Environ. Microbiol.* **73**, 1073–1078 (2007).
424. Mead, G. C. The amino acid-fermenting clostridia. *J. Gen. Microbiol.* **67**, 47–56 (1971).
425. Chong, P. P. *et al.* The Microbiome and Irritable Bowel Syndrome - A Review on the Pathophysiology, Current Research and Future Therapy. *Front. Microbiol.* **10**, 1136 (2019).
426. Tap, J. *et al.* Identification of an Intestinal Microbiota Signature Associated With Severity of Irritable Bowel Syndrome. *Gastroenterology* **152**, 111–123.e8 (2017).
427. Baj, A. *et al.* Glutamatergic Signaling Along The Microbiota-Gut-Brain Axis. *Int. J. Mol. Sci.* **20**, (2019).
428. Campbell, L. L., Jr. Reductive degradation of pyrimidines. III. Purification and properties of dihydrouracil dehydrogenase. *J. Biol. Chem.* **227**, 693–700 (1957).
429. Patel, B. N. & West, T. P. Degradation of the pyrimidine bases uracil and thymine by Escherichia coli B. *Microbios* **49**, 107–113 (1987).
430. Finnicum, C. T. *et al.* Cohabitation is associated with a greater resemblance in gut microbiota which can impact cardiometabolic and inflammatory risk. *BMC Microbiol.* **19**, 230 (2019).
431. Mavel, S. *et al.* 1H-13C NMR-based urine metabolic profiling in autism spectrum disorders. *Talanta* **114**, 95–102 (2013).

432. Duncan, S. H., Louis, P. & Flint, H. J. Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product. *Appl. Environ. Microbiol.* **70**, 5810–5817 (2004).
433. Allen-Vercoe, E. *et al.* Anaerostipes hadrus comb. nov., a dominant species within the human colonic microbiota; reclassification of Eubacterium hadrum Moore et al. 1976. *Anaerobe* **18**, 523–529 (2012).
434. Magnúsdóttir, S. *et al.* Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat. Biotechnol.* **35**, 81–89 (2017).
435. Saulnier, D. M. *et al.* Gastrointestinal microbiome signatures of pediatric patients with irritable bowel syndrome. *Gastroenterology* **141**, 1782–1791 (2011).
436. Mégraud, F., Bébéar, C., Dabernat, H. & Delmas, C. Haemophilus species in the human gastrointestinal tract. *Eur. J. Clin. Microbiol. Infect. Dis.* **7**, 437–438 (1988).
437. Braune, A. & Blaut, M. Bacterial species involved in the conversion of dietary flavonoids in the human gut. *Gut Microbes* **7**, 216–234 (2016).
438. Chiou, Y.-S. *et al.* Metabolic and colonic microbiota transformation may enhance the bioactivities of dietary polyphenols. *J. Funct. Foods* **7**, 3–25 (2014).
439. Matthies, C., Evers, S., Ludwig, W. & Schink, B. Anaerovorax odorimutans gen. nov., sp. nov., a putrescine-fermenting, strictly anaerobic bacterium. *Int. J. Syst. Evol. Microbiol.* **50 Pt 4**, 1591–1594 (2000).
440. Hoskins, L. C. *et al.* Mucin degradation in human colon ecosystems. Isolation and properties of fecal strains that degrade ABH blood group antigens and oligosaccharides from mucin glycoproteins. *J. Clin. Invest.* **75**, 944–953 (1985).
441. Flynn, J. M., Niccum, D., Dunitz, J. M. & Hunter, R. C. Evidence and Role for Bacterial Mucin Degradation in Cystic Fibrosis Airway Disease. *PLoS Pathog.* **12**, e1005846 (2016).
442. Ahn, J. *et al.* Human gut microbiome and risk for colorectal cancer. *J. Natl. Cancer Inst.* **105**, 1907–1911 (2013).
443. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).
444. Liu, S. *et al.* Altered gut microbiota and short chain fatty acids in Chinese children with autism spectrum disorder. *Sci. Rep.* **9**, 287 (2019).
445. Brahe, L. K. *et al.* Specific gut microbiota features and metabolic markers in postmenopausal women with obesity. *Nutr. Diabetes* **5**, e159 (2015).
446. Lun, H. *et al.* Altered gut microbiota and microbial biomarkers associated with chronic kidney disease. *Microbiologyopen* **8**, e00678 (2019).
447. Gargari, G., Taverniti, V. & Gardana, C. Fecal Clostridiales distribution and short‐chain fatty acids reflect bowel habits in irritable bowel syndrome. *Environmentalist* (2018).
448. Schulpis, K. H. *et al.* Maternal-neonatal amino acid blood levels in relation to the mode of delivery. *Acta Obstet. Gynecol. Scand.* **88**, 71–76 (2009).
449. Newlan, L. & Geraghty, S. Efficacy of oral glucose tolerance testing of pregnant women post bariatric surgery. *British Journal of Midwifery* **26**, 704–707 (2018).
450. Lewis, S. J. & Heaton, K. W. Stool form scale as a useful guide to intestinal transit time. *Scand. J. Gastroenterol.* **32**, 920–924 (1997).
451. Whiley, L. *et al.* Ultrahigh-Performance Liquid Chromatography Tandem Mass

Spectrometry with Electrospray Ionization Quantification of Tryptophan Metabolites and Markers of Gut Health in Serum and Plasma-Application to Clinical and Epidemiology Cohorts. *Anal. Chem.* **91**, 5207–5216 (2019).

452. Holmes, E. *et al.* Detection of urinary drug metabolite (xenometabolome) signatures in molecular epidemiology studies via statistical total correlation (NMR) spectroscopy. *Anal. Chem.* **79**, 2629–2640 (2007).

Appendix

Appendix A - Metabolite Identification

Table 1: Metabolite identification for Chapter 3

Serum metabolite	Chemical shifts identified	Confirmed by
1-methylhistidine	7.06 (s), 7.79 (s)	J-res, STOCSY/STORM, HSQC
Tyrosine	3.06 (dd), 3.20 (dd), 3.94 (dd), 6.9 (d), 7.19 (d)	J-res, STOCSY/STORM, HSQC, TOCSY
D-β-hydroxybutyrate	1.22 (d), 2.37 (m), 4.17 (m)	J-res, STOCSY/STORM, HSQC
Choline	3.21 (s)	J-res, HSQC
Unknown-1	3.1 (q), 3.29 (t)	J-res, STOCSY/STORM, HSQC, TOCSY
Creatine	3.04 (s), 3.94 (s)	J-res, STOCSY/STORM, TOCSY, HSQC
Glutamine	2.13 (m), 2.45 (m), 3.76 (t)	J-res, STOCSY/STORM, TOCSY, HSQC
N-acetyl glycoproteins	2.04 (s)	STOCSY/STORM, J-res, HSQC
Acetate	1.91 (s)	J-res, STOCSY/STORM
Alanine	1.47 (d), 3.78 (q)	J-res, STOCSY/STORM, TOCSY, HSQC
Unknown-2	1.06 (d)	J-res
Valine	0.98 (d), 1.03 (d), 2.28 (m), 3.62 (d)	J-res, STOCSY/STORM, TOCSY, HSQC
Isoleucine	0.93 (t), 1.00 (d)	J-res, STOCSY/STORM, HSQC

Urine metabolite	Chemical shifts identified	Confirmed by
Hippurate	3.97 (d), 7.56 (t), 7.64 (t), 7.84 (d)	J-res, STOCSY/STORM, HSQC
Unknown-3	4.3 (s), 4.69 (dd), 7.68 (s)	J-res, STOCSY/STORM, HSQC, COSY, TOCSY
Phenylacetylglutamine	1.92 (m), 2.11 (m), 2.27 (m), 3.67 (m), 4.19 (m), 7.36 (t), 7.43 (t)	J-res, STOCSY/STORM, HSQC
Creatinine	3.05 (s), 4.06 (s)	J-res, STOCSY/STORM, COSY
Unknown-4	2.78 (s)	J-res, STOCSY/STORM, HSQC
Dimethylamine	2.72 (s)	J-res, STOCSY/STORM, HSQC

Oxaloacetate	2.38 (s)	J-res, STOCSY/STORM
N-acetyl glycoproteins	2.04 (s)	J-res, HSQC
Alanine	1.48 (d), 3.79 (q)	J-res, STOCSY/STORM, HSQC, COSY, TOCSY
2-hydroxyisobutyrate	1.36 (s)	J-res, STOCSY/STORM, HSQC
3-hydroxyisovalerate	1.27 (s), 2.37 (s)	J-res, STOCSY/STORM, HSQC, TOCSY
Methylmalonate	1.25 (d), 3.17 (q)	J-res, STOCSY/STORM, HSQC, COSY
4-deoxythreonic acid (4-DTA)	1.24 (d), 4.06 (m)	J-res, STOCSY/STORM, HSQC, COSY, TOCSY
4-deoxyerythronic acid (4-DEA)	1.11 (d), 4.08 (d), 4.12 (m)	J-res, STOCSY/STORM, HSQC, COSY, TOCSY
Valine	0.99 (d), 1.05 (d), 2.27 (m), 3.62 (d)	J-res, STOCSY/STORM, HSQC, COSY, TOCSY

**s=singlet, d=doublet, t=triplet, m=multiplet*
STOCSY Statistical Total Correlation Spectroscopy; STORM Subset Optimisation by Reference Matching; HSQC Hetero-nuclear Single Quantum Coherence (2D NMR); TOCSY Total Correlation Spectroscopy; COSY Correlation Spectroscopy

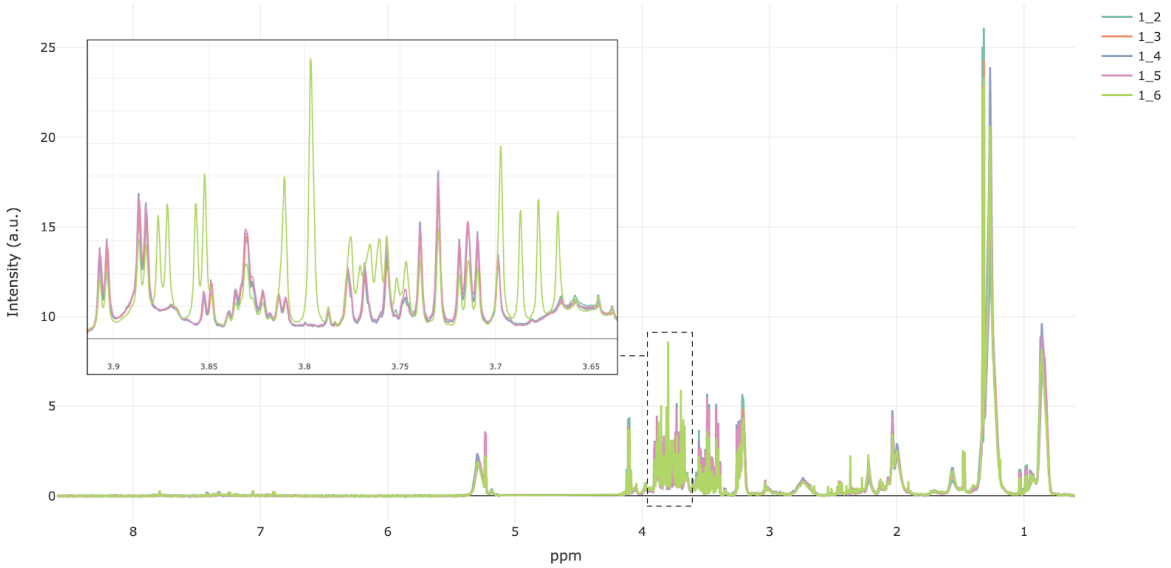


Figure 1: Mannitol in serum. One example of a patient with high concentrations of mannitol in the serum. The serum sample from patient 1 at time point 6 was excluded from feature selection due to mannitol. The other samples from patient 1 are plotted for comparison. *a.u. arbitrary units; ppm parts per million.*

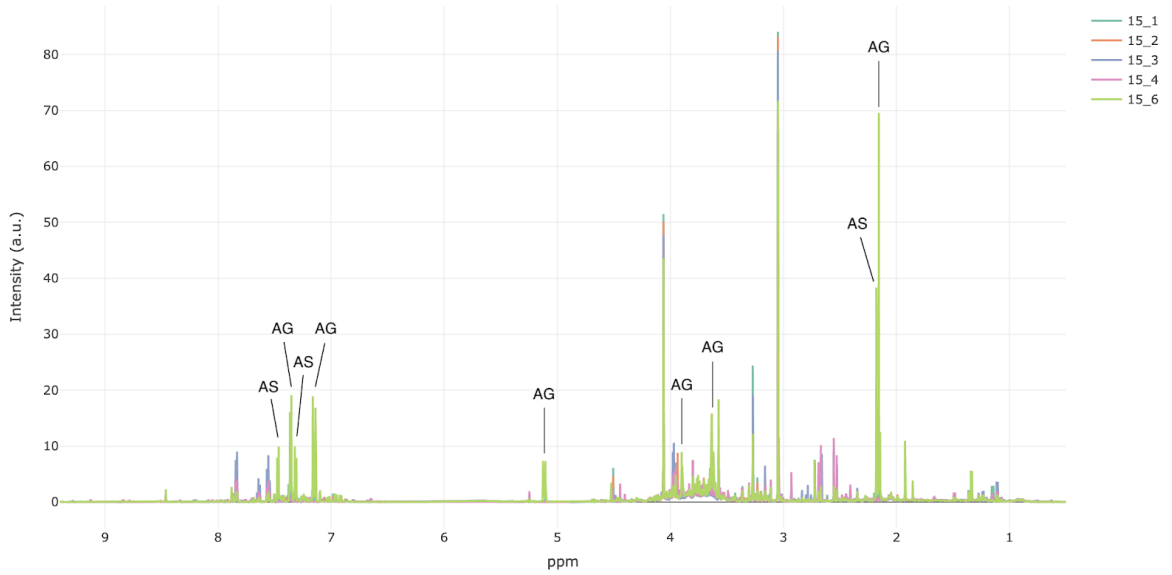


Figure 2: Paracetamol metabolites in urine. One example of a patient with high concentrations of paracetamol, also called acetaminophen, metabolites in the urine. The urine sample from patient 15 at time point 6 was excluded from feature selection due to the presence of paracetamol metabolites[452]. The other samples from patient 15 are plotted for comparison. *a.u. arbitrary units; ppm parts per million; AS acetaminophen sulfate; AG acetaminophen glucuronide.*

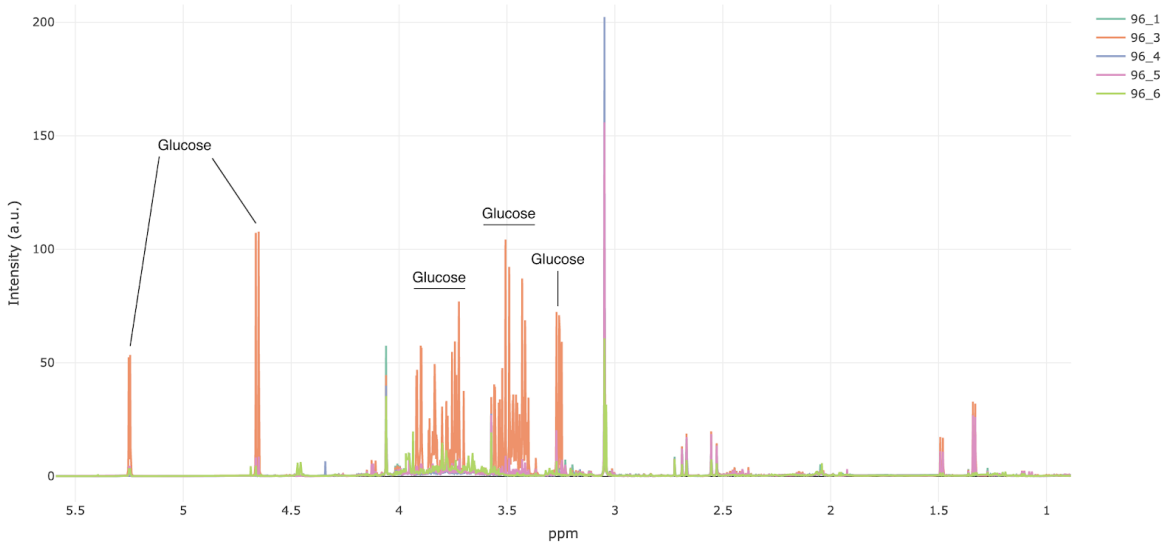


Figure 3: Glucose in urine. One example of a patient with high concentration of glucose in the urine. The urine sample from patient 96 at time point 3 was excluded from feature selection due to glucose. The other samples from patient 96 are plotted for comparison. *a.u. arbitrary units; ppm parts per million.*

Table 2: Metabolite identification for Chapter 4

Serum metabolite	Chemical shifts identified	Confirmed by
Leucine	0.96 (2xd), 1.71 (m), 3.73 (t)	STOCSY/STORM, J-res
Isoleucine	0.94 (t), 1.01 (d)	STOCSY/STORM, J-res
Isobutyrate	1.09 (d)	STOCSY/STORM, J-res, HSQC
N–acetyl glycoproteins	2.05 (s)	STOCSY/STORM, J-res, HSQC
Glutamine	2.14 (m), 2.46 (m), 3.77 (t)	STOCSY/STORM, J-res, HSQC
D-β-hydroxybutyrate	1.22 (d), 2.37 (m), 4.17 (m)	STOCSY/STORM, J-res

Urine metabolite	Chemical shifts identified	Confirmed by
Valine	0.99 (d), 1.04 (d)	STOCSY/STORM, J-res
α-ketoisovalerate	1.11 (d)	STOCSY/STORM, J-res
Methylmalonate	1.24 (d)	STOCSY/STORM, J-res
Creatinine	3.05 (s), 4.06 (s)	STOCSY/STORM, J-res
p-cresyl sulfate	2.35 (s), 7.21 (d), 7.28 (d)	STOCSY/STORM, J-res, HSQC
p-hydroxyphenylacetate	3.45 (s), 6.86 (d), 7.17 (d)	STOCSY/STORM, J-res, HSQC, spike in
Phenylacetylglutamine	1.92 (m), 2.11 (m), 2.27 (m), 3.67 (m), 4.19 (m), 7.36 (t), 7.43 (t)	STOCSY/STORM, J-res, HSQC
Unknown	7.34 (d), 7.48 (t)	STOCSY/STORM, J-res, HSQC
Indoxyl sulfate	7.5 (d), 7.7 (d)	STOCSY/STORM, J-res, HSQC, spike-in

**s=singlet, d=doublet, t=triplet, m=multiplet*
STOCSY Statistical Total Correlation Spectroscopy; STORM Subset Optimisation by Reference Matching; HSQC Hetero-nuclear Single Quantum Coherence (2D NMR); TOCSY Total Correlation Spectroscopy; COSY Correlation Spectroscopy

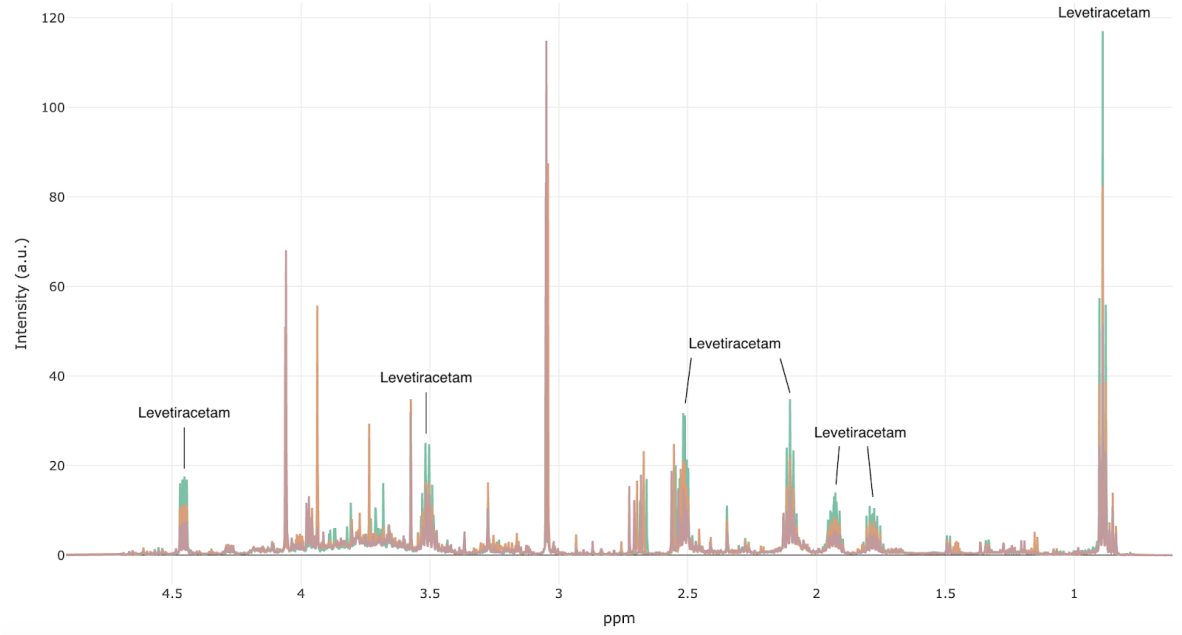


Figure 4: Levetiracetam in urine. Children taking the epilepsy medication Levetiracetam excrete high concentrations of the drug in their urine. The affected samples were excluded from statistical analyses. *a.u. arbitrary units; ppm parts per million.*

