

University of Nottingham
UK | CHINA | MALAYSIA

OMICS APPROACH TO INVESTIGATE MUSCULOSKELETAL, LIVER AND METABOLIC TRAITS

Ojasvi Mehta B.Sc, M.Sc

Thesis submitted to the University of Nottingham for the degree of

Doctor of Philosophy (PhD)

UNIVERSITY OF NOTTINGHAM

SCOPE AND PURPOSE OF THESIS

"All disease begins in the gut"

- Hippocrates, Father of Medicine

Various acute or chronic diseases such as undernourishment, musculoskeletal and liver diseases have been linked to the microbiome[1] and metabolome and these two factors are directly or indirectly reported to be affected by various intrinsic and extrinsic factors such as diet or other lifestyle habits and genetics. This thesis aims to study the effect and interplay of the gut microbiome and metabolome in osteoarthritis (OA) and the same in association of diet in undernourishment and non-alcoholic fatty liver (NAFLD) diseases. Although the phenotypes, pathology, and underlying mechanisms of these diseases are different, the gut is considered central to all the diseases. This is because the gut is harboured with trillions of bacteria and the most diverse microbiome among all the body regions. This makes the gut capable of metabolizing the indigestible dietary compounds and synthesizing a range of vitamins and amino acids that are not encoded by the human genome. These metabolites are vital for the human body and the interaction between the microbiome, metabolome and diet is required to maintain homeostasis.

As a whole, this thesis presents a consensus on how the shared harmony of the gut microbiome and the metabolome in combination with other factors exert an effect on health and a varied range of disorders and metabolic diseases across a diverse population. This has been achieved by investigating the relationship of the gut microbiome and metabolome with the disease and healthy state, impact of diet on the same and their implications in clinical practice by exploring various multi-omics-based analytical approaches.

1 Microbiome comprises of the bacteria, virus, fungi, archaea. But most commonly, the microbiome term has been often implied in reference to the bacterial domain as there were well-established protocols for screening, isolation, and characterization of bacteria from various ecosystems and also the one that were earlier considered to be constituting 99% of the microbiome (Turnbaugh et al., 2007; Huttenhower et al., 2012). However, researchers have not studied the other microbial domains of life inhabiting humans back then. The concept of a 99% ratio of the bacterial community in the microbiome has been recently challenged by emerging technologies. However, for the current work, we have focussed on the bacterial domain and the term microbiome refers to the bacterial constituents of the microbiome unless specified.

COVID-19 IMPACT STATEMENT

From the start of the lockdown due to the Covid-19 pandemic (from 23rd March 2020), my PhD had been affected in terms of pattern and productivity. Being into 5 months of my PhD candidature at that time, there have been challenges impacting the opportunities and trajectory of work. The prospects of innovations and accomplishments was limited due to non-availability of facilities and personnel due government restrictions affecting University operations despite everybody's best efforts.

The Webex cohort was designed to investigate the impact of web-based exercise on microbiome and metabolome on knee pain in patients with OA and began recruiting in 2019. Data was available after being collected at baseline visit for 100 patients, but the follow-up visits planned were limited due to the unexpected introduction of lockdowns and social distancing restrictions due to the Covid-19 pandemic which made the study underpowered. As there are scare studies investigating the microbiome and metabolome in knee pain, my supervisors and I agreed to change the analytical approach and investigate the gut microbes and serum metabolites associated with knee pain scores and pressure pain detection thresholds. We included a meta-analysis comparing two independent studies conducted in the department which formed Chapter 4 of this thesis (Study 2-knee pain in OA).

Apart from impacting research work, Covid-19 affected my study schedule and focus. The pandemic has been a setback consequent to the new variants, travel halts, lockdowns, remote studies, worries following various episodes of health issues of family members with serious underlying comorbidities and meeting and speaking to colleagues rarely. But I would like to mention that I am very pleased that my supervisors and colleagues have provided me a lot of support through online meetings and discussions.

ACKNOWLEDGEMENTS

Ph.D. for me has been like building a monument that commemorate all the learnings and efforts of your research journey. It represents as an asset for the entire professional life. Embarking on this Ph.D. has indeed been a life-changing experience for me, and it would not have been possible to do without the guidance and support that I received from many people.

I am extremely grateful to all my supervisors, for providing much-needed guidance, support, encouragement and constant motivation. Specifically, I am grateful to Prof. Ana M. Valdes for giving me the opportunity to work as Research Assistant with her, by which I could manage my sustenance as an international student in the UK. She has been the source of illumination for me during my studies. I am thankful to Prof. Guruprasad Aithal for providing me with the intellectual support and insights which served as the skylight during my studies. The support from Dr. Jane I. Grove has been the pillars of my Ph.D. and she has always been particularly available for regular and face-to-face meetings along with providing me with all the administrative support necessary for the management of my studies. I highly value Dr. Amrita Vijay and Dr. Stuart Astbury who had been continuous support for my day-to-day discussions and enthusiasm and their support had formed the walls of my Ph.D.

I can't thank enough my supervisors for providing their constant guidance and feedback efficiently during Teams meetings scheduled at different time zones over the months I spent undertaking my remote studies in India. I would like to thank the clinical team, lab assistants, managers and collaborators for doing the data collection, and laboratory-based experiments.

I gratefully acknowledge the University of Nottingham for providing me with the Vice Chancellor's Scholarship for Research and Excellence for my Ph.D., without which this Ph.D. would not have been achievable.

My deep appreciation goes out to Amina Al-Awadi, a colleague and a companion for her friendship and warmth that she extended to me during the ups and lows of my Ph.D. I would like to extend my deepest gratitude to Arun

Khilnani for his unconditional support and friendship during tough times of stay in UK.

I am highly indebted to my Mum-Usha Rani, my Dad-Prem Kumar and my Brothers-Amandeep and Ankur who always motivated me to chase my dreams and persuaded me that I should accomplish my passion of earning doctorate degree. Special thanks to my dear friends, Meenu Chitkara, Mattapalli Rama Gowtham, V. Sucharita, Manisha Chugh and Samta Srivastava for their encouragement to make progress towards Ph.D. in tough times of lock-down during the covid-era.

And finally, to my husband - Davinder Kumar and my in-laws for their support without which I would not be able to complete my studies. And to my darling son Devaj for being such a good little baby being of at the age of seven months managing well in my absence back in India, while I am in the UK, making it possible for me to complete what I started.

I dedicate my thesis to my dad, my mum and my dearest son.

DECLARATIONS

This is to certify that the work submitted in this thesis is the result of original research. It has not already been accepted for any degree, diploma, or other qualification. It has been conducted substantially by me with assistance and collaboration as mentioned below. All the collaborators, authors and technical staff are fully acknowledged.

Study 1 (Undernourishment in Adults) was conducted in a population of India in the Bangalore district in rural and slum geographic locations. This study was designed by Prof. Ana Valdes and run by the University of Nottingham in collaboration with the Department of Community Health at Bangalore Baptist Hospital. Patient recruitment and data collection were done by the team in India including Dr. Gift Norman, and Dr. Leeberk Raja. Sample handling and general administration of microbiome sequencing were done by the team at the University of Nottingham including Prof. Ana M Valdes, Prof. Guruprasad Aithal, Dr. Jane I Grove, Dr. Amrita Vijay, Dr. Stuart Astbury and me. Data analysis and writing were done by me.

Data for **Study 2** (Knee Pain in OA) were utilised from the Knee Pain in the Community (KPIC) cohort and Web-based exercise intervention (Webex) cohort. Study design, ethical application, data collection, and general administration for KPIC study data were conducted by the KPIC team, including Dr. Gwen S. Fernandes, Prof. Ana M. Valdes, Prof. Weiya Zhang, Prof. David A. Walsh and Prof. Michael Doherty and Dr. Anthony Kelly. Webex study was designed by Prof. Ana Valdes and conducted by the team including Dr. Sameer Gohir, Dr. Anthony Kelly, Dr. Amrita Vijay. Study 2 of this thesis which utilised data from these two cohorts was designed by Professor Ana M. Valdes and me. Data analysis and writing were done by me.

Study 3 (NAFLD) utilised data from Fatty Liver and Steatohepatitis study and Omega3 diet intervention study. The NAFLD study was designed by Prof. Guruprasad Aithal and delivered by the NIHR Nottingham Biomedical Research Centre team with support from Dr. Jane I Grove, Dr. Naaventhan Palaniyappan and Dr. Robert Scott. Patient biopsies were scored by Dr. Philip Kaye. The omega3 diet intervention study was designed by Prof. Ana Valdes and Dr.

Amrita Vijay. Study 3 of this thesis that utilised data from these two cohorts was designed by me. Food Frequency Questionnaires (FFQs) were inputted jointly by Amina Al-Awadi and me. Data analysis and writing were done by me.

Technical Assistance I am grateful to Melanie Lingaya for technical assistance including DNA extraction.

All data presented in this thesis were analysed by me with the supervision of Prof. Ana M. Valdes, Prof. Guruprasad Aithal, Dr. Jane I Grove, Dr. Amrita Vijay and Dr. Stuart Astbury.

PUBLICATIONS and PRESENTATIONS

Publications

- **Ojasvi Mehta***, Leeberk Raja Inbaraj*, Stuart Astbury, Jane I. Grove, Gift Norman, Guruprasad P. Aithal, Ana M. Valdes, and Amrita Vijay. Gut Microbial Profile Is Associated with Residential Settings and Not Nutritional Status in Adults in Karnataka, India. Front Nutr. 2021. 23;8:595756. (* **indicates Joint First Author**).

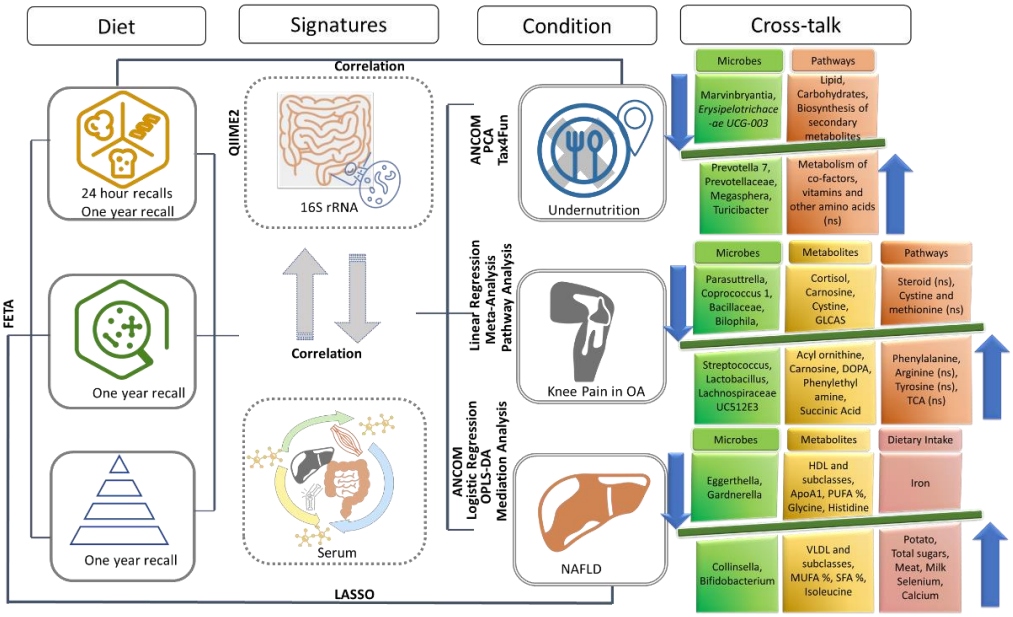
- **Ojasvi Mehta**, Amrita Vijay, Sameer Akram Gohir, Tony Kelly, Weiya Zhang, Michael Doherty, David Andrew Walsh, Guruprasad Aithal, Ana. M. Valdes. Serum Metabolome Analysis identified amino-acid metabolism associated with pain in people with symptomatic knee Osteoarthritis – a cross-sectional study (under communication in Journal of Pain- JPAIN-D-22-00543).

- Ojasvi Mehta et. al. Multi-omics approach to describe NAFLD: Correlation to Causation (under preparation).

Conferences

- One Day Symposium on "Metabolic Associated fatty Liver Disease (MAFLD)" - 25th June, 2022 at Translational Health Sciences and Technology Institute, Faridabad, India.
- "Investigating and Managing Infectious Diseases in the Microbiome Era" conducted by Microbiome Virtual International Conference (MVIF) – 13th September, 2022.

ABSTRACT



GRAPHICAL ABSTRACT

BACKGROUND

Human health is maintained by crosstalk between the gut microbiome and the gastrointestinal tract. The gut microbiome breaks down complex dietary compounds to produce metabolites which serve a multitude of functions and interact with host molecular pathways. The gut microbiome of an individual changes with host and environmental factors including location of residence and diet.

AIMS

To utilise open source software and apply tailored statistical approaches to investigate the role of the omics features of undernourishment, knee pain in osteoarthritis (OA), non-alcoholic fatty liver disease (NAFLD) and the effect of location of residence and dietary intakes. The objectives were: [1] To explore the relationship of the gut microbiome and its functional metabolic pathways with nutrition status and location of residence. [2] To explore the relationship of current knee pain with the serum metabolites and gut microbiome in people with OA. [3] To investigate the interplay of the gut microbiome, serum metabolome and dietary intakes in NAFLD compared to healthy participants.

METHODOLOGY

This research incorporated analysis of 16S rRNA microbiome sequencing, serum metabolome, serum cytokines, and dietary recalls. Data analysis was performed using QIIME2 for microbiome sequence analysis, MicrobiomeAnalyst (Tax4Fun) for functional profiling from 16S rRNA, Metaboanalyst (Pathway Analysis) for investigating the pathways and FETA for calculating dietary intakes. The differential abundance (ANCOM), statistical significance inferences, correlation tests, regression modelling (linear, logistic and LASSO), classification modelling (OPLS-DA), meta-analysis (REML) and mediation analysis (causal mediation) were performed using statistical packages in R.

RESULTS

[1] Microbiome analysis showed an increased abundance of *Prevotella* 7 and uncultured members belonging to Prevotellaceae, *Turicibacter*, *Megasphaera* and decreased abundance of *Marvinbryantia* and *Erysipelotrichaceae UCG-003* in participants residing in an urban location compared to participants from rural Indian population. Also, the functional profiling showed the KEGG Orthologs involved in the carbohydrate metabolism, lipid metabolism and biosynthesis of secondary metabolites to be significantly altered based on the location regardless of the difference in the macronutrient intakes. This suggests that location of residence and not the nutritional status assessed based on BMI is the stronger determining factor modulating the gut microbiome and the functional metabolic pathways.

[2] Meta-analysis of two independent UK studies showed levels of acyl ornithine, carnosine, cortisol, cortisone, cystine, dihydroxyphenylalanine, glycolithocholic acid sulphate, phenylethylamine and succinic acid were significantly associated with knee pain. The pathway analysis further demonstrated amino acids and cholesterol metabolism pathways are involved in knee pain. These metabolites, irrespective of the association with the pain scores were also found to be significantly associated with the serum levels of cytokines including IL-10, IL-13, IL-1β, IL-2, IL-8 and TNF-α suggesting that they may have immunomodulatory properties. Significant associations of increased abundance of *Parasutrella* and decreased abundance of *Streptococcus* with an

increase in pain and their significant associations with biogenic amines such as taurine, histamine and spermidine suggest their key role in secondary metabolism related to pain pathways.

[3] Microbiome analysis revealed that an increased abundance of mainly *Collinsella* and *Bifidobacterium* is associated with an increased likelihood of developing NAFLD. Changes in levels of serum lipoproteins such as HDL and VLDL in NAFLD and their association with specific microbes suggests a potential role of microbes in secondary metabolism of lipids. The differences in the intakes of calcium, iron, selenium, total sugars and different food items classified under potatoes, meat and milk suggests that diet plays an important role in NAFLD. The significant mediation effect scores of the microbiome as mediator suggests its causal link between the diet and disease status.

CONCLUSION

Taken together, this work has demonstrated the utility of integration and analysis of multi-omics data to establish the contribution of the gut microbiome, metabolome and diet towards understanding and improving human health. The omics features associated with the pathophysiological conditions or diseases identified in this research may serve as the molecular signatures or diagnostic biomarkers. This thesis paves a way for the discovery of new therapeutic targets or intervention strategies. This research makes an important contribution towards a deeper understanding and better management of nutrition, knee OA and NAFLD.

INDEX

SCOPE AND PURPOSE OF THESIS i
COVID-19 IMPACT STATEMENT ii
ACKNOWLEDGEMENTS iii
DECLARATIONS v
PUBLICATIONS and PRESENTATIONS vii
Publications vii
Conferences vii
ABSTRACT viii
GRAPHICAL ABSTRACT viii
BACKGROUND viii
AIMS viii
METHODOLOGY ix
RESULTS ix
CONCLUSION x
INDEX xi
LIST OF FIGURES xvi
LIST OF TABLES xviii
ABBREVIATIONS xix
Chapter 1 GUT MICROBIOME AND METABOLOME: AN INTRODUCTION 1
1.1 CONCEPTS OF GUT MICROBIOME AND METABOLOME 2
1.2 DYNAMICS OF GUT MICROBIOME AND METABOLOME 5
1.2.1 GUT MICROBIOME: COMPOSITION AND VARIABILITY 5
1.2.2 METABOLOME AND THE MICROBIOME 7
1.2.3 INFLUENCE OF OTHER FACTORS 11
1.2.3.1 DIET AND NUTRITION 11
1.2.3.2 EXERCISE AND PHYSICAL ACTIVITY 13
1.2.3.3 HOST GENETICS AND PHYSIOLOGY 14
1.3 RESEARCH PROBLEM AND RESEARCH QUESTIONS 15
1.4 ROLE OF GUT MICROBIOME AND METABOLOME IN SPECIFIC DISEASES/CONDITIONS 17
1.4.1 UNDERNOURISHMENT 18
1.4.2 OSTEOARTHRITIS: 19
1.4.3 NON-ALCOHOLIC FATTY LIVER DISEASE (NAFLD): 21
1.5 HYPOTHESIS AND AIMS 23

Chapter 2 METHODOLOGY 25
2.1 STUDY COHORTS 26
2.1.1 STUDY 1: UNDERNOURISHMENT IN ADULTS 29
2.1.2 STUDY 2: KNEE OSTEOARTHRITIS (OA) 29
2.1.2.1 WEBEX 30
2.1.2.2 KPIC 30
2.1.3 STUDY 3: NON-ALCOHOLIC FATTY LIVER DISEASE 31
2.1.3.1 NAFLD-UK COHORT 31
2.1.3.2 OMEGA-3 DIET INTERVENTION COHORT 32
2.2 MICROBIOME 32
2.2.1 SAMPLE COLLECTION AND TRANSPORTATION 33
2.2.2 DNA EXTRACTION 34
2.2.3 16S rRNA SEQUENCING 34
2.2.4 16S rRNA ANALYSIS USING QIIME 2 (VERSION 2019.10) 35
2.2.4.1 DATA PRE-PROCESSING 37
2.2.4.2 TAXONOMIC ASSIGNMENT 38
2.2.4.3 RAREFACTION 38
2.2.4.4 DIVERSITY METRICS AND PHYLOGENETIC ANALYSIS 39
2.2.5 DIFFERENTIAL ABUNDANCE (DA) TESTING OF GUT MICROBES 41
2.2.5.1 ANCOM 2.0 42
2.2.6 FUNCTIONAL PROFILING OF 16S rRNA MICROBIOME DATA 42
2.3 METABOLOME 45
2.3.1 METABOLOMICS 45
2.3.2 SAMPLE COLLECTION, PREPARATION AND DATA ACQUISITION 46
2.3.2.1 BIOCRATES PANEL 46
2.3.2.2 NIGHTINGALE PANEL 47
2.3.3 PRE-PROCESSING OF METABOLITE DATA 47
2.4 INFLAMMATORY MARKERS 47
2.5 DIET 48
2.5.1 DIETARY ASSESSMENT 48
2.5.2 DIET DATA ANALYSIS IN FETA SOFTWARE 50
2.5.3 ENERGY ADJUSTMENTS FOR REGRESSION MODELS 50
2.6 STATISTICAL ANALYSIS 50
2.6.1 INFERRING SIGNIFICANCE 50
2.6.2 PREDICTING ASSOCIATIONS. 51
2.6.3 META-ANALYSIS. 51

2.6.4 PATHWAY ANALYSIS. 52
2.6.5. ORTHOGONAL PARTIAL LEAST SQUARE DISCRIMINANT ANALYSIS (OPLS-DA) AND VARIABLE IMPORTANCE 53
2.6.6. LEAST ABSOLUTE SHRINKAGE AND SELECTOR OPERATOR (LASSO) 54
2.6.7. MEDIATION ANALYSIS 55
Chapter 3 GUT MICROBIAL PROFILE IS ASSOCIATED WITH RESIDENTIAL SETTINGS AND NOT NUTRITIONAL STATUS IN KARNATAKA, INDIA 57
3.1 ABSTRACT 58
3.2 INTRODUCTION 59
3.3 METHODS 61
3.3.1 STUDY SETTINGS 61
3.3.1.1 URBAN SLUM 61
3.3.1.2 RURAL VILLAGE 62
3.3.2 RECRUITMENT OF STUDY PARTICIPANTS 62
3.3.3 CRITERIA USED FOR ASSESSING UNDERNUTRITION STATUS 63
3.3.4 STUDY ETHICS 63
3.3.5 SAMPLE COLLECTION, PROCESSING AND DNA EXTRACTION 63
3.3.6 SEQUENCING DATA ANALYSIS AND DIVERSITY METRICS 64
3.3.7 FUNCTIONAL PROFILING FROM 16S rRNA DATA 64
3.3.8 STATISTICAL ANALYSIS 64
3.4 RESULTS 65
3.4.1 SUBJECT AND CLINICAL CHARACTERISTICS: 65
3.4.2 OVERALL DIFFERENCE AND DIVERSITY INDICES IN THE MICROBIAL COMMUNITY 66
3.4.3 LOCATION IS THE STRONGEST DETERMINANT OF THE GUT MICROBIOTA VARIATION ACROSS THE SAMPLES 72
3.4.4 KEGG Orthologs and Functional Pathways 73
3.5 DISCUSSION 77
Chapter 4 GUT MICROBIOME AND SERUM METABOLOME SIGNATURES OF PAIN INTENSITY IN KNEE OSTEOARTHRITIS 81
4.1 ABSTRACT 82
4.2 INTRODUCTION 83
4.3 METHODS 85
4.3.1 STUDY DESIGN AND PARTICIPANTS RECRUITMENT 85
4.3.2 CLINICAL EVALUATION AND DATA COLLECTION 87
4.3.3 SAMPLE ACQUISITION AND PROFILING 87
4.3.3.1 METABOLITE PROFILING 87

4.3.3.1 DNA EXTRACTION, MICROBIOME SEQUENCING and ANALYSIS 88
4.3.3.3 CYTOKINE MEASUREMENTS 88
4.3.4 METABOLITE DATA PRE-PROCESSING 89
4.3.6 STATISTICAL ANALYSIS 89
4.4 RESULTS 90
4.4.1 ANTHROPOMETRIC AND CLINICAL CHARACTERISTICS 90
4.4.2 METABOLOME AND PAIN 91
4.4.3 METABOLOME AND PRESSURE PAIN THRESHOLDS (PPTs) 95
4.4.4 AMINO-ACID METABOLISM AND KNEE PAIN 98
4.4.5 PAIN AND PRO OR ANTI-INFLAMMATORY CYTOKINES IN THE TEST COHORT 102
4.4.6 GUT MICROBIOME FEATURE ANALYSIS 107
4.4.7 GUT MICROBIOME AND PAIN IN WEBEX DATASET 107
4.4.7 GUT MICROBIOME AND PRESSURE PAIN THRESHOLDS (PPTs) IN WEBEX STUDY 107
4.4.7 GUT MICROBIOME AND METABOLOME ASSOCIATIONS IN WEBEX STUDY 108
4.5 DISCUSSION AND CONCLUSION 111
Chapter 5 MULTI-OMICS APPROACH TO DESCRIBE NAFLD: FROM CORRELATION TO CAUSATION 116
5.1 ABSTRACT 117
5.2 INTRODUCTION 118
5.3 METHODS 121
5.3.1 STUDY DESIGN, PARTICIPANTS RECRUITMENT AND SAMPLE COLLECTION 121
5.3.2 CLINICAL EVALUATION OF NAFLD 123
5.3.3 DIETARY ASSESSMENT AND ANALYSIS 123
5.3.4 SAMPLE COLLECTION, PROCESSING AND DNA EXTRACTION 123
5.3.5 MICROBIOME SEQUENCING AND ANALYSIS 123
5.3.6 STATISTICAL ANALYSIS 124
5.4 RESULTS 125
5.4.1 CLINICAL CHARACTERISTICS 125
5.4.2 MICROBIOME, METABOLOME AND DIET DATA DETAILS 125
5.4.3 THE GUT MICROBIOME IS ALTERED IN NAFLD PARTICIPANTS 126
5.4.4 ASSOCIATION OF THE GUT MICROBIOME WITH NAFLD AND SEVERITY SUB-GROUPS 127
5.4.5 METABOLIC PROFILE DETERMINES THE DISEASE STATUS 129
5.4.6 METABOLIC SIGNATURES OF NAFLD AND SEVERITY SUB-GROUPS 134

5.4.7 ASSOCIATION OF SERUM METABOLOME AND GUT MICROBIOME IN NAFLD AND SEVERITY SUB-GROUPS 140
5.4.8 NUTRIENT INTAKES DIFFER BETWEEN HEALTHY AND NAFLD PARTICIPANTS 143
5.4.9 DIETARY INTAKE IS ASSOCIATED WITH DISEASE STATUS 149
5.4.10 EFFECT OF DIET ON NAFLD MEDIATED BY GUT MICROBIOME 151
5.5 DISCUSSION 153
Chapter 6 DISCUSSION AND CONCLUSION 159
6.1 OVERVIEW 160
6.2 THE GUT MICROBIOME MODULATES METABOLIC HEALTH 161
6.2.1 DIFFERENT GUT MICROBIAL PHYLUM AFFECTS METABOLISM VIA DIFFERENT METABOLIC AND MOLECULAR PATHWAYS 163
6.2.2 LIVER METABOLISM IS DETERMINED BY LONG-TERM NUTRIENTS INTAKE 164
6.2.3 GUT MICROBES MEDIATE THE EFFECT OF DIETARY INTAKE ON LIVER METABOLISM 164
6.3 STATISTICAL CHALLENGES AND APPROACHES 165
6.4 LIMITATIONS 166
6.5 FUTURE ASPECTS 166
6.6 RESEARCH IMPLICATIONS AND CONCLUSION 169
REFERENCES 170
Appendices 199
Appendix 1: Commands for Microbiome Sequence Analysis using QIIME2 version 2019.10 200
Appendix 3: Supplementary Material Supporting Chapter 3 203
Appendix 3.1: Diet Data Collection Form 203
Appendix 3.2: Additional Information on dietary details of the participants collected from the 24-hour recall food frequency questionnaire 207
Appendix 4: Supplementary Material Supporting Chapter 4 210
Appendix 4.1: Biomarkers identified using Biocrates metabolic platform 210
Appendix 5: Supplementary Material Supporting Chapter 5 212
Appendix 5.1: Biomarkers identified using Nightingale NMR-metabolic platform 212
Appendix 5.2: EPIC -NORFOLK-UK 225
Appendix 5.3: FFQs - TWINS-UK 236
Appendix 5.4: Omega-3 Wellness Questionnaire 281

LIST OF FIGURES

Figure 1.1: Factors associated to maintain homeostasis and being accountable for human health.....4
Figure 1.2: Link between Microbiome and Metabolome.....10
Figure 1.3: The gut microbiome and role in various diseases.....18
Figure 2.1: Microbiome Sequencing approach.....33
Figure 2.2: 16S rRNA gene sequencing.....35
Figure 2.3: Flow Diagram of 16S rRNA sequence analysis using QIIME2.....36
Figure 2.4: FASTQ File format of Raw Reads.....37
Figure 2.5: Rarefaction plot example.....39
Figure 2.6: Shannon Diversity.....40
Figure 2.7: UniFrac Distance Measures......41
Figure 2.8: Tax4Fun Workflow.....44
Figure 2.9: Metabolomics Workflow.....46
Figure 2.10: Pathway Analysis Workflow.....53
Figure 2.11: Mediation Analysis.....56
Figure 3.1: Schematic representation of methods for this study.....61
Figure 3.2: Box and Whisker plot depicting relative abundance of significant phylum.....69
Figure 3.3: Pie Chart representation of the relative abundance of the most dominant bacterial phyla.....70
Figure 3.4: Box and Whisker plot depicting Beta Diversity (unweighted Unifrac distances) based on location.....71
Figure 3.5: Principal Coordinates Analysis (PCoA) of the similarity across the study groups based on the nutritional status (nourished vs undernourished) and location (rural vs urban slum) based on the Unweighted UniFrac distances.....72
Figure 3.6: Box and Whisker plot depicting the relative abundance of significant taxa associated with undernourished and nourished from urban slum and rural settings.....73
Figure 3.7: Box and Whisker plot depicting the pathways significantly associated with undernourished and nourished from urban slum and rural settings.....76
Figure 4.1: Flow diagram showing study design, conduct and analysis.....86
Figure 4.2: Metabolites associated with Pain scores in the Webex and KPIC study datasets.....92
Figure 4.3: Metabolites associated with Pain in test (Webex) and replicate (KPIC) cohort.....95
Figure 4.4: Meta-analysis of test (Webex) and replicate (KPIC) datasets......100
Figure 4.5: Relevant pathways involved in current knee pain based on the significantly associated metabolites from the meta-analysis.....101
Figure 4.6: Correlation between replicated metabolites and circulating cytokine levels.....104
Figure 4.7: Regression of microbiome with pain and PPTs in the Webex dataset.....108
Figure 4.8: Associations between metabolites, microbes and Pain.....109
Figure 5.1: Study design and Methodology.....122

Figure 5.2: Differentially abundant microbes in NAFLD patients when compared with Healthy Controls. 127
Figure 5.3: Forest plots of odds ratio (95% CI) of the significant microbes (FDR <0.05) associated with NAFLD when compared with HV. 129
Figure 5.4: OPLS-DA for Healthy vs NAFLD participants. 131
Figure 5.5: Variable Importance Projection 131
Figure 5.6: OPLS-DA for NASH vs CIRRHOSIS participants. 134
Figure 5.7: Forest plots of odds ratio (95% CI) of the significant metabolites (FDR <0.05) associated with NAFLD when compared with HV. 136
Figure 5.8: Micronutrients showing significantly different intakes. 143
Figure 5.9: Macronutrients showing significantly different intake. 144
Figure 5.10: Food groups showing significantly different intakes. 145
Figure 5.11: Cross-validation plot representing the lambda for the LASSO regression model. 149
Figure 5.12: ROC curve for the LASSO regression model. 151
Figure 5.13: Causal Mediation Analysis Model 152
Figure 6.1: Schematic representation of conceptual framework, key findings related to dietary factors, microbiome and as well as metabolome and interpretations from the studies conducted in this thesis. 162

Figure 1.1: Factors associated to maintain homeostasis and being accountable for human health. The figure depicts that there are various intrinsic (host) and extrinsic factors that affect human health. Host factors like genetics and immunity are responsible for directly affecting the host depending upon the environment/demography, the host has been staying in. External factors such as diet, exposure to drugs, exercises and other factors such as gut microbiome, metabolome, also affects human health. 4
Figure 1.2: Link between Microbiome and Metabolome. This figure depicts that there is an interplay of the gut microbiome and metabolome depending upon the composition, abundance and function of the gut microbes. 10
Figure 1.3: The gut microbiome and role in various diseases. This figure (adapted from Das et al., 2018) depicts that an altered gut microbiome may induce inflammation and health disorders like IBD, malnutrition, colorectal cancer and several others. 18
Figure 2.1: Microbiome Sequencing approach 33
Figure 2.2: 16S rRNA gene sequencing. This figure is inherited from (Logue *et al.*, 2016) (i) A first PCR amplification is performed on DNA extracted from each sample targeting 16S rRNA V3-V4 (grey) using primers modified with a 5'-end tail complementary to the Illumina sequencing primers (red). (ii) A second PCR amplification incorporates the Illumina adaptors and a 6-nucleotide barcode unique to each sample's PCR product. (iii) Normalised pooled PCR products are simultaneously sequenced on an MiSeq Illumina platform generating paired-end reads (grey in (iv)) and barcode sequences (grey-dashed). (v) Paired-end reads are then merged to provide consensus sequence reads. 35
Figure 2.3: Flow Diagram of 16S rRNA sequence analysis using QIIME2. 36
Figure 2.4: FASTQ File format of Raw Reads: An example of the raw reads fastq format. The raw reads are demultiplexed, denoised with a Phred score of 30 and joined for further analysis. 37

Figure 2.5: Rarefaction plot example: The figure depicts that the Shannon diversity is same with the increase in the number of sequences in each sample. This indicates that at a certain sequencing depth (here ~13,000) same microbes are sequenced over and over and this doesn't impact the Shannon diversity. This figure is taken from Study 2. 39
Figure 2.6: Shannon Diversity: This diagram presents the Shannon Diversity. Each diamond represents a different microbe, with colours corresponding to distinct taxa. 40
Figure 2.7: UniFrac Distance Measures. A) for two samples B) for all pairwise combinations in a tree to make a distance matrix-Figure B is inherited from (Lozupone and Knight, 2005). 41
Figure 2.8: Tax4Fun Workflow. This figure is adapted from (Wemheuer et al., 2018). 16S rRNA feature table is mapped with the KEGG orthologs (KO) functional reference database generating an association matrix and thus predicting metagenome. The predicted metagenome is further mapped with the KEGG BRITE (functional pathways) database containing the precomputed functional profile of KEGG organisms. KOs are then assigned together to functional categories..... 44
Figure 2.9: Metabolomics Workflow. This figure is adapted from (Nalini Kanta et al., 2020). The biological fluids are prepared and tagged with the label for the identification and quantification of the metabolites. The intensities of the signal are relative to the concentration of the metabolites. 46
Figure 2.10: Pathway Analysis Workflow: This figure represents the steps implemented in Pathway Analysis tool. Each hexagon represents a metabolite. 53
Figure 2.11: Mediation Analysis. 56
Figure 3.1: Schematic representation of methods for this study. 61
Figure 3.2: Box and Whisker plot depicting relative abundance of significant phylum. (A) Firmicutes and (B) Proteobacteria (C) Shannon Diversity (α diversity index) (D) Beta Diversity (unweighted Unifrac distances) in undernourished and nourished from rural and urban-slum settings. p-values indicate statistical significance from independent t-tests comparing nourished and undernourished study groups in both the locations (i.e. rural normal vs rural undernourished and urban normal vs urban-undernourished) and independent t-tests based on location (i.e. rural vs urban-slum). Box-and-whisker plots show high, low, and median values, with lower and upper edges of each box denoting first and third quartiles, respectively. ns indicates non-significant p-values. Black dots represent the outliers..... 69
Figure 3.3: Pie Chart representation of the relative abundance of the most dominant bacterial phyla. A) Across all the study groups, B) Across the undernourished and nourished from rural and urban slum settings, C) Across the location (rural and urban slum). 70
Figure 3.4: Box and Whisker plot depicting Beta Diversity (unweighted Unifrac distances) based on location. p-values indicate independent t-tests statistical significance comparison between nourished and undernourished study groups in both the locations (i.e. rural vs urban slum) and then categorizing the study groups based on the location. Box-and-whisker plots show high, low, and median values, with lower and upper edges of each box denoting first and third quartiles, respectively. ns indicates non-significant p-values. Black dots represent the outliers. 71
Figure 3.5: Principal Coordinates Analysis (PCoA) of the similarity across the study groups based on the nutritional status (nourished vs undernourished) and location (rural vs urban slum) based on the Unweighted UniFrac distances. Each point corresponds to a sample coloured according to the nutritional status and corresponding

location and the size of the point corresponds to the Shannon diversity index. The principal Coordinates (PC) axis expresses the percent variance across the samples... 72
Figure 3.6: Box and Whisker plot depicting the relative abundance of significant taxa associated with undernourished and nourished from urban slum and rural settings. p-values indicate statistical significance from independent t-tests comparing nourished and undernourished study groups in both the locations (i.e. rural normal vs rural undernourished and urban normal vs urban-undernourished) and independent t-tests based on location (i.e. rural vs urban-slum). Box-and-whisker plots show high, low, and median values, with lower and upper edges of each box denoting first and third quartiles, respectively. ns indicates non-significant p-values. Black dots represent the outliers. 73
Figure 3.7: Box and Whisker plot depicting the pathways significantly associated with undernourished and nourished from urban slum and rural settings. (A) Carbohydrate metabolism ($p < 0.001$), (B) Lipid metabolism ($p = 0.006$), (C) Biosynthesis of secondary metabolites ($p < 0.001$). p-values indicate statistical significance from independent t-tests comparing nourished and undernourished study groups in both the locations (i.e. rural normal vs rural undernourished and urban normal vs urban-undernourished) and independent t-tests based on location (i.e. rural vs urban-slum). Box-and-whisker plots show high, low, and median values, with lower and upper edges of each box denoting first and third quartiles, respectively. ns indicates non-significant p-values. Black dots represent the outliers..... 76
Figure 4.1: Flow diagram showing study design, conduct and analysis. For metabolic signature analysis, associations with pain and pain traits were compared between two independent studies where the Webex study was used as test dataset and KPIC study was used as replication dataset. For gut microbiome signatures in knee pain, associations were tested for gut microbiome data from only the Webex data set. 86
Figure 4.2: Metabolites associated with Pain scores in the Webex and KPIC study datasets. Circos plot depicting the metabolites significantly (FDR < 0.1) associated with current knee pain in either or both the datasets (Webex as red bars and KPIC as blue bars). The bars facing outward depicts the positive associations and bars facing inward depicts the negative associations ranging within +0.6 to -0.6. The data were inverse normalized and then regressed using a linear model 1. FDR values <0.05 are marked with an asterisk. 92
Figure 4.3: Metabolites associated with Pain in test (Webex) and replicate (KPIC) cohort. Heatmap depicting the significant metabolites ($p \leq 0.05$) associated with current knee pain scores in two study datasets. The data was inverse normalized and then regressed using a linear model 1. Blue to red highlighted values indicates the beta coefficients from negative to positive associations. p-values in parenthesis indicate the statistical significance of the associations. 95
Figure 4.4: Meta-analysis of test (Webex) and replicate (KPIC) datasets. The forest plots A. Acyl Ornithine, B. Carnosine, C. Cortisone, D. Cortisol, E. Cystine, F. DOPA, G. GLCAS, H. PEA, I. Succinic Acid depict the metabolites significantly associated with pain in replicate (KPIC) and test (Webex) datasets. Values from linear regression model 2 were inputted for the fixed random effect model-based meta-analysis. p-values indicate the statistical significance with FDR < 0.1. 100
Figure 4.5: Relevant pathways involved in current knee pain based on the significantly associated metabolites from the meta-analysis. Scatter plot of the pathways enriched due to the significant metabolites. The x-axis and size of the node

depict the impact of pathways from topology analysis and the y-axis and colour of the node depict the statistical significance in terms of the log_{10}(p-values)..... 101
Figure 4.6: Correlation between replicated metabolites and circulating cytokine levels. Heatmap depicting the correlation of cytokines levels in Webex dataset with the replicated metabolites in the Webex and KPIC datasets from meta-analysis. Blue to red highlighted values indicates the spearman's coefficients from negative to positive associations. p-values in parenthesis indicate the statistical significance of the associations after multiple testing correction (FDR ≤0.1) and are marked with an asterisk. 104
Figure 4.7: Regression of microbiome with pain and PPTs in the Webex dataset. Heatmap depicting the significantly (p-value < 0.05) associated microbial taxa with the pain and PPTs in the Webex dataset. Blue to red highlighted values indicates the regression coefficients from negative to positive associations. p-values in parenthesis indicate the statistical significance and are marked with an asterisk (*** p-value ≤ 0.001, **p-value ≤ 0.05, *p-value ≤ 0.1)..... 108
Figure 4.8: Associations between metabolites, microbes and Pain. Blue coloured circles represent the metabolites. Pain trait is represented as grey coloured circle and the yellow circles represent the microbes. Red colour lines indicate the negative correlations, while green colour lines indicate positive correlation. Dashed lines indicate the association with p-value ≤0.05 while solid lines indicate the associations with p-value ≤0.001. 109
Figure 5.1: Study design and Methodology: This figure presents the consort diagram for the study. *= adjusted for age, gender and BMI, ** = adjusted for age, gender and BMI for rarefied and Inverse Normalised Microbiome or Metabolome or diet data, ^ = Inverse Normalised data, ^^ Energy adjusted inverse normalised data, ^^^ = *Collinsella* or *Bifidobacterium* as a mediator, diet as predictor and disease status as outcome and model adjusted for age, gender and BMI. HV = Healthy Controls. 122
Figure 5.2: Differentially abundant microbes in NAFLD patients when compared with Healthy Controls. The y-axis presents the log of relative abundance of the microbes. 127
Figure 5.3: Forest plots of odds ratio (95% CI) of the significant microbes (FDR <0.05) associated with NAFLD when compared with HV. 129
Figure 5.4: OPLS-DA for Healthy vs NAFLD participants. The x-axis presents the t1 score (predicted principal components-score value of main components) representing the difference between HV and NAFLD participants (intergroup). The y axis presents to1 score (orthogonal principal component-score value of orthogonal components) representing the difference within HV and NAFLD participants (intragroup)..... 131
Figure 5.5: Variable Importance Projection: The x-axis presents the Variable Importance Scores from the OPLS-discriminant analysis model comparing HV versus NAFLD participants for the metabolites presented on the y-axis..... 131
Figure 5.6: OPLS-DA for NASH vs CIRRHOSIS participants. The x-axis presents t1 score (predicted principal components-score value of main components) representing the difference between HV and NAFLD participants. The y axis presents to1 score (orthogonal principal component-score value of orthogonal components) representing the difference within HV and NAFLD participants..... 134
Figure 5.7: Forest plots of odds ratio (95% CI) of the significant metabolites (FDR <0.05) associated with NAFLD when compared with HV...... 136
Figure 5.8: Micronutrients showing significantly different intakes. The y-axis represents the absolute intake in milligrams except for Selenium which is in

micrograms. Box-and-whisker plots show high, low, and median values, with lower and upper edges of each box denoting first and third quartiles, respectively. Asterisk indicates statistical significance from independent t-tests comparing energy adjusted inverse normalised intakes in NAFLD vs HV. Black dots represent the outliers. *** pval ≤ 0.001, ** pval < 0.05. 143
Figure 5.9: Macronutrients showing significantly different intake. The y-axis represents the absolute intake in grams. Box-and-whisker plots show high, low, and median values, with lower and upper edges of each box denoting first and third quartiles, respectively. Asterisk indicates statistical significance from independent t-tests comparing energy adjusted inverse normalised intakes in NAFLD vs HV. Black dots represent the outliers. *** p-val ≤ 0.001, ** p-val ≤ 0.05. 144
Figure 5.10: Food groups showing significantly different intakes. The y-axis represents the absolute intake in grams. Box-and-whisker plots show high, low, and median values, with lower and upper edges of each box denoting first and third quartiles, respectively. Asterisk indicates statistical significance from independent t-tests comparing intakes in NAFLD vs HV. Black dots represent the outliers. *** p-val ≤ 0.001, ** p-val ≤ 0.05. 145
Figure 5.11: Cross-validation plot representing the lambda for the LASSO regression model. The x-axis represents the log-transformed values of the penalty term and the y-axis represents the deviation in the predicted model. The dashed vertical lines present the value of lambda at which the binomial deviance is minimum and the value of the largest lambda at which the binomial deviance is within 1 standard error. 149
Figure 5.12: ROC curve for the LASSO regression model. The x-axis presents the sensitivity of the model in terms of false positive rate and the y -axis presents the specificity of the model in terms of true positive rate. The grey colour lines present the ROC curves for 100 random models computed for the analysis. The red colour line presents the average of the ROC curve. 151
Figure 5.13: Causal Mediation Analysis Model: Mediation analysis between dietary intake (food groups, macronutrients and micronutrients) and disease outcome. Data in the tables is expressed as the ACME or ADE estimate scores and the p values are presented in parenthesis. Significant values are marked with asterisk (*** p-value ≤ 0.001, ** p-value ≤ 0.05). 152
Figure 6.1: Schematic representation of conceptual framework, key findings related to dietary factors, microbiome and as well as metabolome and interpretations from the studies conducted in this thesis. Red colour boxes represent complex dietary components, green coloured boxes represent simpler form of dietary components, yellow boxes represent bacterial phyla, bright green colour boxes represent the metabolites, purple coloured boxes represent the metabolic pathways. Boxes and arrows with dashed outline represent the metabolites or pathway not analysed for the particular study but suggest the mechanistic insights towards the disease/condition. 162

LIST OF TABLES

Table 1.1 Summary of the abundance of dominant bacterial phyla based on different factors. 7
Table 2.1: Overview of the Study Cohorts in this thesis. 28
Table 2.2: Details of the changes made in TWINS-FFQ to make it comparable to the EPIC-Norfolk FFQs. 49

Table 3.1: Characteristics of the study participants based on nutritional status (normal vs undernourished) and location (rural vs urban slum).....67
Table 3.2: Characteristics of the study participants based on the location (rural vs urban slum).....68
Table 3.3: Abundances of taxa at the family level based on location (i.e. rural vs urban slum).....71
Table 3.4: Correlation between the pathways identified and the study groups based on the nutritional status (normal vs undernourished) or location (rural vs urban slum).....75
Table 4.1: Anthropometric and Clinical Characteristics of the participants in study datasets.....91
Table 4.2: Metabolites associated with Pain in Webex(test) and KPIC (replicate) study datasets.....94
Table 4.3: Descriptive of significant serum metabolites (FDR < 0.1) from linear model 1......94
Table 4.4: Metabolites associated with Pressure Pain detection Thresholds (PPTs) in Webex and KPIC study datasets.....97
Table 4.5: Meta-analysis of metabolites significantly associated with current knee pain in the KPIC and Webex cohorts.....100
Table 4.6: Pathway Analysis of significant metabolites associated with knee pain in meta-analysis.....102
Table 4.7: Lack of association between circulating cytokine levels and Pain scores and Pressure Pain detection Thresholds (PPTs) in Webex study.....103
Table 4.8: Correlation of serum cytokines and replicated metabolites from meta-analysis of two studies in the Webex study......106
Table 4.9: Association of gut microbes and metabolites.....110
Table 5.1: Clinical Characteristics of participants in the study.....126
Table 5.2: Association of gut microbes and disease status (HV vs NAFLD).....128
Table 5.3: VIP score. This table presents the metabolites with variable importance scores (VIP) >1.....133
Table 5.4: Association of metabolites and disease status (HV vs NAFLD).....139
Table 5.5: Associations between significant metabolites reported to be associated in this study and Bifidobacterium or Collinsella.....142
Table 5.6: Descriptive of Absolute Dietary Nutrients intake in Healthy and NAFLD participants.....148
Table 5.7: The estimated coefficients for logistic least absolute shrinkage and selection operator (LASSO) regression between dietary data including the covariates (age, gender and BMI).....150

ABBREVIATIONS

ACME	Average Causal Mediation Effect
ADE	Average Direct Effect
AIH	Auto-immune Hepatitis
ALD	Alcohol Dehydrogenase
ALT	Alanine Transaminase
ANCOM	Analysis of Composition of Microbiome
ANOVA	Analysis of Variance Analysis
ASH	Alcoholic steatohepatitis
AST	Aspartate Amino Transferase
ASV	Amplicon Sequence Variant
AUC	Area Under the Curve
BCAA	Branched Chain Amino Acid
BLAST	Basic Local Alignment Search Tool
BMI	Basal Metabolic Index
CAP	Controlled Attenuated Parameter
CAZyme	Carbohydrate Activating Enzymes
CI	Confidence Interval
CSS	Centred Sum Scaling
DA	Differential Abundance
DALY	Disability Adjusted Life Years
DILI	Drug Induced Liver Injury
DNA	Deoxyribose Nucleic Acid
EDTA	Ethylene Diamine Tetra Acetic Acid
EPIC	European Prospective Investigation into Cancer
FA	Fatty Acids
FAS	Fatty Acid Synthetase
FDR	False Discovery Rate

FETA	FFQ EPIC Tool for Analysis
FFQ	Food Frequency Questionnaire
FMT	Faecal Microbiota Transplantation
GGT	gamma-glutamyl transferase
GI	Glycaemic index
GP	General Practitioner
HV	Healthy Controls
HCC	Hepato-cellular Carcinoma
HDL	High Density Lipoprotein
HMP	Human Microbiome Project
HT	High Throughput
HTS	High Throughput Sequencing
IBD	Inflammatory Bowel Syndrome
IFN	Interferon
IL	Interleukin
KEGG	Kyoto Encyclopaedia for Genes and Genomes
KL	Kellgren Lawrence
KO	KEGG Orthologs
KPIC	Knee Pain in Community
LASSO	Least Absolute Shrinkage and Selector Operator
LC	Liquid Chromatography
LCAT	lecithin cholesterol acyl transferase
LCI	Lower limit Confidence Interval
LDL	Low Density Lipoprotein
LME	Linear Mixed Effect
LMIC	Low- and Middle-Income Countries
LPS	Lipopolysaccharides
LV	Latent Variables
MAFLD	Metabolic Associated Fatty Liver Disease

ML	Machine Learning
MRI	Magnetic Resonance Imaging
MS	Mass Spectroscopy
MSD	Musculoskeletal Disorder
MWAS	Metagenome Wide Association Study
NASH	Non-Alcoholic Steatohepatitis
NCBI	National Centre of Biotechnology
NGS	Next Generation Sequencing
NMR	Nuclear Magnetic Resonance
NNM	National Nutrition Mission
NRS	Numerical rating Scale
NSAID	Non-Steroidal Anti-inflammatory Drugs
OA	Osteoarthritis
OPLS	Orthogonal Partial Least Square
ORA	Over-representation Analysis
OTU	Operational Taxonomic Unit
PA	Pathway Analysis
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PDFF	Proton Density Fat Fraction
PEA	Phenylethylamine
PPM	Parts per Million
PPT	Pressure Pain Detection threshold
QIIME	Quantitative Insights into Microbial Ecology
QTL	Quantitative Trait Loci
RDP	Ribosomal Database Project
REML	Restricted Maximum Likelihood
RMSEE	Root Mean Square Error Estimate
RNA	Ribonucleic Acid

ROC	Receiver Operating Curve
RUTF	Ready to use Therapeutic Foods
SCFA	Short Chain Fatty Acids
SD	Standard Deviation
SNP	Single Nucleotide Polymorphism
TNF	Tumor Necrosis Factor
TSS	Total Sum Scaling
UCI	Upper limit Confidence Interval
USG	Ultra-sonography
VIP	Variable Importance Score
VLDL	Very Low-Density Lipoprotein
WEBEX	Web-based Exercise
WHO	World Health Organisation

1

Chapter 1
GUT MICROBIOME AND METABOLOME: AN INTRODUCTION

CONCEPTS, DYNAMICS, INFLUENCE OF EXTERNAL AND INTERNAL FACTORS, ROLE IN HEALTH AND DISEASE

1.1 CONCEPTS OF GUT MICROBIOME AND METABOLOME

The multitude of microbes residing in the human gastrointestinal tract which have co-evolved over time with humans is referred to as the gut microbiome. Microbiome comprises of the bacteria, virus, fungi, archaea. But most commonly, the microbiome term has been often implied in reference to the bacterial domain as there were well-established protocols for screening, isolation, and characterization of bacteria from various ecosystems. Previously bacteria were considered to be constituting 99% of the microbiome **(Turnbaugh *et al.*, 2007; Huttenhower *et al.*, 2012),** however, researchers have not studied the other microbial domains of life inhabiting humans back then. The concept of a 99% ratio of the bacterial community in the microbiome has been recently challenged by emerging technologies. The claim has already been assessed and revisited by researchers and it is reported that viruses outnumber bacteria by ten-fold **(Breitbart *et al.*, 2003; Minot *et al.*, 2013; Manrique *et al.*, 2016)** and fungus constitutes 0.1% of microbiome collectively. Virus and fungus contribute to many major metabolic functions but remained relatively understated **(Qin *et al.*, 2010; Cui, Morris and Ghedin, 2013; Huffnagle and Noverr, 2013)** and this has heightened the need for new 'omics' field i.e. virome and mycobiome. This thesis has focussed on the bacterial domain and the term microbiome refers to the bacterial taxa of the microbiome unless specified. The bacterial kingdom is further classified as phylum, class, order, family, genus, and species.

Microbes influence human health by maintaining either a mutual, symbiotic, or pathogenic relationship **(Eloe-Fadrosh and Rasko, 2013).** Historically, the role of microbes was assumed to be only in the perspective of causing infections and diseases. However, microbes are fundamental to maintain gut permeability and thus provide a better immune system and health.

The growing body of literature indicates the role of the gut microbiome in various complex diseases such as fatty liver diseases, inflammatory bowel diseases (IBD), diabetes, malnourishment, cognitive diseases and various other metabolic syndromes **(Sanz *et al.*, 2015; Marchesi *et al.*, 2016; Dabke, Hendrick and Devkota, 2019)**. These and numerous other studies have

provided information about the change in diversity (the amount of individual microbe from each of the microbial species present in microbiome environment), abundance (the count of the specific microbe in the microbiome environment), and composition (the individual microbes constituting the microbial environment) of the microbiome in and across various populations based on demography, ethnicity, diseases, disorders and other factors. Even after years of extensive research studies, it is still not very clear whether the change in the microbial composition is a cause for or is the effect of various diseases.

The gut microbiome produces and also influences the small compounds present within the human cells, tissues, organs, and biofluids (**Wikoff *et al.*, 2009**). These small compounds are called the microbial metabolites which together with the human metabolites are referred to as the metabolome. Metabolome analysis allows the identification and characterization of the metabolites deciphering the interactions and functional status of the gut microbiome and host cells (**Marcobal *et al.*, 2013**). This eventually reveals the molecular pathways and mechanisms of the disease state and supports discovering biomarkers associated with host physiology. The host metabolome shares many functions with the microbiome's metabolic activities and collectively forms a component of many metabolic pathways which are then either upregulated or downregulated in the various diseases (**Agus, Clément and Sokol, 2021**).

This interplay between the microbiome and host metabolome is the key regulator of host homeostasis **(McHardy *et al.*, 2013; Zierer *et al.*, 2018)**. The presence of a vast functional and biochemical, precisely pool of metabolites in humans has long been a major area of interest in shifting the concept of prognosis and therapeutics for nearly all sorts of diseases and disorders **(Lemon *et al.*, 2012; Vázquez-Baeza *et al.*, 2018)**.

Published research on the microbiome and metabolome to date addresses immunity, genetics and extrinsic factors such as diet, exercise, environment and drugs (**Cotillard *et al.*, 2013; Castro *et al.*, 2019**) collectively impacting human health via altered host physiology. This is outlined in **Figure 1.1**. With the development of high throughput sequencing (16S rRNA sequencing, shotgun metagenomics, whole-genome sequencing), studying the roles of commensal

microbes has expanded. Recently, advanced multi-omics analytical profiling techniques such as metabolomics and bioinformatics analysis of huge datasets are pouring into this research field with key questions linked to the understanding of,

- The effect of interactions between host factors and microbiome
- The effect of diet and exercise on the microbiome and metabolome
- Which is more influential - the human metabolome or the microbiome metabolome and other similar questions **(Cotillard *et al.*, 2013; Castro *et al.*, 2019; Fan and Pedersen, 2021)**.

With the advancement in the field of high throughput omics technologies such as microbiome sequencing and metabolic profiling, studying the role of microbes and metabolites has been expanded over time and has paved for better research approaches.

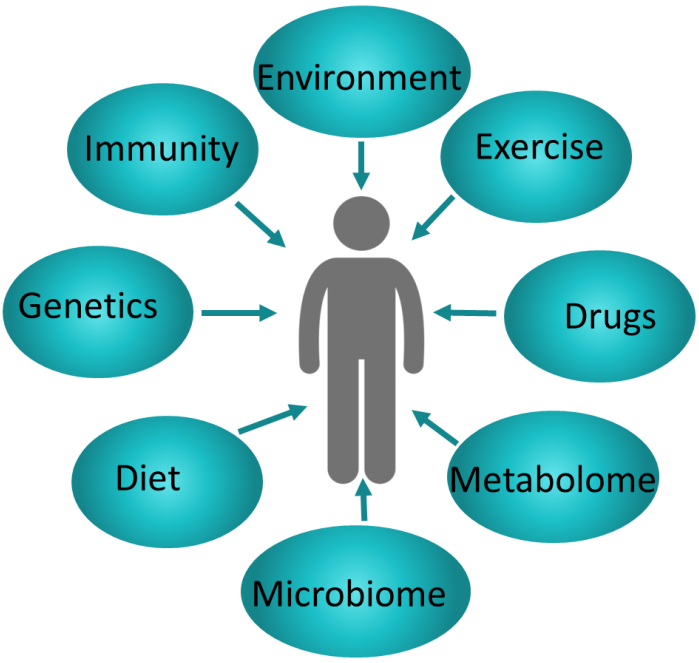


Figure 1.1: Factors associated to maintain homeostasis and being accountable for human health. The figure depicts that there are various intrinsic (host) and extrinsic factors that affect human health. Host factors like genetics and immunity are responsible for directly affecting the host depending upon the environment/demography, the host has been staying in. External factors such as diet, exposure to drugs, exercises and other factors such as gut microbiome, metabolome, also affects human health.

1.2 DYNAMICS OF GUT MICROBIOME AND METABOLOME

1.2.1 GUT MICROBIOME: COMPOSITION AND VARIABILITY

Intra-person variation is a fundamental feature of the gut microbiome making microbial composition unique to an individual **(Turnbaugh, Hamady, *et al.*, 2009**). The microbial composition changes temporally depending upon age, diet, hygiene, and spatially depending upon location and travel. There is no concrete definition of a healthy gut microbiome, however, a healthy individual's gut microbiome has Firmicutes, Bacteroidetes, Proteobacteria, and Actinobacteria as the most dominant phyla (**Mahowald *et al.*, 2009**).

The first line of colonizing bacteria in humans depends on various factors such as mode of delivery i.e. vaginally delivered or caesarean delivered, breastfed or formula-fed infants, gestational age, type of food during the weaning period **(Stearns *et al.*, 2017; Reyman *et al.*, 2019; Wang *et al.*, 2020)**. Microbial composition changes drastically in the first few weeks **(Bäckhed *et al.*, 2015)** and frequently during the first three years of life **(Yatsunenko *et al.*, 2012**) and becomes considerably stable over the years as in adults **(Radjabzadeh *et al.*, 2020)**. Facultative microbes such as *E.coli* and *Enterococcus* are the first to colonize the gut of the new-born infants and later shift towards anaerobic microbes such as *Bifidobacterium* and *Enterobacter* **(Favier *et al.*, 2002)**. The presence of microbes such as *Bifidobacterium*, *Staphylococcus*, and *Lactobacillus* converts breast milk oligosaccharides into short-chain fatty acids (SCFA's) lowering the pH. These factors are favourable for the colonization of beneficial bacteria by creating a symbiotic environment and first line of defence against pathogens **(Cox *et al.*, 2009; Park *et al.*, 2015**). Also, these microbes help in generating hypoxic conditions in the gut by utilizing oxygen and generating an anaerobic environment in gut pockets **(Palmer *et al.*, 2007; Hill *et al.*, 2017)**. These fight against the growth of various enteric pathogens and enhance the growth of Firmicutes and Bacteroidetes in the mature gut (Mahowald *et al.*, 2009). A previous study by Agans. R *et al* **(Agans *et al.*, 2011)** reported that the *Ruminococcus*, *Fecalibacterium*, and *Roseburia* remain the same among adolescents and adults serving as the core microbiome in two cohorts, however, there was a statistically higher abundance of *Bifidobacterium*

and *Clostridium* in adolescents compared to adults. A recent study by Radjabzadeh *et al* **(Radjabzadeh *et al.*, 2020)**, compared the gut microbiome of children and adults and showed that children had significantly lower diversity than adults. The dominance of catabolic pathways in children and biosynthetic pathways in adults indicates that gut microbiota becomes stable in later years of life.

Geographical location and ethnicity are also reported to be determining factors for the composition of the microbiome. These are linked to temperature, access to water, and nutrients, so indirectly these factors are considered to have an impact on gut microbial composition. A study by Nam *et al* **(Nam *et al.*, 2011)** compared the gut microbiome of the Korean population with the population from the United States of America, China, and Japan. Intra-country differences in microbial composition were less in comparison to inter-country differences. The gut of the Korean and Chinese populations seemed to be slightly overlapped showing the dominance of Bacteroidetes. Americans showed a higher abundance of Firmicutes and the Japanese had Actinobacteria as the dominant phylum. A more recent large study conducted on four ethnicities in the United States (**Brooks *et al.*, 2018**) revealed the differences in the microbial community among African Americans, Asian - Pacific Islanders, Caucasians, and Hispanics similar to the Human Microbiome Project (HMP) cohort. Another study conducted on immigrants belonging to different ethnicities in the population of the Netherlands **(Deschasaux *et al.*, 2018)**, suggests that inter-individual diversity depends upon the ethnicities more than the place of residence. The conclusion of this study however showed contrasting results to another study conducted on immigrants in the USA **(Vangay *et al.*, 2018)** which reports that the gut microbiome steadily and over time has adjusted comparable to the native population. The key differences in these studies may be attributed to various change in routine related factors such as exercises, stress, exposure to industrial effluents etc. that has not been considered in these studies. A recent study from the Dutch Microbiome Project **(Gacesa *et al.*, 2022**) involving 8,208 individuals (clustering of 4,745 people into 2,756 families) indicates that microbial taxa is mainly shaped by the environment and cohabitation (~49%) and is least

impacted by the heritability (~7%). This is apparently associated with the childhood environment, parental smoking and socio-economic status.

Taken together, these studies support the notion that mode of delivery, age, genetics, environment, demography, diet, physical activities and socioeconomic status are the factors affecting the structure of the gut microbiome. In addition to these, studies have reported the role of stress/anxiety, medications, xenobiotics exposure and alcohol consumption in shaping the gut microbial communities.

Mode of Delivery *	**Weaning Mode****
• VD: *Bifidobacterium* ↑ • CD: *Enterococcus* ↑ *Klebsiella* spp.↑	• BF: *Veillonella* ↑ *Escherichia /Shigella* ↑ • FF: *Streptococcus* ↑
Infants***	**Adults******
Bifidobacterium ↑ *Lactobacillus* ↑	*Bifidobacterium* ↓ *Blautia* ↑
Ethnicity^	**Environment^^**
• Caucasians: *Christensenellaceae* ↑ • Hispanics: Clostridales ↓ • Islanders: Odoribacteriaceae ↓ • African American: *Veillonella* ↑	• Rural: *Bacteroides* ↑ *Vibrio* ↑ • Urban: *Bifidobacterium* ↑ *Prevotella* ↓

Abbreviations: VD= Vaginally Delivered; CD= Caesarean Delivered; BF= BreastFeed; FF= Formula Feed. Source: * (**Reyman *et al.*, 2019**); ** **(Wang *et al.*, 2020**); ***(**Bäckhed *et al.*, 2015**); **** **(Yatsunenko *et al.*, 2012; Radjabzadeh *et al.*, 2020**); ^(**Reynolds, Kuska and Curtis, 2012**); ^^ **(Das *et al.*, 2018; Hansen *et al.*, 2019**)

Table 1.1 Summary of the abundance of dominant bacterial phyla based on different factors. This table shows the dynamics of the gut microbiome during the growth phases of humans and also based on the location and ethnicity.

1.2.2 METABOLOME AND THE MICROBIOME

The gut microbes have evolved with greater metabolic potential and versatile signalling mechanisms due to the trillion of different genes than humans as discussed in the review by Rowland *et al*. (**Rowland *et al.*, 2018**). Studies over the past two decades have shown that microbial genes code for certain essential

biological functions of producing and converting dietary nutrients and other compounds into essential biochemical compounds called metabolites. Many studies have postulated that these metabolites serve as the substrate and/or by-products of diverse physiological and chemical pathways and are considered to be responsible to maintain homeostasis assisting various metabolic activities. The mechanisms such as energy metabolism, lipid metabolism, and biosynthesis of secondary metabolites contribute to the major interactions between the microbes and the host **(Bäckhed *et al.*, 2004; Velagapudi *et al.*, 2010; Ghazalpour *et al.*, 2016)**. Microbiome composition and health status have been mediated through these molecular and signalling pathways **(Marcobal *et al.*, 2013**). Several studies have reported that an altered microbiome affects the physiological functions of the gut resulting in metabolic disorders such as obesity, diabetes and fatty liver (**Cope, Risby and Diehl, 2000; Dumas *et al.*, 2006; Ridaura *et al.*, 2013**)

Studies comparing germ-free mice and conventional mice have shown close links between the microbiome composition and the metabolome (**Matsumoto *et al.*, 2012**). A study showed a shift in the gut metabolome after the introduction of one or two bacterial species in germ-free mice (**Marcobal *et al.*, 2013**). Similar studies involving human cecum samples suggest that microbial community is predictive of the metabolite pools (**McHardy *et al.*, 2013**). Another population-based study from the Twins UK cohort reflects the association of faecal metabolome with the gut microbial composition accounting 67.7% of the variance in comparison to host genetics which accounted for only 17.9% of the variance **(Zierer *et al.*, 2018**). There is a clear relationship between disease phenotype, gut microbiome structure and metabolome depicting their interplay as a major factor in contributing to health and disease. This is evident from a research study conducted by Jansson *et al*. (**Jansson *et al.*, 2009**) which showed various metabolites to be positively or negatively associated with the gut microbiome and Crohn's disease (CD) status. The microbes in the gut performs functions that humans have not had evolved on their own. There are reports **(Bäckhed *et al.*, 2004; Krajmalnik-Brown *et al.*, 2012)** about the impact of the microbiome on key metabolic functions such as nutrient harvest, immune modulation, and maturity, protection against pathogens, drug

metabolism, blood glucose levels, and brain-behaviour. This is implicated via production of various metabolites by these microbes which are linked to molecular and signalling pathways contributing directly or indirectly to human health.

A considerable amount of literature has been published describing that the main metabolite class produced and transformed by the gut microbiome are short-chain fatty acids (SCFAs) such as butyrate, acetate and propionate, organic acids such as pyruvate and lactate, branched-chain amino acids (BCAAs) such as iso-leucine and valine, fatty acids (FAs) such as phospholipids, neurotransmitters such as GABA and dopamine, vitamins and others **(Heijtz *et al.*, 2011; Blanton *et al.*, 2016; Gu and Li, 2016)**. A schematic representation is depicted in **Figure 1.2**. These metabolites enter the host tissues, cells, organs, and bloodstream becoming an integral part of the host systemic biology. In a mouse model, SCFAs were shown to inhibit autophagy in the colonocytes (**Donohoe *et al.*, 2011**). SCFAs are also known to modulate immune responses through chemokine production and neutrophil recruitment **(Cox *et al.*, 2009; Park *et al.*, 2015**). More recently, SCFAs are found to be associated with the neuropathic pain by regulating microglial activation and pro-inflammatory phenotype polarization (**Zhou *et al.*, 2021; Li *et al.*, 2022**). Many indirect or secondary mechanisms such as the production of bile acids, indole propionic acid (IPA), endocannabinoids (ECs), etc. are reliant on the gut microbiome **(Ridaura *et al.*, 2013; Menni *et al.*, 2017, 2019)**. A study has reported that bile acid metabolism disruption is associated with *C. difficile* infection (**Buffie *et al.*, 2015**) and administration of a bile-acid hydroxylating intestinal bacterium enhances the resistance to infection via secondary bile -acid metabolic pathways. In another study, bile acid composition analysis showed decreased bile acid diversity in germ-free mice when compared with conventionally raised mice **(Sayin *et al.*, 2013**). Bile acids have also found to influence the inflammation in the gut via regulating gut-residing T cells in diseases such as inflammatory bowel disease (IBD) **(Duboc *et al.*, 2013**). More recently, two derivatives of lithocholic acid (LCA), 3-oxoLCA and isoalloLCA have been shown to inhibit the differentiation of T_H17 by binding to RORγt and promoting differentiation of T_{reg} via upregulating FoxP3 expression **(Hang *et al.*, 2019**). Endocannabinoids

(ECs), the lipid-based endogenous metabolites that act as neurotransmitters by binding to the cannabinoid receptors are also reported to be modulating the gut permeability in the cell culture models through the increased expression of mRNA's of tight junctions proteins such as ZO-1, claudin-1 and occludin and thus regulation of adipose tissue plasticity that is causing the reversible and flexible changes to the cellular characteristics restricted to adipose tissue altering cell permeability (**Alhamoruni *et al.*, 2010; Muccioli *et al.*, 2010**). Another recent study showed that the anti-inflammatory effects of SCFAs are partly mediated by the endocannabinoid system **(Vijay, Kouraki, *et al.*, 2021**).

Data from several other studies (**Zhai *et al.*, 2010; Daghestani, Pieper and Kraus, 2015; Ajmera *et al.*, 2017; X. Zhou *et al.*, 2020**) suggest that metabolic profiling could be a prognostic marker of the disease in the future. Despite extensive research, there is still a significant fraction of metabolites that need to be characterised and interlinked to microbial metabolism. The systematic study of microbiome-conferred metabolome requires detailed further research.

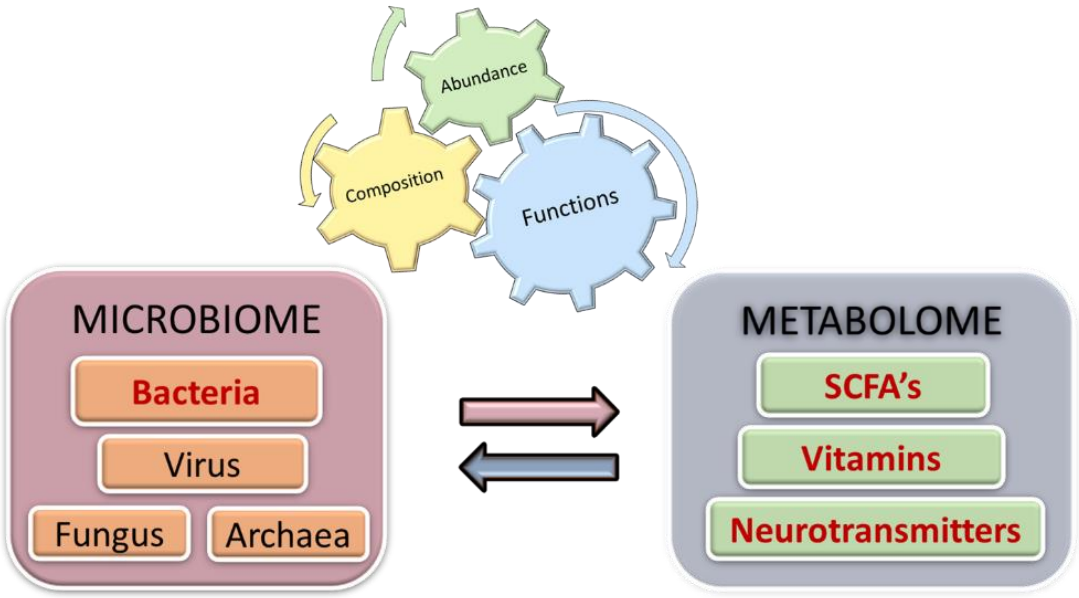


Figure 1.2: Link between Microbiome and Metabolome. This figure depicts that there is an interplay of the gut microbiome and metabolome depending upon the composition, abundance and function of the gut microbes.

1.2.3 INFLUENCE OF OTHER FACTORS

1.2.3.1 DIET AND NUTRITION

Diet and nutrition are essential external factors and known risk modulators for various diseases impacting human health. Microbes degrade the complex dietary compounds and produce micro or macronutrients. For instance, microbes have around 250 homologs of CAZymes, which are not encoded by humans and that convert the indigestible form of dietary components into the forms absorbed by humans **(Kaoutari *et al.*, 2013)**. As already discussed in previous sections, microbes assimilate these compounds and convert them to other metabolites such as SCFAs, vitamins, neurotransmitters and amino acids which may act as energy substrates or immune-modulators **(Diamond *et al.*, 2011)**. Diet has been shown to have a profound effect on the microbial communities affecting their number, composition, and functional repertoire. A study **(Davenport *et al.*, 2014)** investigating seasonal diet in healthy individuals, showed *Bacteroides* increased, while Actinobacteria reduced in the summer. Individuals having diet-associated metabolic syndromes such as obesity tend to have fewer Bacteroidetes, which increase with the intake of a low caloric diet associated with weight loss **(Ley *et al.*, 2006)**.