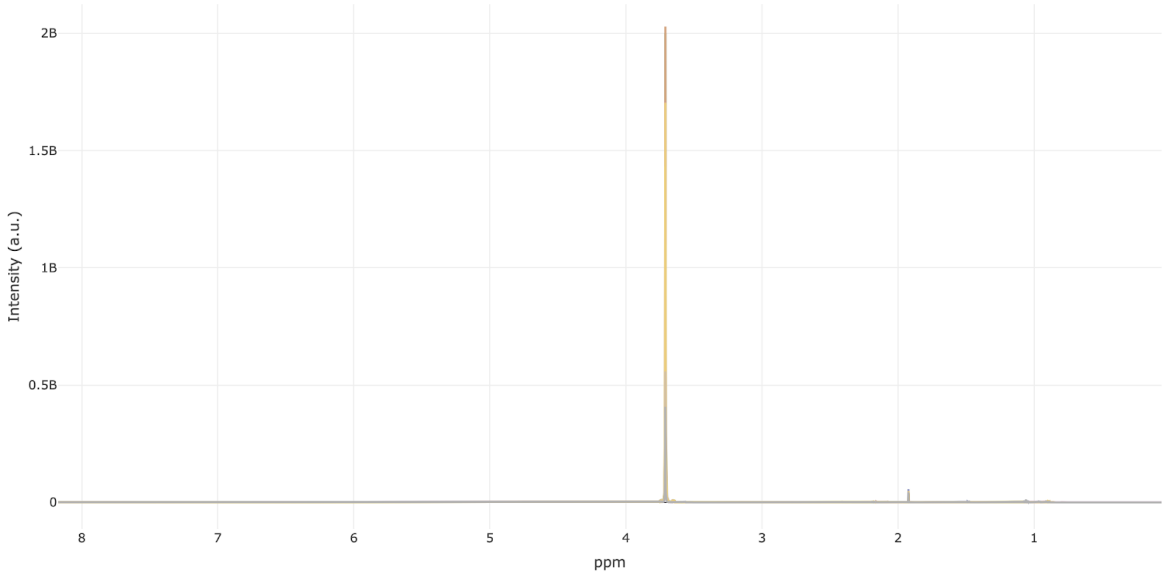


Figure 5: Polyethylene glycol in faeces. Some faecal samples had high concentrations of polyethylene glycol. This signal dominates the spectrum, therefore the affected samples must be excluded from statistical analyses. *a.u. arbitrary units; ppm parts per million.*

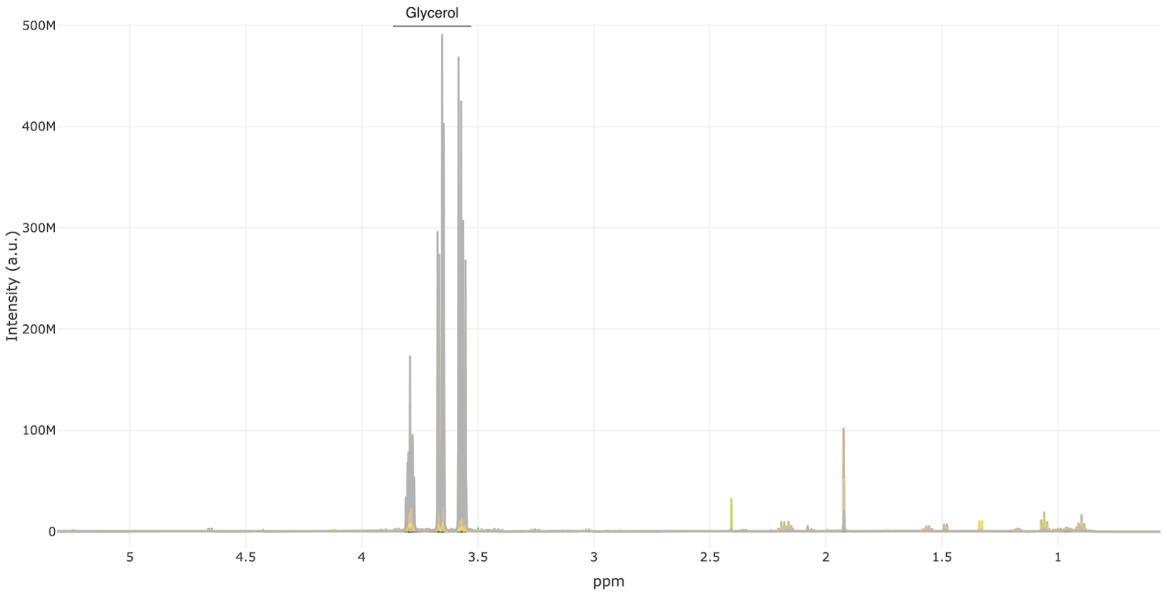


Figure 6: Glycerol in faeces. Some faecal samples had high concentrations of glycerol. These signals dominate the spectrum, therefore the affected samples must be excluded from statistical analyses. *a.u. arbitrary units; ppm parts per million.*

Table 3: Metabolite identification for Chapter 5

Faecal metabolite	Chemical shifts identified	Confirmed by
Tryptophan	7.21 (t), 7.29 (t), 7.33 (s), 7.55 (d), 7.74 (d)	STOCSY/STORM, J-res
Uracil	5.81 (d), 7.54 (d)	STOCSY/STORM, J-res
Phenylalanine	3.13 (dd), 3.28 (dd), 3.37 (s), 4.01 (dd), 7.33 (m), 7.38 (m), 7.43 (m)	STOCSY/STORM, J-res
Tyrosine	3.06 (dd), 3.2 (dd), 3.94 (dd), 6.91 (d), 7.2 (d)	STOCSY/STORM, J-res
Glycine	3.57 (s)	STOCSY/STORM, J-res
Lysine	1.48 (m), 1.72 (m), 1.89 (m), 3.03 (t), 3.77 (t)	STOCSY/STORM, J-res
Aspartate	2.69 (dd), 2.82 (dd), 3.91 (dd)	STOCSY/STORM, J-res
Methionine	2.14 (s), 2.65 (t), 3.87 (dd)	STOCSY/STORM, J-res
Alanine	1.47 (d), 3.78 (q)	STOCSY/STORM, J-res
Glutamate	2.04 (m), 2.12 (m), 2.36 (m), 3.74 (dd)	STOCSY/STORM, J-res
Valine	0.99 (d), 1.05 (d), 2.28 (m), 3.62 (d)	STOCSY/STORM, J-res
Leucine	0.96 (2xd), 1.72 (m), 3.73 (t)	STOCSY/STORM, J-res
Isoleucine	0.94 (t), 1.01 (d)	STOCSY/STORM, J-res

Urine metabolite	Chemical shifts identified	Confirmed by
p-cresyl sulfate	2.35 (s), 7.21 (d), 7.28 (d)	STOCSY/STORM, J-res
Phenylacetylglutamine	1.92 (m), 2.11 (m), 2.27 (m), 3.67 (m), 4.19 (m), 7.36 (t), 7.43 (t)	STOCSY/STORM, J-res

**s=singlet, d=doublet, t=triplet, m=multiplet*
STOCSY Statistical Total Correlation Spectroscopy; STORM Subset Optimisation by Reference Matching

Appendix B - Additional data

Chapter 3

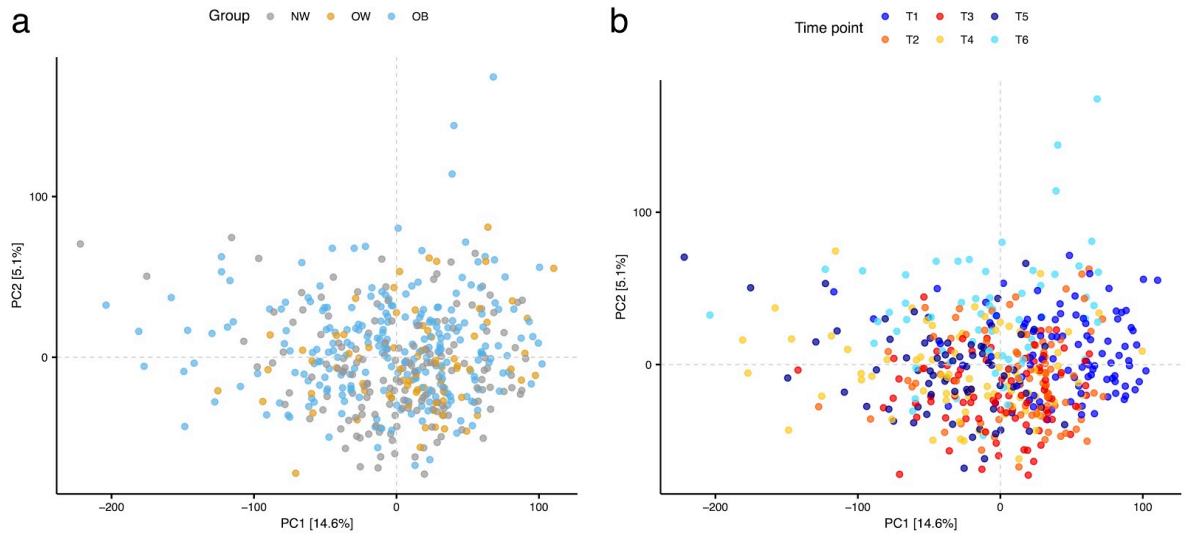


Figure 3.S1: Principal component analysis of all maternal serum samples. Points are coloured by **a)** study group and **b)** time point. *PC principal component; NW normal weight; OW overweight; OB obese; T1 11-14 weeks gestation; T2 20-24 weeks gestation; T3 28-30 weeks gestation; T4 30-33 weeks gestation; T5 35-37 weeks gestation; T6 delivery.*

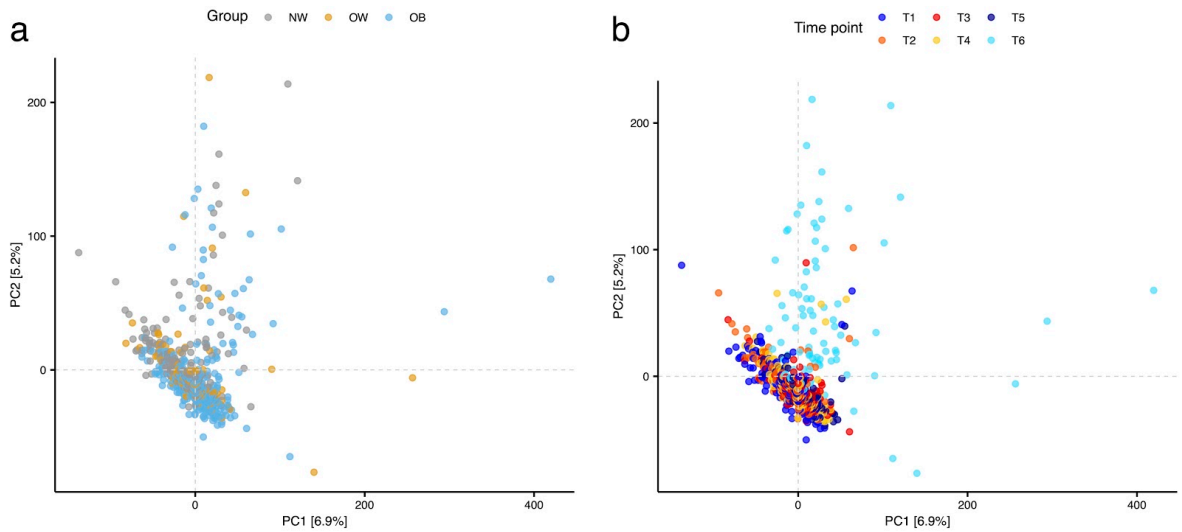


Figure 3.S2: Principal component analysis of all maternal urine samples. Points are coloured by a) study group and b) time point. Many urine samples taken at the delivery time point (T6) contained drug metabolites. *PC principal component; NW normal weight; OW overweight; OB obese; T1 11-14 weeks gestation; T2 20-24 weeks gestation; T3 28-30 weeks gestation; T4 30-33 weeks gestation; T5 35-37 weeks gestation; T6 delivery.*

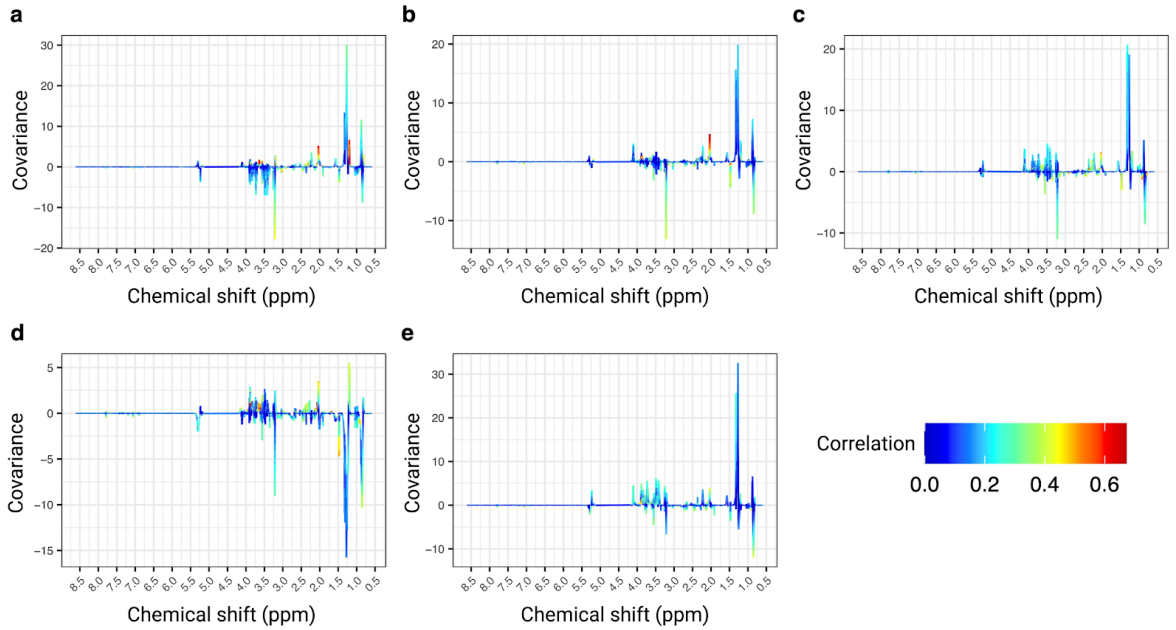


Figure 3.S3: Loadings from orthogonal partial least squares regression models for serum samples. For each time point, **a)** T1 (11-14 weeks gestation), **b)** T2 (20-24 weeks gestation), **c)** T3 (28-30 weeks gestation), **d)** T4 (30-33 weeks gestation) and **e)** T5 (35-37 weeks gestation), ^{1}H NMR serum data was regressed against maternal BMI to select spectral features associated with BMI during pregnancy.

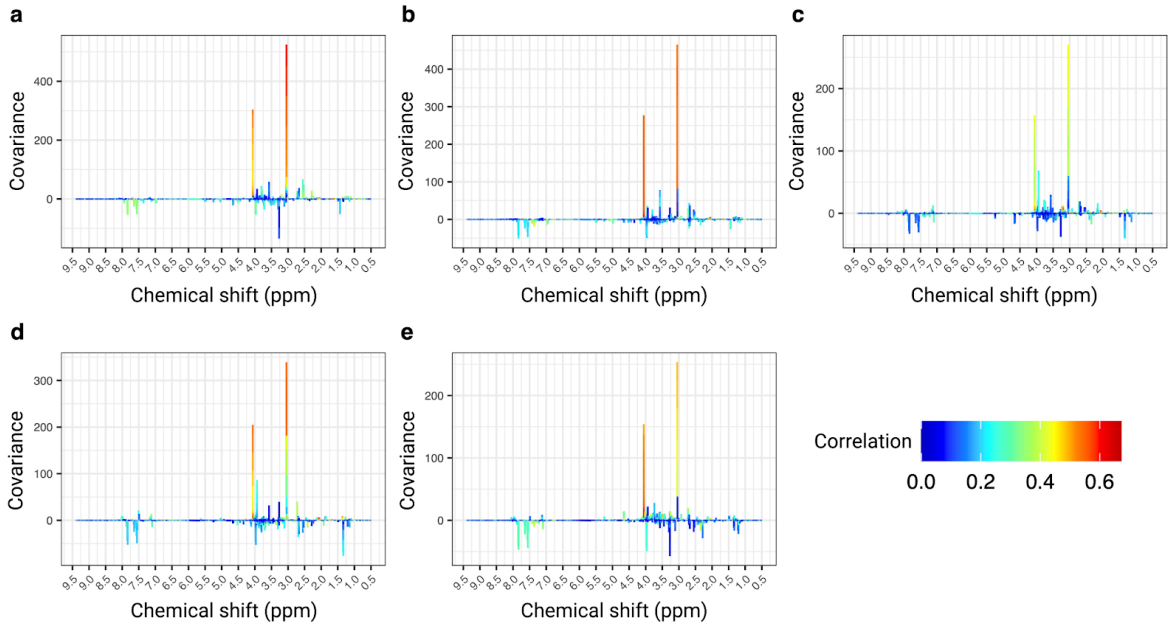


Figure 3.S4: Loadings from orthogonal partial least squares regression models for urine samples. For each time point, a) T1 (11-14 weeks gestation), b) T2 (20-24 weeks gestation), c) T3 (28-30 weeks gestation), d) T4 (30-33 weeks gestation) and e) T5 (35-37 weeks gestation), ^{1}H NMR urine data was regressed against maternal BMI to select spectral features associated with BMI during pregnancy.

Table 3.S1 Integrative analysis of maternal metabolites and clinical data. Multi-block (each block is a time point) partial least squares regression models were calculated for serum and urine separately. Correlations between metabolites (*X*) and clinical parameters (*Y*) for each time point were approximated based on model loadings. Significant (*FDR* <0.05, permutation test) correlations were summarised for each metabolite-clinical parameter pair.

Metabolite	Biofluid	Clinical parameter	Number of time points*	Direction of association
1-methylhistidine	Serum	Maternal HOMA-IR	4	Negative
1-methylhistidine	Serum	Maternal insulin	4	Negative
Acetate	Serum	Maternal HbA1c	4	Negative
Alanine	Serum	Maternal HOMA-IR	4	Negative
Alanine	Serum	Maternal insulin	5	Negative
Glutamine	Serum	Maternal HOMA-IR	5	Negative
Glutamine	Serum	Maternal insulin	5	Negative
GlycA	Serum	Maternal HbA1c	4	Positive

GlycA	Serum	Maternal glucose	2	Positive
GlycA	Serum	Maternal HOMA-IR	5	Positive
GlycA	Serum	Maternal insulin	5	Positive
Tyrosine	Serum	Maternal HOMA-IR	3	Negative
Tyrosine	Serum	Maternal insulin	3	Negative
2-hydroxyisobutyrate	Urine	Maternal glucose	5	Positive
2-hydroxyisobutyrate	Urine	Maternal HOMA-IR	5	Positive
2-hydroxyisobutyrate	Urine	Maternal insulin	5	Positive
3-hydroxyisovalerate	Urine	Birth weight	1	Negative
3-hydroxyisovalerate	Urine	Neonatal HOMA-IR	1	Negative
3-hydroxyisovalerate	Urine	Neonatal insulin	1	Negative
Acetate	Urine	Maternal glucose	2	Negative
Acetate	Urine	Maternal HOMA-IR	5	Negative
Acetate	Urine	Maternal insulin	5	Negative
Acetone	Urine	Maternal HOMA-IR	4	Negative
Acetone	Urine	Maternal insulin	4	Negative
Alanine	Urine	Maternal glucose	2	Negative
Alanine	Urine	Maternal HOMA-IR	5	Negative
Alanine	Urine	Maternal insulin	5	Negative
Alanine	Urine	Neonatal fat mass	2	Negative
Alanine	Urine	Neonatal insulin	2	Negative
Betaine	Urine	Maternal glucose	3	Negative
Betaine	Urine	Maternal HOMA-IR	5	Negative
Betaine	Urine	Maternal insulin	5	Negative
Betaine	Urine	Neonatal fat mass	3	Negative
Citrate	Urine	Maternal glucose	5	Negative
Citrate	Urine	Maternal HOMA-IR	5	Negative
Citrate	Urine	Maternal insulin	5	Negative
Citrate	Urine	Neonatal fat mass	1	Negative
Citrate	Urine	Neonatal insulin	1	Negative
Creatinine	Urine	Maternal HOMA-IR	5	Positive
Creatinine	Urine	Maternal insulin	5	Positive
Dimethylamine	Urine	Maternal glucose	1	Positive

Dimethylamine	Urine	Maternal HOMA-IR	5	Positive
Dimethylamine	Urine	Maternal insulin	5	Positive
Formate	Urine	Maternal glucose	3	Negative
Formate	Urine	Maternal HOMA-IR	5	Negative
Formate	Urine	Maternal insulin	5	Negative
Formate	Urine	Neonatal fat mass	1	Negative
Guanidinoacetate	Urine	Maternal HOMA-IR	4	Negative
Guanidinoacetate	Urine	Maternal insulin	3	Negative
Methylmalonate	Urine	Birth weight	4	Negative
Methylmalonate	Urine	Birth weight percentile	3	Negative
Methylmalonate	Urine	Neonatal fat mass	1	Negative
Methylmalonate	Urine	Neonatal HOMA-IR	4	Negative
Methylmalonate	Urine	Neonatal insulin	3	Negative
Oxaloacetate	Urine	Maternal HOMA-IR	1	Negative
Oxaloacetate	Urine	Maternal insulin	1	Negative
Oxaloacetate	Urine	Neonatal fat mass	1	Negative
Oxaloacetate	Urine	Neonatal insulin	1	Negative
PAG	Urine	Maternal HOMA-IR	2	Negative
PAG	Urine	Maternal insulin	2	Negative
Succinate	Urine	Maternal glucose	1	Negative
Succinate	Urine	Maternal HOMA-IR	4	Negative
Succinate	Urine	Maternal insulin	4	Negative
Succinate	Urine	Neonatal fat mass	1	Negative
Unknown-3	Urine	Maternal glucose	3	Positive
Unknown-3	Urine	Maternal HOMA-IR	5	Positive
Unknown-3	Urine	Maternal insulin	5	Positive

**number of time points in which the correlation was significant (FDR <0.05)*
PAG phenylacetylglutamine; GlycA N-acetylglycoproteins; HOMA-IR homeostatic model assessment for insulin resistance; HbA1c glycated haemoglobin

Table 3.S2 Integrative analysis of umbilical cord metabolites and neonatal clinical data. Correlations were approximated from loadings of a partial least squares regression model (*X*: metabolite, *Y*: clinical parameter) and *P* values were calculated from permutation testing.