Understanding the microbiome composition and abundance, however, alone is not sufficient to describe the systemic effect of diet and nutrition although the role of microbes in converting the dietary compounds into metabolites is already well established. A strong relationship between microbiome, metabolome and diet has been reported in the dietary intervention study by Russell *et al*. **(Russell *et al.*, 2011**) where the weight loss diet comprising low or moderate carbohydrate and high protein intake reduced the proportion of phenolic compounds along with *Roseburia* and *E. rectale* belonging to Firmicutes phyla microbes and an increase in the phenylacetic acid, and nitroso compounds. Dietary intervention in another animal model study conducted by Turnbaugh *et al*. **(Turnbaugh, Ridaura, *et al.*, 2009)** compared the effect of a low-fat, high fibre versus high-fat, high-sugar diet in gnotobiotic mice reporting the increase of Bacilli and Erypsotrichi class belonging to Firmicutes phyla. Another similar dietary intervention study **(David *et al.*, 2014)** investigating short-term dietary

intake showed a decrease in certain genera of Firmicutes involved in plant-based polysaccharide metabolism and an increase in bile acid-tolerant genera belonging to Bacteroidetes in subjects having animal-based dietary products vs plant-based dietary products. These studies altogether describe that the diet majorly impacts the Firmicutes and Bacteroidetes phyla which majorly contributes to the production of SCFAs such as butyrate, acetate and propionate.

.

However, much of the research has been quantitative but the understanding of the mechanisms behind these is still in its infancy. This is supported by the concept that the genus *Bacteroides* is known to convert the xyloglucans **(Larsbrink *et al.*, 2014; Hemsworth *et al.*, 2016)** with the help of carbohydrate-active enzymes which belongs to the family of glycosyl hydrolases **(Kaoutari *et al.*, 2013; McNulty *et al.*, 2013)**. However, these enzymes have a diverse range of paralogues present in different species which explains why other genera of the same phylogeny level have different catabolic efficacies and utilization of complex polysaccharides.

There are limited studies that have explored the extent of change in metabolic profile with change in the diet. This suggests that there is a need for monitoring the individual's metabolomic responses to the diet interventions longitudinally rather than just referring to the abundance and composition of the microbiome. Recent studies have investigated how dietary intake shapes the microbiome and affects human health by manipulating microbiome-related metabolites. Metabolic profile deciphering nutritional intake could help deepen our knowledge of metabolic pathways relevant to human nutrition (**Gibney *et al.*, 2005**) and this could help to identify the dietary biomarkers in relation to health and disease.

The insights about the molecular signatures, both microbial and metabolomics together may robustly serve as potential indicators of health and disease. Analytical approaches investigating the change in microbiome, microbial metabolites and host metabolites with change in the diet could provide insights into the mechanisms by which diet affects host physiology . This could form the basis for designing robust randomized intervention trials.

1.2.3.2 EXERCISE AND PHYSICAL ACTIVITY

For a very long time, exercise and physical activities have been known to reduce weight, strengthen the heart and improve blood circulation, which directly impacts health and wellbeing. Physical activities help transport oxygen and nutrients across the tissues and body cells which if not assimilated may accumulate to form body fat and thus be one of the reasons for inflammation (**Sarin *et al.*, 2019**). As discussed in earlier sections, the gut microbiome is also crucial in maintaining homeostasis and health mediated via metabolites, so it seems likely that the microbiome and metabolome may be directly or indirectly affected by exercises. Yet there is a lack of information on the effect of physical activities on these two factors. Recently, studies linking exercise with the gut microbiome and metabolome are emerging.

Physical activity is found to enrich microbial flora **(Clarke *et al.*, 2014**) and thus improve the colonization of commensal bacteria, improving health status (**Monda *et al.*, 2017**). Both, exercise and gut microbiome are reported to stabilize gut permeability, optimizing the nutrient flow and thus reducing inflammation (**Keirns *et al.*, 2020**). Previous research has established that exercise increases insulin sensitivity and reduces systemic inflammation via the gut microbiome (**Allen *et al.*, 2018**).

The cross-link between microbiome and metabolome is already discussed in the previous sections, that, how these two factors run in parallel and play an important role towards better health status. However, only a few studies draw combinatorial approach research linking microbiome and exercise-induced change in metabolites. Fecal metabolic profiling of human athletes (**Barton *et al.*, 2018**) showed increase in SCFAs associated with enhanced muscle fitness and overall health when compared with sedentary controls.. Other studies revealed that metabolic functions associated with the administration of certain microbes such as *Veillonella atypica* and *Bifidobacterium longum* enhance performance in athletes **(Scheiman *et al.*, 2019; C. L. Lin *et al.*, 2020**).

There is a need to explore exercise-induced metabolite patterns to understand metabolic health. Future studies on the effect of physical activity on gut

microbiota composition or restoration of optimal microbial composition or metabolome towards the improvement of metabolic disorders hold potential for using these as intervention strategies in medicine.

1.2.3.3 HOST GENETICS AND PHYSIOLOGY

Human physiology depends upon the collective effect of the human genetic factors, microbial genomics residing in the gastrointestinal tract as well as the pool of metabolites. Human genetics is linked to the environment, demographic area (**National Institute of Health (US), 2007; Bind *et al.*, 2014; Davenport *et al.*, 2014**) and food preferences (**Smith *et al.*, 2016**). Early examples of research showed that gene mutations and polymorphisms are associated with the pathogenesis of hereditary diseases, however, this has been limited when it comes to complex and multifactorial diseases as the knowledge about the function of nearly 50% of the human genome is still unknown.

Few studies report strong evidence that host genetics regulates the gut microbiome through the secretion of immunoglobulins (IgA) and miRNA **(Suzuki *et al.*, 2004; Liu *et al.*, 2016)**. A study reports that the amount and type of fat intake change HDL-C concentrations through *LIPIC* gene polymorphism and lipoprotein levels with *ApoE* genotype (**Ordovas *et al.*, 2002; Thifault *et al.*, 2013**). Population-based studies have shown that predisposition to metabolic diseases is linked to genetic differences and mediated via an altered gut microbiome. For instance, various quantitative trait loci (QTL) for *PNPLA3* (**Kravetz *et al.*, 2020**), *SLIT3* are associated with the relative abundance of microbes (**Benson *et al.*, 2010; Goodrich *et al.*, 2016**).

It is debated whether it is the microbiome that shapes human physiology or is it the host factor that shapes the gut microbiome. Several studies are discovering and characterizing the genotypes associated with the change in the gut microbiome or metabolome, however, mechanisms linking SNPs with the metabolic axis are limited. This can be achieved by association analysis of genetic polymorphisms relevant to metabolites and relevant pathways in large-scale studies.

1.3 RESEARCH PROBLEM AND RESEARCH QUESTIONS

After being pioneered by Carl Woos in 1977, 16S rRNA has become a phylogenetic marker for bacterial analysis. Being a conserved gene among the bacteria, 16S rRNA is also embellished with hypervariable regions that provide signature sequences. Right from capillary sequencing technologies such as Sanger sequencing to pyrosequencing based next generation sequencing (NGS) technologies such as Roche 454, to using reversible terminators technology such as Illumina-MiSeq/HiSeq and more recent electric current size detection technology based nanopore sequencing, 16S rRNA sequencing have become the mainstay for the screening, identification and characterization of bacteria for studies involving microbiome-based analysis towards human health.

Initial human microbiome-centred comprehensive cohorts such as HMP **(Huttenhower *et al.*, 2012**) and MetaHIT (**Qin *et al.*, 2010**) were followed over for many years. These studies have provided answers to many questions crucial to human health evidencing that individuals carry specific microbial signatures and that there are differences in the microbial diversities among healthy and diseased states. Research in gut microbiome studies for almost a decade led to the conclusion that the contribution of the microbiome is far beyond just the microbial diversity in the given environment. This has served as a basis for many more questions related to the understanding of mechanistic insights such as interaction with host factors, the impact of the external factors and causal mechanisms.

Despite the differences in the pathophysiology of the various diseases, there is a commonality of being associated with both the shared and disease-specific microbial signatures and functions regulating the metabolism of the body (**Rizzatti *et al.*, 2017**) (). As discussed in previous sections, it is known that the interaction of gut microbes with the host metabolism is regulated mainly via immune modulation, eliminating pathogens, maintaining gut barrier and nutrient conversion from external factors, especially the diet. Dietary intakes alter the gut microbiome and thus also bring change in the metabolites. The metabolites may cross the gut barrier and the ones with inflammatory properties cause the

inflammation of the gut mucosal surface. These metabolites are also released in the bloodstream, triggering the change in the metabolic responses which causes many other metabolic syndromes not directly linked to the gut.

Based on these evidences, various qualitative and quantitative research have previously been designed to provide answers to gaps in microbiome-related studies. In the past decade, individual studies investigating the effect of these factors on various health and disease states have been conducted. However, research over the years has surfaced the need for multi-omics studies connecting these factors across a varied range of disorders to understand the overall systemic effect of the same on human health. A better understanding of the crosstalk between the gut microbiome and contributing factors will pave a way for better treatment therapies. The microbiome or metabolome association studies serve as the biomarker identification basis for diagnosis and prognosis of a varied range of diseases and also as potential therapeutic targets after further validation experimental studies.

1.4 ROLE OF GUT MICROBIOME AND METABOLOME IN SPECIFIC DISEASES/CONDITIONS

The role of the gut microbiome and metabolome is pivotal in human health and is influenced by both extrinsic and intrinsic factors as stated in previous sections of this chapter. However, the relationship between the gut microbiome, metabolome, factors such as diet and disease state is still not clear in the majority of diseases. The high scale 'omics' data with the advent of advanced omics tools and technologies have facilitated understanding the biological complexities and define the novel strategies to treat the diseases. This section will highlight the role of the gut microbiome and metabolome in diseases/conditions investigated in this thesis which includes undernourishment, osteoarthritis and non-alcohol fatty liver disease.

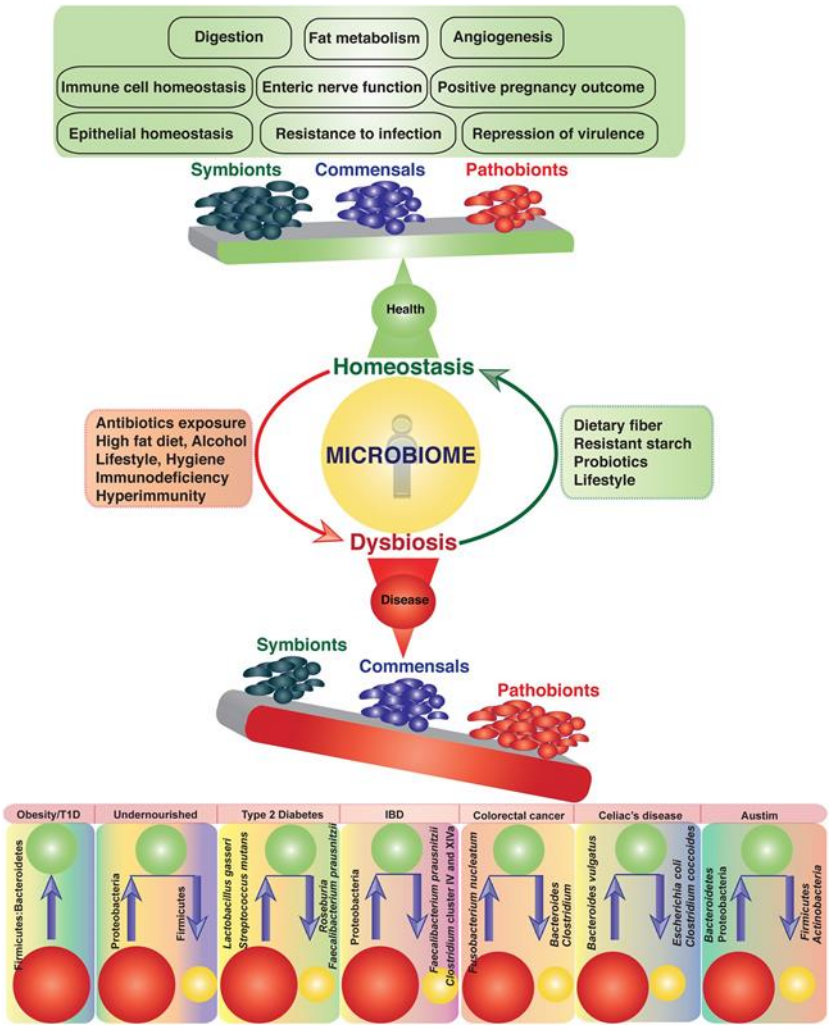


Figure 1.3: The gut microbiome and role in various diseases. This figure (adapted from **Das et al., 2018**) depicts that an altered gut microbiome may induce inflammation and health disorders like IBD, malnutrition, colorectal cancer and several others.

1.4.1 UNDERNOURISHMENT

In broader terms, undernourishment is defined as a nutritional deficiency of energy, protein, and other nutrients causing measurable adverse effects on body composition, function, and clinical outcomes **(Saunders and Smith, 2010; Wells *et al.*, 2020**). As per WHO reports, undernourishment is the leading global disease burden affecting children under the age of five years, estimating 149 million children to be stunted and 45 million to be wasted. Undernutrition in many Low and Middle Income Countries (LMICs) comprises 45% of deaths linked to undernutrition among children under 5 years of age **(*Fact sheets - Malnutrition*, 2022-04-15**). Persistent childhood undernutrition is considered

part of a vicious cycle of recurrent infections, impaired immunity, and worsening undernutrition, compounded by food insecurity and, likely, host genetic factors as well **(Kau *et al.*, 2012).** Undernutrition in adults is multifactorial with causes ranging from inadequate dietary intake to the presence of underlying co-morbidities which may impair the normal assimilation and absorption of nutrients.

Recently, alterations in the gut microbiome have been recognized as part of this cycle **(Gordon et al., 2012; Kau et al., 2012)**. As discussed in the preceding sections, the absorption of dietary nutrients is largely determined by the gut microbiome which if altered impairs immune regulation. There is substantial evidence that indicates a high prevalence of infectious diseases amongst the undernourished and this could be attributed to an impaired immune system (**Bourke, Berkley and Prendergast, 2016**).

Current therapeutics such as ready-to-use therapeutic food (RUTF) prepared with powdered milk, peanuts, vegetable, sugar and mineral supplements **(UNICEF, 2013)** have aided in reducing the percentage of mortality in children. However, this has not helped overcome the risk factors such as hypophosphatemia (Yoshimatsu *et al.*, 2012; Pourhassan *et al.*, 2019) associated with undernutrition during later years of life. Several studies provide evidence that diet improves gut health via colonization of beneficial gut microbes, but research to understand the restoration of a healthy gut microbiome has had limited success. This is attributable to the fact that there is not an absolute definition that which microbes likely represent a healthy gut microbiome.

A research study investigating the impact of undernutrition on the gut microbiome among adults residing at locations separated based on socioeconomic status (rural vs urban slum) in India is presented in Chapter 3.

1.4.2 OSTEOARTHRITIS:

Osteoarthritis (OA) is the most common form of musculoskeletal disorder (MSD) impacting mostly the weight-bearing joints such as knees and feet. One in five people (including children) suffers from musculoskeletal pain **(World Health Organisation, *Accessed 2020-07-19*).** The genetic and epigenetic factors such as sex, ethnicity and age are reported to be associated with OA along

with overweight/obesity, lack of exercise and injuries. Pain, stiffness, and limited movements are the early symptoms of OA. Deterioration of cartilage accompanied by pain and chronic inflammation results in osteoarthritis which progresses to deformity of bones, and total disability at advanced stages. The insights on disease pathology related to musculoskeletal disorders are not entirely understood even though it is the leading contributor to disability.

It is already established that microbes produce a wide range of proinflammatory molecules such as lipopolysaccharides (LPS). LPS is known to be causing inflammation, so it can be hypothesised that the microbiome plays a major role in the pathophysiological mechanisms underlying OA. This is also supported by the presence of microbial DNA and bacteria-based metabolites in the synovial fluid of osteoarthritis and rheumatoid disease patients **(Berthelot *et al.*, 2019) (Zhang *et al.*, 2015**). Studies focused on the altered microbiome in relation to OA have been discussed in reviews previously **(Szychlinska *et al.*, 2019; Favazzo *et al.*, 2020)**.

The systemic effect of the gut microbiome and metabolome as mentioned before has been examined in OA conditions also. For instance, a study on high fat/sucrose diet-induced obese rats found joint degradation was associated with an increased abundance of *Lactobacillus* and *Methanobrevibacter* and an increase in lipopolysaccharides (LPS) levels and body fat **(Collins *et al.*, 2015).** This is supported by associations between macrophage abundance and LPS and LPS-binding protein levels in the synovial fluid of knee OA patients. **(Huang *et al.*, 2016)**. More recently, the first report investigating the association of gut microbiome in human osteoarthritis patients showed that an increase in *Streptococcus* is associated with knee joint pain (**Boer *et al.*, 2019a**). Physical activities, as mentioned in the previous section, impacts the gut microbiome and thus the metabolome. This is also evident from a short-term exercise-based study **(Griffin *et al.*, 2012)** in a high-fat diet-induced obese mice model that reports improvement in glucose tolerance and reduction of the pro-inflammatory cytokines with no association with body fat. There is, however, sparse literature informing about the mechanistic association of metabolites produced by the gut microbiome with osteoarthritis and so there are not any definite results to comment on the direct influence of metabolome in OA.

To date, pain and inflammation are maintained mostly using non-steroidal anti-inflammatory drugs (NSAIDs) which come with side effects such as gastric ulcers, kidney dysfunctions, and others **(Vitetta *et al.*, 2013)**. Identifying gut microbes and metabolites associated with OA may suggest a natural way to promote health in the form of probiotics or fortified foods.

A study investigating the relationship of the gut microbiome and metabolites with pain and pressure pain thresholds in UK-based participants having OA is discussed in Chapter 4.

1.4.3 NON-ALCOHOLIC FATTY LIVER DISEASE (NAFLD):

NAFLD is the hepatic manifestation of metabolic syndrome which arises due to energy surplus resulting in the deposition of additional fat in the liver. NAFLD is a type of fatty liver that occurs in people who don't consume alcohol excess (above recommended limit). Nearly 25% of adults are affected by NAFLD globally **(Younossi *et al.*, 2018)**. NAFLD refers to the spectrum of pathological liver conditions ranging from steatosis with inflammation (called NASH) up to cirrhosis and the latter can progress to hepatocellular carcinoma (HCC). High BMI and obesity are the primary risk factors underlying NAFLD **(Abeysekera *et al.*, 2020)** along with other factors such as diabetes, CVDs, insulin resistance and dyslipidemia (**Ruhl and Everharty, 2003; Matsubayashi *et al.*, 2022**). However, among Asians NAFLD manifests at a lower BMI range and even with insulin sensitivity **(Loomba and Sanyal, 2013).** The genetic susceptibility to NAFLD and ethnic differences are reported in various population-based genetic studies (**Fan *et al.*, 2005; Mohan *et al.*, 2009**). Also, understanding of the factors promoting the progression of fatty liver to the advanced stages of NAFLD is still incomplete.

Dysbiosis in the gut microbiome is already known to be linked to NAFLD **(Li *et al.*, 2003; Bäckhed *et al.*, 2004)**. Metabolites such as butyrate, a short-chain fatty acid (SCFA) and phospholipid choline have already been suggested to have a protective effect on NAFLD **(Spencer *et al.*, 2011; Ye *et al.*, 2018; Zhou *et al.*, 2018**). Mediterranean diet **(Esposito *et al.*, 2009; Ryan *et al.*, 2013; Kontogianni *et al.*, 2014**), hypocaloric and hyperproteic diets **(Pataky *et al.*, 2016)** and physical activities have also been shown to be having a considerable

effect on the NAFLD and progression of other associated diseases **(Baba *et al.*, 2006; St. George *et al.*, 2009)**.

The gold standard to assess the severity of NAFLD is histological evaluation of liver biopsy. Besides this, diagnosis by measuring liver biomarker enzymes in serum such as AST and ALT is not considered to be disease-specific and imaging techniques such as MRI and fibro scan are not increasingly used. With the rapid development of high throughput techniques, over the past five years, non-invasive '*omics*-based biomarkers such as microbiome community profiling, and metabolomics have been able to identify the molecular signatures associated with NAFLD.

The association and influence of diet, gut microbiome and metabolome on NAFLD and its severity in the populations from the UK is discussed in Chapter 5.

1.5 HYPOTHESIS AND AIMS

As discussed in the previous sections of this chapter, the gut microbiome composition and the metabolites produced by gut microbes are associated with disease outcomes, and intrinsic factors such as genetics, immunity and extrinsic factors such as geography, diet and exercise, affect disease status. Although extensive research has been carried out investigating the link between the disease outcome or pathophysiological states and the gut microbiome or metabolome or diet individually, there is not enough published research integrating the analysis of data from multi-omics platforms. Recently, in the discipline of microbiome studies, the concept of the contribution of the systemic effect of all these factors has been explored.

This thesis is based on the hypothesis that altered omics features i.e. gut microbiome, metabolites and dietary intakes affect the disease states or pathophysiological conditions such as undernourishment in adults (Chapter 3), knee pain in osteoarthritis (Chapter 4) and NAFLD (Chapter 5). The detailed hypothesis of each study is mentioned in the respective chapters of this thesis.

The overall aim of this thesis is to apply statistical approaches which integrate microbiome, metabolomic, clinical, nutritional and residential location data to provide an extended understanding of the role of the gut microbiome and metabolites on the incidence and progression of the nutritional status, knee pain in osteoarthritis (OA) and NAFLD. This was investigated based on the objectives as mentioned below:

1. To investigate the role of the gut microbiome by exploring difference in microbes based on nutritional status or location (chapter 3), significant gut microbes predicting knee pain in OA (chapter 4) and differentially abundant microbes with likelihood of NAFLD (chapter 5).
2. To infer the functional metabolic pathways of the gut microbes (chapter 3).
3. To investigate the relationship between serum metabolites and pathophysiological conditions of knee OA i.e. pain scores and pressure pain detection thresholds and to infer the molecular pathways involved in knee pain (chapter 4).

4. To investigate the correlations between the serum metabolites and gut microbes (chapter 4, 5).
5. To investigate the metabolic signatures of NAFLD compared to healthy volunteers (chapter 5).
6. To estimate the average dietary intakes of participants from previous one year and to investigate its influence in determining NAFLD (chapter 5).
7. To determine the proportion of contribution mediated by microbiome towards the influence of diet on the disease (chapter 5).

2

Chapter 2
METHODOLOGY

Study Cohorts, Tools Software, Pipelines, Analytical packages in R

2.1 STUDY COHORTS

This chapter presents an overview of different studies conducted for this thesis and details of the main cohorts that contributed data to the results chapters of this thesis. This is also outlined in **Table 2.1.** Further to this, this chapter details the statistical approaches, methods opted for data analysis.

Study	Analytical Methods	Participants	Observation	Comparison	Outcome
Study 1: Undernourishment in Adults	• 16S rRNA sequence Analysis • Functional Profiling of 16S rRNA sequencing • ANCOM • Spearman's correlations • PCA	• Undernourished- Rural and Urban-Slum • Nourished- Rural and Urban-Slum	• Stool-microbiome 24-hour recall dietary intake (food diary)	Undernourishment Study (case-control)	• Introduces the impact of nourishment and location of residence, the role of gut microbiome in nourishment conditions in adult population of Bangalore (India) • Presents findings on the difference in the gut microbiome and functional pathways based on the nourishment status (undernourished vs nourished) and location of residence (rural vs urban slum)
Study 2: Knee OA and Pain	• Metabolome Profiling Analysis • 16S rRNA sequence Analysis • Linear Regressions • Meta-analysis • Pathway Analysis • Spearman's correlations	Knee-OA patients	• Serum-metabolites • Serum-cytokines • Stool-microbiome	• KPIC • Webex (cross-sectional) (replication study)	• Introduces the impact of knee pain in OA patients, role of the gut microbiome and metabolome in knee pain in the UK population • Presents findings from ○ Meta-analysis of serum metabolome with pain and Pressure Pain detection Thresholds (PPTs) in KPIC and Webex datasets. ○ Association of cytokines with pain, PPTs and metabolites in Webex ○ Association of the gut microbiome with pain, PPTs, metabolites and cytokines in Webex

Study 3: NAFLD-UK	• Dietary intake Analysis • Metabolome Profiling Analysis • ANCOM • Logistic Regressions • Spearman's correlations • OPLS-DA • LASSO Regression • Mediation Analysis	NAFLD patients and Healthy Participants	• Serum-metabolites • Stool-microbiome • Average dietary Intakes during last year (validated Food Frequency Questionnaires (FFQs))	• NAFLD-UK • Omega3 (case-control)	• Introduces the impact of liver fat in NAFLD patients in the UK population, effect of dietary intake, role of gut microbiome and metabolome • Presents findings on ○ Differences in the gut microbiome of NAFLD patients and healthy participants ○ Differences in nutrient intake among the two groups ○ Difference in the metabolic profile of two groups ○ Association of gut microbiome, metabolites, nutrient intakes with disease status ○ Impact of diet intake mediated via gut microbiome

Table 2.1: Overview of the Study Cohorts in this thesis.

2.1.1 STUDY 1: UNDERNOURISHMENT IN ADULTS

This study is a case-control study and participants were recruited based on BMI as the nourishment criteria. Participants with a BMI ≤ 18.5 kg/m^2 were classified as undernourished and participants in the BMI range of 18.5-25 kg/m^2 were classified as nourished. Participants were sub-classified based on the location of residence in the Bangalore district of India as dwelling in rural (Thindlu village) or urban slum (Devarajeevanahalli (DJ Halli)).

Stool samples were collected for gut microbiome sequencing and 24-hour dietary recall food diary data was analysed for the intake of carbohydrates, proteins, fats and total calories. Participant recruitment, sample collection, diet data collection and measurement of intakes were carried out by Dr. Gift Norman and the team in India. Details on study design are elaborated in Chapter 3.

2.1.2 STUDY 2: KNEE OSTEOARTHRITIS (OA)

For this study, data from participants having knee OA were included from two independent cohorts. A subset of the web-based exercise intervention (Webex) cohort was treated as the test dataset and a subset of the Knee Pain in Community (KPIC) cohort was treated as the replicate dataset. Details of both studies are mentioned in sections 2.1.2.1 and 2.1.2.2. For the first part of the analysis, metabolome data from participants at baseline visit of Webex cohort (knee OA) and data from participants reporting knee OA from KPIC cohort were compared and meta-analysed. This study investigated links between the metabolome and knee pain and pressure pain detection thresholds (PPTs). For the second part of the analysis, gut microbiome and cytokine data from Webex study were tested for associations with pain and PPTs.

Pain was measured as the current pain using the Numerical Rating Scale (NRS) (**Hartrick, Kovan and Shapiro, 2003**) on a scale of 0-10 where 0 is no pain and 10 is the worst pain. PPTs were measured in triplicate with an algometer and were assessed by applying a certain pressure over a region of the body and then steadily increasing nonpainful pressure stimulus turning into a painful pressure sensation. PPTs help to determine the pain perception and mechanical pain sensitivity. The test sites for PPTs were superomedial patella, superolateral

patella, medial joint and anterior tibialis. The mean PPT value for each body site was computed from three PPT rounds for analysis.

2.1.2.1 WEBEX

The Webex cohort (**Gohir *et al.*, 2019**) was designed to study the effect of an internet-based exercise intervention on the microbiome, metabolome, and pain and pressure pain thresholds at baseline and follow-up at 6 weeks. Participants who had agreed to be contacted in future studies from the previous/ongoing studies **(Millar *et al.*, 2020**) conducted in the Department of Academic Rheumatology at the University of Nottingham were screened for the Webex cohort. Participants aged ≥ 45 years reporting knee pain and having radiographic structural knee OA (Kellgren Lawrence (KL) score ≥ 1) were recruited for the study. Individuals with terminal or mental illness, neurological conditions, sleep apnea, unstable heart condition and soft tissue knee injury within the last 3 months were excluded from the study. The study was approved by Research Ethics Committee (REC) (ref: 18/EM/0154), Health Research Authority (HRA) (protocol no: 18021) and the Nottingham University Hospitals NHS Trust Research & Innovation (R&I) department (ref: 18RH004).

In this thesis, participants from baseline visits (n=75) were included for analysis as the test dataset. For this cohort stool samples and serum samples were collected for gut microbiome sequencing and metabolic profiling respectively.

2.1.2.2 KPIC

The KPIC cohort study **(Fernandes *et al.*, 2017)** was conducted from 2014-2018 in the Department of Academic Rheumatology to identify the possible phenotypes (clinical or structural or laboratory) and associated risk factors of knee osteoarthritis in the community. Participants aged ≥ 40 years, with or without self-reported knee pain and questionnaire-based GP diagnosed OA (if ever or not), were recruited for this study. Individuals suffering from any known terminal illnesses, psychiatric illness, or pregnancy were excluded from the study. The postal questionnaires asked about the presence of current knee pain with a validated question: “Have you had knee pain for most days of the past

one month?" (Yes/No) **(O'Reilly, Muir and Doherty, 1996)**. Those reporting knee pain were invited for further assessments including standardised knee radiographs, pressure pain detection thresholds along with other clinical traits. The KPIC study protocol was approved by the Nottingham Research Ethics Committee 1 (NREC Ref: 14/EM/0015) and registered (clinicaltrials.gov portal: NCT02098070).

The KPIC cohort did not have data for gut microbiome sequencing or inflammatory markers. The data for metabolic profiling was provided by the Pain Centre Versus Arthritis team at the University of Nottingham. Participants who reported knee pain and OA (n=79) were included in the analysis.

2.1.3 STUDY 3: NON-ALCOHOLIC FATTY LIVER DISEASE

This study is a case-control study, investigating associations between the gut microbiome, dietary nutrient intakes and the metabolome with disease status (NAFLD and severity sub-groups), and comparing it with healthy volunteers. Stool samples, serum samples, and a dietary recall of the previous 1 year were collected and analysed. The data for this study was taken from two individually conducted studies detailed below.

2.1.3.1 NAFLD-UK COHORT

NAFLD participants were recruited from hepatology clinics of Nottingham University Hospitals NHS Trust. The study was approved by NHS Regional Ethics Committees (REC reference 12/WM/0288 and GM010201). Participants aged >18 years reporting <14 units of alcohol consumption and clinically diagnosed with NAFLD were included in this study. Patients with other forms of liver diseases such as ALD, ASH andAIH were excluded in the present analysis. The aims and objectives of the studies were explained to participants and informed consent for using the data for following research studies was obtained.

NAFLD was diagnosed based on the combined approach of testing the presence of fat-deposition analysed by liver tests, abdominal ultrasound (USG), fibro scan, MRI-PDFF and liver biopsy. Initial screening of patients was done based on the liver enzyme tests (ALT levels) and abdominal USG. Patients at low or

intermediate risk were accessed by fibroscan. A CAP (Controlled Attenuated Parameter) score of > 302 db/m was used as the threshold for the presence of >67% of liver fat. A fibrosis score of F0-F4 was indicative of liver scarring where F0 and F1 (2-7kPa) represent no or little scarring, F2 (7.5-20kPa) indicates moderate scarring, F3 (10-14kPa) indicates severe scarring and F4 (>14kPa) indicates late stage scarring or cirrhosis and irreversible liver damage. In high-risk patients, NAFLD was assessed by liver biopsy. A diagnosis of NASH in the liver biopsies was made based on the presence of hepatocellular ballooning and lobular inflammation (**Bedossa *et al.*, 2012)**..

2.1.3.2 OMEGA-3 DIET INTERVENTION COHORT

Data for the healthy volunteers (HV) were included from the participants of an Omega3 diet intervention study (**Vijay, Astbury, *et al.*, 2021**) at a baseline visit which was hosted by the Department of Twin Research and Genetic Epidemiology, King's College London registered under UK Adult Twin Registry (Twins-UK). The study was approved by the West Midlands Black Country Research Ethics Committee (18/WM/0066) and is registered under the clinicaltrials.gov database (NCT03442348). Participants were recruited based on the inclusion criteria of age > 18 years and BMI in the range of 20-39.9 kg/m^2. Participants with regular use of omega3 supplements, NSAIDs, chronic inflammatory diseases and pregnancy were excluded from the study.

2.2 MICROBIOME

Gut microbes exist in a community either in symbiotic, mutualistic or pathogenic relationship and this shapes the composition and abundance of these microbes in an environment. Microbiome sequencing refers to the genomic analysis of microbes present in an environment by determining the nucleotide sequence of a segment of DNA or complete genome to study the abundance, composition, and functions. The conventional microbial culture techniques could detect only 30% of gut microbiota and were not enough for identifying and characterizing a large number of microbes present in an ecosystem. So, over the past few years microbiome sequencing has significantly emerged as a crucial aspect to study human health. 16S rRNA gene sequencing has been the most commonly used technology to identify and characterize bacteria for decades and

is still the foundation for this era of microbiome studies. High throughput sequencing (HTS) technologies have evolved during the past few years due to advance molecular biology techniques **(Hayashi, Sakamoto and Benno, 2002; Goodman *et al.*, 2011; Rettedal, Gumpert and Sommer, 2014)**. The availability of cost-effective Illumina sequencing has made sequencing commonplace for studies with microbiome as an aspect of human health. With the advent of these HTS techniques, sequencing of 16S rRNA gene has become more realistic and a commonly followed approach. For this thesis, the 16S rRNA gene was PCR amplified and processed following the library preparation protocols for MiSeq Illumina platforms (**Illumina Inc., 2013**). The overview of steps for the microbiome sequencing approach is diagrammatically shown in **Figure 2.1.**

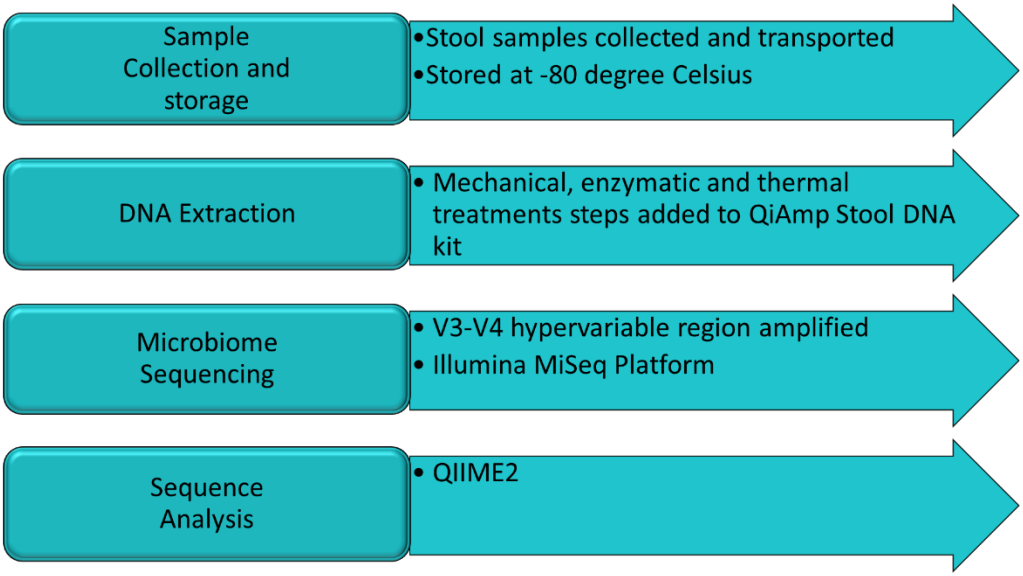


Figure 2.1: Microbiome Sequencing approach

2.2.1 SAMPLE COLLECTION AND TRANSPORTATION

As a source for gut microbiome studies, generally, stool specimens are collected as it is non-invasive and can be sampled repeatedly. For this thesis stool samples were collected by the participants at home on the day of the visit in the kits provided. Sample collection and transportation details are mentioned in respective chapters as per the protocols followed.

2.2.2 DNA EXTRACTION

High-quality DNA extraction is an essential part of microbiome research. The main disadvantage of the kit-based DNA isolation protocols in microbiome studies was that the microbial composition and abundances' representation were found to be different based on different protocols used for isolation. This was due to the different cell wall structures of gram-positive and gram-negative bacteria, so additional steps of the microbial cell lysis had become a crucial step of bacterial DNA isolation protocols.

For this thesis, cell lysis was done using a combinatorial approach of mechanical homogenization, enzymatic treatments, thermal treatment and bead-beating steps to achieve a better yield and better diversity representation of microbes in the sample. This was followed by the Qiagen kit (QiAmp Stool DNA isolation) or MoBio kit (Power Soil DNA extraction).

Details of DNA extraction for each study are specified in respective chapters.

2.2.3 16S rRNA SEQUENCING

The bacteria are identified and characterized by 16S rRNA sequencing, which is a universal and conserved marker gene of about 1,550 bp size and composed of hypervariable regions (referred to as V1-V9) interspersed between constant regions (referred to as C1-C9). The size and polymorphisms in the hypervariable regions of the 16S rRNA gene make it statistically suitable for differentiating bacteria at the genus level across all bacterial phyla. Also, the availability of 16S rRNA gene full-length sequence information in repositories like Greengenes **(DeSantis *et al.*, 2006**), SILVA (**Quast *et al.*, 2013**), 16S Ribosomal database (RDP) (**Cole *et al.*, 2007**) allow the comparison and links to the bacterial taxonomic and phylogenetic information. Sequencing and analysing the full-length 16S rRNA gene allow a better classification of bacterial taxa, however, the cost per sequence is huge and cannot be met when sequencing a large number of samples simultaneously as per the statistical power requirement for microbiome studies. Therefore, the V3-V4 region (2*250 bp) of 16S rRNA was amplified for paired-end sequencing using universal primers 355F (CCAGACTCCTACGGGAGGCAGC) and 806R (GGACTACHVGGGTWTCTAAT). An overview of 16S rRNA sequencing is

presented in **Figure 2.2**. Amplicons were barcoded in the second stage PCR and then these were pooled, quantified using qPCR, normalised and processed following the library preparation protocols for the MiSeq Illumina platform (**Illumina Inc., 2013**). The library preparation concentration and finally pooled concentration are detailed in the respective sections of each Study Chapter.

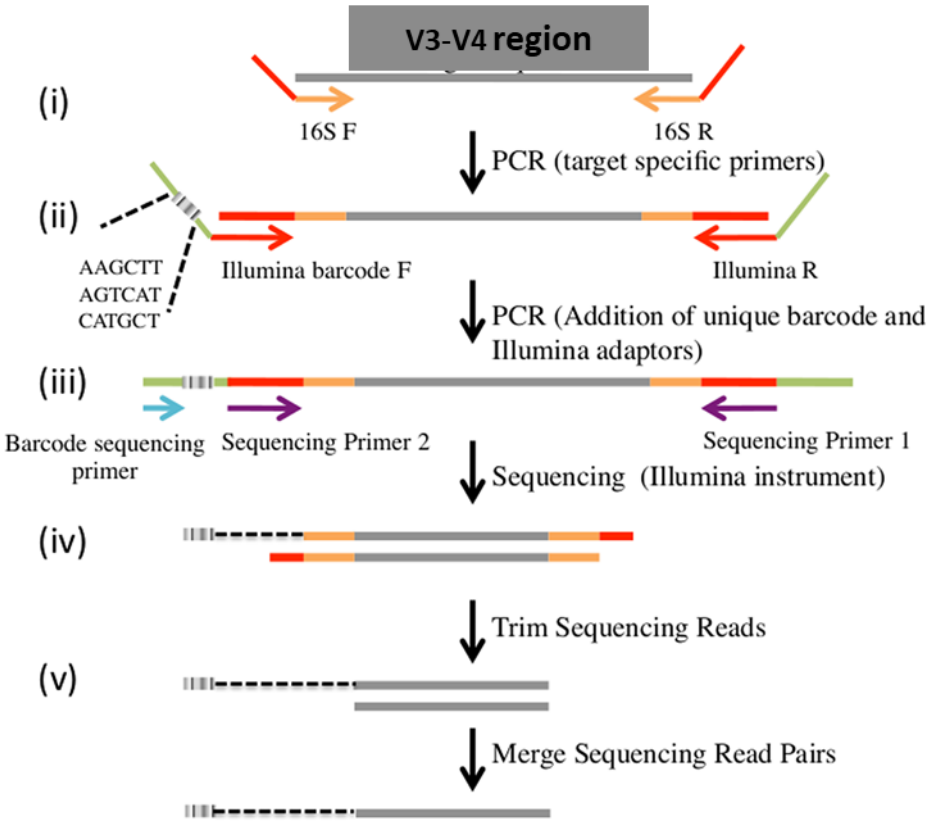


Figure 2.2: 16S rRNA gene sequencing. This figure is inherited from **(Logue *et al.*, 2016)** (i) A first PCR amplification is performed on DNA extracted from each sample targeting 16S rRNA V3-V4 (grey) using primers modified with a 5'-end tail complementary to the Illumina sequencing primers (red). (ii) A second PCR amplification incorporates the Illumina adaptors and a 6-nucleotide barcode unique to each sample's PCR product. (iii) Normalised pooled PCR products are simultaneously sequenced on an MiSeq Illumina platform generating paired-end reads (grey in (iv)) and barcode sequences (grey-dashed). (v) Paired-end reads are then merged to provide consensus sequence reads.

2.2.4 16S rRNA ANALYSIS USING QIIME 2 (VERSION 2019.10)

There are various tools and pipelines used for 16S rRNA sequencing data analysis. For instance, MEGAN **(Huson *et al.*, 2007)**, Mothur **(Schloss *et al.*,**

2009), SILVAngs **(Quast *et al.*, 2013)**, MG-RAST **(Keegan, Glass and Meyer, 2016)**, RDPpipeline **(Cole *et al.*, 2014)**, QIIME2 **(Bolyen *et al.*, 2018)**. These platforms use different methods for demultiplexing (splitting out the sequencing reads belonging to individual samples from the pool of sequencing reads), denoising (removing the sequence errors), clustering as OTUs (grouping the related sequencing reads into operational taxonomic units) or estimating the ASV's (amplicon sequence variants). The field of these bioinformatics-centred approaches is rapidly evolving and there is no consensus about the limitations or advantages of using one method over the other. We chose to analyse our data with QIIME2 (Quantitative Insight into Microbial Ecology) as it is the most widely used and cited open-source tool with self-contained statistical tools. QIIME is implemented in the Python programming language and has been designed for an end-to-end microbiome analysis. It is integrated with the tools for importing raw sequencing data to generate the taxonomy tables and statistical analysis features to display figures and graphs. It is a plugin-supported pipeline with artifacts (output from one step is inputted in the next or another step) and visualizers. **Figure 2.2** depicts the approach followed for the analysis and steps are elaborated in sections below. QIIME commands are provided in **Appendix 1.**

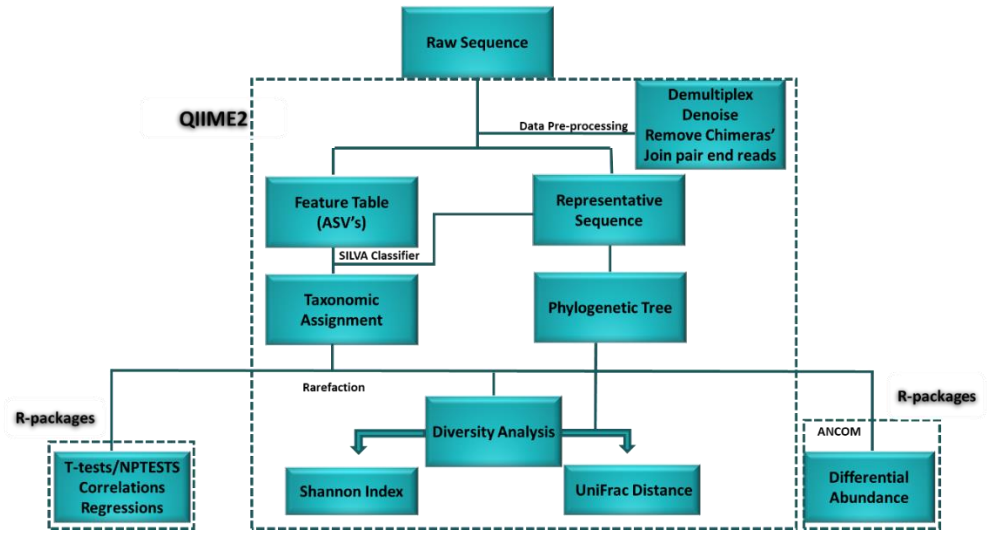


Figure 2.3: Flow Diagram of 16S rRNA sequence analysis using QIIME2.

2.2.4.1 DATA PRE-PROCESSING

Illumina demultiplexed paired end read sequences were trimmed for primers using the Cutadapt plugin (version 2020.6.0). An example of the raw read fastq format is shown in **Figure 2.4**. Denoising of the reads and adapter removal was done using the DADA2 plugin. It joined paired end reads, corrected errors and dereplicated reads by removing chimeras (sequences generated due to incomplete PCR short templates that act as primers to extend and form the incorrectly overlapped sequences) and singletons, merging similar or unique sequences. DADA2 generated a feature table and a representative sequence table. A feature table is a matrix of each sample containing OTUs or ASVs and a representative sequence table contains a list of inferred representative output sequences and the frequency for each OTU or an ASV. OTU is referred to as the operational taxonomic unit and is the consensus sequence of the reference-based clustering with the 97% sequence similarity thresholds. The feature table approach paradigm is shifting from OTU clustering to ASV picking. ASVs are the exact sequence variants that are estimated by calculating the statistical support for the presence of the real sequence. This is done by accounting for the number of times the exact sequence was read and determining the probability that the given sequence was not due to sequencing error. As ASVs are not clustering or reference based unlike OTUs, ASVs, therefore, are less prone to sequencing error effects and these can be compared and combined for multiple studies. Additionally, as ASVs are exact sequences, these can be compared to taxonomy databases at a much higher resolution. **(Callahan, McMurdie and Holmes, 2017)**.

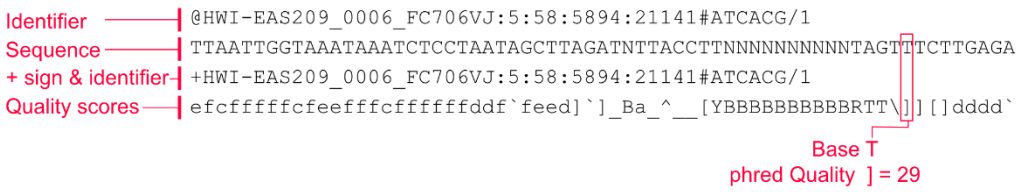


Figure 2.4: FASTQ File format of Raw Reads: An example of the raw reads fastq format. The raw reads are demultiplexed, denoised with a Phred score of 30 and joined for further analysis.

2.2.4.2 TAXONOMIC ASSIGNMENT

To identify which taxa are present in each of the samples, the taxonomic assignment was done by comparing the query sequence to a trained classifier reference database of known taxonomic assignments. QIIME has its resources of scikit-learn naive Bayes machine-learning classifier (q2-feature-classifier) which is an alignment-based consensus taxonomy incorporating VSEARCH (global alignment), and BLAST+ (local alignment). The classifiers are trained on two reference databases which are Greengenes and SILVA. In this thesis taxonomy was assigned using a pre-trained classifier based on the SILVA Database (**QIIME, February 1, 2020; Bokulich et al., 2018**). The abundance of microbes at the genus level was used for further analysis.

2.2.4.3 RAREFACTION

Rarefaction is the normalization method for microbe abundance data adjusting the absolute count of microbes to an equal number of sequences for each sample. The threshold or the rarefaction value is the minimum number of reads where most of the samples have a greater number of reads than that value. Rarefaction considers the errors of difference in library size and sequencing depth across the samples unlike other normalisation procedures like scaling and Aitchison's log ratio. Rarefaction plots (an example is shown in **Figure 2.5)** depict that this approach captures the most possible samples whilst still retaining the overall diversity. It is also independent of dealing with the sparsity of the data; however, this approach loses a fraction of data that may result in a loss of significance of precise taxa across samples. We have used rarefied data where the statistical data inference tests don't have a built-in normalization approach.

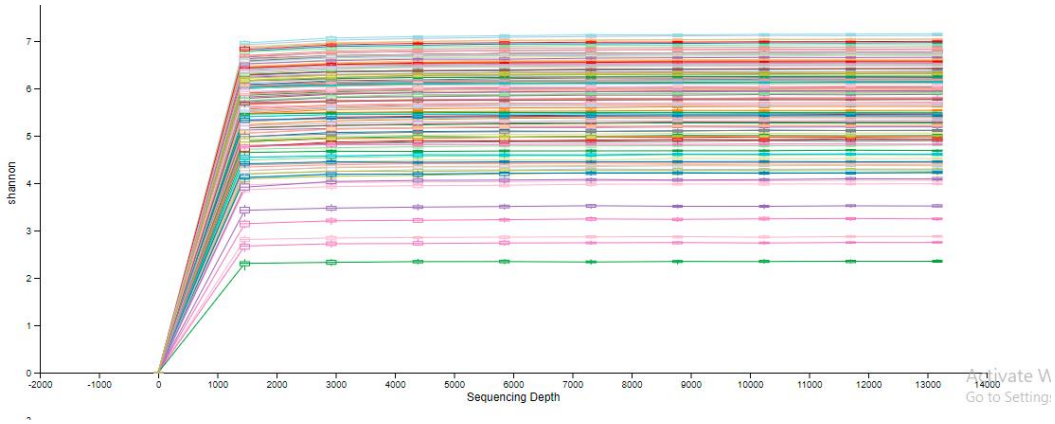


Figure 2.5: Rarefaction plot example: The figure depicts that the Shannon diversity is same with the increase in the number of sequences in each sample. This indicates that at a certain sequencing depth (here ~13,000) same microbes are sequenced over and over and this doesn't impact the Shannon diversity. This figure is taken from Study 2.

2.2.4.4 DIVERSITY METRICS AND PHYLOGENETIC ANALYSIS

Alpha diversity defined as the difference in the microbes of a single sample, taking into account the number of different taxa and their relative abundances and Beta diversity defined as the difference in microbial community composition between individuals were calculated to describe the number and distribution of the microbes. As rarefaction is a random sampling normalisation method, so the diversity measurements have been calculated by taking the average of the rarefied feature table with 50 iterations.

For alpha diversity estimations Shannon index **(Shannon, 1948; Magurran, 1994)** was calculated which accounts for both the abundance and the evenness of the species present. An example explaining Shannon Diversity is presented in **Figure 2.6**. It is calculated as the sum of the proportion of species i, relative to the total number of species (Pi) multiplied by the natural logarithm of this proportion (ln(Pi)) multiplied by -1. It is represented as

$$H' = -\sum Pi \ln(Pi) ,$$

where Pi is the proportion of individuals belonging to species i.

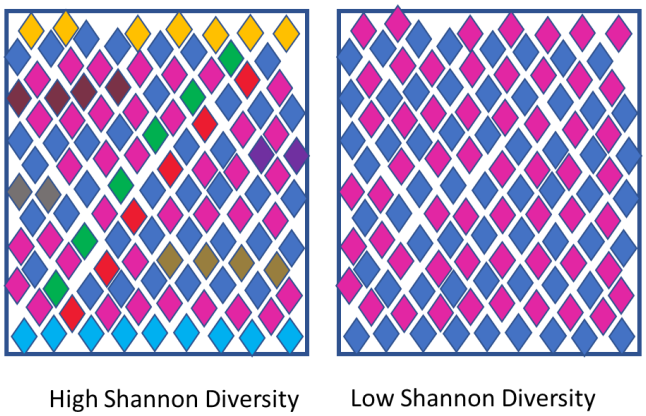


Figure 2.6: Shannon Diversity: This diagram presents the Shannon Diversity. Each diamond represents a different microbe, with colours corresponding to distinct taxa.

Beta diversity was calculated by comparing the unweighted UniFrac distances **(Lozupone and Knight, 2005)** which is a phylogenetic diversity metric and takes the distance between the unique branch lengths into account. The phylogenetic tree was generated by aligning the representative sequences by multiple sequence alignment using FastTree **(Price, Dehal and Arkin, 2010)**. UniFrac distance between two samples is calculated as the ratio of the unique branch length and the total branch length. It is represented as U_{AB}=sum of unique branch length in A and B/total branch length where A and B are the two samples; unique is the branch length that leads to OTUs observed in samples A or B as explained in an example **Figure 2.7A**. In this example, OTU 1 is observed in only A, OTUs 2 and 4 are observed in both A and B, OTU 3 is not observed in either of the sample and OTU 5 is observed only in B. So, U_{AB}= 0.5+0.75/ 0.5+0.5+0.5+1.0+0.75+0.75+1.25 = 0.238. . UniFrac distance matrix can also be used to describe the pairwise phylogenetic distances between the sets of sequences collected from many different microbial communities as explained in an example **Figure 2.7B**. Here, pair A has 7, B has 3 and C has 1 shared branch respectively which scores as 0.7, 0.3 and 0.1. Distance matrix between two pairs is calculated as number of root ancestor clade (here 1) minus number of shared clades score. So, distances results to be 0.3, 0.7 and 0.9 for pair A, B and C respectively. This distance metrics was then used for the multivariate data plotting (PCA) and hierarchical clustering.

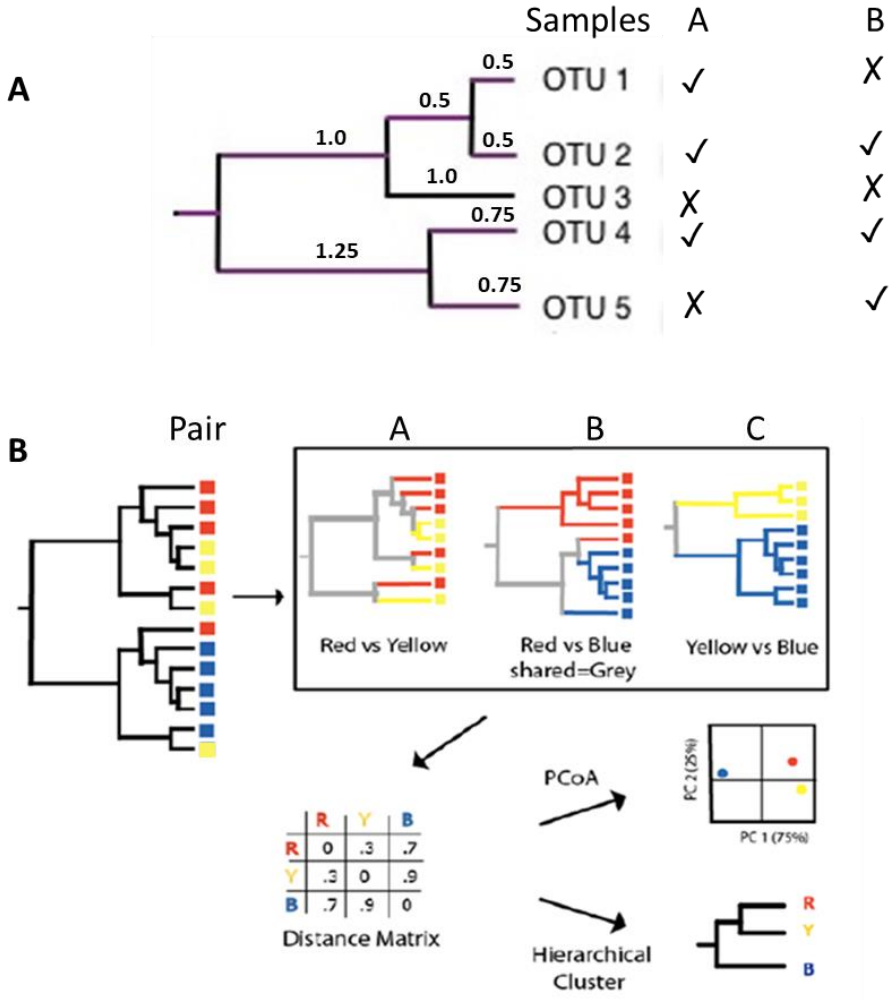


Figure 2.7: UniFrac Distance Measures. A) for two samples B) for all pairwise combinations in a tree to make a distance matrix-Figure B is inherited from (**Lozupone and Knight, 2005**). .

2.2.5 DIFFERENTIAL ABUNDANCE (DA) TESTING OF GUT MICROBES

The data generated from HTS (High Throughput Sequencing) techniques are in proportions and referred to as compositional data. Statistical methods for finding differentially abundant microbes make some assumptions about the distribution of the count of microbes. For example, DESeq **(Anders and Huber, 2010)** assumes a negative binomial distribution and metagenomeSeq **(Paulson *et al.*, 2013)** assumes zero-inflated Gaussian distributions. For this thesis, differentially abundant microbes were identified using ANCOM **(Mandal *et al.*, 2015)**, adjusted for covariates (age, gender and BMI).It takes into account the multinomial distribution and compositionality of the data as explained below.

2.2.5.1 ANCOM 2.0

ANCOM (stands for ANalysis of COmposition of Microbes) is a framework to find out the composition of the microbiome and compute the differentially abundant microbes in a dataset of two or more populations. It is an R language-coded program and gives the output based on a linear modeling analysis. ANCOM runs ANOVA with Linear Mixed Effect (LME) or Kruskal Wallis or Friedman's or Wilcoxon tests on the log ratio across features rather than directly on the individual features. Unlike other tools available, ANCOM is very sensitive to even a smaller number of samples (>20 samples per group) and effectively controls the false discovery rate. ANCOM does not make assumptions about the independence of features but assumes that 1. not more than p-2 features are changing (p is the total number of features in the environment) and 2. if all taxa are differentially abundant then the mean abundance of every feature does not change by the same amount between two or more groups. It works around the log ratio of the abundances and so it adds a pseudo count of constant 1 to replace every zero in the dataset. This is one of the limitations of ANCOM but is a mandate as the log of zero is undefined. One of the advantages of ANCOM is that it calculates a *W* statistic (the number of times the log ratio of taxa was detected to be significantly different when compared with every other taxon) at detection thresholds of 90%, 80%, 70% and 60%. These detection thresholds indicate the percentage of samples in which the microbial taxa are differentially abundant meaning that there was a significant change in the ASV compared to the rest of the ASVs in the community in at least 90% or 80% or 70% or 60% of the comparisons. ANCOM was performed in Rv3.5.2 with the required libraries - nlme, exactRankTests.

2.2.6 FUNCTIONAL PROFILING OF 16S rRNA MICROBIOME DATA

Significant changes in the functioning of the gut microbiome with the change in the composition and abundance across the study groups were predicted from the16S rRNA data using the marker gene-based functional profiling analytical tool "Tax4fun" available at the MicrobiomeAnalyst Pipeline **(Chong *et al.*, 2020).** Tax4Fun transforms the SILVA-based ASVs into a taxonomic profile of KEGG (Kyoto Encyclopaedia of Genes and Genomics) organisms normalized by the 16S rRNA copy number from the National Centre for Biotechnology

Information (NCBI) annotations. These abundances were then linearly combined with functional abundances of the KEGG organisms' functional profiles to compute the KEGG orthologs (KO) based on the phylogenetic differences between the microbial species. The workflow of Tax4Fun is presented in **Figure 2.8**. KEGG orthologs were assigned to KEGG pathways based on the automated annotation through KEGG mapping. Since KO can regulate various categories of pathways, these were narrowed down to higher functional KEGG metabolic pathways.

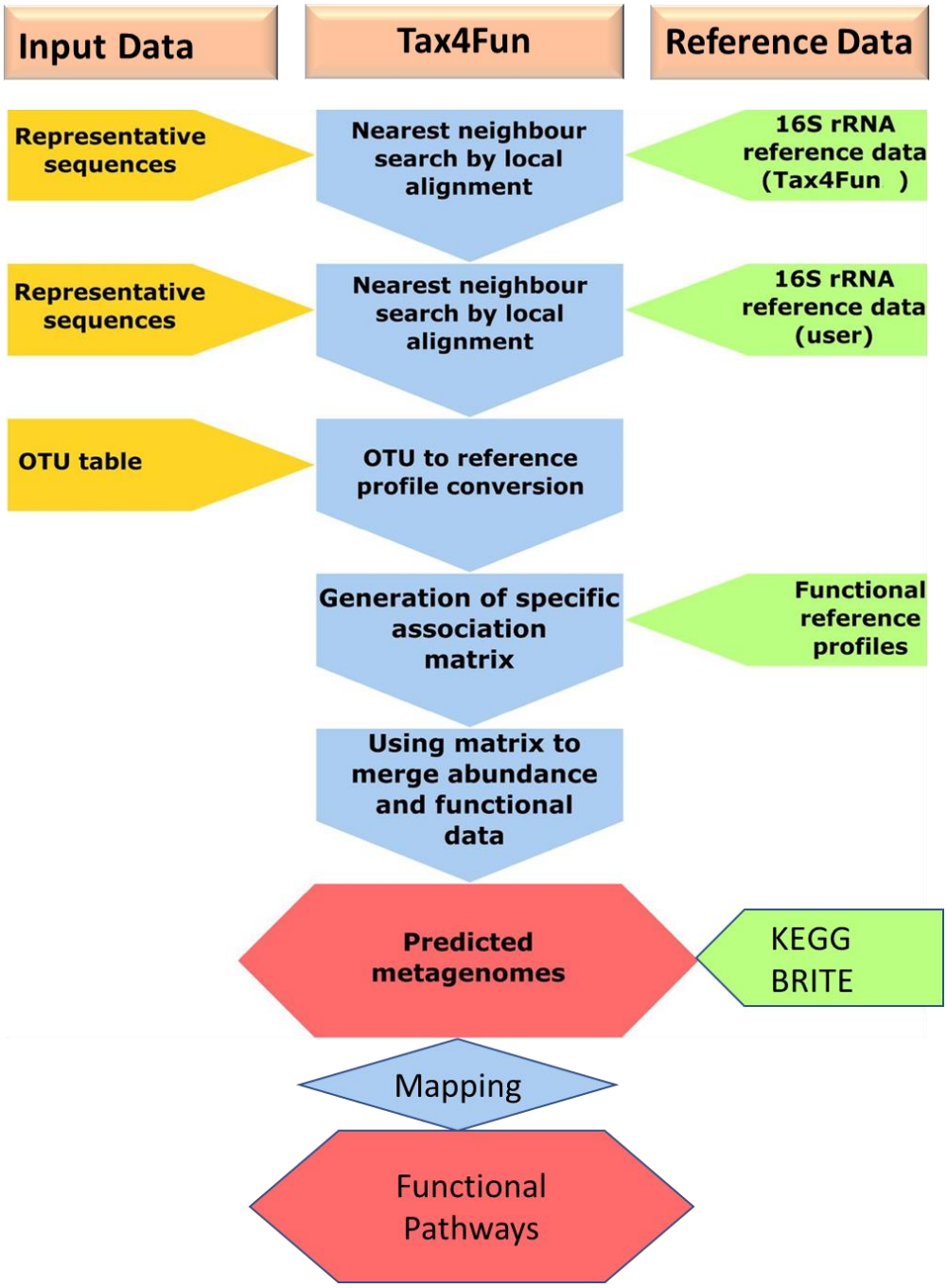


Figure 2.8: Tax4Fun Workflow. This figure is adapted from **(Wemheuer et al., 2018)**. 16S rRNA feature table is mapped with the KEGG orthologs (KO) functional reference database generating an association matrix and thus predicting metagenome. The predicted metagenome is further mapped with the KEGG BRITE (functional pathways) database containing the precomputed functional profile of KEGG organisms. KOs are then assigned together to functional categories.

2.3 METABOLOME

2.3.1 METABOLOMICS

The altered gut microbiome codes different metabolites which are seen to be associated with many diseases and disorders. Metabolomics refers to the approach studying microbial and host-associated small molecules (called metabolites) within the cells and tissues. The capability to identify hundreds to thousands of metabolites simultaneously in the given sample is based on the advance technologies such as high throughput liquid chromatography (LC), mass spectrometry (MS/MS), and High Nuclear Magnetic resonance (H-NMR) technology.

High-performance LC coupled with MS is the most comprehensive method used for metabolomics for several reasons: 1) LC-MS separates molecules based on different migration rates and is followed by structural identification of individual compounds based on charge to mass ratio of ions. LC-MS can thus provide high sensitivity, specificity and resolution as compared with methods like LC or MS alone. 2) It can separate and identify solutes in low concentrations (which are in parts per million- PPM) in a complex mixture. In addition, it is applied to a range of samples and even identifies different types of impurities. An overview of the metabolomic workflow is presented schematically in **Figure 2.9**.

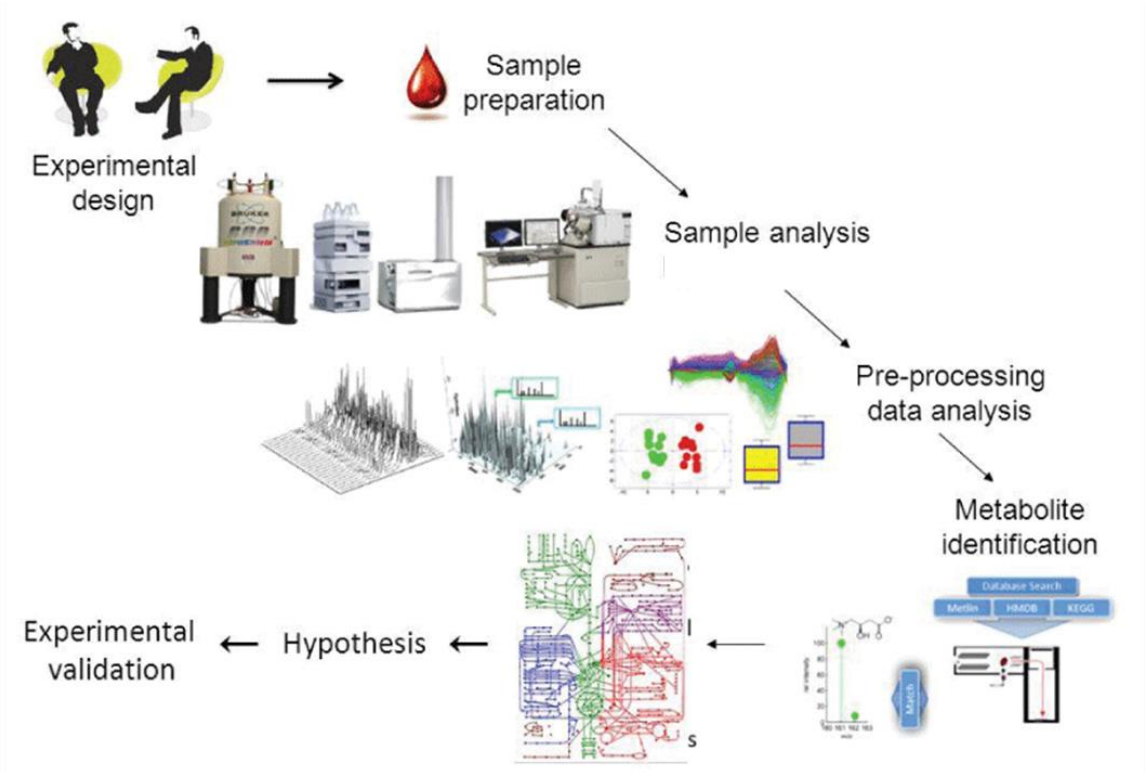


Figure 2.9: Metabolomics Workflow. This figure is adapted from **(Nalini Kanta et al., 2020**). The biological fluids are prepared and tagged with the label for the identification and quantification of the metabolites. The intensities of the signal are relative to the concentration of the metabolites.

2.3.2 SAMPLE COLLECTION, PREPARATION AND DATA ACQUISITION

Blood metabolic profiling presents most of the variation in response to the factors associated with the host health. Plasma and serum are the most widely used components of blood for metabolic profiling.

For this thesis, fasted whole blood samples were collected using serum separator tubes and were processed within 2–3 hours of collection. Blood was allowed to clot left undisturbed at room temperature for ~30 minutes. This was followed by centrifuging steps at 1,000-2,000 x g for 10 minutes at 4°C. The liquid fraction of the blood i.e. serum was separated and aliquoted for storage at −80°C for metabolic profiling and cytokine analysis.

2.3.2.1 BIOCRATES PANEL

Metabolites were profiled based on the integrated approach of high throughput magnetic resonance mass spectrometry and liquid gas chromatography (LC-MS/MS) at Biocrates Life Science AG, Innsbruck, Austria using the commercially available Biocrates MxP Quant 500 kit **(Roberts *et al.*, 2020**).

Metabolite quantitation was analysed using Sciex Analyst® and data was imported into Biocrates MetIDQ™ software for annotation of the metabolites and exported as output file for further analysis. A total of 630 metabolites were quantified belonging to the class of acylcarnitine, alkaloids, amine oxide, amino acids, bile acids, biogenic amines, carboxylic acids, ceramides, cholesterol esters, cresols, diacylglycerols, dihydroceramide, fatty acids, glycerophospholipids, glycosylceramides, hormones, indole derivatives, nucleobases, sphingolipids, sugars, triglycerols, vitamins and cofactors. A list of metabolites included in this study is provided as **Appendix 4.1.**

2.3.2.2 NIGHTINGALE PANEL

Fasted serum samples were sent to Nightingale Health, Helsinki, Finland and were processed using the high-throughput LC- H-NMR metabolomics platform as described previously **(Soininen *et al.*, 2009).** A total of 228 metabolites were quantified belonging to the class of amino acids, apolipoproteins, cholesterol, cholesteryl esters, fatty acids, fluid balance, free cholesterol, glycolysis-related metabolites, inflammation, ketone bodies, lipoprotein particle concentrations, lipoprotein particle sizes, sub-classes of lipoprotein, relative lipoprotein lipid concentration, other lipids, phospholipids, total lipids, triglycerides (**Details in Appendix 5.1**).

2.3.3 PRE-PROCESSING OF METABOLITE DATA

Acquired metabolite concentration data were pre-processed by imputing the missing values and replacing them with the minimum positive value in the data **(Faquih *et al.*, 2020**). The data was inverse normalised before the downstream analysis.

2.4 INFLAMMATORY MARKERS

Pro and anti-inflammatory serum cytokine measurements were performed by Affinity Biomarkers, London using the standardized Human Proinflammatory panel 1 assay kit (cat number K151A0H-1), distributed by Meso Scale Discovery and levels of IFN-γ, IL10, IL-12P70, IL-13, IL-1β, IL2, IL4, IL6, IL8, TNFα were measured and concentrations are reported in ng/L.

2.5 DIET

2.5.1 DIETARY ASSESSMENT

Dietary information was collected based on the extensively validated and widely used **(Bingham *et al.*, 2001; Mulligan *et al.*, 2014)**, European Prospective Investigation into Cancer and Nutrition Food Frequency Questionnaires **(Kroke *et al.*, 1999**) (EPIC-Norfolk FFQs) and TWINS-UK FFQs **(Teucher *et al.*, 2007**) adapted from EPIC-Norfolk. EPIC-Norfolk is a semi-quantitative, paper-based FFQ that includes 130 food items, a separate question for milk intake (exclusive of other dairy products) and an open section for recording an individual's average habitual food intake during the previous year. A copy of both FFQs is attached as Appendix 5.2 and 5.3. Briefly, for each food item, the portion or amount is mentioned in parenthesis and columns with headers representing the frequency of specified intake, on average, over the last 12 months for recording. Frequency intakes were coded from 1-9 where 1 represents never or less than once a month and 9 represents 6+ per day.

For this thesis, dietary intake was calculated and compared between the participants having NAFLD and healthy volunteers (Study 3, Chapter 5). Dietary intakes for healthy volunteers were calculated from the participants recruited from the Omega3 intervention study which used TWINS-UK-FFQ and dietary intakes for NAFLD participants were calculated from EPIC-Norfolk FFQs.

TWINS-UK-FFQs had a few additional dietary items which were excluded to make them comparable to the EPIC-Norfolk FFQs. Some of the dietary items such as type of chocolates, nuts and cereals were asked separately in TWINS-UK-FFQs, therefore the intake for these dietary items was summed up and the frequency of intake was calculated. So, if the sum of the frequencies for the aforementioned items were counting higher than 9, it was replaced with 9 (as 9 is the code for the highest frequency intake). Details of the differences and amendments are mentioned in **Table 2.2**.

TWINS-UK variable	EPIC-NORFOLK variable		Amendment
White or milk chocolates, single or squares (one)	CHOCOLATE	Frequencies of Twins-UK variable summed together	Frequencies greater than 9 were replaced with 9.
Dark chocolates, single or squares (one)			
Salted nuts, e.g. peanuts, cashews (handful)	NUTS	Frequencies of Twins-UK variable summed together	Frequencies greater than 9 were replaced with 9.
Unsalted nuts, e.g. brazil, walnuts (handful)			
Seed e.g. Sunflower, pumpkin (tablespoon)			
Chips, roast potatoes	CHIPS	Frequencies of EPIC-Norfolk variable were halved	Frequencies 0.5 replaced by 1 for either chips or roast potatoes and either was left blank
	ROAST POTATOES		Frequencies 1.5 replaced by 1 for either chips or roast potatoes and either was put 2
Breakfast cereal e.g. Cornflakes, Rice Krispies	CEREAL	Frequencies of Twins-UK variable summed together	Frequencies greater than 9 were replaced with 9.
Sugar topped cereals e.g. Frosties			
Muesli			
High Fibre cereals e.g. Branflakes, All Bran, Fruit and Fibre			
Olive oil spread (teaspoon)	OTHER Margarine	Frequencies of Twins-UK variable summed together	Frequencies greater than 9 were replaced with 9.
Other soft margarine, dairy spreads, e.g. Blue Band, Clover (teaspoon)			

Table 2.2: Details of the changes made in TWINS-FFQ to make it comparable to the EPIC-Norfolk FFQs.

2.5.2 DIET DATA ANALYSIS IN FETA SOFTWARE

The FFQs were inputted in the FETA (FFQ EPIC Tool for Analysis) compatible format and data for the participants not recording more than 10 questions were removed before further analysis. Zero and blank values for milk food was removed and replaced with a2002, for cereals these were changed to a1362, for fat frying column these were replaced with 17046 and for the fat baking column these were replaced with 17018. The aforementioned codes are the default codes indicating that the specified item was used but that the type was unknown)Intake of a total of 17 macronutrients, 27 micronutrients and energy intakes were calculated in the study from FETA software. For association tests, energy adjusted nutrient intakes was inverse normalised.

2.5.3 ENERGY ADJUSTMENTS FOR REGRESSION MODELS

Energy intake adjustment was done by calculating the residuals of nutrients regressed onto energy intake in kcal **(Dietary Assessment Primer, Learn More about Energy Adjustment, National Institutes of Health, National Cancer Institute. (https://dietassessmentprimer.cancer.gov/) [Accessed on June 06, 2022])**. This give an estimate of each nutrient intake uncorrelated with total energy intake and directly related to overall variation in food choice and composition.

2.6 STATISTICAL ANALYSIS

In this thesis, statistical approaches in different studies were chosen based on the study sample size, data structure, type of variables and the research question.

2.6.1 INFERRING SIGNIFICANCE

For the understanding of biological relevance, statistical inferences were made by performing t-tests or non-parametric Mann-Whitney U test or ANOVA. The distribution of data was determined by performing the Kolmogorov Smirnov test and by plotting histograms. For t-tests and regressions, normalization was done following inverse normal transformation **(Beasley, Erickson and Allison, 2009**) in R using function qnorm() which is built-in in R statistical packages. Mann-Whitney tests were performed when the data was shown to have a non-normal distribution. Statistical significance inferring tests were performed in R and presented as boxplots or tables in Microsoft Excel. Descriptive data are

presented as mean (SD) for continuous data and categorical data are presented as percentages. P-values < 0.05 were considered to be significant.

To control the overall significance level, multiple testing correction was done using q-value method (Study 1 and 2) or Bonferroni method (Study 3). Q-value method exert a less stringent control onto the p-values compared to Bonferroni increasing power at the cost of increasing the rate of type I errors. The q-value method was chosen for Study 1 and 2 as the p-value distribution was homogenous. However, in the Study 3 p-value distribution was skewed towards 0, therefore, turning the FDR values to be significant even though their p-value ≧ 0.05, henceforth, Bonferroni methods was performed. The multiple testing threshold was chosen to be of <0.1 for q-value method and <0.05 for Bonferroni method.

2.6.2 PREDICTING ASSOCIATIONS.

To understand the potential relationship between two variables correlation analysis and regression analysis were performed. Spearman's correlation coefficients determined the strength and direction of associations while regression coefficients determined the quantitative association between the two variables, specifically, how much the change in one variable changes the other variable. Multivariate linear and logistic regression were adjusted for possible confounders such as age, gender, BMI, drugs (NSAIDs and opioids) or energy intakes as per the models specified in each chapter. The significance of the relationship was concluded with the p-value < 0.05 and corrected for multiple testing with a threshold of FDR <0.1. The results from correlations and regressions are presented either as heatmaps or circos plots or forest plots or as tables in Microsoft Excel. Correlations were used to predict the direction and strength of the relationship between two variables for a smaller dataset (Study 1) or to define relationships between two variables (Study 2 and 3). Regressions (Study 2 and 3) were used for a more quantitative prediction of the effect of explanatory variable on the dependent variable.

2.6.3 META-ANALYSIS.

Meta-analysis is a qualitative method for checking the overall trend, significance and precision of the results of independent research studies. It provides a more

precise estimate of the effect size and increases the generalizability of the results of individual studies. Therefore, it may enable the resolution of conflicts between studies, and yield conclusive results from all included individual studies.

In this thesis, significant metabolites from association tests in KPIC and Webex study datasets were meta-analyzed using a random-effect model. Differences in the methods or sample characteristics or other factors may introduce variability in the true effects of the outcomes of compared studies. The random effect model works by taking into account the heterogeneity or variance by treating it as random assuming that the individual specific effects are uncorrelated with the independent variables. The regression coefficients of significant metabolites from KPIC and Webex studies were compared using the Residual/Restricted Maximum Likelihood Method (REML) method using the "metafor" package in R. The results are presented as forest plots.

2.6.4 PATHWAY ANALYSIS.

Pathway analysis was done using the Pathway Analysis tool implemented in the MetaboAnalyst pipeline (**Pang et al., 2021)** based on the KEGG metabolic pathways library for humans. Pathway relevance was calculated based on the over-representation analysis (ORA) of metabolites and pathway topology estimation methods. ORA method detects the relevant pathway if the proportion of user-input metabolites exceeds the proportion of metabolites that could be randomly expected, within the given pathway. ORA was tested based on hypergeometric tests which calculate the p-value as the probability of randomly drawing k or more successes from the population in n total draws giving the significance of pathways. The topology measurement of the pathways is an added statistic for computing the significance of the metabolites to the pathways considering the interaction of metabolites with other metabolites in the pathways (direct or indirect). Topology measurement for each pathway was done using relative-betweenness centrality among the metabolites. Betweenness centrality tests the extent of the integration and importance of the metabolite towards the global pathway. This is measured by calculating the number of shortest paths going through the metabolite focussing on global network topology giving the

impact of the pathways. The pathway Analysis workflow is presented schematically in **Figure 2.10**

In this thesis, the results of pathway analysis are presented as scatter plots.

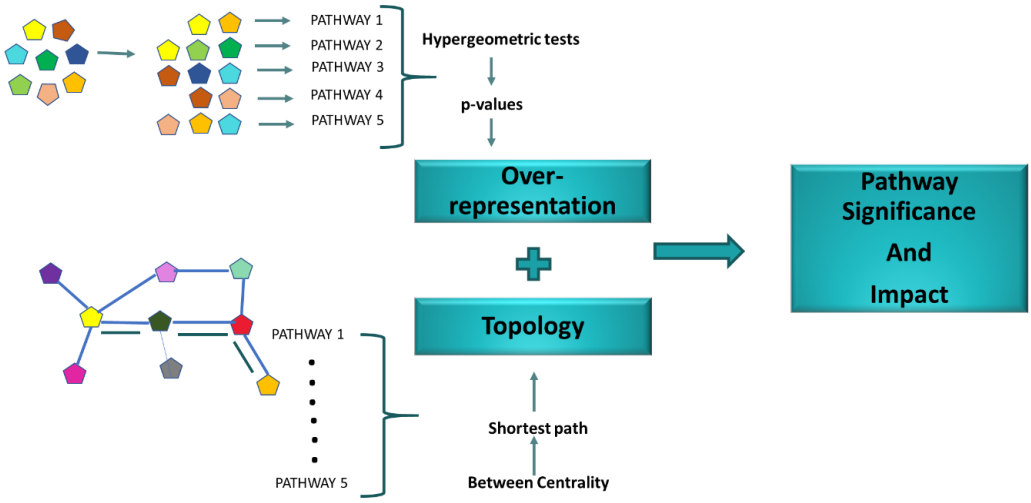


Figure 2.10: Pathway Analysis Workflow: This figure represents the steps implemented in Pathway Analysis tool. Each hexagon represents a metabolite.

2.6.5. ORTHOGONAL PARTIAL LEAST SQUARE DISCRIMINANT ANALYSIS (OPLS-DA) AND VARIABLE IMPORTANCE

For accessing the differences in the metabolic profiles of the healthy and diseased participants, OPLS-DA was performed (Study 3). Linear discriminant analysis (LDA) is based on finding the linear combinations of variables able to separate classes of a categorical outcome in best possible way making it preferred statistical approach for omics workflow. As with PCA, LDA models considers the differences between the classes of data and also can provide a measure of variable importance that can be used for metabolite ranking.

OPLS-DA **(Trygg and Wold, 2002)** is a supervised, dimension reduction technique used for the classification and visualisation of the difference between the two groups. It is an advanced form of latent variable regression method focussing on the orthogonal variability. Latent variables (LV) are the predictive components calculated from the covariance between dependent and the independent variables. Orthogonal components are calculated as the orthogonal projection to LV explaining the maximum covariance and thus computing the systemic variation that are un-correlated to the group separation i.e reporting for

the effect of the uncorrelated independent variables on a particular dependent variable. And thus, OPLS-DA's discrimination algorithm separates the two groups accounting both the intergroup variation as well as the intragroup variation. This separation method leads to improved predictive model by lowering the number of LVs and including the orthogonal variation.

OPLS-DA was performed using 'ropls' package in R. The model was validated using default model validation parameters using permutation testing and calculating root mean square error of estimation (RMSEE) for assessing the overfitting of the data. RMSEE calculates the distance between the real value of the dependent variable and the estimated value of the same. The R2X, R2Y was calculated to indicate the goodness of fit of the model.

Variables contributing the most to the separation of groups were identified by computing the variable importance in projection (VIP) scores. These are calculated as a weighted sum of the squared correlations between the OPLS-DA component and the original variable indicating the percentage variation explained by the components in the model. Metabolites with high VIP scores (cut of >1) are more important in providing class separation, while those with small VIP provide less contribution.

2.6.6. LEAST ABSOLUTE SHRINKAGE AND SELECTOR OPERATOR (LASSO)

For accessing the relationship between the nutrient intakes, food-groups and disease status, LASSO was performed (Study 3). LASSO (**Tibshiranit, 1996**) is a regularized regression model based on shrinkage method. It predicts the relationship between the dependent and independent variable from a set of highly correlated variables by shrinking the estimate of coefficients to zero and thus investigating the most strongly associated variables with the response variable. The variables belonging to the intakes of different nutrients in Study 3 were found to be highly correlated which doesn't meet the assumptions of simple regressions and therefore, LASSO was opted to investigate the association of nutrients and disease status.

In a multicollinear dataset, the coefficients from the simple linear or logistic regression model are inflated due to overfitting of the data. This overfitting of data is improved by lowering the variance at the cost of increasing the bias by adding a penalty and this is called regularisation. LASSO uses L1 regularisation technique adding a cost to the magnitude of coefficients along with minimizing error between predicted and actual observations. This is called penalty or lambda (λ) which is calculated as sum of the squared coefficients. With increase in the value of lambda, estimate of coefficients started to shrink and becomes closed to zero or zero for the less important variables. Lambda was calculated after default 10-fold cross-validation of the model. Cross validation is a technique where first models are trained on the subset of the input data and then evaluated on the main data. The performance of the model was predicted by computing the binomial deviance which indicates the variance between the predictive model and the perfectly fitted model. Model accuracy was predicted by calculating the Area under the Curve (AUC) indicating the specificity and sensitivity of the model.

LASSO was performed using the 'glmnet' package in R. This function gives two values for λ, a) λ- at which binomial deviance is minimum and b) λ- at which the binomial deviance was within one standard error of the minimum binomial deviance of the largest λ. The diet-disease status model was further fitted using the later value of λ.

2.6.7. MEDIATION ANALYSIS

Mediation analysis (**Z. Zhang *et al.*, 2016**) was performed to test if the gut microbiome mediates the disease state in NAFLD via dietary intake. Mediation models identifies the interaction (direct or indirect) between the observed relationship between an independent variable and a dependent variable via the mediator variable. It is performed to estimate the causal mediation effects of treatments on the outcome.

Mediation is implicated when a) significant association is seen between the mediator and independent variable b) significant or non-significant association is seen between dependent variable and independent variable, c) association is

not seen between dependent and independent variable after regression of dependent variable on independent variable + mediator.

For mediation analysis, a set of linear regression models are fitted and then the estimates of "mediation effects" are computed from the fitted models.

Model 1: $lm\ (M \sim X)$
Model 2: $lm\ (Y \sim M + X + Covariate\ 1 + Covariate\ 2 + \cdots)$

where M is mediator variable, Y is dependent variable and X is independent variable.

Indirect effect is tested using the non-parametric bootstrapping and significance is calculated.

Diet, gut microbiome and metabolome all three have shown to exert an influence in NAFLD. The dietary intakes (nutrients and food groups) were treated as the independent variables and disease outcome was treated as the dependent variable. ACME (Average Causal Mediation effect) represents the difference in the outcome in presence of mediator. ADE (Average Direct Effect) represents the expected difference in the outcome in absence of the mediator.

Mediation was performed using 'mediation' package in R with 1000 Monte Carlo simulations.

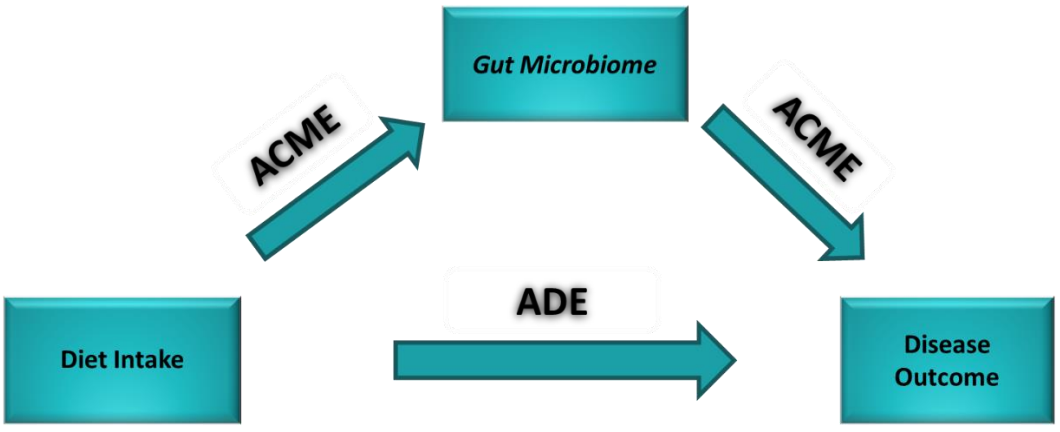


Figure 2.11: Mediation Analysis.

3

Chapter 3

GUT MICROBIAL PROFILE IS ASSOCIATED WITH RESIDENTIAL SETTINGS AND NOT NUTRITIONAL STATUS IN KARNATAKA, INDIA

Frontiers in Nutrition | 8:595756 | doi: 10.3389/fnut.2021.595756

Ojasvi Mehta[†], Leeberk Raja Inbaraj[*†], Stuart Astbury, Jane I. Grove, Gift Norman, Guruprasad P. Aithal, Ana M. Valdes[‡] and Amrita Vijay[*‡]

*Joint first authors, # Joint corresponding authors, $ Joint senior authors

3.1 ABSTRACT[2]

BACKGROUND

Undernutrition is a leading contributor to disease and disability in people of all ages. Several studies have reported a significant association between nutritional status and gut microbiome composition but other factors such as demographic settings may also influence the adult microbiome. The relationship between undernourishment and gut microbiome in adults has not been described to date.

METHODS

We compared the gut microbiome in fecal samples of 48 individuals, from two demographic settings (rural and urban slum) in Karnataka, India using 16S rRNA sequencing. Nutritional status was assessed based on BMI, with a BMI of less than 18.5 kg/m^2 classified as undernourished, and a BMI in the range of 18.5-25 kg/m^2 as nourished. We analyzed 25 individuals from rural settings (12 undernourished and 13 nourished) and 23 individuals from urban slum settings (11 undernourished and 12 nourished).

RESULTS

We found no significant difference in overall gut microbial diversity (Shannon and Unweighted UniFrac) between undernourished and nourished individuals in either geographical setting, however, microbial taxa at the phylum level (i.e. Firmicutes and Proteobacteria) and beta diversity (unweighted UniFrac) differed significantly between the rural and urban slum settings. By predicting microbial function from 16S data profiling we found significant differences in metabolic pathways present in the gut microbiota from people residing in different settings; specifically, those related to carbohydrate and lipid metabolism. The weighted sum of the KEGG Orthologs associated with carbohydrate metabolism (Spearman's correlation coefficient, $\rho = -0.707$, $p<0.001$), lipid metabolism (Spearman's correlation coefficient, $\rho = -0.330$, $p<0.022$) and biosynthesis of secondary metabolites (Spearman's correlation coefficient, $\rho = -0.507$ $p<0.001$) were decreased in the urban slum group compared to the rural group.

[2] Excerpts of this chapter is from the publication cited as "Mehta, O. et al. (2021) 'Gut Microbial Profile Is Associated with Residential Settings and Not Nutritional Status in Adults in Karnataka, India', Frontiers in Nutrition. Frontiers, 8, p. 595756. doi: 10.3389/fnut.2021.595756."

CONCLUSIONS

We report that the geographical location of residence is associated with differences in gut microbiome composition in adults. We found no significant differences in microbiome composition between nourished and undernourished adults from urban slum or rural settings in India.

3.2 INTRODUCTION

Undernourishment is a multifactorial condition caused due to inadequate dietary intake, lack of the essential nutrients in the diet and/or presence of underlying co-morbidities which impair the normal absorption of nutrients. This results in an ill effect on immune health, disease susceptibility and cognitive disabilities (**Saunders and Smith, 2010; Wells et al., 2020**). Over the decade, undernourishment has become a global problem and as per WHO reports it affects 462 million people worldwide **(Fact sheets - Malnutrition, Accessed June 29, 2020**). In India alone, undernutrition is the leading disease burden contributing 17% of the total Disability Adjusted Life Years (DALYs) in all age groups (**Swaminathan et al., 2019**). As per the POSHAN Abhiyan's (National Nutrition Mission - NNM) trend reports (1990-2017) in India **(*POSHAN Abhiyaan - PM's Overarching Scheme for Holistic Nourishment | National Portal of India***), by 2022 there would be an increase of 13·8% for anaemia in women. This could lead to defects in the immunoepigenome, contributing directly or indirectly to the defects in the child through an intergenerational cycle (**Bourke, Berkley and Prendergast, 2016**). Undernutrition in adults has also been shown to have a restraining effect on healthy aging and reducing the life span as this is the forefront mechanistic basis of various diet-related metabolic syndromes like diabetes, heart attacks, strokes etc. Despite the disease burden and impact of undernourishment in adults and onto the next generation, there have been limited studies on undernourished adults.

In LMIC countries, although, poverty associated with food insecurity, is a major risk factor for undernutrition, however, the aetiology of undernutrition is far more complex than the simple lack of food. The gut microbiome which is considered as an organ of the human body provides metabolic functions that are not encoded in the human genome, for instance, fermentation of complex

polysaccharides, metabolism of proteins and peptides, biosynthesis of vitamins, absorption of ions and regulation of several metabolic pathways in the host (**Young, 2017; Liang *et al.*, 2018**). These facilitate the pre-processing of dietary nutrients and efficient harvest of dietary energy for the host. Alterations in the gut microbiome have been recognized as part of the nutritional cycle (**Gordon et al., 2012; Kau et al., 2012**). It is well known that an aberrant microbiota or a microbiota lacking specific micro-organisms could therefore result in reduced nutrient absorption and chronic inflammation (**Belkaid, 2015; Baümler and Sperandio, 2016; Cianci et al., 2018; Lazar et al., 2018**). Several studies have looked at the gut microbiome profile of undernourished children in India and other LMICs **(Gupta et al., 2011; Ghosh et al., 2014; Dinh et al., 2016**). These studies demonstrated an altered microbiota in undernourished children compared to healthy children. Furthermore, randomised controlled nutrition intervention trials targeted specifically to modulate the composition of the gut microbiome have shown to be an effective strategy in improving the nutritional status of undernourished children **(Kerac et al., 2009; Trehan et al., 2013**). There are studies conducted on healthy adults reporting that the residential location i.e. rural vs urban determines the gut microbiome structure **(Tyakht *et al.*, 2013; Ayeni *et al.*, 2018; Hansen *et al.*, 2019**) due to differences in the environment and food availability.

Based on previous findings on the association between the gut microbiome, undernourished children and healthy adults, we hypothesized that in Indian adults the gut microbiome composition will be affected by nutritional status and factors associated with the location. In adults, nutritional assessment is done based on body weight and thinness, where Body Mass Index (BMI) is used as the standard indicator and anthropometric index for assessing undernourishment (International Institute for Population Sciences, 2005). In the current study, we have assessed the gut microbiome composition and 16S rRNA-based functional analysis of 48 adults from rural and urban slum settings in Bangalore district of Karnataka in India with BMI less than 18.5 kg/m^2 categorised as undernourished and BMI ranging from 18.5- 25 kg/m^2 categorised as nourished. This study being the first of its kind in India provides useful insights into how nourishment and demographic settings are related to the gut microbial profile.

3.3 METHODS

The methods opted for this study is presented in **Figure 3.1**.

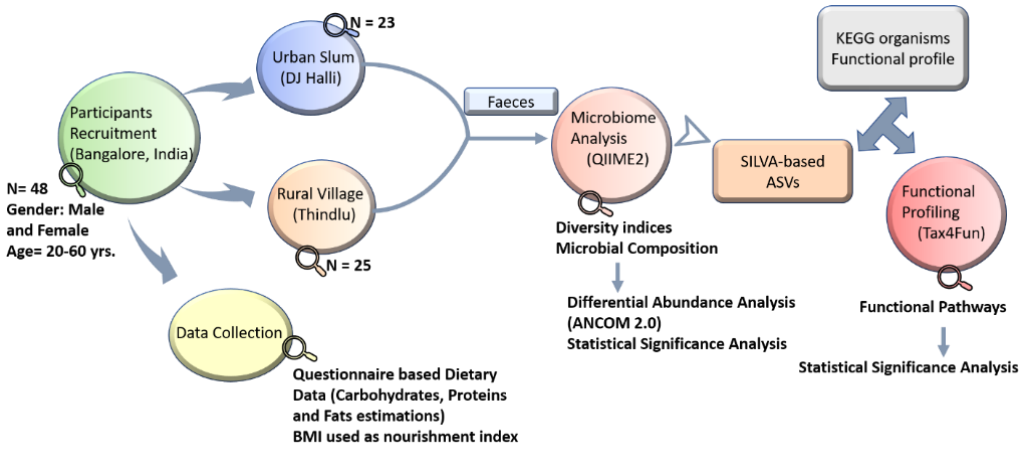


Figure 3.1: Schematic representation of methods for this study.

3.3.1 STUDY SETTINGS

This study was conducted in an urban slum (Devarajeevanahalli (DJ Halli)) and a rural village (Thindlu) in Karnataka. The locations for the study were chosen based on convenience as the community health department has been working in these areas for several years and has a good rapport with the community. The urban and rural settings were chosen based on their distinct and contrasting dietary habits, lifestyle and socioeconomic status (**Wratten, 1995**).

3.3.1.1 URBAN SLUM

DJ Halli is one of the largest government's notified slums in Bangalore, extending over 1.15 km with 420 huts with a 'registered' population of 2463 (***Pages - Slum Statistics***). DJ Halli is served by the Urban Health Centre run by the Department of Community Health, Bangalore Baptist Hospital. Contrary to the official statistics, the population was estimated as 50,000 (approximately 11,000 huts) based on community discussions and observation **(George *et al.*, 2019**). In the urban slum area of DJ Halli, it is reported that 75% of the population falls below the poverty line with only one-third of the population having a regular source of income (**Inbaraj *et al.*, 2020**). This limits the access

to fresh, wholesome produce and comprises predominantly processed foods that are cheaper and therefore affordable.

3.3.1.2 RURAL VILLAGE

According to the 2011 census information, Thindlu has a total population of 786 people. It is located 36km away from Bangalore city (***Thindlu Village in Devanahalli (Bangalore Rural) Karnataka | villageinfo.in***). In contrast to the urban slum, families and individuals in the rural village own pieces of land mainly for agrarian and animal husbandry purposes providing them with a steady source of income and therefore a better socio-economic status compared to the urban slum population. Furthermore, housing in the rural village is built-in with better sanitation and hygiene facilities and ventilation. The population in the rural village normally consumes fresh, farm-grown staples and has a better diet overall.

3.3.2 RECRUITMENT OF STUDY PARTICIPANTS

The current pilot study recruited 25 participants from rural location and 23 participants from the urban slum. Both men and women with no comorbid illnesses and within the age range of 20-60 years were included in the study from both study locations. Potential participants were approached through random door to door home visits in the entire selected area by the Community Health Centre Staff and screened based on the inclusion and exclusion criteria. Individuals with any history of concurrent acute illness; chronic gastrointestinal (GI) disease, chronic constipation, chronic diarrhoea, abdominal tuberculosis; autoimmune disease such as multiple sclerosis or connective tissue disorders; atopic disease like moderate to severe asthma, eczema, eosinophilic disorders; cerebrovascular or peripheral vascular disease, previous antibiotics exposure in past four weeks as four weeks are considered to be adequate for microbiome restoration or who lacked capacity to give informed consent were excluded from the study. Eligible participants were provided with a stool collection kit. Dietary detail was collected using a 24-hour recall questionnaire (**Appendix 3.1**) by interview and total caloric intakes along with the composition of carbohydrates, proteins and fats were determined. Participants who were included in the study

were permanent residents who lived in their respective dwellings for a minimum of one year.

3.3.3 CRITERIA USED FOR ASSESSING UNDERNUTRITION STATUS

BMI was used as the criterion for assessing malnutrition status in the current study according to the standard protocol used to assess nutritional status in India (International Institute for Population Sciences, 2005). BMI in Indian population is categorised as underweight (BMI<18.5 kg/m^2), normal (BMI = 18.5-22.9 kg/m^2), overweight (BMI = 23.0-24.9 kg/m^2) and obese (BMI ≥ 25 kg/m^2) (**Aziz, Kallur and Nirmalan, 2014**). BMI was calculated from the height and weight of the participants. Any male and female in the age group of 20-60 year with a BMI range of 18.5 to 25 kg/m^2 were classified as nourished and BMI less than 18.5 kg/m^2 were classified as undernourished in both rural and urban settings. In the rural area, 12/25 individuals were classified as undernourished and 13/25 were as nourished. In the urban area, 11/23 individuals were classified as undernourished and 12/23 as nourished.

3.3.4 STUDY ETHICS

The study was approved by the Institutional Review Board of Bangalore Baptist Hospital (approval gained on 26.06.2018) and the Health Ministry Screening Committee (HMSC-2018-0548/F1). The current protocol was also approved by the Health Ministry's screening committee (HMSC) through the Indian Council of Medical Research (ICMR), Ministry of Health and Family Welfare.

3.3.5 SAMPLE COLLECTION, PROCESSING AND DNA EXTRACTION

Fresh faecal samples provided by the participants were received at the collection point and then transported on dry ice on the same day of collection to Bangalore Baptist Hospital laboratory and stored at -80°C. Frozen faecal samples were then transported to the Wellcome laboratory in Christian Medical College, Vellore for further analysis. DNA was extracted using the QIAamp Fast DNA Stool Mini Kit (Cat # 51604, Qiagen, Germany). The protocol was modified as per the Standard Operating Procedure of Wellcome Trust Research laboratories. Modification steps were: 180-200mg of stool sample was homogenised in 1ml

InhibitEx buffer. Zirconia beads (0.1 mm, 370 mg) and an enzymatic cocktail (Lysozyme, Mutanolysisn and Lysostaphin) were added and the sample was incubated at a shaker (250 rpm) at 37^0 C for 10 minutes. Cell lysis, protein lysis and RNA degradation were done using SDS and proteinase K and RNAase. The thermal treatment step was performed at 70^0 C for 10 minutes. These steps were followed by the QIAamp kit as per the manufacturer's protocol. Final DNA was eluted in 50ul of buffer and stored at -20^0 C.

3.3.6 SEQUENCING DATA ANALYSIS AND DIVERSITY METRICS

The V3-V5 region of 16S rRNA was amplified and processed following the library preparation protocols for MiSeq Illumina platforms (**Illumina Inc., 2013**). Briefly, the individual libraries were brought down to 4nM and the final pooled library concentration was 4.64nM with 607 as the average library base pair size. Thereafter library denaturation and dilution were carried out considering the pooled library to be 4nM (4000pM). Finally, the library was brought down to 8 pM and was spiked with 20% of PhiX (Internal control) which is also diluted to 8 pM before spiking. The samples were then sequenced to 15 million reads with custom barcoding according to the Fadrosh protocol **(Fadrosh et al., 2014**). Raw reads were demultiplexed, filtered and denoised to derive amplicon sequence variants (ASVs) and taxonomy assignment was done was discussed in the Methodology Chapter section 2.2.4. Microbial abundance was rarefied 37318 reads per sample and microbial diversities were calculated as discussed in Chapter 2 (section 2.2.4).

3.3.7 FUNCTIONAL PROFILING FROM 16S rRNA DATA

The functional profile of microbes from the 16S rRNA data was predicted using the “Tax4fun” tool available at the MicrobiomeAnalyst Pipeline as explained in Chapter-2 (section 2.2.6).

3.3.8 STATISTICAL ANALYSIS

Differential microbial abundance across the study groups was tested by ANalysis of the COmposition of Microbiome (ANCOM 2.0) as discussed in Chapter-2 (section 2.2.5) adjusted for age and gender at 0.9 detection with an

FDR corrected p-value of 0.05. Multiple independent t-tests were performed to determine the statistical differences between microbial diversities, ANCOM-based differentially abundant microbial taxa and functional pathways based on the nutritional status and geographical locations. Spearman correlations were used to determine the association of gut microbial taxa with total calories, estimated nutritional intakes and KEGG functional pathways adjusting for multiple testing (FDR<0.1). All statistical analyses were carried out in R v3.5.2 and SPSS Version 26.

3.4 RESULTS

3.4.1 SUBJECT AND CLINICAL CHARACTERISTICS:

A total of 48 patients were recruited for the study with a BMI of 18.5 to 25 kg/m^2 classified as nourished and a BMI less than 18.5 kg/m^2 classified as undernourished. Participants were categorised as rural-nourished, rural-undernourished, urban slum-nourished and urban slum -undernourished based on their location and nutritional status. The details and clinical characteristics based on location and nutritional status are presented in **Table 3.1A** and based only on the location is presented in **Table 3.2A**. There were no significant differences in age and gender, based on nutritional status (undernourished vs nourished) and location (urban slum vs rural). There were significant differences ($p<0.001$) in the average caloric intakes and estimated intakes of carbohydrates and proteins between undernourished and nourished individuals from both urban slum and rural settings. Based on the 24-hour dietary recalls, the estimated daily intake of calories was less than the recommended daily intake ranging from 1700 kcal to 2200 kcal(Sharma *et al.*, 2020). However, the intakes of carbohydrates, protein and fats followed the recommended daily intakes of carbohydrates (200g- 320g), proteins(42g- 59g) and fats(25g- 30g) in all groups **(*DIETARY GUIDELINES FOR INDIANS-A Manual*)** (**Table 3.1A**). There was no significant difference in the average caloric intake and estimated intake of carbohydrates, proteins and fats based on location (i.e. rural versus urban slum) as shown in **Table 3.2A.**

3.4.2 OVERALL DIFFERENCE AND DIVERSITY INDICES IN THE MICROBIAL COMMUNITY

On average, the most dominant phyla were Firmicutes (68%) followed by Bacteroidetes (18%), Proteobacteria (8%), and Actinobacteria (6%). There were no significant differences between the dominant phyla based on nutritional status as shown in **Table 3.1B**. However, there was a significant increase in the abundance of Firmicutes and Proteobacteria in the rural and urban slum cohort respectively as shown in **Figures 3.2A and 3.2B.** Additional details are provided in **Table 3.2B** and represented in **Figure 3.3**. After taxonomic assignments to the family level, a total of 111 families were found. The independent t-tests on the rarefied table resulted in eight statistically significant families based on location which mostly belonged to the phyla Firmicutes and Proteobacteria (**Table 3.3**). Undernourished adults showed higher alpha diversity (Shannon Index) compared to the nourished in both geographical settings but this was not statistically significant **(Figure 3.2C)**. No significant differences were observed in beta diversity (unweighted UniFrac) between undernourished and nourished in rural or urban slum settings, however, there was a significant difference in the beta diversity based on location (i.e. rural and urban slum) as shown in **Figure 3.2D and Figure 3.4**. Specifically, urban slum settings showed significantly higher beta diversity ($p<0.001$) compared to rural settings. This is mirrored by the Principal Coordinates Analysis PCoA plotted based on the unweighted UniFrac distances (**Figure 3.5)**. The widespread individual samples of urban group demonstrate a greater variation within the urban slum cohort and the narrowly circumscribed individual samples demonstrates the homogeneity in the rural cohort.

Sample-Category	Rural-Nourished	Rural-Undernourished	p -values	Urban Slum Nourished	Urban Slum Undernourished	p-values
A. Clinical Characteristics						
N	13	12		12	11	
Age in years: mean (SD)	31(7.01)	35.33(10.87)	0.244	28.67(8.78)	31.00(12.51)	0.608
Gender- Male	6 (46.16)	6 (50)	0.855	6 (50)	5 (45.45)	0.837
Female: N (%)	7 (53.84)	6 (50)		6 (50)	6 (54.55)	
Weight in Kg: mean (SD)	57.71(9.63)	45.27(5.07)	0.001***	58.42(8.68)	41.94(5.04)	<0.001***
Height in Cm mean (SD)	165.00(12.67)	194.33(112.97)	0.361	162.58(7.28)	158.18(10.30)	0.247
BMI in kg/m^2: mean (SD)	21.37(1.64)	17.45(0.83)	<0.001***	22.05(2.20)	16.74(1.14)	<0.001***
Caloric Intake per day: mean (SD)	1656.53(355.22)	859.66(85.70)	<0.001***	1901.66(517.22)	1150.54(294.62)	<0.001***
Carbohydrates Intake	293.61(45.18)	209.16(34.39)	<0.001***	313.50(67.18)	202.09(10.24)	<0.001***
Protein Intake	51.30(11.45)	42.08(7.12)	0.025**	52.50(11.55)	39.18(4.60)	0.002**
Fat Intake	31.83(3.26)	29.83(3.68)	0.164	31.08(3.62)	28.09(3.50)	0.058
B. Bacterial Phylum Relative Abundance						
Actinobacteria: mean (SD)	0.059(0.032)	0.041(0.032)	0.181	0.045(0.047)	0.078(0.046)	0.110
Bacteroidetes: mean (SD)	0.144(0.082)	0.172(0.082)	0.413	0.225(0.122)	0.165(0.094)	0.201
Firmicutes: mean (SD)	0.744(0.120)	0.746(0.078)	0.957	0.609(0.120)	0.605(0.113)	0.945
Proteobacteria: mean (SD)	0.040(0.052)	0.033(0.034)	0.703	0.111(0.098)	0.139(0.152)	0.610
Others: mean (SD)	0.013(0.026)	0.008(0.014)	0.539	0.008(0.013)	0.011(0.018)	0.561

Table 3.1: Characteristics of the study participants based on nutritional status (normal vs undernourished) and location (rural vs urban slum). A: Clinical Traits, B: Bacterial phylum abundances expressed as relative abundance. The data is expressed as Mean (Standard Deviation). p-values are for independent sample t-tests. ***p-values significant at the level of ≤ 0.001; **p-values significant at the level of ≤ 0.05.

Sample-Category	Rural	Urban Slum	p-values
A. Clinical Characteristics			
N	25	23	
Age in years: mean (SD)	33.08(9.14)	29.78(10.54)	0.252
Gender- Male	12(48)	11 (48)	0.991
Female: N (%)	13 (52)	12 (52)	
Weight in Kg: mean (SD)	51.74(9.91)	50.53(10.96)	0.692
Height in Cm mean (SD)	179.08(78.44)	160.47(8.93)	0.265
BMI in kg/m^2: mean (SD)	19.48(2.37)	19.51(3.22)	0.977
Caloric Intake per day: mean (SD)	1247.04(481.37)	1542.43(566.03)	0.083
Carbohydrates Intake	253.08(58.45)	260.21(74.44)	0.712
Protein Intake	46.88(10.53)	46.13(11.07)	0.811
Fat Intake	30.87(3.54)	29.65(3.80)	0.257
B. Bacterial Phylum Relative Abundance			
Actinobacteria mean (SD)	0.050(0.032)	0.061(0.048)	0.390
Bacteroidetes mean (SD)	0.156(0.081)	0.196(0.111)	0.163
Firmicutes mean (SD)	0.745(0.100)	0.607(0.114)	<0.001***
Proteobacteria mean (SD)	0.036(0.044)	0.124(0.124)	0.002**
Others mean (SD)	0.011(0.020)	0.009(0.015)	0.817

Table 3.2: Characteristics of the study participants based on the location (rural vs urban slum). A: Clinical Traits, B: Bacterial phylum abundances expressed as relative abundance. The data is expressed as Mean (Standard Deviation). p-values are for independent sample t-tests. ***p-values significant at the level of ≤ 0.001; **p-values significant at the level of < 0.05. The p values stayed significant after correcting for multiple testing at the threshold of FDR ≤ 0.1

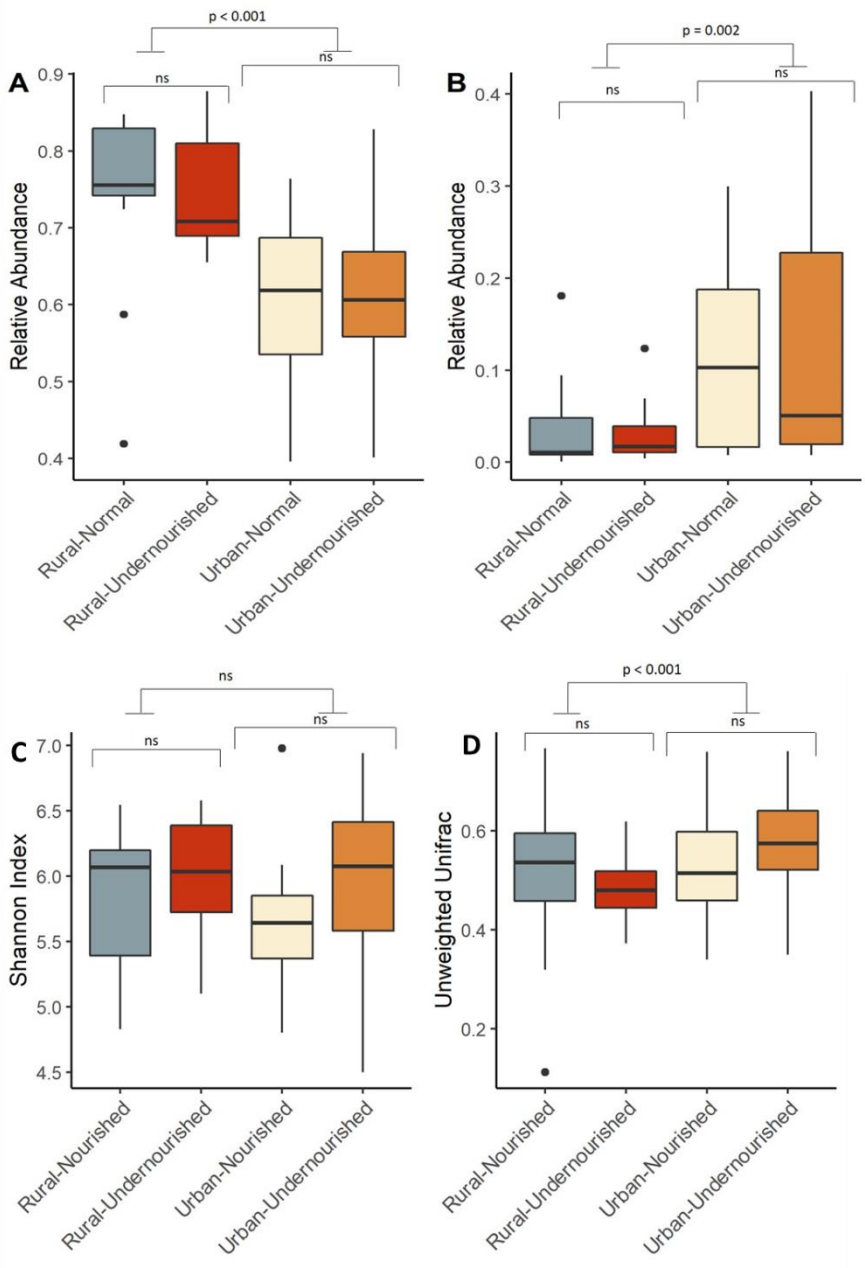


Figure 3.2: Box and Whisker plot depicting relative abundance of significant phylum. (A) Firmicutes and (B) Proteobacteria (C) Shannon Diversity (α diversity index) (D) Beta Diversity (unweighted Unifrac distances) in undernourished and nourished from rural and urban-slum settings. p-values indicate statistical significance from independent t-tests comparing nourished and undernourished study groups in both the locations (i.e. rural normal vs rural undernourished and urban normal vs urban-undernourished) and independent t-tests based on location (i.e. rural vs urban-slum). Box-and-whisker plots show high, low, and median values, with lower and upper edges of each box denoting first and third quartiles, respectively. ns indicates non-significant p-values. Black dots represent the outliers.

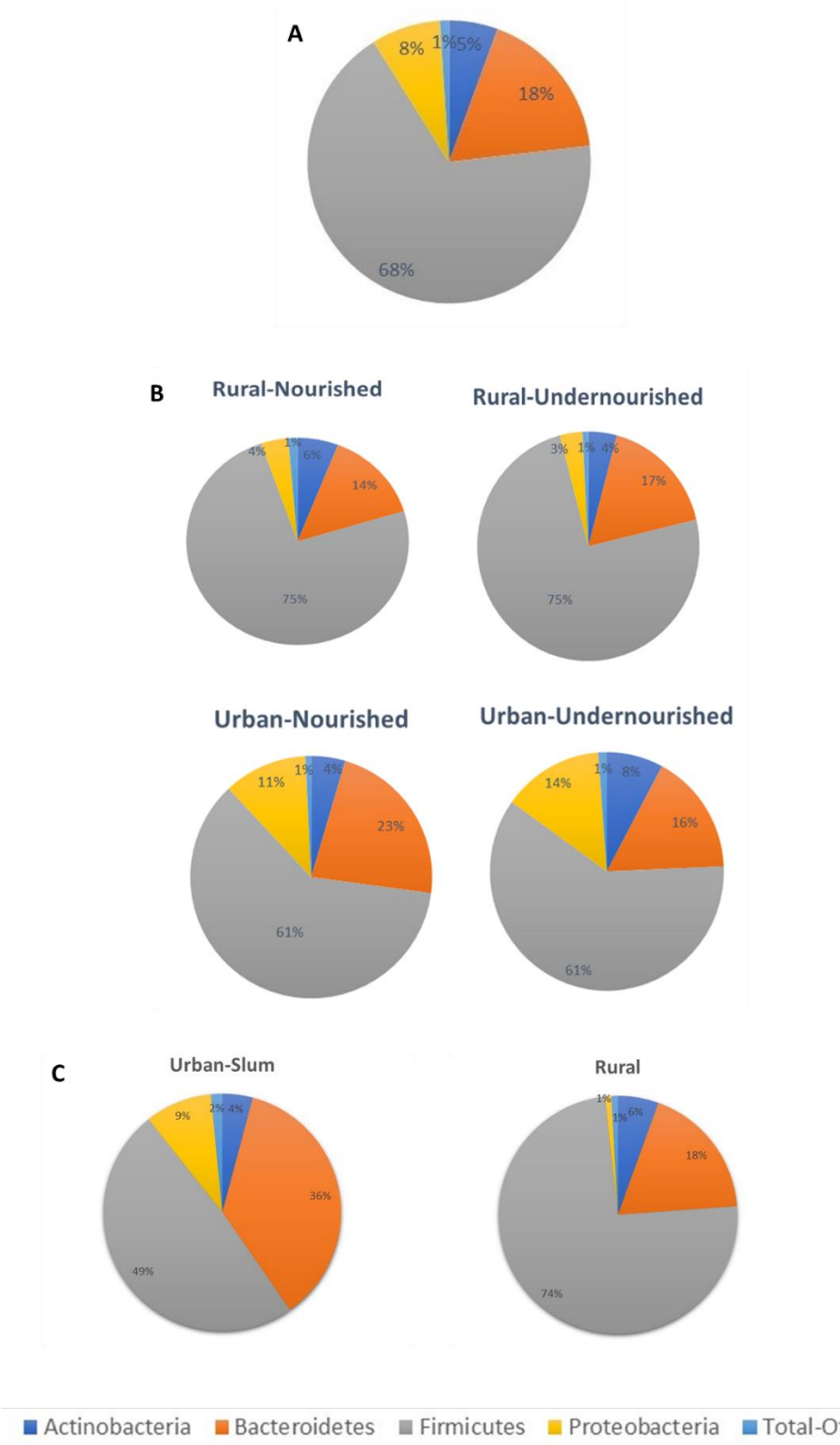


Figure 3.3: Pie Chart representation of the relative abundance of the most dominant bacterial phyla. A) Across all the study groups, B) Across the undernourished and nourished from rural and urban slum settings, C) Across the location (rural and urban slum).

Family	Rural	Urban Slum	p-value
Enterococcaceae mean(SD)	06.24(15.70)	137.74(237.06)	0.008**
Clostridiaceae 1 mean(SD)	962.64(1053.01)	226.43(404.31)	0.003**
Lachnospiraceae mean(SD)	8122.64(2638.45)	4362.35(1709.08)	<0.001***
Peptococcaceae mean(SD)	03.28(5.48)	14.26(26.61)	0.049**
Peptostreptococcaceae mean(SD)	1067.84(1258.60)	194.83(251.84)	0.002**
Ruminococcaceae mean(SD)	10651.92(2944.97)	8377.35(3663.59)	0.022**
Veillonellaceae mean(SD)	1650.56(1582.55)	3410.87(2616.60)	0.007**
Enterobacteriaceae mean(SD)	994.40(1352.90)	4402.35(4728.93)	0.001***

Table 3.3: Abundances of taxa at the family level based on location (i.e. rural vs urban slum). The data is expressed as Mean (Standard Deviation). p-values are for independent sample t-tests. ***p-values significant at the level of ≤ 0.001; **p-values significant at the level of < 0.05. The p values stayed significant after correcting for multiple testing at the threshold of FDR ≤ 0.1

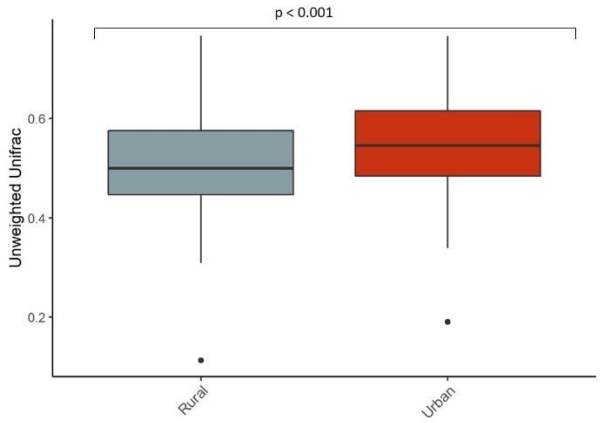


Figure 3.4: Box and Whisker plot depicting Beta Diversity (unweighted Unifrac distances) based on location. p-values indicate independent t-tests statistical significance comparison between nourished and undernourished study groups in both the locations (i.e. rural vs urban slum) and then categorizing the study groups based on the location. Box-and-whisker plots show high, low, and median values, with lower and upper edges of each box denoting first and third quartiles, respectively. ns indicates non-significant p-values. Black dots represent the outliers.

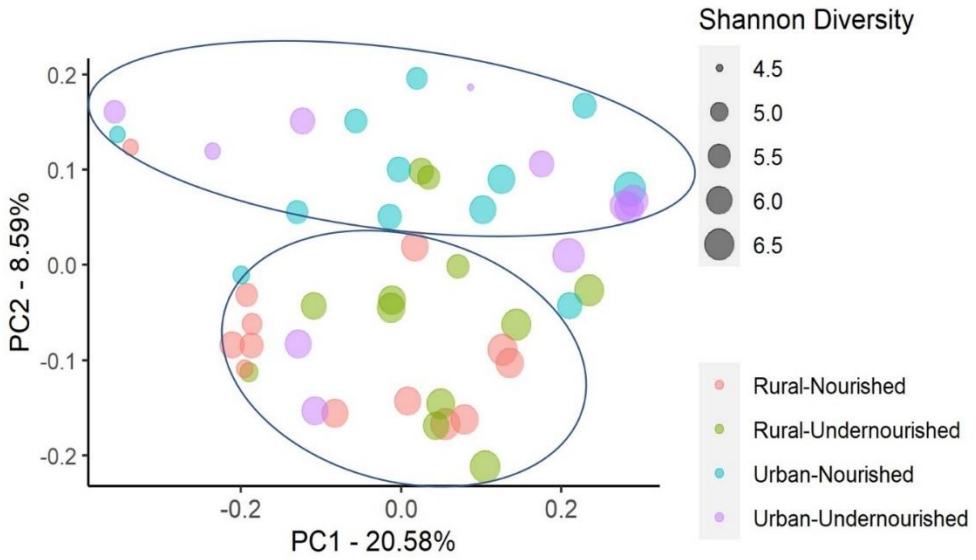


Figure 3.5: Principal Coordinates Analysis (PCoA) of the similarity across the study groups based on the nutritional status (nourished vs undernourished) and location (rural vs urban slum) based on the Unweighted UniFrac distances. Each point corresponds to a sample coloured according to the nutritional status and corresponding location and the size of the point corresponds to the Shannon diversity index. The principal Coordinates (PC) axis expresses the percent variance across the samples.

3.4.3 LOCATION IS THE STRONGEST DETERMINANT OF THE GUT MICROBIOTA VARIATION ACROSS THE SAMPLES

There were no differentially abundant taxa when tested based on nutritional status. However, taxa significantly associated with the rural settings were *Erysipelotrichaceae UCG-003*, *Turicibacter* and *Marvinbryantia*. In the urban slum region, the taxa that were significantly associated belong to the genus *Megasphaera*, *Prevotella* and *Prevotellaceae* family as shown in **Figure 3.6**. There were no taxa significantly associated with the estimated intakes of carbohydrates, proteins and fats after correcting for multiple testing at the threshold of FDR <0.1.

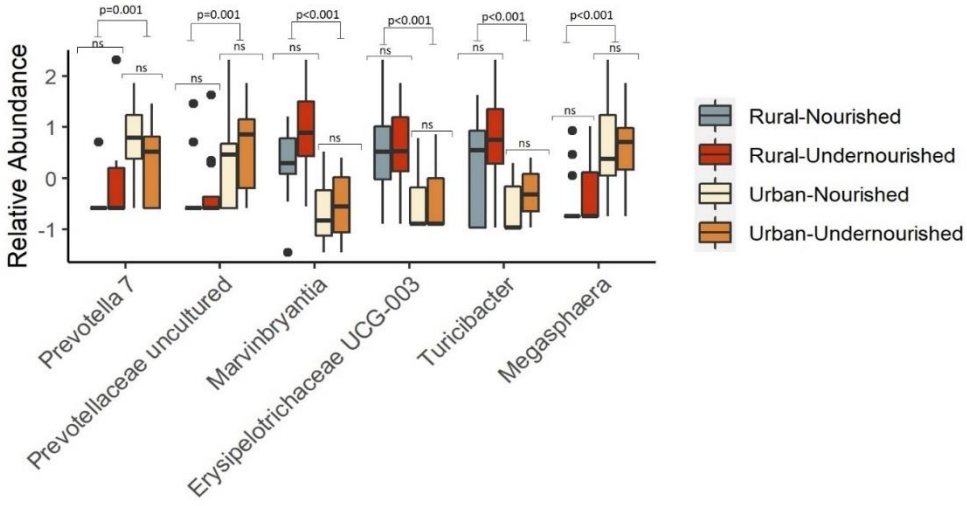


Figure 3.6: Box and Whisker plot depicting the relative abundance of significant taxa associated with undernourished and nourished from urban slum and rural settings. p-values indicate statistical significance from independent t-tests comparing nourished and undernourished study groups in both the locations (i.e. rural normal vs rural undernourished and urban normal vs urban-undernourished) and independent t-tests based on location (i.e. rural vs urban-slum). Box-and-whisker plots show high, low, and median values, with lower and upper edges of each box denoting first and third quartiles, respectively. ns indicates non-significant p-values. Black dots represent the outliers.

3.4.4 KEGG Orthologs and Functional Pathways

Based on the 16S rRNA marker gene amplicon data, we compared the functional gene content based on nutritional status and geographical locations respectively. Although alpha diversity based on the genus level did not show a statistical difference in either of the study groups, reports suggest there may be differences at the functional level **(Das *et al.*, 2018; Ortiz-Estrada *et al.*, 2019**). To gain a deeper insight into whether microbiota compositional differences may have effects within the subgroups, we computed the abundance of higher functional categories (based on KEGG ortholog abundances).

A KEGG ortholog can be a part of many functional pathways, so the functional profiles across the study groups were computed based on the sum of weighted hits of the KEGG orthologs present across the samples in the study groups. Based on nutritional status, there were no significant associations of specific functions, however, pathways for carbohydrate and lipid metabolism, glycan biosynthesis and nucleotide metabolism along with biosynthesis of other amino acids, metabolism of terpenoids and polyketides, xenobiotic biodegradation and metabolism were found to be negatively associated with the nourished cohort. Functions for the metabolism of amino-acid and energy, metabolism of cofactors and vitamins, and biosynthesis of secondary metabolites were found to be positively associated with the nourished cohort as shown in **Table 3.4**. We then

looked at the associations with geographical location and found significant associations for the functional pathways of biosynthesis of other secondary metabolites, carbohydrate metabolism and lipid metabolism ($p<0.05$), as shown in **Figure 3.7**. Specifically, pathways identified for amino acid, nucleotide, lipid, carbohydrate and glycan biosynthesis metabolism along with xenobiotics biodegradation, biosynthesis for secondary metabolites and metabolism of terpenoids and polyketides were negatively associated with the rural settings and functions for energy metabolism, metabolism of cofactors and vitamins and other amino acids were found to be positively associated with the rural settings as shown in **Table 3.4**.

Pathways	Nutritional Status (Nourished vs Undernourished)		Location (Rural vs Urban Slum)	
	Coefficient	p-value	Coefficient	p-value
Amino-acid metabolism	0.071	0.633	-0.083	0.576
Biosynthesis of secondary metabolites	0.056	0.707	-0.507 ***	< 0.001
Carbohydrate metabolism	-0.113	0.445	-0.727 ***	< 0.001
Energy metabolism	0.011	0.943	0.089	0.548
Glycan biosynthesis and metabolism	-0.122	0.409	-0.059	0.692
Lipid metabolism	-0.113	0.445	-0.330 **	0.022
Metabolism of co-factors and vitamins	0.017	0.911	0.074	0.618
Metabolism of other amino- acids	-0.047	0.753	0.059	0.692
Metabolism of terpenoids and polyketides	-0.047	0.753	-0.254	0.081
Nucleotide metabolism	-0.011	0.943	-0.173	0.239
Xenobiotic biodegradation and metabolism	-0.014	0.927	-0.05	0.737

Table 3.4: Correlation between the pathways identified and the study groups based on the nutritional status (normal vs undernourished) or location (rural vs urban slum). Data is expressed as Spearman's coefficient and raw p-values in parentheses. **correlation is significant at the level of <0.05; ***correlation is significant at the level of <0.001. The p values stayed significant after correcting for multiple testing at the threshold of FDR ≤ 0.1.

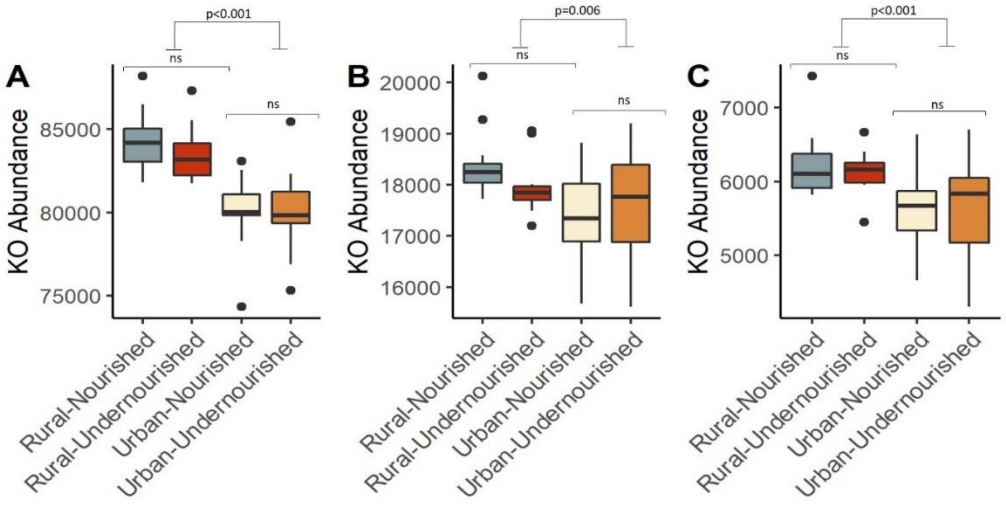


Figure 3.7: Box and Whisker plot depicting the pathways significantly associated with undernourished and nourished from urban slum and rural settings. (A) Carbohydrate metabolism ($p < 0.001$), (B) Lipid metabolism ($p = 0.006$), (C) Biosynthesis of secondary metabolites ($p < 0.001$). p-values indicate statistical significance from independent t-tests comparing nourished and undernourished study groups in both the locations (i.e. rural normal vs rural undernourished and urban normal vs urban-undernourished) and independent t-tests based on location (i.e. rural vs urban-slum). Box-and-whisker plots show high, low, and median values, with lower and upper edges of each box denoting first and third quartiles, respectively. ns indicates non-significant p-values. Black dots represent the outliers.

3.5 DISCUSSION

In the current study, we looked at the impact of nutritional status and geographical location of residence as possible factors influencing the composition and function of the gut microbiome of adults in Karnataka, India. We find no significant differences in microbial diversities (Shannon and Unweighted UniFrac) between undernourished (BMI < 18.5 kg/m^2) and nourished individuals (BMI in the range 18.5-25 kg/m^2) from either rural or urban slum locations even though calorie, carbohydrate and protein intake was significantly lower in the undernourished group (Table 1A). This is consistent with several studies reporting that the reduction in caloric intake can change the gut microbiome without significantly changing alpha diversity **(Pataky et al., 2016; Ott et al., 2017**). This may also be because, in adults, the core microbiome becomes stable over time **(Faith et al., 2013; Radjabzadeh et al., 2020**), which tends to change with the effect of the environment and other factors **(Li, Bihan and Methé, 2013; Mortensen et al., 2018**). Beta diversity was however found to be significantly different between locations (rural and urban slum settings) reflecting that the variation of microbial taxa across the participants from rural settings is homogenous i.e. less diverse in comparison to the variation across the participants in urban settings which had more outliers. This is paralleled with the findings that there are several differentially abundant microbial taxa based on rural vs urban settings. This suggests that location has a significant effect on modulating the composition of the microbiome and associated metabolic pathways as reported previously **(Monaghan *et al.*, 2020**). These findings reflect the substantial differences in food availability, environment and social practices between the urban slum and rural settings. This difference is also supported by the fact that the economic structure of developing countries like India has been shifting from agriculture towards industry and urbanization, resulting in a nutritional transition of switching from traditional staple foods to nutrient-lacking non-staple food, which may be contributing to nutritional differences (**Pingali *et al.*, 2019**). For example, in the study settings of the current study, urban slum settings comprise predominantly overcrowded and ill-ventilated houses with poor sanitation practices and lower socioeconomic status compared to the rural settings. Furthermore, the people in the urban slum generally have a

diet comprising mainly processed foods while the rural population has a higher socioeconomic status with a healthier living environment and a diet comprising fresh traditional staple foods **(Galler *et al.*, 2012; Tette, Sifah and Nartey, 2015**). This is also reflected in lower average calorie intakes per day in the rural settings compared to the urban slum, however, we observe similar protein, carbohydrates and fat intakes. Our work supports the previous suggestion that extrinsic factors beyond nutritional status appear to be associated with microbial profiles **(Price, Dehal and Arkin, 2010; Tyakht *et al.*, 2013; Ayeni *et al.*, 2018; Das *et al.*, 2018; Hansen *et al.*, 2019**).

In the current study, a significant increase in Proteobacteria and a decrease in Firmicutes in the urban slum settings may indicate associations with the risk of diseases as reported previously **(Rizzatti *et al.*, 2017; Vester-Andersen *et al.*, 2019; Doumatey *et al.*, 2020**). Also, the significant association of genera such as *Prevotella* and the Prevotellaeceae family in the urban slum settings may reflect the differences in the enzymes responsible for the breakdown of complex indigestible polysaccharides (**Wu *et al.*, 2011**). The family Prevotellaeceae and different oligotypes of the genus *Prevotella* have been previously shown to be associated with different dietary patterns (**De Filippis *et al.*, 2016**) and with rural and urban locations **(Bhute *et al.*, 2016; Das *et al.*, 2018; Monaghan *et al.*, 2020**). In the current study, we also find significant increases in butyrate-producing genera such as *Marvinbryantia* and *Turicibacter* associated with the rural location which may indicate a fibre-rich diet. The presence or absence of a set of microorganisms may predict the healthy vs diseased state or susceptibility to disease but microbial diversity cannot predict differences in functions. To understand differences at the functional level, we used the 16S data to extrapolate differences at the level of biological functions and associated pathways. The higher-level functional pathways related to carbohydrate, protein and lipid metabolism were also found to be significantly associated with the rural location compared to urban slum location. The differences in the content of carbohydrate and protein intake across the two demographic areas may be a possible factor as studies have reported that the carbohydrate and protein metabolism affect the expression of branched-chain amino acids catabolic pathways and energy demands through the TCA cycle **(Kuzuyama, 2017**). These differences could be attributed to migration from rural to urban settings

which have been previously reported to impair metabolic pathways (**Tahapary *et al.*, 2018**) and gut microbiome composition **(Fujisaka *et al.*, 2018; Jiang *et al.*, 2019**). Although, the gut microbiome has been shown to affect the metabolism of macronutrients with the help of enzymes such as of CAZymes, deaminases, lipases and others maintaining the homeostasis of metabolic pathways (**Rowland *et al.*, 2018**), however, the causal factors contributing to this remain speculative. In addition, the positive association of the xenobiotic degradation pathways and the pathways for the metabolism of terpenoids and polyketides with the urban slum population likely reflects the exposure to industrial products as these molecules exist as glycosides in their natural forms and are chemically engineered in the pharmaceutical and nutraceutical industries **(Mrudulakumari Vasudevan and Lee, 2020**).

Various previous studies have reported no change in diversity metrics but a change in functional profiles that are associated with specific metabolic pathways **(Lozupone *et al.*, 2012; Wagner *et al.*, 2018**). Our observed variation in beta diversity, differences in specific microbial taxa and functional pathways probably reflect the intake of substantially different dietary resources in rural and urban slum dwellers (e.g. processed foods more commonly available in urban location). However, this study only collected data from 24-hour dietary recall and therefore, the speculation of difference in diet needs to be confirmed in large dataset with more information on the dietary pattern over a longer period of time.

There are substantial strengths to our study. This is the first study that has investigated the role of nutritional status and location of the residence on the composition of the gut microbiome in undernourished and nourished adults. The inclusion of undernourished and nourished individuals from two distinct locations within the same region enabled us to gain a deeper understanding of the impact of nutritional status and environment on the composition of the gut microbiome. Also, we have inferred the functioning of the gut microbiome from the 16s rRNA marker-based functional profiling tools. Few studies have compared the outputs of pipelines such as PICRUST and Tax4Fun **(Langille *et al.*, 2013; Aßhauer *et al.*, 2015**) with shotgun metagenome sample sequencing generating comparable results thus making the process cost-effective against the shotgun metagenomic sequences **(Das *et al.*, 2018; Sun, Jones and Fodor,**

2020). Although there are many caveats to the functional profile generated from the 16S rRNA marker gene-based microbial profile, the current analysis provides some indication that the gut microbiome although not significantly different in diversity may still alter metabolic capacity by producing secondary metabolites that alters gene expression and associated metabolic pathways. Our study has several limitations. Firstly, the small sample size reduces the power to show significant associations and therefore needs to be replicated on a larger sample size. Secondly, dietary data was limited to a single administration of a 24-hour food recall diary rather than a standardised Food Frequency Questionnaire (FFQ) restricting analysis of dietary patterns including the information only on macronutrients intake but not on micronutrients intake and duration of nutritional status between urban slum and rural populations. Furthermore, the study also lacked detailed information on lifestyle as a specific questionnaire on lifestyle was not administered. The current study also lacked biomarkers for the cross-validation of nutritional status and the composition of the gut microbiome.

In conclusion, the current study suggests that in adults, the composition of the gut microbiome is driven more by the demographic settings of residence than nutritional status.

Contrary to previous studies that have shown significant differences in the microbial diversities of the gut microbiome amongst undernourished children **(Gupta *et al.*, 2011; Ghosh *et al.*, 2014**), no associations were observed in undernourished adults from either rural or urban settings. Microbiome differences associated with the location of residence and differences in host metabolic pathways may have important consequences on human health influencing metabolism, immunity, development and behaviour **(Sharon *et al.*, 2014**). It may therefore be of value to consider environmental and social factors to address dysbiosis in adults in LMICs. This may enable some health improvement in undernourished individuals. Further studies will be required to understand the interplay between the nutritional status and residual location, socioeconomic factors and gut microbiome for improved health in adults.

4

Chapter 4

GUT MICROBIOME AND SERUM METABOLOME SIGNATURES OF PAIN INTENSITY IN KNEE OSTEOARTHRITIS

METBOLOMIC SIGNATURES RESULTS ARE COMMUNICATED IN JOURNAL OF PAIN

4.1 ABSTRACT[3]

INTRODUCTION

Osteoarthritis (OA) is the most common arthritis affecting synovial joints such as knees and hips of millions of people globally. Usage-related joint pain and reduced function are the most common symptoms experienced by people with OA. To improve pain management, there is a need to identify validated biomarkers predicting therapeutic responses in targeted clinical trials or microbial community associated with pain. This study aimed to identify the gut metabolic biomarkers and microbial signatures for pain and pressure pain detection thresholds (PPTs) in participants with knee pain and symptomatic OA using gut microbiome community profiling and metabolic phenotyping.

METHODS

Metabolite and cytokine measurements were done on serum samples using LC-MS/MS (liquid gas chromatography integrated magnetic resonance mass spectrometry) and Human Proinflammatory panel 1 kit respectively. Regression analysis was done in a test (n = 75) and replication study (n = 79) to investigate the metabolites associated with current knee pain scores and pressure pain detection thresholds (PPTs). Meta-analysis and correlation were done estimating precision of associated metabolites and identifying relationship between significant metabolites and cytokines respectively. Gut microbiome sequencing for Webex datset (n = 75) was performed on the MiSeq platform and analysed using QIIME2.

RESULTS

Acyl ornithine, carnosine, cortisol, cortisone, cystine, DOPA, glycolithocholic acid sulphate (GLCAS), phenylethylamine (PEA) and succinic acid were found to be significantly (FDR < 0.1) associated with pain scores in meta-analysis of both studies. IL-10, IL-13, IL-1β, IL2, IL8 and TNF-α were also found to be associated with the significant metabolites. Low abundance of *Parasutrella* and

[3] Excerpts of this chapter explaining role of metabolome is from the publication entitled as "Serum Metabolome Analysis identified amino-acid metabolism associated with pain in people with symptomatic knee Osteoarthritis – a cross-sectional study."

high abundance of *Streptococcus* were found to be most strongly and significantly associated with increase in the knee pain.

CONCLUSION

Significant associations of ten metabolites and inflammatory markers with knee pain suggests that targeting relevant pathways of amino acid and cholesterol metabolism may modulate cytokines and these could be targeted as novel therapeutics development to improve knee pain and OA management. Associations of *Parasuttrella* and *Streptococcus* with increased knee pain needs to be further explored in larger cohort study.

4.2 INTRODUCTION

Osteoarthritis (OA) is the most common form of arthritis and overall 44 % of the UK adult population suffers from OA pain (**Swain *et al.*, 2020**). Altered metabolism, inflammation, joint injury and other factors contribute to the development of OA (**Deveza *et al.*, 2017**). The most common clinical manifestation leading OA patients to seek medical attention is pain **(Guccione *et al.*, 1994**) which in 4 out of 10 people, is not controlled by simple medication. Certain risk factors such as older age, female gender, and high BMI associate with OA, however, the pathophysiological mechanisms of pain in OA remain unclear. Independent of the degree of peripheral tissue or nerve damage, pain sensitivity is reported to be different between individuals **(Ossipov, Dussor and Porreca, 2010**). This is attributed to the fact that pain experience in OA is not only through peripheral nociception but can be modified through central modulation. These subjective pain behaviour cause difficulty in assessment of pain and at later stages may contribute to the high prevalence of chronic pain.

Current pharmacological therapies for pain have limited efficacies, restricting functional recovery in OA and encouraging eventual reliance upon opioids which are known to have multiple adverse effects (**Reinecke *et al.*, 2015**). This is likely to be related to the lack of understanding of the pain pathways and targeted therapeutics against disease progression. Various studies linking the change in metabolites and/or the gut microbiome is already reported to be associated with the inflammation in various other diseases. Metabolic changes

mediated through host factors, such as heritability, Single Nucleotide Polymorphisms (SNPs) of genes, and microbial metabolites are also reported to change pain perception **(Evangelou *et al.*, 2009; Raine *et al.*, 2013; Rea *et al.*, 2019**).This predicts the potential role of the microbes and metabolites in the inflammation causing knee OA and associated pain when other factors such as injury or ageing is not associated. There is a growing body of literature supporting evidence of serum metabolites being linked to chronic back pain, abdominal pain, joint pain and neuropathic pain (**Finco *et al.*, 2016; Grimby-Ekman *et al.*, 2017; B. Lin *et al.*, 2020**). However, there has been no detailed investigation on role of metabolites in specifically knee pain in OA patients.

The hypothesis of microbes being involved in the pathophysiology of knee OA was first generated when the microbial DNA was isolated from the synovium of the knee joints. Gut microbiome is reported to be signalling various neuropeptides which directly or indirectly controls or provides tolerance to pain through nervous system (**Rea *et al.*, 2019**). However, the mechanistic insights that how the microbes cause the disease however remains speculative. The growing evidences of crosstalk between gut microbiome and serum metabolome reflects the contribution of both these factors to the homeostasis and knee health. These factors highlight the importance of development of certain tools such as microbes and/or metabolites signifying pain and that can be targeted as potential therapeutics. Differences in the gut microbiome (**Boer *et al.*, 2019a**) and serum metabolites in the joint tissues, such as neurotransmitters, amino acids, hormones and various others **(Vincent and Tracey, 2008; Shell *et al.*, 2016; Yang and Chang, 2019**) potentially stimulate the immune response and thus are likely to be involved in the pathophysiology of OA pain. Based on these evidences, the current study hypothesized that the gut microbiome and biochemical change in serum metabolites such as amino acids and neurotransmitters are associated with pain in OA via related molecular pathways linked to chronic pain.