Metabolite	Clinical parameter	Correlation	*P*	*FDR*
Acetate	Birth weight percentile	0.291	**3.00E-03**	**7.99E-03**
Alanine	Birth weight percentile	0.261	**0.018**	**0.018**
1-methylhistidine	Birth weight percentile	0.267	**5.99E-03**	**0.011**
Valine	Birth weight percentile	0.311	**5.99E-03**	**0.011**
Glutamine	Neonatal insulin	0.271	**7.99E-03**	**0.011**
Acetate	Neonatal insulin	0.302	**7.99E-03**	**0.011**
Alanine	Neonatal insulin	0.441	**9.99E-04**	**5.33E-03**
1-methylhistidine	Neonatal insulin	0.310	**2.00E-03**	**7.99E-03**
Valine	Neonatal insulin	0.318	**6.99E-03**	**0.011**
Acetate	Neonatal glucose	-0.274	**5.99E-03**	**0.011**
Tyrosine	Neonatal glucose	-0.261	**9.99E-03**	**0.011**
Isoleucine	Neonatal glucose	-0.315	**3.00E-03**	**7.99E-03**
Lactate	Neonatal glucose	0.264	**0.015**	**0.016**
Pyruvate	Neonatal glucose	0.388	**9.99E-04**	**5.33E-03**
Valine	Neonatal glucose	-0.306	**9.99E-03**	**0.011**
Alanine	Neonatal HOMA-IR	0.495	**9.99E-04**	**5.33E-03**

HOMA-IR homeostatic model assessment for insulin resistance

Chapter 4

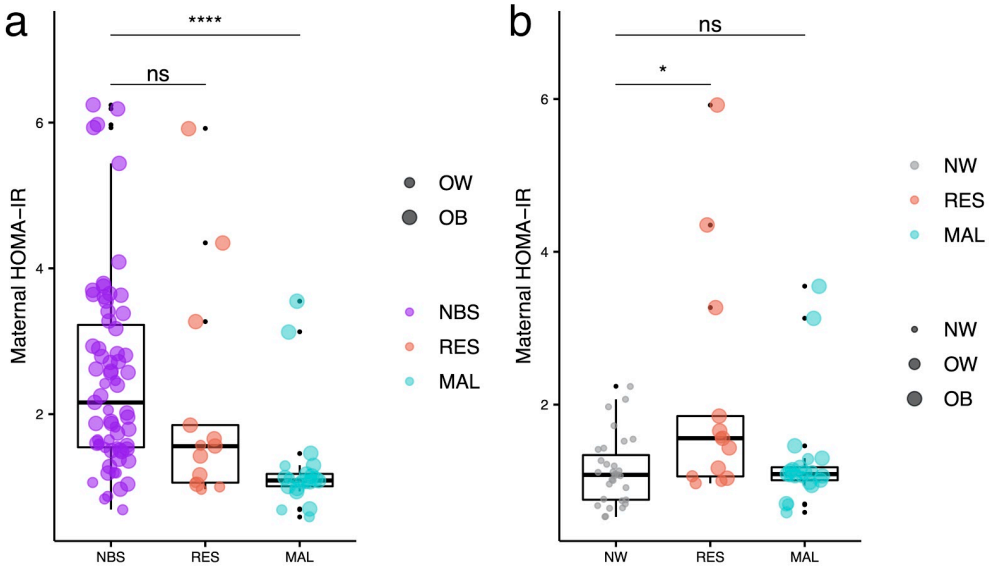


Figure 4.S1: Maternal insulin resistance mid-pregnancy. a) Insulin resistance, measured using the homeostatic model assessment for insulin resistance (HOMA-IR), was elevated in women without bariatric surgery (NBS) compared to those with a previous restrictive (RES) or malabsorptive (MAL) surgery in spite of the similarities in maternal BMI. **b)** Women with a previous MAL surgery maintained insulin sensitivity comparable to normal weight (NW) women. *OW overweight; OB obese; ns not significant; *P<0.05; ****P<0.0001*.

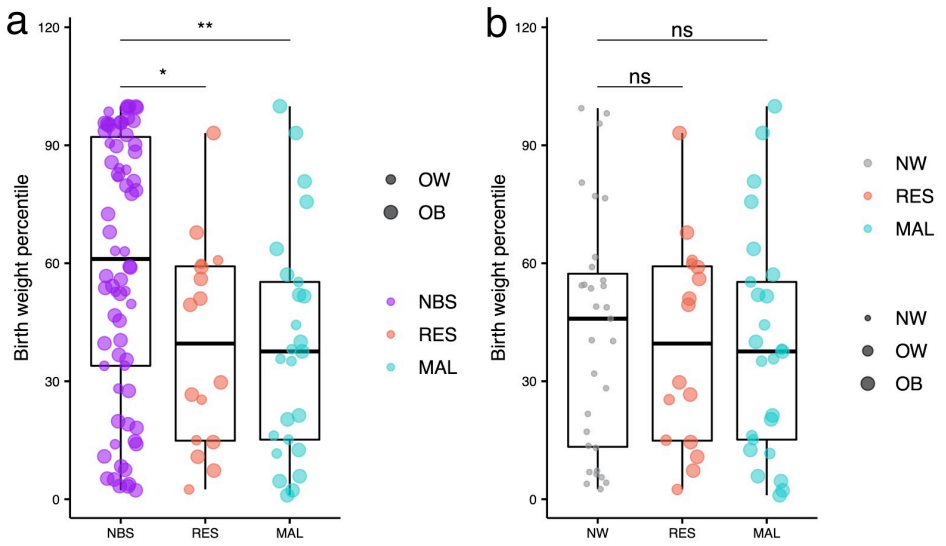


Figure 4.S2: Birth weight percentiles. a) Women without bariatric surgery (NBS) gave birth to larger babies than those with restrictive (RES) or malabsorptive (MAL) surgery in spite of similarities in maternal BMI. **b)** Babies born from bariatric mothers had birth weights comparable with those born from normal weight (NW) women. *OW overweight; OB obese; ns not significant; *P<0.05; **P<0.01.*

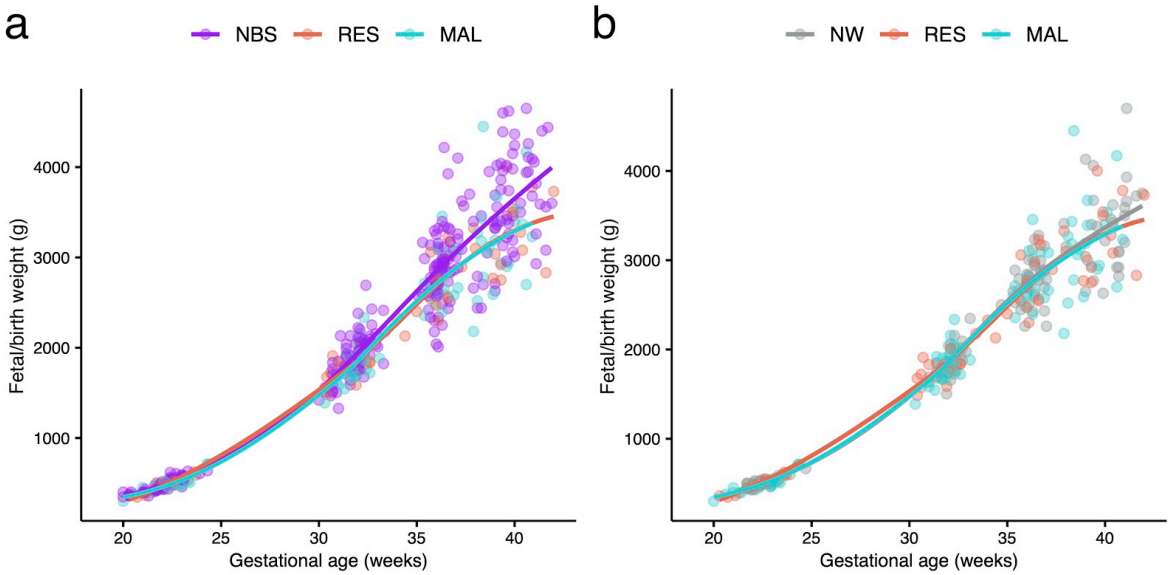


Figure 4.S3: Fetal growth over gestation. a) Estimated fetal weight or birth weight is plotted by week of gestation for women with no (NBS), restrictive (RES) or malabsorptive (MAL) bariatric surgery. **b)** Comparison of growth trajectories between bariatric and normal weight (NW) patients. Curves represent local polynomial regressions.

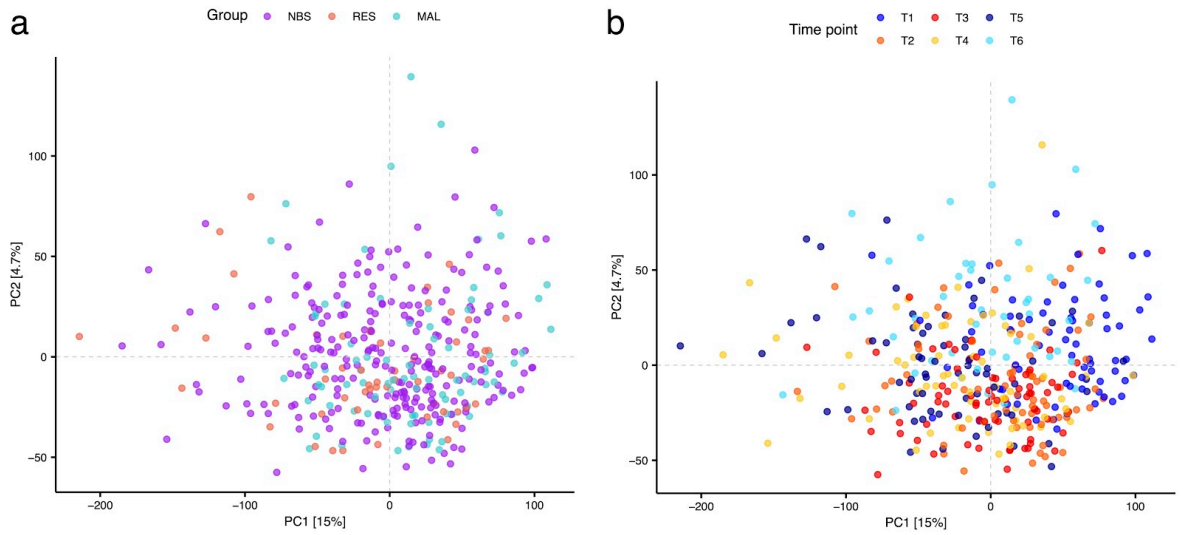


Figure 4.S4: Principal component analysis of all maternal serum samples. Points are coloured by **a)** study group and **b)** time point. *PC principal component; NBS no bariatric surgery; RES restrictive; MAL malabsorptive; T1 11-14 weeks gestation; T2 20-24 weeks gestation; T3 28-30 weeks gestation; T4 30-33 weeks gestation; T5 35-37 weeks gestation; T6 delivery.*

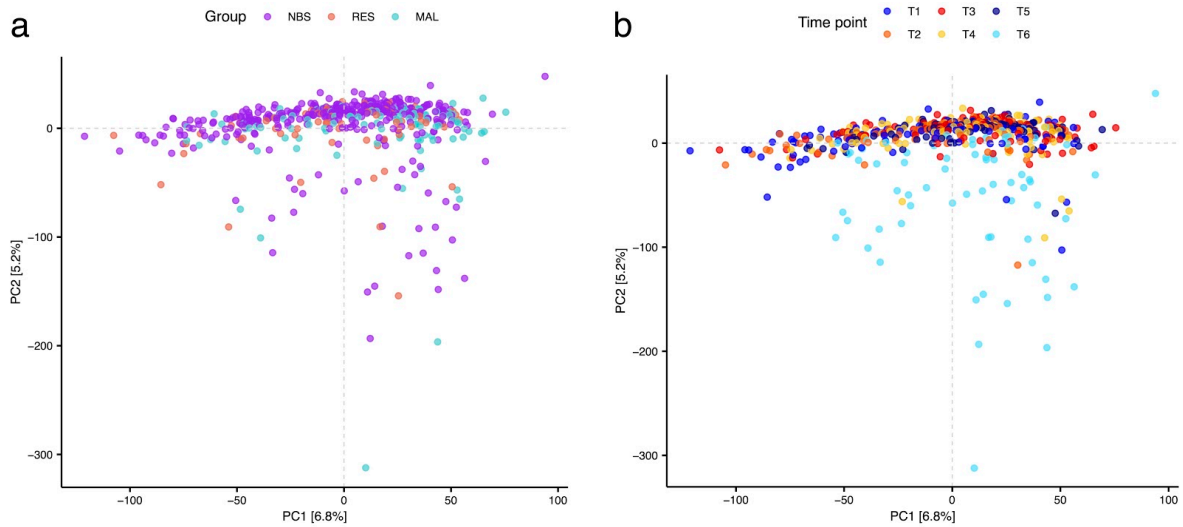


Figure 4.S5: Principal component analysis of all maternal urine samples. Points are coloured by **a)** study group and **b)** time point. Many urine samples collected at the delivery time point (T6) contained drug metabolites. *PC principal component; NBS no bariatric surgery; RES restrictive; MAL malabsorptive; T1 11-14 weeks gestation; T2 20-24 weeks gestation; T3 28-30 weeks gestation; T4 30-33 weeks gestation; T5 35-37 weeks gestation; T6 delivery.*

Table 4.S1: OPLS-DA model statistics

Time point	Biofluid	Samples	Outliers Excluded	Class	R^2Y	Q^2	*P* value
11-14 weeks (T1)	Urine	NBS n = 63	NBS n = 3	NBS vs RES	0.776	-0.117	NS
		RES n = 6	RES n = 1	RES vs MAL	0.935	-0.025	NS
		MAL n = 16	MAL n = 2	**NBS vs MAL**	**0.830**	**0.159**	**0.015388**
	Serum	NBS n = 54	NBS n = 0	NBS vs RES	0.817	-0.274	NS
		RES n = 6	RES n = 0	RES vs MAL	0.929	-0.044	NS
		MAL n = 12	MAL n = 0	NBS vs MAL	0.749	-0.156	NS
20-24 weeks (T2)	Urine	NBS n = 61	NBS n = 1	NBS vs RES	0.810	0.094	NS
		RES n = 11	RES n = 1	RES vs MAL	0.941	0.297	NS
		MAL n = 16	MAL n = 0	**NBS vs MAL**	**0.843**	**0.324**	**1.17E-05**
	Serum	NBS n = 47	NBS n = 0	NBS vs RES	0.830	-0.067	NS
		RES n = 10	RES n = 0	RES vs MAL	0.900	0.056	NS
		MAL n = 15	MAL n = 0	NBS vs MAL	0.825	0.114	NS
28-30 weeks (T3)	Urine	NBS n = 58	NBS n = 0	NBS vs RES	0.826	0.046	NS
		RES n = 11	RES n = 1	RES vs MAL	0.893	0.343	NS
		MAL n = 17	MAL n = 3	**NBS vs MAL**	**0.843**	**0.407**	**3.71E-07**
	Serum	NBS n = 47	NBS n = 0	NBS vs RES	0.748	-0.218	NS
		RES n = 12	RES n = 0	RES vs MAL	0.948	0.216	NS
		MAL n = 15	MAL n = 0	NBS vs MAL	0.744	0.091	NS
30-33 weeks (T4)	Urine	NBS n = 52	NBS n = 2	NBS vs RES	0.790	0.052	NS
		RES n = 12	RES n = 1	RES vs MAL	0.921	0.086	NS
		MAL n = 18	MAL n = 1	**NBS vs MAL**	**0.812**	**0.344**	**2.42E-05**
	Serum	NBS n = 43	NBS n = 0	NBS vs RES	0.742	-0.098	NS
		RES n = 8	RES n = 0	RES vs MAL	0.881	0.214	NS
		MAL n = 15	MAL n = 0	**NBS vs MAL**	**0.775**	**0.178**	**0.027**
35-37 weeks (T5)	Urine	NBS n = 51	NBS n = 2	NBS vs RES	0.802	0.023	NS
		RES n = 12	RES n = 0	RES vs MAL	0.935	0.001	NS
		MAL n = 18	MAL n = 0	**NBS vs MAL**	**0.852**	**0.483**	**2.11E-08**

	Serum	NBS n = 38	NBS n = 0	NBS vs RES	0.908	-0.377	NS
		RES n = 10	RES n = 0	RES vs MAL	0.916	0.224	NS
		MAL n = 17	MAL n = 0	**NBS vs MAL**	**0.847**	**0.245**	**0.006**
Delivery (T6)	Urine	NBS n = 41	NBS n = 2	NBS vs RES	0.844	0.055	NS
		RES n = 12	RES n = 0	RES vs MAL	0.907	0.110	NS
		MAL n = 16	MAL n = 4	NBS vs MAL	0.767	0.067	NS
	Serum	NBS n = 27	NBS n = 0	NBS vs RES	0.895	-0.116	NS
		RES n = 7	RES n = 0	RES vs MAL	0.903	-0.103	NS
		MAL n = 11	MAL n = 0	NBS vs MAL	0.789	0.028	NS
	Cord Blood	NBS n = 30	NBS n = 0	NBS vs RES	0.886	-0.022	NS
		RES n =10	RES n =0	RES vs MAL	0.925	-0.025	NS
		MAL n = 14	MAL n = 0	NBS vs MAL	0.885	0.031	NS
	Newborn urine	NBS n = 18	NBS n = 0	Univariate analysis only			
		RES n = 6	RES n = 0				
		MAL n = 4	MAL n = 0				

NBS no bariatric surgery; RES restrictive; MAL malabsorptive; NS not significant (P <0.05)

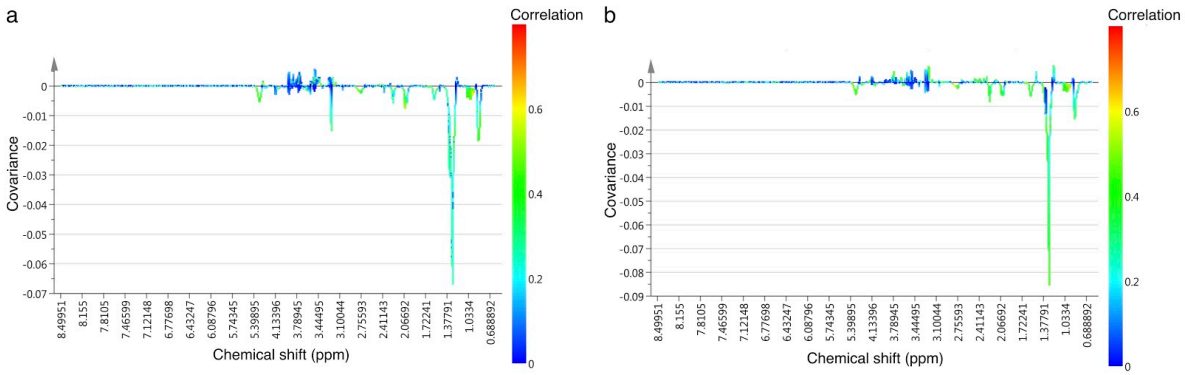


Figure 4.S6: Loadings from orthogonal partial least squares discriminant analysis of serum samples. Discrimination between malabsorptive and control patients was achieved at two time points. Loadings plots for models of ^{1}H NMR serum data at **a)** T4 (30-33 weeks gestation) and **b)** T5 (35-37 weeks gestation) are shown. Signals in the positive direction are associated with malabsorptive surgery.

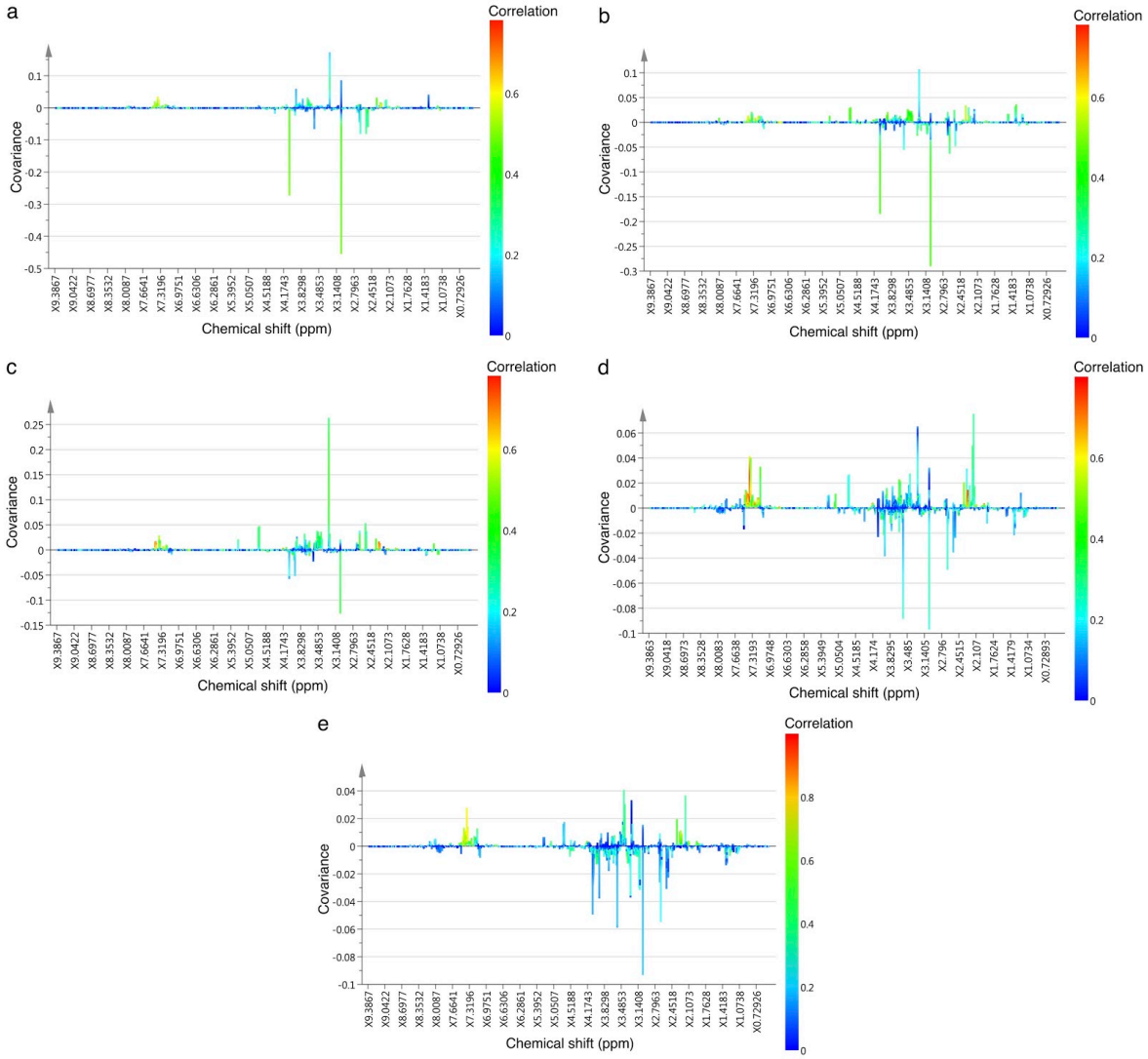


Figure 4.S7: Loadings from orthogonal partial least squares discriminant analysis of urine samples. Loadings plots for models of ^{1}H NMR urine data at **a)** T1 (11-14 weeks gestation), **b)** T2 (20-24 weeks gestation), **c)** T3 (28-30 weeks gestation), **d)** T4 (30-33 weeks gestation) and **e)** T5 (35-37 weeks gestation) are shown. Signals in the positive direction are associated with malabsorptive surgery.

Table 4.S2: Differential abundance in malabsorptive patients. For each taxonomic rank, log_2(FC) of taxa relative abundances at each time point are reported with *P* values from Mann-Whitney U tests. Taxa with *FDR* <0.05 are reported.

Rank	Time point	Taxon	Mean relative abundance malabsorptive	Mean relative abundance control	*P* value	log_2(FC)	*FDR*
Phylum	T1	*Proteobacteria*	0.13632	0.02701	0.0017	2.34	0.0237
Phylum	T4	*Bacteroidetes*	0.39265	0.52280	0.0016	-0.41	0.0237
Phylum	T4	*Proteobacteria*	0.16318	0.03249	5.45E-05	2.33	0.0020
Class	T1	*Gammaproteobacteria*	0.12379	0.01714	0.0022	2.85	0.0248
Class	T1	*Clostridia*	0.28025	0.40068	0.0048	-0.52	0.0357
Class	T1	*Bacilli*	0.05978	0.00789	0.0007	2.92	0.0144
Class	T2	*Bacilli*	0.03774	0.00694	0.0011	2.44	0.0202
Class	T4	*Bacteroidia*	0.39459	0.52538	0.0019	-0.41	0.0248
Class	T4	*Gammaproteobacteria*	0.13910	0.01069	2.61E-05	3.70	0.0020
Class	T4	*Bacilli*	0.07171	0.00987	5.45E-05	2.86	0.0020
Order	T1	*Enterobacteriales*	0.12097	0.01627	0.0048	2.89	0.0357
Order	T1	*Clostridiales*	0.28024	0.40059	0.0048	-0.52	0.0357
Order	T1	*Lactobacillales*	0.05906	0.00789	0.0007	2.90	0.0144
Order	T2	*Enterobacteriales*	0.11434	0.02743	0.0066	2.06	0.0425
Order	T2	*Lactobacillales*	0.03772	0.00689	0.0011	2.45	0.0202
Order	T4	*Bacteroidales*	0.39470	0.52539	0.0019	-0.41	0.0248
Order	T4	*Enterobacteriales*	0.12698	0.00888	4.54E-05	3.84	0.0020
Order	T4	*Lactobacillales*	0.07137	0.00982	5.45E-05	2.86	0.0020
Order	T4	*Pasteurellales*	0.01194	0.00181	0.0043	2.72	0.0357
Order	T4	*Actinomycetales*	0.00084	0.00025	0.0043	1.76	0.0357
Family	T1	*Enterobacteriaceae*	0.12514	0.01639	0.0048	2.93	0.0357

Family	T1	*Streptococcaceae*	0.05582	0.00795	0.0029	2.81	0.0285
Family	T1	*Micrococcaceae*	0.00028	8.76E-05	0.0042	1.66	0.0357
Family	T2	*Enterobacteriaceae*	0.11799	0.02810	0.0066	2.07	0.0425
Family	T2	*Enterococcaceae*	0.00290	1.30E-05	0.0003	7.80	0.0086
Family	T2	*Streptococcaceae*	0.03420	0.00689	0.0021	2.31	0.0248
Family	T2	*Micrococcaceae*	0.00027	4.95E-05	0.0028	2.45	0.0285
Family	T4	*Enterobacteriaceae*	0.12905	0.00927	4.54E-05	3.80	0.0020
Family	T4	*Enterococcaceae*	0.00179	2.78E-05	0.0003	6.01	0.0086
Family	T4	*Streptococcaceae*	0.07027	0.00983	5.45E-05	2.84	0.0020
Family	T4	*Pasteurellaceae*	0.01203	0.00183	0.0043	2.72	0.0357
Family	T4	*Carnobacteriaceae*	0.00032	6.41E-05	0.0013	2.30	0.0222
Family	T4	*Micrococcaceae*	0.00023	8.65E-05	0.0079	1.40	0.0497
Genus	T1	*Escherichia.Shigella*	0.11566	0.01274	0.0022	3.18	0.0248
Genus	T1	*Streptococcus*	0.06395	0.00895	0.0022	2.84	0.0248
Genus	T1	*Anaerostipes*	0.00281	0.01592	0.0037	-2.50	0.0350
Genus	T1	*Holdemanella*	0.02151	0.00238	0.0066	3.18	0.0425
Genus	T1	*Rothia*	0.00031	9.97E-05	0.0055	1.64	0.0380
Genus	T2	*Enterococcus*	0.00323	1.45E-05	0.0003	7.80	0.0086
Genus	T2	*Streptococcus*	0.03881	0.00794	0.0028	2.29	0.0285
Genus	T2	*Butyricimonas*	0.00191	0.00029	0.0050	2.70	0.0357
Genus	T2	*Rothia*	0.00030	5.59E-05	0.0028	2.43	0.0285
Genus	T4	*Escherichia.Shigella*	0.12798	0.00838	5.44E-05	3.93	0.0020
Genus	T4	*Enterococcus*	0.00202	3.12E-05	0.0003	6.02	0.0086
Genus	T4	*Streptococcus*	0.07909	0.01096	5.45E-05	2.85	0.0020
Genus	T4	*Anaerostipes*	0.00395	0.01148	0.0064	-1.54	0.0425

Genus	T4	*Haemophilus*	0.01321	0.00202	0.0049	2.71	0.0357
Genus	T4	*Veillonella*	0.00982	0.00120	0.0004	3.04	0.0097
Genus	T4	*Acidaminococus*	0.00350	7.71E-05	0.0044	5.50	0.0357
Genus	T4	*Solobacterium*	0.00118	0.00029	0.0014	2.01	0.0223
Genus	T4	*Granulicatella*	0.00035	7.13E-05	0.0015	2.31	0.0233

FC fold change; T1 11-14 weeks gestation; T2 20-24 weeks gestation; T4 30-33 weeks gestation

Table 4.S3: Correlation and partial correlation between maternal metabolites and clinical variables. Spearman's correlation coefficients are reported between metabolite concentrations at a specific time point (Variable 1) and clinical or biochemical measures (Variable 2). For each statistically significant (*FDR* <0.05) correlation, the partial correlation is also detailed. Maternal measures were adjusted for maternal age and BMI. Neonatal measures were adjusted for maternal age, BMI and HOMA-IR.

Variable 1	Variable 2	Spearman	*P*	*FDR*	Partial Spearman	*P*	*FDR*
PAG T1	Fasting Maternal Insulin (microU/ml)	-0.37	6.26E-04	7.51E-03	-0.28	1.34E-02	**2.11E-02**
PCS T1	Fasting Maternal Insulin (microU/ml)	-0.36	1.11E-03	9.74E-03	-0.28	1.23E-02	**2.03E-02**
IS T1	Fasting Maternal Insulin (microU/ml)	-0.31	4.47E-03	2.35E-02	-0.30	7.53E-03	**1.61E-02**
PAG T2	Fasting Maternal Insulin (microU/ml)	-0.46	5.15E-06	4.01E-04	-0.32	3.27E-03	**8.79E-03**
PCS T2	Fasting Maternal Insulin (microU/ml)	-0.35	7.61E-04	7.96E-03	-0.23	3.35E-02	**3.98E-02**
Ile T2	Fasting Maternal Insulin (microU/ml)	0.4	5.19E-04	7.27E-03	0.38	1.19E-03	**6.64E-03**
Leu T2	Fasting Maternal Insulin (microU/ml)	0.38	1.16E-03	9.74E-03	0.37	1.84E-03	**8.65E-03**
PAG T3	Fasting Maternal Insulin (microU/ml)	-0.46	1.04E-05	4.36E-04	-0.33	2.69E-03	**8.65E-03**
PCS T3	Fasting Maternal Insulin (microU/ml)	-0.42	5.53E-05	1.55E-03	-0.33	2.52E-03	**8.65E-03**
PAG T4	Fasting Maternal Insulin (microU/ml)	-0.4	2.07E-04	3.48E-03	-0.36	1.01E-03	**6.64E-03**

PCS T4	Fasting Maternal Insulin (microU/ml)	-0.4	1.76E-04	3.28E-03	-0.39	4.14E-04	**4.66E-03**
PAG T5	Fasting Maternal Insulin (microU/ml)	-0.48	6.85E-06	4.01E-04	-0.39	4.91E-04	**4.66E-03**
PCS T5	Fasting Maternal Insulin (microU/ml)	-0.41	2.31E-04	3.53E-03	-0.30	8.58E-03	**1.72E-02**
IS T5	Fasting Maternal Insulin (microU/ml)	-0.41	1.75E-04	3.28E-03	-0.36	1.22E-03	**6.64E-03**
Leu T5	Fasting Maternal Insulin (microU/ml)	0.36	4.44E-03	2.35E-02	0.34	7.61E-03	**1.61E-02**
PCS T6	Fasting Maternal Insulin (microU/ml)	-0.32	7.19E-03	3.55E-02	-0.22	7.81E-02	8.25E-02
PAG T1	Maternal HOMA-IR	-0.35	1.48E-03	1.07E-02	-0.25	2.70E-02	**3.31E-02**
PCS T1	Maternal HOMA-IR	-0.35	1.61E-03	1.08E-02	-0.27	1.64E-02	**2.30E-02**
IS T1	Maternal HOMA-IR	-0.29	8.22E-03	3.94E-02	-0.27	1.49E-02	**2.26E-02**
PAG T2	Maternal HOMA-IR	-0.46	7.17E-06	4.01E-04	-0.32	3.47E-03	**8.79E-03**
PCS T2	Maternal HOMA-IR	-0.33	1.53E-03	1.07E-02	-0.20	6.70E-02	7.48E-02
Ile T2	Maternal HOMA-IR	0.38	8.87E-04	8.28E-03	0.35	2.61E-03	**8.65E-03**
Leu T2	Maternal HOMA-IR	0.33	4.18E-03	2.34E-02	0.31	9.94E-03	**1.81E-02**
PAG T3	Maternal HOMA-IR	-0.41	9.80E-05	2.35E-03	-0.26	1.81E-02	**2.46E-02**
PCS T3	Maternal HOMA-IR	-0.35	8.05E-04	7.96E-03	-0.23	3.46E-02	**3.99E-02**
Ile T3	Maternal HOMA-IR	0.3	9.91E-03	4.50E-02	0.20	9.75E-02	1.00E-01
PAG T4	Maternal HOMA-IR	-0.37	6.84E-04	7.66E-03	-0.32	3.40E-03	**8.79E-03**
PCS T4	Maternal HOMA-IR	-0.35	1.48E-03	1.07E-02	-0.31	4.70E-03	**1.12E-02**
PAG T5	Maternal HOMA-IR	-0.44	5.19E-05	1.55E-03	-0.34	2.73E-03	**8.65E-03**
PCS T5	Maternal HOMA-IR	-0.33	2.98E-03	1.86E-02	-0.21	7.01E-02	7.61E-02
IS T5	Maternal HOMA-IR	-0.36	1.38E-03	1.07E-02	-0.29	1.05E-02	**1.81E-02**
Leu T5	Maternal HOMA-IR	0.32	1.08E-02	4.79E-02	0.29	2.49E-02	**3.26E-02**

PAG T2	EFW T2	-0.31	3.21E-03	1.92E-02	-0.26	1.58E-02	**2.30E-02**
PCS T2	EFW T2	-0.29	6.57E-03	3.34E-02	-0.28	1.01E-02	**1.81E-02**
Ile T2	EFW T2	0.3	9.49E-03	4.43E-02	0.20	1.01E-01	1.01E-01
IS T5	EFW T5	-0.32	4.02E-03	2.33E-02	-0.26	2.68E-02	**3.31E-02**
PAG T6	Birth weight	-0.36	2.70E-03	1.74E-02	-0.46	1.95E-04	**3.71E-03**
IS T6	Birth weight	-0.4	5.88E-04	7.51E-03	-0.49	5.22E-05	**1.98E-03**

PAG phenylacetylglutamine; PCS p-cresol sulfate; IS indoxyl sulfate; Ile isoleucine; Leu leucine; HOMA-IR Homeostatic Model Assessment of Insulin Resistance; EFW estimated fetal weight; T1 11-14 weeks gestation; T2 20-24 weeks gestation; T3 28-30 weeks gestation; T4 30-33 weeks gestation; T5 35-37 weeks gestation; T6 delivery

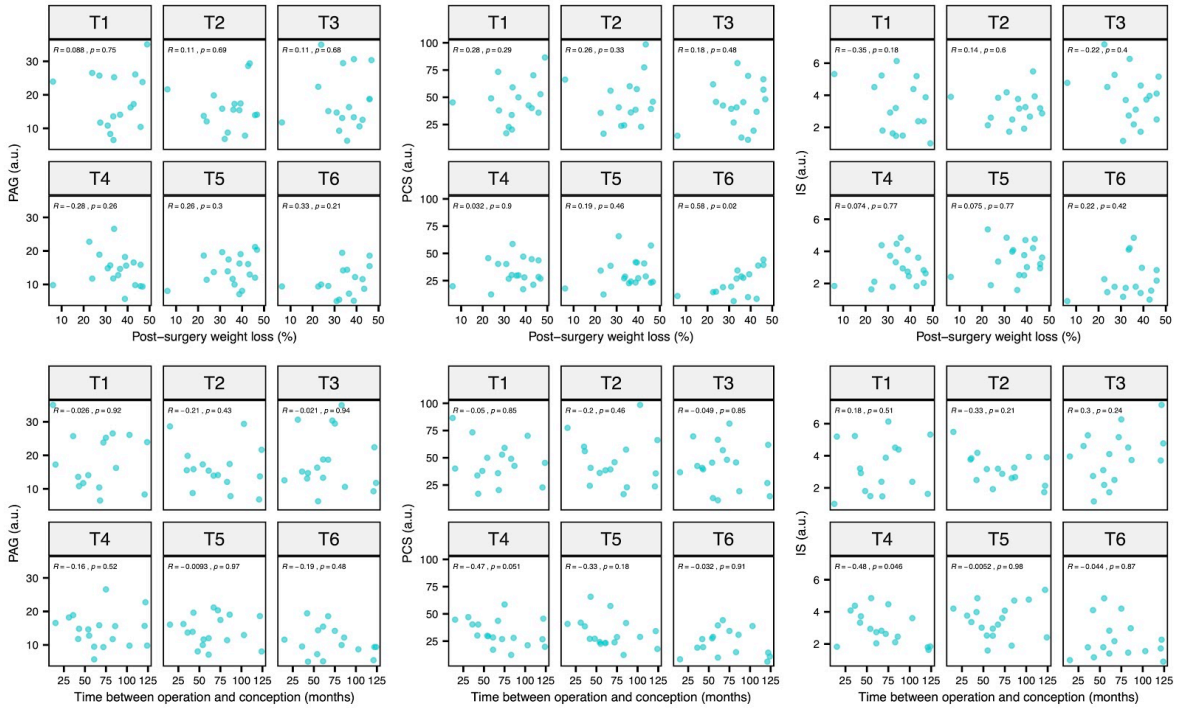


Figure 4.S8: Protein fermentation metabolites are not related to percentage of weight lost or time since surgery. Correlation coefficients reported are Spearman. Relative concentrations are reported in arbitrary units (a.u.). *PAG phenylacetylglutamine; PCS p-cresyl sulfate; IS indoxyl sulfate. T1 11-14 weeks gestation; T2 20-24 weeks gestation; T3 28-30 weeks gestation; T4 30-33 weeks gestation; T5 35-37 weeks gestation; T6 delivery*

Chapter 5

Table 5.S1 Associations between metabolites and faecal genera in ASD. Correlations were approximated from the loadings of a partial least squares regression model (*X*: genera, *Y*: metabolites). *P* values were calculated by permutation testing. Insignificant correlations (*P* <0.05) are omitted from the table. All metabolites were measured in faecal samples except phenylacetylglutamine (PAG) and *p*-cresyl sulfate (PCS), which were measured in urine samples.

Order	Family	Genus	Metabolite	Correlation	*FDR*
Actinomycetales	*Micrococcaceae*	*Rothia*	alanine	-0.115	0.068
Actinomycetales	*Micrococcaceae*	*Rothia*	glutamate	-0.187	**5.50E-03**
Actinomycetales	*Micrococcaceae*	*Rothia*	uracil	-0.301	**1.97E-03**
Bacillales	*Bacillales Incertae Sedis XI*	*Gemella*	alanine	-0.110	0.057
Bacillales	*Bacillales Incertae Sedis XI*	*Gemella*	glutamate	-0.162	**0.014**
Bacillales	*Bacillales Incertae Sedis XI*	*Gemella*	uracil	-0.246	**1.97E-03**
Bacillales	*Staphylococcaceae*	*Staphylococcus*	uracil	-0.164	**1.97E-03**
Bacteroidales	*Bacteroidaceae*	*Bacteroides*	uracil	-0.145	**0.025**
Bacteroidales	*Porphyromonadaceae*	*Butyricimonas*	alanine	-0.146	**4.30E-03**
Bacteroidales	*Porphyromonadaceae*	*Butyricimonas*	glutamate	-0.129	**3.33E-03**
Bacteroidales	*Porphyromonadaceae*	*Butyricimonas*	glycine	-0.124	**4.30E-03**
Bacteroidales	*Porphyromonadaceae*	*Butyricimonas*	isoleucine	-0.128	**7.60E-03**
Bacteroidales	*Porphyromonadaceae*	*Butyricimonas*	leucine	-0.137	**4.30E-03**
Bacteroidales	*Porphyromonadaceae*	*Butyricimonas*	methionine	-0.108	**8.82E-03**
Bacteroidales	*Porphyromonadaceae*	*Butyricimonas*	tyrosine	-0.112	**6.58E-03**
Bacteroidales	*Porphyromonadaceae*	*Butyricimonas*	valine	-0.129	**1.97E-03**
Bacteroidales	*Porphyromonadaceae*	*Odoribacter*	alanine	-0.110	**0.023**
Bacteroidales	*Porphyromonadaceae*	*Porphyromonas*	uracil	-0.164	**1.97E-03**
Bacteroidales	*Prevotellaceae*	*Paraprevotella*	alanine	-0.111	**6.58E-03**
Bacteroidales	*Prevotellaceae*	*Paraprevotella*	leucine	-0.108	**3.33E-03**
Bacteroidales	*Prevotellaceae*	*Prevotella*	alanine	-0.246	**1.97E-03**
Bacteroidales	*Prevotellaceae*	*Prevotella*	aspartate	-0.135	**3.33E-03**
Bacteroidales	*Prevotellaceae*	*Prevotella*	glutamate	-0.184	**1.97E-03**
Bacteroidales	*Prevotellaceae*	*Prevotella*	glycine	-0.206	**1.97E-03**
Bacteroidales	*Prevotellaceae*	*Prevotella*	isoleucine	-0.231	**1.97E-03**
Bacteroidales	*Prevotellaceae*	*Prevotella*	leucine	-0.247	**1.97E-03**
Bacteroidales	*Prevotellaceae*	*Prevotella*	methionine	-0.194	**1.97E-03**
Bacteroidales	*Prevotellaceae*	*Prevotella*	PAG	-0.131	**1.97E-03**
Bacteroidales	*Prevotellaceae*	*Prevotella*	PCS	-0.163	**1.97E-03**
Bacteroidales	*Prevotellaceae*	*Prevotella*	phenylalanine	-0.212	**1.97E-03**
Bacteroidales	*Prevotellaceae*	*Prevotella*	tryptophan	-0.144	**1.97E-03**

Bacteroidales	*Prevotellaceae*	*Prevotella*	tyrosine	-0.218	**1.97E-03**
Bacteroidales	*Prevotellaceae*	*Prevotella*	valine	-0.243	**1.97E-03**
Clostridiales	*Catabacteriaceae*	*Catabacter*	phenylalanine	0.105	**0.017**
Clostridiales	*Catabacteriaceae*	*Catabacter*	uracil	-0.125	**4.30E-03**
Clostridiales	*Christensenellaceae*	*Christensenella*	isoleucine	0.114	**0.025**
Clostridiales	*Christensenellaceae*	*Christensenella*	leucine	0.120	**0.016**
Clostridiales	*Christensenellaceae*	*Christensenella*	phenylalanine	0.142	**9.94E-03**
Clostridiales	*Christensenellaceae*	*Christensenella*	tryptophan	0.120	**0.012**
Clostridiales	*Christensenellaceae*	*Christensenella*	tyrosine	0.130	**0.015**
Clostridiales	*Christensenellaceae*	*Christensenella*	uracil	-0.111	**0.027**
Clostridiales	*Christensenellaceae*	*Christensenella*	valine	0.132	**0.015**
Clostridiales	*Clostridiaceae 1*	*Clostridium sensu stricto*	glutamate	-0.137	**0.024**
Clostridiales	*Clostridiaceae 1*	*Clostridium sensu stricto*	uracil	-0.240	**1.97E-03**
Clostridiales	*Clostridiales Incertae Sedis XI*	*Anaerococcus*	glutamate	-0.178	**7.60E-03**
Clostridiales	*Clostridiales Incertae Sedis XI*	*Anaerococcus*	uracil	-0.299	**1.97E-03**
Clostridiales	*Clostridiales Incertae Sedis XI*	*Finegoldia*	alanine	-0.110	0.069
Clostridiales	*Clostridiales Incertae Sedis XI*	*Finegoldia*	glutamate	-0.180	**5.50E-03**
Clostridiales	*Clostridiales Incertae Sedis XI*	*Finegoldia*	uracil	-0.292	**1.97E-03**
Clostridiales	*Clostridiales Incertae Sedis XI*	*Parvimonas*	alanine	-0.105	**0.045**
Clostridiales	*Clostridiales Incertae Sedis XI*	*Parvimonas*	glutamate	-0.158	**4.30E-03**
Clostridiales	*Clostridiales Incertae Sedis XI*	*Parvimonas*	uracil	-0.243	**1.97E-03**
Clostridiales	*Clostridiales Incertae Sedis XII*	*Fusibacter*	uracil	-0.184	**1.97E-03**
Clostridiales	*Clostridiales Incertae Sedis XIII*	*Anaerovorax*	alanine	0.372	**1.97E-03**
Clostridiales	*Clostridiales Incertae Sedis XIII*	*Anaerovorax*	aspartate	0.216	**1.97E-03**
Clostridiales	*Clostridiales Incertae Sedis XIII*	*Anaerovorax*	glutamate	0.312	**1.97E-03**
Clostridiales	*Clostridiales Incertae Sedis XIII*	*Anaerovorax*	glycine	0.314	**1.97E-03**
Clostridiales	*Clostridiales Incertae Sedis XIII*	*Anaerovorax*	isoleucine	0.332	**1.97E-03**
Clostridiales	*Clostridiales Incertae Sedis XIII*	*Anaerovorax*	leucine	0.356	**1.97E-03**
Clostridiales	*Clostridiales Incertae Sedis XIII*	*Anaerovorax*	methionine	0.281	**1.97E-03**
Clostridiales	*Clostridiales Incertae Sedis XIII*	*Anaerovorax*	PAG	0.204	**1.97E-03**
Clostridiales	*Clostridiales Incertae Sedis XIII*	*Anaerovorax*	PCS	0.245	**1.97E-03**
Clostridiales	*Clostridiales Incertae Sedis XIII*	*Anaerovorax*	phenylalanine	0.280	**1.97E-03**
Clostridiales	*Clostridiales Incertae Sedis XIII*	*Anaerovorax*	tryptophan	0.175	**1.97E-03**
Clostridiales	*Clostridiales Incertae Sedis XIII*	*Anaerovorax*	tyrosine	0.299	**1.97E-03**
Clostridiales	*Clostridiales Incertae Sedis XIII*	*Anaerovorax*	uracil	0.192	**3.33E-03**
Clostridiales	*Clostridiales Incertae Sedis XIII*	*Anaerovorax*	valine	0.341	**1.97E-03**
Clostridiales	*Clostridiales Incertae Sedis XIII*	*Mogibacterium*	alanine	0.108	**0.018**
Clostridiales	*Clostridiales Incertae Sedis XIII*	*Mogibacterium*	isoleucine	0.112	**0.013**
Clostridiales	*Clostridiales Incertae Sedis XIII*	*Mogibacterium*	leucine	0.118	**7.60E-03**