To date, there have not been any study investigating both the microbiome and metabolome specifically into the gut microbiome and serum metabolome of the patients with OA. The aim of this study was to investigate the relationship of gut microbiome and metabolites with both self-reported pain scores and pressure pain threshold traits in people with OA to help identify microbial community

and metabolic signatures associated with pain and pathways related to pain sensitivity in knee pain in OA patients.

4.3 METHODS

4.3.1 STUDY DESIGN AND PARTICIPANTS RECRUITMENT

In the current study, we compared the relationship of the metabolome with pain traits in two independent community-derived cohorts, the Webex (Web-based exercise intervention) cohort (n = 75) (**Gohir *et al.*, 2019**) which was used as the test study and the KPIC (Knee Pain in the Community) cohort (n = 79) (**Fernandes *et al.*, 2017**) which was used for the replication study. Details of two studies are elaborated in section 2.1.2 (Chapter 2).

For investigating the role of microbiome in pain and pain traits in knee OA, microbiome data from Webex study was analysed and associations with pain and pain traits were investigated. Study flowchart is depicted in **Figure 4.1**.

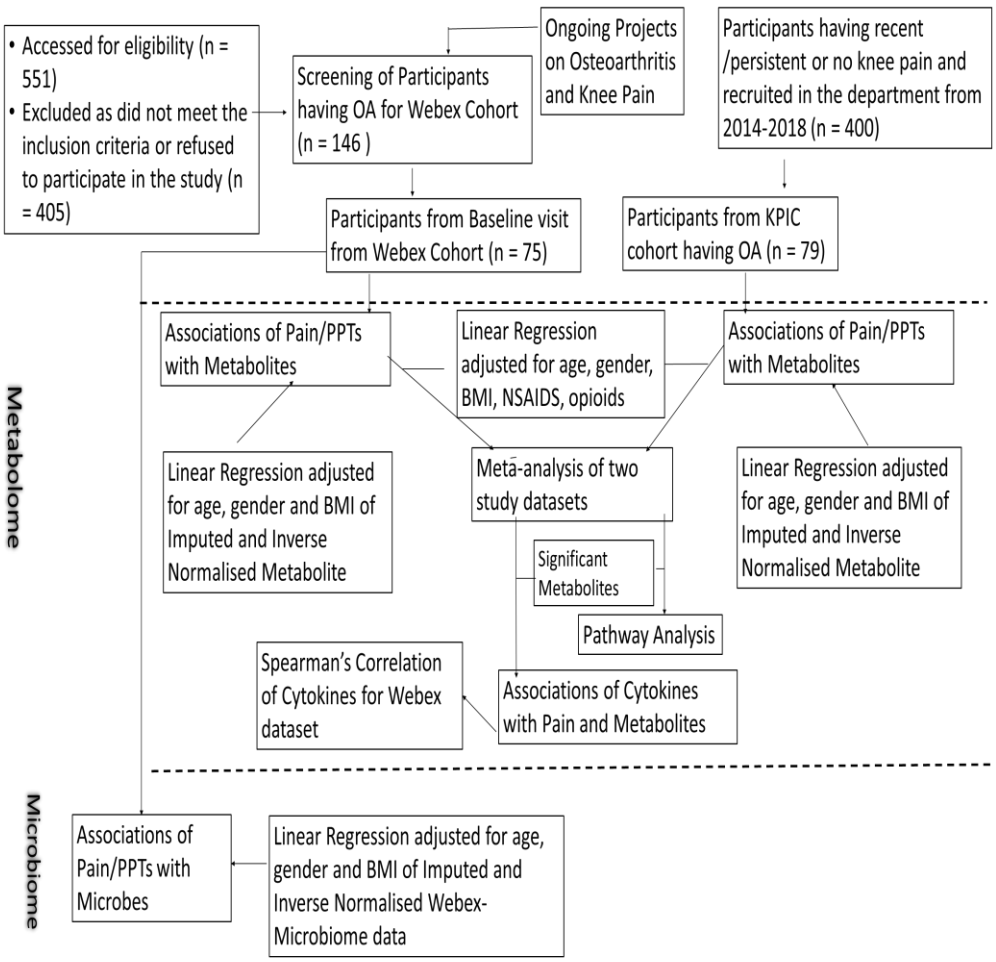


Figure 4.1: Flow diagram showing study design, conduct and analysis. For metabolic signature analysis, associations with pain and pain traits were compared between two independent studies where the Webex study was used as test dataset and KPIC study was used as replication dataset. For gut microbiome signatures in knee pain, associations were tested for gut microbiome data from only the Webex data set.

4.3.2 CLINICAL EVALUATION AND DATA COLLECTION

Clinical traits data was evaluated based on the self-reported medical history as mentioned in the questionnaire following the inclusion and exclusion criteria as discussed in the protocols published (**Fernandes *et al.*, 2017; Gohir *et al.*, 2019)**. Briefly, pain was measured as current knee pain while sitting at the time of visit using Numerical Rating Scale (NRS) on the scale of 0-10 where 0 is no pain and 10 is the worst pain. Pressure Pain detection Thresholds (PPTs) were measured in triplicate with an algometer and were assessed by applying certain pressure over a region of body and then steadily increasing nonpainful pressure stimulus turning into a painful pressure sensation, determining the pain perception and mechanical pain sensitivity. Our study measured these PPTs using test sites at the superomedial patella, superolateral patella, medial joint and anterior tibialis body regions were measured. The mean of PPT values for each body site were computed from all three PPT rounds of measurement for further analysis. Details on use of medications such as NSAIDs and opioids were recorded from the questionnaires.

4.3.3 SAMPLE ACQUISITION AND PROFILING

4.3.3.1 METABOLITE PROFILING

Serum from participants who provided early morning fasting blood samples was extracted for metabolomic profiling using an integrated approach of high throughput magnetic resonance mass spectrometry and liquid gas chromatography (LC-MS/MS) using the commercially available Biocrates MxP Quant 500 kit (Biocrates Life Science AG, Innsbruck, Austria) **(Roberts *et al.*, 2020**). The metabolites concentration was analysed using Sciex Analyst® and data are imported into Biocrates MetIDQ™ software for further analysis. A total of 630 metabolites were quantified belonging to the class of acylcarnitine, alkaloids, amine oxide, amino acids, bile acids, biogenic amines, carboxylic acids, ceramides, cholesterol esters, cresols, diacylglycerols, dihydroceramide, fatty acids, glycerophospholipids, glycosylceramides, hormones, indole derivatives, nucleobases, sphingolipids, sugars, triglycerols, vitamins and cofactors.

4.3.3.1 DNA EXTRACTION, MICROBIOME SEQUENCING and ANALYSIS

Stool samples were collected by participants at home and were brought to the study visit or were sent by post using previously provided specimen collection kit and transportation kit (Safebox^{TM,} Royal Mail) and were stored at −80°C. DNA was extracted from each sample using the PowerSoil® - htp DNA isolation kit (MoBio Laboratories Ltd, Carlsbad, CA) using 100 mg of the stool sample (**Goodrich *et al.*, 2014**). There was no homogenization prior to this step. Gut microbiome composition was determined by 16S rRNA gene sequencing carried out as previously described (Goodrich et al., 2016; Jackson et al., 2016). Briefly, the V3-V4 region of the 16S rRNA gene was amplified using universal primers 355F (CCAGACTCCTACGGGAGGCAGC) and 806R (GGACTACHVGGGTWTCTAAT). Amplified DNA was sequenced following the library preparation protocols for MiSeq platform. Briefly, PCR reactions, carried out consisting 2.5 U Easy-A high-fidelity enzyme, 1 × buffer (Stratagene, La Jolla, CA), 10-100 ng DNA template, and 0.05 μM of each primer. PCR amplicons were quantified using the QuantiT PicoGreen dsDNA Assay Kit (Invitrogen, Carlsbad, CA). Aliquots of amplicons (at equal masses) were combined for a final concentration of approximately 15 ng/μl.

Raw reads were demultiplexed, filtered and denoised to derive amplicon sequence variants (ASVs) and taxonomy assignment was done was discussed in the Methodology Chapter section 2.2.4. Microbial abundance was rarefied 13167 reads per sample and microbial diversities were calculated as discussed in Chapter 2 (section 2.2.4).

4.3.3.3 CYTOKINE MEASUREMENTS

Pro and anti-inflammatory serum cytokines measurement were performed by Affinity Biomarkers, London using the standardized Human Proinflammatory panel 1 assay kit (cat number K151A0H-1), distributed by Meso Scale Discovery and levels of IFN-γ, IL10, IL-12P70, IL-13, IL-1β, IL2, IL4, IL6, IL8, TNFα were measured and concentrations are reported in ng/L.

4.3.4 METABOLITE DATA PRE-PROCESSING

The metabolite data were pre-processed to impute missing data. This was done by replacing missing values with half of the minimum positive value excluding zero. Data were tested for normality and was inverse normalised before downstream statistical analysis.

4.3.6 STATISTICAL ANALYSIS

The current study investigated associations of metabolites (n=93) belonging to the class of amino acid related, bile acids, biogenic amines, carboxylic acids, cresol, hormones, indole derivative, vitamins and cofactor, nucleobase, amine-oxide and amino acids. Normality tests for clinical traits was done using Kolmogorov-Smirnov tests and significance comparison between the two studies was performed with the independent-samples Student t-test, or the Mann–Whitney *U* test as applicable. To investigate the metabolites or microbes associated with change in pain scores or PPTs, linear regression analysis adjusted for age, gender, BMI as confounding factors was performed with metabolites as the outcome and pain traits as the predictor (model 1). Regression analysis for significant metabolites (n =26) was then adjusted for age, gender, BMI and intakes of NSAIDs and opioids (model 2).

$$\text{Model 1: lm (metabolites} \sim \text{pain trait} + \text{age} + \text{gender} + \text{bmi)}$$

$$\text{Model 2: lm(metabolites} \sim \text{pain trait} + \text{age} + \text{gender} + \text{bmi} + \text{NSAIDs} + \text{opioids)}$$

Meta-analysis of the two studies using the fixed random effect model and comparing the regression coefficients from model 2 was done using the REML model using the “metafor” package in R (**Partlett and Riley, 2017**). Pathway analysis was done using the Pathway Analysis tool implemented in the MetaboAnalyst pipeline (**Pang *et al.*, 2021**) based on the KEGG metabolic pathways library (**Kanehisa and Goto, 2000**) for humans. The over-representation was tested based on hypergeometric tests which calculate p-value as the probability of randomly drawing k or more successes from the population in n total draws. The topology measurement was done using relative-betweenness centrality which measures the number of shortest paths going

through the node focussing on global network topology **(Aittokallio and Schwikowski, 2006**). Spearman's correlation was done to analyse the relationship between significant metabolites and cytokines Linear regression and Spearman's correlations were done using "lm. beta" and "Hmisc" package respectively in R. p-value < 0.05 and multiple testing correction using Benjamini-Hochberg (FDR) correction < 0.1. were chosen as the significance thresholds.

4.4 RESULTS

4.4.1 ANTHROPOMETRIC AND CLINICAL CHARACTERISTICS

Clinical and anthropometric details for the study are shown in **Table 4.1**. Overall, the Webex study (n =75) with 53 women (71%) and 22 males (29%) had fasting serum metabolite profiling for all participants except one. The mean age of the participants was 68 years (range 47-84 years). Pain measures on an NRS scale of 0-10 ranged from 0-8 with an average of 3.27.

The KPIC study consisted of 79 participants with 34 men (43%) and 45 women (57%). The average age of the participants (63 years) was matched to a criterion of ± 5 years accounting to be in the range of 41 years to 80 years. The pain score in the replication study measured was 2.67 on average.

	Webex dataset (N= 75)	KPIC dataset (N= 79)	p-value
	Mean(SD)	**Mean(SD)**	
Gender	71% F	57% F	0.14
	29% M	43% M	
Age^ (years)	68.21 (9.3)	62.75 (8.80)	0.0003***
BMI (kg/m^2)	31.13 (7.08)	29.32 (6.40)	0.128
Pain (scale 0-10) $^+$	3.27 (2.48)	2.67 (2.80)	0.208
Superolateral patella (kPa) $^+$	296.05 (144.14)	425.43 (192.05)	5.00E-06***
Superomedial patella (kPa) $^+$	321.62 (134.71)	404.88 (176.74)	0.004**
Medial joint line (kPa)	329.87 (172.40)	389.70 (191.71)	0.047**
Tibialis Anterior Muscle (kPa) $^+$	351.21 (157.13)	428.22 (191.28)	0.004**

Table 4.1: Anthropometric and Clinical Characteristics of the participants in study datasets. Results are presented as mean (sd) except for gender which is presented as percentages. ^Age-matched, criteria for age ± 5 years. $^+$Independent Student t-test was done for statistical significance inference. Significant values are marked with asterisk (*** p-value ≤ 0.001, ** p-value ≤0.05, * p-value ≤ 0.1).

4.4.2 METABOLOME AND PAIN

Higher levels of acyl ornithine (Ac.Orn), carnosine, dihydroxyphenylalanine (DOPA), histamine, hypoxanthine nitro-tyrosine (Nitro.Tyr), phenylethylamine (PEA), spermidine, spermine, taurine, taurolithocholic acid (TLCA), and lower levels of alpha-amino adipic acid (alpha-AAA), anserine, cysteine, cystine, cortisol, cortisone, dehydroepiandrosterone sulphate (DHEAS), glycolithocholic acid sulphate (GLCAS), symmetric dimethylarginine (SDMA), 3-indoleacetic acid (3-IAA) were found to be significantly (all FDR < 0.1) associated with higher current knee pain scores in the test (Webex) dataset. In replicate (KPIC) dataset, higher pain scores were significantly (FDR < 0.1) associated with higher levels of succinic acid, phenyl ethyl amine and lower levels of, asparagine, beta alanine, spermine, tryptophan and 3-indolepropionic acid (**Figure 4.2, Table 4.2**). Descriptive summary is presented in **Table 4.3**. PEA was found to be significantly replicated in both study datasets. Overall, a total of twenty-eight and seventeen metabolites showed significant associations

(p <0.05) with current knee pain score from regression model 1 in test (Webex) and replicate (KPIC) dataset respectively (**Figure 4.3**). These results indicate that although most of the metabolites in the replicate study dataset (KPIC) were not significant after multiple correction but showed a similar direction of associations with pain as in the test study dataset (Webex).

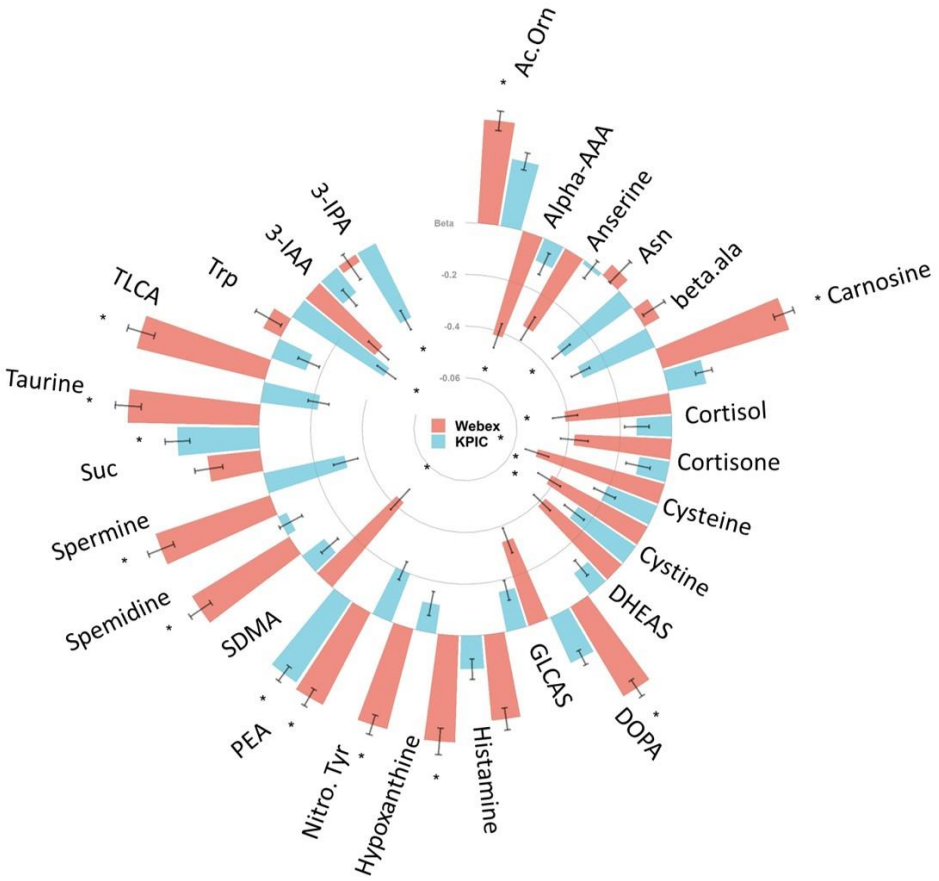


Figure 4.2: Metabolites associated with Pain scores in the Webex and KPIC study datasets. Circos plot depicting the metabolites significantly (FDR < 0.1) associated with current knee pain in either or both the datasets (Webex as red bars and KPIC as blue bars). The bars facing outward depicts the positive associations and bars facing inward depicts the negative associations ranging within +0.6 to -0.6. The data were inverse normalized and then regressed using a linear model 1. FDR values <0.05 are marked with an asterisk.

Abbreviations: Ac-Orn: acyl ornithine, alpha-AAA: alpha-amino adipic acid, Asn: Asparagine, beta-ala: beta-alanine, DHEAS: dehydroepiandrosterone sulphate, DOPA: dihydroxyphenylalanine, GLCAS: glycolithocholic acid sulphate, Nitro.Tyr.: nitro-tyrosine, PEA: phenylethylamine, SDMA: symmetric dimethylarginine, Suc: Succinic Acid, TLCA: taurolithocholic acid, Trp: Tryptophan, 3-IAA: 3-indoleacetic acid, 3-IPA: 3-Indole Propionic Acid

	Webex-Pain				**KPIC-Pain**			
traits	beta	pvalue	std.error	FDR	beta	pvalue	std.error	FDR
Asn	0.054	0.69	0.057	0.559	-0.323	0.005**	0.043	0.039**
Cys	-0.508	0.00008***	0.048	0.012**	-0.207	0.077	0.045	0.296
Trp	0.076	0.571	0.057	0.501	-0.423	0.0001***	0.044	0.003**
Ac.Orn +	0.401	0.002**	0.038	0.034**	0.263	0.024**	0.034	0.144
alpha.AAA	-0.417	0.001**	0.05	0.032**	-0.092	0.439	0.045	0.753
Anserine	-0.341	0.01**	0.052	0.09*	-0.014	0.906	0.042	0.856
Carnosine	0.512	0.00006***	0.04	0.014**	0.15	0.211	0.033	0.546
Cystine +	-0.422	0.001**	0.051	0.033**	-0.27	0.021**	0.048	0.128
DOPA	0.402	0.002**	0.037	0.033**	0.193	0.098	0.031	0.355
Nitro.Tyr	0.4	0.002**	0.039	0.035**	-0.201	0.095	0.038	0.346
SDMA	-0.429	0.001***	0.054	0.03**	-0.108	0.352	0.042	0.692
Taurine	0.509	0.00008***	0.05	0.009**	-0.222	0.062	0.041	0.261
GLCAS	-0.34	0.011**	0.051	0.089*	-0.156	0.198	0.039	0.527
TLCA	0.511	0.0001***	0.054	0.011**	-0.14	0.249	0.044	0.594
beta.Ala	0.067	0.584	0.049	0.506	-0.302	0.01**	0.039	0.07*
Histamine	0.333	0.009**	0.043	0.087*	0.129	0.292	0.039	0.65
PEA ^	0.402	0.002**	0.037	0.036**	0.379	0.001***	0.035	0.011**
Spermidine	0.441	0.001***	0.048	0.022**	-0.035	0.771	0.049	0.823
Spermine	0.467	0.0003***	0.053	0.016**	-0.32	0.006**	0.049	0.05*
Suc	0.205	0.12	0.055	0.274	0.316	0.006**	0.048	0.05**
Cortisol	-0.411	0.002**	0.049	0.033**	-0.136	0.238	0.047	0.573
Cortisone	-0.376	0.005**	0.053	0.054*	-0.117	0.34	0.048	0.693
DHEAS	-0.389	0.002**	0.045	0.033**	-0.101	0.35	0.038	0.691
3.IAA	-0.346	0.009**	0.052	0.087*	-0.121	0.268	0.041	0.61
3.IPA	-0.033	0.812	0.059	0.606	-0.318	0.005**	0.042	0.042**
Hypoxanthine	0.413	0.002**	0.051	0.035**	-0.112	0.36	0.048	0.7

Table 4.2: Metabolites associated with Pain in Webex(test) and KPIC (replicate) study datasets. The table shows beta coefficients, p-value, standard error and FDR values from the regression model 1for the significant metabolites associated with pain. Significant p-values are marked with an asterisk (*** p-value ≤ 0.001, **p-value ≤ 0.05, *p-value ≤ 0.1). ^ indicates metabolites replicated in two studies with FDR < 0.1 and $^{+}$ indicated metabolites replicated in two studies with p-value < 0.05.

traits	Webex	KPIC	P-value	FDR
Asn	50.239(8.93)	50.529(14.74)	0.882	0.552
Cys	70.438(37.95)	89.752(31.71)	0.001***	0.005**
Trp	52.677(9.99)	56.851(15.19)	0.045**	0.113
Ac.Orn	0.003(0.01)	0.001(0)	0.275	0.310
alpha.AAA	0.906(0.6)	1.274(0.58)	<0.001***	0.001***
Anserine	0.018(0.02)	0.011(0.01)	0.005**	0.022**
Carnosine	0.028(0.04)	0.081(0.06)	<0.001***	<0.001***
Cystine	16.092(8.74)	44.851(23)	<0.001***	<0.001***
DOPA	0.009(0.04)	0.053(0.04)	<0.001***	<0.001***
Nitro.Tyr	0.028(0.1)	0.171(0.07)	<0.001***	<0.001***
SDMA	0.55(0.2)	0.609(0.18)	0.055	0.128
Taurine	148.564(32.01)	157.153(45.92)	0.179	0.248
GLCAS	0.178(0.18)	0.273(0.2)	0.002**	0.011**
TLCA	0.092(0.22)	0.013(0.01)	0.003**	0.013**
beta.Ala	3.912(1.35)	3.394(1.98)	0.060	0.136
Histamine	0.036(0.13)	0.07(0.06)	0.036**	0.098
PEA	0.002(0.01)	0.008(0.02)	0.004**	0.017**
Spermidine	0.232(0.08)	0.245(0.1)	0.401	0.384
Spermine	0.184(0.07)	0.302(0.05)	<0.001***	<0.001***
Suc	7.018(0.82)	7.044(2.24)	0.921	0.565
Cortisol	0.266(0.16)	0.307(0.13)	0.080	0.159
Cortisone	0.06(0.03)	0.07(0.02)	0.018**	0.057
DHEAS	1.598(1.44)	2.148(1.21)	0.012**	0.040**
3.IAA	2.252(2.58)	2.056(1.12)	0.549	0.435
3.IPA	0.987(0.71)	0.847(0.58)	0.186	0.251
Hypoxanthine	14.203(3.32)	11.562(4.64)	<0.001***	0.001

Table 4.3: Descriptive of significant serum metabolites (FDR < 0.1) from linear model 1. Mean of absolute concentrations (uM) of metabolites significantly associated with pain scores are presented and values in parenthesis indicate the standard deviations(SD). Independent Student t-test was done for statistical significance inference. Significant values are marked with asterisk (*** p-value ≤ 0.001, ** p-value ≤ 0.05, * p-value ≤ 0.1).

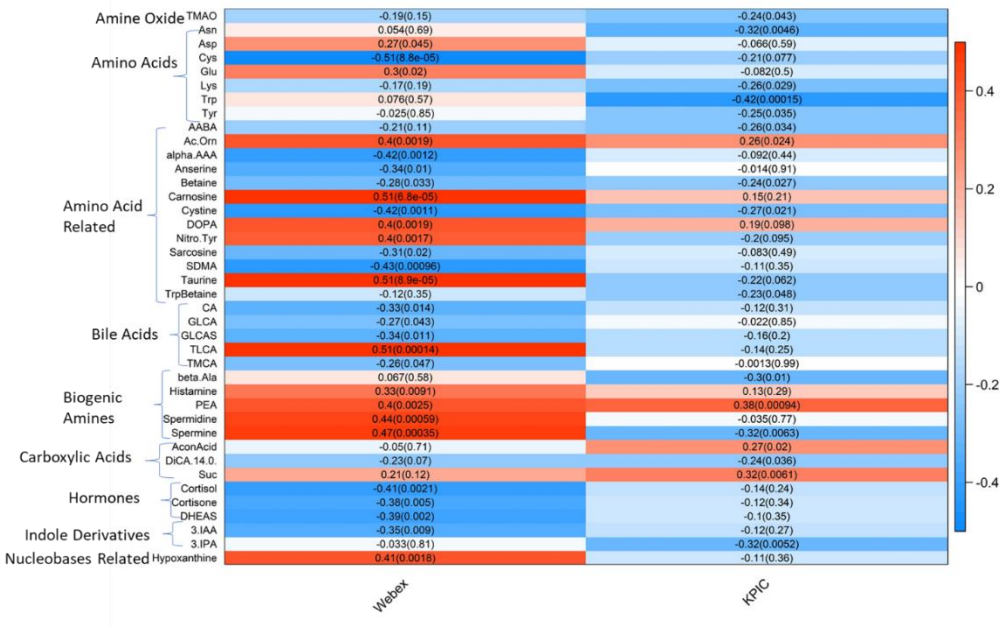


Figure 4.3: Metabolites associated with Pain in test (Webex) and replicate (KPIC) cohort. Heatmap depicting the significant metabolites ($p \leq 0.05$) associated with current knee pain scores in two study datasets. The data was inverse normalized and then regressed using a linear model 1. Blue to red highlighted values indicates the beta coefficients from negative to positive associations. p-values in parenthesis indicate the statistical significance of the associations.

4.4.3 METABOLOME AND PRESSURE PAIN THRESHOLDS (PPTs)

In the test (Webex) dataset, altogether seventeen metabolites and in the replicate (KPIC) dataset altogether twelve metabolites were found to be significantly ($p < 0.05$) associated with PPTs measured at four different regions of the body. However, none of the metabolites were found to be significant (FDR <0.1) after multiple testing correction. Nonetheless, the most common metabolite found to be associated with PPTs in the Webex study was SDMA, while cortisone and carnosine were commonly associated with PPTs in the KPIC study **Table 4.4)**.

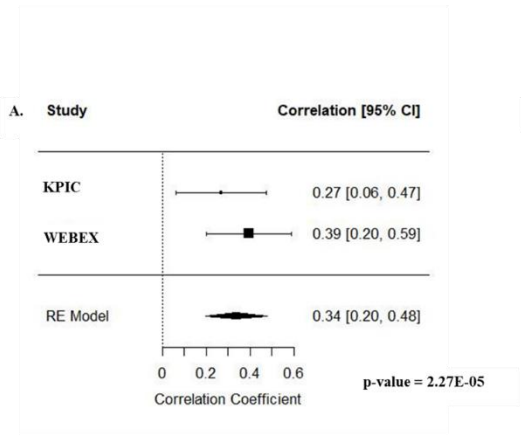
Webex	Superolateral				Superomedial				Tibialis Anterior				Medial Joint Line			
traits	beta	p-value	std. error	FDR	beta	p-value	std. error	FDR	Beta	p-value	std. error	FDR	beta	p-value	std. error	FDR
His	0.115	0.384	0.001	1.439	0.129	0.305	0.001	1.126	0.272	0.041**	0.001	0.962	0.136	0.341	0.001	0.959
Lys	0.062	0.636	0.001	1.227	0.239	0.050**	0.001	1.750	0.271	0.038**	0.001	0.965	0.219	0.114	0.001	1.214
Pro	0.183	0.149	0.001	2.092	0.180	0.135	0.001	1.435	0.169	0.192	0.001	1.004	0.010	0.944	0.001	0.939
Thr	-0.170	0.198	0.001	1.863	0.005	0.970	0.001	1.071	0.067	0.618	0.001	1.023	-0.082	0.566	0.001	0.917
Tyr	0.103	0.422	0.001	1.374	0.291	0.015**	0.001	1.950	0.258	0.045**	0.001	0.946	-0.039	0.779	0.001	0.910
3.Met.His	-0.055	0.666	0.001	1.228	0.015	0.899	0.001	1.063	-0.085	0.513	0.001	1.020	0.043	0.756	0.001	0.907
5.AVA	0.070	0.592	0.001	1.276	0.101	0.416	0.001	1.103	0.197	0.136	0.001	1.026	0.132	0.347	0.001	0.959
alpha.AAA	0.275	0.032**	0.001	2.089	0.335	0.006**	0.001	1.231	0.294	0.024**	0.001	1.716	0.159	0.255	0.001	0.983
Anserine	0.211	0.107	0.001	2.323	0.234	0.060	0.001	1.612	0.276	0.038**	0.001	0.993	0.203	0.152	0.001	0.979
Carnosine	-0.158	0.229	0.001	1.782	-0.105	0.403	0.001	1.125	-0.224	0.094	0.001	1.079	-0.100	0.481	0.001	0.908
Creatinine	0.250	0.024**	0.001	3.496	0.190	0.074	0.001	1.538	0.155	0.175	0.001	0.991	0.124	0.308	0.001	0.967
DOPA	-0.218	0.092	0.001	2.430	-0.179	0.145	0.001	1.406	-0.249	0.058	0.001	0.995	-0.112	0.426	0.001	0.916
HArg	0.254	0.030*	0.001	2.671	0.229	0.040**	0.001	2.033	0.235	0.049**	0.001	0.915	0.195	0.124	0.001	1.190
PAG	0.174	0.183	0.001	2.021	0.249	0.044**	0.001	2.098	0.163	0.222	0.001	1.007	0.081	0.569	0.001	0.915
ProBetaine	-0.003	0.981	0.001	1.144	0.081	0.508	0.001	1.134	0.310	0.016**	0.001	2.484	0.091	0.509	0.001	0.899
SDMA	0.386	0.002**	0.001	0.848	0.361	0.003**	0.001	1.929	0.219	0.098	0.001	1.074	0.372	0.007**	0.001	2.089
t4.OH.Pro	-0.153	0.221	0.001	1.885	0.052	0.664	0.001	1.141	-0.005	0.969	0.001	1.006	-0.068	0.617	0.001	0.885
Taurine	-0.129	0.334	0.001	1.542	-0.281	0.024*	0.001	2.333	-0.209	0.121	0.001	1.067	-0.105	0.464	0.001	0.912
CDCA	0.106	0.427	0.001	1.364	0.041	0.748	0.001	1.120	0.110	0.419	0.001	0.996	-0.045	0.753	0.001	0.911
GLCAS	0.339	0.009**	0.001	1.550	0.190	0.128	0.001	1.494	0.047	0.727	0.001	1.003	0.258	0.067	0.001	1.480
TCA	0.246	0.058	0.001	2.582	0.172	0.166	0.001	1.303	0.279	0.034**	0.001	1.083	-0.025	0.861	0.001	0.923
PEA	-0.203	0.124	0.001	1.896	-0.177	0.161	0.001	1.314	-0.290	0.030**	0.001	1.120	-0.117	0.414	0.001	0.918
Spermidine	-0.163	0.209	0.001	1.872	-0.186	0.130	0.001	1.496	-0.098	0.460	0.001	1.023	-0.096	0.492	0.001	0.908
Spermine	-0.207	0.116	0.001	2.034	-0.190	0.131	0.001	1.484	-0.265	0.048**	0.001	0.942	-0.136	0.341	0.001	0.955
AconAcid	0.255	0.044**	0.001	2.647	0.209	0.085	0.001	1.583	0.148	0.258	0.001	1.035	0.287	0.036**	0.001	2.127
DiCA.12.0.	0.193	0.124	0.001	1.935	0.050	0.678	0.001	1.143	0.102	0.425	0.001	0.993	0.063	0.642	0.001	0.883
Cortisone	0.198	0.136	0.001	1.952	0.175	0.167	0.001	1.297	0.200	0.140	0.001	1.022	0.151	0.295	0.001	0.954
3.IAA	0.098	0.457	0.001	1.336	0.126	0.314	0.001	1.108	0.265	0.046**	0.001	0.932	-0.013	0.928	0.001	0.943
KPIC	Superolateral				Superomedial				Tibialis Anterior				Medial Joint Line			
traits	beta	p-value	std. error	FDR	beta	p-value	std. error	FDR	beta	p-value	std. error	FDR	beta	p-value	std. error	FDR
His	-0.140	0.250	0.001	1.199	-0.183	0.160	0.001	1.052	0.010	0.939	0.001	1.320	-0.022	0.866	0.001	0.963

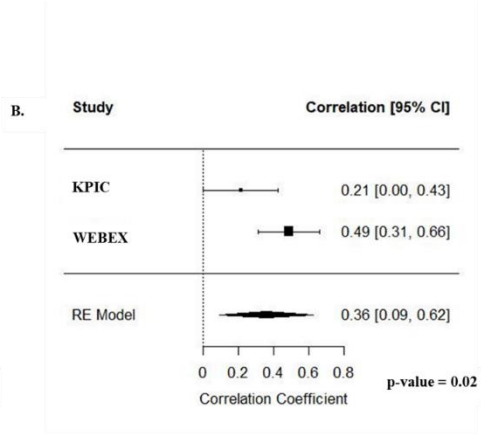
Lys	-0.017	0.889	0.001	1.058	-0.071	0.591	0.001	0.814	0.085	0.533	0.001	1.651	0.056	0.669	0.001	0.943
Pro	-0.237	0.049**	0.001	1.562	-0.278	0.032**	0.001	1.293	-0.189	0.164	0.001	1.708	-0.229	0.078	0.001	1.348
Thr	-0.259	0.036**	0.001	1.493	-0.212	0.112	0.001	1.178	-0.167	0.229	0.001	1.830	-0.159	0.235	0.001	1.119
Tyr	-0.031	0.799	0.001	1.069	0.011	0.935	0.001	0.896	0.140	0.301	0.001	1.934	0.037	0.774	0.001	0.931
3.Met.His	-0.092	0.459	0.001	1.092	-0.251	0.058	0.001	1.495	-0.298	0.029**	0.001	1.546	-0.262	0.047**	0.001	1.771
5.AVA	-0.114	0.360	0.001	1.205	-0.149	0.265	0.001	0.910	-0.373	0.006**	0.001	2.333	-0.251	0.058	0.001	1.587
alpha.AAA	-0.105	0.379	0.001	1.190	-0.181	0.157	0.001	1.047	-0.057	0.669	0.001	1.545	-0.161	0.207	0.001	1.220
Anserine	-0.085	0.488	0.001	1.117	-0.037	0.782	0.001	0.871	0.185	0.174	0.001	1.743	0.033	0.804	0.001	0.940
Carnosine	-0.304	0.011**	0.000	1.006	-0.379	0.003**	0.001	0.798	-0.132	0.328	0.001	1.836	-0.287	0.025**	0.000	1.539
Creatinine	0.021	0.851	0.001	1.065	0.075	0.535	0.001	0.811	0.062	0.620	0.001	1.584	0.094	0.435	0.001	0.916
DOPA	-0.242	0.039**	0.000	1.538	-0.215	0.090	0.001	1.248	-0.057	0.670	0.001	1.534	-0.147	0.248	0.000	1.115
HArg	-0.044	0.721	0.001	1.070	-0.106	0.420	0.001	0.818	-0.151	0.269	0.001	1.851	-0.163	0.213	0.001	1.207
PAG	0.130	0.296	0.001	1.188	0.269	0.043**	0.001	1.337	0.400	0.003**	0.001	2.481	0.201	0.132	0.001	1.478
ProBetaine	-0.212	0.073	0.001	1.397	-0.205	0.108	0.001	1.219	-0.144	0.280	0.001	1.874	-0.190	0.136	0.001	1.377
SDMA	0.146	0.212	0.001	1.274	0.194	0.121	0.001	1.101	0.098	0.452	0.001	1.811	0.133	0.289	0.001	1.067
t4.OH.Pro	-0.226	0.057	0.001	1.580	-0.292	0.021**	0.001	1.349	-0.261	0.049**	0.001	1.829	-0.270	0.033	0.001	1.443
Taurine	0.072	0.554	0.001	1.083	0.123	0.345	0.001	0.814	0.097	0.475	0.001	1.780	0.142	0.275	0.001	1.070
CDCA	0.205	0.090	0.001	1.330	0.199	0.127	0.001	1.076	0.310	0.020**	0.001	1.462	0.242	0.061	0.001	1.491
GLCAS	-0.013	0.913	0.001	1.060	-0.086	0.517	0.001	0.809	0.014	0.916	0.001	1.348	-0.028	0.832	0.001	0.944
TCA	-0.106	0.395	0.001	1.149	-0.119	0.374	0.001	0.808	0.074	0.595	0.001	1.616	-0.202	0.128	0.001	1.462
PEA	0.022	0.855	0.001	1.066	-0.072	0.577	0.001	0.820	0.096	0.475	0.001	1.786	0.002	0.989	0.001	0.967
Spermidine	0.223	0.064	0.001	1.526	0.221	0.089	0.001	1.291	0.184	0.174	0.001	1.723	0.303	0.018**	0.001	1.579
Spermine	-0.001	0.996	0.001	1.056	0.159	0.219	0.001	1.001	0.053	0.698	0.001	1.506	0.113	0.384	0.001	0.924
AconAcid	-0.043	0.721	0.001	1.072	-0.081	0.534	0.001	0.813	0.006	0.962	0.001	1.297	-0.048	0.714	0.001	0.937
DiCA.12.0.	0.219	0.071	0.001	1.485	0.201	0.124	0.001	1.082	0.278	0.040**	0.001	1.837	0.236	0.070	0.001	1.368
Cortisone	-0.354	0.003**	0.001	1.099	-0.400	0.002**	0.001	1.132	-0.342	0.012**	0.001	1.544	-0.381	0.003**	0.001	1.972
3.IAA	0.144	0.192	0.001	1.360	0.214	0.069	0.001	1.451	0.152	0.217	0.001	1.823	0.193	0.101	0.001	1.397

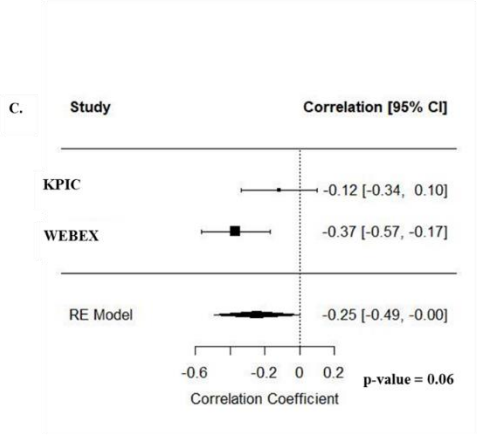
Table 4.4: Metabolites associated with Pressure Pain detection Thresholds (PPTs) in Webex and KPIC study datasets. The table shows beta coefficients, p-value from the regression model 1 (adjusted for age, gender and BMI) for the significant metabolites associated with PPTs. Significant values are marked with an asterisk (*** p-value ≤ 0.001, **p-value ≤ 0.05, *p-value ≤ 0.1)

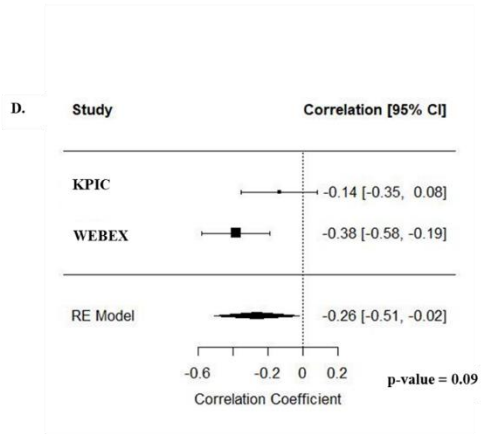
4.4.4 AMINO-ACID METABOLISM AND KNEE PAIN

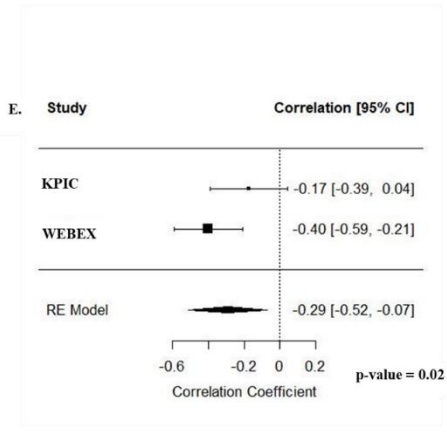
Meta-analysis of standardized linear regression coefficients adjusted for age, gender, BMI, NSAIDs and opioids (model 2) of the two cohorts (Webex and KPIC) showed acyl ornithine, carnosine, DOPA, PEA and succinic acid to increase significantly with an increase in pain while cystine, GLCAS, cortisol and cortisone were significantly decreased with increase in pain. Results are shown in **Figure 4.4 and Table 4.5**. Pathway analysis showed phenylalanine metabolism to be most significantly involved ($p<0.05$) in current knee pain. Other pathways involved were steroid hormone biosynthesis, arginine biosynthesis, butanoate metabolism, histidine metabolism, citrate cycle, beta-alanine metabolism, propanoate metabolism, alanine aspartate and glutamate metabolism, glutathione metabolism, cysteine and methionine metabolism, arginine and proline metabolism, tyrosine metabolism (**Figure 4.5, Table 4.6**).

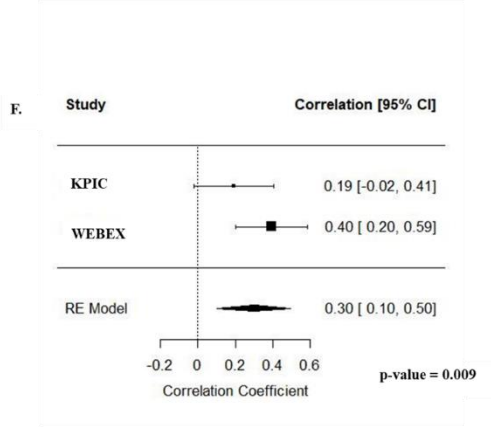

A.
Study
Correlation [95% CI]
KPIC
0.27 [0.06, 0.47]
WEBEX
0.39 [0.20, 0.59]
RE Model
0.34 [0.20, 0.48]
0 0.2 0.4 0.6
Correlation Coefficient
p-value = 2.27E-05

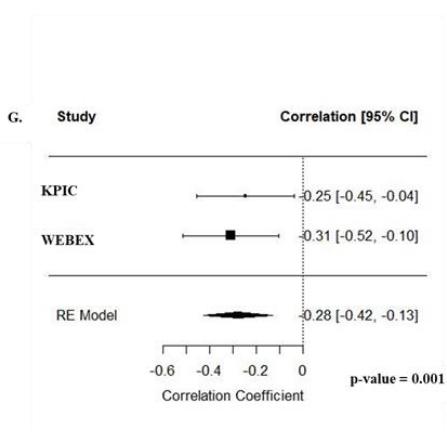

B.
Study
Correlation [95% CI]
KPIC
0.21 [0.00, 0.43]
WEBEX
0.49 [0.31, 0.66]
RE Model
0.36 [0.09, 0.62]
0 0.2 0.4 0.6 0.8
Correlation Coefficient
p-value = 0.02

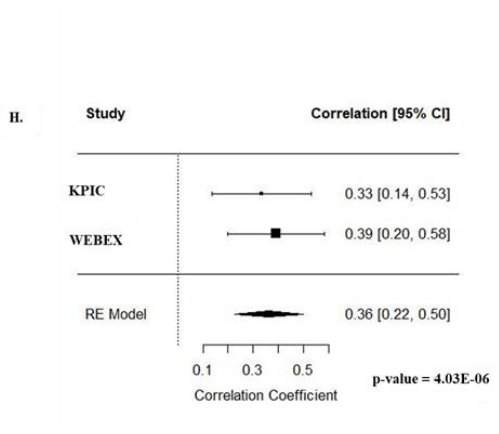

C.
Study
Correlation [95% CI]
KPIC
-0.12 [-0.34, 0.10]
WEBEX
-0.37 [-0.57, -0.17]
RE Model
-0.25 [-0.49, -0.00]
-0.6 -0.2 0 0.2
Correlation Coefficient
p-value = 0.06


D.
Study
Correlation [95% CI]
KPIC
-0.14 [-0.35, 0.08]
WEBEX
-0.38 [-0.58, -0.19]
RE Model
-0.26 [-0.51, -0.02]
-0.6 -0.2 0 0.2
Correlation Coefficient
p-value = 0.09


E.
Study
Correlation [95% CI]
KPIC
-0.17 [-0.39, 0.04]
WEBEX
-0.40 [-0.59, -0.21]
RE Model
-0.29 [-0.52, -0.07]
-0.6 -0.2 0 0.2
Correlation Coefficient
p-value = 0.02


F.
Study
Correlation [95% CI]
KPIC
0.19 [-0.02, 0.41]
WEBEX
0.40 [0.20, 0.59]
RE Model
0.30 [0.10, 0.50]
-0.2 0 0.2 0.4 0.6
Correlation Coefficient
p-value = 0.009


G.
Study
Correlation [95% CI]
KPIC
-0.25 [-0.45, -0.04]
WEBEX
-0.31 [-0.52, -0.10]
RE Model
-0.28 [-0.42, -0.13]
-0.6 -0.4 -0.2 0
Correlation Coefficient
p-value = 0.001


H.
Study
Correlation [95% CI]
KPIC
0.33 [0.14, 0.53]
WEBEX
0.39 [0.20, 0.58]
RE Model
0.36 [0.22, 0.50]
0.1 0.3 0.5
Correlation Coefficient
p-value = 4.03E-06

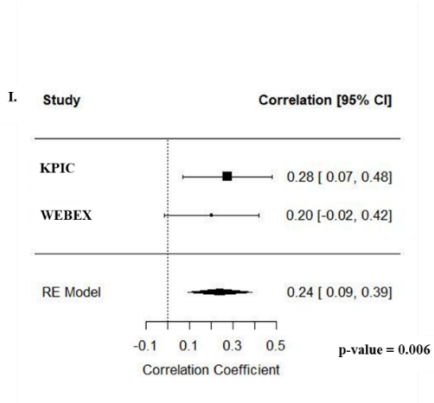


Figure 4.4: Meta-analysis of test (Webex) and replicate (KPIC) datasets. The forest plots A. Acyl Ornithine, B. Carnosine, C. Cortisone, D. Cortisol, E. Cystine, F. DOPA, G. GLCAS, H. PEA, I. Succinic Acid depict the metabolites significantly associated with pain in replicate (KPIC) and test (Webex) datasets. Values from linear regression model 2 were inputted for the fixed random effect model-based meta-analysis. p-values indicate the statistical significance with FDR < 0.1.

Metabolite	beta	se	p-value	ci.ll	ci.ul	FDR
Ac.Orn	0.336	0.072	< 0.001 ***	0.195	0.476	<0.001***
Carnosine	0.357	0.136	0.009**	0.091	0.623	0.023**
Cortisol	-0.265	0.124	0.032**	-0.508	-0.022	0.065*
Cortisone	-0.247	0.126	0.049**	-0.494	-0.001	0.087*
Cystine	-0.293	0.113	0.01**	-0.516	-0.071	0.022**
DOPA	0.3	0.101	0.003**	0.102	0.497	0.01**
GLCAS	-0.278	0.075	< 0.001***	-0.424	-0.131	0.001***
PEA	0.363	0.07	< 0.001***	0.225	0.501	<0.001***
Suc	0.241	0.076	0.002**	0.092	0.391	0.006**

Table 4.5: Meta-analysis of metabolites significantly associated with current knee pain in the KPIC and Webex cohorts. Meta-analysis results of the significant metabolites tested in Webex and KPIC study datasets. Beta coefficients for the regression model 2were inputted for the random-fixed effect meta-analysis model. The table shows beta coefficients, standard error, p-values, confidence intervals (upper and lower limit) and FDR corrected p-values. Significant values are marked with an asterisk (*** p-value ≤ 0.001, **p-value ≤ 0.05, *p-value ≤ 0.1).

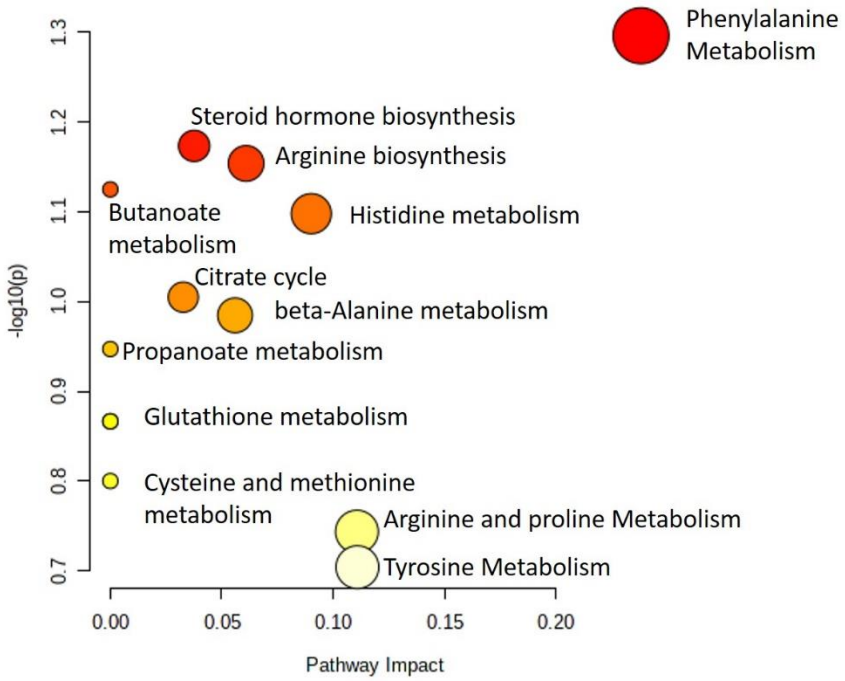


Figure 4.5: Relevant pathways involved in current knee pain based on the significantly associated metabolites from the meta-analysis. Scatter plot of the pathways enriched due to the significant metabolites. The x-axis and size of the node depict the impact of pathways from topology analysis and the y-axis and colour of the node depict the statistical significance in terms of the $\log_{10}$(p-values).

Pathways	Total	Expected	Hits	Raw p	FDR	Impact
Phenylalanine metabolism	10	0.052	1	0.051**	1	0.238
Steroid hormone biosynthesis	85	0.439	2	0.067*	1	0.038
Arginine biosynthesis	14	0.072	1	0.07*	1	0.061
Butanoate metabolism	15	0.077	1	0.075*	1	<0.001
Histidine metabolism	16	0.083	1	0.08*	1	0.090
Citrate cycle (TCA cycle)	20	0.103	1	0.099*	1	0.033
beta-Alanine metabolism	21	0.108	1	0.104	1	0.056
Propanoate metabolism	23	0.119	1	0.113	1	<0.001
Alanine, aspartate and glutamate metabolism	28	0.145	1	0.136	1	<0.001
Glutathione metabolism	28	0.145	1	0.136	1	<0.001
Cysteine and methionine metabolism	33	0.17	1	0.158	1	<0.001
Arginine and proline metabolism	38	0.196	1	0.18	1	0.111
Tyrosine metabolism	42	0.217	1	0.1977	1	0.111

Table 4.6: Pathway Analysis of significant metabolites associated with knee pain in meta-analysis. The table shows perturbed pathways with the nine significant metabolites from the meta-analysis of Webex (test) and KPIC (replicate) study. The values depict the total metabolites in the pathway, expected p-values after enrichment (i.e. over-representation), the number of metabolites from the study involved in the pathway, raw p-values, impact of the pathway from topology analysis. Significant values are marked with an asterisk (*** p-value ≤ 0.001, **p-value ≤ 0.05, *p-value ≤ 0.1).

4.4.5 PAIN AND PRO OR ANTI-INFLAMMATORY CYTOKINES IN THE TEST COHORT

Certain pro-inflammatory cytokines such as IL-1β, IL-2, IL-6, TNFα were found to be negatively associated with pain scores while certain anti-inflammatory cytokines such as IL-4, IL-8, IL-13 were found to be positively associated. However, no cytokine was found to be significantly (FDR < 0.1) associated with pain scores or PPTs in the test (Webex) study after adjusting for age, gender, BMI, NSAIDs and opioids (**Table 4.7**).

	Pain			Superolateral			Superomedial			Tibialis Anterior			Medial JointLine		
Cytokine	beta	std.error	p-value	beta	std.error	p-value	beta	std.error	p-value	beta	std.error	p-value	beta	std.error	p-value
IFN.gamma	0.074	0.845	0.576	-0.072	0.014	0.579	-0.097	0.015	0.448	-0.238	0.013	0.066 *	-0.167	0.013	0.238
IL10	-0.087	0.071	0.525	0.097	0.001	0.466	0.141	0.001	0.286	0.038	0.001	0.781	0.031	0.001	0.831
IL.12P70	0.011	0.093	0.936	-0.023	0.002	0.862	0.093	0.002	0.471	-0.010	0.001	0.937	0.013	0.001	0.929
IL.13	0.086	0.243	0.531	-0.174	0.004	0.192	-0.111	0.004	0.404	-0.097	0.004	0.473	-0.180	0.004	0.220
IL.1beta	-0.072	0.022	0.597	-0.055	0.000	0.679	0.030	0.000	0.821	-0.254	0.000	0.058 *	-0.043	0.000	0.768
IL2	-0.027	0.063	0.843	0.011	0.001	0.933	0.128	0.001	0.327	-0.044	0.001	0.744	0.051	0.001	0.728
IL4	0.155	0.034	0.254	-0.080	0.001	0.547	0.148	0.001	0.261	0.156	0.001	0.244	-0.020	0.001	0.889
IL6	-0.049	0.085	0.703	-0.086	0.001	0.491	0.055	0.001	0.657	-0.232	0.001	0.061 *	-0.108	0.001	0.431
IL8	0.116	3.259	0.399	-0.009	0.054	0.948	-0.093	0.057	0.489	-0.193	0.051	0.156	-0.071	0.049	0.631
TNF.alpha	-0.081	0.309	0.553	-0.058	0.005	0.666	-0.097	0.005	0.465	-0.098	0.005	0.469	-0.121	0.005	0.411

Table 4.7: Lack of association between circulating cytokine levels and Pain scores and Pressure Pain detection Thresholds (PPTs) in Webex study. The table shows beta coefficients, p-values, standard error from the regression model 2 for serum cytokines associated with pain and PPTs in the test study (Webex). * indicates metabolites significantly associated with PPT scores with p-values ≤0.1.

The association of significant metabolites from the meta-analysis was tested with the cytokine concentrations (**Figure 4.6, Table 4.8**). IL-10 and TNFα were significantly correlated with at least six metabolites. IL- 13 was found to be significantly associated ($p < 0.05$) with acyl ornithine, DOPA, cortisol and PEA. IL-1β was found to be associated with PEA and carnosine. IL-2 is associated significantly with, cysteine. IL-8 was significantly associated with PEA, cortisol and carnosine. IFN-γ, IL-4, IL-6 and IL-12p70 was not found to be having significant associations with any of the significant metabolites.

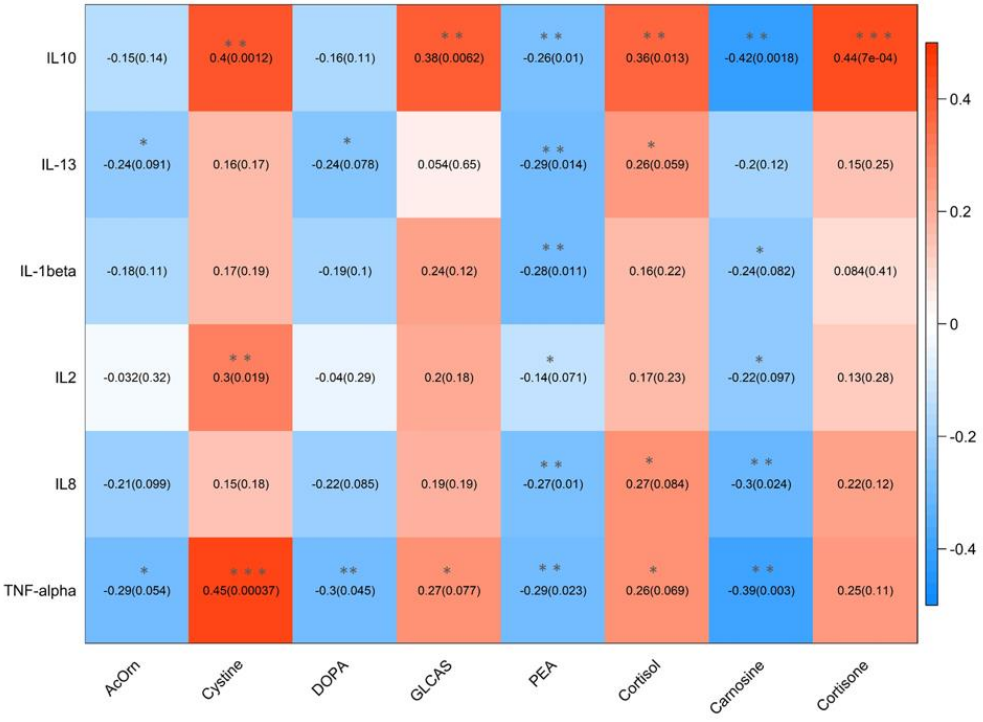


Figure 4.6: Correlation between replicated metabolites and circulating cytokine levels. Heatmap depicting the correlation of cytokines levels in Webex dataset with the replicated metabolites in the Webex and KPIC datasets from meta-analysis. Blue to red highlighted values indicates the spearman's coefficients from negative to positive associations. p-values in parenthesis indicate the statistical significance of the associations after multiple testing correction (FDR ≤0.1) and are marked with an asterisk.

	Ac.Orn			Cystine			DOPA		
	Coefficient	p-value	FDR	Coefficient	p-value	FDR	Coefficient	p-value	FDR
IFN-gamma	-0.116	0.332	0.166	0.177	0.136	0.205	-0.111	0.351	0.176
IL10	-0.148	0.213	0.142	0.403	< 0.001***	0.001***	-0.161	0.172	0.115
IL-12P70	-0.179	0.13	0.13	-0.015	0.901	0.541	-0.182	0.123	0.099
IL-13	-0.235	0.045**	0.091*	0.161	0.174	0.174	-0.242	0.039**	0.078*
IL-1beta	-0.177	0.133	0.107	0.166	0.161	0.194	-0.193	0.101	0.101
IL2	-0.032	0.788	0.315	0.302	0.009**	0.019**	-0.04	0.734	0.294
IL4	0.067	0.574	0.255	0.068	0.565	0.424	0.056	0.636	0.283
IL6	-0.117	0.323	0.185	0.055	0.646	0.431	-0.117	0.322	0.184
IL8	-0.21	0.074	0.099	0.15	0.204	0.175	-0.218	0.064	0.085
TNF-alpha	-0.288	0.013**	0.054*	0.451	< 0.001**	<0.001***	-0.295	0.011**	0.045**
	GLCAS			**PEA**			**Cortisol**		
	Coefficient	p-value	FDR	Coefficient	p-value	FDR	Coefficient	p-value	FDR
IFN-gamma	-0.03	0.802	0.641	-0.04	0.74	0.148	-0.052	0.663	0.589
IL10	0.385	0.001***	0.006**	-0.261	0.026*	0.01**	0.363	0.002**	0.013**
IL-12P70	-0.085	0.476	0.635	-0.206	0.08	0.027	-0.075	0.531	0.607
IL-13	0.054	0.649	0.649	-0.288	0.014*	0.014**	0.255	0.029*	0.059*
IL-1beta	0.236	0.045**	0.119	-0.281	0.016*	0.011**	0.163	0.168	0.224
IL2	0.2	0.089	0.178	-0.137	0.247	0.071	0.172	0.147	0.235
IL4	-0.043	0.719	0.639	-0.101	0.395	0.088	-0.067	0.572	0.572

IL6	-0.066	0.578	0.66	-0.123	0.3	0.075	-0.004	0.974	0.779
IL8	0.186	0.116	0.185	-0.271	0.02*	0.01**	0.27	0.021**	0.084*
TNF-alpha	0.274	0.019**	0.077*	-0.294	0.012*	0.023	0.26	0.026**	0.069*
	Suc			**Carnosine**			**Cortisone**		
	Coefficient	p-value	FDR	Coefficient	p-value	FDR	Coefficient	p-value	FDR
IFN-gamma	0.247	0.037**	0.293	-0.07	0.558	0.637	0.019	0.871	0.653
IL10	-0.011	0.93	0.744	-0.418	<0.001 ***	0.002**	0.436	<0.001***	0.001***
IL-12P70	0.145	0.22	0.352	0.05	0.674	0.599	-0.154	0.194	0.291
IL-13	0.012	0.917	0.815	-0.199	0.091	0.121	0.15	0.207	0.248
IL-1beta	-0.085	0.476	0.635	-0.24	0.041**	0.082*	0.084	0.478	0.41
IL2	-0.026	0.828	0.828	-0.221	0.06	0.097	0.128	0.279	0.279
IL4	0.207	0.079	0.211	0.067	0.575	0.575	0.009	0.937	0.625
IL6	0.231	0.049**	0.195	-0.006	0.963	0.77	0.003	0.982	0.589
IL8	0.184	0.12	0.24	-0.304	0.009***	0.024**	0.221	0.06	0.121
TNF-alpha	-0.048	0.686	0.784	-0.386	0.001**	0.003	0.246	0.036**	0.107

Table 4.8: Correlation of serum cytokines and replicated metabolites from meta-analysis of two studies in the Webex study. This table shows spearman's correlation coefficients and p-values for the significant metabolites from meta-analysis and cytokines for the Webex study. Significant values are marked with an asterisk (*** p-value ≤ 0.001, **p-value ≤ 0.05, *p-value ≤ 0.1).

4.4.6 GUT MICROBIOME FEATURE ANALYSIS

After quality control and demultiplexing, 6950 features were identified with a total frequency of 5,522,110. Taxonomic assignments using SILVA classifier based on SILVA database, this study identified overall 390 taxonomies. After rarefying and removal of taxa present in less than 0.01% in 95% of samples, 224 taxa remained (200 genus- level taxa, 14 family-level, 5 order-level taxa, 3 class-level, 1 phylum and 1domain level taxa) and included for further analysis.

4.4.7 GUT MICROBIOME AND PAIN IN WEBEX DATASET

High abundance of *Parasutterella* was most strongly and significantly (β = -0.387, p-val =0.003) associated with low levels of current knee pain in the Webex dataset. Other microbial taxa that were found significantly (p-value < 0.05) associated with pain were Bacillaceae (β = -0.270), *Bilophila* (β = -0.298), *Coprococcus1* (β = -0.321), *Lactobacillus* (β = 0.310), Lachnospiraceae UC512E3 (β = 0.287), *Streptococcus* (β = 0.280). However, none of the microbe stayed significant after multiple testing correction. Results shown in **Figure 4.7**.

4.4.7 GUT MICROBIOME AND PRESSURE PAIN THRESHOLDS (PPTs) IN WEBEX STUDY

In the Webex dataset, altogether forty-eight microbial taxa were found to be significantly ($p < 0.05$) associated with PPTs measured at four different regions of the body. Most commonly associated taxa with all four PPTs were uncultured Bacteroidales, Enterobacteriaceae and Ruminiclostridium 1 (**Figure 4.7**).

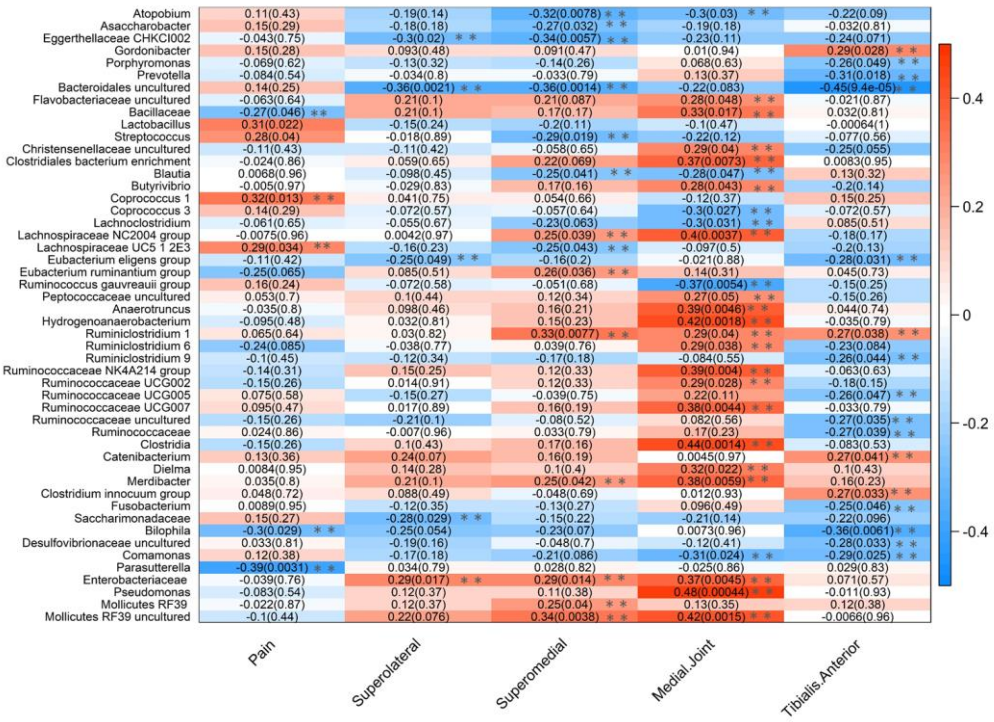


Figure 4.7: Regression of microbiome with pain and PPTs in the Webex dataset. Heatmap depicting the significantly (p-value < 0.05) associated microbial taxa with the pain and PPTs in the Webex dataset. Blue to red highlighted values indicates the regression coefficients from negative to positive associations. p-values in parenthesis indicate the statistical significance and are marked with an asterisk (*** p-value ≤ 0.001, **p-value ≤ 0.05, *p-value ≤ 0.1).

4.4.7 GUT MICROBIOME AND METABOLOME ASSOCIATIONS IN WEBEX STUDY

To explore the associations between gut microbiome composition and the metabolome, Spearman's correlations were performed. The metabolites found significantly associated with pain in the Webex study were tested with the microbes significantly associated with pain. Mainly, *Parasutrella* was found to be significantly associated with low levels of taurine while *Streptococcus* was found to be significantly associated with high levels of histamine, spermidine and hypoxanthine. The associations are plotted as a network graph in **Figure 4.8** and details are in **Table 4.9**.

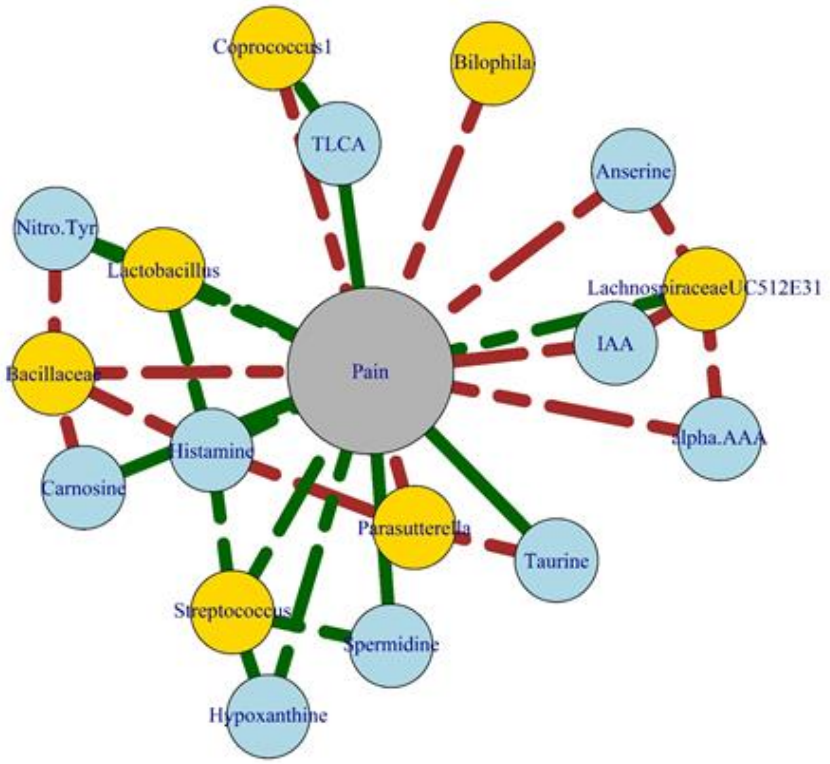


Figure 4.8: Associations between metabolites, microbes and Pain. Blue coloured circles represent the metabolites. Pain trait is represented as grey coloured circle and the yellow circles represent the microbes. Red colour lines indicate the negative correlations, while green colour lines indicate positive correlation. Dashed lines indicate the association with p-value ≤0.05 while solid lines indicate the associations with p-value ≤0.001.

Bacterial genus	Bacillaceae		Lactobacillus		Streptococcus		Coprococcus1		Lachnospiraceae UC512E31		Bilophila		Parasutterella	
metabolite	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
alpha.AAA	0.155	0.188	-0.115	0.331	-0.077	0.512	0.051	0.665	-0.258	0.026*	-0.055	0.640	0.043	0.719
Anserine	-0.132	0.261	-0.007	0.954	0.084	0.479	0.016	0.889	-0.259	0.026*	-0.003	0.983	0.048	0.686
Carnosine	-0.255	0.028*	0.114	0.332	0.068	0.564	0.096	0.414	0.134	0.254	0.021	0.859	-0.141	0.230
Nitro.Tyr	-0.250	0.032**	0.247	0.034*	0.159	0.175	0.211	0.071	0.124	0.294	0.021	0.857	-0.233	0.046*
Taurine	0.038	0.747	0.090	0.444	0.088	0.453	0.132	0.263	0.040	0.736	-0.124	0.293	-0.308	0.008*
TLCA	-0.184	0.121	0.170	0.152	0.097	0.417	0.331	0.005*	0.166	0.162	0.001	0.995	-0.107	0.370
Histamine	-0.307	0.008*	0.297	0.010*	0.277	0.017*	0.194	0.098	-0.043	0.716	-0.057	0.628	-0.231	0.047*
Spermidine	0.090	0.448	0.085	0.473	0.270	0.020*	-0.091	0.440	0.054	0.645	-0.098	0.408	-0.354	0.002*
IAA	0.074	0.530	0.063	0.592	-0.083	0.483	0.063	0.595	-0.273	0.018*	-0.163	0.165	0.029	0.808
Hypoxanthine	-0.023	0.849	0.037	0.755	0.306	0.008	0.117	0.320	-0.005	0.965	-0.029	0.808	-0.170	0.148

Table 4.9: Association of gut microbes and metabolites: This table shows spearman's correlation results depicting associations between the microbial taxa and metabolites found significantly associated with Pain. * = $p \leq 0.05$

4.5 DISCUSSION AND CONCLUSION

The current study reports reproducible associations of serum metabolites with current knee pain scores and PPTs in individuals with knee OA using the metabolic profiles of 93 metabolites in the two independent studies (KPIC and Webex). The metabolites tested belonged to the class of amino acid related, bile acids, biogenic amines, carboxylic acids, cresol, hormones, indole derivative, vitamins and cofactor, nucleobase, amine-oxide and amino acids. Consistent with previous findings, meta-analysis of two independent studies demonstrates the association of acyl ornithine, carnosine, cortisol, cortisone, cystine, DOPA, GLCAS, PEA and succinic acid in knee pain in OA patients. Acyl ornithine, a downstream metabolite of arginine catabolism, carnosine, dipeptide of beta-alanine and L-histidine, DOPA, a precursor of dopamine, PEA, a trace amine and succinic acid have been reported to be increased in patients with OA, postural disorders and ligaments rupture, respectively (**Adams *et al.*, 2012; Ponist *et al.*, 2016; Rabin *et al.*, 2016; W. Zhang *et al.*, 2016; Maerz *et al.*, 2018**).