Clostridiales	*Clostridiales Incertae Sedis XIII*	*Mogibacterium*	phenylalanine	0.118	**9.94E-03**
Clostridiales	*Clostridiales Incertae Sedis XIII*	*Mogibacterium*	tyrosine	0.115	**0.016**
Clostridiales	*Clostridiales Incertae Sedis XIII*	*Mogibacterium*	valine	0.122	**0.012**
Clostridiales	*Eubacteriaceae*	*Anaerofustis*	tryptophan	0.108	0.063
Clostridiales	*Eubacteriaceae*	*Anaerofustis*	uracil	-0.218	**1.97E-03**
Clostridiales	*Lachnospiraceae*	*Anaerostipes*	alanine	0.128	**0.020**
Clostridiales	*Lachnospiraceae*	*Anaerostipes*	glutamate	0.108	**0.038**
Clostridiales	*Lachnospiraceae*	*Anaerostipes*	glycine	0.108	**0.045**
Clostridiales	*Lachnospiraceae*	*Anaerostipes*	isoleucine	0.115	**0.039**
Clostridiales	*Lachnospiraceae*	*Anaerostipes*	leucine	0.123	**0.034**
Clostridiales	*Lachnospiraceae*	*Anaerostipes*	valine	0.118	**0.028**
Clostridiales	*Lachnospiraceae*	*Blautia*	alanine	0.162	**0.017**
Clostridiales	*Lachnospiraceae*	*Blautia*	glutamate	0.142	**0.029**
Clostridiales	*Lachnospiraceae*	*Blautia*	glycine	0.137	**0.038**
Clostridiales	*Lachnospiraceae*	*Blautia*	isoleucine	0.142	**0.032**
Clostridiales	*Lachnospiraceae*	*Blautia*	leucine	0.153	**0.026**
Clostridiales	*Lachnospiraceae*	*Blautia*	methionine	0.120	**0.0496**
Clostridiales	*Lachnospiraceae*	*Blautia*	PCS	0.107	0.069
Clostridiales	*Lachnospiraceae*	*Blautia*	phenylalanine	0.115	**0.043**
Clostridiales	*Lachnospiraceae*	*Blautia*	tyrosine	0.125	**0.047**
Clostridiales	*Lachnospiraceae*	*Blautia*	valine	0.144	**0.026**
Clostridiales	*Lachnospiraceae*	*Clostridium XlVb*	alanine	0.169	**1.97E-03**
Clostridiales	*Lachnospiraceae*	*Clostridium XlVb*	glutamate	0.125	**6.58E-03**
Clostridiales	*Lachnospiraceae*	*Clostridium XlVb*	glycine	0.141	**3.33E-03**
Clostridiales	*Lachnospiraceae*	*Clostridium XlVb*	isoleucine	0.159	**3.33E-03**
Clostridiales	*Lachnospiraceae*	*Clostridium XlVb*	leucine	0.170	**1.97E-03**
Clostridiales	*Lachnospiraceae*	*Clostridium XlVb*	methionine	0.133	**1.97E-03**
Clostridiales	*Lachnospiraceae*	*Clostridium XlVb*	PCS	0.112	**0.011**
Clostridiales	*Lachnospiraceae*	*Clostridium XlVb*	phenylalanine	0.147	**1.97E-03**
Clostridiales	*Lachnospiraceae*	*Clostridium XlVb*	tyrosine	0.150	**4.30E-03**
Clostridiales	*Lachnospiraceae*	*Clostridium XlVb*	valine	0.167	**1.97E-03**
Clostridiales	*Lachnospiraceae*	*Coprococcus*	glutamate	0.112	**1.97E-03**
Clostridiales	*Lachnospiraceae*	*Coprococcus*	uracil	0.123	**3.33E-03**
Clostridiales	*Lachnospiraceae*	*Dorea*	alanine	0.233	**1.97E-03**
Clostridiales	*Lachnospiraceae*	*Dorea*	aspartate	0.147	**7.60E-03**
Clostridiales	*Lachnospiraceae*	*Dorea*	glutamate	0.229	**1.97E-03**
Clostridiales	*Lachnospiraceae*	*Dorea*	glycine	0.198	**1.97E-03**
Clostridiales	*Lachnospiraceae*	*Dorea*	isoleucine	0.191	**1.97E-03**
Clostridiales	*Lachnospiraceae*	*Dorea*	leucine	0.207	**1.97E-03**

Clostridiales	*Lachnospiraceae*	*Dorea*	methionine	0.163	**1.97E-03**
Clostridiales	*Lachnospiraceae*	*Dorea*	PAG	0.133	**7.60E-03**
Clostridiales	*Lachnospiraceae*	*Dorea*	PCS	0.152	**6.58E-03**
Clostridiales	*Lachnospiraceae*	*Dorea*	phenylalanine	0.136	**9.94E-03**
Clostridiales	*Lachnospiraceae*	*Dorea*	tyrosine	0.157	**4.30E-03**
Clostridiales	*Lachnospiraceae*	*Dorea*	uracil	0.211	**3.33E-03**
Clostridiales	*Lachnospiraceae*	*Dorea*	valine	0.188	**1.97E-03**
Clostridiales	*Lachnospiraceae*	*Lactonifactor*	glutamate	-0.148	**0.030**
Clostridiales	*Lachnospiraceae*	*Lactonifactor*	uracil	-0.274	**1.97E-03**
Clostridiales	*Lachnospiraceae*	*Roseburia*	alanine	-0.209	**1.97E-03**
Clostridiales	*Lachnospiraceae*	*Roseburia*	aspartate	-0.115	**7.60E-03**
Clostridiales	*Lachnospiraceae*	*Roseburia*	glutamate	-0.159	**1.97E-03**
Clostridiales	*Lachnospiraceae*	*Roseburia*	glycine	-0.176	**1.97E-03**
Clostridiales	*Lachnospiraceae*	*Roseburia*	isoleucine	-0.195	**1.97E-03**
Clostridiales	*Lachnospiraceae*	*Roseburia*	leucine	-0.208	**1.97E-03**
Clostridiales	*Lachnospiraceae*	*Roseburia*	methionine	-0.164	**1.97E-03**
Clostridiales	*Lachnospiraceae*	*Roseburia*	PAG	-0.112	**6.58E-03**
Clostridiales	*Lachnospiraceae*	*Roseburia*	PCS	-0.139	**3.33E-03**
Clostridiales	*Lachnospiraceae*	*Roseburia*	phenylalanine	-0.177	**1.97E-03**
Clostridiales	*Lachnospiraceae*	*Roseburia*	tryptophan	-0.118	**3.33E-03**
Clostridiales	*Lachnospiraceae*	*Roseburia*	tyrosine	-0.183	**1.97E-03**
Clostridiales	*Lachnospiraceae*	*Roseburia*	valine	-0.204	**1.97E-03**
Clostridiales	*Lachnospiraceae*	*Ruminococcus2*	alanine	0.165	**1.97E-03**
Clostridiales	*Lachnospiraceae*	*Ruminococcus2*	glutamate	0.155	**1.97E-03**
Clostridiales	*Lachnospiraceae*	*Ruminococcus2*	glycine	0.140	**1.97E-03**
Clostridiales	*Lachnospiraceae*	*Ruminococcus2*	isoleucine	0.140	**1.97E-03**
Clostridiales	*Lachnospiraceae*	*Ruminococcus2*	leucine	0.150	**3.33E-03**
Clostridiales	*Lachnospiraceae*	*Ruminococcus2*	methionine	0.119	**3.33E-03**
Clostridiales	*Lachnospiraceae*	*Ruminococcus2*	PCS	0.108	**3.33E-03**
Clostridiales	*Lachnospiraceae*	*Ruminococcus2*	phenylalanine	0.105	**5.50E-03**
Clostridiales	*Lachnospiraceae*	*Ruminococcus2*	tyrosine	0.118	**3.33E-03**
Clostridiales	*Lachnospiraceae*	*Ruminococcus2*	uracil	0.130	**3.33E-03**
Clostridiales	*Lachnospiraceae*	*Ruminococcus2*	valine	0.139	**1.97E-03**
Clostridiales	*Peptoniphilaceae*	*Peptoniphilus*	uracil	-0.214	**1.97E-03**
Clostridiales	*Peptostreptococcaceae*	*Intestinibacter*	glutamate	-0.161	**0.013**
Clostridiales	*Peptostreptococcaceae*	*Intestinibacter*	uracil	-0.268	**1.97E-03**
Clostridiales	*Peptostreptococcaceae*	*Romboutsia*	alanine	-0.114	0.062
Clostridiales	*Peptostreptococcaceae*	*Romboutsia*	glutamate	-0.180	**6.58E-03**
Clostridiales	*Peptostreptococcaceae*	*Romboutsia*	uracil	-0.286	**1.97E-03**

Clostridiales	*Ruminococcaceae*	*Acetanaerobacterium*	glutamate	-0.168	**0.012**
Clostridiales	*Ruminococcaceae*	*Acetanaerobacterium*	uracil	-0.282	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Anaerofilum*	alanine	-0.118	0.070
Clostridiales	*Ruminococcaceae*	*Anaerofilum*	glutamate	-0.189	**5.50E-03**
Clostridiales	*Ruminococcaceae*	*Anaerofilum*	uracil	-0.304	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Butyricicoccus*	alanine	-0.164	**4.30E-03**
Clostridiales	*Ruminococcaceae*	*Butyricicoccus*	glutamate	-0.106	**0.016**
Clostridiales	*Ruminococcaceae*	*Butyricicoccus*	glycine	-0.137	**4.30E-03**
Clostridiales	*Ruminococcaceae*	*Butyricicoccus*	isoleucine	-0.163	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Butyricicoccus*	leucine	-0.173	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Butyricicoccus*	methionine	-0.135	**4.30E-03**
Clostridiales	*Ruminococcaceae*	*Butyricicoccus*	PCS	-0.109	**0.012**
Clostridiales	*Ruminococcaceae*	*Butyricicoccus*	phenylalanine	-0.161	**4.30E-03**
Clostridiales	*Ruminococcaceae*	*Butyricicoccus*	tryptophan	-0.117	**4.30E-03**
Clostridiales	*Ruminococcaceae*	*Butyricicoccus*	tyrosine	-0.160	**4.30E-03**
Clostridiales	*Ruminococcaceae*	*Butyricicoccus*	valine	-0.175	**3.33E-03**
Clostridiales	*Ruminococcaceae*	*Clostridium IV*	alanine	0.172	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Clostridium IV*	glutamate	0.155	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Clostridium IV*	glycine	0.146	**4.30E-03**
Clostridiales	*Ruminococcaceae*	*Clostridium IV*	isoleucine	0.149	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Clostridium IV*	leucine	0.160	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Clostridium IV*	methionine	0.126	**5.50E-03**
Clostridiales	*Ruminococcaceae*	*Clostridium IV*	PCS	0.113	**4.30E-03**
Clostridiales	*Ruminococcaceae*	*Clostridium IV*	phenylalanine	0.117	**6.58E-03**
Clostridiales	*Ruminococcaceae*	*Clostridium IV*	tyrosine	0.129	**4.30E-03**
Clostridiales	*Ruminococcaceae*	*Clostridium IV*	uracil	0.117	**4.30E-03**
Clostridiales	*Ruminococcaceae*	*Clostridium IV*	valine	0.150	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Ethanoligenens*	glutamate	-0.151	**0.026**
Clostridiales	*Ruminococcaceae*	*Ethanoligenens*	uracil	-0.278	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Faecalibacterium*	alanine	-0.221	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Faecalibacterium*	aspartate	-0.119	**7.60E-03**
Clostridiales	*Ruminococcaceae*	*Faecalibacterium*	glutamate	-0.160	**4.30E-03**
Clostridiales	*Ruminococcaceae*	*Faecalibacterium*	glycine	-0.185	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Faecalibacterium*	isoleucine	-0.210	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Faecalibacterium*	leucine	-0.224	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Faecalibacterium*	methionine	-0.176	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Faecalibacterium*	PAG	-0.117	**7.60E-03**
Clostridiales	*Ruminococcaceae*	*Faecalibacterium*	PCS	-0.147	**5.50E-03**
Clostridiales	*Ruminococcaceae*	*Faecalibacterium*	phenylalanine	-0.197	**1.97E-03**

Clostridiales	*Ruminococcaceae*	*Faecalibacterium*	tryptophan	-0.136	**3.33E-03**
Clostridiales	*Ruminococcaceae*	*Faecalibacterium*	tyrosine	-0.201	**3.33E-03**
Clostridiales	*Ruminococcaceae*	*Faecalibacterium*	valine	-0.222	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Flavonifractor*	alanine	0.410	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Flavonifractor*	aspartate	0.234	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Flavonifractor*	glutamate	0.332	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Flavonifractor*	glycine	0.345	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Flavonifractor*	isoleucine	0.372	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Flavonifractor*	leucine	0.398	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Flavonifractor*	methionine	0.313	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Flavonifractor*	PAG	0.223	**3.33E-03**
Clostridiales	*Ruminococcaceae*	*Flavonifractor*	PCS	0.271	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Flavonifractor*	phenylalanine	0.322	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Flavonifractor*	tryptophan	0.207	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Flavonifractor*	tyrosine	0.339	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Flavonifractor*	uracil	0.182	**3.33E-03**
Clostridiales	*Ruminococcaceae*	*Flavonifractor*	valine	0.384	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Intestinimonas*	leucine	0.107	**4.30E-03**
Clostridiales	*Ruminococcaceae*	*Intestinimonas*	phenylalanine	0.106	**5.50E-03**
Clostridiales	*Ruminococcaceae*	*Intestinimonas*	valine	0.111	**3.33E-03**
Clostridiales	*Ruminococcaceae*	*Ruminococcus*	alanine	0.136	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Ruminococcus*	glutamate	0.132	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Ruminococcus*	glycine	0.116	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Ruminococcus*	isoleucine	0.112	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Ruminococcus*	leucine	0.121	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Ruminococcus*	uracil	0.119	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Ruminococcus*	valine	0.111	**3.33E-03**
Clostridiales	*Ruminococcaceae*	*Sporobacter*	glutamate	-0.173	**0.011**
Clostridiales	*Ruminococcaceae*	*Sporobacter*	uracil	-0.290	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Subdoligranulum*	alanine	0.197	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Subdoligranulum*	aspartate	0.120	**4.30E-03**
Clostridiales	*Ruminococcaceae*	*Subdoligranulum*	glutamate	0.181	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Subdoligranulum*	glycine	0.167	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Subdoligranulum*	isoleucine	0.168	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Subdoligranulum*	leucine	0.181	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Subdoligranulum*	methionine	0.143	**3.33E-03**
Clostridiales	*Ruminococcaceae*	*Subdoligranulum*	PAG	0.110	**6.58E-03**
Clostridiales	*Ruminococcaceae*	*Subdoligranulum*	PCS	0.129	**4.30E-03**
Clostridiales	*Ruminococcaceae*	*Subdoligranulum*	phenylalanine	0.129	**4.30E-03**

Clostridiales	*Ruminococcaceae*	*Subdoligranulum*	tyrosine	0.144	**1.97E-03**
Clostridiales	*Ruminococcaceae*	*Subdoligranulum*	uracil	0.144	**4.30E-03**
Clostridiales	*Ruminococcaceae*	*Subdoligranulum*	valine	0.168	**1.97E-03**
Coriobacteriales	*Coriobacteriaceae*	*Collinsella*	uracil	-0.194	**1.97E-03**
Coriobacteriales	*Coriobacteriaceae*	*Eggerthella*	glutamate	-0.160	**0.015**
Coriobacteriales	*Coriobacteriaceae*	*Eggerthella*	uracil	-0.281	**1.97E-03**
Erysipelotrichales	*Erysipelotrichaceae*	*Clostridium XVIII*	uracil	-0.131	**4.30E-03**
Erysipelotrichales	*Erysipelotrichaceae*	*Solobacterium*	alanine	-0.130	**0.024**
Erysipelotrichales	*Erysipelotrichaceae*	*Solobacterium*	glutamate	-0.183	**3.33E-03**
Erysipelotrichales	*Erysipelotrichaceae*	*Solobacterium*	glycine	-0.114	**0.039**
Erysipelotrichales	*Erysipelotrichaceae*	*Solobacterium*	uracil	-0.266	**1.97E-03**
Lactobacillales	*Aerococcaceae*	*Abiotrophia*	alanine	-0.129	**0.043**
Lactobacillales	*Aerococcaceae*	*Abiotrophia*	aspartate	-0.105	0.075
Lactobacillales	*Aerococcaceae*	*Abiotrophia*	glutamate	-0.193	**3.33E-03**
Lactobacillales	*Aerococcaceae*	*Abiotrophia*	glycine	-0.113	0.065
Lactobacillales	*Aerococcaceae*	*Abiotrophia*	uracil	-0.294	**1.97E-03**
Lactobacillales	*Carnobacteriaceae*	*Granulicatella*	uracil	-0.137	**0.015**
Lactobacillales	*Enterococcaceae*	*Enterococcus*	alanine	-0.105	**4.30E-03**
Pasteurellales	*Pasteurellaceae*	*Haemophilus*	alanine	-0.281	**1.97E-03**
Pasteurellales	*Pasteurellaceae*	*Haemophilus*	aspartate	-0.158	**1.97E-03**
Pasteurellales	*Pasteurellaceae*	*Haemophilus*	glutamate	-0.222	**1.97E-03**
Pasteurellales	*Pasteurellaceae*	*Haemophilus*	glycine	-0.236	**1.97E-03**
Pasteurellales	*Pasteurellaceae*	*Haemophilus*	isoleucine	-0.258	**1.97E-03**
Pasteurellales	*Pasteurellaceae*	*Haemophilus*	leucine	-0.275	**1.97E-03**
Pasteurellales	*Pasteurellaceae*	*Haemophilus*	methionine	-0.217	**1.97E-03**
Pasteurellales	*Pasteurellaceae*	*Haemophilus*	PAG	-0.152	**1.97E-03**
Pasteurellales	*Pasteurellaceae*	*Haemophilus*	PCS	-0.186	**1.97E-03**
Pasteurellales	*Pasteurellaceae*	*Haemophilus*	phenylalanine	-0.227	**1.97E-03**
Pasteurellales	*Pasteurellaceae*	*Haemophilus*	tryptophan	-0.148	**1.97E-03**
Pasteurellales	*Pasteurellaceae*	*Haemophilus*	tyrosine	-0.238	**1.97E-03**
Pasteurellales	*Pasteurellaceae*	*Haemophilus*	uracil	-0.110	**6.58E-03**
Pasteurellales	*Pasteurellaceae*	*Haemophilus*	valine	-0.268	**1.97E-03**
Selenomonadales	*Veillonellaceae*	*Veillonella*	alanine	-0.171	**1.97E-03**
Selenomonadales	*Veillonellaceae*	*Veillonella*	glutamate	-0.150	**1.97E-03**
Selenomonadales	*Veillonellaceae*	*Veillonella*	glycine	-0.145	**1.97E-03**
Selenomonadales	*Veillonellaceae*	*Veillonella*	isoleucine	-0.149	**1.97E-03**
Selenomonadales	*Veillonellaceae*	*Veillonella*	leucine	-0.160	**1.97E-03**
Selenomonadales	*Veillonellaceae*	*Veillonella*	methionine	-0.126	**1.97E-03**
Selenomonadales	*Veillonellaceae*	*Veillonella*	PCS	-0.112	**1.97E-03**

Selenomonadales	*Veillonellaceae*	*Veillonella*	phenylalanine	-0.120	**1.97E-03**
Selenomonadales	*Veillonellaceae*	*Veillonella*	tyrosine	-0.131	**1.97E-03**
Selenomonadales	*Veillonellaceae*	*Veillonella*	uracil	-0.108	**4.30E-03**
Selenomonadales	*Veillonellaceae*	*Veillonella*	valine	-0.151	**1.97E-03**
Verrucomicrobiales	*Verrucomicrobiaceae*	*Akkermansia*	uracil	-0.130	**5.50E-03**

Appendix C - R packages and package versions

Chapter 3

Session Info

```
## R version 3.5.0 (2018-04-23)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS  10.15.1
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_GB.UTF-8/en_GB.UTF-8/en_GB.UTF-8/C/en_GB.UTF-8/en_GB.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] dplyr_0.8.0.1   igraph_1.2.4    ppcor_1.1       rmcorr_0.3.0
##  [5] santaR_1.0      pbkrtest_0.4-7  jtools_2.0.1    lme4_1.1-21
##  [9] Matrix_1.2-17   mixOmics_6.6.2  lattice_0.20-38 MASS_7.3-51.4
## [13] ropls_1.14.1    plotly_4.8.0    ggpubr_0.2      magrittr_1.5
## [17] ggplot2_3.2.1   phyloseq_1.26.1 tidyr_0.8.3     stringr_1.4.0
## [21] readr_1.3.1
##
## loaded via a namespace (and not attached):
##  [1] nlme_3.1-138        matrixStats_0.54.0  doParallel_1.0.14
##  [4] RColorBrewer_1.1-2  httr_1.4.0          tools_3.5.0
##  [7] R6_2.4.0            vegan_2.5-4         lazyeval_0.2.2
## [10] BiocGenerics_0.28.0 mgcv_1.8-28         colorspace_1.4-1
## [13] permute_0.9-5       ade4_1.7-13         withr_2.1.2
## [16] tidyselect_0.2.5    gridExtra_2.3       compiler_3.5.0
## [19] Biobase_2.42.0      labeling_0.3        scales_1.0.0
## [22] digest_0.6.18       minqa_1.2.4         rmarkdown_1.12
## [25] XVector_0.22.0      pkgconfig_2.0.2     htmltools_0.3.6
## [28] htmlwidgets_1.3     rlang_0.3.4         shiny_1.3.0
## [31] jsonlite_1.6        crosstalk_1.0.0     biomformat_1.10.1
## [34] Rcpp_1.0.1          munsell_0.5.0       S4Vectors_0.20.1
## [37] Rhdf5lib_1.4.3      ape_5.3             stringi_1.4.3
## [40] yaml_2.2.0          zlibbioc_1.28.0     rhdf5_2.26.2
## [43] plyr_1.8.4          grid_3.5.0          ggrepel_0.8.0
## [46] promises_1.0.1      parallel_3.5.0      crayon_1.3.4
## [49] Biostrings_2.50.2   splines_3.5.0       multtest_2.38.0
## [52] pander_0.6.3        hms_0.4.2           knitr_1.22
## [55] pillar_1.3.1        boot_1.3-20         ggsignif_0.5.0
## [58] corpcor_1.6.9       reshape2_1.4.3      codetools_0.2-16
## [61] stats4_3.5.0        glue_1.3.1          evaluate_0.13
## [64] data.table_1.12.2   nloptr_1.2.1        httpuv_1.5.1
## [67] foreach_1.4.4       gtable_0.3.0        purrr_0.3.2
## [70] assertthat_0.2.1    xfun_0.6            mime_0.6
## [73] xtable_1.8-3        RSpectra_0.14-0     later_0.8.0
## [76] survival_2.44-1.1   viridisLite_0.3.0   rARPACK_0.11-0
## [79] tibble_2.1.1        shinythemes_1.1.2   iterators_1.0.10
## [82] IRanges_2.16.0      ellipse_0.4.1       cluster_2.0.8
```

Chapter 4

Session Info

```
## R version 3.5.0 (2018-04-23)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS  10.15.1
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_GB.UTF-8/en_GB.UTF-8/en_GB.UTF-8/C/en_GB.UTF-8/en_GB.UTF-8
##
## attached base packages:
## [1] parallel  stats4    stats     graphics  grDevices utils     datasets
## [8] methods   base
##
## other attached packages:
##  [1] forcats_0.5.0               stringr_1.4.0
##  [3] dplyr_0.8.5                 purrr_0.3.3
##  [5] readr_1.3.1                 tidyr_1.0.2
##  [7] tibble_2.1.3                tidyverse_1.3.0
##  [9] plotly_4.9.2.1              reshape2_1.4.3
## [11] ppcor_1.1                   igraph_1.2.5
## [13] Hmisc_4.3-1                 Formula_1.2-3
## [15] survival_3.1-11             ropls_1.14.1
## [17] mixOmics_6.6.2              MASS_7.3-51.5
## [19] DESeq2_1.22.2               SummarizedExperiment_1.12.0
## [21] DelayedArray_0.8.0          BiocParallel_1.16.6
## [23] matrixStats_0.56.0          Biobase_2.42.0
## [25] GenomicRanges_1.34.0        GenomeInfoDb_1.18.2
## [27] IRanges_2.16.0              S4Vectors_0.20.1
## [29] BiocGenerics_0.28.0         vegan_2.5-6
```

```
## [31] lattice_0.20-40             permute_0.9-5
## [33] ggpubr_0.2.5                magrittr_1.5
## [35] ggplot2_3.3.0               phyloseq_1.26.1
## [37] santaR_1.0
##
## loaded via a namespace (and not attached):
##   [1] readxl_1.3.1           backports_1.1.5        fastmatch_1.1-0
##   [4] plyr_1.8.6             lazyeval_0.2.2         splines_3.5.0
##   [7] digest_0.6.25          foreach_1.5.0          htmltools_0.4.0
##  [10] fansi_0.4.1            checkmate_2.0.0        memoise_1.1.0
##  [13] cluster_2.1.0          doParallel_1.0.15      DECIPHER_2.10.2
##  [16] Biostrings_2.50.2      annotate_1.60.1        modelr_0.1.6
##  [19] rARPACK_0.11-0         colorspace_1.4-1       rvest_0.3.5
##  [22] blob_1.2.1             haven_2.2.0            xfun_0.12
##  [25] crayon_1.3.4           RCurl_1.98-1.1         jsonlite_1.6.1
##  [28] genefilter_1.64.0      phangorn_2.5.5         iterators_1.0.12
##  [31] ape_5.3                glue_1.3.2             gtable_0.3.0
##  [34] zlibbioc_1.28.0        XVector_0.22.0         Rhdf5lib_1.4.3
##  [37] scales_1.1.0           DBI_1.1.0              Rcpp_1.0.4
##  [40] viridisLite_0.3.0      xtable_1.8-4           htmlTable_1.13.3
##  [43] tidytree_0.3.3         foreign_0.8-76         bit_1.1-15.2
##  [46] htmlwidgets_1.5.1      httr_1.4.1             RColorBrewer_1.1-2
##  [49] ellipsis_0.3.0         acepack_1.4.1          farver_2.0.3
##  [52] pkgconfig_2.0.3        XML_3.99-0.3           nnet_7.3-13
```

```
##  [55] dbplyr_1.4.2           locfit_1.5-9.1         labeling_0.3
##  [58] tidyselect_1.0.0       rlang_0.4.5            later_1.0.0
##  [61] AnnotationDbi_1.44.0   cellranger_1.1.0       munsell_0.5.0
##  [64] tools_3.5.0            cli_2.0.2              generics_0.0.2
##  [67] RSQLite_2.2.0          ade4_1.7-15            broom_0.5.5
##  [70] evaluate_0.14          biomformat_1.10.1      fastmap_1.0.1
##  [73] yaml_2.2.1             ggtree_1.14.6          fs_1.3.2
##  [76] knitr_1.28             bit64_0.9-7            nlme_3.1-145
##  [79] mime_0.9               xml2_1.2.5             compiler_3.5.0
##  [82] shinythemes_1.1.2      rstudioapi_0.11        ggsignif_0.6.0
##  [85] treeio_1.6.2           reprex_0.3.0           geneplotter_1.60.0
##  [88] stringi_1.4.6          RSpectra_0.16-0        Matrix_1.2-18
##  [91] multtest_2.38.0        vctrs_0.2.4            pillar_1.4.3
##  [94] lifecycle_0.2.0        BiocManager_1.30.10    cowplot_1.0.0
##  [97] data.table_1.12.8      bitops_1.0-6           corpcor_1.6.9
## [100] httpuv_1.5.2           R6_2.4.1               latticeExtra_0.6-28
## [103] promises_1.1.0         gridExtra_2.3          codetools_0.2-16
## [106] assertthat_0.2.1       rhdf5_2.26.2           withr_2.1.2
## [109] GenomeInfoDbData_1.2.0 mgcv_1.8-31            hms_0.5.3
## [112] quadprog_1.5-8         grid_3.5.0             rpart_4.1-15
## [115] rvcheck_0.1.8          rmarkdown_2.1          shiny_1.4.0.2
## [118] lubridate_1.7.4        base64enc_0.1-3        ellipse_0.4.1
```

Chapter 5

Session Info

```
## R version 3.5.0 (2018-04-23)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS  10.15.1
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_GB.UTF-8/en_GB.UTF-8/en_GB.UTF-8/C/en_GB.UTF-8/en_GB.UTF-8
##
## attached base packages:
## [1] parallel  stats4    stats     graphics  grDevices utils     datasets
## [8] methods   base
##
## other attached packages:
##  [1] rfUtilities_2.1-5           caret_6.0-86
##  [3] randomForest_4.6-14         forcats_0.5.0
##  [5] stringr_1.4.0               dplyr_0.8.5
##  [7] purrr_0.3.3                 readr_1.3.1
##  [9] tidyr_1.0.2                 tibble_2.1.3
## [11] tidyverse_1.3.0             pbkrtest_0.4-7
## [13] jtools_2.0.3                lme4_1.1-21
## [15] Matrix_1.2-18               plotly_4.9.2.1
## [17] reshape2_1.4.3              ppcor_1.1
## [19] igraph_1.2.5                Hmisc_4.3-1
## [21] Formula_1.2-3               survival_3.1-11
## [23] mixOmics_6.6.2              MASS_7.3-51.5
## [25] DESeq2_1.22.2               SummarizedExperiment_1.12.0
## [27] DelayedArray_0.8.0          BiocParallel_1.16.6
## [29] matrixStats_0.56.0          Biobase_2.42.0
## [31] GenomicRanges_1.34.0        GenomeInfoDb_1.18.2
## [33] IRanges_2.16.0              S4Vectors_0.20.1
## [35] BiocGenerics_0.28.0         vegan_2.5-6
## [37] lattice_0.20-40             permute_0.9-5
## [39] networkD3_0.4               ggpubr_0.2.5
## [41] magrittr_1.5                ggplot2_3.3.0
## [43] phyloseq_1.26.1             santaR_1.0
```

```
##
## loaded via a namespace (and not attached):
##   [1] ggstance_0.3.4         mQTL_1.0               tidyselect_1.0.0
##   [4] RSQLite_2.2.0          AnnotationDbi_1.44.0   htmlwidgets_1.5.1
##   [7] grid_3.5.0             pROC_1.16.2            munsell_0.5.0
##  [10] codetools_0.2-16       withr_2.1.2            colorspace_1.4-1
##  [13] qtl_1.46-2             knitr_1.28             rstudioapi_0.11
##  [16] ggsignif_0.6.0         labeling_0.3           GenomeInfoDbData_1.2.0
##  [19] bit64_0.9-7            farver_2.0.3           rhdf5_2.26.2
##  [22] coda_0.19-3            vctrs_0.2.4            generics_0.0.2
##  [25] ipred_0.9-9            xfun_0.12              itertools_0.1-3
##  [28] R6_2.4.1               doParallel_1.0.15      locfit_1.5-9.1
##  [31] RcppZiggurat_0.1.5     bitops_1.0-6           assertthat_0.2.1
##  [34] promises_1.1.0         scales_1.1.0           nnet_7.3-13
##  [37] gtable_0.3.0           timeDate_3043.102      rlang_0.4.5
##  [40] genefilter_1.64.0      splines_3.5.0          TMB_1.7.16
##  [43] lazyeval_0.2.2         impute_1.56.0          ModelMetrics_1.2.2.2
##  [46] acepack_1.4.1          broom_0.5.5            checkmate_2.0.0
##  [49] yaml_2.2.1             modelr_0.1.6           crosstalk_1.1.0.1
##  [52] backports_1.1.5        Rfast_1.9.9            httpuv_1.5.2
##  [55] MassSpecWavelet_1.48.1 tools_3.5.0            lava_1.6.7
##  [58] ellipsis_0.3.0         biomformat_1.10.1      RColorBrewer_1.1-2
##  [61] Rcpp_1.0.4             plyr_1.8.6             base64enc_0.1-3
##  [64] zlibbioc_1.28.0        RCurl_1.98-1.1         rpart_4.1-15
##  [67] cowplot_1.0.0          haven_2.2.0            ggrepel_0.8.2
```

```
##  [70] cluster_2.1.0          colorRamps_2.3         fs_1.3.2
##  [73] data.table_1.12.8      RSpectra_0.16-0        reprex_0.3.0
##  [76] pcaMethods_1.74.0      hms_0.5.3              mime_0.9
##  [79] evaluate_0.14          xtable_1.8-4           XML_3.99-0.3
##  [82] broom.mixed_0.2.4      readxl_1.3.1           gridExtra_2.3
##  [85] compiler_3.5.0         ptw_1.9-15             ellipse_0.4.1
##  [88] crayon_1.3.4           minqa_1.2.4            htmltools_0.4.0
##  [91] mgcv_1.8-31            corpcor_1.6.9          later_1.0.0
##  [94] snow_0.4-3             geneplotter_1.60.0     lubridate_1.7.4
##  [97] DBI_1.1.0              dbplyr_1.4.2           boot_1.3-24
## [100] ade4_1.7-15            cli_2.0.2              gower_0.2.1
## [103] pkgconfig_2.0.3        foreign_0.8-76         recipes_0.1.10
## [106] xml2_1.2.5             foreach_1.5.0          rARPACK_0.11-0
## [109] annotate_1.60.1        missForest_1.4         multtest_2.38.0
## [112] XVector_0.22.0         speaq_2.6.1            prodlim_2019.11.13
## [115] rvest_0.3.5            digest_0.6.25          Biostrings_2.50.2
## [118] rmarkdown_2.1          cellranger_1.1.0       htmlTable_1.13.3
## [121] shiny_1.4.0.2          nloptr_1.2.2.1         outliers_0.14
## [124] lifecycle_0.2.0        nlme_3.1-145           jsonlite_1.6.1
## [127] Rhdf5lib_1.4.3         viridisLite_0.3.0      fansi_0.4.1
## [130] pillar_1.4.3           fastmap_1.0.1          httr_1.4.1
## [133] glue_1.3.2             shinythemes_1.1.2      iterators_1.0.12
## [136] pander_0.6.3           bit_1.1-15.2           class_7.3-15
## [139] stringi_1.4.6          blob_1.2.1             doSNOW_1.0.18
## [142] latticeExtra_0.6-28    memoise_1.1.0          e1071_1.7-3
## [145] ape_5.3
```

Appendix D - Publication

This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

OPEN ACCESS

ORIGINAL ARTICLE

Longitudinal metabolic and gut bacterial profiling of pregnant women with previous bariatric surgery

Kiana Ashley West ,[1] Chidimma Kanu,[2] Tanya Maric,[2] Julie Anne Kathryn McDonald ,[1] Jeremy K Nicholson,[3] Jia V Li,[1] Mark R Johnson,[2] Elaine Holmes,[1,3] Makrina D Savvidou[2]

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/gutjnl-2019-319620).

[1]Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, Imperial College London, London, UK
[2]Obstetrics & Gynaecology, Chelsea and Westminster Hospital, Institute of Reproductive Developmental Biology, Department of Metabolism, Digestion & Reproduction, Faculty of Medicine, Imperial College London, London, UK
[3]Australian National Phenome Centre, Health Futures Institute, Murdoch University, Murdoch, Western Australia, Australia

Correspondence to
Elaine Holmes, Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, Imperial College London, London SW7 2AZ, UK; elaine.holmes@imperial.ac.uk

KAW and CK contributed equally.

Received 8 August 2019
Revised 3 December 2019
Accepted 4 December 2019

Check for updates

© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: West KA, Kanu C, Maric T, *et al*. *Gut* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/gutjnl-2019-319620

ABSTRACT

Objective Due to the global increase in obesity rates and success of bariatric surgery in weight reduction, an increasing number of women now present pregnant with a previous bariatric procedure. This study investigates the extent of bariatric-associated metabolic and gut microbial alterations during pregnancy and their impact on fetal development.

Design A parallel metabonomic (molecular phenotyping based on proton nuclear magnetic resonance spectroscopy) and gut bacterial (16S ribosomal RNA gene amplicon sequencing) profiling approach was used to determine maternal longitudinal phenotypes associated with malabsorptive/mixed (n=25) or restrictive (n=16) procedures, compared with women with similar early pregnancy body mass index but without bariatric surgery (n=70). Metabolic profiles of offspring at birth were also analysed.

Results Previous malabsorptive, but not restrictive, procedures induced significant changes in maternal metabolic pathways involving branched-chain and aromatic amino acids with decreased circulation of leucine, isoleucine and isobutyrate, increased excretion of microbial-associated metabolites of protein putrefaction (phenylacetlyglutamine, *p*-cresol sulfate, indoxyl sulfate and *p*-hydroxyphenylacetate), and a shift in the gut microbiota. The urinary concentration of phenylacetylglutamine was significantly elevated in malabsorptive patients relative to controls (p=0.001) and was also elevated in urine of neonates born from these mothers (p=0.021). Furthermore, the maternal metabolic changes induced by malabsorptive surgery were associated with reduced maternal insulin resistance and fetal/birth weight.

Conclusion Metabolism is altered in pregnant women with a previous malabsorptive bariatric surgery. These alterations may be beneficial for maternal outcomes, but the effect of elevated levels of phenolic and indolic compounds on fetal and infant health should be investigated further.

Significance of this study

What is already known on this subject?
- ► Bariatric surgery, particularly the malabsorptive type, alters the gut microbiota and host metabolism.
- ► Pregnancy following bariatric surgery, especially a malabsorptive procedure, is associated with a reduced prevalence of maternal gestational diabetes and increased prevalence of small for gestational age babies.

What are the new findings?
- ► We demonstrated that malabsorptive surgery-induced changes in the maternal gut microbiota persist throughout pregnancy compared with pregnancies with similar maternal body mass index but no history of such surgery, and may have a trans-generational impact.
- ► Gut microbial changes in malabsorptive patients are associated with increased excretion of protein putrefaction metabolites during pregnancy.

How might it impact on clinical practice in the foreseeable future?
- ► The impact of maternal malabsorption on fetal health warrants further investigation to determine the most appropriate type of weight loss surgery for women of reproductive age.

INTRODUCTION

Currently more than a third of women are classified as overweight or obese[1] with obesity reaching epidemic levels globally. Obesity is associated with a number of adverse metabolic effects resulting in cardiovascular disease, metabolic syndrome, diabetes and cancer.[2] Bariatric surgery (BS) has proven a successful treatment modality for lasting weight loss and has been shown to reduce obesity-related morbidity and mortality.[3 4] Altered bile flow, reduction in gastric capacity, anatomical gut rearrangement and altered nutrient flow, vagal manipulation and modulation of enteric gut hormones, collectively referred to as the BRAVE effects,[3 5] have been proposed to contribute to the mechanisms of weight loss and diabetes resolution in bariatric patients. There are two main types of BS: the restrictive (RES) procedures such as gastric banding and sleeve gastrectomy, which reduce the stomach size, and the malabsorptive/mixed (MAL) type, such as bilio-pancreatic diversion and Roux-en-Y-gastric bypass, which aim to decrease caloric absorption.[6] BS has been shown to impact metabolic profiles reflecting altered metabolism and changes in the gut microbiome that play a role in improved lipid and glucose metabolism.[7–11] The contribution of the gut mictobiota to host metabolism and

Gut: first published as 10.1136/gutjnl-2019-319620 on 21 January 2020. Downloaded from http://gut.bmj.com/ on May 27, 2020 at BVA. Protected by copyright.

metabolic control is well-established. Microbes degrade dietary components that are undigested by the host, convert dietary molecules into bioactive metabolites, influence host response to drugs and participate in many homeostatic processes and signalling pathways.[12 13] However, the mechanisms by which BS confers metabolic benefit and the potential role of the microbiome in achieving this are poorly understood.

Evidence continues to accumulate that pregnancy post-BS is associated with lower prevalence of gestational diabetes mellitus (GDM) and large for gestational age neonates but higher risk of small for gestational age neonates and late preterm delivery compared with pregnancies in obese women without a previous BS.[14] To investigate the impact of previous BS on maternal phenotype during pregnancy, we adopted a longitudinal metabolic (serum and urine) and gut bacterial (faeces) profiling strategy using proton nuclear magnetic resonance (^{1}H NMR) spectroscopy and 16S rRNA gene amplicon sequencing, respectively, to characterise women with previous BS (RES and MAL) throughout the course of their pregnancy. Profiles were compared against women with similar early pregnancy body mass index (BMI) and no history of BS (NBS). For a subset of the cohort, we also characterise neonatal cord serum and urine metabolomes to assess potential transgenerational metabolic effects of BS.

METHODS

Study population, sampling and clinical data

The population is part of an ongoing prospective cohort study investigating the impact of maternal BS on perinatal outcomes. All women gave written, informed consent for their data and samples to be used. Pregnant women with (n=47) and without (n=118) previous BS were recruited from May 2015 to April 2017 at Chelsea and Westminster Hospital (London, UK) as previously described.[15] Women were seen at five time points during pregnancy (**T1**: 11^{+0}–14^{+0}, **T2**: 20^{+0}–24^{+0}, **T3**: 28^{+0}–30^{+0}, **T4**: 30^{+0}–33^{+0} and **T5**: 35^{+0}–37^{+6} weeks gestation) and within 72 hours of delivery (**T6**). Maternal blood (serum) and urine samples were collected at each visit while faecal samples were requested at the T1, T2 and T4 visits (see online supplementary figure S1 and table S1). A full oral glucose tolerance test (2 hour, 75 g) was conducted at T3 and maternal insulin resistance was calculated using the homeostatic model assessment for insulin resistance (HOMA-IR=fasting serum insulin (μU/L) × fasting glucose (mmol/L)/22.5). Estimated fetal weight was calculated by trans-abdominal ultrasound scans at T2, T4 and T5. At T6, birth weight was recorded, percentiles for the gestation were calculated and, where possible, neonatal samples (cord serum and urine) were collected. All samples were stored at −80°C for future analysis. For the study population, women with diagnosis of type 2 diabetes mellitus or GDM (due to the effect of diabetes on the metabolic profile) and those that had a miscarriage were excluded. Only NBS participants with a BMI of 25–50 kg/m^2 at T1 were included to match the BMI range of the included bariatric patients at T1.

Metabolic profiling of biofluid samples

Serum and urine samples were prepared according to an established protocol.[16] ^{1}H NMR spectra were acquired on a Bruker 600 MHz spectrometer (Bruker BioSpin) following a published method[16] (see online supplementary methods for further detail). Methodology for preprocessing of spectral data is described in online supplementary methods. Multivariate modelling of the spectral data was performed in the software package SIMCA V.14.1 (Sartorius Stedim Biotech). Principal component analysis (PCA) was used to assess variation in metabolic profiles over all time points and to identify extreme outliers to exclude from supervised models for each time point. Orthogonal partial least squares discriminant analysis (OPLS-DA) was used to identify spectral variables (relating to specific chemical compounds) that contributed to discrimination of clinical classes (pairwise comparisons between NBS, MAL and RES groups) at each time point. From the discriminatory NMR peaks, metabolite identities were confirmed using statistical tools,[17 18] 2D NMR experiments and by spiking in authentic standards (see online supplementary methods). Relative concentrations of each discriminatory metabolite were calculated by integrating a representative peak of that metabolite. Downstream analysis was performed in the R software environment.[19] For each discriminatory metabolite identified in the cross-sectional analysis, time series curves were generated through a spline-fitting method implemented in 'santaR' (see online supplementary methods) to visualise their behaviour in different groups over the pregnancy time course. Individuals with at least five data points were included. Correlation analysis including partial correlation to adjust for confounders is detailed in the online supplementary methods along with a list of all R packages used. Missing data were excluded from calculations. All p values were adjusted (p_{adj}) where necessary to control for the false discovery rate according to the Benjamini-Hochberg method.[20] An alpha of 0.05 was used for p and p_{adj} values.