Cystine, a dimer of amino acid cysteine, showed a negative association with pain in the current study and is reported to attenuate synovial fluid thickening and erosion of cartilage in joints with OA (**Shao *et al.*, 2020**). This strengthens the evidence of cystine possessing anti-inflammatory properties and supports its use as a therapeutic target in intervention studies. GLCAS, a sulphate ester with bile acid group and the stress hormone cortisol and carnosine are the intermediates of cholesterol metabolism pathways. There are numerous studies reporting cortisol to heal and improve pain by recovering the tissues, however, studies on cortisol association in OA are scarce. In the current study, the decreased level of cortisol with increase in pain is likely to be attributed to tissue damage progression. However, role of cortisol and its dysfunction could be better understood by evaluating its potential specifically in OA and pain.

Pathway analysis showed amino acids metabolism and cholesterol metabolism pathways to be involved at a higher level of classification. This could be attributed to the fact that the significant metabolites in our study are synthesised from amino acids like tyrosine, glycine, arginine, cystine, histidine, tryptophan,

phenylalanine and steroid as another precursor. These results are likely to be related to pain modulation which is primarily mediated through neurotransmitters and amino acids being the main substrate for the production of the neurotransmitters. Similar results were shown in a study **(Shell *et al.*, 2016**), where treatment with amino acid precursors was associated with substantial improvement in chronic back pain, reduction in inflammation, and improvement in back pain correlated with increased amino acid precursors to neurotransmitters in the blood. Meta-analysis with pain and PA results in this study are consistent with the literature and support the heterogeneity of these metabolites to be used as metabolic markers and potential therapeutic targets for OA and knee pain. Reproducibility of significant metabolites associated with pain in osteoarthritis in two independent studies, supports the validation of these metabolic biomarkers in *in-vitro* and/or *in-vivo* studies. It is notable that there are fewer significant metabolites associated with pain in the replicate study. This may partially be explained by the high standard deviation in the metabolite concentration in the study dataset. However, the metabolome data was normalised prior to the statistical inferences. The pain pressure thresholds (PPTs') are a measure of pain perception, however, studies are depicting that the transduction of information signals are processed differently within individuals in the presence or absence of pain or stimuli. The different direction of associations of pain and PPTs with a few metabolites in test and replication study may be due to the significant difference in the PPTs in the two study datasets. This also corroborates that there is a lack of true associations because the level of pain experienced amongst individuals varies, and henceforth these results need to be interpreted with caution.

Prior studies have noted the importance of cytokines in pain modulation through various mechanisms/neural pathways (**Miller *et al.*, 2009**), increasing pain and thus changes in alteration of sensory neuron behaviour. A review **(White, Bhangoo and Miller, 2005**) collating evidence from other studies confirms that cytokines are associated with not only inflammatory pain but also in neuropathic pain. This depends upon whether cytokines act upon the receptors in the peripheral or the central sensory system. Similar to other studies **(Lindenlaub *et al.*, 2000; Chapman, Tuckett and Song, 2008**), the evidence of significant positive correlation of TNF-alpha with Cystine and negative association of

cortisol with IL-10 in the current study reflects the immune disorder that appears with pain and tissue damage in patients with OA. Synovial fluid from OA patients shows increased expression of D1-like dopamine receptors and this increased expression may lead to anti-inflammatory effects (**Capellino *et al.*, 2014**). These observations accord with findings from the current study which showed that metabolites are significantly associated with immune-modulatory markers. The different direction of associations of PPTs with the metabolites in test and replication study may be due to the significant difference in the PPTs in the two study datasets. This also corroborates that there is a lack of true associations because the level of pain experienced amongst individuals varies, and henceforth these results need to be interpreted with caution.

We also observed a number of cytokines like IL-10 and TNFα to be significantly associated with the pain-related metabolites, although the cytokines were not associated with the pain scores. This supports the hypothesis that pain in OA is not predominantly an inflammatory condition but a degenerative joint disorder with musculoskeletal pain (**Ayral *et al.*, 2005; Bondeson *et al.*, 2006; Scanzello, Plaas and Crow, 2008; Daghestani, Pieper and Kraus, 2015**). Inflammation in OA is reported to be of varying degrees and progressing with time making it is different from other inflammatory forms of arthritis such as rheumatoid arthritis, psoriatic arthritis and spondylarthritis **(Fu, Robbins and McDougall, 2018**).

The empirical results reported herein are subject to the statistical approaches opted. The effect of estimates in the current study are based on the change in the pain predicting the metabolites alterations. However, it is unclear whether the changes in the metabolites result in pain or the body's response to pain results in the metabolite alterations. Therefore, an alternative approach with pain as the outcome would give another insight into the etiological effects of pain in OA. Another potential alternative approach could be to pool the two studies and use a bootstrapping or cross-validation. This would aim to derive and test the associations to fully use the data and provide a measure of uncertainty.

There are certain limitations to this study. The main limitation is that serum concentration of metabolites may not reflect of the actual concentration of these metabolites at the site of pathology (i.e. the knee) or in the central nervous system. However, measuring serum concentration is easily feasible and hence

can be tested and validated in many cohorts and used for clinical applications. The neurotransmitter related metabolic markers are measured in blood but there no studies demonstrating the correlation of levels of these markers in blood vs brain. As the current study did not measure the neurotransmitters or their receptors in brain it lacks a clearer picture of the metabolism and transportation of the neurotransmitter related metabolites such as DOPA. Also, the bioavailability of Cortisol follows certain circadian rhythms of the body, and the current study did not take into account the measurement time of these markers in the blood. Furthermore, methods used to determine the presence of OA was not similar in the two studies. Also, all the study participants had OA, so this restricts the speculation of these findings to pain experienced in other diseases. Another potential limitation is that this study didn't take into account the use of steroids, dietary intakes, exercise details therefore subject to confounding that may have influenced the estimates. The present study is a cross-sectional study so the causal nature of the associations found cannot be determined and prospective studies are required to determine causation.

While exploring the relationship between gut microbial community and knee pain, the most interesting finding in the current study was the association of pain with *Parasutterella* which is shown to be decreased with an increase in pain. *Parasutterella* is identified as a succinate producer and plays a role in bile acid maintenance and cholesterol metabolism (**Ju et al., 2019**). Other studies have reported it as a low-abundance and core microbial member in participants with spared nerve injury-induced neuropathic pain (**Brandon-Mong et al., 2020**). Also, mice deficient in interleukin-1 receptor antagonists (Il1rn −/− mice) spontaneously develop autoimmune arthritis and IL1rn −/− intestinal microbial composition was characterized with a decrease in *Parasutterella* by 10 fold **(Rogier et al., 2017**). *Streptococcus* was shown to be significantly associated with pain ($p<0.05$) in the current study and is previously reported to be strongly associated with joint pain (**Boer *et al.*, 2019b)**. The non-significance of these two microbes after correcting for multiple testing demonstrates to further investigate the associations of gut microbiome and knee pain in larger cohort. This also accords to the fact that knee pain doesn't have a direct association with gut microbiome unlike the gut or liver related diseases/conditions. The significant association of these gut microbes with the significantly associated

serum metabolites with pain demonstrates the systemic effect produced by the microbes via the molecular pathways which needs future work.

In conclusion, the current study shows metabolites such as acyl ornithine, carnosine, cortisol, cortisone, cystine, DOPA, GLCAS, PEA, succinic acid and microbes such as *Parasuttrella* and *Streptococcus* are correlated with pain in knee OA. The current study results indicate the potential of *Parasuterella* and *Streptococcus* to be explored/harnessed and used as supplements to existing therapies against OA and pain. Association of metabolites suggest their potential use as metabolic markers in the two independent cohorts. Also, these metabolites are associated with immune responses via inflammatory markers such as IL-10, IL-8, TNFα and IL-1β. Although the current study sheds light on the associations between metabolites circulating in serum and pain traits , future work will be required to interpret and understand the mechanisms involved in pain modulation. Investigating these modulating factors impacting pain may eventually translate into metabolic marker development for use in the clinical trials after *in vivo*/*in vitro* validations. Given the efficacy of the drugs in pain relief, metabolite therapy holds a promise as potential therapeutics against pain and have anticipated benefits in the treatment and management of OA.

5

Chapter 5

MULTI-OMICS APPROACH TO DESCRIBE NAFLD: FROM CORRELATION TO CAUSATION

GUT MICROBIOME, SERUM METABOLOME AND DIET EFFECTS IN NAFLD

5.1 ABSTRACT

BACKGROUND

The global incidence and prevalence of NAFLD is rising rapidly. Dietary factors have been associated with the development of NAFLD. Gut microbiome composition has been implicated in increasing the risk of NAFLD. The aim of the current study is to investigate the extent to which gut microbiome composition and gut microbial metabolites mediate the effect of dietary intakes on prevalence of NAFLD.

METHODS

This is an observational case-control study comparing NAFLD vs healthy volunteers (HV). This study analysed multi-omics data from 16S rRNA sequencing of the gut microbiome (n= 98 NAFLD vs 68 HV), serum metabolic profiling (n = 83 NAFLD vs 62 HV), dietary factors (n = 80 NAFLD vs 73 HV) to identify the molecular and nutritional signatures of NAFLD and its severity sub-groups. Statistical approaches such as ANCOM, logistic regression, OPLS-DA and LASSO were utilised to determine the omics features associated with the NAFLD status. Correlations were tested to investigate the relationship between the gut microbiome and metabolites. Causal relationship between the dietary factors and NAFLD status was analysed utilizing formal mediation analysis, with gut microbes as the mediator.

RESULTS

Collinsella (OR = 6.9, FDR <0.001) and *Bifidobacterium* (OR = 5.2, FDR <0.001) were found to be the most significantly associated with prevalence of NAFLD. Fractions and subfractions of cholesterol belonging to the group of medium HDL ratios, HDL2, HDL3, VLDL and fatty acid ratios were found to be different in the metabolic profiles of healthy volunteers and NAFLD participants. Dietary analysis revealed that the diet of NAFLD patients was significantly deficient in iron, rich in selenium and high in total sugars from carbohydrates and foods classified under potatoes group. The effect of reduced iron intake and increased intake of total sugars from carbohydrates on NAFLD was observed to be acted as potential full mediator with increased abundance of *Collinsella* and *Bifidobacterium*, However, the effect of increased intake of

selenium on NAFLD was found to be acted as potential mediateoronly with the increase in *Collinsella*.

CONCLUSION

Collinsella and Bifidobacterium were found to mediate the effect of increased dietary sugars, potatoes, milk, meat and their products as well as micronutrients such as reduced iron intakes suggesting a causal link between diet and increased risk of NAFLD. In addition, increase in dietary selenium predicting the likelihood of NAFLD was found to be mediated by *Collinsella*. Multi-omics data analysing approaches utilised in this study assist in understanding the systemic effect of diet, gut microbiome and metabolome on the development of NAFLD.

5.2 INTRODUCTION

Non-alcoholic fatty Liver Disease (NAFLD) is one of the most common and chronic liver diseases globally which is primarily identified with the presence of excess fat in the liver among people who don't drink alcohol more than the recommended daily limits. NAFLD represents a spectrum of liver disease which is initiated by fat deposition, progressing to non-alcoholic steatohepatitis (NASH). NASH is manifested with necro-inflammation and fibrosis which then eventually progresses to cirrhosis and hepatocellular carcinoma (HCC) (**Bedossa *et al.*, 2012**). Various risk factors like increased weight or obesity, diabetes and excessive caloric intake are previously found to be strongly associated with NAFLD severity **(Abeysekera *et al.*, 2020).** NAFLD being asymptomatic often goes unnoticed in its initial phases and is by chance diagnosed by aberrant liver enzymes or imaging.

NAFLD is hepatic manifestation of metabolic syndrome occurring due to the quality and quantity of dietary intakes, for instance, higher carbohydrates, lower fibre, higher sugar and excessive intake of saturated fats **(Berná and Romero-Gomez, 2020)**. The accumulation of the extra amount of nutrients or metabolites in the liver is accompanied by inflammation which is later progressed to fibrosis

and cirrhosis as advanced stages. The macronutrients from dietary intakes such as carbohydrates, proteins and fats form the main energy source for humans which are converted to other forms of nutrients and metabolites such as sugars, amino-acids by the gut microbiome (**Flint, Scott, Louis, *et al.*, 2012; Sanchez *et al.*, 2017**). The differences in the gut microbiome of NAFLD and heathy participants is already established which plausibly indicates the improvement of NAFLD based on dietary nutrients. However, clinical evidence of the role of nutrients and causal mechanisms to understand the pathophysiology of NAFLD is still burgeoning.

There is a large volume of published studies describing that the gut microbiome modulates human metabolism through diet. A study on healthy individuals observed the effect of plant nutrients such as fibre and choline and artificial sweeteners such as saccharin on bile acids. **(Tang *et al.*, 2019**). The production of these compounds is mediated via gut enterotypes and act as a substrate or inhibitor of the related metabolic pathways. Other in-vivo or in- vitro studies showed the effect of polyphenols (**Anhê *et al.*, 2015**) and choline (**Romano *et al.*, 2015**) from dietary sources on metabolic factors such as TG accumulation and TMAO production respectively. Gut microbes or gut-derived metabolites **(Ghosh *et al.*, 2020; Hollander and Kaunitz, 2020**) can be transported to the liver via the portal venous system. Existing research reports that fatty acids alter the gut microbiome and the dietary fatty acid composition has been shown to the development of insulin resistance and hepatic steatosis (**Hua *et al.*, 2010**) which recognises the critical interplay of these factors in NAFLD. The interaction of gut microbiome and diet in NAFLD is also evident from a study conducted in a mice model, in which faeces from high-fat diet mice were transplanted to germ-free mice, developing NAFLD independent of obesity (**Le Roy *et al.*, 2013)**. Also, recent dietary intervention studies using dextrin (**Zhang, Chen and Cui, 2021**) which is a low molecular weight carbohydrate and mung bean coat (**Hou *et al.*, 2022**) which is mainly composed of the starch and glycogen have shown to modulate the gut microbiome and metabolome in NAFLD mice models. This may be attributed to the fact that these compounds get hydrolysed and supress the rise in the blood sugar which is a risk-factor contributing to NAFLD.

Over the past decade, there have been reports linking NAFLD with the gut microbiome (**Aron-Wisnewsky *et al.*, 2020**), metabolites (**Dai *et al.*, 2020**) and diet **(Katsagoni *et al.*, 2017; Cheng *et al.*, 2022)**. The available studies on NAFLD, to date, include investigating data from individual omics platform at a time and therefore have been unable to explain how different components link with each other. Also, to have a better understanding of how diet effects NAFLD and their contribution to the composition and diversity of gut microbiome, systematic evaluation of long-term dietary intakes is needed. Diets are very complex making it difficult to be stratified at the ingredients level and also the cumulative effects of food items make it difficult to predict the associations with the risk or severity of NAFLD, henceforth research beyond dietary items and dietary patterns needs to be explored.

Gut microbes possess the ability to metabolise and utilise dietary nutrients. The carbohydrate-activating enzymes in the gut microbes utilizes starch and fibre components and thus helps in carbohydrate digestion. Proteins in the dietary intake gets fermented to produce amino acids by the microbially produced proteases. Gut microbes helps in lipid digestion mediating the production of secondary bile acids from the primary bile acids. These nutrients then get converted to sugars through multiple pathways of gluconeogenesis.

Various studies (**Abdelmalek *et al.*, 2010; Basaranoglu, Basaranoglu and Bugianesi, 2015; Jensen *et al.*, 2018**) have reported the impact of macronutrients such as carbohydrates, fats, proteins and sugar in the NAFLD condition. For instance, increased intake of fructose is associated with the fibrosis, high fat diet is associated with elevated triglycerides and inflammation, however, only fewer studies have investigated the role of micronutrients in NAFLD. The liver is a storehouse of micronutrients i.e. vitamins and minerals and also produces regulatory proteins such as albumin and bile acids which help in the binding of fat-soluble nutrients. The deficiency of micronutrients such as vitamin D and zinc has previously been reported to contribute towards auto-immune diseases (**Prietl *et al.*, 2013; Hojyo and Fukada, 2016**) and NAFLD (**Hariri and Zohdi, 2019; Kim *et al.*, 2020**). Vitamin D intake has been reported to be associated with reduction in inflammation, stellate cell activation and liver enzymes, reduced intake of zinc on the other hand, increases endoplasmic

reticulum stress resulting into chronic inflammation in Type-2 diabetes, an underlying co-morbidity associated with NAFLD. Previous studies have also reported that micronutrients have antioxidant, anti-fibrotic and lipo-protective properties/effects **(Gillespie, 2021)**. Furthermore, micronutrients in diet such as polyphenols **(Singh *et al.*, 2019)**, compounds found in red wine and tea as well as vitamin D **(Singh *et al.*, 2020)** have been reported to improve the growth of beneficial microbes in the gut.

This study hypothesize that dietary components modifies the gut microbiome and thus influences the metabolic conditions predisposing to the development of NAFLD. The current study aims to explore the relationship between NAFLD and gut microbiome, serum metabolome and average dietary intakes in the form of macronutrients, micronutrients and food groups from the previous one-year while comparing it with healthy volunteers in the UK population. This study has been the first study to investigate the interplay of these factors in NAFLD in a single cohort.

5.3 METHODS

5.3.1 STUDY DESIGN, PARTICIPANTS RECRUITMENT AND DATA COLLECTION

This observational case-control study was conducted among patients with NAFLD recruited at hepatology clinics of Nottingham University Hospitals NHS Trust and healthy volunteers (HV) at baseline visit from the Omega3 diet intervention study **(Vijay, Astbury, *et al.*, 2021**). Participants provided the faecal samples, and this was followed by DNA extraction for microbiome sequencing. Serum metabolic profiling was done using Nightingale panel. The study flowchart is presented in **Figure 5.1**.

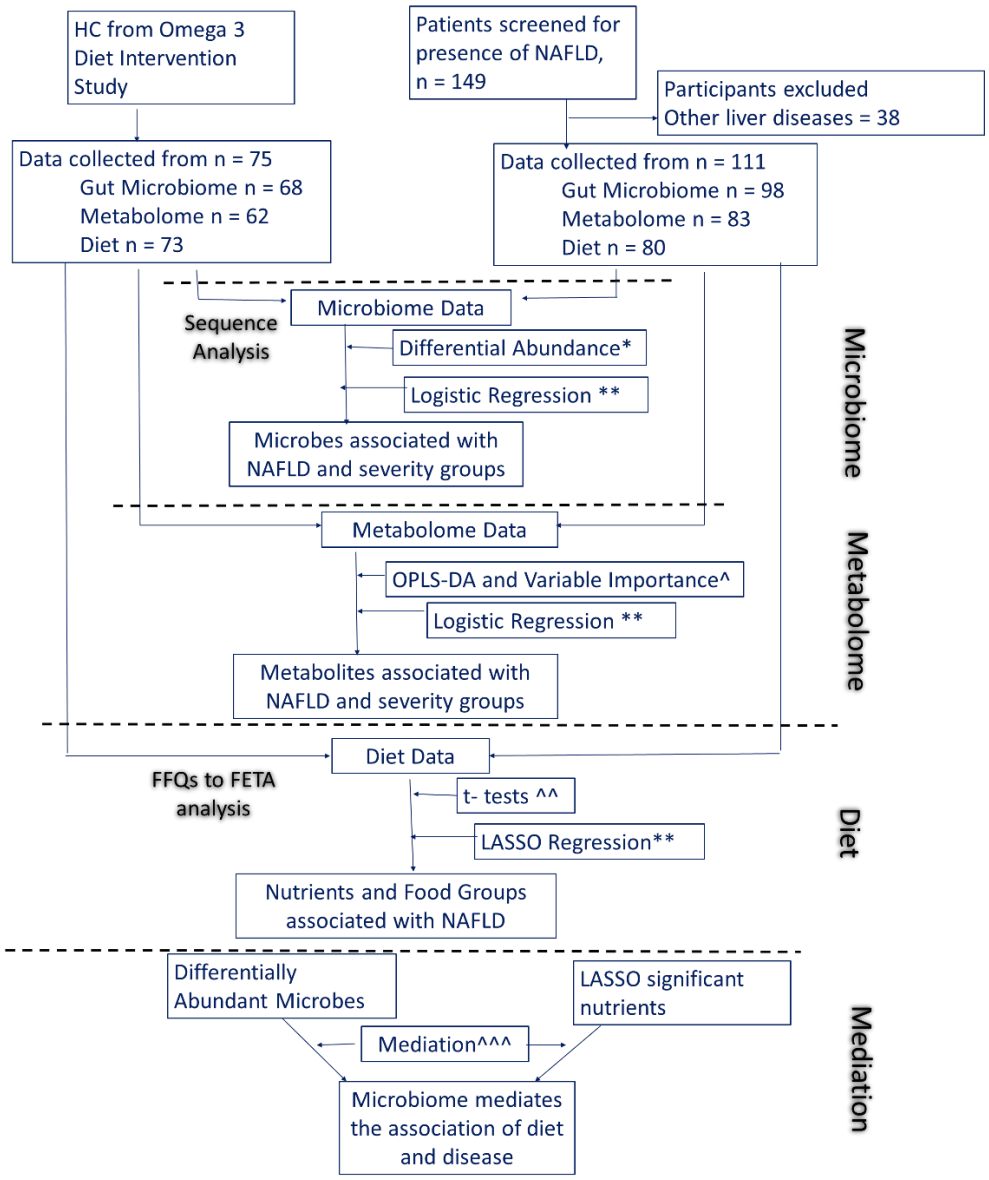


Figure 5.1: Study design and Methodology: This figure presents the consort diagram for the study. *= adjusted for age, gender and BMI, ** = adjusted for age, gender and BMI for rarefied and Inverse Normalised Microbiome or Metabolome or diet data, ^ = Inverse Normalised data, ^^ Energy adjusted inverse normalised data, ^^^ = *Collinsella* or *Bifidobacterium* as a mediator, diet as predictor and disease status as outcome and model adjusted for age, gender and BMI. HV = Healthy Controls.

5.3.2 CLINICAL EVALUATION OF NAFLD

NAFLD was diagnosed based on the combined approach of testing the presence of fat deposition analysed by MRI, liver tests and a fibro scan. NASH was diagnosed based on liver biopsy tests, fibrosis and cirrhosis were measured using MRI and a fibro scan. Details are mentioned in Chapter 2, Section 2.1.3.

5.3.3 DIETARY ASSESSMENT AND ANALYSIS

Dietary information for NAFLD participants and healthy volunteers was collected based on European Prospective Investigation into Cancer and Nutrition Food Frequency Questionnaires (**Kroke *et al.*, 1999**) (EPIC-Norfolk FFQs) and TWINS-UK FFQs (**Teucher *et al.*, 2007**) respectively. Both the FFQs record individual's average habitual food intake during the previous year. TWINS-UK-FFQs and EPIC-Norfolk FFQs were made comparable by excluding a few variables and combining the frequencies of a few variables as described in Chapter 2, Section 2.5.

The FFQs were inputted and analysed using the FETA (FFQ EPIC Tool for Analysis) software calculating the intakes of 17 macronutrients, 27 micronutrients, 14 food groups and energy intakes. Energy-adjusted nutrient intakes were inverse normalised for further analysis.

5.3.4 SAMPLE COLLECTION, PROCESSING AND DNA EXTRACTION

DNA was extracted from stool samples using current standard laboratory techniques using commercially available kits. Briefly, DNA was extracted taking 125 mg of stool sample and homogenizing in 500ul lysis buffer (500mM NaCl, 50mM Tris-HCl, 50mM EDTA, 4% SDS). Bead beating (using MagNA Lyser, Roche) was done by adding 25mg zirconia beads and 3 glass beads (3mm). The protocol was followed with interval steps of 1-minute beating, and 1-minute rest on ice and this was repeated 3 times. Further steps of extraction were done using the QIAmp DNA Stool Mini Kit (Qiagen; cat 51504). DNA was eluted in 100ul of nuclease-free water and stored at -20 ^{0}C.

5.3.5 MICROBIOME SEQUENCING AND ANALYSIS

V3-V4 region of the 16S rRNA gene was amplified using universal primers and was sequenced on the MiSeq platform (Illumina). After quality control and

demultiplexing, 15,710 features were identified with a total frequency of 15,180496. Taxonomic assignments using SILVA classifier based on SILVA database identified 486 taxa at genus level. The genus with a relative abundance of < 0.1% in 95% of samples were removed and abundances were inverse normal transformed before downstream analysis as described in Sections 2.23.

5.3.6 STATISTICAL ANALYSIS

Mean and standard deviation was used to describe the dietary intake data. Differential abundance analysis was performed to compare the gut microbiome of HV with NAFLD patients using ANCOM2. To compare the relationship of the microbiome, metabolome and dietary nutrients with HV and NAFLD groups, association tests were done by running multivariate logistic regression models adjusted for age, gender and BMI, with disease status as the outcome and microbial abundances or metabolome or diet intake as the predictor. The odds ratio was calculated, and p-values were corrected for multiple testing corrections with a threshold of FDR <0.05. The batch effect was checked by plotting PCA for the samples and no batch effect was observed as the samples in batches showed a complete overlap. Orthogonal Partial Least Square Discriminant Analysis (OPLS-DA) was performed on metabolome data to classify the groups based on the diagnosis status or disease severity. Variable Important Projection (VIP) scores indicating the level of contribution to distinguish the metabolic profiles of two groups were computed for the metabolites. LASSO regression was performed on the nutrient intakes for investigating the association with the disease outcome or severity status. Disease status was treated as the dependent variable where HV was coded as 0 and NAFLD participants were coded as 1. Data on a subset of participants having both diet data and microbiome data was investigated for the associations of significant energy-adjusted nutrients and food group intakes with the significant gut microbes using Spearman's correlation. Data from a subset of participants having both metabolome and gut microbiome data was investigated for the association of metabolites with significant microbes using Spearman's correlation. Formal mediation analysis was performed to investigate if the effect of diet mediated via the gut microbiome on NAFLD. All analysis was performed using R packages.

5.4 RESULTS

5.4.1 CLINICAL CHARACTERISTICS

All the participants included in the study were UK based with an average age of <67 years. The average BMI for the NAFLD group was 34kg/m2 which was significantly higher (p-value <0.001) than the average BMI of 26kg/m2 for the HV group. The significant difference in the age of HV and NAFLD groups designate that NAFLD is manifesting even at a very young age of people. NAFLD was divided into two subgroups, NASH irrespective of fibrosis and cirrhosis based on the histological severity of the disease condition.

5.4.2 MICROBIOME, METABOLOME AND DIET DATA DETAILS

Microbiome sequencing was done for 74 samples from HV participants and 149 samples recruited initially for the NASH cohort. After rarefaction, removal of samples with low feature counts and obtaining data for participants having NAFLD, the analysis includes microbiome data for 68 HV subjects and 98 NAFLD subjects with 155 annotated taxa. In the microbiome dataset, there were 103 females (62%) and 63 males (38%) and overall, 52 participants were diagnosed with NASH irrespective of fibrosis and 36 with cirrhosis.

Serum metabolomic profiling was done for 62 HV and 84 NAFLD participants including 50 males and 95 females. Amongst the NAFLD cohort, a total of 43 participants had NASH and 31 had cirrhosis.

Dietary intake data were collected from 163 participants using FFQs. After excluding data from participants having more than 10 missing input variables, the current analysis included data from 73 HV and 80 NAFLD patients. Amongst the 153 participants, 99 (64%) were females and 53 (34%) males, 1 HV chose not to disclose their sex. In the diet dataset, 45 participants had NASH irrespective of fibrosis and 26 participants had cirrhosis. Details are mentioned in **Table 5.1**.

Diet Data	Trait	HV (73)	NAFLD (80)	p-value
	Gender M	9	44	<0.001***
	F	63	36	
	Age*	66.6 (8.92)	59.82 (12.02)	<0.001***
	BMI*	26.38 (4.17)	34.92 (5.29)	<0.001***
	Energy Intake	2170.0 (807.4)	1762.7 (609.1)	0.002**
Microbiome Data		**HV (68)**	**NAFLD (98)**	
	Gender M	9	54	<0.001***
	F	59	44	
	Age*	66.08 (9.07)	59.71 (11.89)	<0.001***
	BMI*	26.46 (4.24)	34.52 (5.34)	<0.001***
Metabolome Data		**HV (62)**	**NAFLD (83)**	
	Gender M	6	44	<0.001***
	F	56	39	
	Age*	66.26 (9.10)	59.25 (12.05)	<0.001***
	BMI*	26.25 (4.38)	34.32 (5.66)	<0.001***

*Age and BMI are represented as mean (SD)

Table 5.1: Clinical Characteristics of participants in the study. *** pval ≤ 0.001, ** pval < 0.05, *

5.4.3 THE GUT MICROBIOME IS ALTERED IN NAFLD PARTICIPANTS

Differential abundance analysis on microbiome data resulted in fifteen altered genera after adjusting for age, gender and BMI in the two groups at the detection threshold of 0.9. *Bifidobacterium*, *Collinsella*, *Lachnoclostridium*, *Lachnospiraceae.ND3007.group*, *Fusicatenibacter* were higher in NAFLD while *Alistipes*, *Christensenellaceae.R.7.group*, *Ruminococcaceae.UCG.014*, *Clostridium.sensu.stricto.1*, *Peptoclostridium*, *Ruminococcaceae.UCG.010*, *Barnesiella* higher in HV (**Figure 5.2**). However, there were no differentially abundant taxa between NAFLD severity groups (NASH vs Cirrhosis).

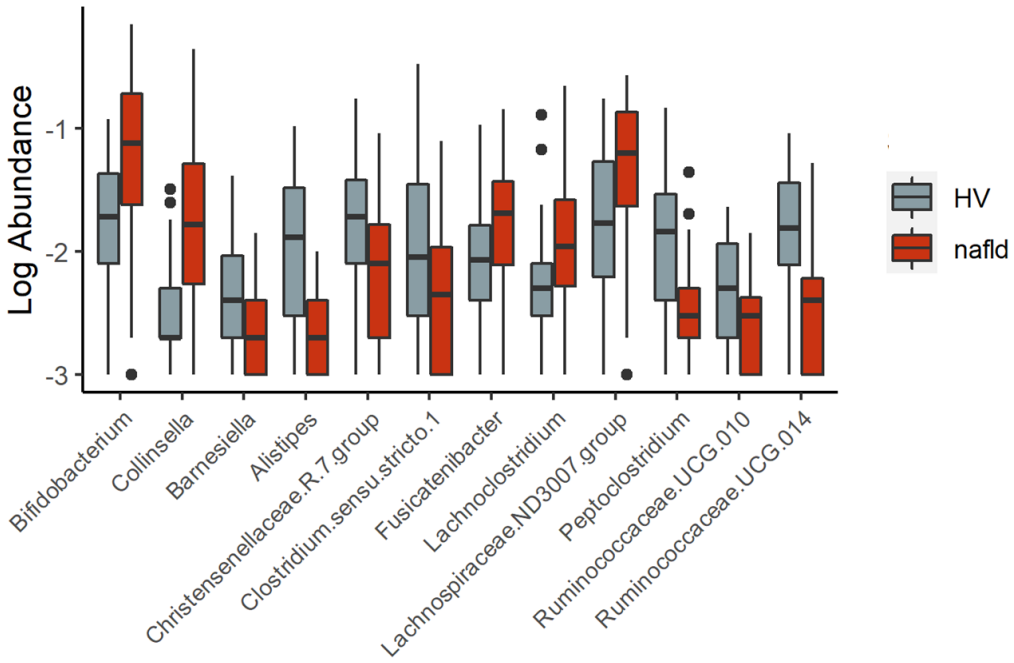


Figure 5.2: Differentially abundant microbes in NAFLD patients when compared with Healthy Controls. The y-axis presents the log of relative abundance of the microbes.

5.4.4 ASSOCIATION OF THE GUT MICROBIOME WITH NAFLD AND SEVERITY SUB-GROUPS

To explore the utility of gut microbiome composition predictive for the disease outcome,NAFLD status or severity subgroups were regressed on relative abundances of microbial taxa adjusted for age, gender and BMI. Along with the differentially abundant taxa, NAFLD was also found to be significantly associated (FDR <0.05) with an increase of *Gardnerella, Eggerthella, Lachnospiraceae.UCG.008, Butyricicoccus,* and decrease of *Bacteroides, Odoribacter, Parabacteroides, Porphyromonadaceae uncultured, Anaerotruncus, Oscillibacter, Ruminiclostridium.6, Desulfovibrio, Mollicutes.RF9 uncultured.bacterium* when compared with HV (**Figure 5.3** and **Table 5.2**). *Alistipes* (OR = 2.29, p = 0.037) and *Ruminococcaceae* uncultured taxa (OR = 2.38, p = 0.005) were found to be significantly associated (p < 0.05) with the likelihood of NASH, showing decreased in cirrhosis group in comparison to NASH participants.

	beta	p-value	std. error	FDR	OR	LCI	UCI
Bifidobacterium	1.66	<0.001***	0.39	<0.001***	5.24	2.45	11.23
Gardnerella	0.91	0.01**	0.37	0.05**	2.49	1.22	5.12
Collinsella	1.94	<0.001***	0.46	<0.001***	6.94	2.83	17.05
Eggerthella	1.03	<0.001***	0.34	0.02**	2.79	1.42	5.47
Bacteroides	-1.12	<0.001***	0.3	<0.001***	0.33	0.18	0.58
Barnesiella	-1.53	<0.001***	0.38	<0.001***	0.22	0.1	0.45
Odoribacter	-1.19	<0.001***	0.36	0.01**	0.3	0.15	0.62
Parabacteroides	-1.27	<0.001***	0.32	<0.001***	0.28	0.15	0.53
Porphyromonadaceae uncultured	-0.98	0.01**	0.35	0.03**	0.37	0.19	0.75
Alistipes	-1.65	<0.001***	0.37	<0.001***	0.19	0.09	0.4
Christensenellaceae.R.7.group	-1.11	<0.001***	0.32	0.01**	0.33	0.18	0.62
Clostridium.sensu.stricto.1	-1.41	<0.001***	0.34	<0.001***	0.24	0.12	0.48
Lachnoclostridium	0.7	0.01**	0.28	0.05**	2	1.17	3.44
Lachnospiraceae.UCG.008	0.81	<0.001***	0.27	0.02**	2.24	1.32	3.79
Peptoclostridium	-1.1	<0.001***	0.32	<0.001***	0.33	0.18	0.62
Anaerotruncus	-0.77	0.01**	0.29	0.03**	0.46	0.26	0.81
Butyricicoccus	0.76	0.01**	0.28	0.03**	2.14	1.24	3.7
Oscillibacter	-1.74	0.01**	0.63	0.03**	0.18	0.05	0.6
Ruminiclostridium.6	-1.32	0.01**	0.48	0.03**	0.27	0.1	0.69
Ruminococcaceae.UCG.010	-1.76	<0.001***	0.42	<0.001***	0.17	0.08	0.39
Ruminococcaceae.UCG.014	-1.64	<0.001***	0.39	<0.001***	0.19	0.09	0.42
Desulfovibrio	-1.52	<0.001***	0.46	0.01**	0.22	0.09	0.54
Mollicutes.RF9 uncultured.bacterium	-1.35	<0.001***	0.42	0.01**	0.26	0.11	0.59

Table 5.2: Association of gut microbes and disease status (HV vs NAFLD). The table shows beta coefficients, p-values, standard error (std.error), FDR values, odds ratio (OR) and upper limit (UCI) and lower limit (LCI) of confidence interval values from the regression model adjusted for age, gender and BMI. *** pval $\leq$ 0.001, ** pval < 0.05.

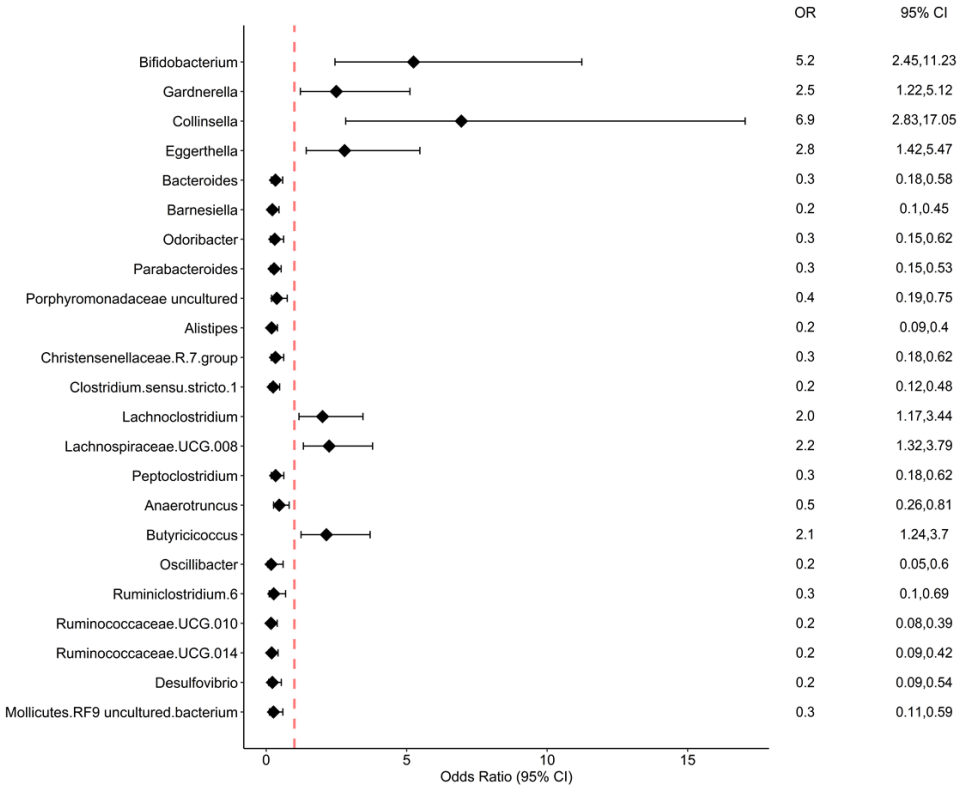


Figure 5.3: Forest plots of odds ratio (95% CI) of the significant microbes (FDR <0.05) associated with NAFLD when compared with HV.

5.4.5 METABOLIC PROFILE DETERMINES THE DISEASE STATUS

To access the difference in the metabolic profile of healthy and NAFLD participants and also the NAFLD severity groups, OPLS-DA was performed on the metabolomics data. **Figure 5.4** presents the score plot of OPLS-DA showing a clear separation between the healthy and NAFLD participants' metabolic profiles. This indicates that there were marked differences in the metabolic profile of HV and NAFLD participants. A high R^2 (cumulative) score (R^2X = 0.743, R^2Y = 0.907) indicating goodness of fit and a high Q^2 (cumulative) score (Q^2 = 0.851) indicating the predictability of model demonstrates that the model constructed is reliable. To investigate the metabolites contributing toward the separation of metabolic profiles in two groups, Variable Importance Projection (VIP) scores were computed. OPLS-DA model showed 89 variables with the VIP score >1 (**Table 5.3)** which indicates the importance of these variables in

the separation of the two groups. These mainly include metabolites belonging to the group of relative lipoprotein lipid concentrations, cholesterol, lipoprotein subclasses, fatty acids, apolipoproteins, amino acids, ketone bodies, glycerides and phospholipids. **Figure 5.5** shows the bubble plot depicting the VIP scores for the top 20 metabolites contributing strongly to the classification of the two groups (HV vs NAFLD). The discriminant analysis test between the NASH and cirrhosis participants' metabolic profiles showed an overlap with no separation (**Figure 5.6**) indicating there is no difference in the metabolic profiles of severity subgroups.

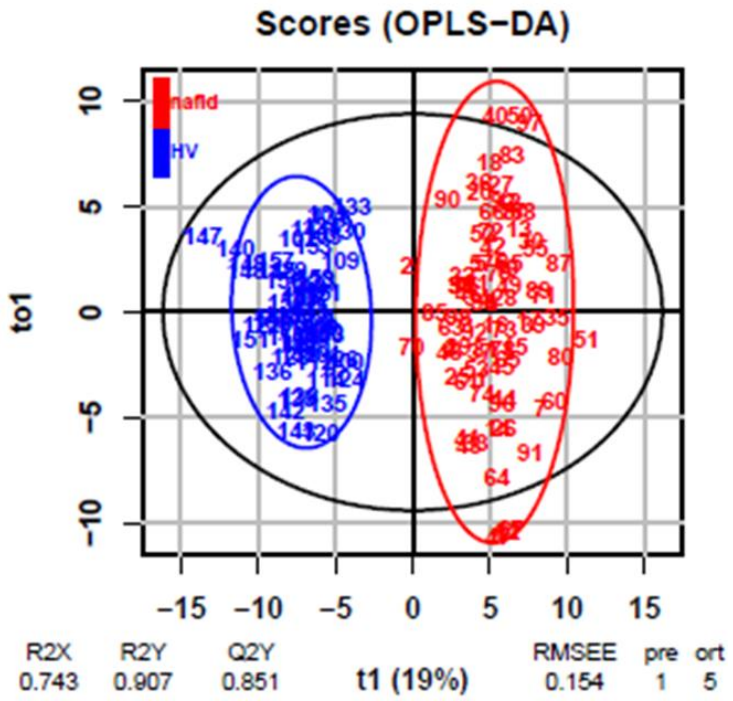


Figure 5.4: OPLS-DA for Healthy vs NAFLD participants. The x-axis presents the t1 score (predicted principal components-score value of main components) representing the difference between HV and NAFLD participants (intergroup). The y axis presents to1 score (orthogonal principal component-score value of orthogonal components) representing the difference within HV and NAFLD participants (intragroup).

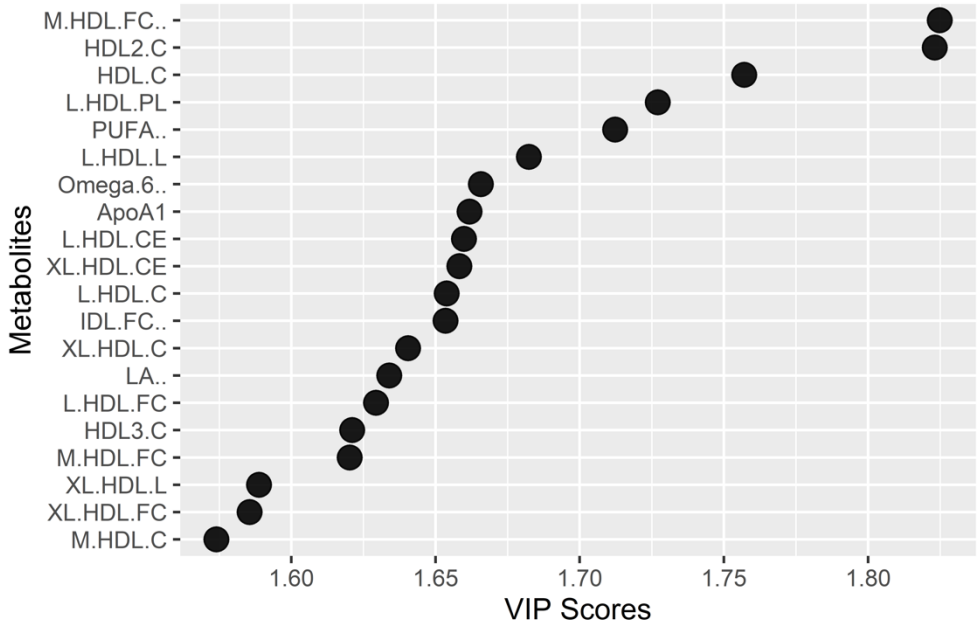


Figure 5.5: Variable Importance Projection: The x-axis presents the Variable Importance Scores from the OPLS-discriminant analysis model comparing HV versus NAFLD participants for the metabolites presented on the y-axis.

	VIP Score
M.HDL.FC %	1.82487
HDL2.C	1.8231
HDL.C	1.75708
L.HDL.PL	1.727062
PUFA %	1.712266
L.HDL.L	1.682354
Omega.6 %	1.665724
ApoA1	1.661781
L.HDL.CE	1.659852
XL.HDL.CE	1.65826
L.HDL.C	1.653873
IDL.FC %	1.653492
XL.HDL.C	1.64046
LA %	1.633969
L.HDL.FC	1.629379
HDL3.C	1.621086
M.HDL.FC	1.620221
XL.HDL.L	1.588826
XL.HDL.FC	1.585555
M.HDL.C	1.574042
HDL.size	1.571854
S.VLDL.FC %	1.56723
M.HDL.CE	1.557749
XL.VLDL.CE	1.530172
XL.HDL.PL	1.516339
L.VLDL.CE	1.514924
Gly	1.51342
VLDL.size	1.51249
XL.VLDL.P	1.49972
XL.VLDL.L	1.487402
XL.VLDL.TG	1.48165
L.VLDL.C	1.474032
L.VLDL.P	1.470128
M.HDL.L	1.467489
XS.VLDL.C %	1.456589
M.HDL.C %	1.453748
L.VLDL.L	1.430884
S.HDL.TG %	1.429911
XL.VLDL.C	1.427713
M.HDL.PL	1.425145
L.VLDL.PL	1.423057
Ile	1.420579
M.VLDL.PL %	1.410758
S.VLDL.PL %	1.393315
XL.VLDL.PL	1.392975
L.VLDL.TG	1.389763
M.VLDL.P	1.387129

L.VLDL.FC	1.380425
XS.VLDL.CE %	1.377445
Sphingomyelins	1.367651
XL.VLDL.PL %	1.349096
XS.VLDL.TG %	1.343257
M.VLDL.TG	1.32649
M.HDL.TG %	1.313371
M.VLDL.L	1.310286
MUFA %	1.297375
M.VLDL.PL	1.287333
XL.VLDL.FC	1.284004
M.VLDL.FC	1.257593
L.LDL.FC %	1.255632
L.HDL.C %	1.247604
VLDL.TG	1.246938
S.HDL.TG	1.245115
S.VLDL.TG %	1.232619
SFA %	1.230683
His	1.209355
L.HDL.FC %	1.204407
M.VLDL.C	1.184678
TG.PG	1.176777
S.VLDL.P	1.167261
S.LDL.FC	1.149537
XXL.VLDL.CE	1.140807
L.HDL.CE %	1.132298
S.VLDL.C %	1.123473
Total.TG	1.117128
S.VLDL.TG	1.11477
XXL.VLDL.C	1.113655
M.HDL.CE %	1.105227
bOHbutyrate	1.104732
DHA %	1.080614
XXL.VLDL.FC	1.076194
XXL.VLDL.TG %	1.072871
L.HDL.TG	1.067365
XS.VLDL.PL %	1.06732
ApoB.ApoA1	1.024438
XS.VLDL.FC %	1.017442
S.HDL.FC	1.00691
S.LDL.PL	1.00076

Table 5.3: VIP score. This table presents the metabolites with variable importance scores (VIP) >1.

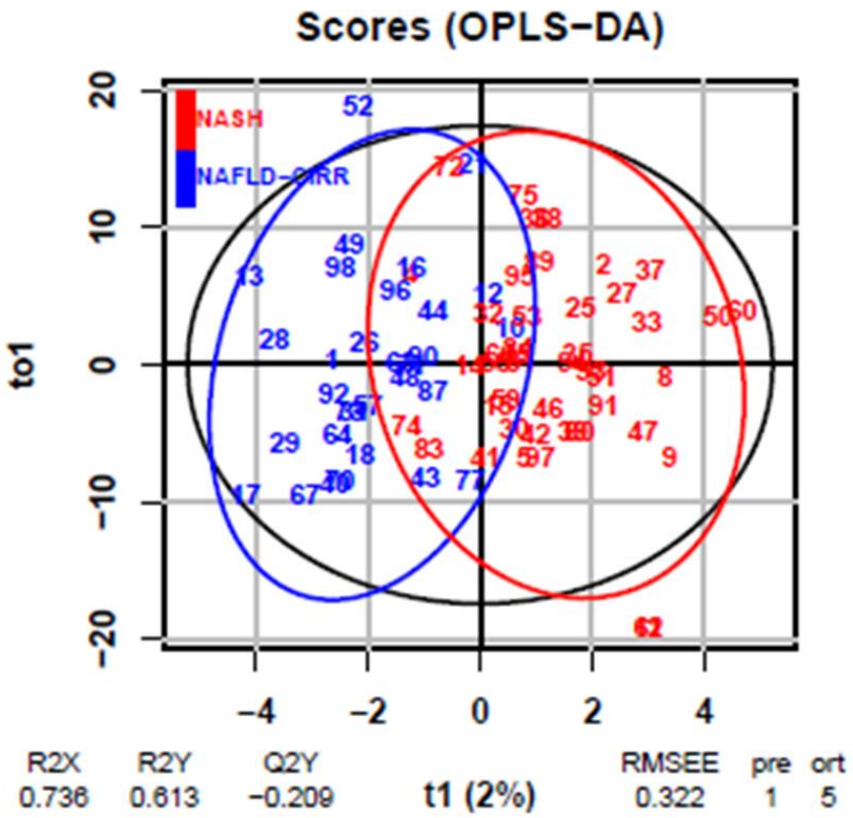


Figure 5.6: OPLS-DA for NASH vs CIRRHOSIS participants. The x-axis presents t1 score (predicted principal components-score value of main components) representing the difference between HV and NAFLD participants. The y axis presents to1 score (orthogonal principal component-score value of orthogonal components) representing the difference within HV and NAFLD participants.

5.4.6 METABOLIC SIGNATURES OF NAFLD AND SEVERITY SUB-GROUPS

The OPLS-DA classification method was further elaborated to see the associations of metabolites with the likelihood of the disease status. For this logistic regression was performed treating the disease status as the dependent variable and metabolites as the independent variable. Metabolite concentration in NAFLD participants when compared with the metabolite concentration with HV participants showed the ratio of MUFA and saturated fatty acid to total fatty acid, isoleucine, the concentration of large and very large VLDL particles and cholesteryl esters in very large VLDL, cholesterol and cholesterol esters in large VLDL, average diameter for VLDL particles, triglycerides to total lipids ratio in very small VLDL, triglycerides to total lipids ratio in small and medium HDL

to be increased in NAFLD patients in comparison to HV patients with >4 times higher likelihood of having NAFLD.

HDL cholesterol and its subfractions (HDL2 and HDL3), phospholipids, cholesterol, cholesteryl esters, and free cholesterol in large and very large HDL, were found to be decreased in NAFLD participants along with Sphingomyelins, apolipoprotein A1, ratio of omega-6 fatty acids, PUFA and docosahexaenoic acid to total fatty acids, glycine and histidine. Details in **Table 5.4** and **Figure 5.7.** The same analysis was done by dividing the NAFLD participants into NASH irrespective of fibrosis and cirrhosis, to determine if there were any metabolites associated with the severity or progression of the disease. Metabolites such as histidine, cholesterol in very large HDL, cholesteryl esters in very large HDL, average diameter for LDL particles, free cholesterol to total lipids ratio in large VLDL, phospholipids to total lipids ratio in small HDL were found to be significantly associated with severity subgroups of NAFLD (p-value <0.05) but turned non-significant after adjusting for multiple testing.

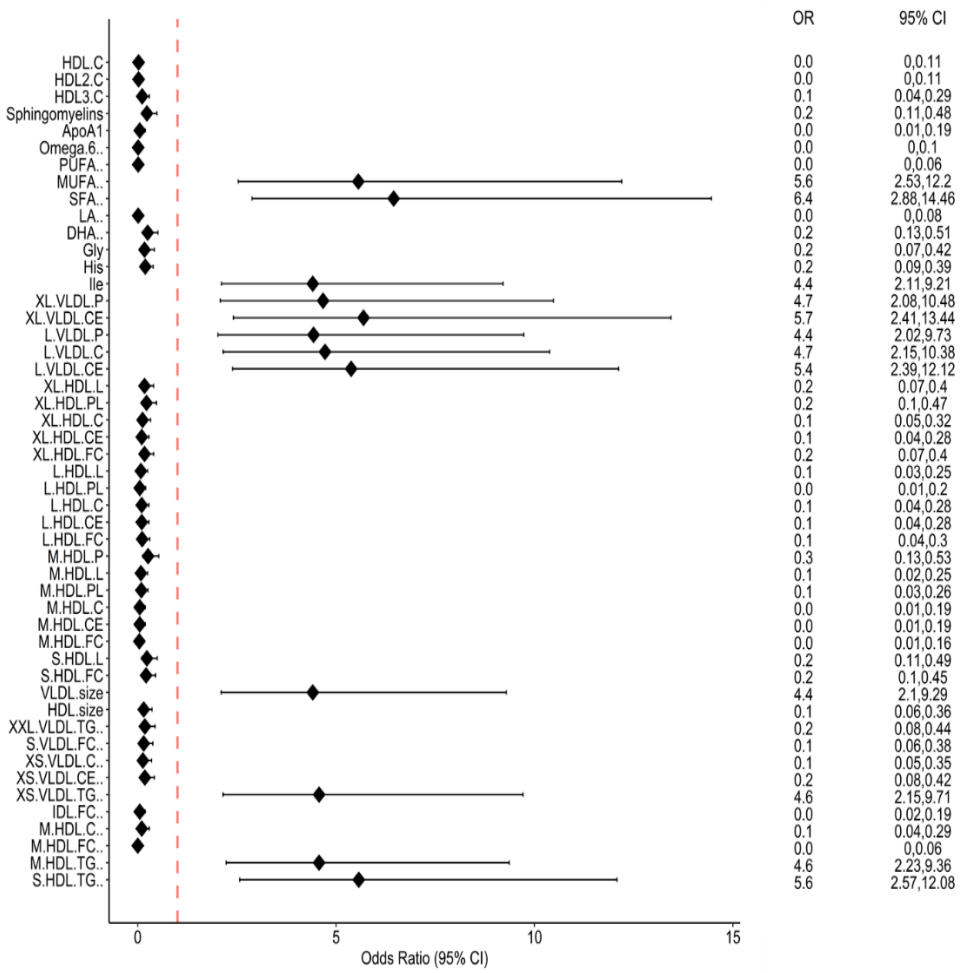


Figure 5.7: Forest plots of odds ratio (95% CI) of the significant metabolites (FDR <0.05) associated with NAFLD when compared with HV.

	beta	P-val	std. error	FDR	OR	LCI	UCI
HDL.C	-3.944	<0.0001***	0.894	0.002**	0.019	0.003	0.112
HDL2.C	-3.992	<0.0001***	0.890	0.002**	0.018	0.003	0.106
HDL3.C	-2.222	<0.0001***	0.497	0.002**	0.108	0.041	0.287
Sphingomyelins	-1.475	0.0001***	0.383	0.027**	0.229	0.108	0.485
ApoA1	-3.013	<0.0001***	0.695	0.003**	0.049	0.013	0.192
Omega.6 %	-4.285	<0.0001***	1.001	0.004**	0.014	0.002	0.098
PUFA %	-5.124	<0.0001***	1.215	0.006**	0.006	0.001	0.064
MUFA %	1.716	<0.0001***	0.401	0.004**	5.562	2.535	12.203
SFA %	1.865	<0.0001***	0.411	0.001**	6.453	2.881	14.456
LA %	-4.726	<0.0001***	1.136	0.007**	0.009	0.001	0.082
DHA %	-1.373	0.0001***	0.360	0.031**	0.253	0.125	0.513
Gly	-1.754	0.0001***	0.448	0.021**	0.173	0.072	0.416
His	-1.667	<0.0001***	0.377	0.002**	0.189	0.090	0.395
Ile	1.484	0.0001***	0.376	0.018**	4.411	2.113	9.210
XL.VLDL.P	1.541	0.0002***	0.412	0.042**	4.671	2.082	10.477
XL.VLDL.CE	1.739	0.0001***	0.439	0.017**	5.689	2.409	13.438
L.VLDL.P	1.488	0.0002***	0.401	0.047**	4.430	2.018	9.727

L.VLDL.C	1.553	0.0001***	0.402	0.025**	4.725	2.150	10.383
L.VLDL.CE	1.682	<0.0001***	0.415	0.011**	5.377	2.385	12.122
XL.HDL.L	-1.760	<0.0001***	0.431	0.010**	0.172	0.074	0.400
XL.HDL.PL	-1.531	0.0001***	0.394	0.024**	0.216	0.100	0.469
XL.HDL.C	-2.093	<0.0001***	0.486	0.004**	0.123	0.048	0.320
XL.HDL.CE	-2.263	<0.0001***	0.513	0.002**	0.104	0.038	0.284
XL.HDL.FC	-1.780	<0.0001***	0.437	0.011**	0.169	0.072	0.397
L.HDL.L	-2.493	<0.0001***	0.571	0.003**	0.083	0.027	0.253
L.HDL.PL	-2.925	<0.0001***	0.661	0.002**	0.054	0.015	0.196
L.HDL.C	-2.294	<0.0001***	0.521	0.002**	0.101	0.036	0.280
L.HDL.CE	-2.308	<0.0001***	0.525	0.003**	0.099	0.036	0.278
L.HDL.FC	-2.202	<0.0001***	0.501	0.003**	0.111	0.041	0.295
M.HDL.P	-1.337	0.0002***	0.355	0.038**	0.263	0.131	0.527
M.HDL.L	-2.563	<0.0001***	0.594	0.004**	0.077	0.024	0.247
M.HDL.PL	-2.433	<0.0001***	0.563	0.004**	0.088	0.029	0.265
M.HDL.C	-3.007	<0.0001***	0.680	0.002**	0.049	0.013	0.187
M.HDL.CE	-2.957	<0.0001***	0.668	0.002**	0.052	0.014	0.192
M.HDL.FC	-3.250	<0.0001***	0.726	0.002**	0.039	0.009	0.161

S.HDL.L	-1.456	0.0001***	0.376	0.025**	0.233	0.111	0.487
S.HDL.FC	-1.579	0.0001***	0.394	0.014**	0.206	0.095	0.446
VLDL.size	1.485	0.0001***	0.380	0.021**	4.414	2.096	9.294
HDL.size	-1.911	<0.0001***	0.460	0.007**	0.148	0.060	0.364
XXL.VLDL.TG %	-1.699	0.0001***	0.444	0.030**	0.183	0.077	0.437
S.VLDL.FC %	-1.908	0.0001***	0.476	0.014**	0.148	0.058	0.377
XS.VLDL.C %	-2.016	<0.0001***	0.487	0.008**	0.133	0.051	0.346
XS.VLDL.CE %	-1.712	0.0001***	0.428	0.014**	0.180	0.078	0.417
XS.VLDL.TG %	1.520	0.0001***	0.384	0.017**	4.573	2.154	9.711
IDL.FC %	-2.919	<0.0001***	0.640	0.001**	0.054	0.015	0.189
M.HDL.C %	-2.283	<0.0001***	0.528	0.003**	0.102	0.036	0.287
M.HDL.FC %	-5.369	<0.0001***	1.269	0.005**	0.005	0.000	0.056
M.HDL.TG %	1.520	<0.0001***	0.366	0.007**	4.572	2.232	9.363
S.HDL.TG %	1.717	<0.0001***	0.395	0.003**	5.569	2.568	12.077

Table 5.4: Association of metabolites and disease status (HV vs NAFLD). The table shows beta coefficients, p-values, standard error (std.error), FDR values, odds ratio (OR) and upper limit (UCI) and lower limit (LCI) of confidence interval values from the regression model adjusted for age, gender and BMI. *** pval $\leq$ 0.001, ** pval < 0.05.

5.4.7 ASSOCIATION OF SERUM METABOLOME AND GUT MICROBIOME IN NAFLD AND SEVERITY SUB-GROUPS

To investigate the relationship between the gut microbes and the metabolites, spearman's correlation was done. *Collinsella* and *Bifidobacterium* showed a similar direction of associations with the significantly associated metabolites reported in this study. Mainly, significant negative associations were seen with HDL and its fractions and subfractions, unsaturation, PUFA and its ratio, sphingomyelins, ApoA1, glycine, histidine, lactate and creatinine while significant positive associations with VLDL and its fractions and sub-fractions, saturated fatty acids, MUFA and its ratio, isoleucine, tyrosine (**Table 5.5**).

	Bifidobacterium			*Collinsella*		
	Correlation Coefficient	**p-value**	**FDR**	**Correlation Coefficient**	**p-value**	**FDR**
HDL.C	-0.381	<0.001 ***	<0.001 ***	-0.366	<0.001 ***	<0.001 ***
HDL2.C	-0.326	<0.001 ***	<0.001 ***	-0.356	<0.001 ***	0.001**
HDL3.C	-0.339	<0.001 ***	<0.001 ***	-0.351	<0.001 ***	<0.001 ***
Sphingomyelins	-0.290	0.001**	0.001**	-0.292	0.001**	0.002**
ApoA1	-0.346	<0.001 ***	<0.001 ***	-0.349	<0.001 ***	<0.001 ***
Omega.6 %	-0.312	<0.001 ***	<0.001 ***	-0.325	<0.001 ***	0.001**
PUFA %	-0.337	<0.001 ***	<0.001 ***	-0.340	<0.001 ***	<0.001 ***
MUFA %	0.324	<0.001 ***	<0.001 ***	0.316	<0.001 ***	0.001**
SFA %	0.322	<0.001 ***	0.001**	0.286	0.001**	0.001**
LA %	-0.264	0.002**	<0.001 ***	-0.304	<0.001 ***	0.006**
DHA %	-0.347	<0.001 ***	<0.001 ***	-0.344	<0.001 ***	<0.001 ***
Gly	-0.296	<0.001 ***	0.002**	-0.267	0.002**	0.002**
His	-0.377	<0.001 ***	<0.001 ***	-0.368	<0.001 ***	<0.001 ***
Ile	0.252	0.003**	0.004**	0.248	0.004**	0.008**
XL.VLDL.P	0.227	0.008**	0.010**	0.222	0.010**	0.016**
XL.VLDL.CE	0.267	0.002**	0.005**	0.243	0.005**	0.005**
L.VLDL.P	0.234	0.006**	0.004**	0.248	0.004**	0.014**
L.VLDL.C	0.275	0.001**	0.004**	0.247	0.004**	0.004**
L.VLDL.CE	0.302	<0.001 ***	0.002**	0.264	0.002**	0.001**
XL.HDL.L	-0.240	0.005**	<0.001 ***	-0.320	<0.001 ***	0.012**
XL.HDL.PL	-0.230	0.007**	<0.001 ***	-0.301	<0.001 ***	0.015**
XL.HDL.C	-0.260	0.002**	<0.001 ***	-0.341	<0.001 ***	0.007**
XL.HDL.CE	-0.273	0.001**	<0.001 ***	-0.360	<0.001 ***	0.004**
XL.HDL.FC	-0.241	0.005**	<0.001 ***	-0.310	<0.001 ***	0.011**

L.HDL.L	-0.345	<0.001 ***	<0.001 ***	-0.344	<0.001 ***	<0.001 ***
L.HDL.PL	-0.359	<0.001 ***	<0.001 ***	-0.350	<0.001 ***	<0.001 ***
L.HDL.C	-0.341	<0.001 ***	<0.001 ***	-0.333	<0.001 ***	<0.001 ***
L.HDL.CE	-0.336	<0.001 ***	<0.001 ***	-0.337	<0.001 ***	<0.001 ***
L.HDL.FC	-0.351	<0.001 ***	<0.001 ***	-0.326	<0.001 ***	<0.001 ***
M.HDL.P	-0.393	<0.001 ***	0.006**	-0.237	0.006**	<0.001 ***
M.HDL.L	-0.408	<0.001 ***	<0.001 ***	-0.319	<0.001 ***	<0.001 ***
M.HDL.PL	-0.418	<0.001 ***	<0.001 ***	-0.329	<0.001 ***	<0.001 ***
M.HDL.C	-0.414	<0.001 ***	<0.001 ***	-0.328	<0.001 ***	<0.001 ***
M.HDL.CE	-0.410	<0.001 ***	<0.001 ***	-0.325	<0.001 ***	<0.001 ***
M.HDL.FC	-0.416	<0.001 ***	<0.001 ***	-0.346	<0.001 ***	<0.001 ***
S.HDL.L	-0.342	<0.001 ***	0.023**	-0.195	0.023**	<0.001 ***
S.HDL.FC	-0.399	<0.001 ***	0.005**	-0.241	0.005**	<0.001 ***
VLDL.size	0.238	0.005**	0.001**	0.289	0.001**	0.012**
HDL.size	-0.267	0.002**	<0.001 ***	-0.322	<0.001 ***	0.005**
XXL.VLDL.TG %	-0.098	0.278	0.605	-0.047	0.605	0.230
S.VLDL.FC %	-0.308	<0.001 ***	<0.001 ***	-0.320	<0.001 ***	0.001**
XS.VLDL.C %	-0.191	0.027**	0.002**	-0.263	0.002**	0.040**
XS.VLDL.CE %	-0.174	0.043**	0.004**	-0.245	0.004**	0.059**
XS.VLDL.TG %	0.178	0.039**	0.002**	0.265	0.002**	0.055**
IDL.FC %	-0.256	0.003**	<0.001 ***	-0.309	<0.001 ***	0.008**
M.HDL.C %	-0.323	<0.001 ***	0.003**	-0.251	0.003**	0.001**
M.HDL.FC %	-0.407	<0.001 ***	<0.001 ***	-0.379	<0.001 ***	<0.001 ***
M.HDL.TG %	0.377	<0.001 ***	<0.001 ***	0.312	<0.001 ***	<0.001 ***
S.HDL.TG %	0.360	<0.001 ***	<0.001 ***	0.318	<0.001 ***	<0.001 ***

Table 5.5: Associations between significant metabolites reported to be associated in this study and Bifidobacterium or Collinsella. The table shows Spearman's correlation coefficients, p-values and FDR values. *** pval $\leq$ 0.001, ** pval < 0.05.

5.4.8 NUTRIENT INTAKES DIFFER BETWEEN HEALTHY AND NAFLD PARTICIPANTS

To describe differences in the nutrient intake for differences attributed to total energy intakes, macronutrients, micronutrients and intakes as food groups were compared among healthy volunteers and NAFLD or its severity groups. The intake of nutrients and food groups was adjusted for energy intakes by calculating the residuals from a linear regression model treating total energy intake as the independent variable and absolute intake of nutrients as the dependent variable.

Overall, the intake of 11 nutrients and 5 food groups were found to be significantly different in between NAFLD and HVs. NAFLD participants showed significantly higher intakes of micronutrients such as calcium, selenium and cholesterol and lower intakes of iron, manganese and niacin. Amongst macronutrient intakes, NAFLD participants consumed higher intakes of carbohydrates specifically those derived from lactose, sucrose and total sugars. NAFLD participants also showed higher intakes of food items grouped as meat and meat products, milk and milk products and potatoes while they had lower intakes of saturated fatty acids, alcohol and alcoholic beverages, nuts and seeds compared to HVs. Descriptive analysis of absolute intakes is mentioned in the **Table 5.6** and significant nutrients are plotted as boxplots in **Figure 5.8, 5.9 and 5.10**. There was no significant difference in the intake of dietary nutrients in the severity subgroups of NAFLD patients.

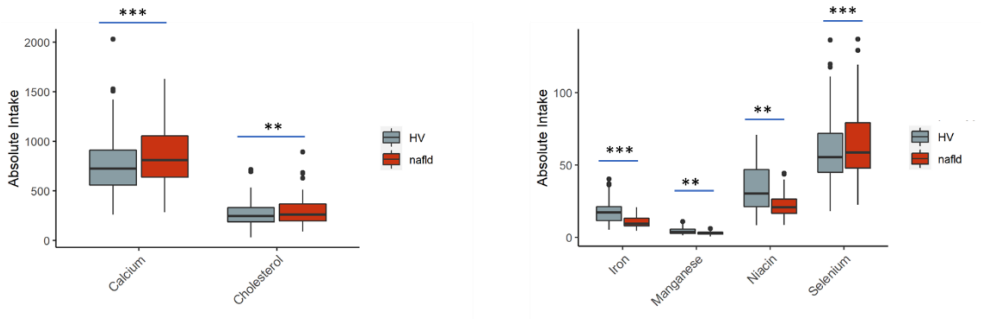


Figure 5.8: Micronutrients showing significantly different intakes. The y-axis represents the absolute intake in milligrams except for Selenium which is in micrograms. Box-and-whisker plots show high, low, and median values, with lower and upper edges of each box denoting first and third quartiles,

respectively. Asterisk indicates statistical significance from independent t-tests comparing energy adjusted inverse normalised intakes in NAFLD vs HV. Black dots represent the outliers. *** pval ≤ 0.001, ** pval < 0.05.

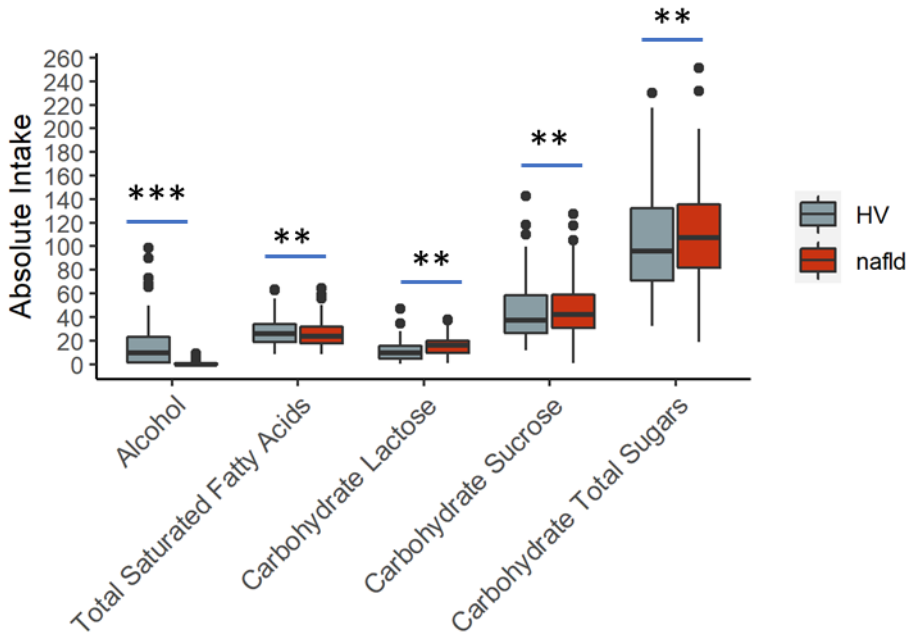


Figure 5.9: Macronutrients showing significantly different intake. The y-axis represents the absolute intake in grams. Box-and-whisker plots show high, low, and median values, with lower and upper edges of each box denoting first and third quartiles, respectively. Asterisk indicates statistical significance from independent t-tests comparing energy adjusted inverse normalised intakes in NAFLD vs HV. Black dots represent the outliers. *** p-val ≤ 0.001, ** p-val ≤ 0.05.

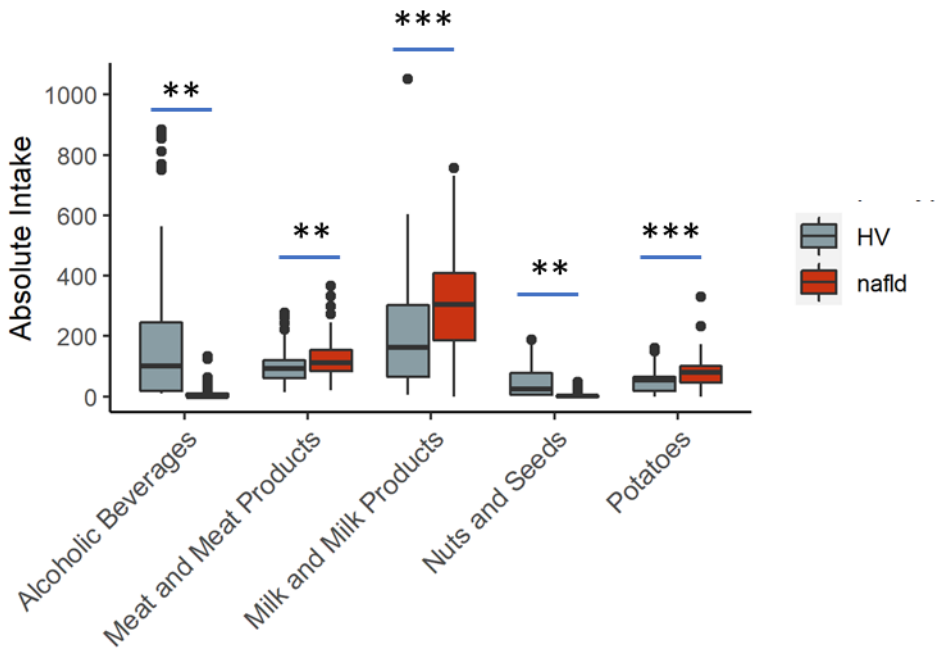


Figure 5.10: Food groups showing significantly different intakes. The y-axis represents the absolute intake in grams. Box-and-whisker plots show high, low, and median values, with lower and upper edges of each box denoting first and third quartiles, respectively. Asterisk indicates statistical significance from independent t-tests comparing intakes in NAFLD vs HV. Black dots represent the outliers. *** p-val ≤ 0.001, ** p-val ≤ 0.05.