Gut bacterial community profiling

Stool samples were randomised for processing and DNA was extracted (see online supplementary methods) using the PowerLyzer PowerSoil DNA Isolation Kit (Mo Bio). 16S rRNA gene amplicon sequencing targeting the V1-V2 regions was performed on the Illumina MiSeq platform as previously described.[21] Raw reads were processed in the R software environment[19] following a published workflow[22] which includes amplicon denoising implemented in 'DADA2'.[23] See online supplementary methods for full details. Functions in the 'vegan' R package were used to calculate Shannon Diversity Indices (α-diversity) on data rarefied to the minimum sequencing depth and Bray-Curtis dissimilarity (β-diversity) on log-transformed data (pseudocount of 1 added to each value). Significance of group separation in β-diversity was assessed by permutational multivariate analysis of variance. Changes in relative abundance were tested at each taxonomic rank from phylum to genus using the Mann-Whitney U test while differentially abundant 16S rRNA gene sequences were identified using 'DESeq2'.[24] For 'DESeq2' analysis, data were pooled for each individual rather than analysing distinct time points.

Integrative analysis of metabolic and taxonomic data

Relationships between the serum, urine and faecal data sets were modelled using the DIABLO method in 'mixOmics'.[25] This is a multi-block latent variable-based approach which aims to identify concordance between multiple data sets. Metabolites significant throughout the time course (serum: leucine, isoleucine, isobutyrate, D-β-hydroxybutyrate; urine: phenylacetylglutamine (PAG), *p*-cresol sulfate (PCS), indoxyl sulfate (IS), *p*-hydroxyphenylacetate (PHPA), unknown, α-ketoisovalerate, creatinine) and a subset of bacterial genera (log-transformed; selected using the least absolute shrinkage and selection operator (LASSO) penalisation method implemented in 'mixOmics') were modelled. Sampling points for each individual, where matching microbiome and metabolite data were obtained (T1, T2 and T4), were included in the model.

Gut: first published as 10.1136/gutjnl-2019-319620 on 21 January 2020. Downloaded from http://gut.bmj.com/ on May 27, 2020 at BVA. Protected by copyright.

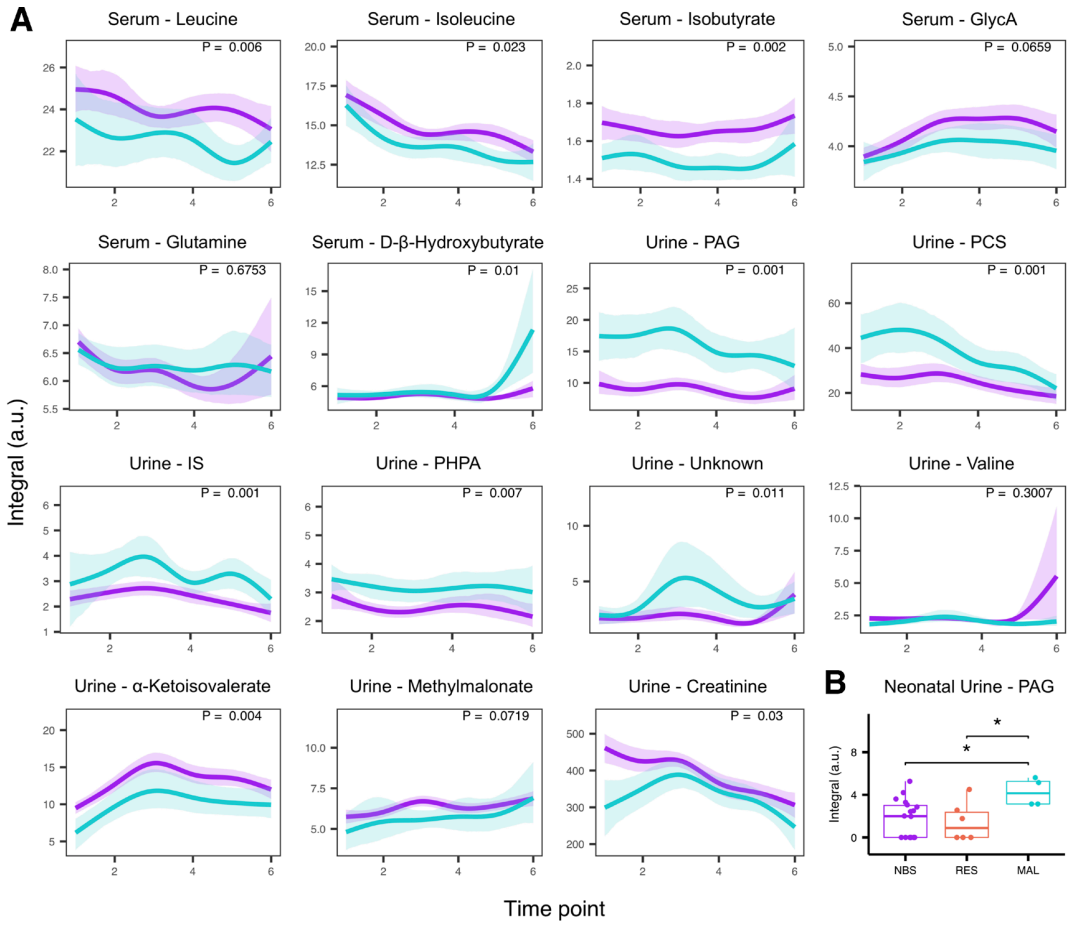


Figure 1 Modelling of metabolite concentrations. (A) Time series analysis of discriminatory metabolites with lines representing group mean curves while shaded areas denote 95% confidence bands. Malabsorptive (MAL, blue) patients relative to No bariatric surgery (NBS, purple) controls. (B) Relative concentration of phenylacetylglutamine (PAG) is increased in neonatal urine from the MAL group compared with the NBS or restrictive (RES) groups. NMR peak integrals are reported in arbitrary units (a.u.). *P<0.05 (Mann-Whitney U test); GlycA, *N*-acetyl glycoprotein; IS, indoxyl sulfate; PCS, *p*-cresol sulfate; PHPA, *p*-hydroxyphenylacetate.

RESULTS

Characteristics of the study participants

Of the 165 women enrolled in the study, 54 were excluded from longitudinal profiling due to BMI (n=36), diagnosis of diabetes (n=16), miscarriage (n=1) or withdrawal from the study (n=1). The final study population included 111 women (NBS n=70, RES n=16, MAL n=25) who contributed 491 maternal urine, 394 maternal blood, 68 maternal stool (online supplementary figure S1), 54 cord blood and 28 neonatal urine samples (online supplementary table S1). Two women (RES n=1, MAL n=1) contributed only neonatal samples. All women in the MAL group had a previous gastric bypass while the RES group had gastric banding (n=8) or sleeve gastrectomy (n=8). Maternal demographics, pregnancy characteristics and outcomes of the study population (online supplementary table S2) were consistent with our earlier report.[15] Compared with the NBS group, women with a previous BS were on average 3.7 years older and, especially those with a MAL procedure, had lower insulin resistance at T3 and delivered smaller babies earlier (online supplementary table S2). Among the BS group, women with a previous MAL procedure had a higher pre-surgery BMI compared with those with a previous RES procedure.

Metabolic differences in maternal and neonatal serum metabolomes associated with bariatric surgery

Clear differences between the serum metabolic profiles of MAL, but not RES, participants and NBS participants were detected in the third trimester of pregnancy (time points T4 and T5; online supplementary table S3) as evidenced by the OPLS-DA models. The serum metabolic profiles associated with BS at the time of labour and in the first and second trimesters could not be differentiated from the control group. Serum samples obtained from MAL patients contained lower concentrations of unsaturated lipids (CH=CH, δ 5.31 (m)) at T5 and the lipids group corresponding to $C{=}CCH_2C{=}C$ (δ 2.79 (m)) at both T4 and T5 (online supplementary figure S2) compared with the NBS group. Other metabolic differences between the MAL and NBS groups included (online supplementary table S4) *N*-acetyl glycoprotein (GlycA), leucine, isoleucine, isobutyrate (decreased in MAL) and glutamine, D-β-hydroxybutyrate (increased in MAL). The behaviour of these metabolites throughout the pregnancy period was investigated further using time series analysis to model individual metabolites. Serum leucine, isoleucine and isobutyrate concentrations were significantly lower in MAL (n=12; NBS n=34) throughout the time course (figure 1A), whereas increases in D-β-hydroxybutyrate were associated with the last few weeks of pregnancy and delivery. GlycA and glutamine were

Gut: first published as 10.1136/gutjnl-2019-319620 on 21 January 2020. Downloaded from http://gut.bmj.com/ on May 27, 2020 at BVA. Protected by copyright.

not significantly changed in MAL during pregnancy. Significance of the differential serum metabolite concentrations at each time point is shown in online supplementary figure S3A. There was no significant difference between NBS and RES participants at any time point (online supplementary table S5). Although OPLS-DA and time series modelling showed clear evidence that previous BS impacted on the metabolic profiles, PCA indicated that the greatest metabolic variation in plasma composition was due to the changes over time during pregnancy, irrespective of surgery, with samples collected at T1 being markedly different from those obtained at later time points in the pregnancy (online supplementary figure S4). Thus no inherent structure associated with study group, age or BMI was evident in the ^{1}H NMR maternal serum data. No differences in the metabolic profiles of cord serum from neonates in each study group were detected (online supplementary table S3).

Surgery-associated metabolic differences in maternal and neonatal urine metabolomes

Similar to the serum samples, PCA models indicated that the metabolic variation driven by temporal changes throughout pregnancy was stronger than that associated with study group, age or BMI (online supplementary figure S4). Urine samples collected at delivery (T6) clearly differed from those taken at any other time point during pregnancy. In addition to endogenous changes, drug metabolites of paracetamol as well as mannitol, a common ingredient in drug tablet coating, were detected in urine obtained at T6. Outliers identified from PCA models at each time point were samples containing high concentrations of glucose or drug metabolites and were excluded from subsequent supervised models (online supplementary table S3). Based on the OPLS-DA models, no significant differences were identified between NBS and RES groups or between RES and MAL groups at any of the six time points; however, urine metabolic profiles of MAL patients were significantly altered compared with the NBS group as early as T1 and this difference persisted throughout pregnancy (online supplementary table S3). The strongest discriminatory metabolites (online supplementary table S4 and online supplementary figure S5) contributing to the differences between MAL and NBS groups were host-gut bacterial co-metabolites PAG, PCS, IS, PHPA, as well as an unidentified aromatic metabolite (unknown) with structural similarities to PCS (based on both ^{1}H and ^{13}C shifts; online supplementary figure S6). These metabolites were significantly elevated in MAL (n=14; NBS n=50) patients during pregnancy (figure 1A). Creatinine and α-ketoisovalerate were present in lower concentrations in MAL patients; methylmalonate and valine were not significantly associated with the MAL group over the time course (figure 1A) but were significantly lower in MAL patients at specific time points (online supplementary figure S3B). None of these metabolites were significantly changed in the RES group after correcting for multiple hypothesis tests (online supplementary table S5). Interestingly, when the RES group was divided into gastric banding (purely mechanical) and sleeve gastrectomy (removal of part of the stomach with metabolic effects), patients with a sleeve tended to have higher PAG and PCS than patients with a band (online supplementary figure S7). Maternal urine samples collected at T6 could not be discriminated for any pairwise comparison by OPLS-DA (online supplementary table S3). However PAG, the metabolite with the largest effect size in maternal samples, retained significance in the offspring's urine at T6 (figure 1B). The relative concentration of PAG was significantly increased in babies born from MAL mothers (n=4) compared with babies born from RES (n=6, p=0.04) and NBS (n=18, p=0.021) mothers, mirroring the difference observed between the MAL and NBS maternal samples.

Taxonomic changes in the maternal gut microbiota of MAL patients

To support the finding of altered microbiota-associated metabolites in MAL patients detected by ^{1}H NMR spectroscopy, faecal microbiome compositions from a subset of the cohort were analysed. A total of 68 stool samples representing 39 mothers (NBS n=25, MAL n=14) were studied. After excluding one maternal NBS sample due to low sequencing depth (2181 reads), the mean number of high-quality, paired-end 16S rRNA gene amplicon sequences per sample was 22 405 (±5535 SD) with a minimum sequencing depth of 12 967.

Women with a previous MAL surgery exhibited greater α-diversity, or a higher number of distinct taxa with more even distributions, compared with the NBS group but this was only statistically significant in the third trimester of pregnancy (figure 2A). β-diversity analysis comparing pairwise dissimilarities between samples revealed that MAL patients had a distinct (p=0.001) gut microbiota from the NBS group throughout pregnancy (figure 2B). MAL patients could be distinguished by an increase in relative abundance of *Escherichia*/*Shigella*, *Streptococcus* and *Enterococcus* genera and these changes manifested at all higher taxonomic ranks (figure 2C). Relative increases in abundances of *Rothia* and its family *Micrococcaceae* were also observed. *Anaerostipes* was the only genus found in significantly lower relative abundance in the MAL group at multiple time points. Only taxa that exhibited differential abundance in the MAL group at more than one time point were included in figure 2C, but taxa displaying differences between groups at a single time point during pregnancy are detailed in online supplementary table S6. Differential abundance at the amplicon sequence level was investigated for sequences assigned to the genera *Streptococcus*, *Enterococcus*, *Escherichia*/*Shigella* and *Rothia*. Species assignments were compared against SILVA[26] and RDB[27] databases. From the 25 16S rRNA gene sequences (*Escherichia*/*Shigella* n=11, *Streptococcus* n=12, *Enterococcus* n=1 and *Rothia* n=1) found in significantly higher abundance in MAL patients (online supplementary table S7), the following species were identified: *Escherichia coli*, *Streptococcus salivarius*, *Streptococcus vestibularis*, *Streptococcus mutans* and *Streptococcus parasanguinis*.

Integrative analysis of phenotypes

To determine if the observed MAL signatures were complementary across the data sets, an integrative approach was used to model the bacterial (n=68) and metabolic (serum n=60; urine n=66) data together. Clinical classes could be discriminated on the first component of the model for both urine and faecal data sets while this difference was not apparent in the serum data (figure 3A). Urine PAG, PCS, and IS concentrations were highly correlated and had the highest contribution to the model followed by leucine, isoleucine (also highly correlated) and isobutyrate (figure 3B). PAG, PCS and IS were most correlated with *Streptococcus*, *Enterococcus*, *Escherichia*/*Shigella*, *Rothia* and *Holdemanella* (figure 3C). Variables in this cluster also had a strong negative correlation with serum leucine, isoleucine and isobutyrate. *Anaerostipes*, the only genus that decreased in abundance in the MAL group, was negatively correlated with PAG, PCS and IS but positively correlated with leucine, isoleucine and isobutyrate (figure 3C).

Gut: first published as 10.1136/gutjnl-2019-319620 on 21 January 2020. Downloaded from http://gut.bmj.com/ on May 27, 2020 at BVA. Protected by copyright.

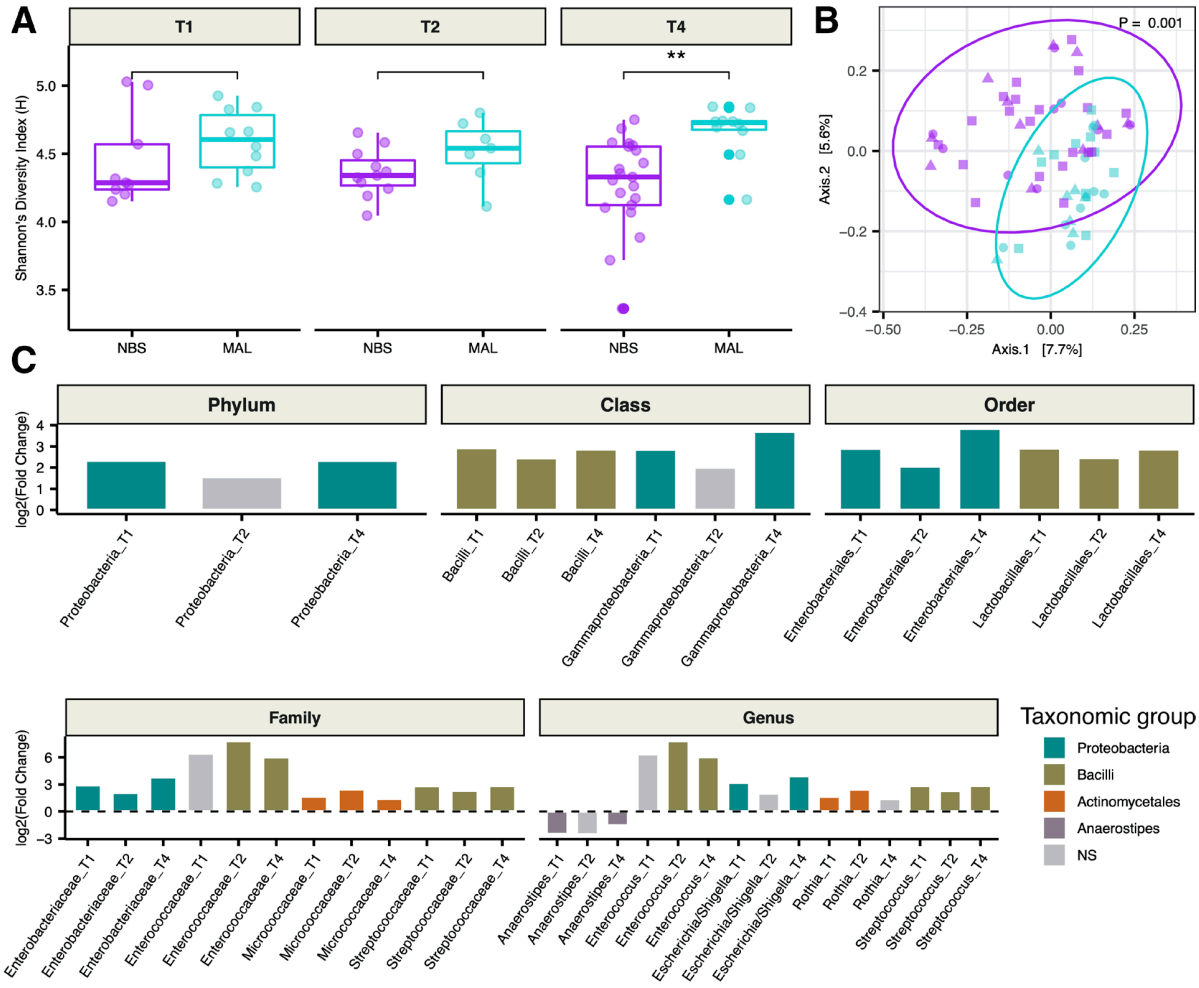


Figure 2 Gut bacterial alterations in malabsorptive patients. (A) Comparison of Shannon diversity between malabsorptive (MAL) and no bariatric surgery (NBS) groups at 11–14 weeks (T1), 20–24 weeks (T2) and 30–33 weeks (T4) gestation. (B) Principal coordinates analysis of Bray-Curtis dissimilarities between MAL (blue) and NBS (purple) during pregnancy (T1: circle, T2: triangle, T4: square). P value obtained from permutational multivariate analysis of variance test. (C) Taxa that exhibited differential abundance (Mann-Whitney U test; p_{adj}<0.05) at more than one time point are plotted. Values are reported in online supplementary table S6. **P<0.01 (Mann-Whitney U test); NS, not significant.

Metabolic profile and clinical outcomes

Spearman's correlation coefficients (ρ) were calculated to assess the relationships between relative concentrations of maternal metabolites and clinical measurements (maternal: fasting insulin, fasting glucose and HOMA-IR measured at T3; neonatal: estimated fetal weight (measured at T2, T4 and T5) or birth weight percentiles, and gestational age at delivery). Throughout pregnancy, urinary host-microbial co-metabolites PAG and PCS were negatively associated with maternal HOMA-IR and fasting insulin while the branched-chain amino acids (mainly at T2) had a positive correlation (figure 4). PAG and PCS relative concentrations at T2, IS and unknown at T5, and IS and PAG at T6 were also negatively correlated with the weight percentiles of the babies at the corresponding time points. Partial correlations controlling for the effect of maternal age and BMI on maternal measurements as well as the effect of maternal age, BMI and HOMA-IR on birthweight percentile are reported in online supplementary table S8, but these confounders did not alter the overall results. Importantly, the concentrations of these metabolites did not correlate (online supplementary figure S8) with percentage of weight lost or time interval between BS and conception (which ranged from 12 months to 11 years in our cohort) in MAL patients, suggesting that the phenotype is persistent.

DISCUSSION

Our results demonstrate that pregnancies following a MAL bariatric procedure are characterised by altered maternal and neonatal metabolic profiles compared with pregnancies without such surgery. In particular, women with previous MAL surgery had lower serum concentrations of branched-chain amino acids (leucine and isoleucine) and branched-chain fatty acids (isobutyrate) and excreted higher concentrations of urinary host-microbial co-metabolites of protein putrefaction (PAG, PCS, IS and PHPA). This signature was correlated with a shift in the gut microbiota which included increases in relative abundance of bacterial genera *Enterococcus*, *Streptococcus*, *Escherichia*/*Shigella* and *Rothia* as well as a decrease in *Anaerostipes*. The observed MAL-associated changes in metabolism were also inversely associated with maternal insulin resistance and the offspring's weight, suggesting that a previous MAL surgery may have both beneficial and detrimental effects on pregnancy. With regard to the RES group, we posit these patients, especially those with a previous sleeve gastrectomy, represent an intermediate phenotype due to the lower risk of malabsorption imposed by these procedures and less extreme changes to the gastrointestinal environment which would likely result in lesser alterations in the gut microbiome.[28]

Gut: first published as 10.1136/gutjnl-2019-319620 on 21 January 2020. Downloaded from http://gut.bmj.com/ on May 27, 2020 at BVA. Protected by copyright.

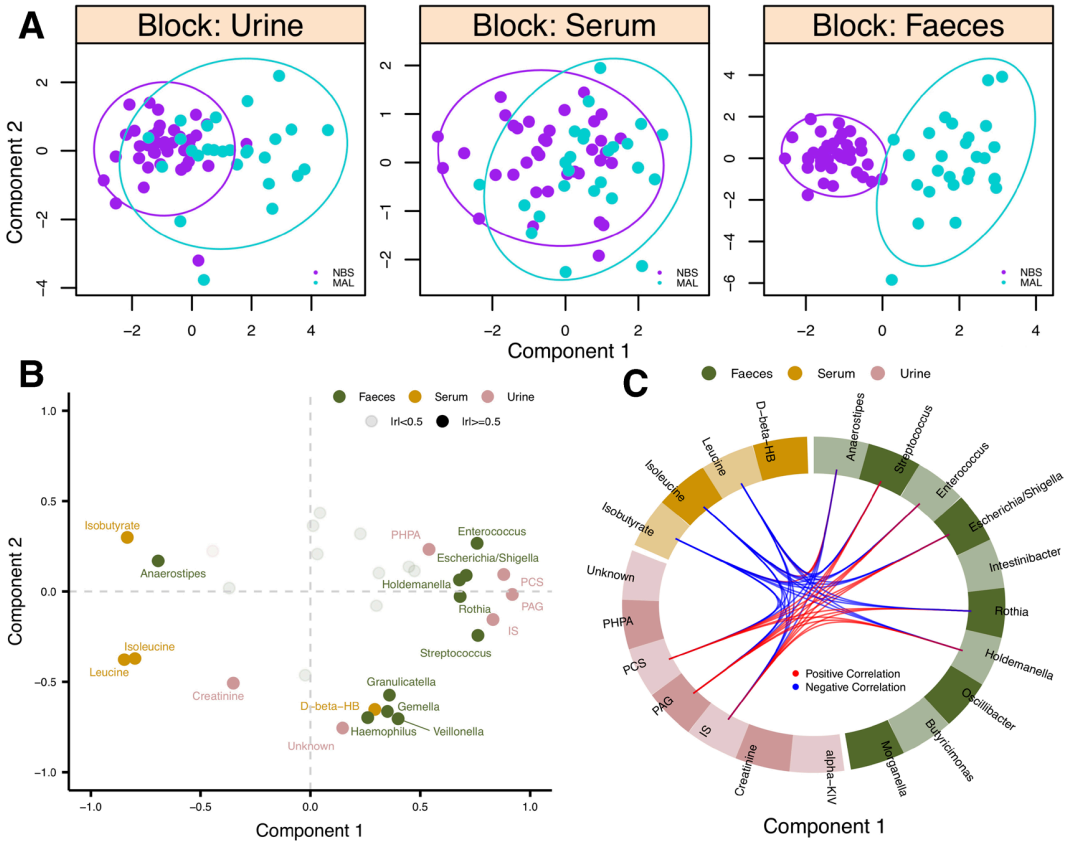


Figure 3 Integration of serum, urine and faecal profiles. (A) Scores plots showing samples from each data set projected in latent space. Clinical classes are discriminated along component 1. (B) Variables are plotted (coloured by biofluid) where each point represents the Pearson correlation between that variable and each component. High correlations with either component (|r|≥0.5) are BOLD and labelled. (C) Pearson correlations (|r|≥0.5) between variables in different data sets along component 1 are plotted. NBS, no bariatric surgery; MAL, malabsorptive; alpha-KIV, α-ketoisovalerate; D-beta-HB, D-β-hydroxybutyrate; IS, indoxyl sulfate; PAG, phenylacetylglutamine; PCS, *p*-cresol sulfate; PHPA, *p*-hydroxyphenylacetate.

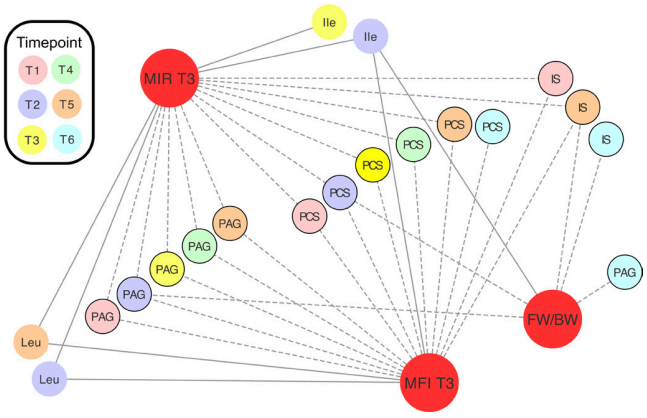


Figure 4 Relative concentrations of metabolites correlate with maternal insulin resistance and fetal/birth weight. Spearman's correlations between clinical measurements (red) and metabolite relative concentrations. Correlations plotted are significant (p_{adj}<0.05). Values are reported in online supplementary table S8. Negative correlations are denoted by dashed lines. Urine metabolites are denoted by black borders. FW/BW, estimated fetal weight or birth weight; Ile, isoleucine; IS, indoxyl sulfate; Leu, leucine; MFI, maternal fasting insulin; MIR, maternal HOMA-IR; PAG, phenylacetylglutamine; PCS, *p*-cresol sulfate.

Our results assessing the influence of BS on the maternal faecal microbiome composition are consistent with previous findings in non-pregnant individuals where MAL, but not RES, surgery resulted in increased relative abundances of *Bacilli* and *Gammaproteobacteria* (including genera *Enterococcus*, *Streptococcus* and *Escherichia*) as well as associated faecal metabolites[28]; abundances of both classes positively correlated with products of protein fermentation. Studies consistently show a gut bacterial signature of facultative anaerobes and oral-associated bacteria in BS (mainly MAL) patients with an increase in relative abundance of enterobacteria being the most consistent bacterial marker across multiple studies in both humans[7 8 10 28–30] and animals.[11 31] This shift in gut microbial community structure is believed to be, at least partially, a result of the increased oxygen and pH in the digestive tract following the anatomical rearrangement of a MAL surgery.[30 32] Enterobacteria are Gram-negative and generally present in very low densities (<<10^8 CFU/g) in the normal gut, but increased abundances of these bacteria, some of which are known pathogens, have been associated with both inflammatory bowel disease and colon cancer.[33] This underscores the need for understanding the mechanisms by which the altered bariatric microbiome can impact on maternal and neonatal metabolism and downstream health.

Increased concentrations of urinary microbial-associated metabolites following MAL BS have been reported in animals[11 34] and in one human study with only two bariatric patients.[9] Carbohydrates are the preferred source of energy for

Gut: first published as 10.1136/gutjnl-2019-319620 on 21 January 2020. Downloaded from http://gut.bmj.com/ on May 27, 2020 at BVA. Protected by copyright.

microbes inhabiting the large intestine.[35] However, we hypothesise that the MAL condition not only changes the gastrointestinal environment resulting in an altered microbial community but also results in more undigested protein reaching the large intestine[36] and both of these factors likely contribute to the shift from carbohydrate to protein metabolism. Fermentation of protein by gut bacteria, or putrefaction, in the colon is generally considered detrimental to health.[37] Gut bacterial metabolism of aromatic amino acids (phenylalanine, tyrosine and tryptophan) results in the production of phenols and indoles,[38] which are inflammatory and potentially toxic compounds.[39 40] Bacterial enzymes are necessary for metabolism of phenylalanine and tyrosine to phenolic compounds phenylacetate and *p*-cresol, respectively, and for the metabolism of tryptophan to indole. Indole and *p*-cresol are sulfated to IS and PCS, respectively, prior to excretion while phenylacetate is converted to PAG by host metabolism. PHPA is an intermediate in the tyrosine degradation pathway but is also formed from phenylacetate.[38] *Enterococcus* spp. contain enzymes necessary for both phenylalanine and tyrosine metabolism; *E. coli* is involved in tyrosine metabolism.[41] Both species are also known as indole producers.[42] Although *Streptococcus* may not be directly involved in the formation of these metabolites, it does have proteolytic activity[43] and could contribute to the observed phenotype by making amino acids available from dietary protein. To our knowledge, this is the first report of upregulation of host-microbe co-metabolism of aromatic amino acids post-MAL surgery in humans. Previous reports have shown that the gut microbiota shifts to protein putrefaction after a MAL BS[10 28 29] but, having focused on faecal samples, the host component of these pathways was overlooked. Lacking dietary data, we cannot exclude the possibility that the MAL metabolic phenotype results from increased protein intake. However, reports of protein malnutrition and protein intolerance post-BS[44 45] suggest that high protein intake is unlikely in these patients. An inverse association between PAG/PCS and BMI has been reported although the relationship between adipiosity and the gut microbiota is still under debate.[46] Other reports, in agreement with our data, have identified decreased concentrations of branched-chain amino acids in non-pregnant BS patients which supports our hypothesis that protein metabolism is altered in these individuals. Branched-chain amino acids have previously been linked to insulin resistance associated with obesity.[47–50]

Our data suggest that PAG and PCS, or their respective metabolic pathways, are also associated with reduction in maternal insulin resistance following BS. We are aware of recent discussion surrounding the suitability of oral glucose tolerance testing in bariatric patients and have since revised our methods. In spite of this caveat, it is conceivable that the gut microbiota plays a role in insulin resistance.[49 51] Our finding of an inverse correlation between maternal urinary host-microbial co-metabolites and fetal/birth weight indicates that the maternal metabolic changes may also be related to the risk of these women delivering small for gestational age neonates. Although malnutrition resulting from maternal malabsorption likely contributes to reduced fetal growth, the downstream metabolic consequences of malabsorption also warrant further investigation given the potential toxicity of the final metabolic products. It has also been shown that higher urinary concentrations of PCS are indicative of reduced sulfation capacity in the host, which presents significant competition for similar reactions, namely acetaminophen detoxification, which also require sulfation prior to excretion.[52] Pharmacokinetics in these patients should be investigated, especially during pregnancy, given that MAL patients have higher urinary concentrations of two sulfated compounds (PCS and IS).

Pregnancy has been shown to induce substantial time-dependent alterations in physiology and metabolism[53 54] under the constantly changing physiological demand as the mother responds to the needs of the growing fetus. Nevertheless, the systematic effect of BS on the metabolic profiles is superimposed on the changing metabolic landscape and can be clearly observed in the metabolic phenotypes. In particular, GlycA, a metabolite associated with systemic inflammation, increased during the first two trimesters but remained lower in MAL patients during the third trimester indicating that these women have a lower grade of pregnancy-related inflammation compared with obese women without BS.[50 55] Drastic metabolic changes could also be seen at the time of delivery in all groups which is likely due to the metabolic consequences of the labour process and masked surgery-related differences.[56] Intriguingly, despite prophylactic administration of antibiotics (given to 59% of the women at delivery), PAG was detected in the neonatal urine soon after birth and in significantly higher concentration in those born from women with previous MAL surgery compared with the offspring of the NBS women. Although this result should be treated with caution due to the low number of newborns enrolled in the study, it suggests transfer of the modified maternal profile to the offspring. Further independent studies are required to validate the generalisability of this observation with longer-term follow-up of the infants to determine whether this neonatal metabolic phenotype is maintained and how it impacts their future risk of obesity and diabetes. BS is not an appropriate intervention at population scale but identification of beneficially altered physiology may allow for development of targeted interventions aimed at specifically modulating key pathways.

Twitter Kiana Ashley West @kiana_a_west

Acknowledgements The authors would like to thank Dr Ivan Serrano Contreras for assistance with carrying out 2D NMR spectroscopy. We would also like to acknowledge that KAW is supported by the STRATiGRAD doctoral training program and JVL is supported by Medical Research Council New Investigator Grant (P62104) and European Research Council Starting Grant (P61481). EH is supported by the Western Australian Government through the Premier's Science Fellowship Program. Infrastructure support was provided by the National Institute for Health Research Imperial Biomedical Research Centre aligned with the gut health theme.

Contributors EH, MDS, JVL and MRJ designed the study. CK and TM recruited patients and collected biospecimens. KAW, CK and JAKM performed the experiments. KAW analysed the data. All authors contributed to the interpretation of results. KAW, MDS and EH wrote the manuscript. All authors read and approved the final manuscript.

Funding The study was supported partially by BORNE (UK Registered Charity number: 1167073).

Disclaimer The funding body did not contribute to the design of the study, collection, analysis, interpretation of data or writing the manuscript.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval The study was approved by the West London Research Ethics Committee (No: 14/LO/0592).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Raw sequencing data are available in a public, open access repository: Sequence Read Archive under BioProject PRJNA493625. Raw NMR data are available upon reasonable request. All pre-processed data and code to reproduce the results are available in a public, open access repository: GitHub (https://github.com/ka-west/PBS_manuscript). The R script is also included as a pdf file created with 'knitr' (see online supplementary methods).

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is

Gut: first published as 10.1136/gutjnl-2019-319620 on 21 January 2020. Downloaded from http://gut.bmj.com/ on May 27, 2020 at BVA. Protected by copyright.

properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Kiana Ashley West http://orcid.org/0000-0002-1896-7936
Julie Anne Kathryn McDonald http://orcid.org/0000-0003-0739-6047

REFERENCES

1 World Health Organization. *Obesity and overweight*, 2017.
2 Abdelaal M, le Roux CW, Docherty NG. Morbidity and mortality associated with obesity. *Ann Transl Med* 2017;5.
3 Ashrafian H, Ahmed K, Rowland SP, *et al*. Metabolic surgery and cancer. *Cancer* 2011;117:1788–99.
4 Noria SF, Grantcharov T. Biological effects of bariatric surgery on obesity-related comorbidities. *Can J Surg* 2013;56:47–57.
5 Ashrafian H, Athanasiou T, Li JV, *et al*. Diabetes resolution and hyperinsulinaemia after metabolic Roux-en-Y gastric bypass. *Obes Rev* 2011;12:e257–72.
6 Elder KA, Wolfe BM. Bariatric surgery: a review of procedures and outcomes. *Gastroenterology* 2007;132:2253–71.
7 Graessler J, Qin Y, Zhong H, *et al*. Metagenomic sequencing of the human gut microbiome before and after bariatric surgery in obese patients with type 2 diabetes: correlation with inflammatory and metabolic parameters. *Pharmacogenomics J* 2013;13:514–22.
8 Kong L-C, Tap J, Aron-Wisnewsky J, *et al*. Gut microbiota after gastric bypass in human obesity: increased richness and associations of bacterial genera with adipose tissue genes. *Am J Clin Nutr* 2013;98:16–24.
9 Calvani R, Miccheli A, Capuani G, *et al*. Gut microbiome-derived metabolites characterize a peculiar obese urinary metabotype. *Int J Obes* 2010;34:1095–8.
10 Tremaroli V, Karlsson F, Werling M, *et al*. Roux-En-Y gastric bypass and vertical banded gastroplasty induce long-term changes on the human gut microbiome contributing to fat mass regulation. *Cell Metab* 2015;22:228–38.
11 Li JV, Ashrafian H, Bueter M, *et al*. Metabolic surgery profoundly influences gut microbial-host metabolic cross-talk. *Gut* 2011;60:1214–23.
12 Nicholson JK, Holmes E, Kinross J, *et al*. Host-gut microbiota metabolic interactions. *Science* 2012;336:1262–7.
13 Holmes E, Li JV, Athanasiou T, *et al*. Understanding the role of gut microbiome-host metabolic signal disruption in health and disease. *Trends Microbiol* 2011;19:349–59.
14 Kwong W, Tomlinson G, Feig DS. Maternal and neonatal outcomes after bariatric surgery; a systematic review and meta-analysis: do the benefits outweigh the risks? *Am J Obstet Gynecol* 2018;218:573–80.
15 Maric T, Kanu C, Johnson MR, *et al*. Maternal, neonatal insulin resistance and neonatal anthropometrics in pregnancies following bariatric surgery. *Metabolism* 2019;97:25–31.
16 Dona AC, Jiménez B, Schäfer H, *et al*. Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping. *Anal Chem* 2014;86:9887–94.
17 Cloarec O, Dumas M-E, Craig A, *et al*. Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic 1H NMR data sets. *Anal Chem* 2005;77:1282–9.
18 Posma JM, Garcia-Perez I, De Iorio M, *et al*. Subset optimization by reference matching (storm): an optimized statistical approach for recovery of metabolic biomarker structural information from 1H NMR spectra of biofluids. *Anal Chem* 2012;84:10694–701.
19 R Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2017. https://www.R-project.org/
20 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc* 1995;57B:289–300.
21 Mullish BH, Pechlivanis A, Barker GF, *et al*. Functional microbiomics: evaluation of gut microbiota-bile acid metabolism interactions in health and disease. *Methods* 2018;149:49–58.
22 Callahan BJ, Sankaran K, Fukuyama JA, *et al*. Bioconductor workflow for microbiome data analysis: from raw reads to community analyses. *Version 2. F1000Res* 2016;5:1492.
23 Callahan BJ, McMurdie PJ, Rosen MJ, *et al*. DADA2: high-resolution sample inference from illumina amplicon data. *Nat Methods* 2016;13:581–3.
24 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol* 2014;15:550.
25 Rohart F, Gautier B, Singh A, *et al*. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput Biol* 2017;13:e1005752.
26 Quast C, Pruesse E, Yilmaz P, *et al*. The Silva ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013;41:D590–6.
27 Cole JR, Wang Q, Fish JA, *et al*. Ribosomal database project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* 2014;42:D633–42.
28 Ilhan ZE, DiBaise JK, Isern NG, *et al*. Distinctive microbiomes and metabolites linked with weight loss after gastric bypass, but not gastric banding. *ISME J* 2017;11:2047–58.
29 Palleja A, Kashani A, Allin KH, *et al*. Roux-En-Y gastric bypass surgery of morbidly obese patients induces swift and persistent changes of the individual gut microbiota. *Genome Med* 2016;8:67.
30 Zhang H, DiBaise JK, Zuccolo A, *et al*. Human gut microbiota in obesity and after gastric bypass. *Proc Natl Acad Sci U S A* 2009;106:2365–70.
31 Liou AP, Paziuk M, Luevano J-M, *et al*. Conserved shifts in the gut microbiota due to gastric bypass reduce host weight and adiposity. *Sci Transl Med* 2013;5:178ra41.
32 Aron-Wisnewsky J, Doré J, Clement K. The importance of the gut microbiota after bariatric surgery. *Nat Rev Gastroenterol Hepatol* 2012;9:590–8.
33 Liu Z, Cao AT, Cong Y. Microbiota regulation of inflammatory bowel disease and colorectal cancer. *Semin Cancer Biol* 2013;23:543–52.
34 Seyfried F, Li JV, Miras AD, *et al*. Urinary phenotyping indicates weight loss-independent metabolic effects of Roux-en-Y gastric bypass in mice. *J Proteome Res* 2013;12:1245–53.
35 Conlon M, Bird A. The impact of diet and lifestyle on gut microbiota and human health. *Nutrients* 2015;7:17–44.
36 Ponsky TA, Brody F, Pucci E. Alterations in gastrointestinal physiology after Roux-en-Y gastric bypass. *J Am Coll Surg* 2005;201:125–31.
37 Windey K, De Preter V, Verbeke K. Relevance of protein fermentation to gut health. *Mol Nutr Food Res* 2012;56:184–96.
38 Smith EA, Macfarlane GT. Formation of phenolic and indolic compounds by anaerobic bacteria in the human large intestine. *Microb Ecol* 1997;33:180–8.
39 Meyer TW, Hostetter TH. Uremic solutes from colon microbes. *Kidney Int* 2012;81:949–54.
40 Sun C-Y, Hsu H-H, Wu M-S. P-Cresol sulfate and indoxyl sulfate induce similar cellular inflammatory gene expressions in cultured proximal renal tubular cells. *Nephrol Dial Transplant* 2013;28:70–8.
41 Gryp T, Vanholder R, Vaneechoutte M, *et al*. p-Cresyl sulfate. *Toxins (Basel)* 2017;9:52.
42 Lee J-H, Lee J. Indole as an intercellular signal in microbial communities. *FEMS Microbiol Rev* 2010;34:426–44.
43 Macfarlane GT, Cummings JH, Allison C. Protein degradation by human intestinal bacteria. *Microbiology* 1986;132:1647–56.
44 Steenackers N, Gesquiere I, Matthys C. The relevance of dietary protein after bariatric surgery: what do we know? *Curr Opin Clin Nutr Metab Care* 2018;21:58–63.
45 Lupoli R, Lembo E, Saldalamacchia G, *et al*. Bariatric surgery and long-term nutritional issues. *World J Diabetes* 2017;8:464–74.
46 Elliott P, Posma JM, Chan Q, *et al*. Urinary metabolic signatures of human adiposity. *Sci Transl Med* 2015;7:285ra62-ra62.
47 Newgard CB, An J, Bain JR, *et al*. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab* 2009;9:311–26.
48 Tai ES, Tan MLS, Stevens RD, *et al*. Insulin resistance is associated with a metabolic profile of altered protein metabolism in Chinese and Asian-Indian men. *Diabetologia* 2010;53:757–67.
49 Pedersen HK, Gudmundsdottir V, Nielsen HB, *et al*. Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature* 2016;535:376–81.
50 Houttu N, Mokkala K, Laitinen K. Overweight and obesity status in pregnant women are related to intestinal microbiota and serum metabolic and inflammatory profiles. *Clin Nutr* 2018;37:1955–66.
51 Kootte RS, Levin E, Salojärvi J, *et al*. Improvement of insulin sensitivity after lean donor feces in metabolic syndrome is driven by baseline intestinal microbiota composition. *Cell Metab* 2017;26:611–9.
52 Clayton TA, Baker D, Lindon JC, *et al*. Pharmacometabonomic identification of a significant Host-microbiome metabolic interaction affecting human drug metabolism. *Proc Natl Acad Sci U S A* 2009;106:14728–33.
53 Liu X, Wang X, Sun H, *et al*. Urinary metabolic variation analysis during pregnancy and application in gestational diabetes mellitus and spontaneous abortion biomarker discovery. *Sci Rep* 2019;9:2605.
54 Wang M, Xia W, Li H, *et al*. Normal pregnancy induced glucose metabolic stress in a longitudinal cohort of healthy women: novel insights generated from a urine metabolomics study. *Medicine* 2018;97:e12417.
55 Otvos JD, Shalaurova I, Wolak-Dinsmore J, *et al*. GlycA: a composite nuclear magnetic resonance biomarker of systemic inflammation. *Clin Chem* 2015;61:714–23.
56 Caboni P, Meloni A, Lussu M, *et al*. Urinary metabolomics of pregnant women at term: a combined GC/MS and NMR approach. *J Matern Fetal Neonatal Med* 2014;27(Suppl 2):4–12.

Gut: first published as 10.1136/gutjnl-2019-319620 on 21 January 2020. Downloaded from http://gut.bmj.com/ on May 27, 2020 at BVA. Protected by copyright.