	HV		NAFLD			
Micronutrients	**mean**	**sd**	**mean**	**sd**	**Pval**	**FDR**
Alpha carotene (mcg)	460.8	353.44	593.76	720.52	0.043**	1
Beta carotene (mcg)	3520.28	2791.77	3280.48	2439.26	0.262	1
Calcium (mg)	778.57	330.25	844.42	290.9	<0.001***	0.001***
Carotene total carotene equivalents (mcg)	3905.79	2952.43	3761.65	2790.79	0.185	1
Cholesterol (mg)	270.52	125.69	293.664	145.13	<0.001***	0.016**
Chloride (mg)	4534.39	1924.62	3990.34	1581.89	0.146	1
Copper (mg)	1.59	0.81	1.21	0.58	0.414	1
Iron (mg)	18.04	8.57	10.62	3.87	<0.001***	<0.001***
Total folate (mcg)	423.71	214.14	287.73	108.12	0.013	0.778
Iodine (mcg)	142.15	65.47	141.68	51.16	0.019	1
Potassium (mg)	3639.87	1201.83	3420.09	1007.43	0.017	0.965
Magnesium (mg)	412.24	196.23	294.88	93.19	0.003**	0.146
Manganese (mg)	4.48	2.22	2.95	1.1	0.001***	0.040**
Sodium (mg)	2984.81	1261.6	2691.49	1095.59	0.062	1
Niacin (mg)	33.84	15.52	22.05	8.27	0.000***	0.005**
Phosphorus (mg)	1484.89	540.32	1358.43	416.51	0.032**	1
Vitamin A-retinol (mcg)	602.75	713.3	696.1	917.63	0.005**	0.308
Vitamin A-retinol equivalents (mcg)	1256.9	1082.02	1333.54	1060.96	0.005**	0.31
Vitamin B2-riboflavin (mg)	2.57	1.23	1.9	0.7	0.050**	1
Selenium (mcg)	59.76	24.09	65	25.17	0.000***	<0.001***
Vitamin B1-thiamin (mg)	1.97	0.86	1.5	0.53	0.153	1
Vitamin B12-cobalamin (mcg)	7.23	4.56	7.23	4.52	0.134	1

Vitamin B6-pyridoxine (mg)	3.23	1.67	2.18	0.76	0.013**	0.728
Vitamin C-ascorbic acid (mg)	108.28	63.26	122.95	61	0.001***	0.084
Vitamin D-ergocalciferol (mcg)	4.56	3.25	3.21	1.87	0.449	1
Vitamin E-alpha tocopherol equivalents (mg)	14.84	7.69	11.2	5.11	0.644	1
Zinc (mg)	10.47	4.54	9.16	3.02	0.126	1
Macronutrients	**mean**	**sd**	**mean**	**sd**	**Pval**	**FDR**
Alcohol (g)	19.3	25.7	0.704	1.64	0.000***	0.001***
Carbohydrate total (g)	232.2	100.01	215.781	81.54	0.012**	0.685
Englyst Fibre Non-Starch Polysaccharides (g)	21.95	11.25	17.187	7.1	0.431	1
Carbohydrate fructose (g)	22.29	11.5	22.627	12.77	0.040**	1
Carbohydrate galactose (g)	0.79	1.46	0.535	0.68	0.814	1
Carbohydrate glucose (g)	20.39	10.86	20.021	10.64	0.012**	0.699
Carbohydrate lactose (g)	11.21	8.79	16.462	9.33	0.000***	0.002**
Carbohydrate maltose (g)	2.51	2.16	2.107	1.58	0.153	1
Protein (g)	88.02	30.1	82.974	27.3	0.007**	0.399
Carbohydrate starch (g)	123.48	63.56	102.124	46.29	0.932	1
Carbohydrate sucrose (g)	44.98	26.61	45.558	24.3	0.000***	0.022**
Nitrogen (g)	14.38	5.03	13.36	4.42	0.013**	0.762
Carbohydrate sugars total (g)	104.77	46.56	109.404	45.08	0.000***	0.002**
Fat total (g)	90.38	46.66	68.695	29.48	0.834	1
Monounsaturated fatty acids-MUFA total (g)	39.42	24.72	24.82	11.62	0.042**	1
Polyunsaturated fatty acids-PUFA total (g)	16.59	9.39	12.038	5.81	0.511	1
Saturated fatty acids-SFA total (g)	27.27	11.93	25.627	11.63	0.000***	0.012**

Food Groups	**mean**	**sd**	**mean**	**sd**	**Pval**	**FDR**
Alcoholic beverages (g)	222.48	286.59	12.46	30.18	0.000***	0.002**
Cereals and cereal products (g)	257.46	146.92	212.385	143.93	0.499	1
Eggs and egg dishes (g)	21.05	24.37	19.438	23.75	0.697	1
Fats and oils (g)	18.67	17.2	18.581	14.27	0.010**	0.566
Fish fish products (g)	46.95	36.5	45.551	32.32	0.391	1
Fruit (g)	217.31	154.38	285.887	228.87	0.022**	1
Meat and meat products (g)	98.9	56.97	125.529	67.54	0.000***	0.002**
Milk and milk products (g)	203.98	175.12	328.42	178.85	0.000***	0.001***
Non-alcoholic beverages (g)	808.47	414.3	840.453	486.93	0.332	1
Nuts and seeds (g)	49.98	61.47	4.424	7.42	0.000***	0.004**
Potatoes (g)	50.46	40.52	83.481	50.97	0.000***	<0.001***
Soups sauces (g)	54.25	59.22	61.03	58.25	0.006**	0.323
Sugars preserves and snacks (g)	40.78	41.03	40.368	41.53	0.087	1
Vegetables (g)	278.34	179.32	275.487	151.15	0.167	1

Table 5.6: Descriptive of Absolute Dietary Nutrients intake in Healthy and NAFLD participants. This table presents the mean of nutrients intake per day and values in parenthesis indicate the standard deviations(SD). Independent Student t-test was done for statistical significance inference. Significant values are marked with asterisk (*** p-value ≤ 0.001, ** p-value ≤ 0.05).

5.4.9 DIETARY INTAKE IS ASSOCIATED WITH DISEASE STATUS

In order to analyse if the differences observed in the intake of dietary nutrients were predictive of NAFLD, least absolute shrinkage and selection operator (LASSO) regression analysis was performed on inverse normalised energy-adjusted nutrient intakes including covariates such as age, gender and BMI. The binomial deviance (**Figure 5.11**) was computed for the test dataset using a default ten-fold cross-validation method utilizing the penalty term, lambda (λ). The regression model was further fitted using the λ = 0.06415, which is the largest λ giving the minimum binomial deviance within 1 standard error.

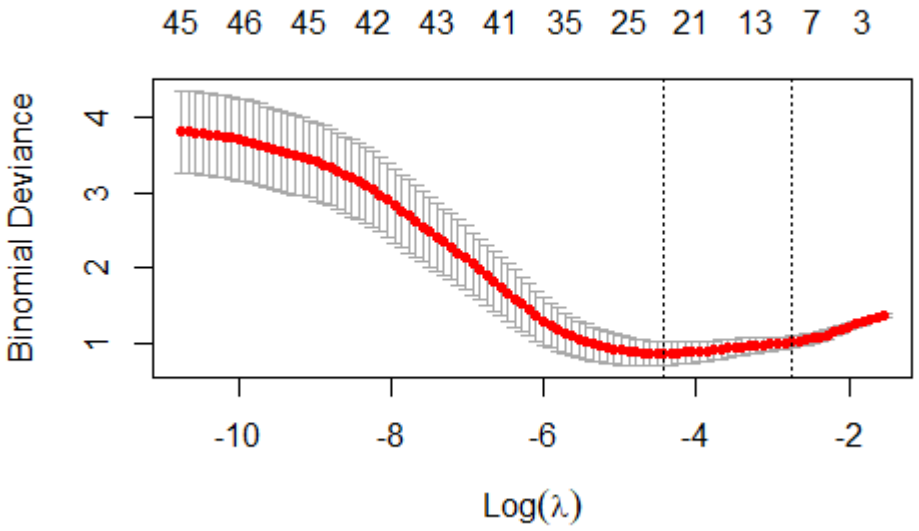


Figure 5.11: Cross-validation plot representing the lambda for the LASSO regression model. The x-axis represents the log-transformed values of the penalty term and the y-axis represents the deviation in the predicted model. The dashed vertical lines present the value of lambda at which the binomial deviance is minimum and the value of the largest lambda at which the binomial deviance is within 1 standard error.

LASSO regression showed that Calcium (β = 0.12), Iron (β = -0.51), Selenium (β = 0.30), total carbohydrates in sugars (β = 0.38), meat and meat products (β = 0.18), milk and milk products (β = 0.15) , potatoes, (β = 0.39), soups and sauces (β = 0.04) were associated with the disease status. **Table 5.7** shows the coefficients for all the energy-adjusted nutrient intakes, age, gender and BMI.

Figure 5.12 shows the ROC curve depicting the performance of the LASSO model.

Nutrients/Food groups	**LASSO coefficients**
Calcium (mg)	0.118648
Iron (mg)	-0.50654
Selenium (mcg)	0.296495
Carbohydrate sugars total (g)	0.378038
Meat and meat products (g)	0.184555
Milk and milk products (g)	0.145501
Potatoes (g)	0.39358
Soups and sauces (g)	0.04252

Table 5.7: The estimated coefficients for logistic least absolute shrinkage and selection operator (LASSO) regression between dietary data including the covariates (age, gender and BMI).

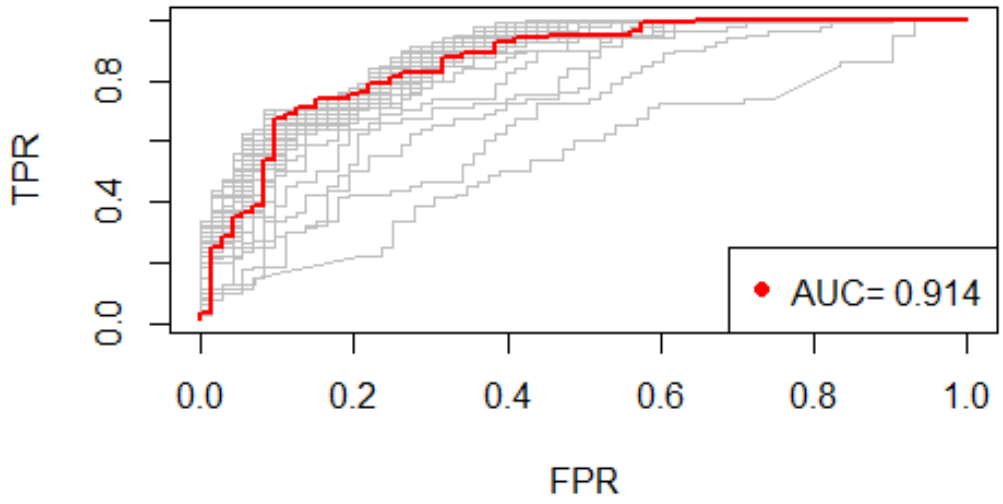
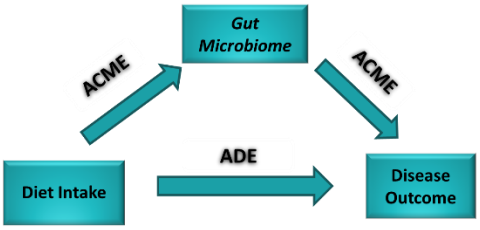


Figure 5.12: ROC curve for the LASSO regression model. The x-axis presents the sensitivity of the model in terms of false positive rate and the y -axis presents the specificity of the model in terms of true positive rate. The grey colour lines present the ROC curves for 100 random models computed for the analysis. The red colour line presents the average of the ROC curve.

5.4.10 EFFECT OF DIET ON NAFLD MEDIATED BY GUT MICROBIOME

To further investigate the causal relationship between dietary intake (nutrients and food groups) and disease outcome, mediation analysis was performed. Mediation analysis provides information on the transmittance of change from cause to its effects either directly or indirectly via mediators. *Collinsella* and *Bifidobacterium* (two most significantly associated gut microbes associated with NAFLD in this study) were treated as the mediator variables and the significant dietary nutrients and food groups found associated with disease status were treated as the predictor variable. The mediation model was adjusted for covariates such as age, gender and BMI. Mediation analysis (**Figure 5.13**) showed that the *Collinsella and Bifidobacterium* taxa acted as full mediator (ACME $p < 0.05$ and ADE $p > 0.05$) towards the positive association of total sugars in carbohydrates, milk and milk products and calcium and towards a negative association of iron with the disease status. The exception to this was selenium, the positive association of which with the disease status was found

to be acted as mediator by *Collinsella* only. The positive association of soups and sauces and potatoes with the disease outcome showed a direct effect (ACME $p > 0.05$ and ADE $p < 0.05$) i.e. not mediated by either of the microbe. The details are presented in the tables presented in the **Figure 5.13**.



ACME

	Collinsella	*Bifidobacterium*
Calcium	0.03973 (0.012**)	0.02625 (0.030**)
Iron	-0.0481 (0.002**)	-0.0273 (0.020**)
Selenium	0.04304 (0.006**)	0.01032 (0.442)
Total Sugar in Carbohydrates	0.05552 (<2e-16***)	0.03880 (0.006**)
Milk and Milk Products	0.03112 (0.036**)	0.022628 (0.054)
Soups and Sauces	0.0192 (0.228)	0.00855 (0.516)
Potatoes	0.02468 (0.156)	0.0170 (0.248)
Meat and Meat Products	0.02291(0.17)	-0.00228 (0.82)

ADE

	Collinsella	*Bifidobacterium*
Calcium	0.03296 (0.276)	0.04899 (0.158)
Iron	-0.0499 (0.100)	-0.0478 (0.130)
Selenium	0.06227 (0.068)	0.07445 (0.038**)
Total Sugar in Carbohydrates	0.05333 (0.086)	0.05584 (0.104)
Milk and Milk Products	0.02405 (0.392)	0.039463 (0.242)
Soups and Sauces	0.0811 (0.006**)	0.09142 (0.016**)
Potatoes	0.11034 (0.004**)	0.1286 (0.004**)
Meat and Meat Products	0.03397 (0.35)	0.05655 (0.14)

Figure 5.13: Causal Mediation Analysis Model: Mediation analysis between dietary intake (food groups, macronutrients and micronutrients) and disease outcome. Data in the tables is expressed as the ACME or ADE estimate scores and the p values are presented in parenthesis. Significant values are marked with asterisk (*** p-value ≤ 0.001, ** p-value ≤ 0.05).

Abbreviations: ACME: Average Causal Mediated Effect, ADE: Average Direct Effect.

5.5 DISCUSSION

In this study, amongst the fifteen differentially abundant genera, *Collinsella* was identified to be the most significant genus associated with NAFLD status. Previous studies have shown *Collinsella* to be associated with atherosclerosis which has a significant association with NAFLD **(Karlsson *et al.*, 2012; Candela *et al.*, 2016**). Surprisingly, *Bifidobacterium*, a beneficial microbe with anti-inflammatory and probiotic properties was found to be significantly increased in participants with NAFLD in our study, which is contradictory to previous report (**Putignani *et al.*, 2018**). This discrepancy may be likely due to the presence of a huge number of prophages and spreading plasmids in the genome of various *Bifidobacterium* species which makes it susceptible to horizontal gene transfer **(Chaplin *et al.*, 2015)** and that is why species belonging to *Bifidobacterium* genus act in a diversified way **(Wang *et al.*, 2021)**. To understand the link between the *Bifidobacterium* genus and NAFLD, the analysis in the current study was extrapolated to the species level. It was observed that abundance of three uncultured species was increased in NAFLD participants while abundance of *B. adolescentis* was decreased in NAFLD patients compared with HV. *B. adolescentis* has been previously found to be reduced in the mouse fed on a diet low in dietary fibre (**Kovatcheva-Datchary *et al.*, 2019)**, which aligns with the reduced intake of non-starch polysaccharide fibre by NAFLD patients in the current study. However, except for *B. adolescentis*, other species are still unidentified and needs further work of identification and characterization. Shotgun metagenomics may provide insight into the metabolic pathways impacted by the microbes.

Lipid and lipoprotein metabolism which occurs mainly in the liver is altered in NAFLD. Lipid molecules being insoluble in water are being transported to liver and other parts of body with the help of the lipoproteins. Latter is made up of the core composed of the cholesterol esters and triglycerides and the periphery made up of free cholesterols, phospholipids and apolipoproteins. High-Density Lipoprotein (HDL) is found to be reduced (**Hae *et al.*, 2004**) and very low-density lipoproteins (VLDL) is increased in NAFLD (**Adiels, Taskinen and Borén, 2008**). Having considered the role of these metabolites in NAFLD, the results of serum lipid profiling in this study, showed significant difference in the

lipoprotein subclasses of HDL and VLDL among HV and NAFLD participants. The decrease in the fractions and subfractions of HDL while an increase in the VLDL fractions and subfractions in NAFLD patients reflects an increased formation of VLDL carrying lipoprotein subclasses while decrease in HDL carrying lipoproteins. This corroborates to impaired lipid adsorption and transportation from the liver and eventually more accumulation of VLDL in the liver. Serum ApoA1, a structural component of HDL and a co-factor for enzyme plays a pivotal role in HDL transportation and metabolism. This is transferred to HDL during chylomicrons metabolism, thus also plays a role in the *de novo* biogenesis of HDL via activation of the lecithin cholesterol acyl transferase (LCAT) enzyme (**Karavia *et al.*, 2012**). The current study observed a decrease in ApoA1 in NAFLD patients which demonstrates the increased risk of NAFLD similar to a study by Yang et al. (**Yang, Sung and Gwak, 2016**). A very little is currently known about the role of saturation of fatty acids in NAFLD and precisely the influence of serum levels of PUFAs (polyunsaturated fatty acids), MUFAs (monounsaturated fatty acids) and SFAs (saturated fatty acids) in NAFLD remains unclear. In current study, higher OR associated with increased ratio of MUFA to total fatty acids (FA) ratio and ratio of SFA to FA demonstrates they increase the risk of NAFLD. A lower OR associated with reduced ratio of PUFA to FA, demonstrates its protective effect in NAFLD. One possible explanation for this could be that PUFAs impedes the transfer the cholesterol esters from HDL to LDL (Low Density lipoproteins) and VLDL aiding the increase in the HDL concentrations and decreased concentrations of PUFA in the current study describes the formation of less HDL and more VLDL. Furthermore, amino acids such as glycine reportedly lowers insulin resistance and abdominal fat (**Rom *et al.*, 2018**) while isoleucine, a branched-chain amino acid, on the other hand impairs oxidative fat metabolism **(Du *et al.*, 2011; Yu *et al.*, 2021**). The decrease in glycine and increase in isoleucine, in NAFLD participants in this study, reports the close association of these amino acids with lipid metabolism based on previous studies. Furthermore, the higher odds ratio of isoleucine in the present study further supports the harmful impact of the same in NAFLD.

With respect to the aim of investigating the link between dietary nutrient intakes and NAFLD, our study describes that diet consumed by NAFLD participants is mainly rich in potatoes and total sugars from carbohydrates sources. Potatoes are primarily composed of starchy carbohydrates and have a high glycaemic index. The starch in potatoes is converted into glucose which raises the blood glucose rapidly and the excess is stored in the liver similar to the excess of total sugars from the other forms of carbohydrates. A systematic review analysis of dietary intake and NAFLD confirms the evidence of an increase in the intake of potatoes to be associated with higher odds of NAFLD (**Mirmiran *et al.*, 2017**). Another pilot study conducted in NAFLD patients fed on a low carbohydrate diet i.e. restricted starch and glucose sugar diet showed improvements in the fatty liver by the reduction in serum gamma-glutamyl transferase (GGT), a liver enzyme and HbA1c (**Unwin *et al.*, 2015**). The increased intake of potatoes and total sugars from carbohydrates in the current study explains risk and susceptibility of NAFLD as the excess sugar gets converted to fat via *de novo* lipogenesis in the liver which is a direct cause of fatty liver (**Glimcher and Lee, 2009**).

This study showed NAFLD participants having a diet that is mainly deficient in significant amount of micronutrients such as iron, while rich in micronutrient selenium. The decreased intake of iron in participants with NAFLD in our study accords with the finding of hypoferremia in obese patients **(Yanoff *et al.*, 2007)**, which is due to a combination of dietary iron deficiency, inadequate bioavailable iron relative to body weight, as well as diminished intestinal absorption and decreased iron bioavailability induced by inflammatory adipokines found in those with excessive adiposity. Altered iron homeostasis apparently increases the oxidative stress in the liver **(Bresgen and Eckl, 2015; Galaris, Barbouti and Pantopoulos, 2019**) and results in liver fibrosis progressing to cirrhosis (**Vela, 2018**). In an epidemiological study with ~37,000 participants, serum iron was found to be significantly decreased in NAFLD participants in both genders and in females with advanced hepatic fibrosis (**Yang *et al.*, 2021**). Iron status influences NAFLD in obesity through gut microbiome. Different dosages of iron supplementation independent from different dosages of iron with or without high fat diet showed that bacterial diversity decreases under a low iron diet in

comparison to high iron content (**Mayneris-Perxachs and Swann, 2019**). Selenium, an essential trace element is mainly obtained from meat, fish, eggs, dairy products, bread and cereals (**Fairweather-Tait *et al.*, 2011**). The observed increased intake of selenium in this study corresponds with a significant increase in the intakes of meat and milk products by NAFLD patients. Various cross sectional study reports positive associations between levels of selenium in serum and serum ALT or NAFLD (**Yang *et al.*, 2016; Wang, Seo and Park, 2021**). Also, high selenium exposure including high selenium intake has been associated with an increased risk of diabetes, hyperlipidemia (**Stranges *et al.*, 2010**) and NAFLD (**Wu *et al.*, 2020**). The role of selenium in NAFLD patients has been investigated by dietary supplementation studies primarily in animal models (**Xu *et al.*, 2022**). Supplementation of selenium in combination with zinc reduced lipid peroxidation and reduction in liver fat (**Mousavi *et al.*, 2018; Shidfar *et al.*, 2018**) in experimental models of NAFLD. This inconsistency may be due to the difference in the metabolism of different forms of the selenium (**Burk *et al.*, 2006**) and also the discordance between the levels of selenium and SELENOP, a glycoprotein synthesized by the liver cells which helps in the transport of selenium to other cells of the body (**Polyzos *et al.*, 2020**). Selenium exerts its effects mainly through its incorporation into selenoproteins as the amino acid, selenocysteine (**Reeves and Hoffmann, 2009**) which are involved in redox homeostasis (**Stranges *et al.*, 2010**). In addition, dose of selenium supplementation also relies upon the amount sequestered by the gut microbiome **(Kasaikina *et al.*, 2011**). As a result, amount of trace element available for incorporation into host selenoprotein can be limited. This may increase the requirement of selenium for the host. Further research is needed to understand the mechanisms of selenium mediated insights of how these dietary nutrients are metabolised and becomes available to the host.

Formal mediation analysis in this study demonstrates that adherence to selenium rich diet is a risk factor for NAFLD and an increase in the abundance of *Collinsella* act as the potential mediator towards the positive association between the two. Adherence to a diet deficient in iron also increases the likelihood of NAFLD and both *Collinsella* and *Bifidobacterium* act as the potential mediators . Similar findings are reported from recent studies conducted in animal models,

where iron homeostasis is observed to be maintained by the gut microbiome (**Das *et al.*, 2020**). Investigation of selenium nanoparticles in heritage roosters showed a significant reduction in the abundance of pathogenic bacteria in the caecum (**Gangadoo *et al.*, 2019**). However, these results need *in-vivo* validation, as there are trillions of microbes in the gut environment that may mediate the functional capacity to utilise these dietary nutrients independently of the disease outcome.

To the best of our knowledge, the current study is the first to report significant associations between dietary nutrients, microbiome and metabolome with NAFLD and its severity subgroups in a single large cohort. The FFQs used to measure the dietary intakes for the study have been extensively validated in large group of people. Analysis of average micronutrient intakes based on the previous one-year dietary records in NAFLD participants is the first of its kind. There are certain limitations involved in the present study. First, the measured intakes of nutrients could be biased due to the recall reporting of the dietary items over the period of a year. The data from healthy participants used as controls were not screened for the presence or absence of NAFLD as there was no fibro scan procedure or other diagnostic tests done for these participants. The diet assessment tool, FETA, does not calculate the intake of Vitamin D and K which has been found to be associated with NAFLD and obesity, a risk factor of NAFLD (**Shea *et al.*, 2010**). The histological data to assess the severity of the disease (grade of inflammation and stages of fibrosis) was available for a proportion of participants because of which parts of the analysis could not reach sufficient statistical power. It is a reason that this study could not predict the significant differences between the severity sub-groups after correcting for multiple testing.

To sum up, the gut microbiota *Collinsella and Bifidobacterium* act as the potential mediators towards the effect of dietary sugars, as well as micronutrients such as iron to increase the risk of NAFLD. In addition, increase in dietary selenium associated with *Collinsella* resulted in the development of NAFLD. Subsequent to this, *Collinsella* and *Bifidobacterium* were also associated with key metabolites associated with NAFLD, positively with those increasing the risk i.e. VLDL and its sub-fractions, and isoleucine, as well as

negatively with those reducing the risk such as HDL, sphingomyelins, ApoA1, glycine and histidine. The findings of this study assist in the development of dietary interventions for the treatment of NAFLD as well as potential faecal microbiota-based signatures to assess efficacy of effective interventions.

6

Chapter 6
DISCUSSION AND CONCLUSION

6.1 OVERVIEW

"Let food be thy medicine, let medicine be thy food"

- Hippocrates

This thesis sought to examine different diseases understanding the impact of the gut microbiome on human health and its interplay with metabolome and dietary factors. The work presented in this thesis centred around analysing the data from high throughput *omics* technologies such as 16S rRNA gut microbiome sequencing and serum metabolome and the 24-hour dietary recall or average intake of recalls from the previous year. The research includes undernourishment in adults, knee pain in OA, and NAFLD.

The objectives of this research were primarily three-fold:

1. To investigate the gut microbiome of participants with the aforementioned pathophysiological conditions or diseases and access the gut microbial signatures compared with healthy volunteers or replicate cohorts.
2. To identify both microbial and host metabolites or metabolic pathways associated with individual disease states compared with healthy volunteers or replicate cohorts.
3. To investigate the effect of location of residence, 24-hour dietary recalls and long-term dietary recalls as extrinsic factors and to determine the proportion of contribution mediated by the gut microbiome.

The key messages from the current work are:

1. The gut microbiome and metabolic profile and/or pathways are significantly different based on geographical location and disease states. Also, different gut microbial phyla are associated with different diseases.
2. The gut microbiome contributes to metabolism via associations with certain metabolites impacting regulation of different metabolic pathways.

3. The effect of a long-term dietary factors on the disease state is mediated via changes in the gut microbiome.

6.2 THE GUT MICROBIOME MODULATES METABOLIC HEALTH

The analysis of three different diseases extended a comprehensive understanding of the metabolites and/or metabolic pathways demonstrating the crosstalk between the microbiome and host factors. The following sections present the interpretation of findings from the different studies conducted in this thesis which is also schematically presented in **Figure 6.1.**

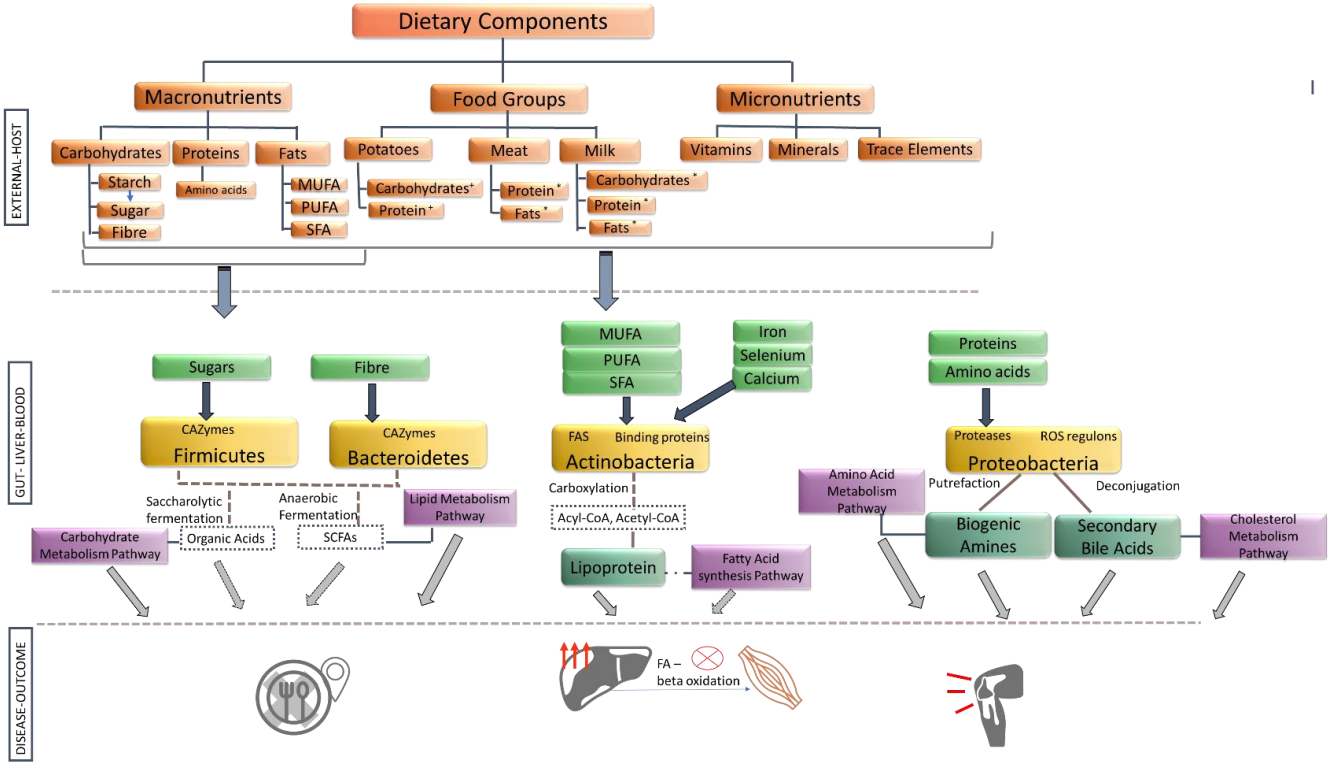


Figure 6.1: Schematic representation of conceptual framework, key findings related to dietary factors, microbiome and as well as metabolome and interpretations from the studies conducted in this thesis. Red colour boxes represent complex dietary components, green coloured boxes represent simpler form of dietary components, yellow boxes represent bacterial phyla, bright green colour boxes represent the metabolites, purple coloured boxes represent the metabolic pathways. Boxes and arrows with dashed outline represent the metabolites or pathway not analysed for the particular study but suggest the mechanistic insights towards the disease/condition.

6.2.1 DIFFERENT GUT MICROBIAL PHYLUM AFFECTS METABOLISM VIA DIFFERENT METABOLIC AND MOLECULAR PATHWAYS

The gut microbes harvest their energy from the type of foods being consumed, fermenting and converting to simpler forms that can be metabolised and utilised by the host. The four dominant phyla in the gut microbiome are Firmicutes, Bacteroidetes, Proteobacteria and Actinobacteria. With respect to the microbial taxa associated with metabolism in different diseases, this research work infers that:

1. Members of Firmicutes and Bacteroidetes regulate carbohydrate and lipid metabolism based on different type of dietary intake across geographic locations (Chapter 3): Most of the taxa belonging to these phyla break large and complex dietary compounds into smaller forms producing organic acids and short-chain fatty acids (SCFAs) (**Flint, Scott, Duncan, *et al.*, 2012**). These organic acids and SCFAs act as a substrate for further respiration by saccharolytic and anaerobic fermentation in cells regulating mainly the carbohydrate and lipid metabolism.
2. Members of Proteobacteria phyla regulate secondary bile acids production and amino-acids metabolic pathway and also favours growth of Firmicutes phyla in knee pain in OA (Chapter 4): Most members of Proteobacteria maintain the redox potential of the gut by de-hydroxylation or deconjugation of the amino-acids moiety of the bile acids, converting these to secondary bile acids (**Moon et al., 2018**) Also, sulphur reduction and denitrification via proteases and peptidase enzymes regulates the amino acid metabolic pathways (**Z. Zhou *et al.*, 2020**). The significant associations of Proteobacteria members and marginally significant associations of members belonging to Firmicutes phyla suggests that Proteobacteria may not directly promote gut health but reportedly favours growth of SCFA producing bacteria (Firmicutes).
3. Members of Actinobacteria phyla regulate fatty acid biosynthesis and micronutrients bioavailability in NAFLD (Chapter 5): Taxa belonging to Actinobacteria associated with the fractions and subfractions of cholesterols and lipids, are involved in fatty acid synthesis due to the presence of the most diverse FAS (fatty acid synthase) **(Gago *et al.*,**

2011). Further carboxylation process converts these fatty acids into Acyl-CoA signalling the formation of lipoproteins and its subclasses. Various micronutrients required by bacteria for growth and cell maintenance are utilized via permeases and binding proteins producing a variety of secondary metabolites. However, there is a limited research about the functions of the members of Actinobacteria phylum isolated from the human gut.

6.2.2 LIVER METABOLISM IS DETERMINED BY LONG-TERM NUTRIENTS INTAKE

The liver is the place for the metabolism of dietary macronutrients and their conversion to simpler forms of metabolites. It stores and also helps in the absorption and transportation of micronutrients from the diet. Findings from analysis of dietary factors in undernourishment and NAFLD studies suggests that:

1. As there was no significant difference in the carbohydrate and protein metabolism pathways regardless of the significant difference in the total caloric, carbohydrates and proteins intakes among the nourished and undernourished, therefore, the 24-hour dietary intake may not show the effect on the liver metabolic health (Chapter 3).
2. As significant differences were also observed in the average intakes of micronutrients along with the macronutrients and food groups, therefore, micronutrients intake and not just calories and macronutrients intake, determine the fat storage in liver cells and thus marked changes on liver adiposity in NAFLD participants (Chapter 5).

6.2.3 GUT MICROBES MEDIATE THE EFFECT OF DIETARY INTAKE ON LIVER METABOLISM

The quantity and quality of dietary intakes alter the gut microbiome, bringing change in the metabolic capacity of the liver. Differential abundance of gut microbes, the difference in the dietary intakes in the disease and healthy states and their association with NAFLD indicate that the modulatory effects of diet on disease are mediated through the microbial community (Chapter 5).

6.3 STATISTICAL CHALLENGES AND APPROACHES

Next-generation microbiome sequencing, comprehensive microbial databases, advanced bioinformatics, machine learning-based advanced modelling and increased computational power have enhanced the capabilities of omics-based research.

The advanced technologies produce high dimensional data where the number of microbial taxa or metabolites outnumber the number of samples. Alongside getting a huge amount of data at cost-effective rates, the heterogeneity of data has become a real issue. The broad spectrum of statistical challenges, such as data processing, integration of data from multiple platforms, data normalisation, cross-study comparisons, multiple testing correction methods and energy adjustment methods for diet data is the root cause of the non-reproducibility of data. The data analysis pipeline completely depends upon the number of samples, sparsity of data, outliers and collinearity between the variables.

Another big challenge in clinical studies is the missing data which adds bias to the results. Observations with missing data cannot always be ignored as this will reduce the statistical power of the study. So, the missing variables need to be handled based on the type of missing data. If the data is missing at random for a clinical trait or anthropometric detail i.e. does not follow a pattern, then these could be ignored completely as these don't affect the outcome directly. If the missing variable is in the data captured from HT platforms, the missing variables could be imputed with various methods such as mean, median, log transformation, total sum scaling (TSS), centred sum scaling (CSS), and half of the minimum positive values. Missing variables from the diet data can be zero-imputed indicating consuming the dietary item as never/rarely.

In this thesis, appropriate statistical methods were chosen to take these factors into account. Data normalisation was done by inverse normalising the data, missing data were imputed with half of the minimum positive values excluding zero, and collinearity of the variables was handled utilizing the LASSO regression. Correction for multiple testing was done utilizing the FDR correction

methods such as q-value and Bonferroni depending upon the distribution of the p-values. Observations with more than 10 missing variables were excluded from the study for the diet analysis, missing frequencies were imputed with code 1 representing the consumption of the item as never or rarely and the missing variables for the type of dietary item were imputed with the default codes which represents the mix composition of the types of specific dietary item. Energy adjustment was done by calculating the residuals from the linear models regressing the nutrients intake on the energy intakes.

6.4 LIMITATIONS

The scope of this thesis is limited in terms of observational studies which serves to generate hypothesis for the future studies rather than exploring the mechanisms. The mechanistic insight is an important aspect to understand the pathophysiology of the disease state/conditions. Also, as observational studies are subject to biases from covariates and selection or technical measures, the work presented in this thesis needs to be replicated in cohorts with a larger sample size. For a depth of understanding the contribution of this factor inferring the causality, studies based on longitudinal designs needs to be implemented. Currently, there is also limited research on the mobile genetic elements of most of the microbes in the gut microbiome, the production time of the metabolites by the gut microbes and parameters to investigate the inflow of these metabolites. An additional uncontrolled factor is the core and fluctuating microbiome of an individual which gives a representation of the stable and different metabolic pools. Determining these factors is crucial to understand the other aspects of mechanistic insights into the gut microbiome and its contribution to overall health.

6.5 FUTURE ASPECTS

The 16S rRNA sequencing analysis analyses only one gene of the bacterial genome and so, the overall functional potential of the microbes can’t be predicted. The falling cost of sequencing technologies and advanced machine learning-based statistical approaches enable more sensitive methods to characterize microbes beyond the abundances and compositions. The frame of microbiome research has shifted from purely dysbiosis analysis to microbial

compositions and now towards a deeper analysis of microbiome such as shotgun metagenomics. The shotgun metagenomics **(Quince *et al.*, 2017**) approach allows to sequence and analyse the entire genomic content of a sample comparing it to reference databases or de-novo assembly. The requirement for a greater depth of sequencing while sequencing a greater number of samples simultaneously, however, makes this approach implausible for large cohort studies. When the advanced technology becomes less expensive large-scale Metagenomic Wide Associated Studies (MWAS) will pave a way to advance knowledge about the mechanistic insights of the gut microbiome (**Petrosino, 2018**). A greater focus on single cell microbe analysis could access the functions of microbes depending upon their environment. This could account for the mechanism of horizontal gene transfer and could shed light on determining these as symbiotic or mutualistic or pathogenic microbe **(Garud and Pollard, 2020**).

Metabolic profiling present one aspect of changes in the system biology and thus the impact on health. The other aspects such as the transcriptional activity of the microbes are not described by such studies. Transcriptomic studies are required to back up the specific answers to the mechanisms by which the microbes contribute as the role-players in the disease outcomes. Limited research to date has investigated how many genes contained in the microbiome are coded and translated to proteins and how many are metabolically inactive. In this respect, microbiome studies should be complemented with RNA sequencing, protein profiling and transcriptome profiling. Investigating the human microbiome integrated with metabolomics, meta-proteomics and meta-transcriptomics, has now been the target for human health- based studies. Such studies may provide knowledge on disease pathogenesis along with the information on which bacteria may drive or protect against the disease enabling targeted probiotic administration, antibiotic treatment, or nutritional interventions to alter the gastrointestinal microbiome and metabolome. However, there is a lot of work to be done in this area such as 1. there is huge data that does not have a proper match in the databases, so work needs to be done to curate databases, 2. There is scarce information about the elicit change in the metabolic profiles due to the extent of changes in the nutrient content of

the diet, so further longitudinal studies accounting these research aims needs to be performed elucidating causation, 3. the reported side-effects of FMT, mainly if the transplant sample has drug resistance bacteria, and ineffectiveness to long-term and directly unrelated metabolic disorders, so studies focussing on the individual strains and single cell microbe should be targeted as potential therapeutics, 4. Metabolic models informing the mechanistic predictions by inferring the response of individual microbial species to changing environmental or dietary conditions and microbial interactions needs to be modelled,

Advanced statistical approaches such as development of curated resource databases, development of algorithms with integrated analysis of transcriptomes and metagenomes, high-resolution mapping of data with these databases etc. will be required to add stringency to the analytical models that might be able to pull out the stronger relationships between the variables from different omics platforms.

The gut microbial taxa are known to utilise the dietary components for their own growth, digesting and converting these to a series of metabolites which are absorbed from the intestine into the circulation. These metabolites serve as substrates in various microbial metabolic pathways. There is a scarce research about the metabolic capabilities of the microbial genes. Considerably more experimental research needs to be done to investigate specific diet and microbial enzymes responsible for metabolite generation.

In this respect, the bacterial taxa identified from the large interventional study cohorts and investigating further into the genomic repertoire and translational activities of these taxa could prove to be more effective as complementary therapeutics. These taxa after a further research and in vitro or in vivo experimental validation studies could be supplemented in the form of probiotics or postbiotics or synbiotics.

6.6 RESEARCH IMPLICATIONS AND CONCLUSION

This thesis demonstrates that there is a direct or indirect interaction between the gut microbiome, metabolites and diet. The integrated multi-omics phenotyping of diseases provided a holistic perspective towards the understanding of the molecular changes and inter-relationships at the level of gut health and overall metabolic health. In the current situation, investigating the omics features are only complementary rather than replacement for the prediction of the disease outcome or therapeutics. These further require validation studies in *in-vitro* or *in-vivo* models. The work presented in this thesis may enable the development of intervention strategies, and various methods of provisioning integration of data analysed from multi-omics.

REFERENCES

Abdelmalek, M. F. *et al.* (2010) 'Increased fructose consumption is associated with fibrosis severity in patients with nonalcoholic fatty liver disease', *Hepatology*. Hepatology, 51(6), pp. 1961–1971. doi: 10.1002/hep.23535.

Abeysekera, K. W. M. *et al.* (2020) 'Prevalence of steatosis and fibrosis in young adults in the UK: a population-based study', *The Lancet Gastroenterology and Hepatology*. Elsevier Ltd, 5(3), pp. 295–305. doi: 10.1016/S2468-1253(19)30419-4.

Adams, S. B. *et al.* (2012) 'Global metabolic profiling of human osteoarthritic synovium', *Osteoarthritis and Cartilage*. W.B. Saunders, 20(1), pp. 64–67. doi: 10.1016/j.joca.2011.10.010.

Adiels, M., Taskinen, M. R. and Borén, J. (2008) 'Fatty liver, insulin resistance, and dyslipidemia', *Current diabetes reports*. Curr Diab Rep, 8(1), pp. 60–64. doi: 10.1007/S11892-008-0011-4.

Agans, R. *et al.* (2011) 'Distal gut microbiota of adolescent children is different from that of adults', *FEMS Microbiology Ecology*. NIH Public Access, 77(2), pp. 404–412. doi: 10.1111/j.1574-6941.2011.01120.x.

Agus, A., Clément, K. and Sokol, H. (2021) 'Gut microbiota-derived metabolites as central regulators in metabolic disorders', *Gut*. BMJ Publishing Group, 70(6), pp. 1174–1182. doi: 10.1136/GUTJNL-2020-323071.

Aittokallio, T. and Schwikowski, B. (2006) 'Graph-based methods for analysing networks in cell biology', *Briefings in Bioinformatics*. Oxford Academic, pp. 243–255. doi: 10.1093/bib/bbl022.

Ajmera, V. *et al.* (2017) 'Novel plasma biomarkers associated with liver disease severity in adults with nonalcoholic fatty liver disease', *Hepatology (Baltimore, Md.)*. Hepatology, 65(1), pp. 65–77. doi: 10.1002/HEP.28776.

Alhamoruni, A. *et al.* (2010) 'Pharmacological effects of cannabinoids on the Caco-2 cell culture model of intestinal permeability', *Journal of Pharmacology and Experimental Therapeutics*. American Society for Pharmacology and Experimental Therapeutics, 335(1), pp. 92–102. doi: 10.1124/jpet.110.168237.

Allen, J. M. *et al.* (2018) 'Exercise Alters Gut Microbiota Composition and Function in Lean and Obese Humans', *Medicine and Science in Sports and Exercise*. Lippincott Williams and Wilkins, 50(4), pp. 747–757. doi: 10.1249/MSS.0000000000001495.

Anders, S. and Huber, W. (2010) 'Differential expression analysis for sequence count data', *Genome Biology*. Genome Biol, 11(10). doi: 10.1186/gb-2010-11-10-r106.

Anhê, F. F. *et al.* (2015) 'A polyphenol-rich cranberry extract protects from diet-induced obesity, insulin resistance and intestinal inflammation in association

with increased Akkermansia spp. population in the gut microbiota of mice', *Gut*. BMJ Publishing Group, 64(6), pp. 872–883. doi: 10.1136/gutjnl-2014-307142.

Aron-Wisnewsky, J. *et al.* (2020) 'Gut microbiota and human NAFLD: disentangling microbial signatures from metabolic disorders', *Nature Reviews Gastroenterology & Hepatology 2020 17:5*. Nature Publishing Group, 17(5), pp. 279–297. doi: 10.1038/s41575-020-0269-9.

Aßhauer, K. P. *et al.* (2015) 'Tax4Fun: Predicting functional profiles from metagenomic 16S rRNA data', *Bioinformatics*, 31(17), pp. 2882–2884. doi: 10.1093/bioinformatics/btv287.

Ayeni, F. A. *et al.* (2018) 'Infant and Adult Gut Microbiome and Metabolome in Rural Bassa and Urban Settlers from Nigeria', *Cell Reports*. Elsevier B.V., 23(10), pp. 3056–3067. doi: 10.1016/j.celrep.2018.05.018.

Ayral, X. *et al.* (2005) 'Synovitis: a potential predictive factor of structural progression of medial tibiofemoral knee osteoarthritis – results of a 1 year longitudinal arthroscopic study in 422 patients', *Osteoarthritis and Cartilage*. Elsevier, 13(5), pp. 361–367. doi: 10.1016/J.JOCA.2005.01.005.

Aziz, N., Kallur, S. D. and Nirmalan, P. (2014) 'Implicationsof the revised consensus body mass indices for Asian Indians on clinical obstetric practice', *Journal of Clinical and Diagnostic Research*. Journal of Clinical and Diagnostic Research, 8(5), p. OC01. doi: 10.7860/JCDR/2014/8062.4212.

Baba, C. S. *et al.* (2006) 'Effect of exercise and dietary modification on serum aminotransferase levels in patients with nonalcoholic steatohepatitis', *Journal of Gastroenterology and Hepatology (Australia)*. Blackwell Publishing, 21(1 PART1), pp. 191–198. doi: 10.1111/j.1440-1746.2005.04233.x.

Bäckhed, F. *et al.* (2004) 'The gut microbiota as an environmental factor that regulates fat storage', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 101(44), pp. 15718–15723. doi: 10.1073/pnas.0407076101.

Bäckhed, F. *et al.* (2015) 'Dynamics and stabilization of the human gut microbiome during the first year of life', *Cell Host and Microbe*. Cell Press, 17(5), pp. 690–703. doi: 10.1016/j.chom.2015.04.004.

Barton, W. *et al.* (2018) 'The microbiome of professional athletes differs from that of more sedentary subjects in composition and particularly at the functional metabolic level', *Gut*. BMJ Publishing Group, 67(4), pp. 625–633. doi: 10.1136/gutjnl-2016-313627.

Basaranoglu, M., Basaranoglu, G. and Bugianesi, E. (2015) 'Carbohydrate intake and nonalcoholic fatty liver disease: fructose as a weapon of mass destruction.', *Hepatobiliary surgery and nutrition*. AME Publications, 4(2), pp. 109–16. doi: 10.3978/j.issn.2304-3881.2014.11.05.

Baümler, A. J. and Sperandio, V. (2016) 'Interactions between the microbiota

and pathogenic bacteria in the gut', *Nature*, 535(7610), pp. 85–93. doi: 10.1038/nature18849.

Beasley, T. M., Erickson, S. and Allison, D. B. (2009) 'Rank-based inverse normal transformations are increasingly used, but are they merited?', *Behavior Genetics*. Behav Genet, 39(5), pp. 580–595. doi: 10.1007/s10519-009-9281-0.

Bedossa, P. *et al.* (2012) 'Histopathological algorithm and scoring system for evaluation of liver lesions in morbidly obese patients', *Hepatology*. John Wiley & Sons, Ltd, 56(5), pp. 1751–1759. doi: 10.1002/HEP.25889.

Belkaid, Y. and T. H. (2015) 'Role of the Microbiota in Immunity and inflammation', *Cell*, 157(1), pp. 121–141. doi: 10.1016/j.cell.2014.03.011.Role.

Benson, A. K. *et al.* (2010) 'Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 107(44), pp. 18933–18938. doi: 10.1073/pnas.1007028107.

Berná, G. and Romero-Gomez, M. (2020) 'The role of nutrition in non-alcoholic fatty liver disease: Pathophysiology and management', *Liver International*. John Wiley & Sons, Ltd, 40(S1), pp. 102–108. doi: 10.1111/LIV.14360.

Berthelot, J. M. *et al.* (2019) 'Cartilage-gut-microbiome axis: A new paradigm for novel therapeutic opportunities in osteoarthritis', *RMD Open*, 5(2), pp. 1–7. doi: 10.1136/rmdopen-2019-001037.

Bhute, S. *et al.* (2016) 'Molecular characterization and meta-analysis of gut microbial communities illustrate enrichment of prevotella and megasphaera in Indian subjects', *Frontiers in Microbiology*. Frontiers Media S.A., 7(MAY), p. 660. doi: 10.3389/fmicb.2016.00660.

Bind, M. A. *et al.* (2014) 'Effects of temperature and relative humidity on DNA methylation', *Epidemiology*. Lippincott Williams and Wilkins, 25(4), pp. 561–569. doi: 10.1097/EDE.0000000000000120.

Bingham, S. A. *et al.* (2001) 'Nutritional methods in the European Prospective Investigation of Cancer in Norfolk', *Public health nutrition*. Public Health Nutr, 4(3), pp. 847–858. doi: 10.1079/PHN2000102.

Blanton, L. V. *et al.* (2016) 'Gut bacteria that prevent growth impairments transmitted by microbiota from malnourished children', *Science*. American Association for the Advancement of Science, 351(6275). doi: 10.1126/science.aad3311.

Boer, C. G. *et al.* (2019a) 'Intestinal microbiome composition and its relation to joint pain and inflammation', *Nature Communications*. Springer US, 10(1), pp. 1–9. doi: 10.1038/s41467-019-12873-4.

Boer, C. G. *et al.* (2019b) 'Intestinal microbiome composition and its relation to joint pain and inflammation', *Nature Communications*. Nature Publishing Group, 10(1), pp. 1–9. doi: 10.1038/s41467-019-12873-4.

Bokulich, N. A. *et al.* (2018) 'Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin', *Microbiome*. BioMed Central Ltd., 6(1), p. 90. doi: 10.1186/s40168-018-0470-z.

Bolyen, E. *et al.* (2018) 'QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science'. PeerJ Inc. doi: 10.7287/peerj.preprints.27295v2.

Bondeson, J. *et al.* (2006) 'The role of synovial macrophages and macrophage-produced cytokines in driving aggrecanases, matrix metalloproteinases, and other destructive and inflammatory responses in osteoarthritis', *Arthritis Research and Therapy*. Arthritis Res Ther, 8(6), p. R187. doi: 10.1186/ar2099.

Bourke, C. D., Berkley, J. A. and Prendergast, A. J. (2016) 'Immune Dysfunction as a Cause and Consequence of Malnutrition', *Trends in Immunology*. Elsevier Ltd, 37(6), pp. 386–398. doi: 10.1016/j.it.2016.04.003.

Brandon-Mong, G. J. *et al.* (2020) 'A network approach to investigating the key microbes and stability of gut microbial communities in a mouse neuropathic pain model', *BMC Microbiology*. BioMed Central Ltd, 20(1), p. 295. doi: 10.1186/s12866-020-01981-7.

Breitbart, M. *et al.* (2003) 'Metagenomic analyses of an uncultured viral community from human feces', *Journal of Bacteriology*. American Society for Microbiology Journals, 185(20), pp. 6220–6223. doi: 10.1128/JB.185.20.6220-6223.2003.

Bresgen, N. and Eckl, P. M. (2015) 'Oxidative Stress and the Homeodynamics of Iron Metabolism', *Biomolecules*. Multidisciplinary Digital Publishing Institute (MDPI), 5(2), p. 808. doi: 10.3390/BIOM5020808.

Brooks, A. W. *et al.* (2018) 'Gut microbiota diversity across ethnicities in the United States', *PLOS Biology*. Edited by K. Cadwell. Public Library of Science, 16(12), p. e2006842. doi: 10.1371/journal.pbio.2006842.

Buffie, C. G. *et al.* (2015) 'Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile', *Nature*. NIH Public Access, 517(7533), pp. 205–208. doi: 10.1038/nature13828.

Burk, R. F. *et al.* (2006) 'Effects of chemical form of selenium on plasma biomarkers in a high-dose human supplementation trial', *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. Cancer Epidemiol Biomarkers Prev, 15(4), pp. 804–810. doi: 10.1158/1055-9965.EPI-05-0950.

Callahan, B. J., McMurdie, P. J. and Holmes, S. P. (2017) 'Exact sequence variants should replace operational taxonomic units in marker-gene data analysis', *ISME Journal*. Nature Publishing Group, 11(12), pp. 2639–2643. doi: 10.1038/ismej.2017.119.

Candela, M. *et al.* (2016) 'Modulation of gut microbiota dysbioses in type 2 diabetic patients by macrobiotic Ma-Pi 2 diet', *British Journal of Nutrition*. Cambridge University Press, 116(1), pp. 80–93. doi: 10.1017/S0007114516001045.

Capellino, S. *et al.* (2014) 'Increased expression of dopamine receptors in synovial fibroblasts from patients with rheumatoid arthritis: Inhibitory effects of dopamine on interleukin-8 and interleukin-6', *Arthritis and Rheumatology*. John Wiley and Sons Inc., 66(10), pp. 2685–2693. doi: 10.1002/art.38746.

Castro, A. *et al.* (2019) *Association of skeletal muscle and serum metabolites with maximum power output gains in response to continuous endurance or high-intensity interval training programs: The TIMES study - A randomized controlled trial*, *PLoS ONE*. doi: 10.1371/journal.pone.0212115.

Chaplin, A. V. *et al.* (2015) 'Intraspecies Genomic Diversity and Long-Term Persistence of Bifidobacterium longum', *PLoS ONE*. Public Library of Science, 10(8). doi: 10.1371/JOURNAL.PONE.0135658.

Chapman, C. R., Tuckett, R. P. and Song, C. W. (2008) 'Pain and Stress in a Systems Perspective: Reciprocal Neural, Endocrine, and Immune Interactions', *The Journal of Pain*. Elsevier, 9(2), pp. 122–145. doi: 10.1016/J.JPAIN.2007.09.006.

Cheng, R. *et al.* (2022) 'A randomized controlled trial for response of microbiome network to exercise and diet intervention in patients with nonalcoholic fatty liver disease', *Nature Communications 2022 13:1*. Nature Publishing Group, 13(1), pp. 1–13. doi: 10.1038/s41467-022-29968-0.

Chong, J. *et al.* (2020) 'Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data', *Nature Protocols*. Springer US, 15(3), pp. 799–821. doi: 10.1038/s41596-019-0264-1.

Cianci, R. *et al.* (2018) 'The microbiota and immune system crosstalk in health and disease', *Mediators of Inflammation*. Hindawi, 2018, pp. 10–12. doi: 10.1155/2018/2912539.

Clarke, S. F. *et al.* (2014) 'Exercise and associated dietary extremes impact on gut microbial diversity', *Gut*. BMJ Publishing Group, 63(12), pp. 1913–1920. doi: 10.1136/gutjnl-2013-306541.

Cole, J. R. *et al.* (2007) 'The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data', *Nucleic acids research*. Nucleic Acids Res, 35(Database issue). doi: 10.1093/NAR/GKL889.

Cole, J. R. *et al.* (2014) 'Ribosomal Database Project: Data and tools for high throughput rRNA analysis', *Nucleic Acids Research*, 42(D1). doi: 10.1093/nar/gkt1244.

Collins, K. H. *et al.* (2015) 'Relationship between inflammation, the gut microbiota, and metabolic osteoarthritis development: Studies in a rat model', *Osteoarthritis and Cartilage*. W.B. Saunders Ltd, 23(11), pp. 1989–1998. doi:

10.1016/j.joca.2015.03.014.

Cope, K., Risby, T. and Diehl, A. M. (2000) 'Increased gastrointestinal ethanol production in obese mice: Implications for fatty liver disease pathogenesis', *Gastroenterology*. Gastroenterology, 119(5), pp. 1340–1347. doi: 10.1053/gast.2000.19267.

Cotillard, A. intervention impact on gut microbial gene richness *et al.* (2013) 'Dietary intervention impact on gut microbial gene richness', *Nature*. Nature, 500(7464), pp. 585–588. doi: 10.1038/nature12480.

Cox, M. A. *et al.* (2009) 'Short-chain fatty acids act as antiinflammatory mediators by regulating prostaglandin E2 and cytokines', *World Journal of Gastroenterology*. Baishideng Publishing Group Inc, 15(44), pp. 5549–5557. doi: 10.3748/wjg.15.5549.

Cui, L., Morris, A. and Ghedin, E. (2013) 'The human mycobiome in health and disease', *Genome Medicine*. BioMed Central, p. 63. doi: 10.1186/gm467.

Dabke, K., Hendrick, G. and Devkota, S. (2019) 'The gut microbiome and metabolic syndrome', *Journal of Clinical Investigation*. American Society for Clinical Investigation, pp. 4050–4057. doi: 10.1172/JCI129194.

Daghestani, H. N., Pieper, C. F. and Kraus, V. B. (2015) 'Soluble Macrophage Biomarkers Indicate Inflammatory Phenotypes in Patients With Knee Osteoarthritis', *Arthritis & Rheumatology*. John Wiley & Sons, Ltd, 67(4), pp. 956–965. doi: 10.1002/ART.39006.

Dai, X. *et al.* (2020) 'Microbial Metabolites: Critical Regulators in NAFLD', *Frontiers in Microbiology*. Frontiers Media S.A., p. 567654. doi: 10.3389/fmicb.2020.567654.

Das, B. *et al.* (2018) 'Analysis of the gut microbiome of rural and urban healthy indians living in sea level and high altitude areas', *Scientific Reports*. Springer US, 8(1), pp. 1–15. doi: 10.1038/s41598-018-28550-3.

Das, N. K. *et al.* (2020) 'Microbial Metabolite Signaling Is Required for Systemic Iron Homeostasis', *Cell Metabolism*. Cell Press, 31(1), pp. 115-130.e6. doi: 10.1016/j.cmet.2019.10.005.

Davenport, E. R. *et al.* (2014) 'Seasonal variation in human gut microbiome composition', *PLoS ONE*. Public Library of Science, 9(3). doi: 10.1371/journal.pone.0090731.

David, L. A. *et al.* (2014) 'Diet rapidly and reproducibly alters the human gut microbiome', *Nature*. NIH Public Access, 505(7484), pp. 559–563. doi: 10.1038/nature12820.

DeSantis, T. Z. *et al.* (2006) 'Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB', *Applied and Environmental Microbiology*. American Society for Microbiology (ASM), 72(7), p. 5069. doi: 10.1128/AEM.03006-05.

Deschasaux, M. *et al.* (2018) 'Depicting the composition of gut microbiota in a population with varied ethnic origins but shared geography', *Nature Medicine*. Nature Publishing Group, 24(10), pp. 1526–1531. doi: 10.1038/s41591-018-0160-1.

Deveza, L. A. *et al.* (2017) 'Knee osteoarthritis phenotypes and their relevance for outcomes: a systematic review', *Osteoarthritis and Cartilage*. W.B. Saunders, pp. 1926–1941. doi: 10.1016/j.joca.2017.08.009.

Diamond, B. *et al.* (2011) 'It takes guts to grow a brain: Increasing evidence of the important role of the intestinal microflora in neuro- and immune-modulatory functions during development and adulthood', *BioEssays*. NIH Public Access, 33(8), pp. 588–591. doi: 10.1002/bies.201100042.

DIETARY GUIDELINES FOR INDIANS-A Manual (no date). doi: Nat Inst Nutrition. 2. 89-117. 2011.

Dinh, D. M. *et al.* (2016) 'Longitudinal analysis of the intestinal microbiota in persistently stunted young children in south India', *PLoS ONE*, 11(5), pp. 1–17. doi: 10.1371/journal.pone.0155405.

Donohoe, D. R. *et al.* (2011) 'The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon', *Cell Metabolism*. Cell Metab, 13(5), pp. 517–526. doi: 10.1016/j.cmet.2011.02.018.

Doumatey, A. P. *et al.* (2020) 'Gut Microbiome Profiles Are Associated With Type 2 Diabetes in Urban Africans', *Frontiers in Cellular and Infection Microbiology*. Frontiers Media S.A., 10, p. 63. doi: 10.3389/fcimb.2020.00063.

Du, Y. *et al.* (2011) 'Isoleucine or valine deprivation stimulates fat loss via increasing energy expenditure and regulating lipid metabolism in WAT', *Amino Acids 2011 43:2*. Springer, 43(2), pp. 725–734. doi: 10.1007/S00726-011-1123-8.

Duboc, H. *et al.* (2013) 'Connecting dysbiosis, bile-acid dysmetabolism and gut inflammation in inflammatory bowel diseases', *Gut*. Gut, 62(4), pp. 531–539. doi: 10.1136/GUTJNL-2012-302578.

Dumas, M. E. *et al.* (2006) 'Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice', *Proceedings of the National Academy of Sciences of the United States of America*. Proc Natl Acad Sci U S A, 103(33), pp. 12511–12516. doi: 10.1073/PNAS.0601056103.

Eloe-Fadrosh, E. A. and Rasko, D. A. (2013) 'The human microbiome: From symbiosis to pathogenesis', *Annual Review of Medicine*, pp. 145–163. doi: 10.1146/annurev-med-010312-133513.

Esposito, K. *et al.* (2009) 'Adherence to a Mediterranean diet and glycaemic control in Type 2 diabetes mellitus', *Diabetic Medicine*. John Wiley & Sons, Ltd, 26(9), pp. 900–907. doi: 10.1111/j.1464-5491.2009.02798.x.

Evangelou, E. *et al.* (2009) 'Large-scale analysis of association between GDF5 and FRZB variants and osteoarthritis of the hip, knee, and hand', *Arthritis and*

Rheumatism. Arthritis Rheum, 60(6), pp. 1710–1721. doi: 10.1002/art.24524.

Fact sheets - Malnutrition (no date). Available at: https://www.who.int/news-room/fact-sheets/detail/malnutrition (Accessed: 15 April 2022).

Fadrosh, D. W. *et al.* (2014) 'An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform', *Microbiome*. BioMed Central Ltd., 2(1), p. 6. doi: 10.1186/2049-2618-2-6.

Fairweather-Tait, S. J. *et al.* (2011) 'Selenium in Human Health and Disease', *https://home.liebertpub.com/ars*. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA , 14(7), pp. 1337–1383. doi: 10.1089/ARS.2010.3275.

Faith, J. J. *et al.* (2013) 'The Long-Term Stability of the Human Gut Microbiota', *Science*. American Association for the Advancement of Science, 341(6141), p. 1237439. doi: 10.1126/science.1237439.

Fan, J. G. *et al.* (2005) 'Prevalence of and risk factors for fatty liver in a general population of Shanghai, China', *Journal of Hepatology*. Elsevier, 43(3), pp. 508–514. doi: 10.1016/j.jhep.2005.02.042.

Fan, Y. and Pedersen, O. (2021) 'Gut microbiota in human metabolic health and disease', *Nature Reviews Microbiology*. Nature Publishing Group, pp. 55–71. doi: 10.1038/s41579-020-0433-9.

Faquih, T. *et al.* (2020) 'A workflow for missing values imputation of untargeted metabolomics data', *Metabolites*, 10(12), pp. 1–23. doi: 10.3390/metabo10120486.

Favazzo, L. J. *et al.* (2020) 'The gut microbiome-joint connection: Implications in osteoarthritis', *Current Opinion in Rheumatology*, pp. 92–101. doi: 10.1097/BOR.0000000000000681.

Favier, C. F. *et al.* (2002) 'Molecular monitoring of succession of bacterial communities in human neonates', *Applied and Environmental Microbiology*. American Society for Microbiology (ASM), 68(1), pp. 219–226. doi: 10.1128/AEM.68.1.219-226.2002.

Fernandes, G. S. *et al.* (2017) 'Knee pain and related health in the community study (KPIC): A cohort study protocol', *BMC Musculoskeletal Disorders*. BioMed Central, 18(1), pp. 1–13. doi: 10.1186/s12891-017-1761-4.

De Filippis, F. *et al.* (2016) 'Unusual sub-genus associations of faecal Prevotella and Bacteroides with specific dietary patterns', *Microbiome*. Microbiome, 4, pp. 1–6. doi: 10.1186/s40168-016-0202-1.

Finco, G. *et al.* (2016) 'Can urine metabolomics be helpful in differentiating neuropathic and nociceptive pain? a proof-of-concept study', *PLoS ONE*, 11(3), p. e0150476. doi: 10.1371/journal.pone.0150476.

Flint, H. J., Scott, K. P., Duncan, S. H., *et al.* (2012) 'Microbial degradation of

complex carbohydrates in the gut', *Gut Microbes*. Taylor & Francis, 3(4), p. 289. doi: 10.4161/GMIC.19897.

Flint, H. J., Scott, K. P., Louis, P., *et al.* (2012) 'The role of the gut microbiota in nutrition and health', *Nature Reviews Gastroenterology & Hepatology 2012 9:10*. Nature Publishing Group, 9(10), pp. 577–589. doi: 10.1038/nrgastro.2012.156.

Fu, K., Robbins, S. R. and McDougall, J. J. (2018) 'Osteoarthritis: the genesis of pain', *Rheumatology*. Oxford Academic, 57(suppl_4), pp. iv43–iv50. doi: 10.1093/RHEUMATOLOGY/KEX419.

Fujisaka, S. *et al.* (2018) 'Diet, Genetics, and the Gut Microbiome Drive Dynamic Changes in Plasma Metabolites', *Cell Reports*. Elsevier B.V., 22(11), pp. 3072–3086. doi: 10.1016/j.celrep.2018.02.060.

Gacesa, R. *et al.* (2022) 'Environmental factors shaping the gut microbiome in a Dutch population', *Nature 2022 604:7907*. Nature Publishing Group, 604(7907), pp. 732–739. doi: 10.1038/s41586-022-04567-7.

Gago, G. *et al.* (2011) 'Fatty acid biosynthesis in actinomycetes', *FEMS Microbiology Reviews*. NIH Public Access, pp. 475–497. doi: 10.1111/j.1574-6976.2010.00259.x.

Galaris, D., Barbouti, A. and Pantopoulos, K. (2019) 'Iron homeostasis and oxidative stress: An intimate relationship', *Biochimica et Biophysica Acta - Molecular Cell Research*. Elsevier, p. 118535. doi: 10.1016/j.bbamcr.2019.118535.

Galler, J. R. *et al.* (2012) 'Socioeconomic outcomes in adults malnourished in the first year of life: A 40-year study', *Pediatrics*, 130(1), pp. 1–7. doi: 10.1542/peds.2012-0073.

Gangadoo, S. *et al.* (2019) 'In vitro growth of gut microbiota with selenium nanoparticles', *Animal Nutrition*. KeAi Publishing, 5(4), p. 424. doi: 10.1016/J.ANINU.2019.06.004.

Garud, N. R. and Pollard, K. S. (2020) 'Population Genetics in the Human Microbiome', *Trends in Genetics*. Elsevier Ltd, pp. 53–67. doi: 10.1016/j.tig.2019.10.010.

St. George, A. *et al.* (2009) 'Independent effects of physical activity in patients with nonalcoholic fatty liver disease', *Hepatology*. Hepatology, 50(1), pp. 68–76. doi: 10.1002/hep.22940.

George, C. E. *et al.* (2019) 'Health issues in a Bangalore slum: Findings from a household survey using a mobile screening toolkit in Devarajeevanahalli', *BMC Public Health*. BMC Public Health, 19(1), pp. 1–12. doi: 10.1186/s12889-019-6756-7.

Ghazalpour, A. *et al.* (2016) 'Expanding role of gut microbiota in lipid metabolism', *Current Opinion in Lipidology*. Lippincott Williams and Wilkins, pp. 141–147. doi: 10.1097/MOL.0000000000000278.

Ghosh, S. S. *et al.* (2020) 'Intestinal barrier dysfunction, LPS translocation, and disease development', *Journal of the Endocrine Society*. Oxford Academic, 4(2). doi: 10.1210/jendso/bvz039.

Ghosh, T. S. *et al.* (2014) 'Gut microbiomes of Indian children of varying nutritional status', *PLoS ONE*, 9(4), pp. 1–13. doi: 10.1371/journal.pone.0095547.

Gibney, M. J. *et al.* (2005) 'Metabolomics in human nutrition: opportunities and challenges', *The American journal of clinical nutrition*. Am J Clin Nutr, 82(3), pp. 497–503. doi: 10.1093/AJCN.82.3.497.

Gillespie, J. (2021) '"You Are What You Eat": The Role of Dietary Macronutrients and Micronutrients in MAFLD', *Clinical Liver Disease*. Wiley-Blackwell, 18(2), p. 67. doi: 10.1002/CLD.1083.

Glimcher, L. H. and Lee, A. H. (2009) 'From sugar to fat: How the transcription factor XBP1 regulates hepatic lipogenesis', *Annals of the New York Academy of Sciences*. NIH Public Access, p. E2. doi: 10.1111/j.1749-6632.2009.04956.x.

Gohir, S. A. *et al.* (2019) 'Evaluating the efficacy of Internet-Based Exercise programme Aimed at Treating knee Osteoarthritis (iBEAT-OA) in the community: A study protocol for a randomised controlled trial', *BMJ Open*, 9(10), p. e030564. doi: 10.1136/bmjopen-2019-030564.

Goodman, A. L. *et al.* (2011) 'Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 108(15), pp. 6252–6257. doi: 10.1073/pnas.1102938108.

Goodrich, J. K. *et al.* (2014) 'Human genetics shape the gut microbiome', *Cell*. Cell Press, 159(4), pp. 789–799. doi: 10.1016/J.CELL.2014.09.053/ATTACHMENT/9E974D9A-E7D0-4CBF-A4FF-77001365A165/MMC2.XLSX.

Goodrich, J. K. *et al.* (2016) 'Genetic Determinants of the Gut Microbiome in UK Twins', *Cell Host and Microbe*. Cell Press, 19(5), pp. 731–743. doi: 10.1016/j.chom.2016.04.017.

Gordon, J. I. *et al.* (2012) 'The human gut microbiota and undernutrition', *Science Translational Medicine*, 4(137), pp. 1–7. doi: 10.1126/scitranslmed.3004347.

Griffin, T. M. *et al.* (2012) 'Induction of osteoarthritis and metabolic inflammation by a very high-fat diet in mice: Effects of short-term exercise', *Arthritis and Rheumatism*. Arthritis Rheum, 64(2), pp. 443–453. doi: 10.1002/art.33332.

Grimby-Ekman, A. *et al.* (2017) 'Different DHEA-S Levels and Response Patterns in Individuals with Chronic Neck Pain, Compared with a Pain Free Group-a Pilot Study', *Pain medicine (Malden, Mass.)*. Oxford Academic, 18(5), pp. 846–855. doi: 10.1093/pm/pnw162.

Gu, Q. and Li, P. (2016) 'Biosynthesis of Vitamins by Probiotic Bacteria', in *Probiotics and Prebiotics in Human Nutrition and Health*. InTech. doi: 10.5772/63117.

Guccione, A. A. *et al.* (1994) 'The effects of specific medical conditions on the functional limitations of elders in the Framingham Study', *American journal of public health*. Am J Public Health, 84(3), pp. 351–358. doi: 10.2105/AJPH.84.3.351.

Gupta, S. *et al.* (2011) 'Metagenome of the gut of a malnourished child', *Gut Pathogens*. BioMed Central Ltd, 3(1), p. 7. doi: 10.1186/1757-4749-3-7.

Hae, J. K. *et al.* (2004) 'Metabolic significance of nonalcoholic fatty liver disease in nonobese, nondiabetic adults', *Archives of internal medicine*. Arch Intern Med, 164(19), pp. 2169–2175. doi: 10.1001/ARCHINTE.164.19.2169.

Hang, S. *et al.* (2019) 'Bile acid metabolites control TH17 and Treg cell differentiation', *Nature 2019 576:7785*. Nature Publishing Group, 576(7785), pp. 143–148. doi: 10.1038/s41586-019-1785-z.

Hansen, M. E. B. *et al.* (2019) 'Population structure of human gut bacteria in a diverse cohort from rural Tanzania and Botswana', *Genome Biology*. BioMed Central, 20(1), pp. 1–21. doi: 10.1186/s13059-018-1616-9.

Hariri, M. and Zohdi, S. (2019) 'Effect of Vitamin D on Non-Alcoholic Fatty Liver Disease: A Systematic Review of Randomized Controlled Clinical Trials', *International Journal of Preventive Medicine*. Wolters Kluwer -- Medknow Publications, 10(1), p. 14. doi: 10.4103/ijpvm.ijpvm_499_17.

Hartrick, C. T., Kovan, J. P. and Shapiro, S. (2003) 'The Numeric Rating Scale for Clinical Pain Measurement: A Ratio Measure?', *Pain Practice*. John Wiley & Sons, Ltd, 3(4), pp. 310–316. doi: 10.1111/J.1530-7085.2003.03034.X.

Hayashi, H., Sakamoto, M. and Benno, Y. (2002) 'Phylogenetic analysis of the human gut microbiota using 16S rDNA clone libraries and strictly anaerobic culture-based methods', *Microbiology and Immunology*. Center for Academic Publications Japan, 46(8), pp. 535–548. doi: 10.1111/j.1348-0421.2002.tb02731.x.

Heijtz, R. D. *et al.* (2011) 'Normal gut microbiota modulates brain development and behavior', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 108(7), pp. 3047–3052. doi: 10.1073/pnas.1010529108.

Hemsworth, G. R. *et al.* (2016) 'Structural dissection of a complex Bacteroides ovatus gene locus conferring xyloglucan metabolism in the human gut', *Open Biology*. Royal Society of London, 6(7). doi: 10.1098/rsob.160142.

Hill, C. J. *et al.* (2017) 'Evolution of gut microbiota composition from birth to 24 weeks in the INFANTMET Cohort', *Microbiome*. BioMed Central Ltd., 5(1), p. 4. doi: 10.1186/s40168-016-0213-y.

Hojyo, S. and Fukada, T. (2016) 'Roles of Zinc Signaling in the Immune System',

Journal of Immunology Research. Hindawi Limited, 2016. doi: 10.1155/2016/6762343.

Hollander, D. and Kaunitz, J. D. (2020) 'The "Leaky Gut": Tight Junctions but Loose Associations?', *Digestive Diseases and Sciences*. NIH Public Access, pp. 1277–1287. doi: 10.1007/s10620-019-05777-2.

Hou, D. *et al.* (2022) 'Alteration of fecal microbiome and metabolome by mung bean coat improves diet-induced non-alcoholic fatty liver disease in mice', *Food Science and Human Wellness*. Elsevier, 11(5), pp. 1259–1272. doi: 10.1016/J.FSHW.2022.04.023.

Hua, J. *et al.* (2010) 'Dietary fatty acids modulate antigen presentation to hepatic NKT cells in nonalcoholic fatty liver disease', *Journal of Lipid Research*. Elsevier, 51(7), pp. 1696–1703. doi: 10.1194/JLR.M003004.

Huang, Z. Y. *et al.* (2016) 'Both systemic and local lipopolysaccharide (LPS) burden are associated with knee OA severity and inflammation', *Osteoarthritis and Cartilage*. W.B. Saunders Ltd, 24(10), pp. 1769–1775. doi: 10.1016/j.joca.2016.05.008.

Huffnagle, G. B. and Noverr, M. C. (2013) 'The emerging world of the fungal microbiome', *Trends in Microbiology*. Elsevier Current Trends, pp. 334–341. doi: 10.1016/j.tim.2013.04.002.

Huson, D. H. *et al.* (2007) 'MEGAN analysis of metagenomic data', *Genome Research*. Cold Spring Harbor Laboratory Press, 17(3), pp. 377–386. doi: 10.1101/gr.5969107.

Huttenhower, C. *et al.* (2012) 'Structure, function and diversity of the healthy human microbiome', *Nature*. Nature Publishing Group, 486(7402), pp. 207–214. doi: 10.1038/nature11234.

Illumina Inc. (2013) '16S Metagenomic Sequencing Library Preparation - Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System', *16S Metagenomic Sequencing Library Preparation Manual*, pp. 1–28.

Inbaraj, L. R. *et al.* (2020) 'Paternal involvement in feeding and its association with nutritional status of children in an urban slum in a low-resource setting: A cross-sectional study', *Nutrition*. Elsevier Inc., 74. doi: 10.1016/j.nut.2020.110735.

International Institute for Population Sciences (2005) 'India', *International Journal of Health Care Quality Assurance*, 18(7), p. 765. doi: 10.1108/ijhcqa.2005.06218gab.007.

Jansson, J. *et al.* (2009) 'Metabolomics reveals metabolic biomarkers of Crohn's disease', *PLoS ONE*. PLoS One, 4(7). doi: 10.1371/journal.pone.0006386.

Jensen, V. S. *et al.* (2018) 'Dietary fat stimulates development of NAFLD more potently than dietary fructose in Sprague-Dawley rats', *Diabetology and Metabolic Syndrome*. BioMed Central Ltd., 10(1), pp. 1–13. doi: 10.1186/s13098-018-0307-8.

Jiang, P. *et al.* (2019) 'Reproducible changes in the gut microbiome suggest a shift in microbial and host metabolism during spaceflight', *Microbiome*. BioMed Central Ltd., 7(1), p. 113. doi: 10.1186/s40168-019-0724-4.

Ju, T. *et al.* (2019) 'Defining the role of Parasutterella, a previously uncharacterized member of the core gut microbiota', *ISME Journal*. Nature Publishing Group, 13(6), pp. 1520–1534. doi: 10.1038/s41396-019-0364-5.

Kanehisa, M. and Goto, S. (2000) 'KEGG: Kyoto Encyclopedia of Genes and Genomes', *Nucleic Acids Research*. Oxford University Press, 28(1), p. 27. doi: 10.1093/NAR/28.1.27.

Kaoutari, A. El *et al.* (2013) 'The abundance and variety of carbohydrate-active enzymes in the human gut microbiota', *Nature Reviews Microbiology*. Nature Publishing Group, 11(7), pp. 497–504. doi: 10.1038/nrmicro3050.

Karavia, E. A. *et al.* (2012) 'Apolipoprotein A-I Modulates Processes Associated with Diet-Induced Nonalcoholic Fatty Liver Disease in Mice', *Molecular Medicine*. The Feinstein Institute for Medical Research, 18(1), p. 901. doi: 10.2119/MOLMED.2012.00113.

Karlsson, F. H. *et al.* (2012) 'Symptomatic atherosclerosis is associated with an altered gut metagenome', *Nature Communications 2012 3:1*. Nature Publishing Group, 3(1), pp. 1–8. doi: 10.1038/ncomms2266.

Kasaikina, M. V. *et al.* (2011) 'Dietary selenium affects host selenoproteome expression by influencing the gut microbiota', *The FASEB Journal*. FASEB J, 25(7), pp. 2492–2499. doi: 10.1096/fj.11-181990.

Katsagoni, C. N. *et al.* (2017) 'A "healthy diet-optimal sleep" lifestyle pattern is inversely associated with liver stiffness and insulin resistance in patients with nonalcoholic fatty liver disease', *Applied physiology, nutrition, and metabolism = Physiologie appliquee, nutrition et metabolisme*. Appl Physiol Nutr Metab, 42(3), pp. 250–256. doi: 10.1139/APNM-2016-0492.

Kau, A. L. *et al.* (2012) 'Human nutrition, the gut microbiome, and immune system: envisioning the future', *Nature*, 474(7351), pp. 327–336. doi: 10.1038/nature10213.Human.

Keegan, K. P., Glass, E. M. and Meyer, F. (2016) 'MG-RAST, a metagenomics service for analysis of microbial community structure and function', in *Methods in Molecular Biology*. Humana Press Inc., pp. 207–233. doi: 10.1007/978-1-4939-3369-3_13.

Keirns, B. H. *et al.* (2020) 'Exercise and intestinal permeability: another form of exercise-induced hormesis?', *American Journal of Physiology - Gastrointestinal and Liver Physiology*. American Physiological Society, pp. G512–G518. doi: 10.1152/AJPGI.00232.2020.

Kerac, M. *et al.* (2009) 'Probiotics and prebiotics for severe acute malnutrition (PRONUT study): a double-blind efficacy randomised controlled trial in Malawi',

The Lancet. Elsevier Ltd, 374(9684), pp. 136–144. doi: 10.1016/S0140-6736(09)60884-9.

Kim, M. C. *et al.* (2020) 'Serum zinc level and hepatic fibrosis in patients with nonalcoholic fatty liver disease', *PLoS ONE*. Public Library of Science, 15(10 October), p. e0240195. doi: 10.1371/journal.pone.0240195.

Kontogianni, M. D. *et al.* (2014) 'Adherence to the Mediterranean diet is associated with the severity of non-alcoholic fatty liver disease', *Clinical Nutrition*. Churchill Livingstone, 33(4), pp. 678–683. doi: 10.1016/j.clnu.2013.08.014.

Kovatcheva-Datchary, P. *et al.* (2019) 'Simplified Intestinal Microbiota to Study Microbe-Diet-Host Interactions in a Mouse Model', *Cell Reports*. Elsevier, 26(13), p. 3772. doi: 10.1016/J.CELREP.2019.02.090.

Krajmalnik-Brown, R. *et al.* (2012) 'Effects of gut microbes on nutrient absorption and energy regulation', *Nutrition in Clinical Practice*, 27(2), pp. 201–214. doi: 10.1177/0884533611436116.

Kravetz, A. M. *et al.* (2020) 'Effect of gut microbiota and PNPLA3 rs738409 variant on nonalcoholic fatty liver disease (NAFLD) in obese youth', *Journal of Clinical Endocrinology and Metabolism*. Endocrine Society, 105(10). doi: 10.1210/clinem/dgaa382.

Kroke, A. *et al.* (1999) 'Validation of a self-administered food-frequency questionnaire administered in the European Prospective Investigation into Cancer and Nutrition (EPIC) Study: comparison of energy, protein, and macronutrient intakes estimated with the doubly labeled water, urinary nitrogen, and repeated 24-h dietary recall methods', *The American journal of clinical nutrition*. Am J Clin Nutr, 70(4), pp. 439–447. doi: 10.1093/AJCN/70.4.439.

Kuzuyama, T. (2017) 'Biosynthetic studies on terpenoids produced by Streptomyces', *Journal of Antibiotics*. Nature Publishing Group, 70(7), pp. 811–818. doi: 10.1038/ja.2017.12.

Langille, M. G. I. *et al.* (2013) 'Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences', *Nature Biotechnology*. Nature Publishing Group, 31(9), pp. 814–821. doi: 10.1038/nbt.2676.

Larsbrink, J. *et al.* (2014) 'A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes', *Nature*. Nature, 506(7489), pp. 498–502. doi: 10.1038/nature12907.

Lazar, V. *et al.* (2018) 'Aspects of gut microbiota and immune system interactions in infectious diseases, immunopathology, and cancer', *Frontiers in Immunology*, 9(AUG), pp. 1–18. doi: 10.3389/fimmu.2018.01830.

Lemon, K. P. *et al.* (2012) 'Microbiota-targeted therapies: An ecological perspective', *Science Translational Medicine*, pp. 137–142. doi: 10.1126/scitranslmed.3004183.

Ley, R. E. *et al.* (2006) 'Microbial ecology: Human gut microbes associated with obesity', *Nature*. Plenum, 444(7122), pp. 1022–1023. doi: 10.1038/4441022a.

Li, K., Bihan, M. and Methé, B. A. (2013) 'Analyses of the Stability and Core Taxonomic Memberships of the Human Microbiome', *PLoS ONE*, 8(5). doi: 10.1371/journal.pone.0063139.

Li, Z. *et al.* (2003) 'Probiotics and antibodies to TNF inhibit inflammatory activity and improve nonalcoholic fatty liver disease', *Hepatology*. W.B. Saunders, 37(2), pp. 343–350. doi: 10.1053/jhep.2003.50048.

Li, Zhen *et al.* (2022) 'SCFAs Ameliorate Chronic Postsurgical Pain–Related Cognition Dysfunction via the ACSS2-HDAC2 Axis in Rats', *Molecular Neurobiology*. Springer, 59(10), pp. 6211–6227. doi: 10.1007/S12035-022-02971-8/FIGURES/9.

Liang, D. *et al.* (2018) 'Involvement of gut microbiome in human health and disease: Brief overview, knowledge gaps and research opportunities', *Gut Pathogens*. BioMed Central, 10(1), pp. 1–9. doi: 10.1186/s13099-018-0230-4.

Lin, B. *et al.* (2020) 'Gut microbiota regulates neuropathic pain: Potential mechanisms and therapeutic strategy', *Journal of Headache and Pain*. BioMed Central, p. 103. doi: 10.1186/s10194-020-01170-x.

Lin, C. L. *et al.* (2020) 'Bifidobacterium longum subsp. Longum olp-01 supplementation during endurance running training improves exercise performance in middle-and long-distance runners: A double-blind controlled trial', *Nutrients*. Multidisciplinary Digital Publishing Institute, 12(7), pp. 1–14. doi: 10.3390/nu12071972.

Lindenlaub, T. *et al.* (2000) 'Effects of neutralizing antibodies to TNF-alpha on pain-related behavior and nerve regeneration in mice with chronic constriction injury', *Brain Research*. Brain Res, 866(1–2), pp. 15–22. doi: 10.1016/S0006-8993(00)02190-9.

Liu, S. *et al.* (2016) 'The Host Shapes the Gut Microbiota via Fecal MicroRNA', *Cell Host and Microbe*. Cell Press, 19(1), pp. 32–43. doi: 10.1016/j.chom.2015.12.005.

Logue, K. *et al.* (2016) 'Unbiased Characterization of Anopheles Mosquito Blood Meals by Targeted High-Throughput Sequencing', *PLOS Neglected Tropical Diseases*. Public Library of Science, 10(3), p. e0004512. doi: 10.1371/JOURNAL.PNTD.0004512.

Loomba, R. and Sanyal, A. J. (2013) 'The global NAFLD epidemic', *Nature Reviews Gastroenterology and Hepatology*. Nature Publishing Group, pp. 686–690. doi: 10.1038/nrgastro.2013.171.

Lozupone, C. A. *et al.* (2012) 'Diversity, stability and resilience of the human gut microbiota', *Nature*, 489(7415), pp. 220–230. doi: 10.1038/nature11550.

Lozupone, C. and Knight, R. (2005) 'UniFrac: A new phylogenetic method for

comparing microbial communities', *Applied and Environmental Microbiology*. American Society for Microbiology (ASM), 71(12), pp. 8228–8235. doi: 10.1128/AEM.71.12.8228-8235.2005.

Maerz, T. *et al.* (2018) 'Metabolomic serum profiling after ACL injury in rats: A pilot study implicating inflammation and immune dysregulation in post-traumatic osteoarthritis', *Journal of Orthopaedic Research*. John Wiley and Sons Inc., 36(7), pp. 1969–1979. doi: 10.1002/jor.23854.

Magurran (1994) 'Measuring biological diversity', *Environmental and Ecological Statistics*, 1(2), pp. 95–103. doi: 10.1007/BF02426650.

Mahowald, M. A. *et al.* (2009) 'Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 106(14), pp. 5859–5864. doi: 10.1073/pnas.0901529106.

Mandal, S. *et al.* (2015) 'Analysis of composition of microbiomes: a novel method for studying microbial composition', *Microbial Ecology in Health & Disease*. Co-Action Publishing, 26(0). doi: 10.3402/mehd.v26.27663.

Manrique, P. *et al.* (2016) 'Healthy human gut phageome', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 113(37), pp. 10400–10405. doi: 10.1073/pnas.1601060113.

Marchesi, J. R. *et al.* (2016) 'The gut microbiota and host health: A new clinical frontier', *Gut*. BMJ Publishing Group, 65(2), pp. 330–339. doi: 10.1136/gutjnl-2015-309990.

Marcobal, A. *et al.* (2013) 'A metabolomic view of how the human gut microbiota impacts the host metabolome using humanized and gnotobiotic mice', *ISME Journal*. Nature Publishing Group, 7(10), pp. 1933–1943. doi: 10.1038/ismej.2013.89.

Matsubayashi, Y. *et al.* (2022) 'Impact of metabolic syndrome and metabolic dysfunction-associated fatty liver disease on cardiovascular risk by the presence or absence of type 2 diabetes and according to sex', *Cardiovascular Diabetology*. BioMed Central, 21(1), pp. 1–10. doi: 10.1186/s12933-022-01518-4.

Matsumoto, M. *et al.* (2012) 'Impact of intestinal microbiota on intestinal luminal metabolome', *Scientific reports*. Sci Rep, 2. doi: 10.1038/SREP00233.

Mayneris-Perxachs, J. and Swann, J. R. (2019) 'Metabolic phenotyping of malnutrition during the first 1000 days of life', *European Journal of Nutrition*. Dr. Dietrich Steinkopff Verlag GmbH and Co. KG, pp. 909–930. doi: 10.1007/s00394-018-1679-0.

McHardy, I. H. *et al.* (2013) 'Integrative analysis of the microbiome and metabolome of the human intestinal mucosal surface reveals exquisite inter-relationships', *Microbiome*. BioMed Central Ltd., 1(1), p. 17. doi: 10.1186/2049-

2618-1-17.

McNulty, N. P. *et al.* (2013) 'Effects of Diet on Resource Utilization by a Model Human Gut Microbiota Containing Bacteroides cellulosilyticus WH2, a Symbiont with an Extensive Glycobiome', *PLoS Biology*. PLoS Biol, 11(8). doi: 10.1371/journal.pbio.1001637.

Menni, C. *et al.* (2017) 'Gut microbiome diversity and high-fibre intake are related to lower long-term weight gain', *International Journal of Obesity*. Nature Publishing Group, 41(7), pp. 1099–1105. doi: 10.1038/ijo.2017.66.

Menni, C. *et al.* (2019) 'Circulating levels of the anti-oxidant indoleproprionic acid are associated with higher gut microbiome diversity', *Gut Microbes*. Taylor and Francis Inc., 10(6), pp. 688–695. doi: 10.1080/19490976.2019.1586038.

Millar, B. *et al.* (2020) 'Investigating musculoskeletal health and wellbeing; A cohort study protocol', *BMC Musculoskeletal Disorders*. BioMed Central Ltd., 21(1), pp. 1–10. doi: 10.1186/S12891-020-03195-4/TABLES/1.

Miller, R. J. *et al.* (2009) 'Cytokine and chemokine regulation of sensory neuron function', *Handbook of Experimental Pharmacology*. Handb Exp Pharmacol, pp. 417–449. doi: 10.1007/978-3-540-79090-7_12.

Minot, S. *et al.* (2013) 'Rapid evolution of the human gut virome', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 110(30), pp. 12450–12455. doi: 10.1073/pnas.1300833110.

Mirmiran, P. *et al.* (2017) 'Relationship between Diet and Non-alcoholic Fatty Liver Disease: A Review Article', *Iranian Journal of Public Health*. Tehran University of Medical Sciences, 46(8), p. 1007.

Mohan, V. *et al.* (2009) 'Prevalence of non-alcoholic fatty liver disease in urban south Indians in relation to different grades of glucose intolerance and metabolic syndrome', *Diabetes Research and Clinical Practice*. Elsevier, 84(1), pp. 84–91. doi: 10.1016/j.diabres.2008.11.039.

Monaghan, T. M. *et al.* (2020) 'Metagenomics reveals impact of geography and acute diarrheal disease on the Central Indian human gut microbiome', *Gut Microbes*. Taylor and Francis Inc., 12(1), pp. 1–24. doi: 10.1080/19490976.2020.1752605.

Monda, V. *et al.* (2017) 'Exercise Modifies the Gut Microbiota with Positive Health Effects', *Oxidative medicine and cellular longevity*. Oxid Med Cell Longev, 2017. doi: 10.1155/2017/3831972.

Mortensen, M. S. *et al.* (2018) 'Stability and resilience of the intestinal microbiota in children in daycare - A 12 month cohort study', *BMC Microbiology*. BMC Microbiology, 18(1), pp. 1–11. doi: 10.1186/s12866-018-1367-5.

Mousavi, S. N. *et al.* (2018) 'Zinc and Selenium Co-supplementation Reduces Some Lipid Peroxidation and Angiogenesis Markers in a Rat Model of NAFLD-

Fed High Fat Diet', *Biological Trace Element Research*. Biol Trace Elem Res, 181(2), pp. 288–295. doi: 10.1007/s12011-017-1059-2.

Mrudulakumari Vasudevan, U. and Lee, E. Y. (2020) 'Flavonoids, terpenoids, and polyketide antibiotics: Role of glycosylation and biocatalytic tactics in engineering glycosylation', *Biotechnology Advances*. Elsevier Inc, 41, p. 107550. doi: 10.1016/j.biotechadv.2020.107550.

Muccioli, G. G. *et al.* (2010) 'The endocannabinoid system links gut microbiota to adipogenesis', *Molecular Systems Biology*. John Wiley & Sons, Ltd, 6(1), p. 392. doi: 10.1038/msb.2010.46.

Mulligan, A. A. *et al.* (2014) 'A new tool for converting food frequency questionnaire data into nutrient and food group values: FETA research methods and availability', *BMJ Open*. British Medical Journal Publishing Group, 4(3), p. e004503. doi: 10.1136/BMJOPEN-2013-004503.

Nalini Kanta, S. *et al.* (2020) 'An overview on NMR spectroscopy based metabolomics', *International Journal of Pharmaceutical Sciences and Developmental Research*. Peertechz Publications Private Limited, 6(1), pp. 016–020. doi: 10.17352/ijpsdr.000029.

Nam, Y. Do *et al.* (2011) 'Comparative analysis of korean human gut microbiota by barcoded pyrosequencing', *PLoS ONE*. Edited by E. Dias-Neto. Public Library of Science, 6(7), p. e22109. doi: 10.1371/journal.pone.0022109.

National Institute of Health (US) (2007) 'Understanding Human Genetic Variation - NIH Curriculum Supplement Series - NCBI Bookshelf'. National Institutes of Health (US).

O'Reilly, S. C., Muir, K. R. and Doherty, M. (1996) 'Screening for pain in knee osteoarthritis: Which question?', *Annals of the Rheumatic Diseases*. Ann Rheum Dis, 55(12), pp. 931–933. doi: 10.1136/ard.55.12.931.

Ordovas, J. M. *et al.* (2002) 'Dietary fat intake determines the effect of a common polymorphism in the hepatic lipase gene promoter on high-density lipoprotein metabolism: Evidence of a strong dose effect in this gene-nutrient interaction in the Framingham study', *Circulation*. Lippincott Williams & Wilkins, 106(18), pp. 2315–2321. doi: 10.1161/01.CIR.0000036597.52291.C9.

Ortiz-Estrada, Á. M. *et al.* (2019) 'Predictive functional profiles using metagenomic 16S rRNA data: a novel approach to understanding the microbial ecology of aquaculture systems', *Reviews in Aquaculture*, 11(1), pp. 234–245. doi: 10.1111/raq.12237.

Ossipov, M. H., Dussor, G. O. and Porreca, F. (2010) 'Central modulation of pain', *Journal of Clinical Investigation*. American Society for Clinical Investigation, pp. 3779–3787. doi: 10.1172/JCI43766.

Ott, B. *et al.* (2017) 'Effect of caloric restriction on gut permeability, inflammation markers, and fecal microbiota in obese women', *Scientific Reports*, 7(1), pp. 1–10. doi: 10.1038/s41598-017-12109-9.

Pages - Slum Statistics (no date). Available at: https://karunadu.karnataka.gov.in/ksdb/Pages/Slum-Statistics.aspx (Accessed: 29 June 2020).

Palmer, C. *et al.* (2007) 'Development of the human infant intestinal microbiota', *PLoS Biology*. Edited by Y. Ruan. Public Library of Science, 5(7), pp. 1556–1573. doi: 10.1371/journal.pbio.0050177.

Pang, Z. *et al.* (2021) 'MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights', *Nucleic Acids Research*. Oxford Academic, 49(W1), pp. W388–W396. doi: 10.1093/NAR/GKAB382.

Park, J. *et al.* (2015) 'Short-chain fatty acids induce both effector and regulatory T cells by suppression of histone deacetylases and regulation of the mTOR-S6K pathway', *Mucosal Immunology*. Nature Publishing Group, 8(1), pp. 80–93. doi: 10.1038/mi.2014.44.

Partlett, C. and Riley, R. D. (2017) 'Random effects meta-analysis: Coverage performance of 95% confidence and prediction intervals following REML estimation', *Statistics in Medicine*. Wiley-Blackwell, 36(2), pp. 301–317. doi: 10.1002/sim.7140.

Pataky, Z. *et al.* (2016) 'Impact of Hypocaloric Hyperproteic Diet on Gut Microbiota in Overweight or Obese Patients with Nonalcoholic Fatty Liver Disease: A Pilot Study', *Digestive Diseases and Sciences*, 61(9), pp. 2721–2731. doi: 10.1007/s10620-016-4179-1.

Paulson, J. N. *et al.* (2013) 'Differential abundance analysis for microbial marker-gene surveys', *Nature Methods*. Nat Methods, 10(12), pp. 1200–1202. doi: 10.1038/nmeth.2658.

Petrosino, J. F. (2018) 'The microbiome in precision medicine: The way forward', *Genome Medicine*. BioMed Central Ltd. doi: 10.1186/s13073-018-0525-6.

Pingali, P. *et al.* (2019) 'Diet Diversity and the Declining Importance of Staple Grains', pp. 73–91. doi: 10.1007/978-3-030-14409-8_4.

Polyzos, S. A. *et al.* (2020) 'Selenium and selenoprotein P in nonalcoholic fatty liver disease', *Hormones (Athens, Greece)*. Hormones (Athens), 19(1), pp. 61–72. doi: 10.1007/S42000-019-00127-3.

Ponist, S. *et al.* (2016) 'Effect of carnosine in experimental arthritis and on primary culture chondrocytes', *Oxidative Medicine and Cellular Longevity*. Hindawi Publishing Corporation, 2016, p. 11. doi: 10.1155/2016/8470589.

POSHAN Abhiyaan - PM's Overarching Scheme for Holistic Nourishment | National Portal of India (no date). Available at: https://www.india.gov.in/spotlight/poshan-abhiyaan-pms-overarching-scheme-holistic-nourishment (Accessed: 15 August 2021).

Pourhassan, M. *et al.* (2019) 'Hypophosphatemia as a sign of malnutrition in older hospitalized patients', *European Journal of Clinical Nutrition*. Nature

Publishing Group, 73(4), pp. 634–636. doi: 10.1038/s41430-018-0251-6.

Price, M. N., Dehal, P. S. and Arkin, A. P. (2010) 'FastTree 2 - Approximately maximum-likelihood trees for large alignments', *PLoS ONE*. Public Library of Science, 5(3), p. e9490. doi: 10.1371/journal.pone.0009490.

Prietl, B. *et al.* (2013) 'Vitamin D and Immune Function', *Nutrients 2013, Vol. 5, Pages 2502-2521*. Multidisciplinary Digital Publishing Institute, 5(7), pp. 2502–2521. doi: 10.3390/NU5072502.

Putignani, L. *et al.* (2018) 'Bifidobacteria and lactobacilli in the gut microbiome of children with non-alcoholic fatty liver disease: which strains act as health players?', *Archives of Medical Science*. Termedia Publishing, 14(1), pp. 81–87. doi: 10.5114/aoms.2016.62150.

QIIME (no date) *QIIME- SILVA Data Resource*. Available at: https://www.arb-silva.de/download/archive/qiime (Accessed: 1 February 2020).

Qin, J. *et al.* (2010) 'A human gut microbial gene catalogue established by metagenomic sequencing', *Nature*. Nature Publishing Group, 464(7285), pp. 59–65. doi: 10.1038/nature08821.

Quast, C. *et al.* (2013) 'The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools', *Nucleic Acids Research*, 41(D1). doi: 10.1093/nar/gks1219.

Quince, C. *et al.* (2017) 'Shotgun metagenomics, from sampling to analysis', *Nature Biotechnology 2017 35:9*. Nature Publishing Group, 35(9), pp. 833–844. doi: 10.1038/nbt.3935.

Rabin, M. L. *et al.* (2016) 'Postural, Bone, and Joint Disorders in Parkinson's Disease', *Movement Disorders Clinical Practice*. Wiley-Blackwell, pp. 538–547. doi: 10.1002/mdc3.12386.

Radjabzadeh, D. *et al.* (2020) 'Diversity, compositional and functional differences between gut microbiota of children and adults', *Scientific Reports*. Nature Research, 10(1), pp. 1–13. doi: 10.1038/s41598-020-57734-z.

Raine, E. V. A. *et al.* (2013) 'Allelic expression analysis of the osteoarthritis susceptibility gene COL11A1 in human joint tissues', *BMC Musculoskeletal Disorders*. BMC Musculoskelet Disord, 14(85). doi: 10.1186/1471-2474-14-85.

Rea, K. *et al.* (2019) 'Pain Bugs: Gut Microbiota and Pain Disorders', *Current Opinion in Physiology*, 11, pp. 97–102. doi: 10.1016/j.cophys.2019.10.001.

Reeves, M. A. and Hoffmann, P. R. (2009) 'The human selenoproteome: recent insights into functions and regulation', *Cellular and molecular life sciences : CMLS*. Cell Mol Life Sci, 66(15), pp. 2457–2478. doi: 10.1007/S00018-009-0032-4.

Reinecke, H. *et al.* (2015) 'Analgesic efficacy of opioids in chronic pain: recent meta-analyses', *British journal of pharmacology*. Br J Pharmacol, 172(2), pp.

324–333. doi: 10.1111/BPH.12634.

Rettedal, E. A., Gumpert, H. and Sommer, M. O. A. (2014) 'Cultivation-based multiplex phenotyping of human gut microbiota allows targeted recovery of previously uncultured bacteria', *Nature Communications*. Nature Publishing Group, 5. doi: 10.1038/ncomms5714.

Reyman, M. *et al.* (2019) 'Impact of delivery mode-associated gut microbiota dynamics on health in the first year of life', *Nature Communications*. Nature Publishing Group, 10(1), pp. 1–12. doi: 10.1038/s41467-019-13014-7.

Reynolds, T., Kuska, B. and Curtis, L. (2012) *NIH Human Microbiome Project defines normal bacterial makeup of the body | National Institutes of Health (NIH)*, *National Institutes of Health*. Available at: https://www.nih.gov/news-events/news-releases/nih-human-microbiome-project-defines-normal-bacterial-makeup-body (Accessed: 11 July 2020).

Ridaura, V. K. *et al.* (2013) 'Gut microbiota from twins discordant for obesity modulate metabolism in mice', *Science*. American Association for the Advancement of Science, 341(6150). doi: 10.1126/science.1241214.

Rizzatti, G. *et al.* (2017) 'Proteobacteria: A Common Factor in Human Diseases'. doi: 10.1155/2017/9351507.

Roberts, J. A. *et al.* (2020) 'Blood metabolite signature of metabolic syndrome implicates alterations in amino acid metabolism: Findings from the baltimore longitudinal study of aging (blsa) and the tsuruoka metabolomics cohort study (TMCS)', *International Journal of Molecular Sciences*. Multidisciplinary Digital Publishing Institute, 21(4), p. 1249. doi: 10.3390/ijms21041249.

Rogier, R. *et al.* (2017) 'Aberrant intestinal microbiota due to IL-1 receptor antagonist deficiency promotes IL-17- and TLR4-dependent arthritis', *Microbiome*. BioMed Central, 5(1), p. 63. doi: 10.1186/s40168-017-0278-2.

Rom, O. *et al.* (2018) 'Emerging therapeutic potential of glycine in cardiometabolic diseases: dual benefits in lipid and glucose metabolism', *Current opinion in lipidology*. NIH Public Access, 29(5), p. 428. doi: 10.1097/MOL.0000000000000543.

Romano, K. A. *et al.* (2015) 'Intestinal microbiota composition modulates choline bioavailability from diet and accumulation of the proatherogenic metabolite trimethylamine-N-oxide', *mBio*. American Society for Microbiology, 6(2). doi: 10.1128/mBio.02481-14.

Rowland, I. *et al.* (2018) 'Gut microbiota functions: metabolism of nutrients and other food components', *European Journal of Nutrition*. Dr. Dietrich Steinkopff Verlag GmbH and Co. KG, p. 1. doi: 10.1007/s00394-017-1445-8.

Le Roy, T. *et al.* (2013) 'Intestinal microbiota determines development of non-alcoholic fatty liver disease in mice', *Gut*. BMJ Publishing Group, 62(12), pp. 1787–1794. doi: 10.1136/gutjnl-2012-303816.

Ruhl, C. E. and Everharty, J. E. (2003) 'Determinants of the association of

overweight with elevated serum alanine aminotransferase activity in the United States', *Gastroenterology*. W.B. Saunders, 124(1), pp. 71–79. doi: 10.1053/gast.2003.50004.

Russell, W. R. *et al.* (2011) 'High-protein, reduced-carbohydrate weight-loss diets promote metabolite profiles likely to be detrimental to colonic health', *The American Journal of Clinical Nutrition*. Oxford Academic, 93(5), pp. 1062–1072. doi: 10.3945/ajcn.110.002188.

Ryan, M. C. *et al.* (2013) 'The Mediterranean diet improves hepatic steatosis and insulin sensitivity in individuals with non-alcoholic fatty liver disease', *Journal of Hepatology*. Elsevier, 59(1), pp. 138–143. doi: 10.1016/j.jhep.2013.02.012.

Sanchez, S. *et al.* (2017) 'Our microbes not only produce antibiotics, they also overproduce amino acids', *The Journal of Antibiotics 2018 71:1*. Nature Publishing Group, 71(1), pp. 26–36. doi: 10.1038/ja.2017.142.

Sanz, Y. *et al.* (2015) 'Understanding the role of gut microbiome in metabolic disease risk', *Pediatric Research*, pp. 236–244. doi: 10.1038/pr.2014.170.

Sarin, H. V. *et al.* (2019) 'Substantial fat mass loss reduces low-grade inflammation and induces positive alteration in cardiometabolic factors in normal-weight individuals', *Scientific Reports*. Nature Publishing Group, 9(1), pp. 1–14. doi: 10.1038/s41598-019-40107-6.

Saunders, J. and Smith, T. (2010) 'Malnutrition: Causes and consequences', *Clinical Medicine, Journal of the Royal College of Physicians of London*. Royal College of Physicians, pp. 624–627. doi: 10.7861/clinmedicine.10-6-624.

Sayin, S. I. *et al.* (2013) 'Gut microbiota regulates bile acid metabolism by reducing the levels of tauro-beta-muricholic acid, a naturally occurring FXR antagonist', *Cell Metabolism*. Elsevier, 17(2), pp. 225–235. doi: 10.1016/j.cmet.2013.01.003.

Scanzello, C. R., Plaas, A. and Crow, M. K. (2008) 'Innate immune system activation in osteoarthritis: Is osteoarthritis a chronic wound?', *Current Opinion in Rheumatology*, 20(5), pp. 565–572. doi: 10.1097/BOR.0B013E32830ABA34.

Scheiman, J. *et al.* (2019) 'Meta-omics analysis of elite athletes identifies a performance-enhancing microbe that functions via lactate metabolism', *Nature Medicine 2019 25:7*. Nature Publishing Group, 25(7), pp. 1104–1109. doi: 10.1038/s41591-019-0485-4.

Schloss, P. D. *et al.* (2009) 'Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities', *Applied and Environmental Microbiology*. American Society for Microbiology (ASM), 75(23), pp. 7537–7541. doi: 10.1128/AEM.01541-09.

Shannon, C. E. (1948) 'A Mathematical Theory of Communication', *Bell System Technical Journal*, 27(3), pp. 379–423. doi: 10.1002/j.1538-

7305.1948.tb01338.x.

Shao, Z. *et al.* (2020) 'S-allyl cysteine reduces osteoarthritis pathology in the tert-butyl hydroperoxide-treated chondrocytes and the destabilization of the medial meniscus model mice via the Nrf2 signaling pathway', *Aging*. Impact Journals LLC, 12(19), pp. 19254–19272. doi: 10.18632/aging.103757.

Sharma, M. *et al.* (2020) 'A comparison of the Indian diet with the EAT-Lancet reference diet', *BMC Public Health*. BioMed Central Ltd., 20(1), p. 812. doi: 10.1186/s12889-020-08951-8.

Sharon, G. *et al.* (2014) 'Specialized metabolites from the microbiome in health and disease', *Cell Metabolism*. Cell Press, pp. 719–730. doi: 10.1016/j.cmet.2014.10.016.

Shea, M. K. *et al.* (2010) 'Adulthood obesity is positively associated with adipose tissue concentrations of vitamin K and inversely associated with circulating indicators of vitamin K status in men and women', *Journal of Nutrition*. J Nutr, 140(5), pp. 1029–1034. doi: 10.3945/jn.109.118380.

Shell, W. E. *et al.* (2016) 'Reduction in pain and inflammation associated with chronic low back pain with the use of the medical food theramine', *American Journal of Therapeutics*. Lippincott Williams and Wilkins, 23(6), pp. e1353–e1362. doi: 10.1097/MJT.0000000000000068.

Shidfar, F. *et al.* (2018) 'Regression of Nonalcoholic Fatty Liver Disease with Zinc and Selenium Co-supplementation after Disease Progression in Rats', *Iranian Journal of Medical Sciences*. Shiraz University of Medical Sciences, 43(1), p. 26.

Singh, A. K. *et al.* (2019) 'Beneficial Effects of Dietary Polyphenols on Gut Microbiota and Strategies to Improve Delivery Efficiency', *Nutrients*. Multidisciplinary Digital Publishing Institute (MDPI), 11(9). doi: 10.3390/NU11092216.

Singh, P. *et al.* (2020) 'The potential role of vitamin D supplementation as a gut microbiota modifier in healthy individuals', *Scientific Reports 2020 10:1*. Nature Publishing Group, 10(1), pp. 1–14. doi: 10.1038/s41598-020-77806-4.

Smith, A. D. *et al.* (2016) 'Genetic and environmental influences on food preferences in adolescence', *The American Journal of Clinical Nutrition*. American Society for Nutrition, 104(2), p. 446. doi: 10.3945/AJCN.116.133983.

Soininen, P. *et al.* (2009) 'High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism', *Analyst*. The Royal Society of Chemistry, 134(9), pp. 1781–1785. doi: 10.1039/B910205A.

Spencer, M. D. *et al.* (2011) 'Association Between Composition of the Human Gastrointestinal Microbiome and Development of Fatty Liver With Choline Deficiency'. doi: 10.1053/j.gastro.2010.11.049.

Stearns, J. C. *et al.* (2017) 'Ethnic and diet-related differences in the healthy infant microbiome', *Genome Medicine*. BioMed Central Ltd., 9(1), p. 32. doi: 10.1186/s13073-017-0421-5.

Stranges, S. *et al.* (2010) 'Selenium status and cardiometabolic health: state of the evidence', *Nutrition, metabolism, and cardiovascular diseases : NMCD*. Nutr Metab Cardiovasc Dis, 20(10), pp. 754–760. doi: 10.1016/J.NUMECD.2010.10.001.

Sun, S., Jones, R. B. and Fodor, A. A. (2020) 'Inference-based accuracy of metagenome prediction tools varies across sample types and functional categories', *Microbiome*. Microbiome, 8(1), pp. 1–9. doi: 10.1186/s40168-020-00815-y.

Suzuki, K. *et al.* (2004) 'Aberrant expansion of segmented filamentous bacteria in IgA-deficient gut', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 101(7), pp. 1981–1986. doi: 10.1073/pnas.0307317101.

Swain, S. *et al.* (2020) 'Trends in incidence and prevalence of osteoarthritis in the United Kingdom: findings from the Clinical Practice Research Datalink (CPRD)', *Osteoarthritis and Cartilage*. W.B. Saunders Ltd, 28(6), pp. 792–801. doi: 10.1016/j.joca.2020.03.004.

Swaminathan, S. *et al.* (2019) 'The burden of child and maternal malnutrition and trends in its indicators in the states of India: the Global Burden of Disease Study 1990–2017', *The Lancet Child and Adolescent Health*, 3(12), pp. 855–870. doi: 10.1016/S2352-4642(19)30273-1.

Szychlinska, M. A. *et al.* (2019) 'A correlation between intestinal microbiota dysbiosis and osteoarthritis', *Heliyon*. Elsevier Ltd, 5(1). doi: 10.1016/j.heliyon.2019.e01134.

Tahapary, D. L. *et al.* (2018) 'Impact of rural-urban environment on metabolic profile and response to a 5-day high-fat diet', *Scientific Reports*. Nature Publishing Group, 8(1), pp. 1–9. doi: 10.1038/s41598-018-25092-6.

Tang, Z. Z. *et al.* (2019) 'Multi-omic analysis of the microbiome and metabolome in healthy subjects reveals microbiome-dependent relationships between diet and metabolites', *Frontiers in Genetics*. Frontiers Media S.A., 10(MAY), p. 454. doi: 10.3389/fgene.2019.00454.

Tette, E. M. A., Sifah, E. K. and Nartey, E. T. (2015) 'Factors affecting malnutrition in children and the uptake of interventions to prevent the condition', *BMC Pediatrics*. BMC Pediatrics, 15(1), pp. 1–11. doi: 10.1186/s12887-015-0496-3.

Teucher, B. *et al.* (2007) 'Dietary patterns and heritability of food choice in a UK female twin cohort', *Twin research and human genetics : the official journal of the International Society for Twin Studies*. Twin Res Hum Genet, 10(5), pp. 734–748. doi: 10.1375/TWIN.10.5.734.

Thifault, E. *et al.* (2013) 'Effects of age, sex, body mass index and APOE genotype on cardiovascular biomarker response to an n-3 polyunsaturated fatty acid supplementation', *Journal of Nutrigenetics and Nutrigenomics*. J Nutrigenet Nutrigenomics, pp. 73–82. doi: 10.1159/000350744.

Thindlu Village in Devanahalli (Bangalore Rural) Karnataka | villageinfo.in (no date). Available at: https://villageinfo.in/karnataka/bangalore-rural/devanahalli/thindlu.html (Accessed: 29 June 2020).

Tibshiranit, R. (1996) 'Regression Shrinkage and Selection Via the Lasso', *Journal of the Royal Statistical Society: Series B (Methodological)*. John Wiley & Sons, Ltd, 58(1), pp. 267–288. doi: 10.1111/J.2517-6161.1996.TB02080.X.

Trehan, I. *et al.* (2013) 'Antibiotics as part of the management of severe acute malnutrition', *New England Journal of Medicine*, 368(5), pp. 425–435. doi: 10.1056/NEJMoa1202851.

Trygg, J. and Wold, S. (2002) 'Orthogonal projections to latent structures (O-PLS)', *Journal of Chemometrics*. John Wiley & Sons, Ltd, 16(3), pp. 119–128. doi: 10.1002/CEM.695.

Turnbaugh, P. J. *et al.* (2007) 'The Human Microbiome Project', *Nature*. Nature Publishing Group, pp. 804–810. doi: 10.1038/nature06244.

Turnbaugh, P. J., Hamady, M., *et al.* (2009) 'A core gut microbiome in obese and lean twins', *Nature*. Nature, 457(7228), pp. 480–484. doi: 10.1038/NATURE07540.

Turnbaugh, P. J., Ridaura, V. K., *et al.* (2009) 'The effect of diet on the human gut microbiome: A metagenomic analysis in humanized gnotobiotic mice', *Science Translational Medicine*. American Association for the Advancement of Science, 1(6), p. 6ra14. doi: 10.1126/scitranslmed.3000322.

Tyakht, A. V. *et al.* (2013) 'Human gut microbiota community structures in urban and rural populations in Russia', *Nature Communications*. Nature Publishing Group, 4(1), pp. 1–9. doi: 10.1038/ncomms3469.

UNICEF (2013) 'Ready-to-use therapeutic food for children with severe acute malnutrition', *Position Paper*, (1), p. 4.

Unwin, D. J. *et al.* (2015) 'A pilot study to explore the role of a low-carbohydrate intervention to improve GGT levels and HbA 1c', *Diabesity in Practice*, 4.

Vangay, P. *et al.* (2018) 'US Immigration Westernizes the Human Gut Microbiome', *Cell*, 175(4), pp. 962-972.e10. doi: 10.1016/j.cell.2018.10.029.

Vázquez-Baeza, Y. *et al.* (2018) 'Impacts of the Human Gut Microbiome on Therapeutics', *Annual Review of Pharmacology and Toxicology*. Annual Reviews, pp. 253–270. doi: 10.1146/annurev-pharmtox-042017-031849.

Vela, D. (2018) 'Low hepcidin in liver fibrosis and cirrhosis; A tale of progressive disorder and a case for a new biochemical marker', *Molecular Medicine*. BioMed Central, pp. 1–18. doi: 10.1186/s10020-018-0008-7.

Velagapudi, V. R. *et al.* (2010) 'The gut microbiota modulates host energy and lipid metabolism in mice', *Journal of Lipid Research*. J Lipid Res, 51(5), pp. 1101–1112. doi: 10.1194/jlr.M002774.

Vester-Andersen, M. K. *et al.* (2019) 'Increased abundance of proteobacteria in aggressive Crohn's disease seven years after diagnosis', *Scientific Reports*, 9(1). doi: 10.1038/s41598-019-49833-3.

Vijay, A., Kouraki, A., *et al.* (2021) 'The anti-inflammatory effect of bacterial short chain fatty acids is partially mediated by endocannabinoids', *Gut Microbes*. Taylor and Francis Ltd., 13(1). doi: 10.1080/19490976.2021.1997559.

Vijay, A., Astbury, S., *et al.* (2021) 'The prebiotic effects of omega-3 fatty acid supplementation: A six-week randomised intervention trial) The prebiotic effects of omega-3 fatty acid supplementation: A six-week randomised intervention trial The prebiotic effects of omega-3 fatty acid supple', *Gut Microbes*, 13(1), p. 1863133. doi: 10.1080/19490976.2020.1863133.

Vincent, K. and Tracey, I. (2008) 'Hormones and their Interaction with the Pain Experience', *Reviews in Pain*. SAGE Publications, 2(2), p. 20. doi: 10.1177/204946370800200206.

Vitetta, L. *et al.* (2013) 'The gastrointestinal microbiome and musculoskeletal diseases: A beneficial role for probiotics and prebiotics', *Pathogens*, 2(4), pp. 606–626. doi: 10.3390/pathogens2040606.

Wagner, B. D. *et al.* (2018) 'On the use of diversity measures in longitudinal sequencing studies of microbial communities', *Frontiers in Microbiology*. Frontiers Media S.A., 9(MAY), p. 1037. doi: 10.3389/fmicb.2018.01037.

Wang, Linlin *et al.* (2021) 'Bifidobacterium bifidum Shows More Diversified Ways of Relieving Non-Alcoholic Fatty Liver Compared with Bifidobacterium adolescentis', *Biomedicines*. Biomedicines, 10(1). doi: 10.3390/BIOMEDICINES10010084.

Wang, X., Seo, Y. A. and Park, S. K. (2021) 'Serum selenium and non-alcoholic fatty liver disease (NAFLD) in U.S. adults: National Health and Nutrition Examination Survey (NHANES) 2011–2016', *Environmental Research*. Academic Press, 197, p. 111190. doi: 10.1016/j.envres.2021.111190.

Wang, Z. *et al.* (2020) 'Comparing Gut Microbiome in Mothers' Own Breast Milk- and Formula-Fed Moderate-Late Preterm Infants', *Frontiers in Microbiology*. Frontiers Media S.A., 11, p. 891. doi: 10.3389/fmicb.2020.00891.

Wells, J. C. *et al.* (2020) 'The double burden of malnutrition: aetiological pathways and consequences for health', *The Lancet*. World Health Organization. Published by Elsevier Ltd. All rights reserved., 395(10217), pp. 75–88. doi: 10.1016/S0140-6736(19)32472-9.

Wemheuer, F. *et al.* (2018) 'Tax4Fun2: a R-based tool for the rapid prediction of habitat-specific functional profiles and functional redundancy based on 16S rRNA gene marker gene sequences', *bioRxiv*. Cold Spring Harbor Laboratory, p. 490037. doi: 10.1101/490037.

White, F. A., Bhangoo, S. K. and Miller, R. J. (2005) 'Chemokines: Integrators of

pain and inflammation', *Nature Reviews Drug Discovery*. NIH Public Access, pp. 834–844. doi: 10.1038/nrd1852.

WHO (no date) *Fact sheets - Malnutrition*. Available at: https://www.who.int/news-room/fact-sheets/detail/malnutrition (Accessed: 29 June 2020).

Wikoff, W. R. *et al.* (2009) 'Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 106(10), pp. 3698–3703. doi: 10.1073/pnas.0812874106.

World Health Organisation (no date) *Musculoskeletal conditions*. doi: 10.1007/978-0-387-74337-0_7.

Wratten, E. (1995) 'Conceptualizing urban poverty', *Environment & Urbanization*. Sage PublicationsSage CA: Thousand Oaks, CA, 7(1), pp. 11–38. doi: 10.1177/095624789500700118.

Wu, G. D. *et al.* (2011) 'Linking long-term dietary patterns with gut microbial enterotypes', *Science*, 334(6052), pp. 105–108. doi: 10.1126/science.1208344.

Wu, J. *et al.* (2020) 'Association Between Dietary Selenium Intake and the Prevalence of Nonalcoholic Fatty Liver Disease: A Cross-Sectional Study', *Journal of the American College of Nutrition*. J Am Coll Nutr, 39(2), pp. 103–111. doi: 10.1080/07315724.2019.1613271.

Xu, L. *et al.* (2022) 'The Role and Mechanisms of Selenium Supplementation on Fatty Liver-Associated Disorder', *Antioxidants*. Multidisciplinary Digital Publishing Institute (MDPI). doi: 10.3390/antiox11050922.

Yang, H. H. *et al.* (2021) 'Serum iron and risk of nonalcoholic fatty liver disease and advanced hepatic fibrosis in US adults', *Scientific reports*. Sci Rep, 11(1). doi: 10.1038/S41598-021-89991-X.

Yang, M. H., Sung, J. and Gwak, G. Y. (2016) 'The associations between apolipoprotein B, A1, and the B/A1 ratio and nonalcoholic fatty liver disease in both normal-weight and overweight Korean population', *Journal of Clinical Lipidology*. Elsevier Ltd, 10(2), pp. 289–298. doi: 10.1016/j.jacl.2015.11.017.

Yang, S. and Chang, M. C. (2019) 'Chronic pain: Structural and functional changes in brain structures and associated negative affective states', *International Journal of Molecular Sciences*. Multidisciplinary Digital Publishing Institute (MDPI), p. 3130. doi: 10.3390/ijms20133130.

Yang, Z. *et al.* (2016) 'Plasma selenium levels and nonalcoholic fatty liver disease in Chinese adults: a cross-sectional analysis', *Scientific Reports 2016 6:1*. Nature Publishing Group, 6(1), pp. 1–8. doi: 10.1038/srep37288.

Yanoff, L. B. *et al.* (2007) 'Inflammation and iron deficiency in the hypoferremia of obesity', *International Journal of Obesity*. NIH Public Access, 31(9), pp. 1412–1419. doi: 10.1038/sj.ijo.0803625.

Yatsunenko, T. *et al.* (2012) 'Human gut microbiome viewed across age and geography', *Nature*. NIH Public Access, pp. 222–227. doi: 10.1038/nature11053.

Ye, J. *et al.* (2018) 'Butyrate Protects Mice Against Methionine–Choline-Deficient Diet-Induced Non-alcoholic Steatohepatitis by Improving Gut Barrier Function, Attenuating Inflammation and Reducing Endotoxin Levels', *Frontiers in Microbiology*. Frontiers Media SA, 9, p. 1967. doi: 10.3389/fmicb.2018.01967.

Yoshimatsu, S. *et al.* (2012) 'Hypophosphataemia among severely-malnourished children: Case series', *Journal of Health, Population and Nutrition*. BioMed Central, 30(4), pp. 491–494. doi: 10.3329/jhpn.v30i4.13419.

Young, V. B. (2017) 'The role of the microbiome in human health and disease: An introduction for clinicians', *BMJ (Online)*, 356. doi: 10.1136/bmj.j831.

Younossi, Z. *et al.* (2018) 'Global burden of NAFLD and NASH: Trends, predictions, risk factors and prevention', *Nature Reviews Gastroenterology and Hepatology*. Nature Publishing Group, pp. 11–20. doi: 10.1038/nrgastro.2017.109.

Yu, D. *et al.* (2021) 'The adverse metabolic effects of branched-chain amino acids are mediated by isoleucine and valine', *Cell Metabolism*. Cell Press, 33(5), pp. 905-922.e6. doi: 10.1016/J.CMET.2021.03.025.

Zhai, G. *et al.* (2010) 'Serum branched-chain amino acid to histidine ratio: A novel metabolomic biomarker of knee osteoarthritis', *Annals of the Rheumatic Diseases*. BMJ Publishing Group Ltd, 69(6), pp. 1227–1231. doi: 10.1136/ard.2009.120857.

Zhang, W. *et al.* (2015) 'Relationship between blood plasma and synovial fluid metabolite concentrations in patients with osteoarthritis', *Journal of Rheumatology*. Journal of Rheumatology, 42(5), pp. 859–865. doi: 10.3899/jrheum.141252.

Zhang, W. *et al.* (2016) 'Metabolomic analysis of human plasma reveals that arginine is depleted in knee osteoarthritis patients', *Osteoarthritis and Cartilage*. W.B. Saunders Ltd, 24(5), pp. 827–834. doi: 10.1016/j.joca.2015.12.004.

Zhang, Z. *et al.* (2016) 'Causal mediation analysis in the context of clinical research', *Annals of Translational Medicine*. AME Publishing Company, 4(21), pp. 9–9. doi: 10.21037/ATM.2016.11.11.

Zhang, Z., Chen, X. and Cui, B. (2021) 'Modulation of the fecal microbiome and metabolome by resistant dextrin ameliorates hepatic steatosis and mitochondrial abnormalities in mice', *Food & Function*. The Royal Society of Chemistry, 12(10), pp. 4504–4518. doi: 10.1039/D1FO00249J.

Zhou, D. *et al.* (2018) 'Sodium butyrate reduces high-fat diet-induced non-alcoholic steatohepatitis through upregulation of hepatic GLP-1R expression',

Experimental and Molecular Medicine. Nature Publishing Group, 50(12), p. 157. doi: 10.1038/s12276-018-0183-1.

Zhou, F. *et al.* (2021) 'Short-chain fatty acids contribute to neuropathic pain via regulating microglia activation and polarization', *Molecular pain*. Mol Pain, 17. doi: 10.1177/1744806921996520.

Zhou, X. *et al.* (2020) 'Metabolomics analysis identifies lysine and taurine as candidate prognostic biomarkers for AML-M2 patients', *International Journal of Hematology*. Springer, 111(6), pp. 761–770. doi: 10.1007/s12185-020-02836-7.

Zhou, Z. *et al.* (2020) 'Genome diversification in globally distributed novel marine Proteobacteria is linked to environmental adaptation', *ISME Journal*. Nature Publishing Group, 14(8), pp. 2060–2077. doi: 10.1038/s41396-020-0669-4.

Zierer, J. *et al.* (2018) 'The fecal metabolome as a functional readout of the gut microbiome', *Nature Genetics*. Nature Publishing Group, 50(6), pp. 790–795. doi: 10.1038/s41588-018-0135-7.

Appendices

Appendix 1: Commands for Microbiome Sequence Analysis using QIIME2 version 2019.10

Step 1: Prepare Metadata file

- **If Source is demultiplexed Raw Sequences**

Step 2: Prepare Manifest File

Step 3 : Raw Reads to demux input file with manifest file (prepared)
qiime tools import --type 'SampleData[PairedEndSequencesWithQuality]' --input-path path and name of Manifest file.txt --output-path path and name of output file.qza --input-format PairedEndFastqManifestPhred33V2

- **If Source is biom file from QIIME1**

Step 3 : Raw Reads to demux input file with biom file
qiime tools import --input-path - path and name of input file. --type 'FeatureTable[Frequency]' --input-format BIOMV210Format --output-path path and name of output file.fastq

qiime tools import --input-path - path and name of input file. --type 'FeatureData[Taxonomy]' --input-format BIOMV210Format --output-path path and name of output file.fastq

- **If Source is Raw Reads in one file**

Step 3 : Raw Reads to demux files with bbmap

Install bbmap and run following commands

- sed '/^[+@]/ s/ .*//' input-file- path and name of input file.fastq > path and name of output file.fastq
- demuxbyname.sh in= path and name of input file.fastq out=out_%.fastq prefixmode=t delimiter='_' outu=unmatched-reads.fastq stats=split-statistics.txt

Step 4 : Sequence quality control

qiime quality-filter q-score-joined --i-demux path and name of input file d.qza --o-filtered-sequences path and name of output file.qza --o-filter-stats path and name of output file.qza

Step 5 : Removing Primers and Adaptor sequence

qiime cutadapt trim-paired --i-demultiplexed-sequences paired-end-demux.qza --p-front-f forward primer sequence --p-front-r reverse primer sequence --o-trimmed-sequences path and name of output file.qza

Step 6 : Denoising data using DADA2

qiime dada2 denoise-paired --i-demultiplexed-seqs path and name of input file.qza --p-trim-left-f 0 --p-trim-left-r 0 --p-trunc-len-f length to seq to be retained --p-trunc-len-r length to seq to be retained --o-table path and name of output file.qza --o-representative-sequences path and name of output file.qza --o-denoising-stats denoising- path and name of output file.qza --p-n-threads 0

Step 7 : Feature Table Summary

qiime feature-table summarize --i-table path and name of input file.qza --o-path and name of output file.qzv --m-sample-metadata-file path and name of metadata file.tsv

Step 8 : Filtering data

qiime feature-table filter-samples --i-table path and name of input file.qza --m-metadata-file path and name of metadata file.tsv --p-where "[variable name]='variable value'" --o-filtered-table path and name of output file.qza

Step 9 : Phylogeny Estimations

qiime phylogeny align-to-tree-mafft-fasttree --i-sequences path and name of input file.qza --o-alignment path and name of output file.qza --o-masked-alignment path and name of output file.qza --o-tree path and name of output file.qza --o-rooted-tree path and name of output file.qza

Step 9 : Core metrices calculations

qiime diversity core-metrics-phylogenetic --i-phylogeny path and name of input file.qza --i-table path and name of input file.qza --p-sampling-depth threshold value for rarefaction --m-metadata-file path and name of metadata file.tsv --output-dir core-metrics-results

Step 9 : Alpha Diversity rarefactions

qiime diversity alpha-rarefaction --i-table path and name of input file.qza --p-max-depth threshold value for rarefaction --m-metadata-file path and name of metadata file.tsv --p-metrics 'Metrices method' --p-metrics 'Metrices method' --o-visualization path and name of output file.qzv --p-steps 10 --p-iterations 50

Step 10: Beta Diversity rarefactions

qiime diversity beta-rarefaction --i-table path and name of input file.qza --i-phylogeny path and name of input file.qza --p-metric 'Metrices method' --m-path and name of metadata file.tsv --p-sampling-depth threshold value for

rarefaction --p-iterations 50 --o-visualization path and name of output file.qzv --p-clustering-method 'Clustering method'

Step 11: Install Taxonomy Classifier
wget -O "silva-132-99-515-806-nb-classifier.qza" "https://data.qiime2.org/2019.10/common/silva-132-99-515-806-nb-classifier.qza"

Step 12: Summarise Taxonomies
qiime metadata tabulate --m-input-file taxonomy_515-806.qza --o-visualization taxonomy_515-806.qzv

Step 13 : Assigning taxonomy
qiime feature-classifier classify-sklearn --i-classifier classifier-trainer file.qza --i-reads path and name of input file.qza --o-classification path and name of output file.qza --p-n-jobs 4

Step 14: Combining feature table and taxonomy
qiime taxa collapse --i-table path and name of input file.qza --i-taxonomy generated taxonomy classifier file.qza --p-level level of taxa --o-collapsed-table path and name of output file.qza

Step 15: Rarefying feature table
qiime feature-table rarefy --i-table path and name of input file.qza --p-sampling-depth threshold value for rarefaction --o-rarefied-table path and name of output file.qza

Appendix 3: Supplementary Material Supporting Chapter 3

Appendix 3.1: Diet Data Collection Form

FOOD QUESTIONNAIRE

PERSONAL DATA

NAME:

SEX:

AGE:

Height:

Weight:

Date of Birth:

BMI:

IBM:

Occupation:

Economic Status:

Qualification:

No. of Family Members:

Co-morbidities:

Contact Number:

Food pattern: vegetarian / non-vegetarian

24 HOUR DIET RECALL METHOD

Food	Amount (Serving size)	Sodium	Grains	Vegetables	Fruits	Milk &	Meat, Fish, Poultry	Nuts, Seeds, Legumes	Oils (Fish Oil)	Sweets, Added Sugars
Breakfast										
Lunch										
Snack										
Dinner										
Day's Total										

FOOD FREQUENCY TABLE

Food Group	Daily	Once a week	Twice a week	Fortnightly	Monthly
Cereals					
Wheat					
Ragi					
Rice					
Rava					
Bread					
Vermiceli					
Noodles					
Any Other					
Pulses					
Tur dal					
Moong dal					
Urd dal					
Green gram					
Bengal gram					
Avare					
Soyabean					
Sprouted					
Any Other					
Milk and Milk Products					
Milk					
Curd					
Buttermilk					
Ghee					
Paneer					
Cheese					
Any other					
Roots and Tubers					
Potato					
Carrot					
Radish					
Beetroot					
Knol Knol					
Garlic					
Onion					

Food Group	Daily	Once a week	Twice a week	Fortnightly	Monthly
Sugar & Sweetners					
Sugar					
Jaggery					
Honey					
Non-Veg					
Chicken					
Fish					
Mutton					
Egg					
Pork					
Oil					
Sunflower oil					
Groundnut oil					
Olive oil					
Mustard oil					
Rice bran oil					
Nuts & oil seeds					
Cashewnut					
AlmondCoconut					
Groundnut					
Mutard					
Gingelly					
Miscellaneous					
Biscuits					
Chips					
Chocolate					
Ice creams					
Soft drink					
Sweet					
Any Other					

Food Group	Daily	Once a week	Twice a week	Fortnightly	Monthly
Green Leafy Vegetables					
Spinach					
Fenugreek					
Amaranth					
Lettuce					
Shepu					
Drumstick leaves					
Coriander					
Mint					
Curry Leave					
Any Other					
Other vegetables					
Beans					
Brinjal					
Ladies finger					
Gourds					
Pumpkin					
Cabbage					
Cauliflour					
Cluster bean					
Capsicum					
Cucumber					
Tomato					
Drumstick					
Any Other					
Fruit					
Banana					
Apple					
Pomegranate					
Chickoo					
Margo					
Mosambi					
orange					
Papaya					
Any other					

Appendix 3.2: Additional Information on dietary details of the participants collected from the 24-hour recall food frequency questionnaire

ID	Caloric Intake (kcal/day)	Carbohydrates (g/day)	Protein (g/day)	Fat (g/day)	Location	Status	Sample Arm
1	2100	320	55	30	Rural	nourished	Rural nourished
2	1748	310	43	32	Rural	nourished	Rural nourished
3	2010	279	47	29	Rural	nourished	Rural nourished
4	1921	281	53	33	Rural	nourished	Rural nourished
5	1598	302	50	35	Rural	nourished	Rural nourished
6	2012	339	59	31	Rural	nourished	Rural nourished
7	1823	310	41	30.6	Rural	nourished	Rural nourished
8	760	210	50	28	Rural	under-nourished	Rural undernourished
9	1250	291	55	33	Rural	nourished	Rural nourished
10	910	221	40	30	Rural	under-nourished	Rural undernourished
11	1445	230	65	35.2	Rural	nourished	Rural nourished
12	1060	190	35	28	Rural	nourished	Rural nourished
13	1920	320	62	39	Rural	nourished	Rural nourished
14	1548	280	70	31	Rural	nourished	Rural nourished
15	780	273	41	33	Rural	under-nourished	Rural undernourished
16	802	155	32	27	Rural	under-nourished	Rural undernourished

17	1100	365	32	27	Rural	nourished	Rural nourished
18	890	203	36	28	Rural	under-nourished	Rural undernourished
19	910	198	37	24	Rural	under-nourished	Rural undernourished
20	1010	182	41	33	Rural	under-nourished	Rural undernourished
21	710	268	53	37	Rural	under-nourished	Rural undernourished
22	910	178	35	31	Rural	under-nourished	Rural undernourished
23	870	210	39	28	Rural	under-nourished	Rural undernourished
24	954	190	52	26	Rural	under-nourished	Rural undernourished
25	810	222	49	33	Rural	under-nourished	Rural undernourished
26	2200	310	55	36	urban	nourished	Urban nourished
27	2010	302	51	31	urban	nourished	Urban nourished
28	1870	341	47	29	urban	nourished	Urban nourished
29	2890	319	59	34	urban	nourished	Urban nourished
30	1021	210	44	28	urban	under-nourished	Urban undernourished
31	1097	205	42	27	urban	under-nourished	Urban undernourished
32	1590	378	45	31	urban	nourished	Urban nourished
33	2120	298	55	26	urban	nourished	Urban nourished
34	1780	345	56	27	urban	nourished	Urban nourished
35	823	210	33	29	urban	under-nourished	Urban undernourished
36	1178	209	32	33	urban	under-nourished	Urban undernourished

37	990	199	35	31	urban	under-nourished	Urban undernourished
38	1100	182	36	28	urban	under-nourished	Urban undernourished
39	2200	288	65	37	urban	nourished	Urban nourished
40	925	135	30	32	urban	nourished	Urban nourished
41	1145	387	36	33	urban	nourished	Urban nourished
42	2200	379	61	26	urban	nourished	Urban nourished
43	1200	201	42	22	urban	under-nourished	Urban undernourished
44	1920	200	39	33	urban	under-nourished	Urban undernourished
45	1200	190	40	23	urban	under-nourished	Urban undernourished
46	1890	280	70	31	urban	nourished	Urban nourished
47	1287	219	46	27	urban	under-nourished	Urban undernourished
48	840	198	42	28	urban	under-nourished	Urban undernourished

Appendix 4: Supplementary Material Supporting Chapter 4

Appendix 4.1: Biomarkers identified using Biocrates metabolic platform

Metabolites	Abbreviation	Group
TMAO	Tri-Methyl Amine oxide	Amine Oxides
Ala	Alanine	Amino acids
Arg	Arginine	Amino acids
Asn	Asparagine	Amino acids
Asp	Aspartate	Amino acids
Cys	Cystine	Amino acids
Gln	Glutamine	Amino acids
Glu	Glutamate	Amino acids
Gly	Glycine	Amino acids
His	Histidine	Amino acids
Ile	Isoleucine	Amino acids
Leu	Leucine	Amino acids
Lys	Lysine	Amino acids
Met	Methionine	Amino acids
Phe	Phenylalanine	Amino acids
Pro	Proline	Amino acids
Ser	Serine	Amino acids
Thr	Threonine	Amino acids
Trp	Tryptophan	Amino acids
Tyr	Tyrosine	Amino acids
Val	Valine	Amino acids
1.Met.His	1-Methyl Histidine	Amino acids Related
13.Met.His	3-Methyl Histidine	Amino acids Related
5.AVA	5-Amino Valeric Acid	Amino acids Related
AABA	α-Aminobutyric acid	Amino acids Related
Ac.Orn	Acyl ornithine	Amino acids Related
ADMA	Asymmetric dimethylarginine	Amino acids Related
alpha.AAA	α-Aminoadipic acid	Amino acids Related
Anserine	Anserine	Amino acids Related
BABA	β-Aminobutyric acid	Amino acids Related
Betaine	Betaine	Amino acids Related
c4.OH.Pro	cis-4-Hydroxyproline	Amino acids Related
Carnosine	Carnosine	Amino acids Related
Cit	Citrulline	Amino acids Related
Creatinine	Creatinine	Amino acids Related
Cystine	Cystine	Amino acids Related
DOPA	Dihydroxyphenylalanine	Amino acids Related
HArg	Homoarginine	Amino acids Related
HCys	Homocystine	Amino acids Related
Kynurenine	Kynurenine	Amino acids Related
Met.SO	Methionine Sulfoxide	Amino acids Related
Nitro.Tyr	Nitrotyrosine	Amino acids Related
Orn	Ornithine	Amino acids Related
PAG	Phenylacetylglycine	Amino acids Related
PheAlaBetaine	Phenylalanine Betaine	Amino acids Related

ProBetaine	Proline Betaine	Amino acids Related
Sarcosine	Sarcosine	Amino acids Related
SDMA	Symmetric dimethylarginine	Amino acids Related
t4.OH.Pro	trans-4-Hydroxyproline	Amino acids Related
Taurine	Taurine	Amino acids Related
TrpBetaine	Tryptophan Betaine	Amino acids Related
Ala	Alanine	Amino acids
CA	Cholic Acid	Bile Acids
CDCA	Chenodeoxycholic acid	Bile Acids
DCA	Deoxycholic acid	Bile Acids
GCA	Glycocholic acid	Bile Acids
GCDCA	Glycodeoxycholic acid	Bile Acids
GDCA	Glycochenodeoxycholic acid	Bile Acids
GLCA	Glycolithocholic acid	Bile Acids
GLCAS	Glycolithocholic acid sulfate	Bile Acids
GUDCA	Glycoursodeoxycholic acid	Bile Acids
TCA	Taurocholic acid	Bile Acids
TCDCA	Taurochenodeoxycholic acid	Bile Acids
TDCA	Taurodeoxycholic acid	Bile Acids
TLCA	Taurolithocholic acid	Bile Acids
TMCA	Tauromurocholic acid	Bile Acids
beta.Ala	β-Alanine	Biogenic Amines
Dopamine	Dopamine	Biogenic Amines
GABA	γ-Aminobutyric acid	Biogenic Amines
Histamine	Histamine	Biogenic Amines
PEA	Phenylethylamine	Biogenic Amines
Putrescine	Putrescine	Biogenic Amines
Serotonin	Serotonin	Biogenic Amines
Spermidine	Spermidine	Biogenic Amines
Spermine	Spermine	Biogenic Amines
AconAcid	Aconitic Acid	Carboxylic Acids
DiCA (12:0)	Dodecanedioic acid	Carboxylic Acids
DiCA (14.0)	Tetradecanedioic acid	Carboxylic Acids
HipAcid	Hippuric Acid	Carboxylic Acids
Lac	Lactic Acid	Carboxylic Acids
OH.GlutAcid	3-Hydroxyglutaric Acid	Carboxylic Acids
Suc	Succinic Acid	Carboxylic Acids
p.Cresol.SO4	p-Cresol sulfate	Cresols
AbsAcid	Abscisic acid	Hormones
Cortisol	Cortisol	Hormones
Cortisone	Cortisone	Hormones
DHEAS	Dehydroepiandrosterone sulfate	Hormones
3.IAA	3-Indoleacetic acid	Indoles Derivatives
3.IPA	3-Indole Propionic Acid	Indoles Derivatives
Ind.SO4	Indoxyl sulfate3	Indoles Derivatives
Indole	Indole	Indoles Derivatives
Hypoxanthine	Hypoxanthine	Nucleobases Related
Xanthine	Xanthine	Nucleobases Related
Choline	Choline	Vitamins & Cofactors

Appendix 5: Supplementary Material Supporting Chapter 5

Appendix 5.1: Biomarkers identified using Nightingale NMR-metabolic platform

Biomarker name	Abbreviation	Unit	Group
Total cholesterol	Total-C	mmol/l	Cholesterol
VLDL cholesterol	VLDL-C	mmol/l	Cholesterol
Remnant cholesterol (non-HDL, non-LDL -cholesterol)	Remnant-C	mmol/l	Cholesterol
LDL cholesterol	LDL-C	mmol/l	Cholesterol
HDL cholesterol	HDL-C	mmol/l	Cholesterol
HDL2 cholesterol	HDL2-C	mmol/l	Cholesterol
HDL3 cholesterol	HDL3-C	mmol/l	Cholesterol
Total esterified cholesterol	Esterified-C	mmol/l	Cholesterol
Total free cholesterol	Free-C	mmol/l	Cholesterol
Total triglycerides	Total-TG	mmol/l	Glycerides and phospholipids
Triglycerides in VLDL	VLDL-TG	mmol/l	Glycerides and phospholipids
Triglycerides in LDL	LDL-TG	mmol/l	Glycerides and phospholipids
Triglycerides in HDL	HDL-TG	mmol/l	Glycerides and phospholipids
Phosphoglycerides	Phosphoglyc	mmol/l	Glycerides and phospholipids
Ratio of triglycerides to phosphoglycerides	TG/PG	ratio	Glycerides and phospholipids
Total cholines	Cholines	mmol/l	Glycerides and phospholipids
Phosphatidylcholines	Phosphatidylc	mmol/l	Glycerides and phospholipids
Sphingomyelins	Sphingomyelins	mmol/l	Glycerides and phospholipids

Apolipoprotein B	ApoB	g/l	Apolipoproteins
Apolipoprotein A1	ApoA1	g/l	Apolipoproteins
Ratio of apolipoprotein B to apolipoprotein A1	ApoB/ApoA1	ratio	Apolipoproteins
Total fatty acids	Total-FA	mmol/l	Fatty acids
Degree of unsaturation	Unsaturation	degree	Fatty acids
Omega-3 fatty acids	Omega-3	mmol/l	Fatty acids
Omega-6 fatty acids	Omega-6	mmol/l	Fatty acids
Polyunsaturated fatty acids	PUFA	mmol/l	Fatty acids
Monounsaturated fatty acids	MUFA	mmol/l	Fatty acids
Saturated fatty acids	SFA	mmol/l	Fatty acids
Linoleic acid	LA	mmol/l	Fatty acids
Docosahexaenoic acid	DHA	mmol/l	Fatty acids
Ratio of omega-3 fatty acids to total fatty acids	Omega-3 %	%	Fatty acids
Ratio of omega-6 fatty acids to total fatty acids	Omega-6 %	%	Fatty acids
Ratio of polyunsaturated fatty acids to total fatty acids	PUFA %	%	Fatty acids
Ratio of monounsaturated fatty acids to total fatty acids	MUFA %	%	Fatty acids
Ratio of saturated fatty acids to total fatty acids	SFA %	%	Fatty acids
Ratio of linoleic acid to total fatty acids	LA %	%	Fatty acids
Ratio of docosahexaenoic acid to total fatty acids	DHA %	%	Fatty acids
Alanine	Ala	mmol/l	Amino acids
Glutamine	Gln	mmol/l	Amino acids
Glycine	Gly	mmol/l	Amino acids
Histidine	His	mmol/l	Amino acids
Isoleucine	Ile	mmol/l	Amino acids

Leucine	Leu	mmol/l	Amino acids
Valine	Val	mmol/l	Amino acids
Phenylalanine	Phe	mmol/l	Amino acids
Tyrosine	Tyr	mmol/l	Amino acids
Glucose	Glucose	mmol/l	Glycolysis related metabolites
Lactate	Lactate	mmol/l	Glycolysis related metabolites
Pyruvate	Pyruvate	mmol/l	Glycolysis related metabolites
Citrate	Citrate	mmol/l	Glycolysis related metabolites
Glycerol	Glycerol	mmol/l	Glycolysis related metabolites
3-Hydroxybutyrate	Acetate	mmol/l	Ketone bodies
Acetate	Acetoacetate	mmol/l	Ketone bodies
Acetoacetate	bOHbutyrate	mmol/l	Ketone bodies
Creatinine	Creatinine	mmol/l	Fluid balance
Albumin	Albumin	signal area	Fluid balance
Glycoprotein acetyls	GlycA	mmol/l	Inflammation
Concentration of chylomicrons and extremely large VLDL particles	XXL-VLDL-P	mol/l	Lipoprotein subclasses
Total lipids in chylomicrons and extremely large VLDL	XXL-VLDL-L	mmol/l	Lipoprotein subclasses
Phospholipids in chylomicrons and extremely large VLDL	XXL-VLDL-PL	mmol/l	Lipoprotein subclasses
Cholesterol in chylomicrons and extremely large VLDL	XXL-VLDL-C	mmol/l	Lipoprotein subclasses
Cholesteryl esters in chylomicrons and extremely large VLDL	XXL-VLDL-CE	mmol/l	Lipoprotein subclasses
Free cholesterol in chylomicrons and extremely large VLDL	XXL-VLDL-FC	mmol/l	Lipoprotein subclasses
Triglycerides in chylomicrons and extremely large VLDL	XXL-VLDL-TG	mmol/l	Lipoprotein subclasses
Concentration of very large VLDL particles	XL-VLDL-P	mol/l	Lipoprotein subclasses
Total lipids in very large VLDL	XL-VLDL-L	mmol/l	Lipoprotein subclasses

Phospholipids in very large VLDL	XL-VLDL-PL	mmol/l	Lipoprotein subclasses
Cholesterol in very large VLDL	XL-VLDL-C	mmol/l	Lipoprotein subclasses
Cholesteryl esters in very large VLDL	XL-VLDL-CE	mmol/l	Lipoprotein subclasses
Free cholesterol in very large VLDL	XL-VLDL-FC	mmol/l	Lipoprotein subclasses
Triglycerides in very large VLDL	XL-VLDL-TG	mmol/l	Lipoprotein subclasses
Concentration of large VLDL particles	L-VLDL-P	mol/l	Lipoprotein subclasses
Total lipids in large VLDL	L-VLDL-L	mmol/l	Lipoprotein subclasses
Phospholipids in large VLDL	L-VLDL-PL	mmol/l	Lipoprotein subclasses
Cholesterol in large VLDL	L-VLDL-C	mmol/l	Lipoprotein subclasses
Cholesteryl esters in large VLDL	L-VLDL-CE	mmol/l	Lipoprotein subclasses
Free cholesterol in large VLDL	L-VLDL-FC	mmol/l	Lipoprotein subclasses
Triglycerides in large VLDL	L-VLDL-TG	mmol/l	Lipoprotein subclasses
Concentration of medium VLDL particles	M-VLDL-P	mol/l	Lipoprotein subclasses
Total lipids in medium VLDL	M-VLDL-L	mmol/l	Lipoprotein subclasses
Phospholipids in medium VLDL	M-VLDL-PL	mmol/l	Lipoprotein subclasses
Cholesterol in medium VLDL	M-VLDL-C	mmol/l	Lipoprotein subclasses
Cholesteryl esters in medium VLDL	M-VLDL-CE	mmol/l	Lipoprotein subclasses
Free cholesterol in medium VLDL	M-VLDL-FC	mmol/l	Lipoprotein subclasses
Triglycerides in medium VLDL	M-VLDL-TG	mmol/l	Lipoprotein subclasses
Concentration of small VLDL particles	S-VLDL-P	mol/l	Lipoprotein subclasses
Total lipids in small VLDL	S-VLDL-L	mmol/l	Lipoprotein subclasses
Phospholipids in small VLDL	S-VLDL-PL	mmol/l	Lipoprotein subclasses
Cholesterol in small VLDL	S-VLDL-C	mmol/l	Lipoprotein subclasses
Cholesteryl esters in small VLDL	S-VLDL-CE	mmol/l	Lipoprotein subclasses

Free cholesterol in small VLDL	S-VLDL-FC	mmol/l	Lipoprotein subclasses
Triglycerides in small VLDL	S-VLDL-TG	mmol/l	Lipoprotein subclasses
Concentration of very small VLDL particles	XS-VLDL-P	mol/l	Lipoprotein subclasses
Total lipids in very small VLDL	XS-VLDL-L	mmol/l	Lipoprotein subclasses
Phospholipids in very small VLDL	XS-VLDL-PL	mmol/l	Lipoprotein subclasses
Cholesterol in very small VLDL	XS-VLDL-C	mmol/l	Lipoprotein subclasses
Cholesteryl esters in very small VLDL	XS-VLDL-CE	mmol/l	Lipoprotein subclasses
Free cholesterol in very small VLDL	XS-VLDL-FC	mmol/l	Lipoprotein subclasses
Triglycerides in very small VLDL	XS-VLDL-TG	mmol/l	Lipoprotein subclasses
Concentration of IDL particles	IDL-P	mol/l	Lipoprotein subclasses
Total lipids in IDL	IDL-L	mmol/l	Lipoprotein subclasses
Phospholipids in IDL	IDL-PL	mmol/l	Lipoprotein subclasses
Cholesterol in IDL	IDL-C	mmol/l	Lipoprotein subclasses
Cholesteryl esters in IDL	IDL-CE	mmol/l	Lipoprotein subclasses
Free cholesterol in IDL	IDL-FC	mmol/l	Lipoprotein subclasses
Triglycerides in IDL	IDL-TG	mmol/l	Lipoprotein subclasses
Concentration of large LDL particles	L-LDL-P	mol/l	Lipoprotein subclasses
Total lipids in large LDL	L-LDL-L	mmol/l	Lipoprotein subclasses
Phospholipids in large LDL	L-LDL-PL	mmol/l	Lipoprotein subclasses
Cholesterol in large LDL	L-LDL-C	mmol/l	Lipoprotein subclasses
Cholesteryl esters in large LDL	L-LDL-CE	mmol/l	Lipoprotein subclasses
Free cholesterol in large LDL	L-LDL-FC	mmol/l	Lipoprotein subclasses
Triglycerides in large LDL	L-LDL-TG	mmol/l	Lipoprotein subclasses
Concentration of medium LDL particles	M-LDL-P	mol/l	Lipoprotein subclasses

Total lipids in medium LDL	M-LDL-L	mmol/l	Lipoprotein subclasses
Phospholipids in medium LDL	M-LDL-PL	mmol/l	Lipoprotein subclasses
Cholesterol in medium LDL	M-LDL-C	mmol/l	Lipoprotein subclasses
Cholesteryl esters in medium LDL	M-LDL-CE	mmol/l	Lipoprotein subclasses
Free cholesterol in medium LDL	M-LDL-FC	mmol/l	Lipoprotein subclasses
Triglycerides in medium LDL	M-LDL-TG	mmol/l	Lipoprotein subclasses
Concentration of small LDL particles	S-LDL-P	mol/l	Lipoprotein subclasses
Total lipids in small LDL	S-LDL-L	mmol/l	Lipoprotein subclasses
Phospholipids in small LDL	S-LDL-PL	mmol/l	Lipoprotein subclasses
Cholesterol in small LDL	S-LDL-C	mmol/l	Lipoprotein subclasses
Cholesteryl esters in small LDL	S-LDL-CE	mmol/l	Lipoprotein subclasses
Free cholesterol in small LDL	S-LDL-FC	mmol/l	Lipoprotein subclasses
Triglycerides in small LDL	S-LDL-TG	mmol/l	Lipoprotein subclasses
Concentration of very large HDL particles	XL-HDL-P	mol/l	Lipoprotein subclasses
Total lipids in very large HDL	XL-HDL-L	mmol/l	Lipoprotein subclasses
Phospholipids in very large HDL	XL-HDL-PL	mmol/l	Lipoprotein subclasses
Cholesterol in very large HDL	XL-HDL-C	mmol/l	Lipoprotein subclasses
Cholesteryl esters in very large HDL	XL-HDL-CE	mmol/l	Lipoprotein subclasses
Free cholesterol in very large HDL	XL-HDL-FC	mmol/l	Lipoprotein subclasses
Triglycerides in very large HDL	XL-HDL-TG	mmol/l	Lipoprotein subclasses
Concentration of large HDL particles	L-HDL-P	mol/l	Lipoprotein subclasses
Total lipids in large HDL	L-HDL-L	mmol/l	Lipoprotein subclasses
Phospholipids in large HDL	L-HDL-PL	mmol/l	Lipoprotein subclasses
Cholesterol in large HDL	L-HDL-C	mmol/l	Lipoprotein subclasses

Cholesteryl esters in large HDL	L-HDL-CE	mmol/l	Lipoprotein subclasses
Free cholesterol in large HDL	L-HDL-FC	mmol/l	Lipoprotein subclasses
Triglycerides in large HDL	L-HDL-TG	mmol/l	Lipoprotein subclasses
Concentration of medium HDL particles	M-HDL-P	mol/l	Lipoprotein subclasses
Total lipids in medium HDL	M-HDL-L	mmol/l	Lipoprotein subclasses
Phospholipids in medium HDL	M-HDL-PL	mmol/l	Lipoprotein subclasses
Cholesterol in medium HDL	M-HDL-C	mmol/l	Lipoprotein subclasses
Cholesteryl esters in medium HDL	M-HDL-CE	mmol/l	Lipoprotein subclasses
Free cholesterol in medium HDL	M-HDL-FC	mmol/l	Lipoprotein subclasses
Triglycerides in medium HDL	M-HDL-TG	mmol/l	Lipoprotein subclasses
Concentration of small HDL particles	S-HDL-P	mol/l	Lipoprotein subclasses
Total lipids in small HDL	S-HDL-L	mmol/l	Lipoprotein subclasses
Phospholipids in small HDL	S-HDL-PL	mmol/l	Lipoprotein subclasses
Cholesterol in small HDL	S-HDL-C	mmol/l	Lipoprotein subclasses
Cholesteryl esters in small HDL	S-HDL-CE	mmol/l	Lipoprotein subclasses
Free cholesterol in small HDL	S-HDL-FC	mmol/l	Lipoprotein subclasses
Triglycerides in small HDL	S-HDL-TG	mmol/l	Lipoprotein subclasses
Average diameter for VLDL particles	VLDL size	nm	Lipoprotein particle sizes
Average diameter for LDL particles	LDL size	nm	Lipoprotein particle sizes
Average diameter for HDL particles	HDL size	nm	Lipoprotein particle sizes
Phospholipids to total lipids ratio in chylomicrons and extremely large VLDL	XXL-VLDL-PL %	%	Relative lipoprotein lipid concentrations
Cholesterol to total lipids ratio in chylomicrons and extremely large VLDL	XXL-VLDL-C %	%	Relative lipoprotein lipid concentrations

Cholesteryl esters to total lipids ratio in chylomicrons and extremely large VLDL	XXL-VLDL-CE %	%	Relative lipoprotein lipid concentrations
Free cholesterol to total lipids ratio in chylomicrons and extremely large VLDL	XXL-VLDL-FC %	%	Relative lipoprotein lipid concentrations
Triglycerides to total lipids ratio in chylomicrons and extremely large VLDL	XXL-VLDL-TG %	%	Relative lipoprotein lipid concentrations
Phospholipids to total lipids ratio in very large VLDL	XL-VLDL-PL %	%	Relative lipoprotein lipid concentrations
Cholesterol to total lipids ratio in very large VLDL	XL-VLDL-C %	%	Relative lipoprotein lipid concentrations
Cholesteryl esters to total lipids ratio in very large VLDL	XL-VLDL-CE %	%	Relative lipoprotein lipid concentrations
Free cholesterol to total lipids ratio in very large VLDL	XL-VLDL-FC %	%	Relative lipoprotein lipid concentrations
Triglycerides to total lipids ratio in very large VLDL	XL-VLDL-TG %	%	Relative lipoprotein lipid concentrations
Phospholipids to total lipids ratio in large VLDL	L-VLDL-PL %	%	Relative lipoprotein lipid concentrations
Cholesterol to total lipids ratio in large VLDL	L-VLDL-C %	%	Relative lipoprotein lipid concentrations
Cholesteryl esters to total lipids ratio in large VLDL	L-VLDL-CE %	%	Relative lipoprotein lipid concentrations
Free cholesterol to total lipids ratio in large VLDL	L-VLDL-FC %	%	Relative lipoprotein lipid concentrations

Triglycerides to total lipids ratio in large VLDL	L-VLDL-TG %	%	Relative lipoprotein lipid concentrations
Phospholipids to total lipids ratio in medium VLDL	M-VLDL-PL %	%	Relative lipoprotein lipid concentrations
Cholesterol to total lipids ratio in medium VLDL	M-VLDL-C %	%	Relative lipoprotein lipid concentrations
Cholesteryl esters to total lipids ratio in medium VLDL	M-VLDL-CE %	%	Relative lipoprotein lipid concentrations
Free cholesterol to total lipids ratio in medium VLDL	M-VLDL-FC %	%	Relative lipoprotein lipid concentrations
Triglycerides to total lipids ratio in medium VLDL	M-VLDL-TG %	%	Relative lipoprotein lipid concentrations
Phospholipids to total lipids ratio in small VLDL	S-VLDL-PL %	%	Relative lipoprotein lipid concentrations
Cholesterol to total lipids ratio in small VLDL	S-VLDL-C %	%	Relative lipoprotein lipid concentrations
Cholesteryl esters to total lipids ratio in small VLDL	S-VLDL-CE %	%	Relative lipoprotein lipid concentrations
Free cholesterol to total lipids ratio in small VLDL	S-VLDL-FC %	%	Relative lipoprotein lipid concentrations
Triglycerides to total lipids ratio in small VLDL	S-VLDL-TG %	%	Relative lipoprotein lipid concentrations
Phospholipids to total lipids ratio in very small VLDL	XS-VLDL-PL %	%	Relative lipoprotein lipid concentrations

Cholesterol to total lipids ratio in very small VLDL	XS-VLDL-C %	%	Relative lipoprotein lipid concentrations
Cholesteryl esters to total lipids ratio in very small VLDL	XS-VLDL-CE %	%	Relative lipoprotein lipid concentrations
Free cholesterol to total lipids ratio in very small VLDL	XS-VLDL-FC %	%	Relative lipoprotein lipid concentrations
Triglycerides to total lipids ratio in very small VLDL	XS-VLDL-TG %	%	Relative lipoprotein lipid concentrations
Phospholipids to total lipids ratio in IDL	IDL-PL %	%	Relative lipoprotein lipid concentrations
Cholesterol to total lipids ratio in IDL	IDL-C %	%	Relative lipoprotein lipid concentrations
Cholesteryl esters to total lipids ratio in IDL	IDL-CE %	%	Relative lipoprotein lipid concentrations
Free cholesterol to total lipids ratio in IDL	IDL-FC %	%	Relative lipoprotein lipid concentrations
Triglycerides to total lipids ratio in IDL	IDL-TG %	%	Relative lipoprotein lipid concentrations
Phospholipids to total lipids ratio in large LDL	L-LDL-PL %	%	Relative lipoprotein lipid concentrations
Cholesterol to total lipids ratio in large LDL	L-LDL-C %	%	Relative lipoprotein lipid concentrations
Cholesteryl esters to total lipids ratio in large LDL	L-LDL-CE %	%	Relative lipoprotein lipid concentrations

Free cholesterol to total lipids ratio in large LDL	L-LDL-FC %	%	Relative lipoprotein lipid concentrations
Triglycerides to total lipids ratio in large LDL	L-LDL-TG %	%	Relative lipoprotein lipid concentrations
Phospholipids to total lipids ratio in medium LDL	M-LDL-PL %	%	Relative lipoprotein lipid concentrations
Cholesterol to total lipids ratio in medium LDL	M-LDL-C %	%	Relative lipoprotein lipid concentrations
Cholesteryl esters to total lipids ratio in medium LDL	M-LDL-CE %	%	Relative lipoprotein lipid concentrations
Free cholesterol to total lipids ratio in medium LDL	M-LDL-FC %	%	Relative lipoprotein lipid concentrations
Triglycerides to total lipids ratio in medium LDL	M-LDL-TG %	%	Relative lipoprotein lipid concentrations
Phospholipids to total lipids ratio in small LDL	S-LDL-PL %	%	Relative lipoprotein lipid concentrations
Cholesterol to total lipids ratio in small LDL	S-LDL-C %	%	Relative lipoprotein lipid concentrations
Cholesteryl esters to total lipids ratio in small LDL	S-LDL-CE %	%	Relative lipoprotein lipid concentrations
Free cholesterol to total lipids ratio in small LDL	S-LDL-FC %	%	Relative lipoprotein lipid concentrations
Triglycerides to total lipids ratio in small LDL	S-LDL-TG %	%	Relative lipoprotein lipid concentrations

Phospholipids to total lipids ratio in very large HDL	XL-HDL-PL %	%	Relative lipoprotein lipid concentrations
Cholesterol to total lipids ratio in very large HDL	XL-HDL-C %	%	Relative lipoprotein lipid concentrations
Cholesteryl esters to total lipids ratio in very large HDL	XL-HDL-CE %	%	Relative lipoprotein lipid concentrations
Free cholesterol to total lipids ratio in very large HDL	XL-HDL-FC %	%	Relative lipoprotein lipid concentrations
Triglycerides to total lipids ratio in very large HDL	XL-HDL-TG %	%	Relative lipoprotein lipid concentrations
Phospholipids to total lipids ratio in large HDL	L-HDL-PL %	%	Relative lipoprotein lipid concentrations
Cholesterol to total lipids ratio in large HDL	L-HDL-C %	%	Relative lipoprotein lipid concentrations
Cholesteryl esters to total lipids ratio in large HDL	L-HDL-CE %	%	Relative lipoprotein lipid concentrations
Free cholesterol to total lipids ratio in large HDL	L-HDL-FC %	%	Relative lipoprotein lipid concentrations
Triglycerides to total lipids ratio in large HDL	L-HDL-TG %	%	Relative lipoprotein lipid concentrations
Phospholipids to total lipids ratio in medium HDL	M-HDL-PL %	%	Relative lipoprotein lipid concentrations
Cholesterol to total lipids ratio in medium HDL	M-HDL-C %	%	Relative lipoprotein lipid concentrations

Cholesteryl esters to total lipids ratio in medium HDL	M-HDL-CE %	%	Relative lipoprotein lipid concentrations
Free cholesterol to total lipids ratio in medium HDL	M-HDL-FC %	%	Relative lipoprotein lipid concentrations
Triglycerides to total lipids ratio in medium HDL	M-HDL-TG %	%	Relative lipoprotein lipid concentrations
Phospholipids to total lipids ratio in small HDL	S-HDL-PL %	%	Relative lipoprotein lipid concentrations
Cholesterol to total lipids ratio in small HDL	S-HDL-C %	%	Relative lipoprotein lipid concentrations
Cholesteryl esters to total lipids ratio in small HDL	S-HDL-CE %	%	Relative lipoprotein lipid concentrations
Free cholesterol to total lipids ratio in small HDL	S-HDL-FC %	%	Relative lipoprotein lipid concentrations
Triglycerides to total lipids ratio in small HDL	S-HDL-TG %	%	Relative lipoprotein lipid concentrations

Appendix 5.2: EPIC -NORFOLK-UK

CONFIDENTIAL STUDY NUMBER: ☐☐☐☐☐☐☐☐

FOOD FREQUENCY QUESTIONNAIRE

This questionnaire asks for some background information about you, especially about what you eat.

Your answers will be treated as strictly confidential and will be used only for medical research.

PLEASE COMPLETE USING BLACK INK/PEN

Date of birth: ☐☐ ☐☐ ☐☐☐☐

Please enter: **M if you are Male**
F if you are Female ☐

Please answer every question. If you are uncertain about how to answer a question then do the best you can, but please do not leave a question blank.

CAMB/PQ/6/1205

1. **YOUR DIET LAST YEAR**
For each food there is an amount shown, either a "medium serving" or a common household unit such as a slice or teaspoon. Please put a tick (✓) in the box to indicate how often, on average, you have eaten the specified amount of each food during the past year.

EXAMPLES:

For white bread the amount is one slice, so if you ate 4 or 5 slices a day, you should put a tick in the column headed "4-5 per day".

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
BREAD AND SAVOURY BISCUITS (one slice or biscuit)	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
White bread and rolls								✓	

For chips, the amount is a "medium serving", so if you had a helping of chips twice a week you should put a tick in the column headed "2-4 per week".

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
POTATOES, RICE AND PASTA (medium serving)	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Chips				✓					

For very seasonal fruits such as strawberries and raspberries you should estimate your average use when the fruits are in season, so if you ate strawberries or raspberries about once a week when they were in season you should put a tick in the column headed "once a week"

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
FRUIT (1 fruit or medium serving)	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Strawberries, raspberries, kiwi fruit			✓						

EXAMPLE

2

Please estimate your average food use as best you can, and please answer every question – do not leave ANY lines blank. PLEASE PUT A TICK (✓) ON EVERY LINE

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
MEAT AND FISH (medium serving)	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Beef: roast, steak, mince, stew or casserole									
Beefburgers									
Pork: roast, chops, stew or slices									
Lamb: roast, chops or stew									
Chicken or other poultry eg. turkey									
Bacon									
Ham									
Corned beef, Spam, luncheon meats									
Sausages									
Savoury pies, eg. meat pie, pork pie, pasties, steak & kidney pie, sausage rolls									
Liver, liver paté, liver sausage									
Fried fish in batter, as in fish and chips									
Fish fingers, fish cakes									
Other white fish, fresh or frozen, eg. cod, haddock, plaice, sole, halibut									
Oily fish, fresh or canned, eg. mackerel, kippers, tuna, salmon, sardines, herring									
Shellfish, eg. crab, prawns, mussels									
Fish roe, taramasalata									
	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day

Please check that you have a tick (✓) on EVERY line

3

PLEASE PUT A TICK (✓) ON EVERY LINE

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
BREAD AND SAVOURY BISCUITS (one slice or biscuit)	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
White bread and rolls									
Brown bread and rolls									
Wholemeal bread and rolls									
Cream crackers, cheese biscuits									
Crispbread, eg. Ryvita									
CEREALS (one bowl)									
Porridge, Readybrek									
Breakfast cereal such as cornflakes, muesli etc.									
POTATOES, RICE AND PASTA (medium serving)									
Boiled, mashed, instant or jacket potatoes									
Chips									
Roast potatoes									
Potato salad									
White rice									
Brown rice									
White or green pasta, eg. spaghetti, macaroni, noodles									
Wholemeal pasta									
Lasagne, moussaka									
Pizza									
	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day

Please check that you have a tick (✓) on EVERY line

4

PLEASE PUT A TICK (✓) ON EVERY LINE

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
DAIRY PRODUCTS AND FATS	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Single or sour cream (tablespoon)									
Double or clotted cream (tablespoon)									
Low fat yogurt, fromage frais (125g carton)									
Full fat or Greek yogurt (125g carton)									
Dairy desserts (125g carton)									
Cheese, eg. Cheddar, Brie, Edam (medium serving)									
Cottage cheese, low fat soft cheese (medium serving)									
Eggs as boiled, fried, scrambled, etc. (one)									
Quiche (medium serving)									
Low calorie, low fat salad cream(tablespoon)									
Salad cream, mayonnaise (tablespoon)									
French dressing (tablespoon)									
Other salad dressing (tablespoon)									
The following on bread or vegetables									
Butter (teaspoon)									
Block or hard margarine, eg. Stork, Krona (teaspoon)									
Polyunsaturated margarine, eg. Flora, sunflower, soya spreads (teaspoon)									
Soft margarines,including olive oil based and dairy spreads, eg. Blue Band, Clover (teaspoon)									
Low fat spreads (less than 60% fat), eg. Outline, Gold (teaspoon)									
Very low fat spread (less than 30% fat) (teaspoon)									
	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day

Please check that you have a tick (✓) on EVERY line

5

PLEASE PUT A TICK (✓) ON EVERY LINE

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
SWEETS AND SNACKS (medium serving)	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Sweet biscuits, chocolate , eg. digestive (one)									
Sweet biscuits, plain, eg. Nice, ginger (one)									
Cakes eg. fruit, sponge, home baked									
Cakes eg. fruit, sponge, ready made									
Buns, pastries eg. scones, flapjacks, home baked									
Buns, pastries eg. croissants, doughnuts, ready made									
Fruit pies, tarts, crumbles, home baked									
Fruit pies, tarts, crumbles, ready made									
Sponge puddings, home baked									
Sponge puddings, ready made									
Milk puddings, eg. rice, custard, trifle									
Ice cream, choc ices									
Chocolates, single or squares									
Chocolate snack bars eg. Mars, Crunchie									
Sweets, toffees, mints									
Sugar added to tea, coffee, cereal (teaspoon)									
Crisps or other packet snacks, eg. Wotsits									
Peanuts or other nuts									
SOUPS, SAUCES, AND SPREADS									
Vegetable soups (bowl)									
Meat soups (bowl)									
Sauces, eg. white sauce, cheese sauce, gravy (tablespoon)									
Tomato ketchup (tablespoon)									
Pickles, chutney (tablespoon)									
Marmite, Bovril (teaspoon)									
Jam, marmalade, honey (teaspoon)									
Peanut butter (teaspoon)									
	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day

Please check that you have a tick (✓) on EVERY line

6

PLEASE PUT A TICK (✓) ON EVERY LINE

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
DRINKS	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Tea (cup)									
Coffee, instant or ground (cup)									
Coffee, decaffeinated (cup)									
Coffee whitener, eg. Coffee-mate (teaspoon)									
Cocoa, hot chocolate (cup)									
Horlicks, Ovaltine (cup)									
Wine (glass)									
Beer, lager or cider (half pint)									
Port, sherry, vermouth, liqueurs (glass)									
Spirits, eg. gin, brandy, whisky, vodka (single)									
Low calorie or diet fizzy soft drinks (glass)									
Fizzy soft drinks, eg. Coca cola, lemonade (glass)									
Pure fruit juice (100%) eg. orange, apple juice (glass)									
Fruit squash or cordial (glass)									
FRUIT For seasonal fruits marked *, please estimate your average use when the fruit is in season									
Apples (1 fruit)									
Pears (1 fruit)									
Oranges, satsumas, mandarins (1 fruit)									
Grapefruit (half)									
Bananas (1 fruit)									
Grapes (medium serving)									
Melon (1 slice)									
* Peaches, plums, apricots (1 fruit)									
* Strawberries, raspberries, kiwi fruit (medium serving)									
Tinned fruit (medium serving)									
Dried fruit, eg. raisins, prunes (medium serving)									
	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day

Please check that you have a tick (✓) on EVERY line

7

PLEASE PUT A TICK (✓) ON EVERY LINE

FOODS AND AMOUNTS	AVERAGE USE LAST YEAR								
VEGETABLES Fresh, frozen or tinned (medium serving)	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Carrots									
Spinach									
Broccoli, spring greens, kale									
Brussels sprouts									
Cabbage									
Peas									
Green beans, broad beans, runner beans									
Marrow, courgettes									
Cauliflower									
Parsnips, turnips, swedes									
Leeks									
Onions									
Garlic									
Mushrooms									
Sweet peppers									
Beansprouts									
Green salad, lettuce, cucumber, celery									
Watercress									
Tomatoes									
Sweetcorn									
Beetroot									
Coleslaw									
Avocado									
Baked beans									
Dried lentils, beans, peas									
Tofu , soya meat, TVP, Vegeburger									
	Never or less than once/month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day

Please check that you have a tick (✓) on EVERY line

8

YOUR DIET LAST YEAR, continued

2. Are there any **OTHER** foods which you ate more than once a week? Yes ☐ No ☐

If YES, please list below

Food	Usual serving size	Number of times eaten each week

3. What type of milk did you most often use?

Select one only

Full cream/whole ☐ Semi-skimmed ☐
Skimmed ☐ Channel Islands, gold ☐
Dried milk ☐ Soya ☐
Other, specify ☐ None ☐

4. How much milk did you drink each day, including milk with tea, coffee, cereals etc?

None ☐ Three quarters of a pint ☐
Quarter of a pint ☐ One pint ☐
Half a pint ☐ More than one pint ☐

5. Did you usually eat breakfast cereal (excluding porridge and Ready Brek mentioned earlier)? Yes ☐ No ☐

If YES, which brand and type of breakfast cereal, including muesli, did you usually eat?
List the one or two types most often used

Brand e.g. Kellogg's	Type e.g. cornflakes

6. What kind of fat did you most often use for frying, roasting, grilling etc?

Select one only

Butter ☐ Solid vegetable fat ☐
Lard/dripping ☐ Margarine ☐
Vegetable oil ☐ None ☐

If you used vegetable oil, please give type eg. corn, sunflower ☐

7. What kind of fat did you most often use for baking cakes etc?

Select one only

Butter ☐ Solid vegetable fat ☐
Lard/dripping ☐ Margarine ☐
Vegetable oil ☐ None ☐

If you used margarine, please give name or type eg. Flora, Stork ☐

9

8. How often did you eat food that was fried at home?
   Daily ☐ 1-3 times a week ☐ 4-6 times a week ☐
   Less than once a week ☐ Never ☐

9. How often did you eat fried food away from home?
   Daily ☐ 1-3 times a week ☐ 4-6 times a week ☐
   Less than once a week ☐ Never ☐

10. What did you do with the visible fat on your meat?
    Ate most of the fat ☐ Ate as little as possible ☐
    Ate some of the fat ☐ Did not eat meat ☐

11. How often did you eat grilled or roast meat? ☐☐ times a week

12. How well cooked did you usually have grilled or roast meat?
    Well done /dark brown ☐ Lightly cooked/rare ☐
    Medium ☐ Did not eat meat ☐

13. How often did you add salt to food while cooking?
    Always ☐ Rarely ☐
    Usually ☐ Never ☐
    Sometimes ☐

14. How often did you add salt to any food at the table?
    Always ☐ Rarely ☐
    Usually ☐ Never ☐
    Sometimes ☐

15. Did you regularly use a salt substitute (eg LoSalt)? Yes ☐ No ☐
    If YES, which brand? _____

16. During the course of last year, on average, how many times a week did you eat the following foods?

Food type	Times/week	Portion size
Vegetables (not including potatoes)	☐☐	medium serving
Salads	☐☐	medium serving
Fruit and fruit products (not including fruit juice)	☐☐	medium serving or 1 fruit
Fish and fish products	☐☐	medium serving
Meat, meat products and meat dishes (including bacon, ham and chicken)	☐☐	medium serving

10

17. Have you taken any vitamins, minerals, fish oils, fibre or other food supplements during the past year?

- ☐ Yes
- ☐ No
- ☐ Sometimes
- ☐ Don't know

If **YES** or **SOMETIMES**, please complete the table below.

If you have taken more than 8 types of supplement please put the most frequently consumed brands first.

Example: If you take one tablet of vitamin C two times a day, please write '2' in the amount-column and tick (✓) the 'once a day' box. Most supplements mention a strength value (in our example 500mg), please write this information in the table.

Supplements				Average frequency for the past year Tick (✓) ONE box per line to show how often on average you took the amount consumed as mentioned in 'amount' column.					
Brand	Name	Strength (strength of the supple-ment for each tablet or capsule)	Amount (number of tablets, capsules or teaspoons taken in one day)	Never or less than once a month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day
Boots	High strength vitamin C	500mg	2 tablets						✓

EXAMPLE

Thank you for your help

11

Appendix 5.3: FFQs - TWINS-UK

«STUDY_NO»



TwinsUK

Questionnaire

For 24 years we have been a leader in health-related research.

Thank you for making this happen.

Please continue to help our research efforts by answering the enclosed questions.



YOUR DIET LAST YEAR

For each food there is an amount shown, either a "medium serving" or an example of a serving such as "a slice" or "a teaspoon". Please put a cross (×) in the box to indicate how often, **on average**, you have eaten a specified amount of food over the last 12 months. If you complete a box incorrectly please fill the wrong one. e.g. and put a cross in the correct box.



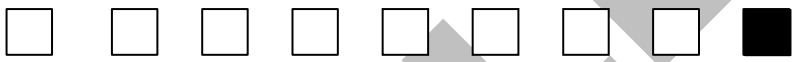
The rest of this page shows examples of how to fill in the questionnaire.

These questions are all about the food you have eaten over the last year.

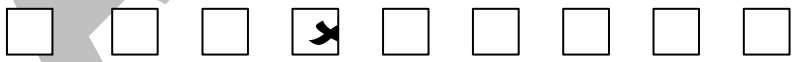
EXAMPLE 1

For white bread the serving size is one slice or roll, so if you ate 2 to 4 slices a week on average, over the last 12 months, you should put a cross in the column headed "2-4 per week".

FOODS AND AMOUNTS **AVERAGE USE IN THE LAST TWELVEMONTHS**



EXAMPLE 2



EXAMPLE 3

For very seasonal fruits such as strawberries and raspberries you should estimate your average intake in summer, so if you ate strawberries or raspberries about once a week in summer you should put a cross in the column headed "once a week".

FOODS AND AMOUNTS	AVERAGE USE IN THE LAST TWELVEMONTHS								
FRUIT	Never or less than once a	1-3 per	Once a	2-4 per	5-6 per	Once a	2-3 per	4-5 per	6+ per

EXAMPLE

(1 fruit or handful)

Strawberries, raspberries, kiwi fruit	☐	☐	☒	☐	☐	☐	☐	☐	☐

START OF DIET QUESTIONS

1a. On average do you think you eat the same, more or less than someone of your own age and sex?

☐ Same ☐ More ☐ Less

1b.On an average day how many portions of fruit do you eat? ☐

1c. On an average day how many portions of vegetables do you eat? ☐

Please estimate your average food use as best you can, and please answer every question– do not leave ANY lines blank. **PLEASE PUT A CROSS (×) ON EVERY LINE**

FOODS AND AMOUNTS AVERAGE USE IN THE LAST TWELVE MONTHS

	Never or less than once a month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day

DAIRY PRODUCTS AND FATS

Single or sour cream (tablespoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Double or clotted cream (tablespoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Low fat yoghurt, fromage frais (small pot)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Full fat or Greek yoghurt (small pot)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Dairy desserts (small pot) e.g. chocolate mousse, cream caramels	☐	☐	☐	☐	☐	☐	☐	☐	☐
Cheese, e.g. cheddar, brie, edam (matchbox size)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Low fat cheese e.g. reduced fat cheddar (matchbox size)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Cottage cheese, low fat soft cheese (2 tablespoons)	☐	☐	☐	☐	☐	☐	☐	☐	☐

Eggs as boiled, fried, scrambled, etc. (one)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Quiche (slice)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Low calorie, low fat salad cream (tablespoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Full fat salad cream, mayonnaise (tablespoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
French dressing (tablespoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Other salad dressing (tablespoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐

AVERAGE USE IN THE LAST TWELVEMONTHS

The following list refers to **DAIRY PRODUCTS AND FATS** that you put on bread or cooked vegetables.	Never or less than once a month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Butter (teaspoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Reduced fat butter (teaspoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Block margarine, e.g. Stork, Krona (teaspoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Polyunsaturated margarine, e.g. Flora sunflower (teaspoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Olive oil spread (teaspoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Other soft margarine, dairy spreads, e.g. Blue Band, Clover (teaspoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Low fat spread, e.g. outline, Gold (teaspoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Very low fat spread (teaspoon) e.g. Diet Flora	☐	☐	☐	☐	☐	☐	☐	☐	☐

	AVERAGE USE IN THE LAST TWELVEMONTHS								
Cholesterol lowering fat spreads e.g. Benzenol (teaspoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐

FRUIT- For seasonal fruits marked*, please estimate your average use in summer

Apples (1 fruit)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Pears (1 fruit)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Oranges, satsumas, mandarins	☐	☐	☐	☐	☐	☐	☐	☐	☐
Grapefruit (half)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Bananas (1 fruit)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Grapes (handful)	☐	☐	☐	☐	☐	☐	☐	☐	☐

	AVERAGE USE IN THE LAST TWELVEMONTHS								
Melon (1 slice)	☐	☐	☐	☐	☐	☐	☐	☐	☐
*Peaches, plums, apricots (1 fruit)	☐	☐	☐	☐	☐	☐	☐	☐	☐
*Strawberries, raspberries, other berries, kiwi fruit (one fruit or handful)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Tinned fruit (handful)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Dried fruit, e.g. raisins, prunes (heaped tablespoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐

AVERAGE USE IN THE LAST TWELVEMONTHS

VEGETABLES Fresh, frozen or tinned (handful)	Never or less than once a month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Carrots	☐	☐	☐	☐	☐	☐	☐	☐	☐
Spinach	☐	☐	☐	☐	☐	☐	☐	☐	☐
Broccoli, spring green, kale	☐	☐	☐	☐	☐	☐	☐	☐	☐
Brussels sprouts	☐	☐	☐	☐	☐	☐	☐	☐	☐
Cabbage	☐	☐	☐	☐	☐	☐	☐	☐	☐
Peas	☐	☐	☐	☐	☐	☐	☐	☐	☐
Green beans, broad beans, runner beans	☐	☐	☐	☐	☐	☐	☐	☐	☐

	AVERAGE USE IN THE LAST TWELVEMONTHS								
Baked beans	☐	☐	☐	☐	☐	☐	☐	☐	☐
Marrow, courgettes	☐	☐	☐	☐	☐	☐	☐	☐	☐
Cauliflower	☐	☐	☐	☐	☐	☐	☐	☐	☐
Parsnips, turnips, swedes	☐	☐	☐	☐	☐	☐	☐	☐	☐
Leeks	☐	☐	☐	☐	☐	☐	☐	☐	☐
Onions	☐	☐	☐	☐	☐	☐	☐	☐	☐
Garlic (clove)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Mushrooms	☐	☐	☐	☐	☐	☐	☐	☐	☐
Sweet peppers	☐	☐	☐	☐	☐	☐	☐	☐	☐
Beansprouts	☐	☐	☐	☐	☐	☐	☐	☐	☐

	AVERAGE USE IN THE LAST TWELVEMONTHS								
Green salad, lettuce, cucumber, celery	☐	☐	☐	☐	☐	☐	☐	☐	☐
Watercress	☐	☐	☐	☐	☐	☐	☐	☐	☐
Tomatoes	☐	☐	☐	☐	☐	☐	☐	☐	☐
Sweetcorn	☐	☐	☐	☐	☐	☐	☐	☐	☐
Beetroot	☐	☐	☐	☐	☐	☐	☐	☐	☐
Coleslaw	☐	☐	☐	☐	☐	☐	☐	☐	☐
Avocado	☐	☐	☐	☐	☐	☐	☐	☐	☐
Pulses e.g. lentils, beans, peas	☐	☐	☐	☐	☐	☐	☐	☐	☐
Meat substitutes e.g. tofu, soyameat, textured vegetable protein, vegeburger	☐	☐	☐	☐	☐	☐	☐	☐	☐

FOODS AND AMOUNTS **AVERAGE USE IN THE LAST TWELVEMONTHS**

MEAT AND FISH

(half a small side plate)	Never or less than once a month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Beef: roast, steak, mince, stew or casserole	☐	☐	☐	☐	☐	☐	☐	☐	☐
Beefburgers	☐	☐	☐	☐	☐	☐	☐	☐	☐
Pork: roast, chops or stew	☐	☐	☐	☐	☐	☐	☐	☐	☐
Lamb: roast, chops or stew	☐	☐	☐	☐	☐	☐	☐	☐	☐
Chicken or other poultry e.g. turkey	☐	☐	☐	☐	☐	☐	☐	☐	☐
Bacon or gammon	☐	☐	☐	☐	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐	☐	☐	☐	☐

FOODS AND AMOUNTS	AVERAGE USE IN THE LAST TWELVEMONTHS								
Ham, cured meats & chorizo									
Corned beef, Spam, luncheon meats	☐	☐	☐	☐	☐	☐	☐	☐	☐
Sausages	☐	☐	☐	☐	☐	☐	☐	☐	☐
Savoury pies, e.g. meat pie, pork pie, pastries, steak & kidney pie, sausage rolls	☐	☐	☐	☐	☐	☐	☐	☐	☐
Liver, liver pate, liver sausage	☐	☐	☐	☐	☐	☐	☐	☐	☐
Fried fish in batter, as fish and chips	☐	☐	☐	☐	☐	☐	☐	☐	☐
Fish fingers, fish cakes & breaded fish	☐	☐	☐	☐	☐	☐	☐	☐	☐
Other white fish, fresh or frozen, e.g. cod, plaice, sole, haddock, halibut	☐	☐	☐	☐	☐	☐	☐	☐	☐
Oily fish, fresh or canned, e.g. tuna, mackerel, kippers, salmon, sardines, herring	☐	☐	☐	☐	☐	☐	☐	☐	☐
Shellfish, e.g. crab, prawns, mussels	☐	☐	☐	☐	☐	☐	☐	☐	☐

FOODS AND AMOUNTS	AVERAGE USE IN THE LAST TWELVEMONTHS								
Fish roe, taramasalata	☐	☐	☐	☐	☐	☐	☐	☐	☐

FOODS AND AMOUNTS **AVERAGE USE IN THE LAST TWELVEMONTHS**

BREAD AND SAVOURY BISCUITS

(One slice, or roll, or biscuit)	Never or less than once a	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
White bread/rolls	☐	☐	☐	☐	☐	☐	☐	☐	☐
Brown bread/rolls	☐	☐	☐	☐	☐	☐	☐	☐	☐
Wholemeal & granary bread/rolls	☐	☐	☐	☐	☐	☐	☐	☐	☐
Cream crackers, savoury biscuits	☐	☐	☐	☐	☐	☐	☐	☐	☐
Crispbread, e.g. Ryvita	☐	☐	☐	☐	☐	☐	☐	☐	☐
Naan, poppadoms, flour tortillas	☐	☐	☐	☐	☐	☐	☐	☐	☐

FOODS AND AMOUNTS **AVERAGE USE IN THE LAST TWELVEMONTHS**

SOUPS, SAUCES AND SPREADS

Vegetable soups (bowl)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Meat soups (bowl) (to include meat and vegetable soups)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Sauces, e.g. white sauce, cheese sauce, gravy (tablespoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Tomato ketchup (tablespoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Pickles, chutney (tablespoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Marmite, Bovril (teaspoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Jam, marmalade, honey (teaspoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐

FOODS AND AMOUNTS	AVERAGE USE IN THE LAST TWELVEMONTHS								
Peanut butter (teaspoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐

FOODS AND AMOUNTS **AVERAGE USE IN THE LAST TWELVEMONTHS**

SWEETS AND SNACKS (medium serving)	Never or less than once a month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Cakes e.g. fruit, sponge, home baked	☐	☐	☐	☐	☐	☐	☐	☐	☐
Cakes e.g. fruit, sponge, ready made	☐	☐	☐	☐	☐	☐	☐	☐	☐
Buns, pastries e.g. scones, flapjacks, croissants, doughnuts, home baked	☐	☐	☐	☐	☐	☐	☐	☐	☐
Buns, pastries e.g. scones, flapjacks, croissants, doughnuts, ready made	☐	☐	☐	☐	☐	☐	☐	☐	☐
Fruit pies, tarts, crumbles, home baked	☐	☐	☐	☐	☐	☐	☐	☐	☐
Fruit pies, tarts, crumbles, ready made	☐	☐	☐	☐	☐	☐	☐	☐	☐
Sponge puddings, home baked	☐	☐	☐	☐	☐	☐	☐	☐	☐

Sponge puddings, ready made	☐	☐	☐	☐	☐	☐	☐	☐	☐
Milk puddings e.g. rice, custard, trifle	☐	☐	☐	☐	☐	☐	☐	☐	☐
Ice cream, choc ices	☐	☐	☐	☐	☐	☐	☐	☐	☐
Sweet biscuits, chocolate, e.g. digestive (one)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Sweet biscuits, plain, e.g. Nice, ginger (one)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Reduced fat biscuits e.g. Go Ahead, Highlights (one small packet or one small bar/biscuit)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Cereal bars (one)	☐	☐	☐	☐	☐	☐	☐	☐	☐
White or milk chocolates, single or squares (one)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Dark chocolates, single or squares (one)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Chocolate snacks bars e.g. Mars, Crunchie (one)	☐	☐	☐	☐	☐	☐	☐	☐	☐

Sweets, toffees, mints (small packet)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Sugar added to tea, coffee, cereal (teaspoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Crips or other packet snacks, e.g. Wotsits (one packet)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Salted nuts, e.g. peanuts, cashews (handful)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Unsalted nuts, e.g. brazil, walnuts (handful)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Seed e.g. Sunflower, pumpkin (tablespoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐

FOODS AND AMOUNTS **AVERAGE USE IN THE LAST TWELVEMONTHS**

CEREAL (one bowl)	Never or less than once a month	1-3 per month		Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Porridge, Readybreak, oats	☐	☐	☐	☐	☐	☐	☐	☐	☐	
Breakfast cereal e.g. Cornflakes, Rice Krispies	☐	☐	☐	☐	☐	☐	☐	☐	☐	
Sugar topped cereals e.g. Frosties	☐	☐	☐	☐	☐	☐	☐	☐	☐	
Muesli	☐	☐	☐	☐	☐	☐	☐	☐	☐	
High Fibre cereals e.g. Branflakes, All Bran, Fruit and Fibre	☐	☐	☐	☐	☐	☐	☐	☐	☐	

POTATOES, RICE AND PASTA

(one small side plate)

Boiled, mashed, instant or one jacket potato	☐	☐	☐	☐	☐	☐	☐	☐	☐
Chips, roast potatoes	☐	☐	☐	☐	☐	☐	☐	☐	☐
Potato salad	☐	☐	☐	☐	☐	☐	☐	☐	☐
White rice	☐	☐	☐	☐	☐	☐	☐	☐	☐
Brown rice	☐	☐	☐	☐	☐	☐	☐	☐	☐
White or green pasta, e.g. spaghetti, macaroni, noodles	☐	☐	☐	☐	☐	☐	☐	☐	☐
Wholemeal pasta	☐	☐	☐	☐	☐	☐	☐	☐	☐
Lasagne, moussaka	☐	☐	☐	☐	☐	☐	☐	☐	☐
Pizza (one slice)	☐	☐	☐	☐	☐	☐	☐	☐	☐

Study No: «STUDY_NO» DOB: «DATE_BORN» Visit date: «VISIT_DATE»

FOODS AND AMOUNTS **AVERAGE USE IN THE LAST TWELVEMONTHS**

DRINKS	Never or less than once a month	1-3 per month	Once a week	2-4 per week	5-6 per week	Once a day	2-3 per day	4-5 per day	6+ per day
Tea (cup)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Green tea (cup)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Fruit tea (cup)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Coffee, instant or ground (cup)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Coffee, decaffeinated (cup)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Coffee whitener, e.g. Coffee-mate (teaspoon)	☐	☐	☐	☐	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐	☐	☐	☐	☐

Cocoa, hot chocolate (cup)									
Low fat hot chocolate (cup)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Horlicks, Ovaltine (cup)	☐	☐	☐	☐	☐	☐	☐	☐	☐
White wine (small glass)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Red wine (small glass)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Beer, lager or cider (half pint)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Port, sherry, vermouth, liqueurs (pub measure)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Spirits, e.g. gin, brandy, whisky, vodka (pub measure)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Low calorie or diet fizzy soft drinks (cup)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Fizzy soft drinks, e.g. Coca Cola, lemonade (cup)	☐	☐	☐	☐	☐	☐	☐	☐	☐

Pure fruit juice (100%) e.g. orange, apple juice (cup)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Fruit squash or cordial (cup)	☐	☐	☐	☐	☐	☐	☐	☐	☐
Smoothies (cup)	☐	☐	☐	☐	☐	☐	☐	☐	☐

Study No: «STUDY_NO» DOB: «DATE_BORN» Visit date: «VISIT_DATE»

YOUR DIET LAST YEAR, continued

Are there any **OTHER** foods which you ate more than once a week over the last year? No ☐ ⟹ **GO TO** question 3 Yes ☐ **If yes**, please list below

Food	Usual serving size	Number of times eaten each week

Have you drunk milk or a milk substitute over the past year? (including in hot drinks) No ☐ ⟹ **GO TO** question 6 Yes ☐ **GO TO** question 4

What type of milk or milk substitute did you <u>most often use</u> over the last year?

Select one only

Full cream ☐ Semi-skimmed ☐

Skimmed ☐ Channel Islands ☐

Dried milk ☐ Soya ☐

Other (please specify) _____

About how much milk did you drink each day, including milk with tea, coffee, cereals etc?

Less than a quarter of pint ☐ One pint ☐

Half a pint ☐ More than one pint ☐

Three quarters of a pint ☐

Did you usually eat breakfast cereal in the last year (excluding Porridge and Ready Brek mentioned earlier)?

If yes, which brand and type of breakfast cereal, including muesli, did you usually eat?

No ☐ ⟹ **GO TO** question 7 Yes ☐

List the one or two types most often used

Brand e.g. Kellogg's Type e.g. cornflakes

Do you use oil or fat for frying, roasting, grilling etc?

No ☐ ⟹ **GO TO** question 9 Yes ☐ ⟹ **GO TO** question8

What kind of fat did you most often use for frying, roasting, grilling etc? **Select one only**

Butter	☐	Olive oil	☐	Margarine	☐	Vegetable oil	☐
Lard/dripping	☐	Solid vegetable fat	☐	Low fat spray	☐	Other	☐

If you used vegetable oil or other oil

Please give type e.g. corn, sunflower, sesame

_____

What kind of fat did you most often use for baking cakes etc? **Select only one**

I don't bake	☐	Lard/dripping	☐	Solid vegetable fat	☐	None	☐
Butter	☐	Vegetable oil	☐	Margarine	☐		

If you used margarine, please give name or type e.g. Flora, Stork_____

How often did you eat food that was fried at home?

Never	☐	Less than once a week	☐	1-3 times a week	☐
4-6 times a week	☐	Daily	☐		

How often did you eat fried food away from home?

Never	☐	Less than once a week	☐	1-3 times a week	☐
4-6 times a week	☐	Daily	☐		

Have you eaten meat in the last year?

No ☐ ⟹ **GO TO** question 17 Yes ☐ ⟹ **GO TO** question 13

What did you do with the visible fat on your meat?

Did not eat meat with visible fat ☐ Ate some of the fat ☐

Ate as little as possible ☐ Ate most of the fat ☐

Did you eat roast/grilled meat over the past year?

No ☐ ⟹ **GO TO** question 17 Yes ☐ ⟹ **GO TO** question 15

How often did you eat grilled or roast meat?

Daily	☐	1-3 times a week	☐
4-6 times a week	☐	Less than once a week	☐

How well cooked did you usually have grilled or roast meat?

Well done/dark brown	☐	Lightly cooked/rare	☐
Medium	☐		

How often did you add salt to food while cooking?

Never	☐	Sometimes	☐	Always	☐
Rarely	☐	Usually	☐		

Study No: «STUDY_NO» DOB: «DATE_BORN» Visit date: «VISIT_DATE»

How often did you add salt to any food at the table?

Never ☐ Sometimes ☐ Always ☐

Rarely ☐ Usually ☐

Do you regularly use a salt substitute (e.g. LoSalt)?

No ☐ Yes ☐ **If yes**, which brand?

We would like to summarise what you have told us. During the course of last year, on average, how many times a week did you eat the following foods?

Food type	**Times/week**	**Portion size**
Vegetables (excluding potatoes but including baked beans)		Medium serving

Salads		Medium serving
Fruit and fruit products (not including fruit juice)		Medium serving or 1 fruit
Oily fish and fish products		Medium serving
Non oily fish and fish products		Medium serving
Meat, meat products and meat dishes (including bacon, ham and chicken)		Medium serving

Have you ever taken any vitamins, minerals, fish oils, fibre or other supplements during the past year?

No ☐ ⟹ **GO TO PAGE** 36 Yes ☐ ⟹ **GO TO** question **2**

Please list any vitamins, minerals, fish oils, fibre or other food supplements taken during the past year. If you have taken more than 6 types of supplement please list the 6 most frequently consumed brands first.

Vitamin Supplements

Average Frequency

Cross **one** box per line to show how often on average you consumed supplements

Name and brand	**Number**	Never or less than once a month	1-3 per month Once a week	2-4 per week	5-6 per week	Once a d ay	2-3 per day	4-5 per day	6+ per day
Please list full name and **brand**	Please state the ~~number of~~ pills, capsules or teaspoons consumed								
		☐	☐	☐	☐	☐	☐	☐	☐
	1-9								

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

1-9

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

1-9

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

1-9

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

1-9

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

1-9

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

1-9

Thank You

We really appreciate the time and effort it has taken for you to answer this questionnaire.

Your answers will help us to discover the important genetic and environmental causes of complex diseases and ageing and how it relates to our diet and lifestyle.

Please bring this completed questionnaire with you for your

Many thanks for your ongoing help and support.



TwinsUK staff at the 21st anniversary party

Keep in touch:

TwinsUK		www.twinsuk.ac.uk/
Facebook		www.facebook.com/twinsuk
YouTube		www.youtube.com/user/DeptTwinResearch
Twitter		twitter.com/twinsukres

Contact Info:

Department of Twin Research & Genetic Epidemiology

King's College London

St Thomas' Hospital Campus

3rd & 4th Floor South Wing Block D Westminster Bridge Road

London SE1 7EH

Tel for twins/general public: +44 (0) 20 7188 5555

Tel for scientific community: +44 (0) 20 7188 6765



SECTION 1: The Way You Feel

1. The following questions are designed to help us to know how you feel. Read each item below and place tick ✓ the box next to the reply which comes closest to how you have been feeling in the past week. Don't take too long over your replies. Your immediate reaction to each item will probably be more accurate than a long, thought-out response.

a. I feel tense or 'wound up'
☐ most of the time
☐ a lot of the time
☐ from time to time, occasionally
☐ not at all

b. I still enjoy the things I used to enjoy
☐ definitely as much
☐ not quite so much
☐ only a little
☐ hardly at all

c. I get a sort of frightened feeling as if something awful is about to happen
☐ very definitely and quite badly
☐ yes, but not too badly
☐ a little, but it doesn't worry me
☐ not at all

d. I can laugh and see the funny side of things
☐ as much as I always could
☐ not quite so much now
☐ definitely not so much now
☐ not at all

e. Worrying thoughts go through my mind
☐ a great deal of the time
☐ a lot of the time
☐ not too often
☐ very little

f. I feel cheerful
☐ never
☐ not often
☐ sometimes
☐ most of the time

g. I can sit at ease and feel relaxed
☐ definitely
☐ usually
☐ not often
☐ not at all

h. I feel as if I am slowed down
☐ nearly all the time
☐ very often
☐ sometimes
☐ not at all

i. I get a sort of frightened feeling like 'butterflies' in the stomach
☐ not at all
☐ occasionally
☐ quite often

☐ very often

j. I have lost interest in my appearance
☐ definitely

☐ I don't take as much care as I should

☐ I may not take quite as much care

☐ I take just as much care as ever

k. I feel restless as if I have to be on the move
☐ very much indeed

☐ quite a lot

☐ not very much

☐ not at all

l. I look forward with enjoyment to things
☐ as much as I ever did

☐ rather less than I used to

☐ definitely less than I used to

☐ hardly at all

m. I get sudden feelings of panic
☐ very often indeed

☐ quite often

☐ not very often

☐ not at all

n. I can enjoy a good book or radio or television programme
☐ often

☐ sometimes

☐ not often

☐ very seldom

SECTION 2: About your sleep quality

The HADS Anxiety & depression questionnaire

2. How long did it usually take you to fall asleep during the past 4 weeks?

0-15 minutes	16-30 minutes	31-45 minutes	46-60 minutes	More than 60 minutes
☐	☐	☐	☐	☐

3. On average, how many hours did you sleep each night during the past 4 weeks?

Write in number of hours per night ☐☐

4. **How often in the past 4 weeks did you…** *(Circle one number on each line)*

	All of the Time	*Most of the Time*	*A Good Bit of the Time*	*Some of the Time*	*A Little of the Time*	*None of the Time*
Feel that your sleep was not quiet (moving restlessly, feeling tense, speaking, etc., while sleeping)?	1	2	3	4	5	6
Get enough sleep to feel rested upon waking in the morning?	1	2	3	4	5	6
Awaken short of breathe or with a headache?	1	2	3	4	5	6
Feel drowsy or sleepy during the day?	1	2	3	4	5	6
Have trouble falling asleep?	1	2	3	4	5	6
Awaken during your sleep time and have trouble falling asleep again?	1	2	3	4	5	6
Have trouble staying awake during the day?	1	2	3	4	5	6
Snore during your sleep?	1	2	3	4	5	6
Taking naps (5 minutes or longer) during the day?	1	2	3	4	5	6

SECTION 3: About your Quality of Life

Medical outcomes survey sleep questionnaire

The next three questions concern your bodily symptoms (for example, pain, aches, pressure in your body, breathing difficulties, tiredness, etc.).

5. Does your pain or other bodily symptoms stop you from working (including both work outside the home and your housework)?

☐ No ☐ Rarely ☐ Sometimes ☐ Often ☐ Most of the time

6. Does your pain or other bodily symptoms stop you from concentrating on what you are doing?

☐ No ☐ Rarely ☐ Sometimes ☐ Often ☐ Most of the time

7. Does your pain or other bodily symptoms stop you from enjoying yourself?

☐ No ☐ Rarely ☐ Sometimes ☐ Often ☐ Most of the time

8. In general, would you say your health currently is:

☐ Excellent ☐ Very good ☐ Good ☐ Fair ☐ Poor

9. The following questions are about activities you may do during a typical day. **Would your health limit you** in these activities? If so, how much?

	Yes, limited a lot	Yes, limited a little	Not, limited at all
Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling or playing golf	☐	☐	☐

Climbing **several** flights of stairs ☐ ☐ ☐

SECTION 4: About Pain Anywhere in your Body

10. During the **past month,** have you had any of the following problems with work or other regular daily activities **as a result of your physical health?**

	Yes	No
Accomplished less than you would like	☐	☐
Were **limited** in the **kind** of work or other activities	☐	☐

11. During the **past month,** have you had any of the following problems with your work or other regular daily activities **as a result of any emotional problems** (e.g. feeling depressed or anxious)?

	Yes	No
Accomplished less than you would like	☐	☐
Didn't do work or other activities as **carefully** as usual	☐	☐

12. During the **past month** how much did **pain** interfere with your normal work (including both work outside the home and housework)?

Not at all	Slightly	Moderately	Quite a bit	Extremely
☐	☐	☐	☐	☐

13. These next questions are about how you feel and how things have been with you **during the past month.** Please give the **one** answer that is closest to the way you have been feeling for **each** item.

How much time **during the last month…**

	None of the time	A little of the time	Some of the time	A good bit of the time	Most of the time	All of the time
Have you felt calm and peaceful?	☐	☐	☐	☐	☐	☐
Did you have a lot of energy?	☐	☐	☐	☐	☐	☐
Have you felt downhearted and blue?	☐	☐	☐	☐	☐	☐

14. During the **past month**, how much of the time has your **physical health or emotional problems** interfered with your social activities (like visiting friends, relatives etc).

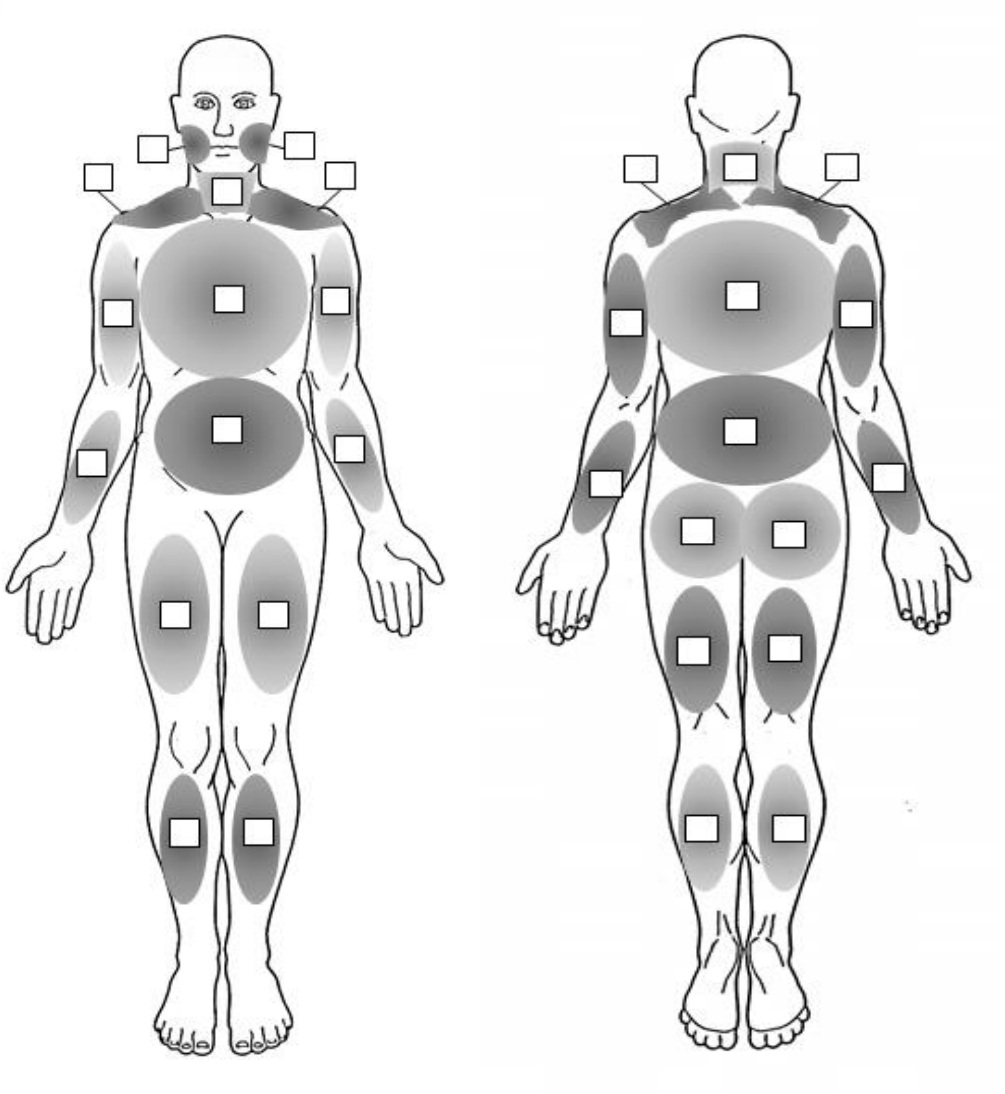
None of the time	A little of the time	Some of the time	A good bit of the time	Most of the time	All of the time
☐	☐	☐	☐	☐	☐

15. **In the past month**, did you feel tired on most days?

☐ Never ☐ Seldom ☐ Sometimes ☐ Often ☐ Always

16. This question is about **recent pain** you may have had in **any part of your body**. In the diagrams below, please tick the boxes where you have experienced pain in the region indicated by the shaded areas for **most days in the previous month**. Please tick all relevant boxes (the boxes are either placed in the area itself or joined to the area by a line). By pain we also mean aching, discomfort and/or stiffness. Please **do not**

include pain due to feverish illness such as flu. If you **do not have any body pain** that has lasted one day or longer in the last 4 weeks, please tick this box and move to **SECTION 5.**



17. How much physical pain have you had during the **past 4 weeks**? *(please tick one)*

None	Very mild	Mild	Moderate	Severe	Very Severe
☐	☐	☐	☐	☐	☐

18. If you answered “moderate” or “severe” or “very severe”, has this pain lasted **more than 3 months?**

☐ Yes ☐ No

19. During the **last 7 days**, on how many days did you do **vigorous** physical activities like heavy lifting, digging, aerobics, or fast bicycling?

_____ **days per week**

No vigorous physical activities ***Skip to question 21***

20. How much time did you usually spend doing **vigorous** physical activities on one of those days?

_____ **hours per day**

_____ **minutes per day**

☐ Don’t know/Not sure

Think about all the **moderate** activities that you did in the **last 7 days**. **Moderate** activities refer to activities that take moderate physical effort and make you breathe somewhat harder than normal. Think only about those physical activities that you did for at least 10 minutes at a time.

21. During the **last 7 days**, on how many days did you do **moderate** physical activities like

carrying light loads, bicycling at a regular pace, or doubles tennis? Do not include walking.

_____ **days per week**

No moderate physical activities ***Skip to question 23***

22. How much time did you usually spend doing **moderate** physical activities on one of those days?

_____ hours per day

_____ minutes per day

☐ Don't know/Not sure

Think about the time you spent **walking** in the **last 7 days**. This includes at work and at home, walking to travel from place to place, and any other walking that you have done solely for recreation, sport, exercise, or leisure.

23. During the **last 7 days**, on how many days did you **walk** for at least 10 minutes at a time?

_____ days per week

No walking ***Skip to question 25***

24. How much time did you usually spend **walking** on one of those days?

_____ hours per day

_____ **minutes per day**

☐ Don’t know/Not sure

The last question is about the time you spent **sitting** on weekdays during the **last 7 days**. Include time spent at work, at home, while doing course work and during leisure time. This may include time spent sitting at a desk, visiting friends, reading, or sitting or lying down to watch television.

25. During the **last 7 days**, how much time did you spend **sitting** on a **week day**?

_____ **hours per day**

_____ **minutes per day**

☐ Don’t know/Not sure

SECTION 5: About your hunger and appetite

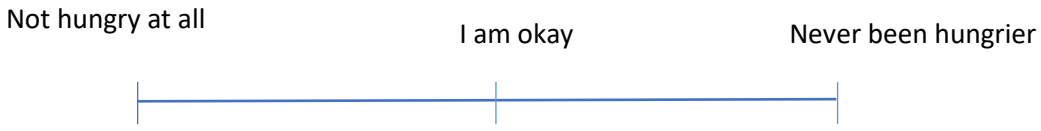
Appetite and desire for specific food types questionnaire

Flint et al 2000

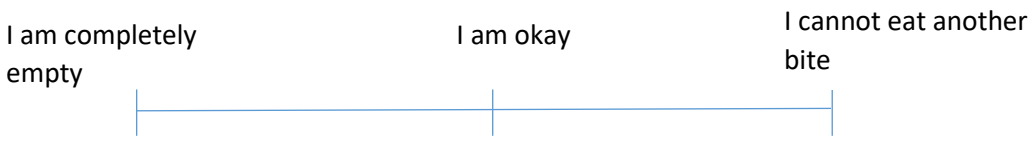
Throughout the questionnaire below you will use the scales below to rate your feelings of hunger, sateity, fullness, appetite and desire for specific food types at spefic times during the day.

The scale ranges from the most positive to the most negative feelings to assess hunger, sateity, fullness, prospective food consumtion and the desire to eat something sweet, salty, savoury and fatty. The scale consists of anchor words that are placed on either ends of the scale which represent either a strong positive or strong negative feeling with regards to each of the above mentioned variables. The words 'I am okay/ A moderate amount' placed in the middle of the scale are 'Nuetral' states which represents neither a strong positive nor a strong negative feeling towards any of the variables. You will be required to mark a vertical line with a pen at any one point along the horizontal lines using the anchor words as references to best match your feelings for each of the below mentioned questions. **Please note that all the questions below are madatory**.

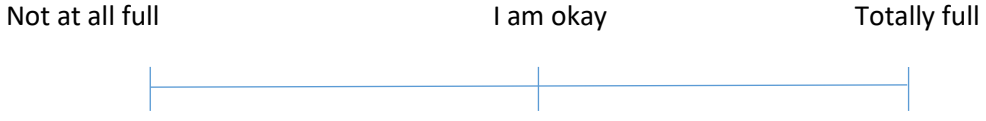
26. How hungry do you feel?



27. How satisfied do you feel?



28. How full do you feel?



29. How much do you think you can eat?

Nothing at all | A moderate amount | A lot



30. Would you like to eat something sweet?

Yes, very much | I am okay | No, not at all



31. Would you like to eat something salty?

Yes, very much | I am okay | No, not at all



32. Would you like to eat something savoury?

Yes, very much | I am okay | No, not at all



33. Would you like to eat something fatty?

Yes, very much | I am okay | No, not at all



This is the end of the questionnaires, thank you for participating.